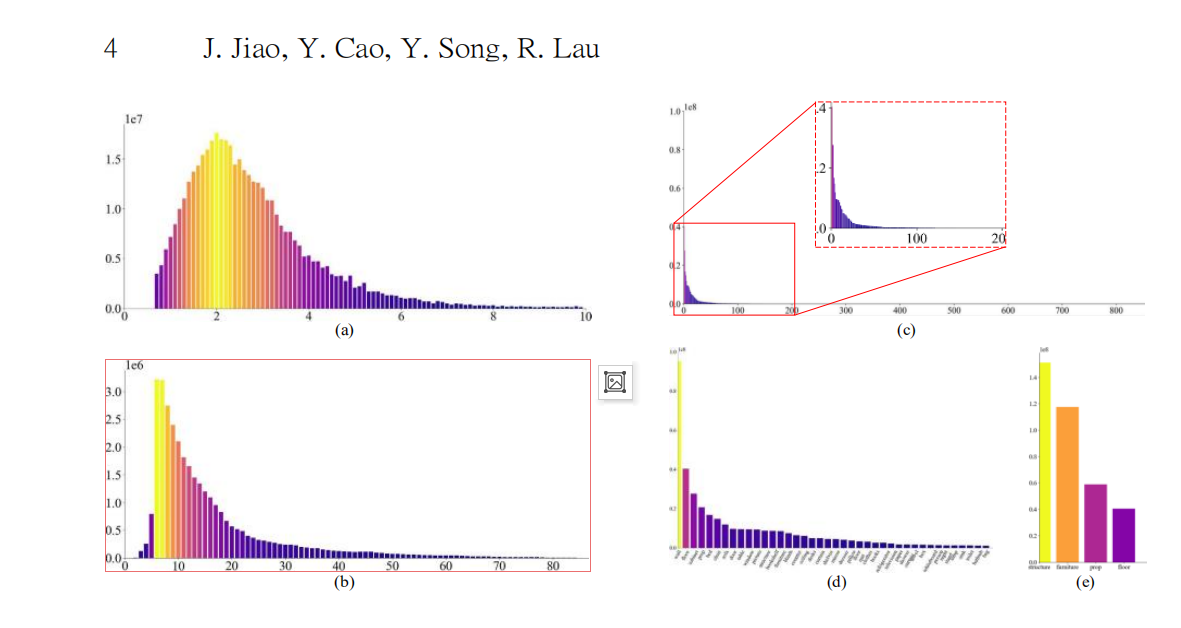
论文阅读：Look Deeper into Depth: Monocular Depth

Estimation with Semantic Booster and Attention-Driven Loss

**内容**：在本文中，研究了长尾特性并深入研究了远距离深度区域（即尾

部），为网络监督提出了注意力驱动的损失。此外，为了更好地利用单眼深度估计的语义信息，提出了一个协同网络来自动学习两个任务之间的信息共享策略。利用所提出的注意力驱动的损失和协同网络，可以相互改善深度估计和语义标记任务。

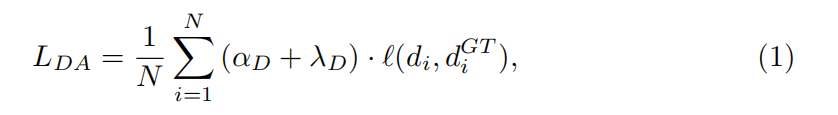


**细节方面的理解：**

**深度感知目标函数**

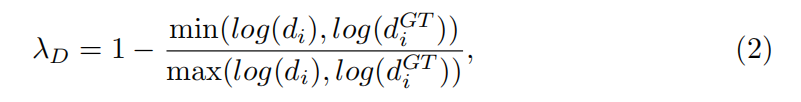
在本文中，作者提到它发现图像的深度值是呈现长尾分布的在特定深度之后，每个深度值的样本数/像素数急剧下降，仅有很小的深度范围支配大量像素。这种数据不平衡问题与物体检测中的相似性但性质不同。这是因为来自成像过程的透视效果的固有自然属性导致深度像素的不均匀分布，这不能通过简单地增加训练数据来消除。

建议引导网络在训练期间更加关注远处深度区域并相应地自适应地调整反向传播流程。深度意识目标表述：



其中IMG_256是像素索引，IMG_257是深度图像素数量，IMG_258和IMG_259分别是预测的深度值和实际的深度值，IMG_260是距离度量如IMG_261或IMG_262范式等，IMG_263是一个深度感知关注项使网络更关注深度值较大的区域。这样，反向传播时梯度更偏重少数远距离区域。因此IMG_264和深度相关，可以定义为真实深度值的一个线性函数。

