南京信息工程大学学报
Journal of Nanjing University of Information Science & Technology
ISSN 1674-7070,CN 32-1801/N

《南京信息工程大学学报》网络首发论文

题目：基于神经网络的 VSLAM 综述
作者：尚光涛，陈炜峰，吉爱红，周铖君，王曦杨，徐崇辉
DOI：10.13878/j.cnki.jnuist.20220420001
收稿日期：2022-04-20
网络首发日期：2024-03-09
引用格式：尚光涛，陈炜峰，吉爱红，周铖君，王曦杨，徐崇辉．基于神经网络的 VSLAM 综述[J/OL]．南京信息工程大学学报．
https://doi.org/10.13878/j.cnki.jnuist.20220420001

基于神经网络的 VSLAM 综述

尚光涛[1]，陈炜峰[1]，吉爱红[2]，周铖君[1]，王曦杨[1]，徐崇辉[1]

1.南京信息工程大学 自动化学院，南京 210044

2.南京航空航天大学 机电学院/运动仿生与智能机器人实验室，南京 210016

摘要：传统的基于视觉的 SLAM 技术成果颇丰，但在具有挑战性的环境中难以取得想要的效果。深度学习推动了计算机视觉领域的快速发展，并在图像处理中展现出愈加突出的优势。将深度学习与基于视觉的 SLAM 结合是一个热门话题，诸多研究人员的努力使二者的广泛结合成为可能。本文从深度学习经典的神经网络入手，介绍了深度学习与传统基于视觉的 SLAM 算法的结合，概述了卷积神经网 (CNN)与循环神经网络（RNN)在深度估计、位姿估计、闭环检测等方面的成就，分析了神经网络在语义信息提取方面的优点，以期为未来自主移动机器人真正自主化提供帮助。最后，对未来 VSLAM 发展进行了展望。

关键词：同时定位和地图构建（SLAM）；深度学习；卷积神经网（CNN）；循环神经网络（RNN)；位姿估计；闭环检测；语义

中图分类号：TP242;TP391.41　　DOI:10.13878/j.cnki.jnuist.20220420001

A review of visual SLAM based on neural networks

SHANG Guangtao[1]　CHEN Weifeng[1]　JI Aihong[2]　ZHOU Chengjun[1]　WANG Xiyang[1]　XU Chonghui[1]

1 School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044

2 Lab of Locomotion Bioinspiration and Intelligent Robots/College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016

Abstract: Traditional vision-based SLAM technologies have achieved impressive results, but they are difficult to achieve in challenging environments. Deep learning promotes the rapid development of computer vision and shows more and more prominent advantages in image processing. Combining deep learning with vision-based SLAM is a hot topic, and the efforts of many researchers have made it possible for the two to be widely combined. Starting from the classical neural network of deep learning, this paper introduces the combination of deep learning and traditional vision-based SLAM algorithm. The achievements of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) in depth estimation, pose estimation and closed-loop detection are summarized. The advantages of neural network in semantic information extraction are pointed out, which provides help for the real autonomy of autonomous mobile robots in the future. The future development of VSLAM is also prospected.

Key words: simultaneous localization and mapping（SLAM）; deep learning; convolutional neural network (CNN); recurrent neural network (RNN); pose estimation; loop closure; semantic

收稿日期：2022-04-20

基金项目：国家自然科学基金（51875281）

作者简介：尚光涛，男，硕士生，研究方向为移动机器人、SLAM。20201222014@nuist.edu.cn

陈炜峰（通信作者），男，博士，教授，研究方向为机器人、机电控制。002021@nuist.edu.cn

0 引言

移动机器人执行任务的首要前提是要确定自己所在环境中的位置[1]。室外空旷环境下，基于 GPS 的定位方法可以基本满足机器人的定位需求，但是有时接收不到 GPS 信号[2]。室内环境中，通常需要提前设立导航信标如二维码、磁条等，这大大限制了移动机器人的应用范围[3]。大多数情况下，移动机器人需要自主完成某些任务，这就要求机器人可以适应足够陌生的环境。因此，能够在未知环境中进行定位和地图构建的 SLAM（Simultaneous Localization and Mapping）[4]技术成为自主移动机器人必备的能力。根据所使用的传感器不同，SLAM 技术主要分为激光 SLAM 与视觉 SLAM（VSLAM）[5]。与激光 SLAM 相比，VSLAM 与人眼类似，主要以图像作为环境感知信息源，更符合人类的认知。近年来，由于相机具有廉价、易安装、可以获得丰富的环境信息、易与其他传感器融合等优势[6]，基于相机的 VSLAM 研究受到了科研人员的广泛关注，大量以视觉为基础的 SLAM 算法应运而生[7]。

随着深度学习的快速发展，不少学者尝试采用深度学习的方法解决视觉 SLAM 所遇到的问题。深度学习可以根据具体问题学习更强大和有效的特征，并成功地展示了一些具有挑战性的认知和感知任务的良好能力。最近的工作尝试包括从单目图像中对场景进行深度估计、视觉里程计和语义映射生成等。权美香等[8]对传统的 VSLAM 进行了详细总结，并对比了不同方法的优缺点。胡凯等[9]从视觉里程计的角度，对 VSLAM 的发展做了概述，并介绍了深度学习在 VSLAM 中的应用。刘瑞军等[10]从里程计、闭环检测等方面介绍了深度学习与 VSLAM 的结合，并与传统方法进行了对比。李少朋等[11]将基于深度学习的 VSLAM 与传统的 VSLAM 进行了对比，并展望了未来发展方向。上述文献大多仅从深度学习角度讲述部分方法，未详细介绍典型神经网络与传统 VSLAM 的结合，也未将整个发展脉络完整展开。本文首先概述了 VSLAM 发展脉络，然后从深度学习的两个主要的神经网络，即卷积神经网络(Convolutional Neural Network, CNN)与循环神经网络（Recurrent Neural Network, RNN）入手，重点阐述了神经网络在 VSLAM 系统中深度估计、位姿估计、回环检测、以及数据融合等方面的贡献，并介绍了神经网络在语义信息提取方面的优势，最后对 VSLAM 的发展做出总结和展望。

本文具体结构如下：第 1 节介绍深度学习中两个典型的神经网络 CNN 和 RNN，并列举了部分优秀的 VSLAM 算法；第 2 节阐述了 CNN 与 VSLAM 的结合，并从单目深度估计、位姿估计、回环检测三个方面详细总结了 VSLAM 的发展进程；第 3 节重点介绍了 RNN 与视觉惯性数据融合方面的优势，并给出了神经网络与传统 VSLAM 结合的部分优秀方案；第 4 节为总结,并对未来 VSLAM 的发展做出了展望。

1 神经网络与 VSLAM 概述

传统的 VSLAM 研究已经取得了诸多令人惊叹的成就。2007 年，Davidson 等[12]提出了首个实时的单目 VSLAM 算法——MonoSLAM, 该算法可实现实时无漂移的运动结构恢复。2011 年，Newcombe 等[13]提出了 DTAM 算法，该算法被认为是第一个实际意义上的直接法 VSLAM。2015 年，Mur-Artal 等[14]提出了 ORB-SLAM 算法，创新地使用跟踪、局部建图和闭环检测三个线程同时进行，有效地降低了累计误差。闭环检测线程采用词袋模型 BoW[15]进行闭环的检测和修正，在处理速度和构建地图的精度上都取得了很好的效果。随后几年，Mur-Artal 团队相继推出了 ORB-SLAM2[16]与 ORB-SLAM3[17]。ORB-SLAM 系列是基于特征点提取方法中的佼佼者,它将传统 VSLAM 方法发展到了十分完善的程度。2016 年，Engel 等[18]提出了可以有效利用任何图像像素的 DSO 算法，它是直接法中的经典,其在无特征的区域中也具有良好的鲁棒性,并得到了广泛使用。2018 年，香港科技大学团队推出了单目惯性紧

