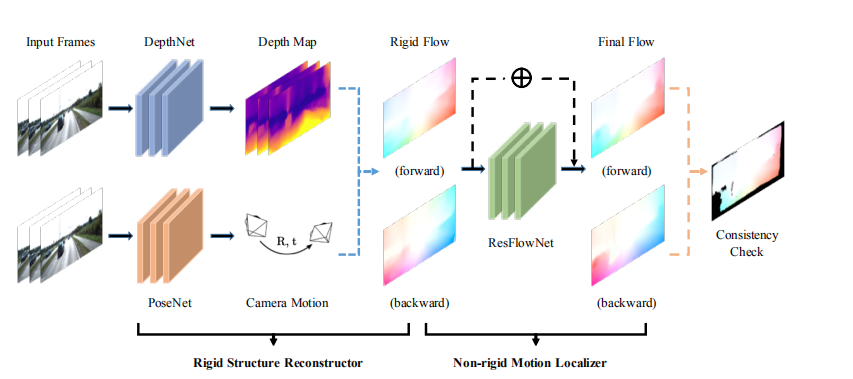
论文阅读笔记：

GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose 2018cvpr

**内容**：一个联合无监督的学习框架， 用于从视频中进行单眼深度，光流和运动估计。在3D场景中这三个是相互联系着，文章提出的框架以端到端的方式共同学习。在各个模块的预测上提取 几何关系，然后将其组合为图像重建损失，分别推断静态和动态场景部分。此外，我们提出了一种自适应几何一致性损失，以增加对异常值和非朗伯区域的鲁棒性， 从而有效地解决了遮挡和纹理模糊。



**理解**：视频或图像是投影到特定维度的3D空间的屏幕截图。3D场景自然由静态背景和移动物体组成。视频中静态 部件的移动仅由相机运动和深度结构引起。动态物体 的运动更复杂，设计了一个由两个阶段组成的级联结构将刚性流动和物体运动分开。们使用单独的组件分别通过刚性结构重建器和非刚性运动定位器来学习刚性流动和物体运动。采用图像外观相似性来指导无监督学习。

**细节方面的理解：**

1.网络结构（刚性结构重构器和运动物体定位器）

提出Geonet包含两个阶段，刚性结构推理阶段和非刚性运动细化阶段。

1. 推断场景布局的第一阶段由两个子网络组成，即DepthNet和PoseNet。深度图和相机姿势分别回归并融合以产生刚性流动。
2. Res-FlowNet完成第二阶段以处理动态对象。ResFlowNet学习的残余非刚性流与刚性流相结合，得出最终流量预测。
3. 的每个子网都针对特定的子任务，因此将复杂的场景几何理解目标分解为一些更简单的子任务。在不同阶段查看综合作为我们无监督学习范式的基本监督。

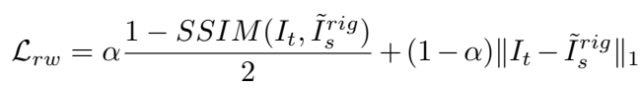
刚性结构重构器

深度估计网络输入为单幅图像，输出为其对应的深度图。姿态估计网络输入为整个图像序列，直接回归出每一个时刻的相机位姿。这两个网络组成了静态场景重构器。其监督信息为像素一致性。

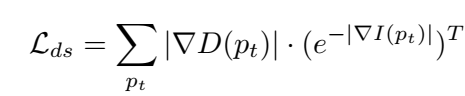
ae3dca275a21443ca2415ca3834f0007

通过上式把深度图和位姿能够生成合成的视图，将其与真实视图做比较，就得到了监督信息。

为了使得深度估计对运动物体和遮挡物体更加鲁棒，采用如下的误差函数：



其中SSIM表示结构相似性指数和通过交叉验证α取0.85。 除了刚性翘曲损失Lrw，过滤掉错误的预测并保留清晰的细节，我们引入了边缘感知深度平滑度损失Lds由图像梯度加权



运动物体定位器：

第一阶段得到的结果作为初始值（将深度图和相机位姿转换为光流图），得到完整的光流图。本阶段的监督信息与上一阶段类似，只需将深度图和姿态的得到的合成视图替换为光流图得到的合成视图，将深度图的平滑替换为光流的平滑约束。

ResFlowNet运动定位器只关注其他非刚性残留物，Res-FlowNet不仅可以在动态中纠正错误的预测对象，但也从第一阶段细化不完美的结果。得益于我们的端到端学习协议，这可能是由高饱和度和极端光照条件引起的。

