

基于深度学习的视觉惯性里程计技术综述

王文森¹, 黄凤荣¹⁺, 王 旭², 刘庆麟¹, 羿博珩¹

1. 河北工业大学 机械工程学院, 天津 300401

2. 中国人民解放军 93756 部队

+ 通信作者 E-mail: 2015016@hebut.edu.cn

摘 要:视觉惯性里程计在很多方面可以很好地实现视觉和惯性传感器的优势互补, 获得高精度的6自由度导航定位, 因此应用领域极为广泛。然而, 传感器自身的误差、异常视觉环境的扰动、多传感器之间的时空校准误差都会干扰导航结果, 导致导航精度下降。近年来, 正在迅速发展的深度学习方法凭借其强大的数据处理和预测能力, 给视觉惯性里程计的发展提供了全新的发展方向。对基于深度学习的视觉惯性里程计的主要发展成果进行了回顾与总结。首先, 按照两种融合策略分别概述研究方法, 包括深度学习与传统模型结合的方法和基于深度学习的端到端的方法。之后, 根据深度学习类型分为监督学习和无监督/自监督学习的方法, 并分别阐述了这些方法的模型结构。然后, 概述了系统的优化与评估方法, 并比较了其中一些具有代表性的方法的性能。最后, 对该领域需要解决的关键难点问题进行了总结, 对未来发展进行了展望。

关键词:视觉惯性里程计; 特征融合; 深度学习; 网络模型; 位姿

文献标志码:A **中图分类号:**V249.32⁺8; TP301.6

Overview of Visual Inertial Odometry Technology Based on Deep Learning

WANG Wensen¹, HUANG Fengrong¹⁺, WANG Xu², LIU Qinglin¹, YI Boheng¹

1. College of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China

2. 93756 Unit of PLA, China

Abstract: Visual inertial odometer can well realize the complementary advantages of vision and inertial sensors, and obtain high precision 6-DOF navigation and positioning, so it has a very wide range of applications. However, the errors of sensors themselves, the disturbance of abnormal visual environment, and the space-time calibration errors between multi-sensor will interfere with the navigation results, leading to the decline of navigation accuracy. In recent years, the deep learning method is developing rapidly. With its powerful data processing and prediction ability, it provides a new direction for the development of visual inertial odometer. This paper reviews the main development achievements of deep learning-based methods. First of all, according to the fusion mode, the research methods are summarized, which are divided into the method combining deep learning with traditional models and the end-to-end method based on deep learning. Then, according to the type of deep learning, visual inertial odometer can be divided into supervised learning and unsupervised/self-supervised learning methods, and the model structures of these methods are described respectively. Next, the optimization and evaluation methods of the system are summarized, and the performance of some of them is compared. Finally, this paper summarizes the key and difficult problems that need to be solved in this field, and looks forward to the future development.

Key words: visual inertial odometry; feature fusion; deep learning; network model; posture

基金项目:国家自然科学基金(61973333)。

This work was supported by the National Natural Science Foundation of China (61973333).

收稿日期:2022-09-05 **修回日期:**2022-11-30

视觉惯性里程计(visual inertial odometry, VIO)^[1-3], 又称为视觉惯性导航系统,是由视觉和惯性传感器构成的组合导航系统。VIO拥有自主性、实时性等特点,传感器的优势互补使VIO的导航精度明显高于由单一传感器组成的惯性导航系统或视觉里程计(visual odometry, VO),低成本、体积小的消费级微机电惯性测量单元(micro electro mechanical systems inertial measurement unit, MEMS-IMU)和相机的使用更促进其发展。VIO研究的主要目的,就是充分利用视觉惯性的优势,实现系统的高精度6自由度(degree of freedom, DoF)位置与姿态估计。

传统的VIO系统的基本框架如图1所示。其中,前端包括基于运动学模型的惯性预处理模块和基于几何学模型的视觉里程计,后端为基于滤波器或优化器的状态估计模块。此外,为了进一步提高导航精度,还可能会添加回环检测等功能。传统方法已经展示了不错的性能^[4-5],但受到建模的局限和真实环境的复杂性使其仍然难以投入实际应用中。近年来,深度学习^[6-7]为VIO的方法研究提供了新的思路。深度学习的方法相比传统方法表现出了更强的鲁棒性。基于深度学习的VIO相比传统方法展现出的优势可以体现在以下方面:

(1)传统方法基于复杂的几何与运动学模型,而且现实中很难建立与真实应用严格相符的数学模型,深度学习模型基于神经网络,可以通过自适应训练实现高精度导航。

(2)由于受到IMU的噪声和偏差的影响,传统方法一般仅对惯性数据进行简单的预处理^[8],基于深度学习的方法使惯性特征也具有了量测的能力,可以使系统不再局限于来自单模态的量测特征。

(3)传统方法提取图像特征局限于特征点、线和平面等低级特征的提取方法^[9],深度学习可以学习潜在的高级特征,有利于实现复杂环境中的导航。

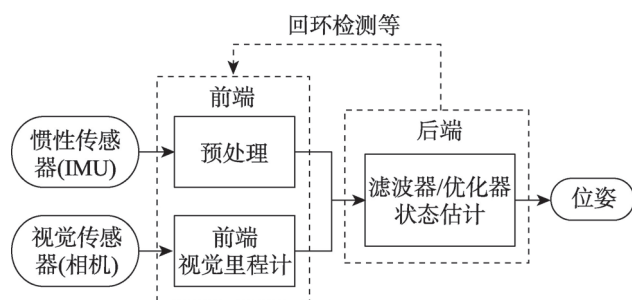


图1 基于几何学与运动学模型VIO的基本框架

Fig.1 Framework of VIO based on geometric and kinematic model

由此,随着越来越多基于深度学习的VIO的研究方法的出现,本文在对基于深度学习的视觉惯性里程计的发展历史、研究现状以及方法梳理的基础上,从融合策略的角度分别对深度学习与传统模型结合的方法和端到端的深度学习方法进行了综述,并分别从监督学习和无监督/自监督学习方面介绍了网络模型,同时分析并阐述了常用数据集、评价指标和方法对比。最后,总结了当前研究中亟待突破的问题并对未来的研究方向进行了展望。

1 基于深度学习的VIO系统融合策略

根据后端是否是以深度学习的方式实现融合,可以将VIO系统按融合策略分为深度学习与传统模型结合的融合和基于深度学习的端到端融合。同时,VIO系统无疑是多模态的融合^[10-11],可分为数据级融合、特征级融合和决策级融合。特征级融合和决策级融合的方法都已经实现,在VIO中一般称之为紧耦合和松耦合。以下将从融合策略概述现有的研究方法。

1.1 深度学习与传统模型结合的融合

在传统方法中,惯性状态计算基于运动学模型,视觉状态和特征点特征计算基于视觉几何模型,最后采用滤波器或优化器实现二者的特征融合。深度学习与传统模型结合的方法完整保留了传统模型的后端,但是在前端则基于深度学习设计了学习状态的新模型。

