

第1章 基于深度学习的路面车辆视觉里程计

在前两章中，我们分别从损失函数（第??章）和网络架构（第??）的角度对基于学习的视觉里程计方法进行了改进。我们知道在深度学习中，存在关键的三个因素：损失函数、网络架构和训练数据。本文我们将从路面车辆的运动约束和数据分布的角度对基于学习的视觉里程计模型进行改进。大部分基于深度的视觉里程计算法尝试学习从连续图像所组成的图像对到相机运动的映射模型，其中相机运动一般由三自由度的平移运动和三自由度旋转运动组成。然而，我们发现，对于路面轮式车辆系统来说，由于运动模式受其自身机械结构和动力机制约束，其在三维空间的运动并不具备完整的六自由度，大部分运动局限在 z 轴方向的平移运动和 y 轴方向的旋转运动，即用于训练网络的数据在各个维度上分布不均，我们认为这是当前基于学习的视觉里程计问题精度较低的原因之一。因为所有基于监督学习的问题都依赖大量的标签数据，如分类的问题的 Imagenet 数据集^[1] 和语义分割问题的 CityScapes 数据集^[2]。而对视觉视觉里程计问题，主流的数据集 KITTI 数据集^[3] 在数据量和数据多样性上都相对较低。在本文中，我们打算利用车辆的运动模型，仅建模车辆的主体运动维度，而忽略运动较小的自由度，并探索运动维度的聚焦和解耦对基于学习的视觉里程计问题的影响。

在基于几何计算的传统视觉里程计问题中，路面车辆的运动模型已经被广泛使用。由于车辆在 y 轴方向的平移运动幅度极小，固定在车辆上的相机的高度不易发生变化，包括本文其第二章在内的很多工作^[4-8] 以此作为绝对尺度参考恢复单目视觉运动估计的绝对尺度。Scaramuzza 等人^[9] 根据车辆 Ackermann 运动模型，简化车辆运动估计，提出基于单特征点的 RANSAC 运动估计，提高算法实时性。Choi 等人^[10] 在基于路面车辆模型的基础之上，考虑车辆震荡，放松了严格平面运动的约束，是算法更加鲁棒。此外 Scaramuzza 等人^[11] 还提出了基于单应性的全景相机视觉里程计。

本文第一次将车辆的运动模型引入到基于学习的单目视觉里程计算法中。在具体实现上，我们首先定量评价在忽略受限制的运动维度后车辆运动轨迹的偏移，并根据车辆的旋转模型对车辆的旋转运动和平移运动解耦，以减弱在忽略受限制维度运动时的轨迹便宜。同时我们设计并构建了用于学习车辆的主要运动的轻量化的卷积神经网络模型。基于对上述工作的研究，本文做出如下贡献：

1. 我们通过实验定量地分析了运动聚焦所造成的轨迹偏移，实验发现轨迹偏移相对很小，证明了运动聚焦的可行性；
2. 我们根据车辆的运动模型，分析了车辆沿 x 轴的平移运动与绕 z 轴旋转运

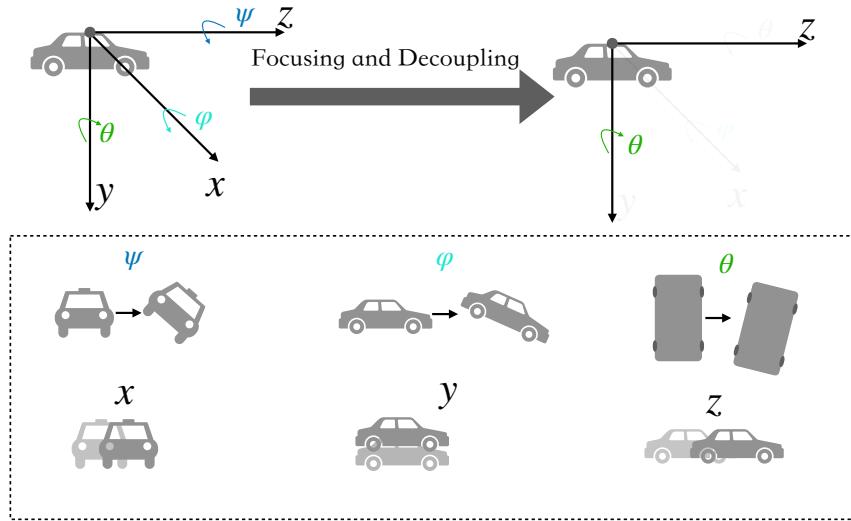


图 1.1 车辆运动简化

动之间的映射关系，并通过运动解耦减低了运动聚焦时的运动偏移；

3. 我们实验证明了提出运动聚焦和运动解耦可以提高基于学习的单目视觉里程计的性能，包括减少训练时间、提升训练精度；
4. 我们构建了一个十分轻量化的运动估计网络，该网络在训练时仅需占用 2G 的 GPU 显存并可以快速收敛，此外并可以在 CPU 上实时运行，且达到与其他复杂网络的精度，算法已开源^①.

本文结构如下：首先在第1.1.1节介绍算法的数学原理和实现方式；然后再第1.2节，我们在 KITTI 数据集^[3] 定性和定量的评价我们算法；最后我们在1.3节总结本章工作。

1.1 路面车辆视觉里程计方法

运动聚焦，即忽略车辆的小幅度运动，让网络聚焦于车辆的主体运动。根据车辆模型的约束，以简化运动估计为目的，我们提出了运动聚焦算法，但运动聚焦过程中会带来姿态偏移，于是我们提出了运动解耦以降低运动聚焦带来的姿态偏移。本节，我们首先在第1.1.1节介绍运动聚焦和运动解耦方法；然后再第1.1.2节，介绍模型架构已经训练和测试方法。