前向一致性检查

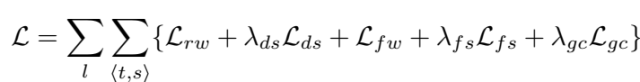
GeoNet采用刚性结构重建器作为静态场景，非刚性运动定位器作为动态对象的补偿。两个阶段都利用视图合成目标作为监督，隐含的光度一致性假设。

一致性约束不应在封闭区域施加翘曲损失。文章中优化了最终运动场的自适应一致性损失。

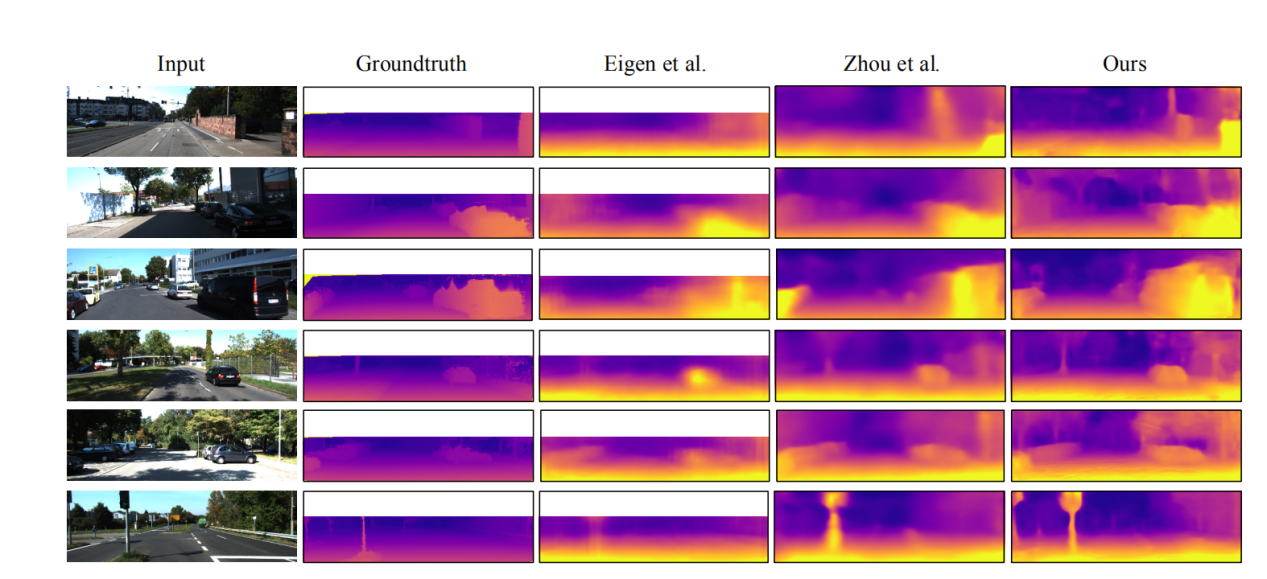
IMG_256

如果同一点在前项通道和后向通道的取值相差太大，那么认为该点很可能是野点。

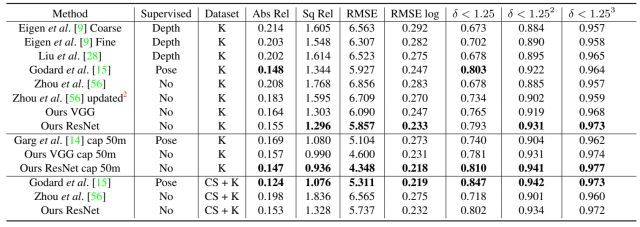
2.损失函数



**实验与结果**：



Eigen等人之间单眼深度估计的比较。[9]（由深度监督），周等人。[56]（无人监督）和我们的（无人监督）。文章的方法捕获薄结构的细节，并在近距离和远距离区域保持始终如一的高质量预测。



文中单目图像的深度估计，以及光流估计、姿态估计三个方面分别设计实验，对算法进行了验证，这里就对深度估计的结果做了分析。

本文中使用VGG框架仅在KITTI上接受了与“周等人”相同的网络架构但没有BN，这显示了损失函数的有效性。本文的方法明显优于两种监督方法和之前无人监督的工作.图中显示了定性比较，但本文的结果略逊于Godard等人。

**问题**：