1. GEN-SLAM - Generative Modeling for Monocular Simultaneous Localization and Mapping

主要内容：移动机器人需要解决定位并估计到周围各种障碍物的距离的双重任务，以规划通过环境的无碰撞路径。我们提出了一种基于深度学习的模型，该模型仅使用单目视觉解决了这两项任务。我们的深度生成模型使用传统的视觉SLAM进行训练，并且在测试时，接收RGB图像，输出其拓扑姿势，并且还根据该姿势生成该图像的深度图

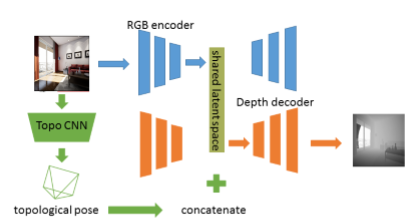


图1  GEN-SLAM架构，一种在测试配置中具有共享潜在空间的条件VAE。配对的RGB和深度VAE分别以蓝色和橙色表示。RGB图像通过TopoCNN以获得拓扑姿势，同时通过RGB编码器获得该图像的潜在矢量。该共享潜在向量与拓扑姿势连接，并通过深度解码器解码，以获得RGB图像的深度图

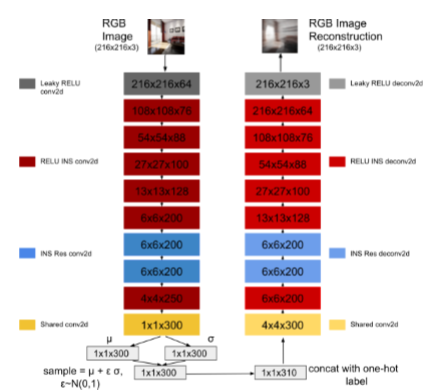


图2  RGB VAE每层的输出张量大小。层根据类型着色，与相应的解码器层（第二列）相比，编码器层（第一列）是更暗的阴影。深度VAE（此处未显示）是相同的，并与RGB VAE共享黄色卷积层。图像所属的地图中的拓扑节点与来自编码器的样本连接为单热矢量。

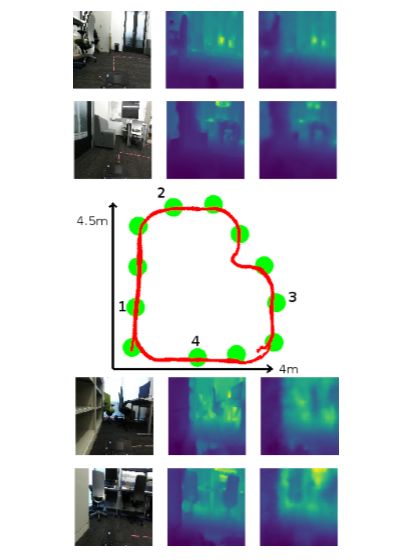
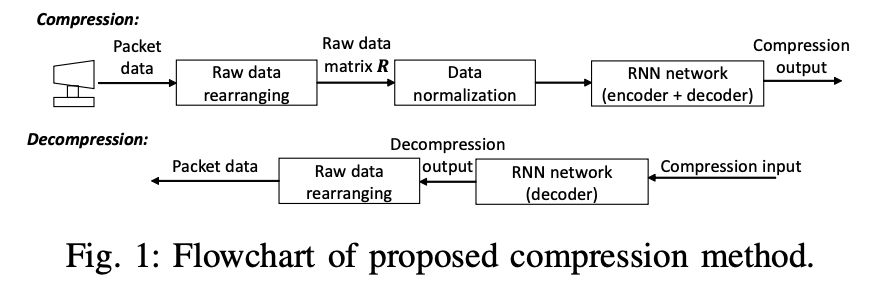
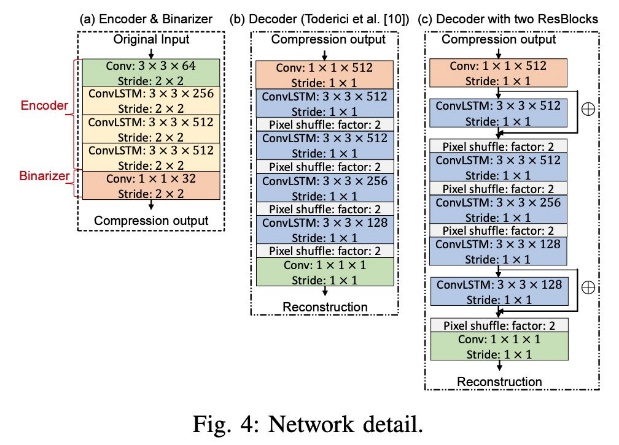
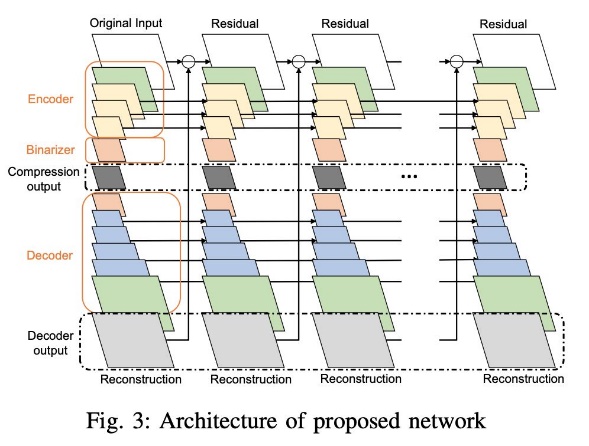