早期的深度学习模型主要用于替换原有的前端传统模型。Rambach等^[12]设计了首个基于深度学习的监督学习VIO模型,模型结构如图2所示。其惯性前端基于长短时记忆网络(long short-term memory, LSTM)^[6]学习位置和姿态,同时加入误差检测器可以实现惯性网络和视觉前端的互相监督,最后以卡尔曼滤波(Kalman filter, KF)作为后端实现了VIO的松耦合。Li等^[13]将基于卷积神经网络(convolutional

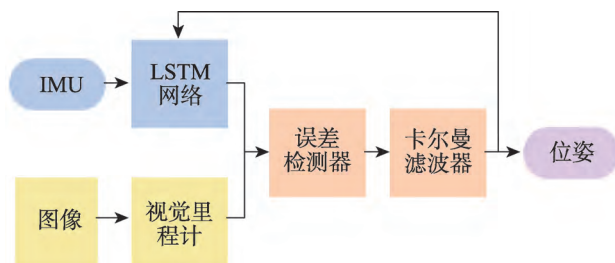


图2 文献[12]的模型结构

Fig.2 Structure of Ref.[12]

neural networks, CNN)^[6]的VO模型DeepVO^[14]作为VIO的视觉前端输出相对位姿,再利用扩展卡尔曼滤波器(extended Kalman filter, EKF)将视觉位姿预处理的惯性状态进行融合。余洪山等^[15]基于改进SuperPoint网络^[16]检测和描述特征点,有效抑制了异常特征点,加强了视觉前端的鲁棒性,在后端则使用了VINS-Mono^[17]的紧耦合融合框架进行融合,实现了高精度导航。

其他方法则会利用深度学习在特征学习中的多样性,在前端建立新的子模块,扩展了后端的特征向量,如行人导航方法RNIN-VIO^[18]使用EKF作为融合后端,在惯性前端的深度学习网络中,利用人体运动的规律性,使用IMU原始数据和滤波器中的姿态学习相对位移和其不确定度。最终,视觉特征、惯性状态和网络输出的惯性特征通过滤波器实现了紧耦合。该方法增强了对惯性特征的利用,提高了系统鲁棒性。系统也可以仅依靠惯性数据进行较高精度的导航。Wang等^[19]同样以EKF作为后端,其视觉前端建立了地标识别模型,通过识别已知位置的地标信息计算比例关系进而实现位姿优化,以缓解位置误差累积的问题。Shan等^[20]和Zuo等^[21]基于MSCKF(multi-state constraint Kalman filter)^[22]的融合框架,前者在前端建模了目标物体的语义特征的网络,系统在几何和语义级别上理解周围环境,以目标物体产生的残差约束视觉惯性的状态,可以实现高精度定位和生成全局地图;后者建模了深度估计网络,将图像深度作为特征向量以实现视觉惯性更紧密的耦合,系统在输出位姿的同时还可以实时地提供密集稠密深度图。以上方法通过建立额外的量测约束,使VIO在一些特定应用场景中拥有更强的鲁棒性。

1.2 基于深度学习的端到端的融合

Clark等^[23]提出了首个使用深度学习框架实现的端到端的监督学习VIO方法VINet,整体可微的CNN-LSTM架构使其可以实现端到端的训练,其中CNN-LSTM架构是由CNN、LSTM网络结合的网络模型架构。系统前端将视觉惯性特征转化为高维特征表达,在后端将视觉特征、惯性特征和上时刻位姿拼接,最后基于LSTM网络 and 全连接层进行特征融合并估计位姿。VINet在应对时间不同步、数据外参标定不准确和校准误差导致的发散时,相比传统方法都表现出更强的鲁棒性。但是其后端没有明确特征选择的建模,隐式的处理方法很难对静态和动态的特征实现有效和灵活的识别,在提取不同表示、不同

分布的数据特征时并不稳定。后续的研究为建模特征选择过程,分别采用基于加法交互作用的方法^[10,24-25]和基于乘法交互作用的方法^[26-28]。对特征选择进行建模进一步提高了系统的鲁棒性,具体可以体现在应对传感器数据丢失、损坏,视觉惯性传感器数据不同步等方面。不同于利用LSTM网络建模特征融合后端的方法,Aslan等^[28]基于高斯过程回归^[29]实现了特征融合。这些方法的原理框图如图3所示。

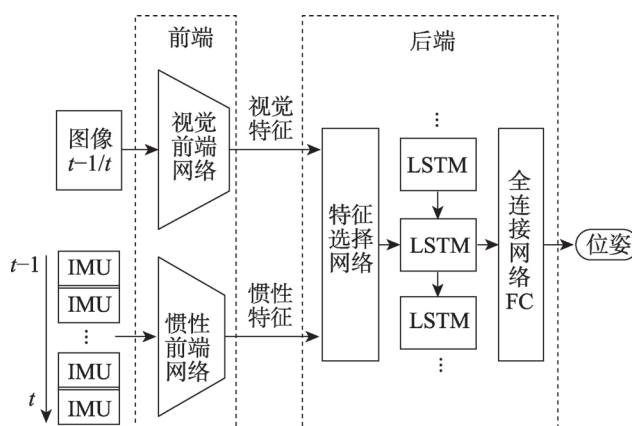


图3 监督学习VIO的基本框架

Fig.3 Basic framework for supervised VIO

为减少对数据集真值的依赖,无监督和自监督的方法^[25,30-35]也被提出,其系统框架如图4所示。无监督与自监督学习的VIO不直接使用数据集真值建立损失函数,而是基于重建的源图像和目标图像的几何约束^[36],建立无监督损失项。无监督VIO中用于建立无监督项的深度图由外部提供,自监督方法的重建图像信息来自相机图像序列,Almalioglu等^[25]使用生成式对抗网络(generative adversarial networks, GAN)和无监督学习方法联合实现姿态估计和生成深度图,实现在未知陌生环境中的定位和建图。Han等^[34]利用立双目图像序列估计深度得到密集的三维点云,进而得到三维光流和6自由度姿态等三维几何约

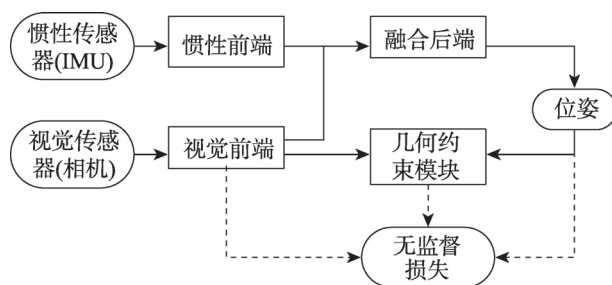


图4 无监督VIO的基本框架

Fig.4 Basic framework of unsupervised VIO

束作为自监督项。无监督 VIO 可以对有尺度轨迹做在线矫正,在面对新环境和恶劣环境时具有更强的适应和泛化能力,同时受错误校准、数据不同步等因素影响相比传统方法要低,有些方法^[25,31-32]还可以在已知传感器外参和视觉惯性数据松散同步的情况下给出载体位姿信息。