耦合的 VINS-Mono[19]算法，该算法是视觉惯性融合 SLAM 中最优秀的算法之一，它充分利用惯性测量单元（Inertial Measurement Unit，IMU）与单目相机的互补性，改善了具有挑战性环境中的定位精度。表 1 根据前端所用传感器不同，从视觉里程计（Visual Odometry,VO）及视觉惯性里程计（Visual-Inertial Odometry,VIO）两方面列举了部分优秀的传统 VSLAM 方案，并给出了其开源地址。

传统方法多采用基于特征提取的间接法或者直接对像素进行操作的直接法。虽然在大多数环境中可以稳定运行，但是在光照强烈、相机快速旋转或是动态物体普遍存在等环境中鲁棒性会大大降低，甚至可能会失效。近年来，深度学习的快速发展，吸引了诸多学者的目光，将深度学习的方法与传统 VSLAM 相结合成为广受关注的研究领域[20]。

表 1　部分优秀的传统 VSLAM 算法

Table.1 Some excellent traditional vision-based SLAM algorithms

	方案	前端	后端	闭环检测	建图	开源地址
视觉里程计（VO）	MonoSLAM[12]	点特征	滤波	无	稀疏	https://github.com/rrg-polito/mono-slam
	PTAM[21]	点特征	优化	无	稀疏	https://github.com/Oxford-PTAM/PTAM-GPL
	ORB-SLAM2[16]	点特征	优化	有	稀疏	https://github.com/raulmur/ORB_SLAM2
	PL-SVO[22]	点线结合	优化	无	稀疏	https://github.com/rubengooj/pl-svo
	PL-SLAM[23]	点线结合	优化	有	稀疏	https://github.com/rubengooj/pl-slam
	DTAM[13]	直接法	优化	无	稠密	https://github.com/anuranbaka/OpenDTAM
	SVO[24]	混合法	优化	无	稀疏	https://github.com/uzh-rpg/rpg_svo
	LSD-SLAM[25]	直接法	优化	有	半稠密	https://github.com/tum-vision/lsd_slam
	DSO[18]	直接法	优化	无	稀疏	https://github.com/JakobEngel/dso
视觉惯性里程计（VIO）	MSCKF[26]	紧耦合	滤波	无	稀疏	https://github.com/daniilidisgroup/msckf_mono
	OKVIS[27]	紧耦合	优化	无	稀疏	https://github.com/ethz-asl/okvis
	ROVIO[28]	紧耦合	滤波	无	稀疏	https://github.com/ethz-asl/rovio
	VINS-Mono[19]	紧耦合	优化	有	稀疏	https://github.com/HKUST-AerialRobotics/VINS-Mono

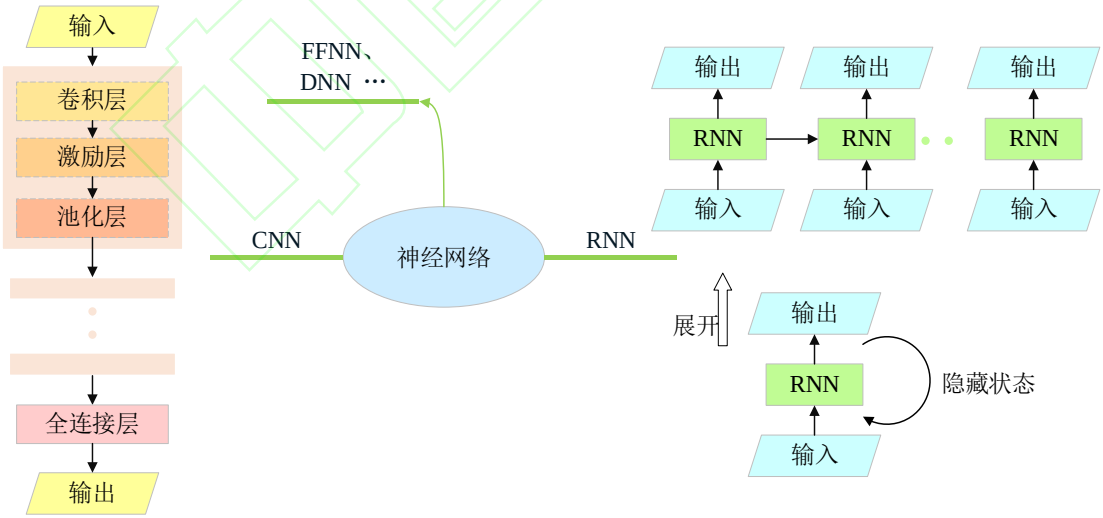

图 1 CNN 和 RNN 结构

Fig.1 CNN and RNN structure and their main performance comparison

深度学习可以学习不同数据中的特征或者是数据之间的某种关联，学习得到的特征属性与关联关系都可以用于不同的任务中[29]。深度学习通过层次化的处理方式，对视觉数据进行学习，得到数据的抽象表达，在图像识别、语义理解、图像匹配、三维重建[30]等任务中取得

了显著的成果[31]。作为深度学习中两个重要的神经网络，CNN 与 RNN 在多个领域取得了很高的成就，图 1 为 CNN 和 RNN 的基本框图，并在表 2 中给出了两者主要特点对比。CNN 可以从图像中捕捉空间特征,准确地识别物体以及它与图像中其他物体的关系[32]。RNN 可以有效地处理图像或数值数据，并且由于网络本身具有记忆能力，因此可以学习具有前后相关的数据类型[33]。此外，其他类型的神经网络如深度神经网络（Deep Neural Networks，DNN），在 VSLAM 领域也有一些尝试性的工作，但尚在起步阶段。如表 3 所示，结合深度学习进行 VSLAM 的研究已经有了许多突破性的进展。部分学者建议使用深度学习的方法替换传统 SLAM 的某些模块，如深度估计、回环检测、位姿估计等，从而改善传统方法。这些方法都取得了一定效果,在不同程度上提高了传统方法的性能。后文将从 CNN 和 RNN 两个神经网络入手，重点讲述它们与传统 VSLAM 的结合。

表 2 CNN 与 RNN 主要特点对比

Table.2　Comparison of main features of CNN and RNN

CNN	RNN
1）输入和输出的结果是固定的（接收固定尺寸的图像, 并将其输出到适当的类别）	1）输入和输出的结果是变化的（接收不同的文本并输出转换——结果句子可以包含更多或更少的单词）
2）理想的使用场景为图片	2）理想的使用场景为连续数据，如视频、文本等
3）可用于图像识别与分类、人脸检测、图像分析等	3）多用于文本转换、自然语言处理等

表 3 部分优秀的神经网络与 VSLAM 结合的算法

Table.3　Some algorithms of excellent neural networks combined with VSLAM

前端	方案	传感器	神经网络	监督方式	贡献
VO	CNN-SLAM[34]	单目	CNN	监督	只在关键帧上进行深度预测，提高了计算效率
	DeepVo[35]	单目	R-CNN	监督	使深度学习的方法在新环境下得到了广泛的应用
	Code-SLAM[36]	单目	U-Net	监督	通过姿态变量和代码对系统进行有效优化
	DVSO[37]	双目	DispNet	自监督	提出了自监督图像重建损失和稀疏深度预测
	UnDeepVo[38]	单目	VGG encoder-decoder	无监督	将深度学习的方法用于位姿和深度估计
	CNN-SVO[39]	单目	CNN	混合	在光照强烈的环境仍能稳定工作
	GANVO[40]	单目	GAN	无监督	通过对抗学习产生深度，避免了复杂的计算
	Li et al. [41]	单目	CNN	监督	提出简化了特征点和描述符的 CNN 结构
	D3VO[42]	单目	CNN	混合	提高了基于几何 VO 方法的性能
	DeepSeqSLAM[43]	单目	CNN+RNN	监督	提出了一个可训练的 CNN+RNN 架构
	DeepSLAM[44]	单目	RCNN	无监督	将深度学习与图优化相结合
	LIFT-SLAM[45]	单目	DNN	监督	消除了匹配阈值的固定值，从而无需微调数据集