图3  Lab数据集中的拓扑定位和深度图估计。沿着行：rgb，地面实况和地图中4个拓扑节点的估计深度。度量本地化路径（红色）从ORBSLAM2获得。

1. Point Cloud Compression for 3D LiDAR Sensor using Recurrent Neural Network with Residual Blocks

主要内容：在自动驾驶领域已经证明了其实用性的3维LiDAR，目前已经将他的使用延伸到了许多其他的领域。从3维LiDAR传感器中分享及传输得到的点云数据在机器人领域有着广阔的前景。然而，由于这些数据的稀疏性以及无序性，想要直接将他压缩到一个很小的体量是非常困难的。一种潜在的解决方法是利用未处理的LiDAR数据。我们可以把这些从每一帧中得到的生数据无损的重新排列程一个2D矩阵，使得数据更加的紧凑和有序。由于3D LiDAR数据的特殊结构，这些重构的2D矩阵的纹理相对相机照片的2D矩阵来说也是不规则的。为了能够有效的压缩这种2D格式LiDAR生数据，作者在本篇论文中提出了一种使用循环神经网络和残差块来渐进的压缩从LiDAR获得的一帧图像信息的方法



1. Localizing Discriminative Visual Landmarks for Place Recognition

主要内容：

本文解决了具有感知变化的视觉位置识别问题。视觉位置识别的基本问题是生成鲁棒的图像表示，该图像表示不仅对环境变化不敏感，而且可以区分不同的位置。利用卷积神经网络（CNN）的特征提取能力，我们进一步研究如何定位对相似性测量有积极贡献的有鉴别性的视觉路标，例如建筑物和植被。特别地，我们设计了一个路标定位网络（LLN）来表示图像的哪些区域用于识别。在具有不同外观和视角变化的开源数据集上进行了详细的实验。与最先进的方法相比，所提方法具有卓越的性能。