2 深度学习 VIO 系统的神经网络模型

深度学习 VIO 的网络模型需依据是否在训练中使用数据集提供的真值,可以分为监督学习模型和无监督/自监督学习模型。

2.1 监督学习模型

惯性前端网络能够利用低精度的 IMU 信息提高整个系统的鲁棒性和精度。Rambach 等^[12]建模的惯性网络包括 1 层 LSTM 网络和 3 层全连接层,虽然网络可以利用有限的的数据得到不错的结果,却存在比较严重的漂移。RNIN-VIO^[18]建模的鲁棒惯性网络由 ResNet18、3 层 LSTM 网络和两个并行的全连接层组成。ResNet18 用于学习人体运动隐藏变量,LSTM 网络将当前的隐藏状态与之前的隐藏状态进行融合,以估计运动的当前隐藏状态。同时 RNIN-VIO 设计了两种不同的损失函数用于保证每个窗口以及长序列的训练精度。视觉前端的网络可以提高系统在快速运动和无纹理场景等特殊环境中的鲁棒性。Li 等^[13]的视觉网络使用 CNN 网络提取视觉特征,将特征排列为时间序列,然后通过双层 LSTM 网络输出相机位姿和不确定度。

VINet^[23]是首个基于深度学习的端到端方法,其模型框架如图 5 所示,其中惯性前端基于 LSTM 网络进行建模,网络每次将图像两帧之间的所有原始数据输入,这样保证了惯性特征的学习和前端视觉惯性

特征的同步输出。光流网络可以利用图像序列中像素在时间域上的变化以及相邻帧之间的相关性来找到图像的对应关系,进而获得载体的运动信息。因此,视觉前端使用预训练的 FlowNetCorr 光流网络^[37-38]的前端卷积部分网络以两张连续的图像作为输入,经过光流网络内 CNN 网络的多次特征提取后输出高维的特征表达。VINet 的后端使用两层的 LSTM 网络建模以实现特征融合。

在特征融合后端,为了进一步提高特征融合网络模型的可解释性和提高系统鲁棒性,Chen 等^[24]提出在视觉惯性特征向量拼接后分别使用具有确定性的软融合和具有随机性的强融合两种具有可解释性的融合模式,以加法交互作用的方式实现特征选择的显示建模。同时,这种方法还采用了轻量级的 FlowNetSimple 网络^[37-38]以加快运行速度。但是这种融合方式依然缺少视觉惯性特征之间的显式联系。为了进一步提高模型的可解释性和可学习性,Shinde 等^[26]基于多头自注意力机制^[39]建模了后端融合模型,以乘法交互作用的方式实现显式融合。ATVIO^[27]在特征选择过程中根据 SENet 网络^[40]构建了注意力生成模块,显式地建模了特征之间的相关性,减少了异常数据对后端特征融合造成的影响。

在特征提取前端也需要准确、高效的模型。早期的惯性网络一般基于 LSTM 网络建模,然而 LSTM 网络内参数较多,训练时间较长。CNN 网络相比 LSTM 网络虽然不能补偿传感器间的时间偏差,但是其建模计算速度更快,网络更稳定和容易收敛^[41]。随着传感器同步校准精度的提高,基于 CNN 的惯性前端网络模型也可以发挥优势。ATVIO^[27]使用了两个并行的 3 层 CNN 网络层分别学习 IMU 中加速度和角速度中的特征。Aslan 等^[28]将平滑和去噪的 IMU 数据

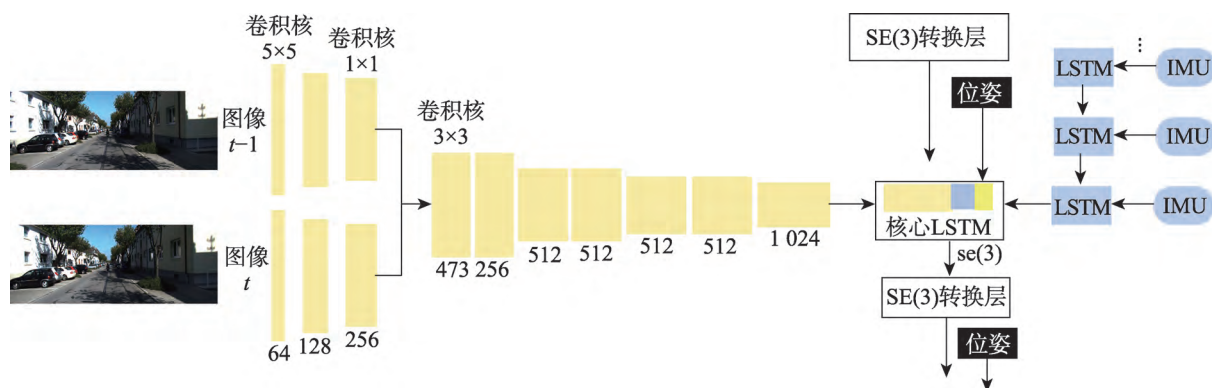


图5 VINet的模型结构

Fig.5 Model structure of VINet

使用预训练的 Inception V3 网络^[42]学习惯性特征。在视觉前端, CNN 网络无法记忆先前的图像信息, 为此 ATVIO^[27] 使用 ConvLSTM 网络建模视觉前端, ConvLSTM 网络是可以同时提取图像时空相关特征的网络, 使视觉前端得以学习来自先前图像特征的约束。此外, 经过合理初始化的视觉前端网络相比未经过训练的网络模型具有更快的收敛速度, 训练过程也更稳定, 因此特征级融合的方法一般会对前端视觉网络进行预训练。

端到端的监督学习模型的损失函数 θ 可以使 k 时刻的真实位姿 (p_k, ϕ_k) 与其估计的地面位姿 $(\hat{p}_k, \hat{\phi}_k)$ 之间的欧氏距离最小化以实现最优结果^[14,43-44], 一般以均方误差 (mean square error, MSE) 计算, 称为 MSE 损失函数。部分数据集的姿态真值以四元数的形式保存, 但直接使用四元数计算损失会因其冗余的维数导致训练难度增加, 同时浪费了计算资源, 因此一般会将四元数转化为欧拉角使用。网络模型复杂的深层结构使 MSE 损失函数在训练中仍受到诸多限制, 模型子网络的平均性能较差。于是 Liu 等^[27] 将自适应损失函数^[45] 应用于训练过程中, 模型在训练过程中自适应地调整参数, 加快了网络收敛, 同时强化了对子网络的训练, 提升了网络整体性能。监督学习 VIO 的损失函数定义为:

$$\theta_{\text{Adapt}} = \theta(\hat{p}_k - p_k) + \beta \cdot \theta(\hat{\phi}_k - \phi_k) \quad (1)$$