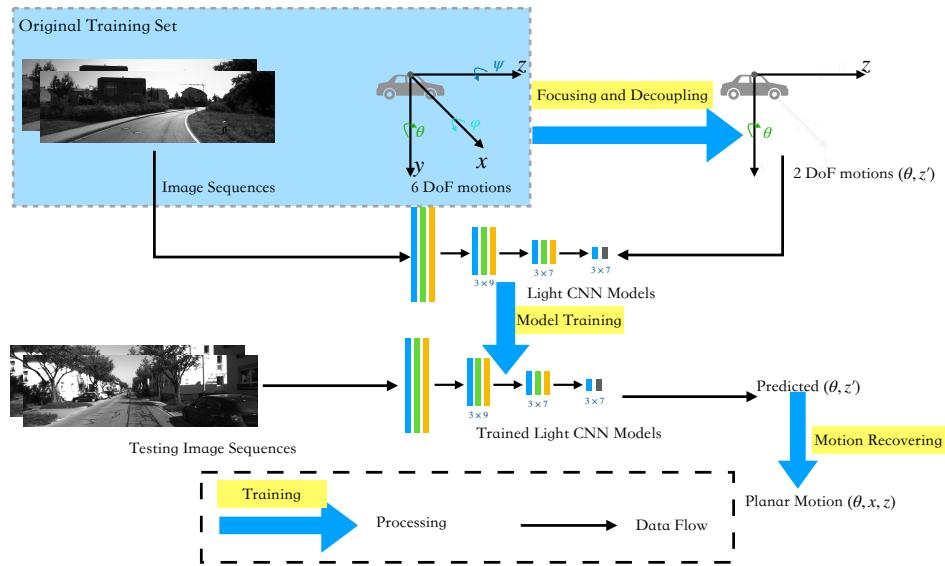


图 1.2 系统架构

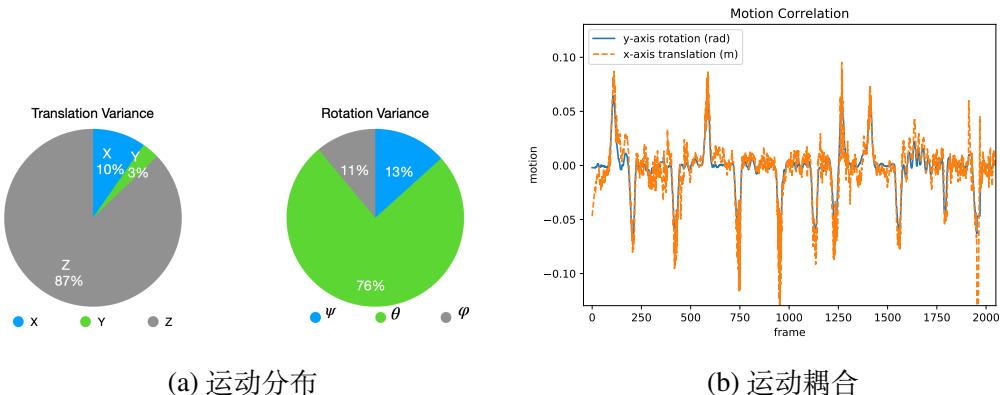


图 1.3 运动模态分析

1.1.1 运动聚焦与解耦

1.1.1.1 运动聚焦

根据车辆的机械结构和动力模型的限制，车辆的主要运动为沿 Z 轴的平移运动（前进）和绕 Y 轴的旋转运动（转向），为了通过数据证明这一论点，我们首先在 KITTI 视觉里程计数据集中定量的分析了车辆在各个维度上的运动方差，如图1.3(a)所示（在图1.3(a)中，我们仅可视化了 KITTI 数据集 00 序列的运动，为了说明其代表性，我们在附录??中可视化了更多的数据）。在运动表征方法，我们选择使用标准的相机坐标系，其为以相机光心为原点的右手坐标系，向前为 Z 轴方向，向右和向下分别 X 轴方向和 X 轴方法。车辆绕 X 轴、Y 轴和 Z 轴的旋转运动的欧拉角分别表示为 ψ , φ , and θ 。从图1.3(a)中可看出，车辆的平移运动确实主要集中在 Z 轴方向，而车辆的旋转运动主要集中在 Y 轴上。所以，我们

① <https://github.com/TimingSpace/DMVOGV>

选择仅建模输入图像到车辆 Z 轴平移运动和 Y 轴旋转运动的映射，但是车辆在 X 轴存在着不可忽视的平移运动，我们称这个过程为运动聚焦，如图1.1。

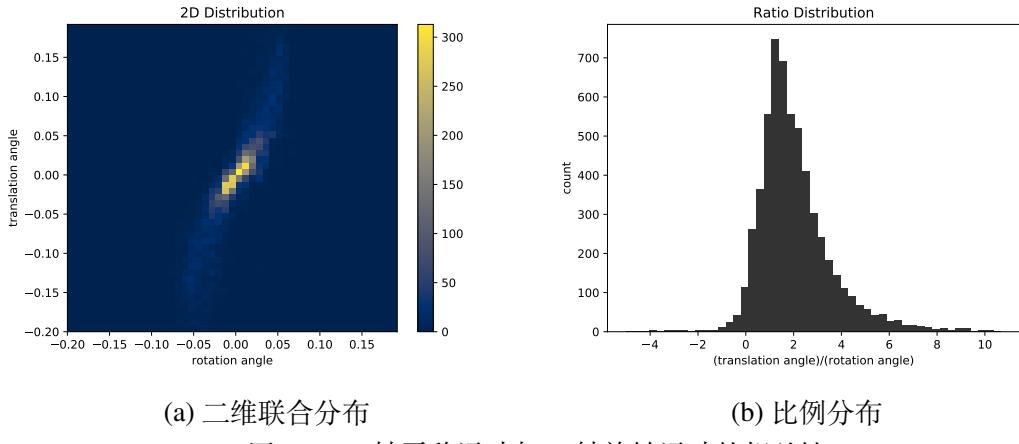


图 1.4 X 轴平移运动与 Y 轴旋转运动的相关性

1.1.1.2 运动解耦

然而，我们发现车辆依然存在着一定幅度（约 10%）的 X 轴平移运动。但由于动力约束，无人车系统从原理上是无法沿 X 轴有大幅度的运动的，那么这 10% 的轴平移运动来自哪里呢？通过仔细分析无人车的旋转模式，我们发现 X 轴的平移运动产生的原因为车辆的运动表征方式：

$$\begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ 0 & 1 \end{pmatrix} \quad (1.1)$$