主要贡献：

1. 提出了一个路标定位网络（LLN）来识别有鉴别性的图像区域以便进行位置识别。

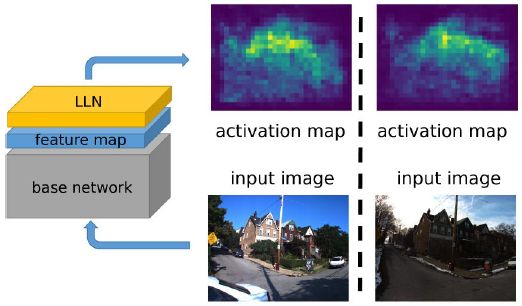
2. 采用度量学习策略和难分样本挖掘策略对网络进行训练。

3. 用一组有鉴别意义的路标来描述每个图像，以利用局部不变特征测量图像相似性。

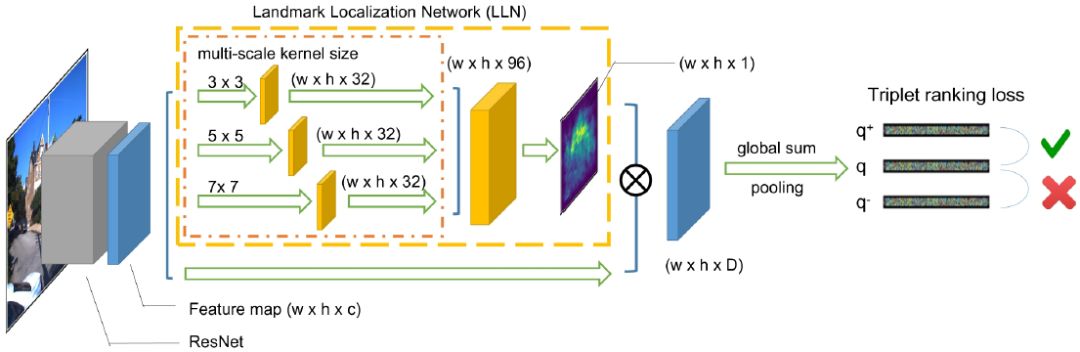
4. 在几个数据集上对所提出的方法进行了评估，并将整个过程集成到ORB-SLAM中。大量的实验结果表明，该方法可以在季节变化情况下有效地提高定位性能。

实现方法：

1. 路标定位网络（LLN）



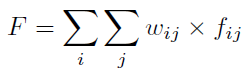
使用ResNet101模型作为基础网络以便进行稠密特征提取，其第三个卷积块的输出特征图（feature map）作为LLN的输入。为了对视点变化和部分遮挡产生更好的不变性，在LLN中使用了具有不同内核大小的多尺度滤波器。将多尺度滤波器的输出级联形成w x h x 96的特征映射。每个空间位置的所有激活与1x1xD的卷积层结合起来，再进行ReLU激活生成激活图。激活图定位有鉴别性的图像区域，用于描述图像。



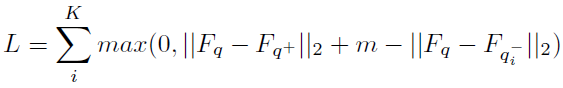
训练

为了计算度量损失，需要生成整个图像的嵌入。这是通过对所有局部特征fij加权求和来完成的：

权重wij通过LLN进行预测。将这些局部特征聚合之后，再进行L2规范化。



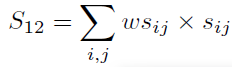
    对于每个查询图像q，我们通过比较聚合的特征F对所有数据集图像进行排序。非匹配图像q-是属于不同位置但与查询图像相似的前K个图像。我们提供图像元组到网络,每个元组包含6个图片,包括1个查询图像q,1个匹配图像q+和4个未匹配图像q-。损失函数基于三元排序损失:



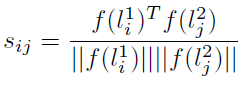
通过最小化该损失函数，该网络可以学习为有鉴别性的视觉路标分配更高的激活度。

3. 相似性测量

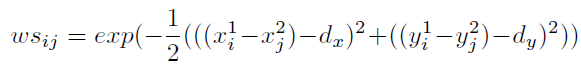
    根据LLN的结果，选择激活度最高的前n个局部特征来产生图像表达。两个图像I1和I2之间的相似性取决于它们匹配的局部特征的加权和：



两个图象局部特征的相似性计算公式：



权重：



1. FlashFusion: Real-time Globally Consistent Dense 3D Reconstruction using CPU Computing

主要内容：针对便携式设备上稠密3D重建的实际应用，我们提出了FlashFusion，一种快速的大规模的、高分辨率（亚厘米级）3D重建系统，并且无需使用GPU。它通过鲁棒且快速的全局集束调整方案实现全局一致的定位。并且，他实现基于空间散列的体积融合，以300Hz运行，还实现了基于过高效的有效块选择和纹理提取方案的渲染，以25Hz运行。现实世界和已有数据集的广泛实验表明，FlashFusion成功实现了在受限的计算能力下（即仅用便携式设备上的CPU），实现实时、全局一致、高分辨率（5mm）的大规模稠密3D重建。