其中, β 是用于平衡位置和姿态的比例因子。

2.2 无监督/自监督学习模型

无监督和自监督学习的 VIO 需要通过在训练过程中建立约束模型以摆脱对数据集真值的依赖或应对没有真值的情况。在深度学习与传统模型结合的方法中, 由于难以提供真实的视觉特征, 利用深度学习对特征点或其他特征进行跟踪匹配, 或者实现深度预测等, 往往需要用无监督或自监督学习解决。余洪山等^[15] 的改进 SuperPoint 网络由轻量级的编码层、特征点检测层和描述符解码层构成, 采用稀疏描述符损失函数进行训练, 但是网络在训练前还需经过预训练获取合适的初始化参数, 以保证后续网络的正常收敛。CodeVIO^[21] 用于深度预测的网络分为两部分: 一部分是修改过的编码网络, 通过原始图像和级联稀疏深度图预测稠密的深度图及其不确定度; 另一部分是变分自编码器, 通过对深度信息进行编码得到用于 VIO 优化的深度向量。

Shamwell 等^[31-32] 提出了首个端到端的无监督方法 VIO Learner, 模型结构如图 6 所示。在 IMU 固有参数和外部校准参数未知的情况下, 网络首先学习

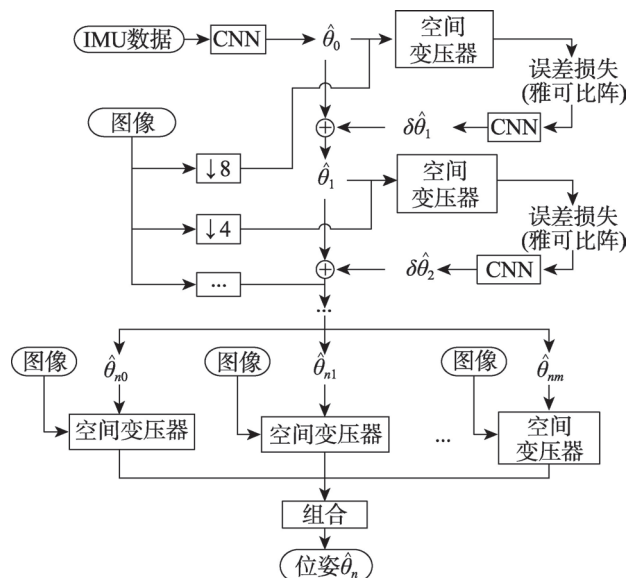


图6 VIO Learner的模型结构

Fig.6 Model structure of VIO Learner

IMU 状态并生成原始轨迹, 然后通过多尺度缩放图像的投影误差的修正, 实现原始轨迹的在线校正。多尺度的缩放不仅有助于克服训练期间的梯度局部性, 而且有助于在运行时进行在线误差校正。Lindgren 等^[33] 提出了 Boom-VIO, 系统包括一个学习相对位移的传统模型、一个深度网络和一个无监督学习模型, 无监督学习模型基于 VIO Learner。其在网络训练过程中加入传统模型的引导, 并得到最终的训练轨迹。DeepVIO^[34] 通过直接结合二维光流特征和 IMU 原始数据来提供绝对轨迹估计。系统包括一个学习视觉特征的 CNN 光流网络, 一个学习惯性特征的 LSTM 网络, 一个用于融合的全连接网络。此外, 还有一个用于建立自监督约束的模块, 能够分别对视觉网络、IMU 网络和整体的网络进行训练, 其模型结构如图 7 所示。SelfVIO^[25] 前端包括基于 CNN 的惯性网络、视

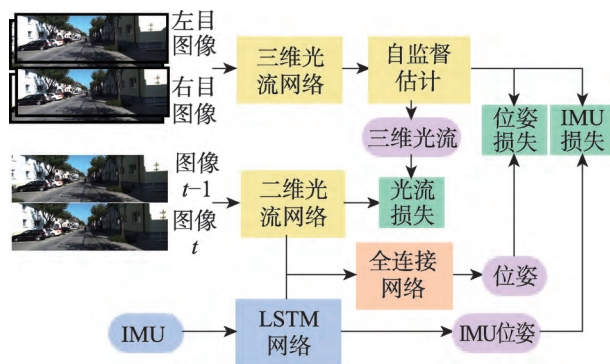


图7 DeepVIO的模型结构

Fig.7 Model structure of DeepVIO

觉网络和深度学习网络,后端由基于多头自注意力机制的融合网络和LSTM网络组成。其中,深度网络学习输出的单目深度图,与网络估计的位姿、源图像共同实现图像重建。UnVIO^[35]同样通过预测图像深度建立无监督约束。此外,UnVIO在训练过程中采用了滑动窗口优化的策略,以克服长期运行中误差累积和尺度模糊的问题。窗口内部通过判断光度一致性建立几何约束,窗口之间利用三维几何一致性和轨迹一致性建立约束,这有效缓解了误差累积的问题。

无监督和自监督损失可以利用图像的时间或空间性质构造^[31-32],以 $\langle I_1, I_2, \dots, I_N \rangle$ 表示一个训练的图像

序列,其中的某一帧 I_t 为目标图像,其余的作为源图像,根据两帧图像间的光度差异可定义损失函数为:

$$\theta_{\text{unsupervised}} = \sum_s \sum_p |I_t(p) - \hat{I}_s(p)| \quad (2)$$

