1、

视觉与 IMU 进行融合之后有何优势？

IMU 测量精度可以很高（成本不同、精度不同，精度高时成本大大增加）；测量频率也快；但是由于自身构造特性和工作特性的原因，由 IMU 测量出的结果进行积分进而获得的平移和旋转测量容易漂移，例如自身温度、零偏、振动等都是产生漂移的原因，同时这种漂移是无法通过更多的测量数据矫正回来的。

视觉的测量过程不需要积分，直接测量旋转与平移，会一定程度上避免 IMU 存在的误差导致漂移的问题，且自身工作原理上也不存在漂移问题，且视觉的测量结果可以通过不同图像的特征提取与匹配纠正测量结果；视觉的测量频率比较低，当环境中有运动的物体时就会对对相机的测量产生影响；而且视觉是通过对图像处理估计相机运动，图像处理的过程容易受到环境因素的影响，例如光照、图像的色彩；同时还有一些特殊问题，如果是成本低的单目相机的话，存在尺度问题，当单目相机是纯旋转运动时，无法通过三角测量确定尺度。

IMU 可以弥补视觉的频率低、动态物体影响测量以及光照条件影响测量的问题（光照在不同场景下变化还是比较大的），而视觉可以弥补 IMU 成本过高、无法避免的漂移问题；我们通过结合二者的测量结果，可以是松耦合的方式，也可以是紧耦合的方式（松耦合是两部分的测量结果已经产生，然后再进行最后的估计；紧耦合是实时结合两部分信息，得出一个测量结果，以上只是个人理解，具体实现过程尚不清楚），这样的互补关系可以让 VIO 对短时间快速的运动和长时间慢速的运动的测量都具有一定的鲁棒性和足够的精准度，而最终产生的测量结果对两部分信息各自的依赖情况则催生了不同的 VIO 方案，也具有各自的特点，这就需要后续的学习。

有哪些常见的视觉+IMU 融合方案？有没有工业界应用的例子？

VIO 方案分为紧耦合和松耦合方案，紧耦合方案为多。紧耦合方案分为基于滤波的方法(Filtering-based approaches)和基于优化的算法(Optimization-based approaches)。

基于滤波的方法：

MSCKF 系列（Multi-State Constraint Kalman Filter（多状态约束下的 Kalman 滤波器））：

MSCKF 在 EKF 框架下融合 IMU 和视觉信息，相较于单纯的 VO 算法，MSCKF 能够适应更剧烈的运动、一定时间的纹理缺失等，具有更高的鲁棒性；相较于基于优化的 VIO 算法（VINS，OKVIS），MSCKF 精度相当，速度更快，适合在计算资源有限的嵌入式平台运行。

ROVIO：

OpenVINS 系列：

基于优化的方法：

VINS-MONO 系列：

ORB-SLAM3：

OKVIS：

工业界应用的例子

Google Project Tango（做 Android 手机原型机进行 3D 建模）用了 MSCKF 进行位姿估计。

苹果的 ARKit（增强现实应用开发平台），谷歌的 ARCore（增强现实开发平台）

无人机利用 VIO 定位，追随规划的轨迹，避障，做悬停，拍照防抖。

移动设备中的应用：如手机端的室内定位，同样也有 AR 增强现实的应用。

在学术界，VIO 研究有哪些新进展？有没有将学习方法用到 VIO 中的例子？

VIO 目前有两个研究方向：

- 研究新的融合模式增强系统的鲁棒性和适用性。如：加入激光雷达，轮速计，GNSS 等传感器，具有代表性的工作有 LVI-SAM，GVINS，R3Live。

- 研究新的视觉约束和新的视觉特征提取来提高系统的精度。如：使用线，曲线，平面特征来构建残差。

将学习应用到 VIO 上有以下例子：

- 使用过深度学习来替换传统 SLAM 中的各个模块，如 SuperPoint 替换特征提取模块，SuperGlue 替换回环检测模块。

- 加入语义信息。

- 基于深度学习的端到端位姿估计。

以下来自 19 年知乎回答：https://zhuanlan.zhihu.com/p/68627439

传统方法新进展：

19 年论文 Visual-Inertial Mapping with Non-Linear Factor Recovery

通过重建非线因子图，将回环约束加入到因子图中，进行全局非线性优化，与 SOTA 的方案进行对比实验（目前看不懂）

19 年论文 VIL-VIO：Stereo Visual Inertial LiDAR Simultaneous Localization and Mapping