主要贡献：

1.通过在预积分阶段使用Huber范数来提高定位的鲁棒性。

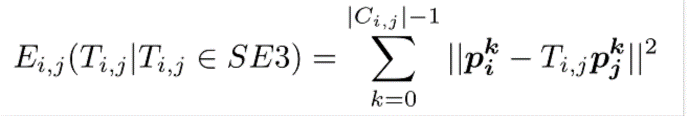
2.加速了TSDF滤波，达到了300HZ。

   3.通过三个阶段加速纹理提取以达到25HZ。

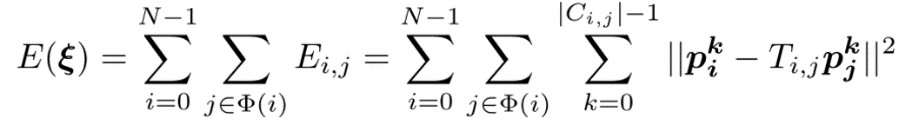
实现细节：

 1.鲁棒的全局一致的定位

提取相邻两帧的ORB特征，通过使用李群最小化流形上的代价函数：

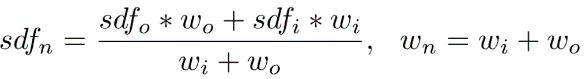


因为该代价函数对异常值敏感，因此，在代价函数中，用Huber范式代替了L2范式。为了达到实时，只选择相对位姿变化最明显的10个相对应的关键帧进行优化。为了实现全局一致的位姿估计，我们最小化所有关联的关键帧对的重投影误差



2、高效的TSDF滤波：

2.1.TSDF表达：3D点映射为到最近表面的距离，记为符号距离函数(SDF)。深度观察使用投影映射方式集成，投影距离和深度地图值的插值记为表面距离d，并且被融合到TSDF域：



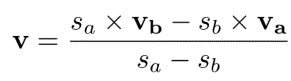
 2.2. 有效的数据选择：所有落入相机视角的视锥的数据块都是候选数据块，然后，滤除掉那些无效的数据块（这些数据块的体素没有有效的sdf值）。

 2.3. 基于关键帧的优化, FlashFusion中颜色值为实际的颜色值和权值相乘的结果。

    3.纹理提取

    3.1.多边形生成：

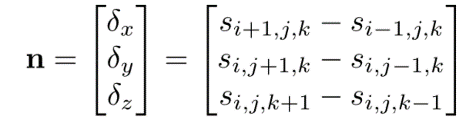
    每个顶点的准确的位置v通过两个端点Va、Vb的线性插值



为了判断一个体素中是否存在多边形，采取了动态阈值的策略

 3.2.法线提取

    每个顶点的发现通过对TSDF域求导：



1. Self-Supervised Monocular Depth Hints

主要内容：

像素点的深度真值是很难获得的，自监督单目深度估计网络的出现是解决这个难题一个很有希望的办法，本文作者提出了一系列的改进，与现有的自监督方法相比，这些改进共同作用得到了定量和定性改进的深度图。自监督单目深度网络的研究通常探究日益复杂的框架，损失函数和图像形成模型，这些有效地缩进了与有监督方法的差距。我们表明一个意外简单的模型和相关的设计选择，可以带来不俗的预测效果。特别是我们提出，（i）最小重投影损失，目的是稳健地处理遮挡情况，（ii）减少视觉伪像的全分辨率多尺度采样方法，（iii）自动掩盖损失以忽略那些违反相机运动假设的训练像素。我们分别展示了每个模块的有效性，并在KITTI基准上显示了高质量的最新结果。

主要贡献：

1）我明已有的自监督递归网络在最小化光度重投影误差时，会陷入无法找到最优解的窘境中。

2）证明了使用深度提示的选择性训练是一种通用的提升手段，它可以改进多个网络，使他们达到最小值。而这个深度提示来自于双目图像。

3）发现带有深度提示的选择性训练结合明智的网络设计选择，使我们的方法超过了其他算法。我们在KITTI数据集上取得了高质量的结果，不仅超过了我们之前的基准模型，也超过了那些已发表的算法。

主要方法

1）训练双目图片对

在训练期间，我们为网络提供每个像素 “深度提示”，即网络自身深度估计的潜在替代假设。关键思想是，我们只想在深度提示的地方提供监控信号， 与使用网络预测相比以使它们具有出色的重新投影图像。否则，提示将被忽略。需要明确的是，提出的目标不是在学习回归暗示深度值的图。

2）计算深度提示

  作者提出使用双目图片对得到“深度提示”，其实就是通过Semi-Global Matching这样一个双目匹配算法去产生深度图。论文中把这种方法叫做“RandomSGM” 。另外又提出对于每个训练图像对，我们可以通过使用每个可能的超参数选择运行SGM来生成深度图的集合。我们将此空间离散化为12个参数选择，这些参数选择由三种块大小和四种分辨率的组合组成  我们将此版本的深度提示称为“FusedSGM”.值得注意的是作者模型使用的就是第二种(Fused SGM)，但是给出了两种匹配算法的效果对比。

3）实现细节

作者的网络使用的是encoder-decoder结构，encoder是比较经典的resnet-18,也是许多单目估计网络的选择。作者在ImageNet进行了预训练并且还表明图片分辨率对精确度有很直接的影响。