其中, p 是像素点坐标值, $\hat{I}_s$ 是基于源图像 I_s 重建后的源图像。

3 深度学习VIO的数据优化与评估

以上方法从学习方式、融合方式、方法特性、方法局限等方面汇总并整理至表1。除建立网络模型外,模型的训练、优化与评估方法也至关重要。深度

表1 基于深度学习的VIO方法概览

Table 1 Overview of deep learning-based VIO methods

方法	时间	学习方式	融合方式	方法特性	方法局限
Rambach等 ^[12]	2016	监督	滤波器	前端的惯性网络与视觉里程计可互相检测异常状态并修正	松耦合方法,传感器互补性不足
VINet ^[23]	2017	监督	深度学习	首个模型端到端的可训练网络模型,减少了传感器之间的时空误差造成的漂移	后端没有显式的特征选择建模过程 ^[24]
VIO Learner ^[31-32]	2018	无监督	深度学习	可以在没有IMU内参和传感器外部校准的情况下实现导航和在线校正轨迹	深度图由外部设备提供,缺少自主性 ^[35]
Chen等 ^[10,24]	2019	监督	深度学习	端到端网络模型,建立了显式的视觉惯性融合模块,增强了系统的稳定性和鲁棒性	后端的特征选择过程时特征融合深度较低 ^[26]
Tian等 ^[30]	2019	半/无监督	深度学习	基于VINet ^[23] 的网络模型建立无监督项实现无监督训练	精度低于VINet且泛化能力差
Boom-VIO ^[33]	2019	无监督	深度学习	利用传统方法引导网络训练,从单目图像中估计轨迹尺度,图像退化环境中高鲁棒性	复杂性较高,模型训练效率低;精度较低 ^[33]
DeepVIO ^[34]	2019	自监督	深度学习	利用立体图像序列计算自监督信息,并用于训练视觉光流和系统的位姿	复杂性较高,模型训练效率低
UnVIO ^[35]	2020	无监督	深度学习	建立了显式视觉惯性融合模块;通过滑动窗口优化的方式克服误差累积和尺度模糊问题	复杂性较高,模型训练效率低
Li等 ^[13]	2020	监督	滤波器	以深度学习的单目VO方法DeepVO作为视觉前端提供视觉位姿	松耦合方法,传感器互补性不足,易受加速度计偏差影响 ^[13]
OrcVIO ^[20]	2020	监督	滤波器	系统拥有语义级别理解环境的能力,可以利用物体的残差做量测约束	无物体重新识别机制,减少依赖语义关键点使精度下降 ^[20]
Shinde等 ^[26]	2020	监督	深度学习	端到端网络模型,基于多头自注意力机制建立了显式的视觉惯性融合模块	复杂性较高
RNIN-VIO ^[18]	2021	监督	滤波器	建立了鲁棒的惯性前端网络,减少系统对视觉信息的依赖	仅适用于行人导航
Wang等 ^[19]	2021	监督	滤波器	测量地标位置实现定位修正,抑制导航误差随时间累积的问题	只能应用于有已定义地标的场景中
CodeVIO ^[21]	2021	自监督	滤波器	可利用稀疏测量更新稠密的深度图	只进行了局部稠密映射,没有全局稠密映射 ^[21]
ATVIO ^[27]	2021	监督	深度学习	端到端网络模型,基于通道域注意力机制建立了显式的视觉惯性融合模块	复杂性较高
余洪山等 ^[15]	2021	自监督	非线性优化	基于网络提取特征点和描述子,在光强变化的场景中有较高鲁棒性	复杂性较高,特征点跟踪的泛化能力有限 ^[15]
SelfVIO ^[25]	2022	自监督	深度学习	GAN网络提供清晰准确的深度图,自适应融合提高了特征的互补性	GAN网络训练过程存在不稳定性,模型训练效率低
VHONet ^[28]	2022	监督	深度学习	端到端的可训练网络模型,惯性数据图像化后建立光流网络实现惯性特征的提取	复杂性较高,信息大量冗余;计算量大,难以实时更新 ^[28]

学习 VIO 模型的训练和测试需要使用数据集。模型优化的最终要求是模型输出的损失达到目标值,这需要选择合适的优化器,并针对不同的融合策略和学习方式建立与之匹配的损失函数等,这里只展开介绍损失函数。评估方法可以用于对比系统因模型的变化,或面对不同的环境,或与不同方法的横向对比中时,表现出这些模型、方法的优秀性能和存在的问题。因此,本章将对 VIO 现有的公开数据集与评估方法进行总结,同时比较部分方法的性能。

3.1 数据集

基于深度学习的 VIO 网络模型需要使用大量数据进行训练以提高泛化能力和提高导航精度。网络模型在训练测试过程中一般使用公共的数据集。公共数据集按采集数据的载体平台分类可分为:驾驶类数据集 KITTI(Odometry 序列)^[46]、Malaga Urban^[47]、UMich NCLT^[48]、Zurich Urban^[49]、Canoe^[50]、CUHK-AHU^[51]等;手持设备数据集 TUM-VI^[52]、PennCOSYVIO^[53]、ADVIO^[54]、CVG ZJU^[55]、NEAR^[56]、UMA-VI^[57]、HAUD^[58]等;微型飞行器(micro air vehicle, MAV)/无人驾驶飞机(unmanned aerial vehicle, UAV)等小型机器人数据集 EuRoC MAV^[59]、AQUALOC^[60]、Blackbird UAV^[61]等;虚拟系统采集的数据集 WHU-RSVI^[62]、VIODE^[63]等。以上数据集的基本属性可见表 2。其中, KITTI、

EuRoC MAV 是常用的公开数据集。

KITTI 数据集^[46]由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合制作,是目前最大的自动驾驶场景中的公开数据集。KITTI 包含市区、乡村和高速公路等室外场景采集的 22 个序列,其中 11 个有真值。图像采集自 2 个灰度相机(FL2-14S3M-C)、2 个彩色相机(FL2-14S3C-C),采集频率为 10 Hz, IMU 采集频率为 100 Hz,真值来自高精度全球定位和惯性导航组成的组合系统 OXTS RT 3003。

EuRoC MAV 数据集^[59]是由苏黎世联邦理工学院制作的微型飞行器数据集,数据采集于一个工厂场景和两个室内场景。整个数据集包含从良好视觉条件下的缓慢飞行到运动模糊和光照差的动态飞行共 11 个序列。图像采集使用双目相机 MT9V034,采集频率为 20 Hz, IMU 使用 ADIS16448,采集频率为 200 Hz,真值来自激光跟踪系统或 Vicon 动捕系统。

3.2 评估方法与指标

深度学习网络通常是模块化设计,可以使用消融实验^[64],即通过删除、修改或替换某些模块以判断网络行为和验证一些提出的方法的有效性。

评估 VIO 最重要的指标就是导航精度。在 VIO 方法的评估实验中,常用的度量标准包括:

(1)绝对轨迹误差(absolute trajectory error, ATE)

表 2 VIO 数据集

Table 2 VIO datasets

数据集	时间	环境	载体平台	真值来源	真值
KITTI(Odometry) ^[46]	2013	室外	汽车	OXTS RT 3003 系统	6DoF 位姿
Malaga Urban ^[47]	2014	室外	汽车	GPS	地面水平位置
UMich NCLT ^[48]	2016	室内/室外	平衡车	GPS/IMU/激光组合	6DoF 位姿
EuRoC MAV ^[59]	2016	室内	MAV	激光或 Vicon 动作捕捉系统	仅位置或 6DoF 位姿
Zurich Urban ^[49]	2017	室外	MAV	Pix4D 与视觉位姿	6DoF 位姿
PennCOSYVIO ^[53]	2017	室内/室外	手持	视觉基准标记	6DoF 位姿
TUM-VI ^[52]	2018	室内/室外	手持设备	OptiTrack 动作捕捉系统	6DoF 位姿
ADVIO ^[54]	2018	室内/室外	手持设备	惯导系统与人工标记点定位	6DoF 位姿
Canoe ^[50]	2018	河边水面	船	GPS 与 IMU 组合	6DoF 位姿
CVG ZJU ^[55]	2019	室内	手持设备	Vicon 动作捕捉系统	6DoF 位姿
AQUALOC ^[60]	2019	水下	潜水器	SfM 离线计算位姿	6DoF 位姿
NEAR ^[56]	2019	室内	手持手机	视觉跟踪系统	6DoF 位姿
WHU-RSVI ^[62]	2020	室内	虚拟场景	虚拟系统	6DoF 位姿
UMA-VI ^[57]	2020	室内/室外	手持设备	SfM 离线计算位姿	6DoF 位姿
Blackbird UAV ^[61]	2020	室内/室外	UAV	OptiTrack 动作捕捉系统	6DoF 位姿
CUHK-AHU ^[51]	2020	室内/室外	汽车	基于图的 SLAM 技术	6DoF 位姿
VIODE ^[63]	2021	室内/室外	虚拟 UAV	虚拟系统	6DoF 位姿
HAUD ^[58]	2021	水下	手持设备	Vicon 动作捕捉系统	6DoF 位姿