	Zhang et al. [46]	双目	U-Net encoder-decoder	无监督	提高了经典的双目视觉算法的性能
VIO	VINet[47]	单目+IMU	CNN+LSTM	监督	将其视为一个序列到序列的回归问题
	VIOLearner[48]	单目+IMU	CNN	无监督	提出了在线校正模块
	DeepVIO[49]	双目+IMU	CNN+LSTM	监督	提出了一个从立体图像和 IMU 学习的 VIO 框架
	Chen et al. [50]	单目+IMU	FlowNet+LSTM	无监督	改进了传感器融合的特征选择
	Kim et al. [51]	单目+IMU	CNN+LSTM	无监督	克服了单传感器学习不确定性的局限性
	Gurturk et al. [52]	单目+IMU	CNN+LSTM	监督	提出的 VIO 框架优于 OKVIS 和 ORB-SLAM2

2 CNN 与 VSLAM

CNN 以一定的模型对事物进行特征提取，而后根据特征对该事物进行分类、识别、预测或决策等优点，可以对 VSLAM 的不同模块有所帮助。

2.1 单目深度估计

基于单目相机的 VSLAM 算法由于传感器成本低，简单实用，受到了诸多学者的喜爱。单目相机只能得到二维的平面图象，无法获得深度信息。简单地说，单目的局限性主要在于无法得到确定尺度[53] 。CNN 在图像处理方面的优势已得到充分验证，使用 CNN 进行视觉深度估计，最大程度上解决了单目相机无法得到可靠的深度信息的问题[54]。

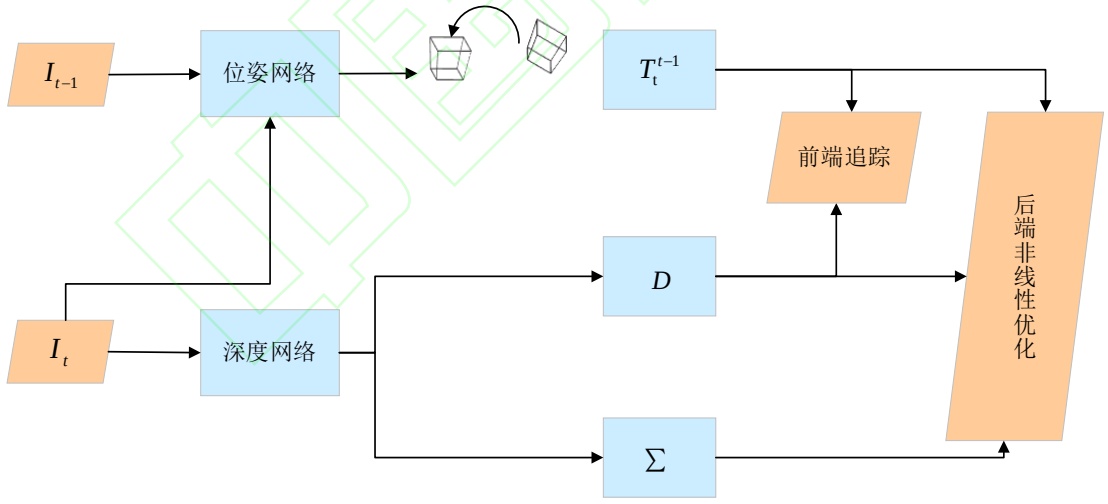

图 2　D3VO 核心流程[42]

Fig.2 The flow chart of D3VO. It utilizes three layers of deep neural networks[42]

2017 年，Tateno 等[34]在 LSD-SLAM 的框架上提出了基于 CNN 的实时 SLAM 算法 CNN-SLAM。该算法用 CNN 做深度预测将其输入到后续的传统位姿估计等模块，用来提升定位和建图精度。此外，该算法利用 CNN 提取环境的语义信息，进行全局地图和语义标签的融合，提高了机器人的环境感知能力。类似利用 CNN 预测深度信息的工作还有 Code-SLAM[36]以及 DVSO[37]等。但上述方法只在某个方面利用了 CNN 的优势，Yang 等[42]提出的 D3VO 则从三个层面利用了 CNN，包括利用深度学习进行深度估计、位姿估计以及

不确定度估计(Σ)。如图 2 所示，D3VO 将预测深度（D）、位姿(T_t^{t-1})以及不确定度紧密结合到一个直接视觉里程计中，来同时提升前端追踪以及后端非线性优化的性能。所提出的单目深度估计网络的核心是自监督训练体制，这种自监督训练是通过最小化时间立体图像和静态立体图像之间的光度重投影误差来实现的，原理如下：

$$L_{self} = \frac{1}{|V|}\sum_{P\in V}\min_{t'} r(I_t, I_{t'\to t}), \quad (1)$$

其中：V 是图片I_t上面所有像素的集合，文中将I_t设置为双目相机中左侧摄像头所得帧；t'是所有源帧的索引（区别于时刻 t 的某一时刻，右上角的′表示将其与 t 区分开）；$I_{t'}$为包含相邻时间的两帧以及右侧摄像头所得帧，即$I_{t'} \in \{I_{t-1}, I_{t+1}, I_t^s\}$（$I_{t-1}$为 t 时刻前一时刻左侧相机所得帧，I_{t+1}为时t刻时后一时刻左侧相机所得帧，I_t^s为双目相机中右侧摄像头所得帧）。

2.2 位姿估计

传统的位姿估计方法，一般采用基于特征的方法或直接法，通过多视图几何来确定相机位姿。但基于特征的方法需要复杂的特征提取和运算[55]，直接法则依赖于像素强度值，这使得传统方法在光照强烈或纹理稀疏等环境中很难取得想要的结果[56]。基于深度学习的方法由于无需提取环境特征，也无需进行特征匹配和复杂的几何运算，因此更加直观简洁[57]。Zhu 等[58]通过利用 CNN 关注光流输入的不同象限来学习旋转和平移，在数据集中测试结果比传统 SLAM 效果更好。表 4 给出了在位姿估计方面传统方法与基于 CNN 方法的不同。由于 CNN 的特征检测层通过训练数据进行学习，所以在使用 CNN 时，避免了显示的特征抽取，而隐式地从训练数据中进行学习，文献[32, 59]在这方面做出了较为详细的总结。相比传统位姿估计方法，CNN 可以替代传统方法复杂的公式计算，无需提取和匹配特征，因此在线运算速度较快[60]。

表 4　CNN 用于 VSLAM 与传统方法对比

Table.4　CNN is used for VSLAM compared with traditional methods

位姿估计	方法	特点
传统方法	对极几何、PnP、ICP、LK 光流	几何特征只能为相机的姿势提供短期的限制, 而且可能在有强烈的光和快速运动的环境中失败，且复杂的特征提取相当耗时
基于 CNN 的方法	数据关联、高级信息提供帮助（如语义信息）	无需提取环境特征，也无需进行特征匹配和复杂的几何运算，当光照强度、观测距离和角度变化时，语义信息保持不变

2.3 闭环检测

闭环检测可以消除累积轨迹误差和地图误差，决定着整个系统的精度，其本质是场景识别问题[61]。在闭环检测方面，传统方法多以词袋模型为基础。如图 3 所示，首先需要从图像中提取出相互独立的视觉词汇，通常经过特征检测、特征表示以及单词本的生成三个步骤；然后再将从新采集到的图像进行词典匹配并分类，过程复杂。而深度学习的强大识别能力，可以提取图像更高层次的稳健特征如语义信息，使得系统能对视角、光照等图像变化具备更