其中 \mathbf{I} 为 3×3 的单位矩阵。在这种表征方式中，机器人进行首先平移运动 \mathbf{t} ，然后进行旋转运动 \mathbf{R} ，那么如果机器人同时进行了旋转和平移运动，那么车辆的参考坐标系会因为旋转运动发生变化，那么本来仅沿 Z 轴的平移运动也会由于旋转的存在映射成 Z 轴平移分量与 X 轴平移分量，如图1.5所示。我们将参考坐标系变化带来的平移偏角 α 定义为：

$$\alpha = \arctan\left(\frac{x}{z}\right) \quad (1.2)$$

其中 x 和 z 分别表示 X 轴和 Z 轴的平移运动。从图1.3(b)中可以看出，X 轴的平移运动与 X 轴的旋转运动相关性很强，这一点恰好证明 X 轴运动源自于旋转后运动映射。由于图1.3(b)只能评价局部的运动相关性，为了得到更具带表性的结果，我们通过直方图可视化旋转角度 θ 和平移偏角 α 的全局关系。图1.4(b)的一维直方图可视化了 α/θ 的分布曲线；图1.4(a)中的二维直方图可视化了 α 和 θ 的

联合概率分布。从两个直方图中都可以看出，车辆旋转角 θ 和平移偏角 α 有着很强的相关性。

那么如何改变运动表征方式来解除 X 轴平移运动与 Y 轴旋转运动之间的耦合呢？一种朴素的方法为调换旋转运动和平移运动的顺序，车辆先进行旋转运动，然后再旋转之后的坐标系下进行平移运动，这样车辆的平移运动就不会被映射到 X 轴上，这种运动表征方式可以公式化为：

$$\begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}' & \mathbf{0} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{t}' \\ 0 & 1 \end{pmatrix} \quad (1.3)$$

公式中 $\mathbf{R}' = \mathbf{R}$, $\mathbf{t}' = \mathbf{R}^{-1}\mathbf{t}$ ，相当于把由于旋转重映射的平移运动在反射回来。然而，如图1.5(a)所示，由旋转运动引发的车辆的平移偏角 α 其实并不等于车辆

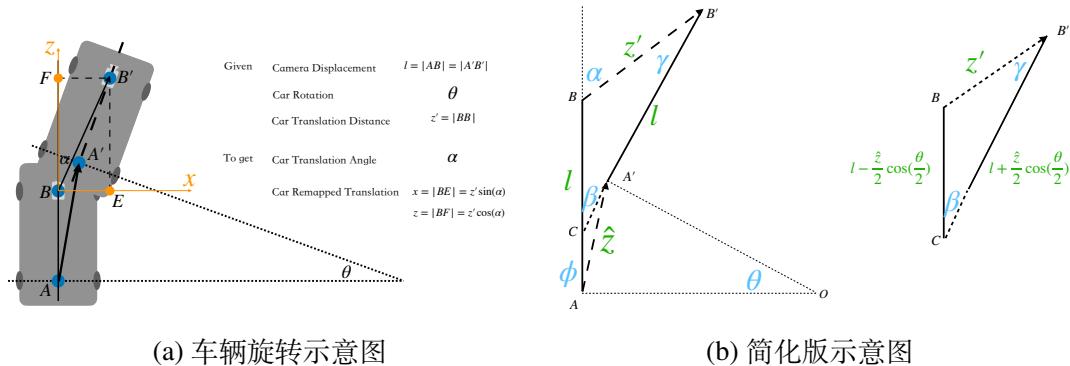


图 1.5 车辆旋转模型

的旋转角 θ 。我们的重映射首先需要确定平移偏角 α 与旋转角 θ 之间的关系，我们定义其间关系为 $\alpha = f(\theta)$ 。在已知这个关系的基础之上，我们可以车辆的前进距离和车辆的旋转角这两个运动参数来表征车车辆的平面运动，平面运动中的平移运动为

$$(x, z) = z(\sin(f(\theta)), \cos(f(\theta))) \quad (1.4)$$

图1.5(a)中，A 点表示车辆后轴的中心，B 点表示相机的安装位置，设定 A 点和 B 点之间的距离为 l ，称之为相机偏距。视觉里程计之间估计的运动为，B 点和 B' 之间的距离，记为 z' 。我们将图1.5(a)简化为图1.5(b)。

根据车辆 Ackermann 运动模型^[12]， $OA \perp AB$ 且 $OA' \perp A'B'$ ，于是可以得到 $\phi = 0.5\beta = 0.5\theta$ 。在三角形 CBB' 中，根据正弦定理：

$$\frac{\sin(\gamma)}{\sin(\beta)} = \frac{l - \frac{z}{2}/\cos(\frac{\theta}{2})}{z'} \quad (1.5)$$

考虑到 θ 很小接近 0，可做如下近似 $\cos(\frac{\theta}{2}) \approx 1$ 且 $\frac{\gamma}{\beta} \approx \frac{\sin(\gamma)}{\sin(\beta)}$ ，于是

$$\frac{\gamma}{\beta} \approx \frac{l - \frac{z}{2}}{z'} \quad (1.6)$$

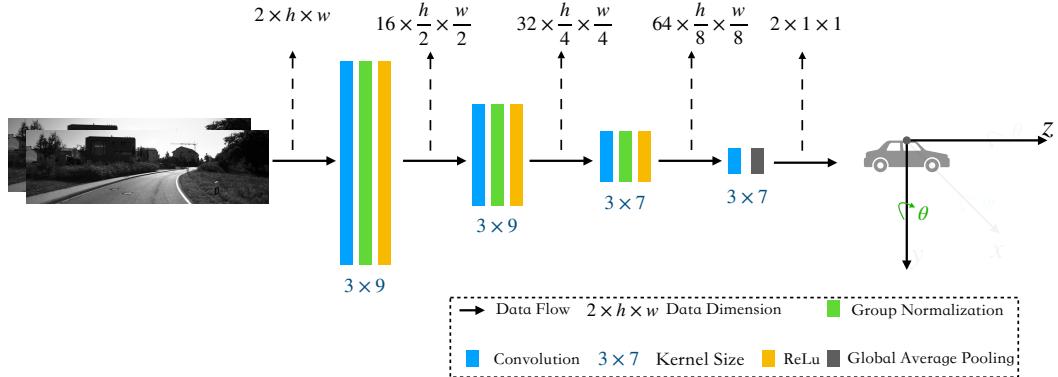


图 1.6 轻量化网络架构

记 $d = |AC| \approx 0.5|AA'| = 0.5\hat{z}$, 在三角形 CBB' 中根据余弦定理:

$$z'^2 = (l+d)^2 + (l-d)^2 - 2(l+d)(l-d)\cos(\beta) = 2l^2 + 2d^2 - 2(l^2 - d^2)\cos(\beta) \approx 4d^2 \quad (1.7)$$