直接计算 VIO 位姿的估计值与真实值之间的差值,可以直观地反映算法的精度。首先将真实值与估计值的时间戳对齐,然后计算每对位姿之间的差值。一般使用均方根误差(root mean square error, RMSE)统计 ATE。

(2)相对位姿误差(relative pose error, RPE)用于衡量运动轨迹中固定长度或时间内的局部准确度。通过位姿真实值与估计值的实时比较,可以估计系统的漂移情况,一般使用 RMSE 统计 RPE。

(3)CPU/GPU 的负载、内存的占用、计算速度等参数也是 VIO 的评价指标, VIO 不仅要实现高精度,也要综合考虑应用环境的成本和实现条件。

表 3 比较了一些重要方法在公开数据集中的性能。评估指标为 KITTI 的 09、10 两个序列在长度为 100~800 m 的平均位移和角度的均方根误差漂移 t_{rel} (%) 和 r_{rel} (°)/hm) 以及 EuRoC 中 Vicon 动捕房间中的前 5 个数据集的绝对轨迹误差。此外表中还添加了经典的传统方法进行对比,包括基于滤波的方法 MSCKF^[22]、S-MSCKF^[65]和基于优化的方法 OKVIS^[66]、VINS-Mono^[17]。其中可以看到,在大部分的测试中深度学习方法具有更高的精度。同时,数据集不同可

能会影响深度学习方法的结果,比如 Li 等^[13]的方法在 KITTI 中具有很高的精度,然而在 IMU 数据的偏差噪声更大的 EuRoC 精度较差。此外,在遇到光照改变、图像模糊、相机运动过快、图像和 IMU 数据丢失等情况时,深度学习的方法表现出更强的鲁棒性。

4 总结与展望

本文简述了深度学习 VIO 的研究现状,对研究方法进行了梳理和概括,总结了基于深度学习的系统融合策略,分析了深度学习 VIO 的模型结构,并对可用于其数据集、损失函数以及评估模型的方法与指标等进行了介绍,以期能对现有的方法进行总结,以及对未来的发展方向提供一些参考。目前可以从两方面总结现有方法的性能。

(1)从融合策略的方面来说。深度学习与传统模型结合的方法利用网络可以针对性地优化子模型的性能,进而提高系统的鲁棒性;同时,系统内部有明确意义的特征可以与其他系统进行一定程度的相互融合。这类方法的局限是其限制了隐藏特征的表达,而且状态量的增多会提高模型的复杂度,使计算量增加。端到端的方法对潜在特征挖掘的能力要高于与传统模型结合的方法,但是复杂网络的训练首先需要高性能的计算机;其次,网络模型内部的不可解释性使得端到端的模型内部的高维特征表达也使其内部的特征难以利用,使系统功能仅局限于输出位姿。

(2)从网络模型的学习方式来说。监督学习与无监督学习的模型都具有很强的鲁棒性,在有挑战性的视觉环境中相比传统方法可以保持更高的导航精度。然而,这些模型需要大量数据进行训练,同时它们都难以在与训练环境不同的场景中继续保持高精度。监督学习模型结构更简单,训练更容易;无监督因无监督项的构建使模型更为复杂,同时训练也相对困难。

深度学习与 VIO 结合的研究正在快速发展,基于深度学习的 VIO 的方法研究正不断地有新的研究成果出现。同样的,依然存在很多可以优化和尚未解决的问题,需要继续深入研究。基于以上存在的问题,未来开展基于深度学习的 VIO 方法研究时可以从初始对准、复杂环境导航、深度融合和多系统融合等方面着手,具体如下:

(1)初始对准。初始对准极大地影响后续位姿估计,初始化的不准确将使后续位姿的回归快速发散,初始化是 VIO 运行过程中非常重要的一步。VIO

表 3 基于深度学习的 VIO 方法比较

Table 3 Comparison of deep learning-based VIO methods

方法	KITTI		EuRoC
	$t_{rel}/\%$	$r_{rel}/((^\circ)/\text{hm})$	ATE/m
MSCKF ^{[22]a}			0.28
S-MSCKF ^{[65]a}			0.12
OKVIS ^{[66]b}	13.91	2.90	0.21
VINS-Mono ^{[17]a}	30.50	2.48	0.12
VINet ^{[23]c}	13.98	4.66	
VIO-Learner ^{[31-32]d}	2.53	1.34	
VIO-soft ^{[10,24]c}	13.65	4.83	
Boom-VIO ^{[33]*}	20.16	2.91	
DeepVIO ^{[34]d}	1.12	1.08	
SelfVIO ^{[25]b}	1.88	1.23	0.10
UnVIO ^{[35]*}	4.82	0.71	
Li 等 ^{[13]a}	1.15	0.25	2.19
OrcVIO ^{[20]*}	5.95		
Shinde 等 ^{[26]c}	10.60	4.30	
CodeVIO ^{[21]*}			0.10
ATVIO ^{[27]d}	4.03	1.72	
余洪山等 ^{[15]*}			0.06
VIIONet ^{[28]b}			0.05

注:a 直接由实验验证,b 来自文献[25],c 来自文献[26],d 来自文献[27],*来自对应文献,没有相关实验结果的部分保留为空白。

的初始对准因系统初始位置的随机性使其难以通过真实数据集进行训练,可以使用无监督学习的方式实现,在保证对准精度和时间的情况下省略传感器标定、IMU与相机校准等的人工校准行为。

(2)复杂环境导航。基于深度学习的方法需要根据数据集进行训练,在面对与训练数据不同的场景中,导航的精度会快速下降。因此,可以建立多场景的大型数据集,通过包含更多场景、更多运动模式的数据集提高模型的鲁棒性,或者使用迁移学习等方法提高模型的泛化性。此外,在深度学习与传统模型结合的方法中可以学习一些高级特征,比如利用语义信息实现语义层面的定位约束,提高系统的鲁棒性和环境适应性。

(3)深度融合。目前的端到端网络模型对多模态特征的冗余性和差异性的理解依然有限。特征融合过程中引入新的深度学习方法可以进一步提高融合深度。在基于滤波和优化的VIO方法中,深度学习使多模态特征的融合不再局限于后端的滤波器或优化器,前端的融合可以增加系统的融合深度,提高系统的精度和鲁棒性。

(4)多系统融合。系统间的协作与融合是一个趋势,VIO可以与其他导航传感器结合以适应某些特殊环境或运动行为。比如与蓝牙、WiFi定位相结合实现行人或机器人的室内导航,与全球定位系统(global positioning system, GPS)结合以提高远距离无人机导航的精度和自主性。也可以投入实际应用中,以输出位姿、深度地图等作为辅助信号,实现系统的路径规划和自动驾驶等研究。

参考文献:

- [1] GEMEINER P, EINRAMHOF P, VINCZE M. Simultaneous motion and structure estimation by fusion of inertial and vision data[J]. *International Journal of Robotics Research*, 2007, 26(6): 591-605.
- [2] SERVANT F, HOULIER P, MARCHAND É. Improving monocular plane-based SLAM with inertial measures[C]// *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, China, Oct 18-22, 2010. Piscataway: IEEE, 2010: 3810-3815.
- [3] HUANG G. Visual-inertial navigation: a concise review[C]// *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, Jun 13-19, 2020. Piscataway: IEEE, 2020: 2575-2075.
- [4] DELMERICO J, SCARAMUZZA D. A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots[C]// *Proceedings of the 2018 IEEE International Conference on Robotics and Automation*, Brisbane, May 21-25, 2018. Piscataway: IEEE, 2018: 2502-2509.
- [5] SERVIÈRES M, RENAUDIN V, DUPUIS A, et al. Visual and visual-inertial slam: state of the art, classification, and experimental benchmarking[J]. *Journal of Sensors*, 2021(1): 1-26.
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [7] FARSAL W, ANTER S, RAMDANI M. Deep learning: an overview[C]// *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, Rabat, Nov 24-26, 2017. New York: ACM, 2018: 1-6.
- [8] FORSTER C, CARLONE L, DELLAERT F, et al. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation[C]// *Proceedings of the Robotics: Science and Systems*, Rome, Jul 13-17, 2015. Cambridge: MIT Press, 2015.
- [9] YANG Y, HUANG G. Aided inertial navigation with geometric features: observability analysis[C]// *Proceedings of the 2018 IEEE International Conference on Robotics and Automation*, Brisbane, May 21-25, 2018. Piscataway: IEEE, 2018: 2334-2340.
- [10] CHEN C, ROSA S, LU C X, et al. SelectFusion: a generic framework to selectively learn multisensory fusion[J]. *arXiv*: 1912.13077, 2019.
- [11] 任泽裕, 王振超, 柯尊旺, 等. 多模态数据融合综述[J]. *计算机工程与应用*, 2021, 57(18): 49-64.
REN Z Y, WANG Z C, KE Z W, et al. A survey of multi-modal data fusion[J]. *Computer Engineering and Applications*, 2021, 57(18): 49-64.
- [12] RAMBACH J, TEWARI A, PAGANI A, et al. Learning to fuse: a deep learning approach to visual-inertial camera pose estimation[C]// *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality*, Merida, Sep 19-23, 2016. Piscataway: IEEE, 2016: 71-76.
- [13] LI C, WASLANDER S L. Towards end-to-end learning of visual inertial odometry with an EKF[C]// *Proceedings of the 17th Conference on Computer and Robot Vision*, Ottawa, May 13-15, 2020. Piscataway: IEEE, 2020: 190-197.
- [14] WANG S, CLARK R, WEN H, et al. DeepVO: towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]// *Proceedings of the 2017 IEEE International Conference on Robotics and Automation*, Singapore, May 29-Jun 3, 2017. Piscataway: IEEE, 2017: 2043-2050.
- [15] 余洪山, 郭丰, 郭林峰, 等. 融合改进SuperPoint网络的

- 鲁棒单目视觉惯性SLAM[J]. 仪器仪表学报, 2021, 42(1): 116-126.
- YU H S, GUO F, GUO L F, et al. Robust monocular visual-inertial SLAM based on the improved SuperPoint network [J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 116-126.
- [16] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Piscataway: IEEE, 2018: 2160-7508.
- [17] QIN T, LI P, SHEN S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [18] CHEN D, WANG N, XU R, et al. RNIN-VIO: robust neural inertial navigation aided visual-inertial odometry in challenging scenes[C]//Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality, Bari, Oct 4-8, 2021. Piscataway: IEEE, 2021: 275-283.
- [19] WANG T, CHANG Z, ZHANG W, et al. Research on the integrated positioning method of inertial/visual aided by convolutional neural network[C]//Proceedings of the 2021 International Conference on Control, Automation and Information Sciences, Xi'an, Oct 14-17, 2021. Piscataway: IEEE, 2021: 2475-7896.
- [20] SHAN M, FENG Q, ATANASOV N. OreVIO: object residual constrained visual-inertial odometry[C]//Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Oct 24, 2020-Jan 24, 2021. Piscataway: IEEE, 2020: 5104-5111.
- [21] ZUO X, MERRILL N, LI W, et al. CodeVIO: visual-inertial odometry with learned optimizable dense depth[C]//Proceedings of the 2021 IEEE International Conference on Robotics and Automation, Xi'an, May 30-Jun 5, 2021. Piscataway: IEEE, 2021: 14382-14388.
- [22] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Rome, Apr 10-14, 2007. Piscataway: IEEE, 2007: 3565-3572.
- [23] CLARK R, WANG S, WEN H, et al. VINet: visual-inertial odometry as a sequence-to-sequence learning problem[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, Feb 4-9, 2017. Menlo Park: AAAI, 2017: 3995-4001.
- [24] CHEN C, ROSA S, MIAO Y, et al. Selective sensor fusion for neural visual-inertial odometry[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 15-20, 2019. Piscataway: IEEE, 2019: 10542-10551.
- [25] ALMALIOGLU Y, TURAN M, SAPUTRA M R U, et al. SelfVIO: self-supervised deep monocular visual-inertial odometry and depth estimation[J]. Neural Networks, 2022, 150(6): 119-136.
- [26] SHINDE K, LEE J, HUMT M, et al. Learning multiplicative interactions with Bayesian neural networks for visual-inertial odometry[J]. arXiv:2007.07630, 2020.
- [27] LIU L, LI G, LI T H. ATVIO: attention guided visual-inertial odometry[C]//Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, Jun 6-11, 2021. Piscataway: IEEE, 2021: 4125-4129.
- [28] ASLAN M F, DURDU A, SABANCI K. Visual-inertial image-odometry (VIIONet): a Gaussian process regression-based deep architecture proposal for UAV pose estimation [J]. Measurement, 2022, 194: 111030.
- [29] SEEGER M. Gaussian processes for machine learning[J]. International Journal of Neural Systems, 2004, 14(2): 69-106.
- [30] TIAN Y, COMPERE M. A case study on visual-inertial odometry using supervised, semi-supervised and unsupervised learning methods[C]//Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality, San Diego, Dec 9-11, 2019. Piscataway: IEEE, 2019: 203-207.
- [31] SHAMWELL E J, LEUNG S, NOTHWANG W D. Vision-aided absolute trajectory estimation using an unsupervised deep network with online error correction[C]//Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Oct 1-5, 2018. Piscataway: IEEE, 2018: 2524-2531.
- [32] 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, 2019, 42(10): 2478-2493.
- [33] LINDGREN K, LEUNG S, NOTHWANG W D, et al. BoomVIO: bootstrapped monocular visual-inertial odometry with absolute trajectory estimation through unsupervised deep learning[C]//Proceedings of the 19th International Conference on Advanced Robotics, Belo Horizonte, Dec 2-6, 2019. Piscataway: IEEE, 2019: 516-522.
- [34] HAN L, LIN Y, DU G, et al. DeepVIO: self-supervised deep learning of monocular visual inertial odometry using 3D geometric constraints[C]//Proceedings of the 2019 IEEE/RSJ