强的适应能力，提高闭环图像识别能力[62]。因此，基于深度学习的场景识别可以提高闭环检测准确率，CNN 用于闭环检测也得到了诸多可靠的效果。

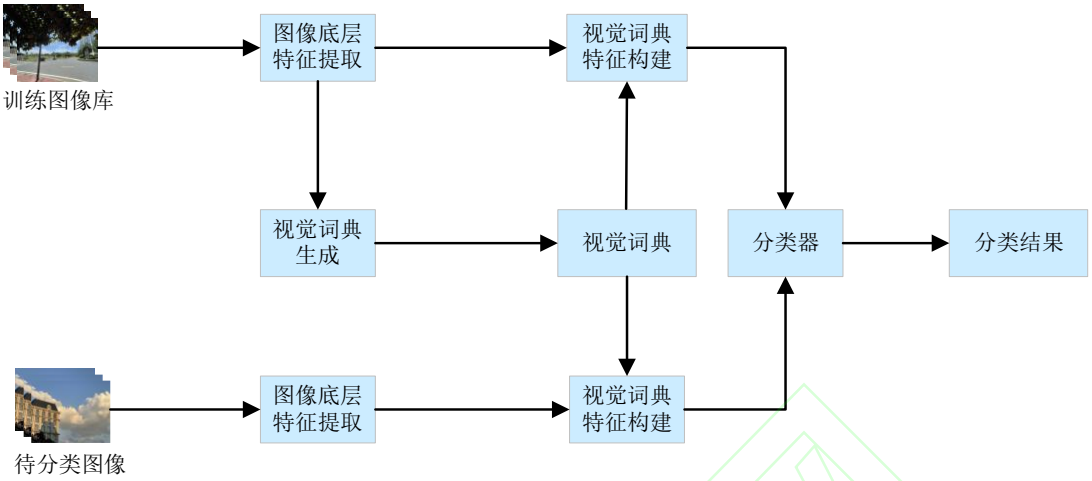

图 3　传统的基于视觉词袋模型的图像分类系统结构

Fig.3 Traditional structure of classification system based on visual word bag model

Memon 等[63]提出一种基于词典的深度学习方法，它不同于传统的 BoW 词典，创新地使用两个 CNN 网络一起工作，以加快闭环检测的速度，并忽略移动对象对闭环检测的影响。其核心如图 4 所示，该方法使用并行线程(标记为虚线框)使闭合检测可以达到更高的速度。将 patch 逐个送入移动对象识别层，从标记为静止的 patch 中提取 CNN 特征，由创新检测层进一步处理。所有不包含任何移动物体的 patch 再经过创新检测层处理来判断是否访问过该场景。在新的场景下，自动编码器在一个单独的线程上并行地训练这些特征。该方法可以鲁棒地执行循环闭环检测，比同类方法拥有更快的运行速度。Li 等[64]使用 CNN 从每帧图像中提取局部特征和全局特征，然后将这些特征输入现代 SLAM 模块，用于姿势跟踪、局部映射和重新定位。与传统的基于 BoW 的方法相比，它的计算效率更高，并且计算成本更低。Qin 等[65]采用 CNN 提取环境语义信息，并将视觉场景建模为语义子图。该方法只保留目标检测中的语义和几何信息，并在数据集中与传统方法进行了比较。结果表明，基于深度学习的特征表示方法，在不提取视觉特征的情况下，可以明显改善闭环检测的效果。

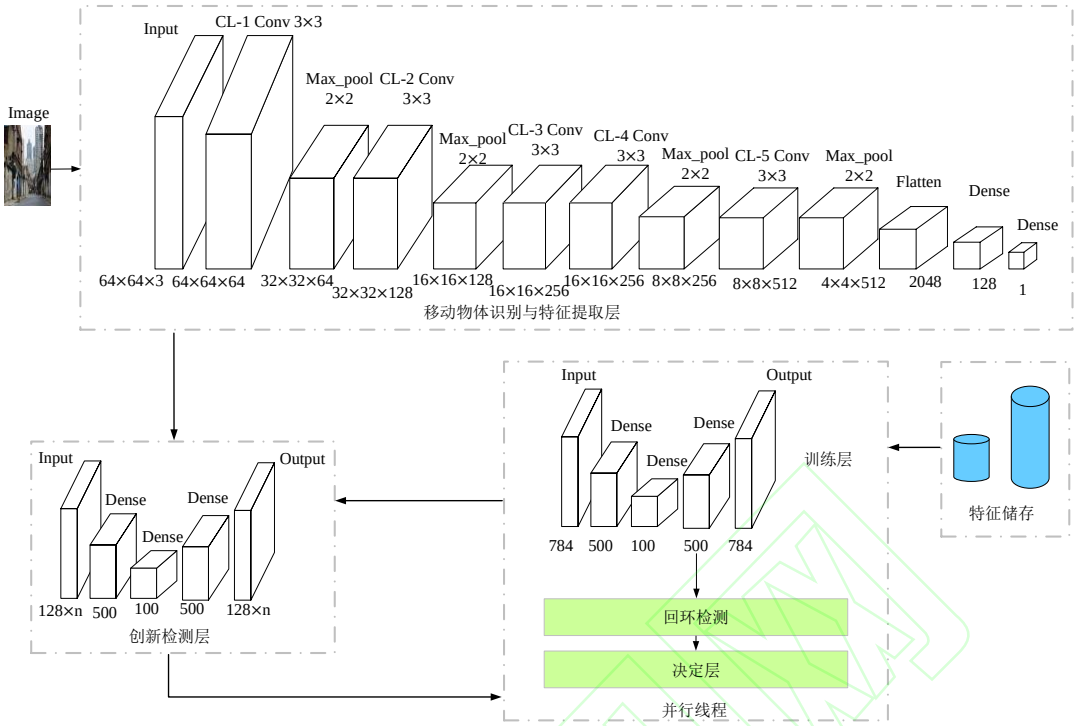

图 4　基于 CNN 的视觉词袋模型回环检测方法[63]

Fig.4 Closed-loop detection method of visual word bag model based on CNN[63]

上述内容主要从单目深度估计、位姿估计、回环检测三个方面举例了 CNN 与 VSLAM 的结合。表 5 给出了传统方法与结合深度学习方法的对比。CNN 在取代传统的特征提取环节取得了不错的效果，改善了传统特征提取环节消耗时间、消耗计算资源的缺点。其次，CNN 有效的提高了单目深度估计的精度。此外，文献[34,66]利用 CNN 提取环境的语义信息，以更高层次的特征来优化传统 VSLAM 的进程，使得传统 VSLAM 获得了更好的效果。采用神经网络提取语义信息，并与 VSLAM 结合将会是一个备受关注的领域，借助语义信息将数据关联从传统的像素级别提升到物体级别，将感知的几何环境信息赋以语义标签，进而得到高层次的语义地图，可帮助机器人进行自主环境理解和人机交互。

表 5　结合深度学习的 VSLAM 与传统方法对比

Tabel.5 Comparison between VSLAM combined with deep learning and traditional methods

环节	传统方法	结合深度学习的方法
单目深度估计	传统方法无法很好的解决单目尺度不确定性问题	CNN 可以在一些挑战性的环境中更有效地估计图像深度，如低纹理区域
相机位姿估计	通过特征提取与匹配，或是基于像素亮度变化，需要复杂的计算环节，并且在具有挑战性的环境中（低纹理区域、光照强烈、快速运动）无法得到可靠的效果	可以取代传统方法复杂的公式计算、特征提取与匹配，速度更快
闭环检测	本质是场景识别问题，传统方法多采用词袋模型。在场景光照变化大，相机视野变化大等环境中，传统的 DBoW 方法能力有限	闭环过程使用深度学习中的图像检索，能有效的减少由于环境光照、季节更替、视角变化引起的匹配问题

语义信息	传统的 VSLAM 算法中，基本不涉及高层次信息的提取，对于移动机器人的真正智能化没有帮助	采用深度学习的方法可以有效地提取环境中的语义信息，高层次的语义信息，可以给系统带来长期稳定的约束，并且使机器人更好的理解周围环境

3 RNN 与 VSLAM

循环神经网络 RNN 的研究从 20 世纪八九十年代开始，并在 21 世纪初发展为深度学习经典算法之一 ，其中长短期记忆网络（Long Short-Term Memory networks，LSTM）是最常见的循环神经网络之一。LSTM 是 RNN 的一种变体，它记忆可控数量的前期训练数据，或是以更适当的方式遗忘[67]。LSTM 基本结构如图 5 所示,从左到右依次为遗忘门、输入门、输出门。采用了特殊隐式单元的 LSTM 可以长期保存输入，LSTM 的这种结构继承了 RNN 模型的大部分特性，同时解决了梯度反传过程由于逐步缩减而产生的 Vanishing Gradient 问题。此外 GRU（Gate Recurrent Unit）相比 LSTM，更容易进行训练，能够很大程度上提高训练效率，因此很多时候会倾向于使用 GRU，但在 VSLAM 领域还只是尝试阶段。