所以 $z' \approx \hat{z}$. 最后我们得到了平移偏角 α 和旋转角 θ 之间的关系

$$\alpha = \beta + \gamma \approx \left(\frac{l}{z'} + 0.5\right)\beta = \left(\frac{l}{z'} + 0.5\right)\theta \quad (1.8)$$

我们根据平移偏角 α 构建旋转矩阵 R_α ,

$$\mathbf{R}_\alpha = \begin{pmatrix} \cos(\alpha) & 0 & \sin(\alpha) \\ 0 & 1 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix} \quad (1.9)$$

然后重映射平移向量 t'

$$\mathbf{t}' = \mathbf{R}_\alpha^{-1} \mathbf{t} \quad (1.10)$$

车辆的运动距离 z 为平移向量 t' 中的第三个分量。至此车辆的平面运动可以被车辆旋转角 θ 和前进距离 z 两个参数近似表示，本文我简化运动估计的目标，仅学习这两个运动参数。网络模型将会在下一章介绍，效果会在1.2.2.3进行评测。

1.1.2 网络模型与训练

我们构建了一个轻量化的网络来学习路面车辆的主要运动，网络架构如图1.6所示。网络主要有卷积层构成，除了最后一个卷积层以外，每个卷积后后面都附加一个归一化层^[13]和ReLU层。同周等人^[14]的架构一样，我们在网络的最后一层没有使用全连接层，而是使用全局均值池化层^[15]，用以降低参数量，增强泛化性。此外，我们观察到，车辆运动引发的图像光流，大部分为水平方向，尤其是在车辆旋转的情况下，如图1.7所示。所以我们没有使用正方形的卷积核，而是通过使用宽大于高的卷积核来扩大水平方向的感受野。此外，我们使用了膨胀卷积（Dilated Convolution）^[16]，进一步在参数量相同的情况下扩大感受野。

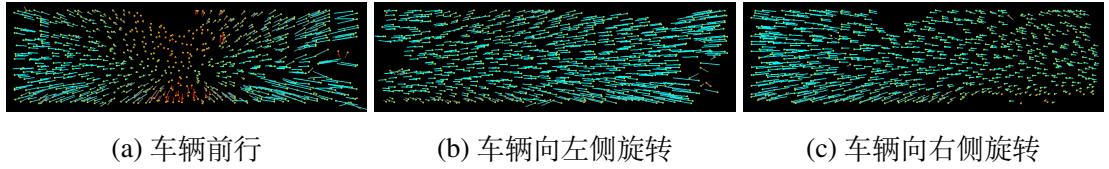


图 1.7 路面车辆运动时的光流分析

模型的输入为由灰度图像叠加而成图像对，我们不仅使用相邻帧图像组合成图图像（图像间隔为 0）对作为输入，我们随机生成闭区间 [-4,4] 之间的一个整数，然后使用这个数作为图像对中两个图像之间的间隔，当做一个数据增加的手段。模型的输出为与图像对对应的相机运动，表示为 Y 轴旋转运动 θ 和前进运动 z' 。其中 Y 轴旋转运动 θ 为旋转矩阵计算得来的欧拉角所对应的 Y 轴分量；前进运动 z' 可使用公式(1.10)计算。我们使用 L2 距离作为损失函数：

$$L_2 = \|\theta_g - \theta_p\|_2 + \|z_g - z_p\|_2 \quad (1.11)$$

其中 θ_p 和 z_p 为模型估计出的运动， θ_g 和 z_g 运动的真值。我们使用 ADAM 优化器^[17] 最小化损失函数获取模型参数，初始学习率设置为 0.001，在第 50 个周期之后，学习率线性衰减，至 100 个周期衰减为 xx。

模型训练成功之后，模型输入连续图像后可以得到与之对应的机器人旋转角 θ 前进距离 z 。我们首先根据公式(1.8)计算平移偏角 α ，并假设其他维度旋转均为 0，然后构建旋转矩阵 \mathbf{R}_θ 和 \mathbf{R}_α ，车辆的平移向量计算为 $\mathbf{t}_\alpha = \mathbf{R}_\alpha(0, 0, z)^T$ 。运动矩阵可表示为：

$$\mathbf{T}_i = \begin{pmatrix} \mathbf{R}_\theta & \mathbf{t}_\alpha \\ 0 & 1 \end{pmatrix} \quad (1.12)$$

车辆的位姿可以通过运动矩阵的累积获取：

$$\mathbf{P}_i = \mathbf{P}_{i-1} \mathbf{T}_i \quad (1.13)$$

1.2 路面车辆模型实验

我们在公开数据集 KITTI 视觉里程计数据^[3] 上对本文提出的算法进行评测和分析。我们首先介绍此数据集和我们的测试环境；然后我们依次介绍：位姿偏移评测、运动模型解耦、运动估计提升和与其他方法对比等四个实验；最后我们讨论和分析实验结果。

1.2.1 实验数据和测试平台

KITTI 数据集^[3] 共有 22 条测试序列，其中前 11 条序列提供位姿的真实值可以用于线下评测。对于每一条测试序列，KITTI 数据集提供了 RGB 图像、灰度图像和激光雷达点云等数据。我们仅使用数据集中的灰度图像以及对应的真实位姿用来测试算法，并使用其提供评价标准定量测量相对位置误差（RPE），其中包括相对平移误差和相对旋转误差^[3]。

本文算法使用 Python 程序语言实现，基于深度学习框架 PyTorch 进行网格模型搭建和模型训练，目前本文算法已经开源^①。本文算法的测试平台为一台具备 16GB 内存、因特尔酷睿 i7 (i7-7700 CPU @ 2.80GHz)、英伟达 GPU (GeForce GTX 1060) 的笔记本电脑上进行测试。测试系统为配备 CUDA10.0，Python 3.6.9 的 Ubuntu 操作系统。由于所提出模型的轻量化，当训练批大小 (Batch Size) 为 30 时，模型训练仅需要 2G 的显存；在预测时，本文算法可以仅在 CPU 上达到 200 帧每秒的实时效果。