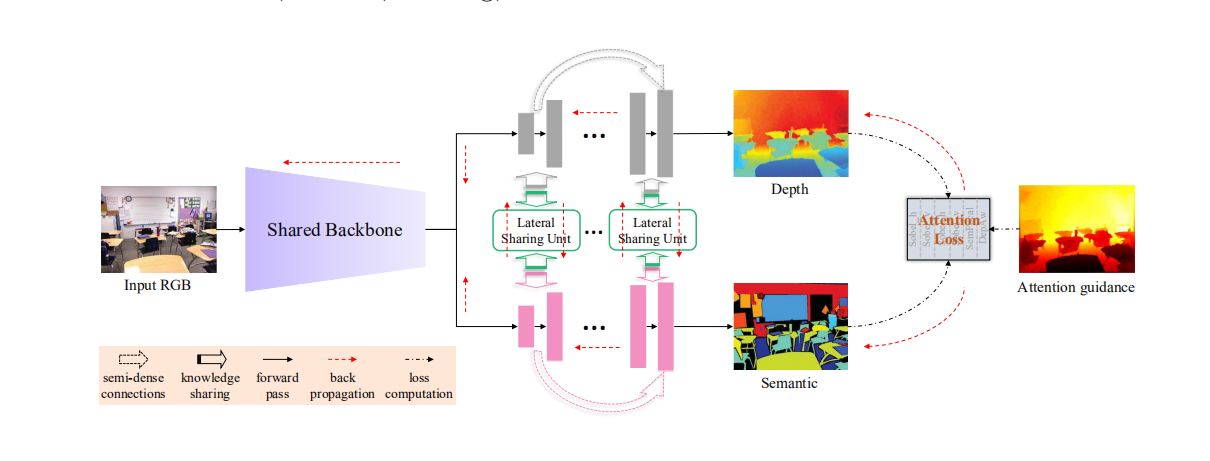
自动引导网络自适应地聚焦于不同区域并自动调整每个训练样本的强度/注意力，从而确保模型的优化方向相对平衡。使网络能够在训练期间更加关注远处的像素样本。



为了避免训练刚开始时出现梯度消失，学习近距离区域时出现中断，引入正则化项描述训练过程中的学习状态。如果网络预测的深度值接近实际值，接近0，否则接近1。这样，即使对近距离区域(接近0)，梯度仍然能通过原始损失函数反向传播。增加了训练稳定性。

**网络结构**（协同网络）

协同网络是一个多任务深度CNN，主要由四部分组成：深度预测子网，语义标记子网，知识共享单元/连接，以及注意力驱动的损失。



1.输入的RGB图像通过骨干编码器，将颜色空间转换为高维特征空间。

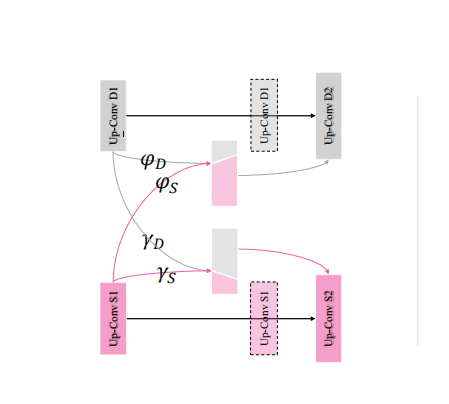
2.通过两个子网络，从共享的高维特征重建深度和语义标签

3.通过所提出的半密集上跳连接（SUC）在每个子网内部执行知识共享。两个任务之间的知识共享是通过横向共享单元（LSU）实现

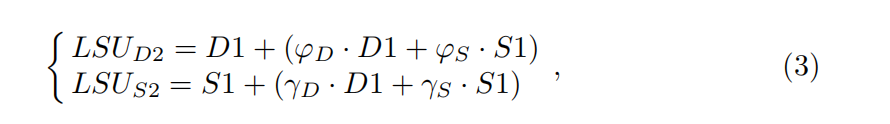
4.整个训练过程由注意力驱动的损失监督，

**共享单元**

一个双向横向共享单元（LSU），以动态路由方式自动学习共享策略。正向传递和反向传播都实现了信息共享。在网络中的每两个上升层之间，除了任务内传播之外，还添加这样的LSU以共享来自其他任务的剩余知识/表示。