论文中指出，尽管 LiDAR-based 的方法本身就已经具有较高的精度，为什么还需要结合其他传感器？因为 LiDAR 在穿越隧道时会失效。系统由紧耦合双目 VIO 和 LiDAR mapping，以及 LiDAR enhanced visual loop closure. 与 SOTA 的 LiDAR 方法比起来，论文提出的方法健壮性和精度都有提升。

基于学习方法的例子：

1 Shamwell, E. Jared, et al. "Unsupervised Deep Visual-Inertial Odometry with Online Error Correction for RGB-D Imagery." IEEE transactions on pattern analysis and machine intelligence (2019).

学会了在没有惯性测量单元（IMU）内在参数或 IMU 和摄像机之间的外部校准的情况下执行视觉惯性里程计（VIO）

网络学习整合 IMU 测量并生成假设轨迹，然后根据相对于像素坐标的空间网格的缩放图像投影误差的雅可比行列式在线校正。

2 Chen, Changhao, et al. "Selective Sensor Fusion for Neural Visual-Inertial Odometry." arXiv preprint arXiv:1903.01534 (2019).

该论文 CVPR2019 已经接收。

论文集中在如何学习多传感器融合策略上。提出了一种针对单目 VIO 的端到端的多传感器选择融合策略。

具体是指提出了两种基于不同掩蔽策略(masking strategies)的融合模态：确定性软融合和随机硬融合，并与先前提出的直接融合 baseline 进行比较。在测试期间，网络能够选择性地处理可用传感器模态的特征并且产生确定尺度下的轨迹。

在 MAV 和 hand-held VIO 数据集上的测试表明了论文提出的融合策略相对直接融合，具有更好的性能，特别是在传感器由损坏数据的情况下。

此外，通过可视化不同场景中的掩蔽层和不同的数据损坏来研究融合网络的可解释性，揭示融合网络与不完美的传感输入数据之间的相关性。

3 Lee, Hongyun, Matthew McCrink, and James W. Gregory. "Visual-Inertial Odometry for Unmanned Aerial Vehicle using Deep Learning." AIAA Scitech 2019 Forum. 2019.

这篇文章针对传统的 VIO 需要进行标定的问题，提出一种网络，可以不需要标定

4 Wang, Chengze, Yuan Yuan, and Qi Wang. "Learning by Inertia: Self-supervised Monocular Visual Odometry for Road Vehicles." ICASSP 2019-2019 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

这篇论文提出了 inertia-embedded deep visual odometry，通过构建惯性误差函数对 VO 进行自监督的训练。

个人理解：

VIO 目前主要的研究方向和进展包括传感器信息的获取、传感器信息的融合、不同场景下利用 VIO 的优劣三方面。

传感器信息的获取可以通过添加新的传感器来进行创新，也可以通过对传感器获取的信息采用新的分析方法。

传感器信息的融合就比较多样，可能还是主要研究紧耦合，虽然主要就是滤波和优化两种方法，但是细分领域还是有很多的，需要更强的数学知识应用能力，也要考虑到计算成本的问题。

目前 VIO 肯定还是无法应用的所有场景的，但是对于解决单一场景下，或者说应用在特定领域中还是可以满足要求的，例如无人机在森林里，汽车在高速路上，停车场自动泊车的场景等等，所以新的论文很可能是针对不同的场景进行方案的尝试，选取最优方案。

而我对于学习的理解是，只要是和数学模型相关的，都可以通过建立数据集，然后建立各种各样的机器学习模型来训练，所以 VIO 自然也不例外，只是学习模型换了，那自然就会有不同的应用例子了。

2、

理解注释在源代码中

代码运行结果：

3、(另一个 PDF“第一章推导作业中”)