1.2.2 实验结果

我们首先评价运动聚焦引发的位置偏移；然后评价运动解耦对位置偏移的减缓效果；之后我们评价运动聚焦和解耦以及本文的轻量化网络架构带来的算法性能提升；最后我们与其他算法进行比较。

1.2.2.1 运动聚集造成的运动偏移的定量评测

由于车辆运动受其机械结构和动力模型的约束，车辆主要运动集中于 Z 中上的平移运动和 Y 轴上的平移运动。此实验将定量分析在去除其他部分和全部次要运动时，车辆位姿的偏移。在忽略次要运动之后，我们重构机器人轨迹，然后使用相对位置误差评价位姿偏移。在 KITTI 序列 00 至序列 10 上的相对位置误差记录于表1.1，并可视化与图1.8中。表1.1中，每一行保留着相同的角度维度，每一列保留着相同的平移维度。但我们仅保留 Y 轴旋转运动和 Z 轴平移运动时，平均位置误差为 2.20%。我们将重构的轨迹可视化于图1.9。可见运动简化后的轨迹在水平面上依然与真实轨迹较为接近，但不同在竖直方向会因不同序列的高度变化产生不同的偏差。

我们求取平均位姿误差 (cost) 与忽略维度的数量 (gain) 求商，如1.8中的黄色线所示。这里比例系数可以作为一个系数平均相对损失。可见在仅保留 Z 轴平移运动和 Y 轴旋转运动时，相对损失比较小。

^① <https://github.com/TimingSpace/DMVOGV>

表 1.1 运动聚焦导致的位置偏移

| R / t | z | xz | yz | xyz |
|-------|------|------|------|------|
| y | 2.20 | 2.06 | 2.45 | 2.34 |
| xy | 1.92 | 1.77 | 1.76 | 1.56 |
| zy | 2.05 | 1.91 | 1.47 | 1.27 |
| xyz | 1.92 | 1.81 | 0.49 | 0 |

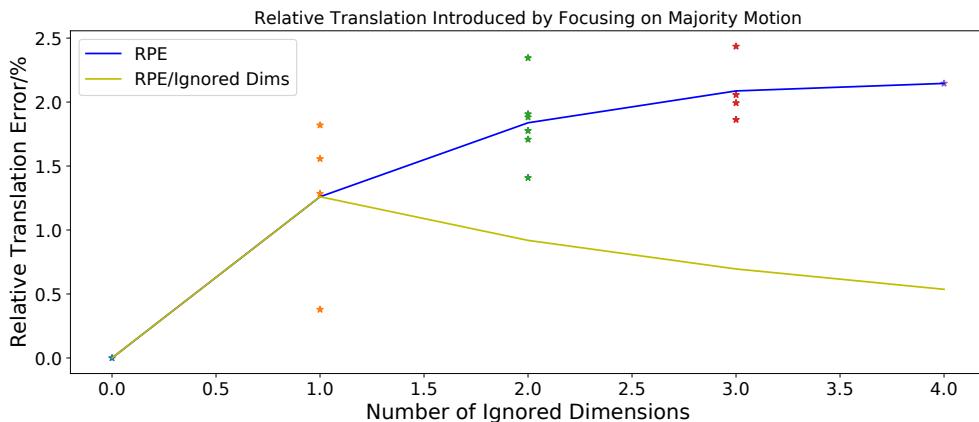
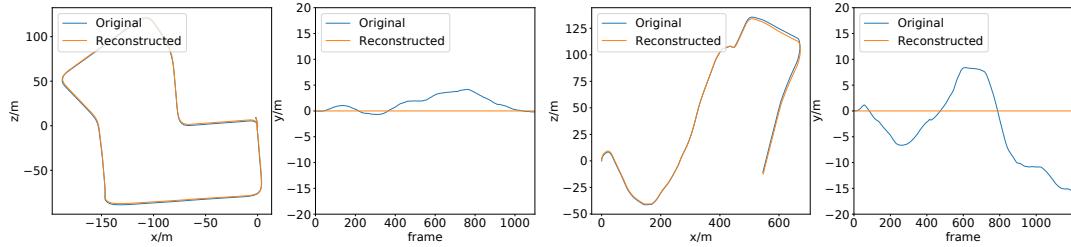


图 1.8 运动简化位置偏移评价

1.2.2.2 运动解耦评测

根据公式(1.8)，车辆的平移偏角 α 和车辆的旋转角 θ 为线性关系。然而，其比例系数依赖于动态的车辆前进距离，并不固定的。我们首先研究使用动态的比例系数的实验效果，我们使用不同的比例系数计算车辆偏移角，然后重映射车辆运动，得到的车辆评价位置误差可视化与图1.10(a)。从图中可以看出，当比例系数为 1.7 时，评价误差最小，说明旋转时车辆的运动距离约为 $\frac{1.7-0.5}{l}$ 米。为了更好的理解图中不同比例系数的物理意义，我们使用了不同颜色标出的几个特殊的比例系数。红色图表示比例系数为 0，此时平移运动不做任何映射，为原始的运动表示；比例系数为 1 时，认为平移偏角和旋转角度相同，根据公式(1.3)的朴素映射解耦；黑色表示平移偏角为旋转角度的一半，这种情况仅在相机安装在后轴中心时才成立，但相机距离后轴中心相对于车辆前进距离较小时，此系数也可近似成立。

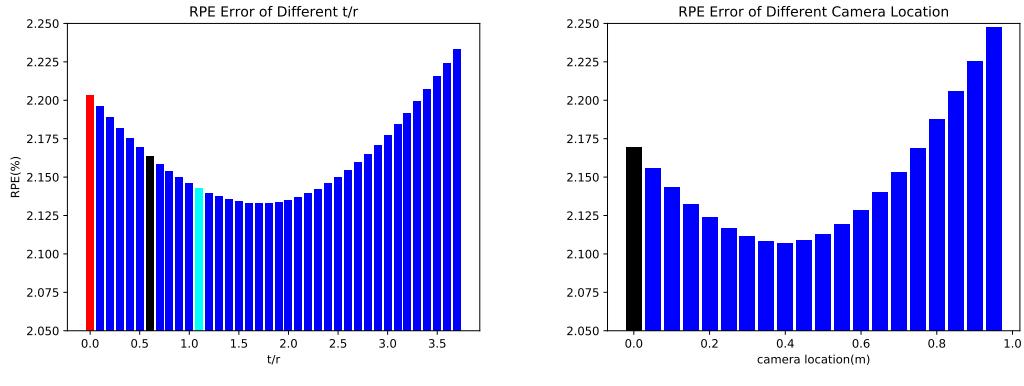
固定的比例系数无视了前进距离对平移偏角的影响，于是我们定量分析动态比例系数的位姿偏移。我们使用不同的相机偏距 l ，根据公式(1.8)，计算比例系数，然后去平均重构轨迹的相对位置误差，可视化于图 1.10(b)，但相机偏距为