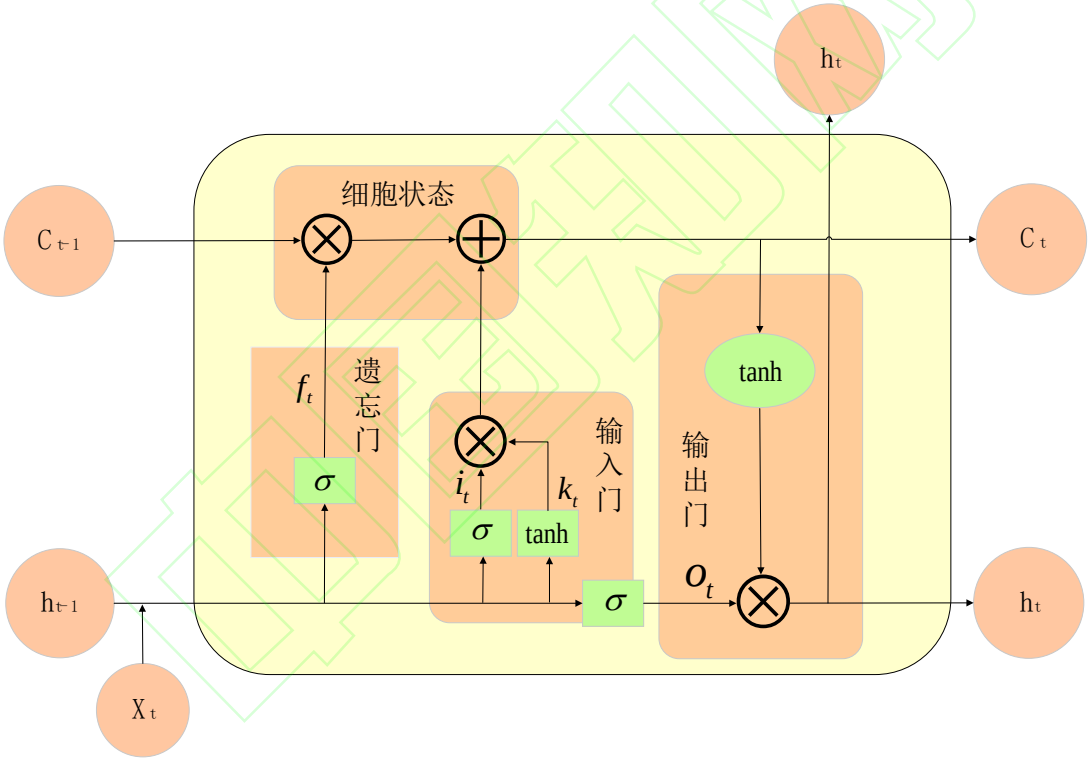

图 5　LSTM 基本框架包含输入门、输出门、遗忘门

Fig.5 The LSTM basic framework includes input gate， out gate, and forget gate

遗忘门状态方程为

$$f_t = \sigma(w_f * [h_{t-1}, x_t] + b_f) \quad 。 \tag{2}$$

输入门状态方程为

$$\begin{aligned} f_t &= \sigma(w_i * [h_{t-1}, x_t] + b_i), \\ k_t &= \tanh(w_k * [h_{t-1}, x_t] + b_k)。 \end{aligned} \tag{3}$$

输出门状态方程为

$$O_t = \sigma(w_o * [h_{t-1}, x_t] + b_0),$$
$$h_t = O_t * \tanh(C_t)。 \tag{4}$$

更新后的细胞状态为

$$C_t = f_t \otimes C_{t-1} + i_t \oplus k_t 。 \tag{5}$$

循环神经网络具有记忆性、参数共享，因此在对序列的非线性特征进行学习时具有一定优势， RNN 在帮助建立相邻帧之间的一致性方面具有很大的优势，高层特征具备更好的区分性 ，可以帮助机器人更好完成数据关联。

3.1 位姿估计

传统的位姿估计方法首先需要特征提取与匹配[68]，或是基于像素亮度变化的复杂计算。其原理如图 6 所示，该问题的核心是求解旋转矩阵和平移向量，需要繁琐的计算过程。基于特征的方法（图 6b）需要十分耗时地提取特征，计算描述子的操作丢失了除了特征点以外的很多信息（图 6a 中$\boldsymbol{R}, \boldsymbol{t}$分别为旋转矩阵和平移向量，红色点为空间中的特征点，黑色点为特征点在不同图像中的投影）。而直接法（图 7）不同于特征点法最小化重投影误差，而是通过最小化相邻帧之间的灰度误差估计相机运动，但是基于灰度不变假设：

$$I(x, y, z) = I(x + \Delta x, y + \Delta y, z + \Delta z) \tag{6}$$

如图 6b 假设空间点P在相邻两帧图像上的投影分别为P_1,P_2两点（用不同颜色的点表示其二者像素强度的差别）。它们的像素强度分别为$I_1(P_{1,i})$和$I_2(P_{2,i})$，其中i表示当前图像中第i个点。则优化目标就是这两点的亮度误差$\boldsymbol{e}_i$的二范数。

$$\min_{\xi} J(\xi) = \sum_{i=1}^{N} \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{e}_i, \tag{7}$$

$$e_i = I_1(p_{1,i}) - I_2(p_{2,i}), \tag{8}$$

$$p_{1,i} = \boldsymbol{T} p_{2,i}, \tag{9}$$

$$\boldsymbol{T} = \exp(\xi^{\wedge}). \tag{10}$$

其中 $\boldsymbol{T}$ 和 ξ 分别是P_1，P_2之间的转换矩阵及其李代数。式(6) ξ 右上角的∧表示把ξ转为一个四维矩阵，从而通过指数映射成为变换矩阵。

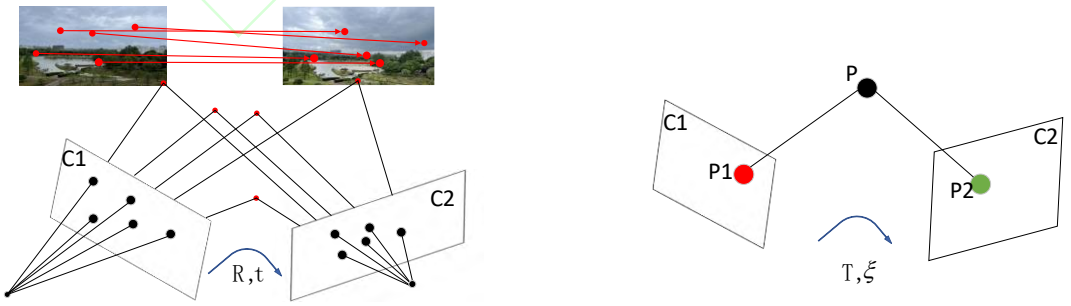

图 6 (a)基于特征法的提取与匹配; (b)基于直接法的像素强度变化计算

Fig.6 (a)Extraction and matching based on feature method (b)the calculation of pixel intensity change based on the direct method

通过引入端对端的深度学习方法，使得视觉图像帧间的位姿参数解算无须特征匹配与复杂的几何运算，可快速得到帧间相对位姿参数[69]。Xue 等[70]基于 RNN 来实现位姿的估计。

在位姿估计过程中，旋转和位移是分开进行训练的，相对于传统方法有更好的适应性。2017年，Wang 等[35]使用深度递归卷积神经网络，提出一种新颖的端到端单目 VO 的框架。由于它是以端到端的方式进行训练和配置的，因此可以直接从一系列原始的 RGB 图像中计算得到姿态，而无需采用任何传统 VO 框架中的模块。该方法做到了视觉里程计的端到端实现，免去了帧间各种几何关系的约束计算，有良好的泛化能力。如图 7 所示，该方案使用 CNN+RNN 对相机的运动进行估计，直接从原始 RGB 图像序列推断姿态。它不仅通过卷积神经网络自动学习 VO 问题的有效特征表示，而且还利用深度回归神经网络隐式建模顺序动力学和关系。