假设当前的反卷积层产生的特征图是和，分享的特征表示可以表示

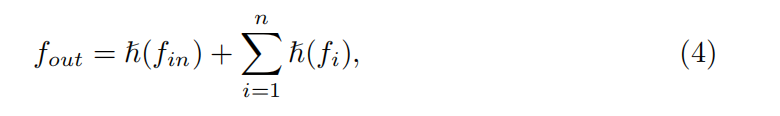


共享表示和传播到下层反卷积层。所有的参数都是学习来的。虽然所有LSU有相同的内部结构，但它们有各自的参数，这样可以更灵活地共享。

**半稠密向上跳跃连接SUC：**

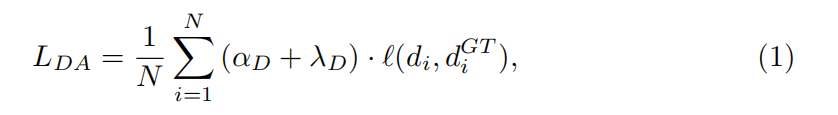
为了执行更好的任务内知识共享并保留长期记忆，在up-conv层之间引入半密集上跳连接（SUC）

作者提出的网络框架在编码器后分为深度估计、语义标注两路，称为两个任务。在任务内添加SUC是为了任务内(相对于任务间)的知识共享和保存长期记忆(前一层反卷积层传过来的反卷积结果称为短期记忆)。*这其实就是跳跃连接。*SUC可以表示为：

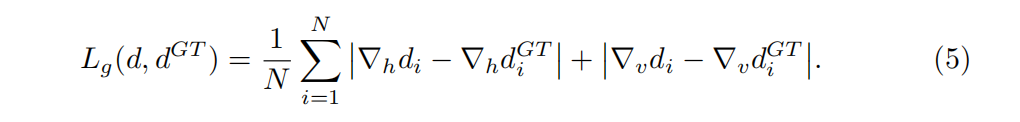
，其中IMG_257和IMG_258分别是解码器的输入输出，IMG_259是第IMG_260层反卷积的输出

**注意力驱动的损失**

关注驱动损失:

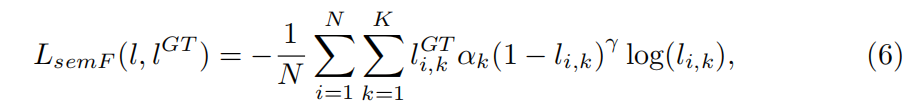


梯度损失:为了更好地保留局部结构和表面区域的细节，我们建议设置梯度约束，并引入梯度损耗层



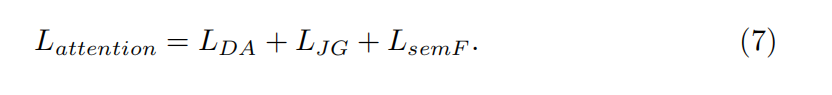
h和v分别代表水平方向和竖直方向。语义信息Lg（d，s）也考虑进来。

语义焦点损失:为了让网络更关注长尾，使用损失项



其中IMG_256是像素IMG_257的预测标签，IMG_258是类别索引。αk和γ是调节关注的平衡权重和焦点参数

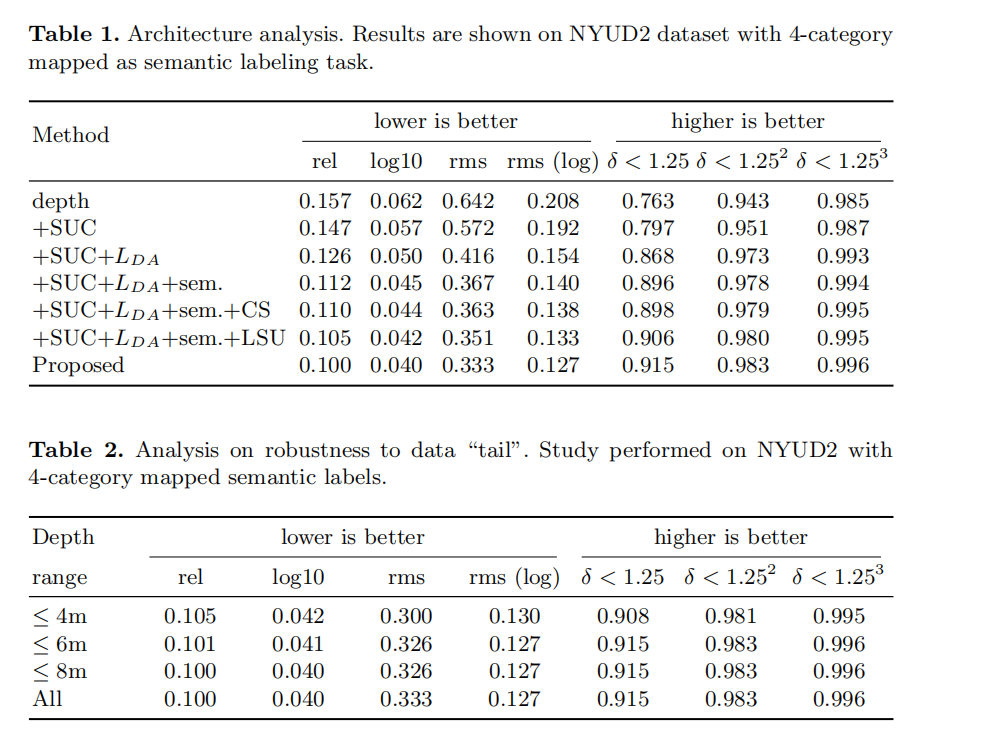
总损失：



**实验与结果**：

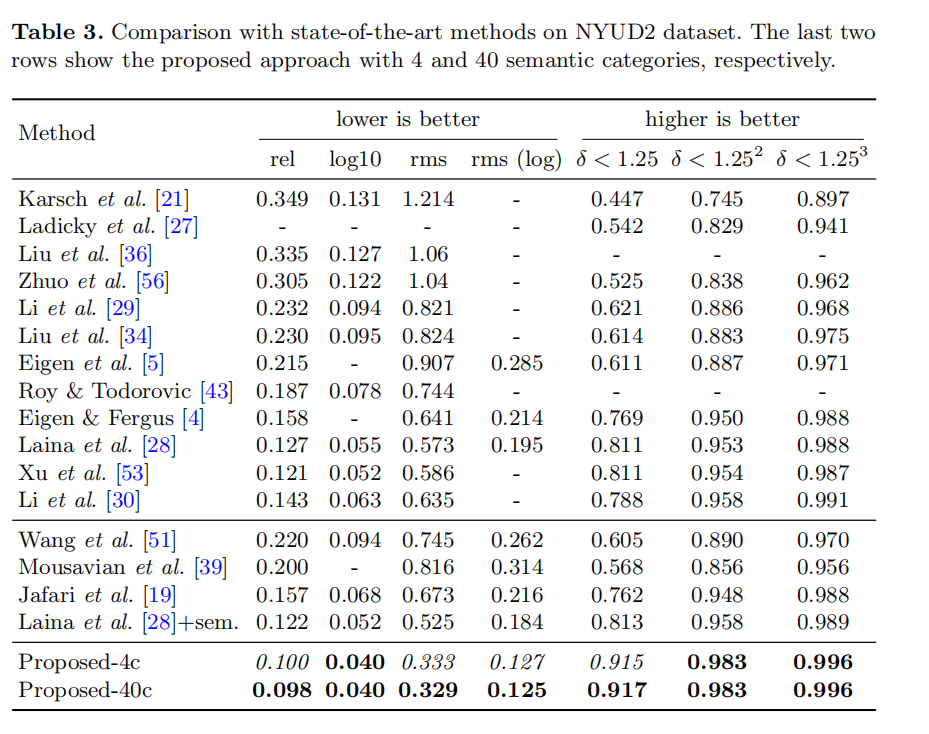
平均绝对相对误差（rel），平均log10 误差（log10），均方根误差（rms），rms（log）

和阈值下的准确度（δ<1.25i，i = 1,2,3）



CS结构确实有利于表示共享，而LSU表现稍好一些。来自语义标记任务的协同增强也对深度估计有很大帮助。注意力驱动的损失对性能的贡献最大，知识共享的辅助贡献来自语义标记。

为了验证所提出的长尾数据方法的稳健性，我们对数据的尾部进行了消融研究。具体地说，通过将相应的尾部切割2米（即≤4m，6m，8m，10m）将测试数据的深度范围分成四个部分。注意力驱动的损失监督网络更多地关注远距离深度，它在短尾数据上表现良好并且在不同范围内一致地表现，这表明所提出的注意力损失能够自适应地根据不同而变化



与NYUD2数据集上的最新方法的比较