(a) KITTI 07

(b) KITTI 10

图 1.9 运动聚焦后的轨迹重构



(a) 静态解耦

(b) 动态解耦

图 1.10 运动解耦效果可视化

0.4 米时，重构误差最小。图中黑色为相机偏距为 0 时的重构误差，此时 $\alpha = 0.5\theta$ ，和图 1.10(a) 中的黑色物理意义相同，数值相等。

我们将动态映射和静态映射的效果进行了比较，记录于图 1.11 中。可以发现，两种运动结构方式都降低了机器人轨迹的相对位置误差，动态结构相比于静态结构效果更好，更接近同时保留 X 轴平移和 Z 轴平移时的精度。

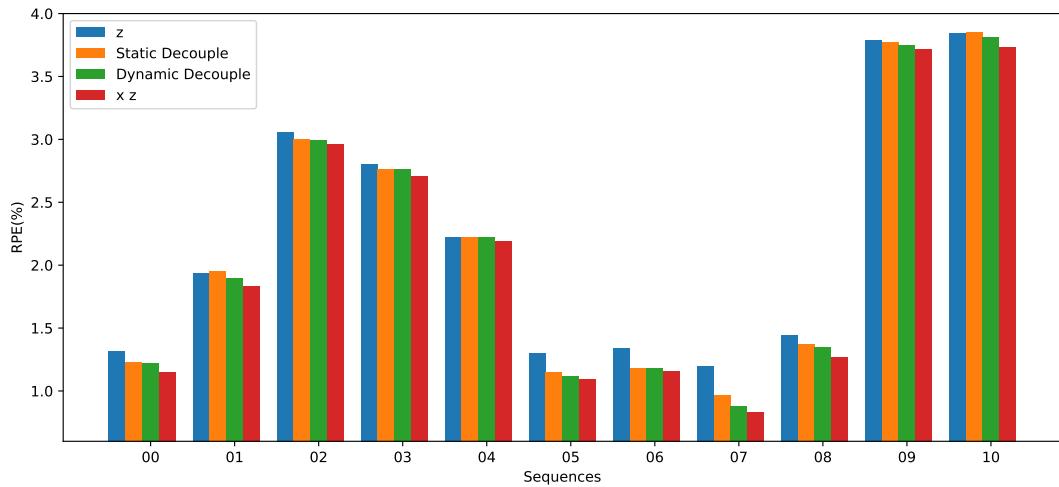


图 1.11 运动解耦效果对比

表 1.2 运动聚焦解耦的性能提升

| Train | Test | Learn All Motion | | Learn R_y, t_z | |
|-------------|----------------|------------------|----------------|------------------|----------------|
| | | Trans (%) | Rot (deg/m) | Trans (%) | Rot (deg/m) |
| 00 | 02 04 06 08 10 | 26.8 | 0.137 | 23.9 | 0.110 |
| 00 02 | 04 06 08 10 | 18.3 | 0.095 | 16.7 | 0.070 |
| 00 02 04 | 06 08 10 | 17.6 | 0.091 | 16.9 | 0.076 |
| 00 02 04 06 | 08 10 | 15.3 | 0.082 | 13.2 | 0.065 |

1.2.2.3 运动简化的有效性验证

我们设计对比实验以证明运动简化的有效性。我们使用相同的训练数据训练了两种模型：1) 运动简化模型 (Motion Focusing Model, MFM) 仅学习车辆的前进运动的旋转运动；2) 全运动模型 (All Motion Model, AMM), 学习车辆六个自由度的全部运动。实验在多个训练和测试数据序列的组合上进行以减低随机性的影响。我们记录了网络训练时的损失函数曲线和测试时的相对平移误差。