3.2 视觉惯性融合

由于惯性测量元件 IMU 能够在短时间内高频地获得精准的估计，减轻动态物体对相机的影响，而相机数据也能有效地修正 IMU 的累积漂移，IMU 被认为是与相机互补性最强的传感器之一[71]。传统方法中，视觉惯性融合按照是否将图像特征信息加入到状态向量中可以分为松耦合和紧耦合[72]。松耦合是指 IMU 和相机分别进行自身的运动估计，然后对其位姿估计输出结果进行融合[73]。紧耦合是指把 IMU 的状态与相机的状态合并在一起，共同构建运动方程和观测方程，然后进行状态估计[74]。图 8 为传统方法典型的视觉惯性融合流程，由于相机和 IMU 频率相差较大，需要先进行严格的同步校准；其次，不同传感器的数据融合，势必会带来计算资源消耗过多、实时性差等问题。

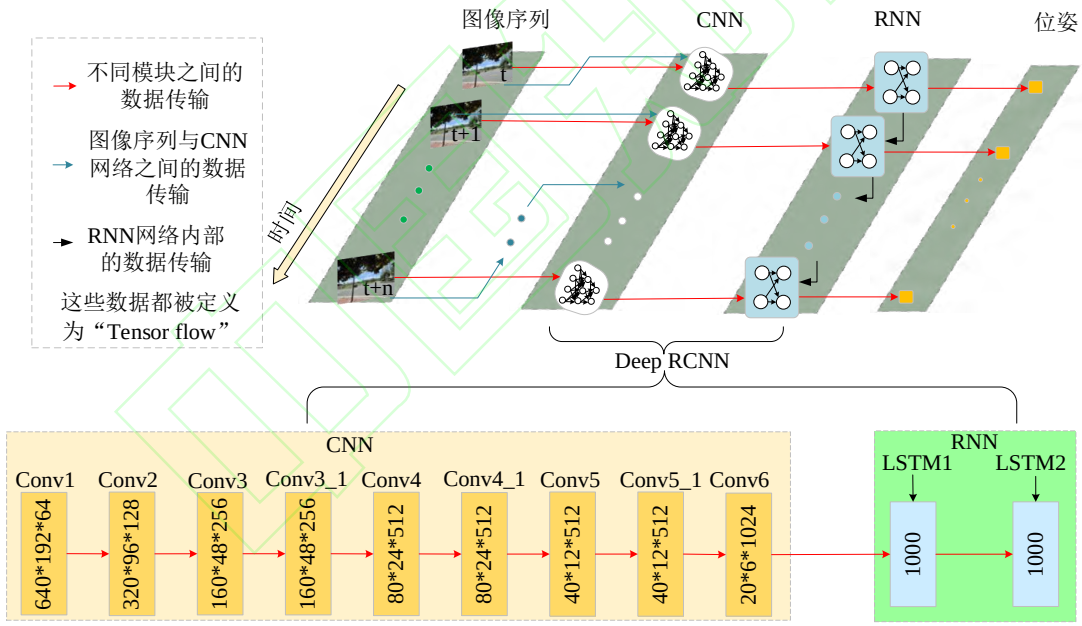

图 7　CNN 与 RNN 结合用来改善传统 VO[35]

Fig.7 CNN is combined with RNN to improve traditional VO[35]

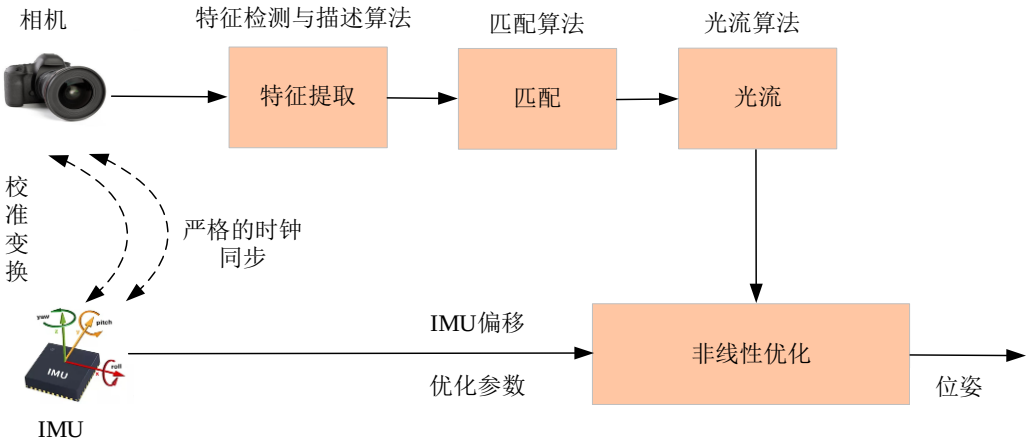

图 8　传统的视觉惯性融合流程

Fig.8 Traditional visual-inertial fusion process

RNN 是深度学习领域数据驱动的时序建模常用方法，IMU 输出的高帧率角速度、加速度等惯性数据，在时序上有着严格的依赖关系，特别适合使用 RNN 这类模型来优化。Clark 等[47]基于此，提出了使用一个常规的小型 LSTM 网络来处理 IMU 的原始数据，得到 IMU 数据下的运动特征。如图 9 所示，在对相机数据和 IMU 数据做一个结合后，送入一个核心的 LSTM 网络进行特征融合和位姿估计。该方法通过神经网络，避免了传统方法复杂的数据融合过程，使得运行效率大大提升。

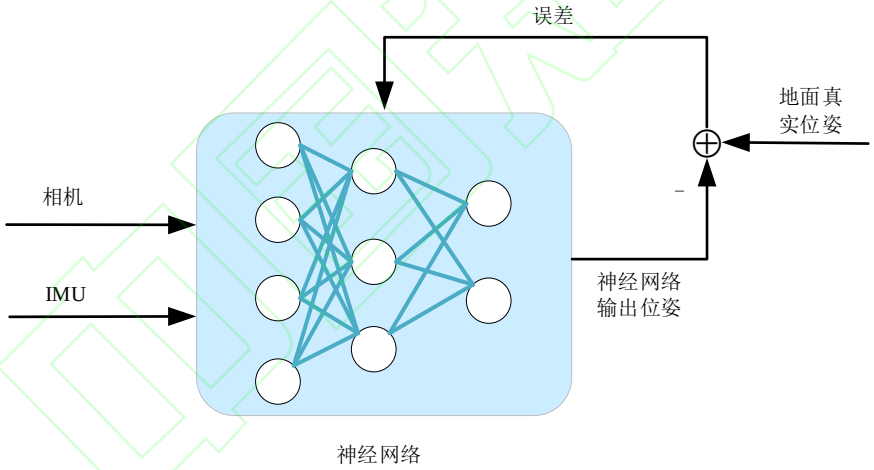

图 9 采用 CNN+LSTM 的神经网络结构，从数据中学习，避免了传统方法复杂的数据融合计算步骤

Fig.9 The neural network structure of CNN+LSTM is adopted to learn from data, which avoids the complicated data fusion calculation steps of traditional methods

与单纯用于位姿估计等方面相比， RNN 在视觉惯性数据融合方面做出的贡献更具吸引力。此类方法非常有效的对视觉惯性数据进行了融合，相比传统方法更便捷，类似的工作有文献[50-51]等。此外，一些工作利用神经网络提取环境中的语义信息，高层特征更具区分性，对于 VSLAM 数据关联有很好的帮助。2017 年，Yu 等[75]使用 RNN 与 KinectFusion 相结合，对 RGB-D 相机采集图像进行语义标注，用来重建三维语义地图。通过在 RNN 中引入了一个新的循环单元，来解决 GPU 计算资源消耗过大的问题。该方法充分利用 RNN 的优点，实现了语义信息的标注，高层特征具备更好的区分性 ，同时帮助机器人更好完成数据关联。

4 总结与展望

本文对深度学习中的两个典型神经网络 CNN 与 RNN 进行了介绍，并详细总结了神经网络在 VSLAM 中的贡献，从深度估计、位姿估计、闭环检测等方面将基于神经网络的方法与传统方法进行对比。从 CNN 与 RNN 各自的特点入手，列举出其对传统 VSLAM 不同模块的改善。并指出利用神经网络一定程度上改善了传统 VSLAM 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。