- International Conference on Intelligent Robots and Systems, Macau, China, Nov 3-8, 2019. Piscataway: IEEE, 2019: 6906-6913.
- [35] WEI P, HUA G, HUANG W, et al. Unsupervised monocular visual-inertial odometry network[C]//Proceedings of the 29th International Joint Conferences on Artificial Intelligence, Yokohama, Jan 7-15, 2020: 2347-2354.
- [36] ZHANG Z, DERICHE R, FAUGERAS O, et al. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry[J]. Artificial Intelligence, 1995, 78(1/2): 87-119.
- [37] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: learning optical flow with convolutional networks[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Piscataway: IEEE, 2015: 2758-2766.
- [38] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Piscataway: IEEE, 2017: 1647-1655.
- [39] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30, Long Beach, Dec 4-9, 2017. Red Hook: Curran Associates, 2017: 5998-6008.
- [40] JIE H, LI S, GANG S. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [41] WEYJTJENS H, WEERDT J D. Process outcome prediction: CNN vs. LSTM (with attention) [C]//Proceedings of the 2020 International Conference on Business Process Management, Seville, Sep 13-18, 2020. Cham: Springer, 2020: 321-333.
- [42] DONG N, ZHAO L, WU C H, et al. Inception v3 based cervical cell classification combined with artificially extracted features[J]. Applied Soft Computing, 2020, 93: 106311.
- [43] XUE F, WANG Q, WANG X, et al. Guided feature selection for deep visual odometry[C]//LNCS 11366: Proceedings of the 14th Asian Conference on Computer Vision, Perth, Dec 2-6, 2018. Cham: Springer, 2018: 293-308.
- [44] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, Jun 16-21, 2012. Piscataway: IEEE, 2012: 3354-3361.
- [45] BARRON J T. A general and adaptive robust loss function [C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 15-20, 2019. Piscataway: IEEE, 2019: 4331-4339.
- [46] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [47] BLANCO-CLARACO J L, MORENO-DUENAS F A, GONZÁLEZ-JIMÉNEZ J. The Málaga urban dataset: high-rate stereo and LiDAR in a realistic urban scenario[J]. International Journal of Robotics Research, 2014, 33(2): 207-214.
- [48] CARLEVARIS-BIANCO N, USHANI A K, EUSTICE R M. University of Michigan North Campus long-term vision and LIDAR dataset[J]. The International Journal of Robotics Research, 2016, 35(9): 1023-1035.
- [49] MAJDIK A L, TILL C, SCARAMUZZA D. The Zurich urban micro aerial vehicle dataset[J]. The International Journal of Robotics Research, 2017, 36(3): 269-273.
- [50] MILLER M, CHUNG S J, HUTCHINSON S. The visual-inertial canoe dataset[J]. The International Journal of Robotics Research, 2018, 37(1): 13-20.
- [51] CHEN W, LIU Z, ZHAO H, et al. CUHK-AHU dataset: promoting practical self-driving applications in the complex airport logistics, hill and urban environments[C]//Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Oct 24, 2020. Piscataway: IEEE, 2020: 4283-4288.
- [52] SCHUBERT D, GOLL T, DEMMEL N, et al. The TUM VI benchmark for evaluating visual-inertial odometry[C]//Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Oct 1-5, 2018. Piscataway: IEEE, 2018: 1680-1687.
- [53] PFROMMER B, SANKET N, DANIILIDIS K, et al. Penn-COSYVIO: a challenging visual inertial odometry benchmark[C]//Proceedings of the 2017 IEEE International Conference on Robotics and Automation, Singapore, May 29-Jun 3, 2017. Piscataway: IEEE, 2017: 3847-3854.
- [54] REINA S C, SOLIN A, RAHTU E, et al. ADVIO: an authentic dataset for visual-inertial odometry[C]//LNCS 11214: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 425-440.
- [55] JINYU L, BANGBANG Y, DANPENG C, et al. Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality[J]. Virtual Reality & Intelligent Hardware, 2019, 1(4): 386-410.
- [56] WANG C, ZHAO Y, GUO J, et al. NEAR: the NetEase AR oriented visual inertial dataset[C]//Proceedings of the 2019

- IEEE International Symposium on Mixed and Augmented Reality Adjunct, Beijing, Oct 10-18, 2019. Piscataway: IEEE, 2019: 366-371.
- [57] ZUÑIGA-NOËL D, JAENAL A, GOMEZ-OJEDA R, et al. The UMA-VI dataset: visual-inertial odometry in low-textured and dynamic illumination environments[J]. The International Journal of Robotics Research, 2020, 39(9): 1052-1060.
- [58] SONG Y, QIAN J, MIAO R, et al. HAUD: a high-accuracy underwater dataset for visual-inertial odometry[C]//Proceedings of the 20th IEEE Sensors, Australia, Oct 31-Nov 3, 2021. Piscataway: IEEE, 2021: 1-4.
- [59] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets[J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [60] FERRERA M, CREUZE V, MORAS J, et al. AQUALOC: an underwater dataset for visual-inertial-pressure localization[J]. The International Journal of Robotics Research, 2019, 38(14): 1549-1559.
- [61] ANTONINI A, GUERRA W, MURALI V, et al. The blackbird UAV dataset[J]. The International Journal of Robotics Research, 2020, 39(10/11): 1346-1364.
- [62] CAO L, LING J, XIAO X. The WHU rolling shutter visual-inertial dataset[J]. IEEE Access, 2020, 8: 50771-50779.
- [63] MINODA K, SCHILLING F, WÜEST V, et al. VIODE: a simulated dataset to address the challenges of visual-inertial odometry in dynamic environments[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1343-1350.
- [64] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [65] SUN K, MOHTA K, PFROMMER B. Robust stereo visual inertial odometry for fast autonomous flight[J]. IEEE Robo-

tics and Automation Letters, 2018, 3(2): 965-972.

- [66] 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.



王文森(1998—),男,山东人,硕士研究生,主要研究方向为惯性导航、信息融合技术等。

WANG Wensen, born in 1998, M.S. candidate. His research interests include inertial navigation, information fusion, etc.



黄凤荣(1969—),女,河北人,博士,高级工程师,主要研究方向为惯性导航、信息融合技术等。

HUANG Fengrong, born in 1969, Ph.D., senior engineer. Her research interests include inertial navigation, information fusion, etc.



王旭(1988—),男,天津人,主要研究方向为导航定位技术。

WANG Xu, born in 1988. His research interest is navigation and positioning.



刘庆麟(1998—),男,河北人,硕士研究生,主要研究方向为行人导航。

LIU Qinglin, born in 1998, M.S. candidate. His research interest is pedestrian navigation.



羿博珩(1998—),男,河北人,硕士研究生,主要研究方向为车辆导航。

YI Boheng, born in 1998, M.S. candidate. His research interest is vehicle navigation.