如图1.12所示，运动简化模型收敛速度相比于全运动模型有着大幅提高。

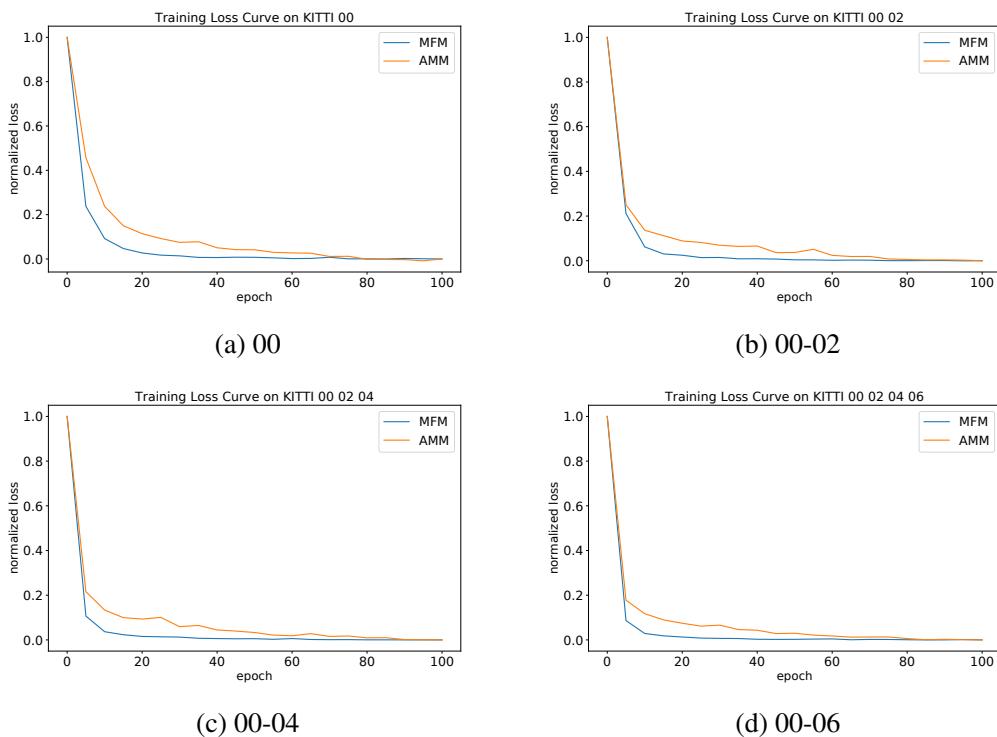


图 1.12 训练损失函数收敛速度比较

不同训练模型的测试误差记录在表1.2，可视化与图1.13。可以看出，运动

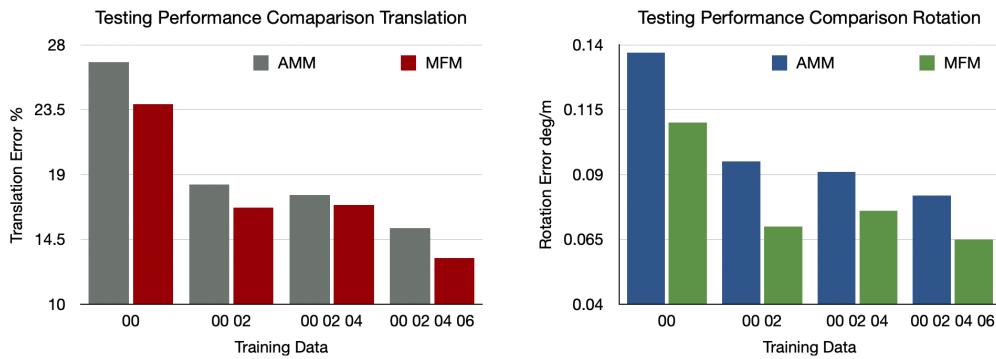


图 1.13 运动聚焦与解耦的性能提升

简化模型的测试精度相比于全运动模型在不同数据组和的情况下均有一定提高。其中相对平移误差提高约 2%，相对旋转误差提高约 0.2 度每米。可见，尽管运动简化是训练轨迹的目标与真实轨迹相比出现了一定的偏移，但整体来看，测试精度是有所提升的。另外，我们发现，无论是运动简化模型还是全运动模型，随着训练数据集数据量的不断提高，测试误差不断降低。

1.2.2.4 与其他算法比较

表 1.3 与其他基于学习的视觉里程计方法比较

| Seq | Zhan et al. (from ^[18]) | | DeepVO (from ^[19]) | | Zhou et al. (from ^[14]) | | GeoNet (from ^[20]) | | Our Method | |
|-----|--|---------|-----------------------------------|---------|--|---------|-----------------------------------|---------|------------|--------------|
| | Trans | Rot | Trans | Rot | Trans | Rot | Trans | Rot | Trans | Rot |
| | (%) | (deg/m) | (%) | (deg/m) | (%) | (deg/m) | (%) | (deg/m) | (%) | (deg/m) |
| 09 | 11.92 | 0.036 | - | - | 17.84 | 0.068 | 26.93 | 0.095 | 9.26 | 0.023 |
| 10 | 12.62 | 0.034 | 8.11 | 0.088 | 37.91 | 0.178 | 24.69 | 0.0843 | 9.10 | 0.022 |
| Avg | 12.27 | 0.035 | 8.11 | 0.088 | 28.88 | 0.123 | 25.81 | 0.090 | 9.18 | 0.023 |

我们将我们的算法与其他基于学习的以及基于几何计算方法进行定量比较。我们的模型使用 KITTI 数据集 00-08 进行训练，在序列 09 和 10 上进行测试，这个训练测试组合与其他基于卷积神经网络的方法^[14, 18, 20]一致。测试平均误差记录与表1.3和表1.4中。

由于 SfM-Learner^[14] 和 GeoNet^[20] 训练时并无绝对尺度信息，所以在评价其误差之前，我们先将其与真实轨迹进行了对齐。

表1.4 与其他传统视觉里程计方法比较

| Seq | LIBVISO2 (from ^[4]) | | ORB-SLAM (from ^[21]) | | Our Method | |
|-----|------------------------------------|---------|-------------------------------------|---------|-------------|---------|
| | Trans | Rot | Trans | Rot | Trans | Rot |
| | (%) | (deg/m) | (%) | (deg/m) | (%) | (deg/m) |
| 09 | 4.04 | 0.0143 | 15.30 | 0.0026 | 9.26 | 0.0229 |
| 10 | 25.20 | 0.0388 | 3.68 | 0.0048 | 9.10 | 0.0221 |
| Avg | 14.62 | 0.0266 | 9.49 | 0.0037 | 9.18 | 0.0225 |

此外 ORB-SLAM^[21] 的单目版本和 LIBVISO^[22] 的单目版本也需要与真实轨迹对齐。

从表1.3中可以看出我们方法的效果优于只基于卷积神经网络的方法^[14, 18, 20] 和基于卷积神经网络和递归神经网络的 DeepVO^[19] 精度不相上下。和传统发方法 LibVISO2^[22] 和 ORB-SLAM 单目^[21] 相比，我们获取了更好的平均精度。

1.2.3 实验结果讨论

本节我将总结实验结果、分析算法性能和局限性。

1.2.3.1 算法有效性

根据上述实验结果，我们可以如下四个方面总结。1) 运动聚焦并不会引入过多的姿态偏移。运动聚焦之后，轨迹平均位置误差为 2%，其意味着在机器人运行 100m 后，其平均偏移量大概为 2m。从可视化的重构轨迹中可以看出，这个偏移量相对很小。2) 移动解耦可以进一步减少姿态偏移。运动解耦通过解耦 y 轴旋转运行和 x 轴平移运动之间的耦合性，减小了去除 x 为平移时的姿态偏移，其中所提出的动态解耦算法优于静态解耦算法。在动态解耦算法中，相机到后轴的距离作为一个主要参数，本文实验中该距离为通过数据计算得到，在实际情况中也可以通过测量获取。3) 运动聚焦和运动解耦提升了算法性能，主要体现在两个方面：运动聚焦和解耦之后的模型可以更快的收敛，平均在 20 个训练周期之后即可收敛，而普通模型需要大约 60 个训练周期；此外，运动聚焦和解耦之后的模型提升了运动估计的精度，在所有的对比实验中，性能都由于普通模型。4) 与其他算法相比，我们取得更好的效果。在对比中可以发现，几何方法并不鲁棒，在不同测试集中效果差异较大，而本文算法取得了更好的平均效果；