综合他人所作研究，本文认为未来 VSLAM 的发展趋势如下：

1）更高层次的环境感知。神经网络可以更加方便地提取环境中高层次的语义信息的优点，可以促进机器人智能化的发展。传统的 VSLAM 算法只能满足机器人基本的定位导航需求，无法完成更高级别的任务，如：“帮我把卧室门关上”、“去厨房帮我拿个苹果”等。借助语义信息将数据关联从传统的像素级别提升到物体级别，将感知的几何环境信息赋以语义标签，进而得到高层次的语义地图，可帮助机器人进行自主环境理解和人机交互，实现真正自主化。

2）更完善的理论支撑体系。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，目前深度学习多应用于 SLAM 局部的子模块，如深度估计、闭环检测等，而如何将深度学习应用贯穿于整个 SLAM 系统仍是一个巨大挑战。

3）更高效的数据融合。CNN 可以与 VLSAM 的诸多环节进行结合，如特征提取与匹配、深度估计、位姿估计等， RNN 的应用范围更小。 但 RNN 在数据融合方面的优势，可以更好的融合多传感器的数据，快速推动传感器融合 SLAM 技术的发展。未来可能会更多的关注 CNN 与 RNN 的结合，来提升 VSLAM 的整体性能。

参考文献：

[1] 任伟建，高强，康朝海，等. 移动机器人同步定位与建图技术综述[J]. 计算机测量与控制, 2022, 30(2): 1-10, 37

REN Weijian, GAO Qiang, KANG Chaohai, et al. Overview of simultaneous localization and mapping technology of mobile robots[J]. Computer Measurement ＆ Control, 2022, 30(2): 1-10, 37

[2] 赵乐文，任嘉倩，丁杨. 基于 GNSS 的空间环境参数反演平台及精度评估[J]. 南京信息工程大学学报(自然科学版), 2021, 13(2): 204-210

ZHAO Lewen, REN Jiaqian, DING Yang. Platform for GNSS real-time space environment parameter inversion and its accuracy evaluation[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2021, 13(2): 204-210

[3] 尹姝，陈元橡，仇翔. 基于 RFID 和自适应卡尔曼滤波的室内移动目标定位方法[J]. 南京信息工程大学学报(自然科学版), 2018, 10(6): 749-753

YIN Shu, CHEN Yuanyuan, QIU Xiang. Indoor moving-target localization using RFID and adaptive Kalman filter[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2018, 10(6): 749-753

[4] 周韦，孙宪坤，吴飞. 基于 SLAM/UWB 的室内融合定位算法研究[J]. 全球定位系统, 2022, 47(1): 36-42, 85

ZHOU Wei, SUN Xiankun, WU Fei. Research on indoor fusion positioning algorithm based on SLAM/UWB[J]. GNSS World of China, 2022, 47(1): 36-42, 85

[5] Bresson G, Alsayed Z, Yu L, et al. Simultaneous localization and mapping: a survey of current trends in autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2017, 2(3): 194-220

[6] 李晓飞，宋亚男，徐荣华，等. 基于双目视觉的船舶跟踪与定位[J]. 南京信息工程大学学报(自然科学版), 2015, 7(1): 46-52

LI Xiaofei, SONG Yanan, XU Ronghua, et al. Tracking and positioning of ship based on binocular vision[J]. Journal

of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2015, 7(1): 46-52

[7] 刘明芹，张晓光，徐桂云，等. 单机器人 SLAM 技术的发展及相关主流技术综述[J]. 计算机工程与应用, 2020, 56(18): 25-35

LIU Mingqin, ZHANG Xiaoguang, XU Guiyun, et al. Review of development of single robot SLAM technology and related mainstream technology[J]. Computer Engineering and Applications, 2020, 56(18): 25-35

[8] 权美香，朴松昊，李国. 视觉 SLAM 综述[J]. 智能系统学报, 2016(6):768-776

QUAN Meixiang, PIAO Songhao, LI Guo. An overview of visual SLAM[J]. CAAI Transactions on Intelligent Systems, 2016(6):768-776

[9] 胡凯，吴佳胜，郑翡，等. 视觉里程计研究综述[J]. 南京信息工程大学学报(自然科学版), 2021, 13(3): 269-280

HU Kai, WU Jiasheng, ZHENG Fei, et al. A survey of visual odometry[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2021, 13(3): 269-280

[10] 刘瑞军，王向上，张晨，等. 基于深度学习的视觉 SLAM 综述[J]. 系统仿真学报, 2020, 32(7): 1244-1256

LIU Ruijun, WANG Xiangshang, ZHANG Chen, et al. A survey on visual SLAM based on deep learning[J]. Journal of System Simulation, 2020, 32(7): 1244-1256

[11] 李少朋，张涛. 深度学习在视觉 SLAM 中应用综述[J]. 空间控制技术与应用, 2019, 45(2): 1-10

LI Shaopeng, ZHANG Tao. A survey of deep learning application in visual SLAM[J]. Aerospace Control and Application, 2019, 45(2): 1-10

[12] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067

[13] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C]//2011 International Conference on Computer Vision. Barcelona, Spain. IEEE,2011: 2320-2327

[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] Galvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197

[16] Mur-Artal R, Tardós J 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

[17] Campos C, Elvira R, Rodríguez J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual–inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890

[18] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625

[19] Qin T, Li P L, Shen S J. VINS-mono: a robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020

[20] 邓晨，李宏伟，张斌，等. 基于深度学习的语义 SLAM 关键帧图像处理[J]. 测绘学报, 2021, 50(11): 1605-1616

DENG Chen, LI Hongwei, ZHANG Bin, et al. Research on key frame image processing of semantic SLAM based on deep learning[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(11): 1605-1616

[21] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, Japan. IEEE,2007: 225-234

[22] Gomez-Ojeda R, Briales J, Gonzalez-Jimenez J. PL-SVO: Semi-direct Monocular Visual Odometry by combining points and line segments[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, Korea (South). IEEE, 2016: 4211-4216

[23] Pumarola A, Vakhitov A, Agudo A, et al. PL-SLAM: real-time monocular visual SLAM with points and

lines[C]//2017 IEEE International Conference on Robotics and Automation. Singapore. IEEE,2017 : 4503-4508

[24] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[C]//2014 IEEE International Conference on Robotics and Automation. Hong Kong, China. IEEE,2014: 15-22

[25] Engel J , Schps T , Cremers D . LSD-SLAM: large-scale direct monocular SLAM[C]// European Conference on Computer Vision. Springer, Cham, 2014

[26] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//Proceedings 2007 IEEE International Conference on Robotics and Automation. Rome, Italy. IEEE, 2007: 3565-3572

[27] Leutenegger S, Lynen S, Bosse M, et al. Keyframe-based visual–inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334

[28] Bloesch M, Omari S, Hutter M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany. IEEE, 2015: 298-304

[29] Xu D, Vedaldi A, Henriques J F. Moving SLAM: fully unsupervised deep learning in non-rigid scenes[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic. IEEE,2021 : 4611-4617

[30] 张彦雯，胡凯，王鹏盛. 三维重建算法研究综述[J]. 南京信息工程大学学报(自然科学版), 2020, 12(5): 591-602

ZHANG Yanwen, HU Kai, WANG Pengsheng. Review of 3D reconstruction algorithms[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2020, 12(5): 591-602

[31] Li J L, Li Z J, Feng Y, et al. Development of a human–robot hybrid intelligent system based on brain teleoperation and deep learning SLAM[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(4): 1664-1674

[32] Mumuni A, Mumuni F. CNN architectures for geometric transformation-invariant feature representation in computer vision: a review[J]. SN Computer Science, 2021, 2(5): 1-23

[33] Ma R B, Wang R, Zhang Y B, et al. RNNSLAM: Reconstructing the 3D colon to visualize missing regions during a colonoscopy[J]. Medical Image Analysis, 2021, 72: 102100