另外，我们的算法优于其他基于卷积神经网络的算法，但与借助递归神经网络的 DeepVO 基本持平，我们取得了更小的角度误差，DeepVO 取得了更小的平移误差

1.2.3.2 算法为什么有效

运动聚焦和运动解耦的有效性可以从三个角度解释：1) 首先，有人地面车辆的运动受其自身机械结构和动力机制的约束，其不具备完善的三维空间内的 6 自由度运动模式，所以在不考虑其非主要运动维度上的运动时，并不会造成过大的位置偏移，这一论点在第1.2.2.1节得到验证，是本文方法的基础。2) 由于非主要维度的运动幅度非常小，导致其信噪比不高，如果尝试去学习这些维度的运动，模型容易被噪声干扰。3) 由于我们仅学习两个主要维度，所学习的问题变得相对简单，进而一个轻量级的模型就可以去学习拟合这个映射，这样数据也就相对充足，而充足的数据会提升算法的性能（如表 1.2）。

1.2.3.3 算法局限性及解决方案

本文所提出的算法在车辆的运动大部分局限在水平面上时可以取得较好的效果，如果车辆有较多的 x 轴转动时，本文提出的算法的精度会下降。如图 1.11 所示，由于序列 09 和 10 存在较大比例的非水平运动，所以序列 09 和 10 的 RPE 误差也相对较大。为了解决这个限制，一种可行的方案为，使用其他传感（如惯性传感器）测量车辆 x 轴的转动，作为本文算法的一个补充。

此外，本文假设视觉传感器水平朝前安装。但相机的安装角度不是水平时，车辆的前进运动会被映射为前进分量和上下分量。在这种情况下，我们需要在初始阶段，标定车相机与水平线的俯仰角 σ ，然后使用如下公式对车辆平移运动进行变换

$$t_\sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\sigma) & -\sin(\sigma) \\ 0 & \sin(\sigma) & \cos(\sigma) \end{pmatrix} t \quad (1.14)$$

1.3 本章小结

本章提出了使用一种轻量化卷积神经网络架构去学习路面车辆的主要运动。我们定量的评价了运动简化造成的轨迹偏移，发现仅保留主要的 Y 轴旋转已经 Z 轴平移运动时，整体轨迹平移相对较小。我们根据车辆的运动模态建立车辆的平移偏角和旋转角之间的数学关系，进一步降低了在运动简化时轨迹偏移。我们

所提出的网络轻量化网络模型可以在 CPU 上实时运行。在 KITTI 上的定量实验证明了我们算法精度不差于现有算法。

参考文献

- [1] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, Florida, USA, 20-25 Jun 2009: 248-255.
- [2] CORDTS M, OMRAN M, RAMOS S, et al. The Cityscapes Dataset for Semantic Urban Scene Understanding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Caesars Palace, Las Vegas, Nevada, United States, Jun 26- Jul 1, 2016.
- [3] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE: 3354-3361.
- [4] SONG S, CHANDRAKER M, GUEST C. High Accuracy Monocular SFM and Scale Correction for Autonomous Driving[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015: 1-1.
- [5] LEE B, DANILIDIS K, LEE D D. Online self-supervised monocular visual odometry for ground vehicles[Z]. Conference Paper. 2015.
- [6] ZHOU D, DAI Y, LI H. Reliable scale estimation and correction for monocular Visual Odometry[C]//Intelligent Vehicles Symposium (IV), 2016 IEEE. IEEE, 2016: 490-495.
- [7] WANG X, ZHANG H, YIN X, et al. Monocular visual odometry scale recovery using geometrical constraint[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). 2018: 988-995.
- [8] YANG S, JIANG R, WANG H, et al. Road Constrained Monocular Visual Localization Using Gaussian-Gaussian Cloud Model[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3449-3456.
- [9] SCARAMUZZA D, FRAUNDORFER F, SIEGWART R. Real-time monocular visual odometry for on-road vehicles with 1-point ransac[C]//2009 IEEE International Conference on Robotics and Automation. 2009: 4293-4299.
- [10] CHOI S, PARK J, YU W. Simplified epipolar geometry for real-time monocular visual odometry on roads[J]. International Journal of Control, Automation and Systems, 2015, 13(6): 1454-1464.
- [11] SCARAMUZZA D, SIEGWART R. Appearance-Guided Monocular Omnidirectional Visual Odometry for Outdoor Ground Vehicles[J]. IEEE Transactions on Robotics, 2008, 24(5): 1015-1026.
- [12] SIEGWART R, NOURBAKHSIR, SCARAMUZZA D. Introduction to autonomous mobile robots[M]. MIT press, 2011.
- [13] WU Y, HE K. Group normalization[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [14] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video[C]//CVPR:vol. 26. 2017: 7.
- [15] LIN M, CHEN Q, YAN S. Network in network[C]//International Conference on Learning Representations (ICLR). Banff, Canada, Apr 14 - 16, 2014.
- [16] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representations (ICLR). San Juan, Puerto Rico, May 2-4, 2016.
- [17] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. ArXiv preprint arXiv:1412.6980, 2014.

-
- [18] ZHAN H, GARG R, WEERASEKERA C S, et al. Unsupervised Learning of Monocular Depth Estimation and Visual Odometry with Deep Feature Reconstruction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 340-349.
 - [19] WANG S, CLARK R, WEN H, et al. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]//Robotics and Automation (ICRA), 2017 IEEE International Conference on. 2017: 2043-2050.
 - [20] YIN Z, SHI J. GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR):vol. 2. 2018.
 - [21] MUR-ARTAL R, MONTIEL J, TARDÓS J D. Orb-slam: a versatile and accurate monocular slam system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
 - [22] GEIGER A, ZIEGLER J, STILLER C. StereoScan: Dense 3D Reconstruction in Real-time[C]// Intelligent Vehicles Symposium (IV). 2011.