[34] Tateno K, Tombari F, Laina I, et al. CNN-SLAM: real-time dense monocular SLAM with learned depth prediction[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6565-6574

[35] Wang S, Clark R, Wen H K, et al. DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks[C]//2017 IEEE International Conference on Robotics and Automation. Singapore. IEEE, 2017: 2043-2050

[36] Bloesch M, Czarnowski J, Clark R, et al. CodeSLAM - learning a compact, optimisable representation for dense visual SLAM[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE,2018 : 2560-2568

[37] Yang N, Wang R, Stückler J, et al. Deep virtual stereo odometry: leveraging deep depth prediction for monocular direct sparse odometry[C]//Computer Vision – ECCV 2018, 2018

[38] Li R H, Wang S, Long Z Q, et al. UnDeepVO: monocular visual odometry through unsupervised deep learning[C]//2018 IEEE International Conference on Robotics and Automation. Brisbane, QLD, Australia. IEEE, 2018: 7286-7291

[39] Loo S Y, Amiri A J, Mashohor S, et al. CNN-SVO: improving the mapping in semi-direct visual odometry using single-image depth prediction[C]//2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada. IEEE, 2019: 5218-5223

[40] Almalioglu Y, Saputra M R U, de Gusmão P P B, et al. GANVO: unsupervised deep monocular visual odometry and depth estimation with generative adversarial networks[C]//2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada. IEEE, 2019: 5474-5480

[41] Li Y, Ushiku Y, Harada T. Pose graph optimization for unsupervised monocular visual odometry[C]//2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada. IEEE, 2019: 5439-5445

[42] Yang N, von Stumberg L, Wang R, et al. D3VO: deep depth, deep pose and deep uncertainty for monocular visual odometry[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 1278-1289

[43] Chancán M, Milford M. Deepseqslam: a trainable CNN+RNN for joint global description and sequence-based place recognition[J]. arXiv preprint arXiv:2011.08518, 2020

[44] Li R H, Wang S, Gu D B. DeepSLAM: a robust monocular SLAM system with unsupervised deep learning[J]. IEEE Transactions on Industrial Electronics, 2021, 68(4): 3577-3587

[45] Bruno H M S, Colombini E L. LIFT-SLAM: a deep-learning feature-based monocular visual SLAM method[J]. Neurocomputing, 2021, 455: 97-110

[46] Zhang S M, Lu S Y, He R, et al. Stereo visual odometry pose correction through unsupervised deep learning[J]. Sensors (Basel, Switzerland), 2021, 21(14): 4735

[47] Clark R, Wang S, Wen H, et al. Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem[C]//Proceedings of the AAAI Conference on Artificial Intelligence，2017:3995-4001

[48] Shamwell E J, Lindgren K, Leung S, et al. Unsupervised deep visual-inertial odometry with online error correction for RGB-D imagery[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2478-2493

[49] Han L M, Lin Y M, Du G G, et al. DeepVIO: self-supervised deep learning of monocular visual inertial odometry using 3D geometric constraints[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macao, China. IEEE,2019 : 6906-6913

[50] Chen C H, Rosa S, Miao Y S, et al. Selective sensor fusion for neural visual-inertial odometry[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE,2019 : 10534-10543

[51] Kim Y, Yoon S, Kim S, et al. Unsupervised balanced covariance learning for visual-inertial sensor fusion[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 819-826

[52] Gurturk M, Yusefi A, Aslan M F, et al. The YTU dataset and recurrent neural network based visual-inertial odometry[J]. Measurement, 2021, 184: 109878

[53] 傅杰，徐常胜．关于单目标跟踪方法的研究综述[J]．南京信息工程大学学报(自然科学版), 2019, 11(6): 638-650

FU Jie, XU Changsheng. A survey of single object tracking methods[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2019, 11(6): 638-650

[54] Steenbeek A, Nex F. CNN-based dense monocular visual SLAM for real-time UAV exploration in emergency conditions[J]. Drones, 2022, 6(3): 79

[55] 唐灿，唐亮贵，刘波．图像特征检测与匹配方法研究综述[J]．南京信息工程大学学报(自然科学版), 2020, 12(3): 261-273

TANG Can, TANG Lianggui, LIU Bo. A survey of image feature detection and matching methods[J]. Journal of Nanjing University of Information Science ＆ Technology (Natural Science Edition), 2020, 12(3): 261-273

[56] Li L, Kong X, Zhao X R, et al. Semantic scan context: a novel semantic-based loop-closure method for LiDAR SLAM[J]. Autonomous Robots, 2022, 46(4): 535-551

[57] Sakkari M, Hamdi M, Elmannai H, et al. Feature extraction-based deep self-organizing map[J]. Circuits, Systems, and Signal Processing, 2022, 41(5): 2802-2824

[58] Zhu R, Yang M K, Liu W, et al. DeepAVO: Efficient pose refining with feature distilling for deep Visual Odometry[J]. Neurocomputing, 2022, 467: 22-35

[59] Kim J J Y, Urschler M, Riddle P J, et al. SymbioLCD: ensemble-based loop closure detection using CNN-extracted objects and visual bag-of-words[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic. IEEE,2021 : 5425

[60] Ai Y B, Rui T, Lu M, et al. DDL-SLAM: a robust RGB-D SLAM in dynamic environments combined with deep learning[J]. IEEE Access, 8: 162335-162342

[61] Javed Z, Kim G W. PanoVILD: a challenging panoramic vision, inertial and LiDAR dataset for simultaneous localization and mapping[J]. The Journal of Supercomputing, 2022, 78(6): 8247-8267

[62] Duan R, Feng Y R, Wen C Y. Deep pose graph-matching-based loop closure detection for semantic visual SLAM[J]. Sustainability, 2022, 14(19): 11864

[63] Memon A R, Wang H S, Hussain A. Loop closure detection using supervised and unsupervised deep neural networks for monocular SLAM systems[J]. Robotics and Autonomous Systems, 2020, 126: 103470

[64] Li D J, Shi X S, Long Q W, et al. DXSLAM: a robust and efficient visual SLAM system with deep features[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, NV, USA. IEEE, : 4958-4965

[65] Qin C, Zhang Y Z, Liu Y D, et al. Semantic loop closure detection based on graph matching in multi-objects scenes[J]. Journal of Visual Communication and Image Representation, 2021, 76: 103072

[66] Godard C, Aodha O M, Brostow G J. Unsupervised monocular depth estimation with left-right consistency[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. IEEE,2017: 6602-6611

[67] Sang H R, Jiang R, Wang Z P, et al. A novel neural multi-store memory network for autonomous visual navigation in unknown environment[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2039-2046

[68] Li G H, Chen S L. Visual slam in dynamic scenes based on object tracking and static points detection[J]. Journal of Intelligent & Robotic Systems, 2022, 104(2): 1-10

[69] Liu L, Tang T H, Chen J, et al. Real-time 3D reconstruction using point-dependent pose graph optimization framework[J]. Machine Vision and Applications, 2022, 33(2): 1-11

[70] Xue F, Wang Q, Wang X, et al. Guided feature selection for deep visual odometry[C]//Asian Conference on Computer Vision. Springer, Cham, 2018: 293-308

[71] Tang Y F, Wei C C, Cheng S L, et al. Stereo visual-inertial odometry using structural lines for localizing indoor wheeled robots[J]. Measurement Science and Technology, 2022, 33(5): 055114

[72] Bucci A, Zacchini L, Franchi M, et al. Comparison of feature detection and outlier removal strategies in a mono visual odometry algorithm for underwater navigation[J]. Applied Ocean Research, 2022, 118: 102961

[73] Wu J F, Xiong J, Guo H. Improving robustness of line features for VIO in dynamic scene[J]. Measurement Science and Technology, 2022, 33(6): 065204

[74] Huang W B, Wan W W, Liu H. Optimization-based online initialization and calibration of monocular visual-inertial odometry considering spatial-temporal constraints[J]. Sensors (Basel, Switzerland), 2021, 21(8): 2673

[75] Xiang Y, Fox D. DA-RNN: semantic mapping with data associated recurrent neural networks[J]. arXiv preprint arXiv:1703.03098, 2017