

文献分享

20/09/25

Bidirectional Attention Network for Monocular Depth Estimation

Shubhra Aich*, Jean Marie Uwabeza Vianney*, Md Amirul Islam, Mannat Kaur, and Bingbing Liu[†]

Introduction

- 提出了一种用于单目深度估计的端到端的双向注意网络(BANet), 改善了卷积神经网络整合局部和全局信息方面的局限性。
- 该机制的结构源于神经机器翻译(NMT, Neural Machine Translation)的强大概念基础。方法引入了双向注意(bidirectional attention)模块。
- 实验表明, 方法在KITTI和DIODE两个具有挑战性的数据集上进行单目深度估计, 优于或至少齐平最先进的单眼深度估计方法, 并可以更少的内存和计算复杂度来运行。

Method

- 双向attention网络含义:

主干网络有多个特征层，利用前向后向两个方向分别分析不同层级的效果，最后结合成一个像素级别的权重，用以结合不同深度特征层的效果

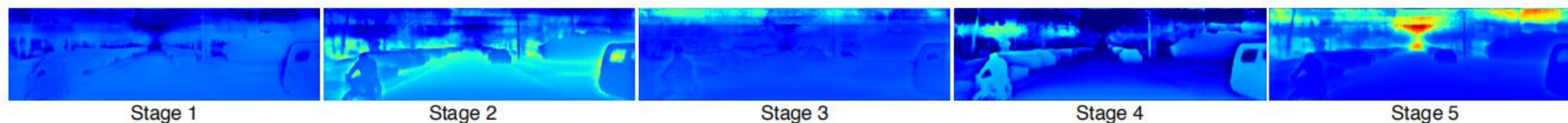
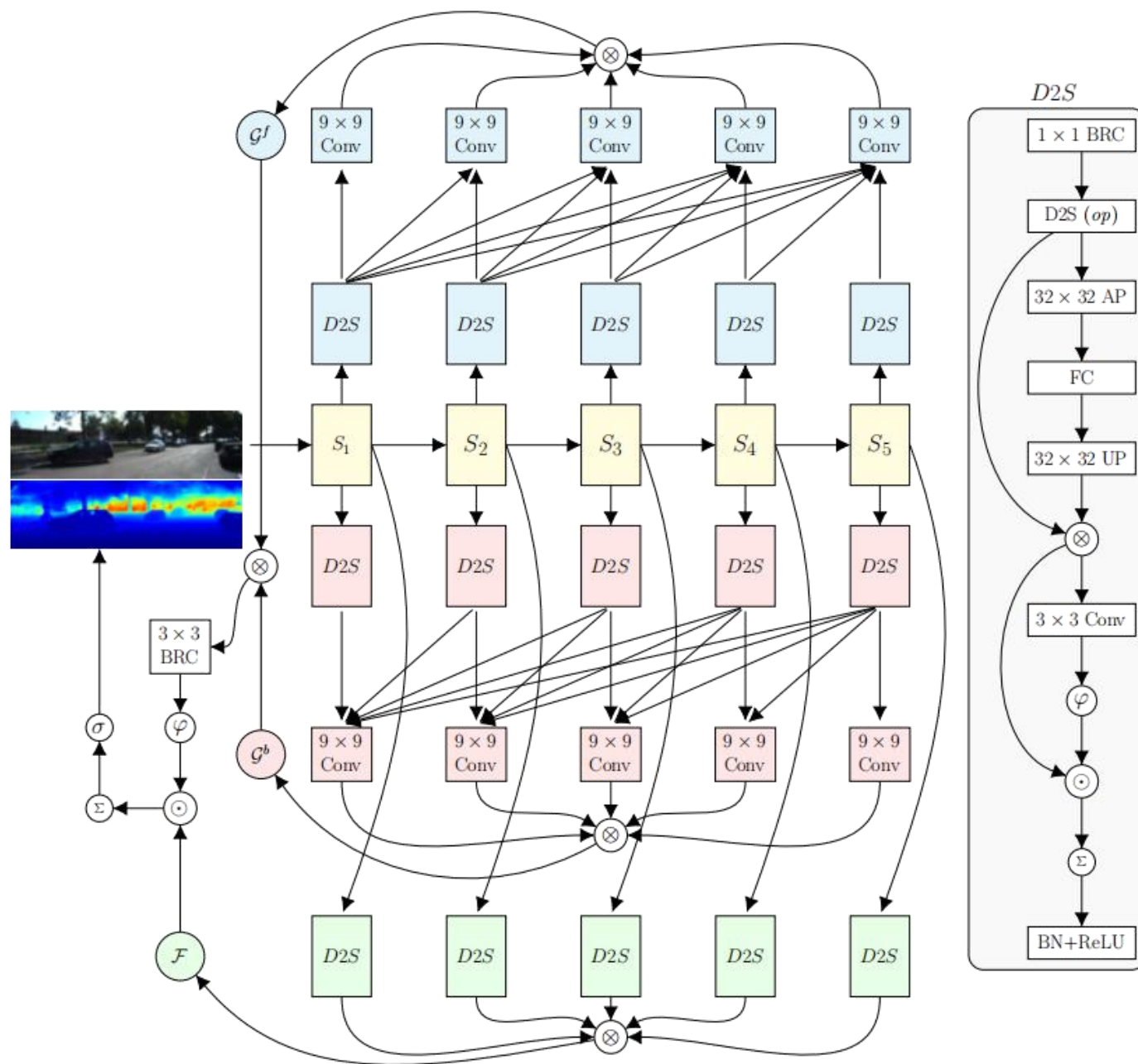
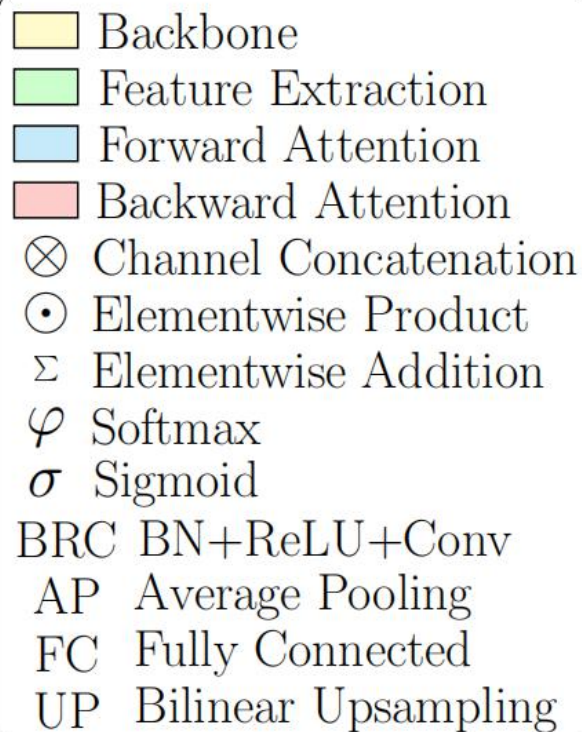


Fig. 3: An illustration of the generated stage-wise attention weights using the forward and backward attention modules.

Method



Method

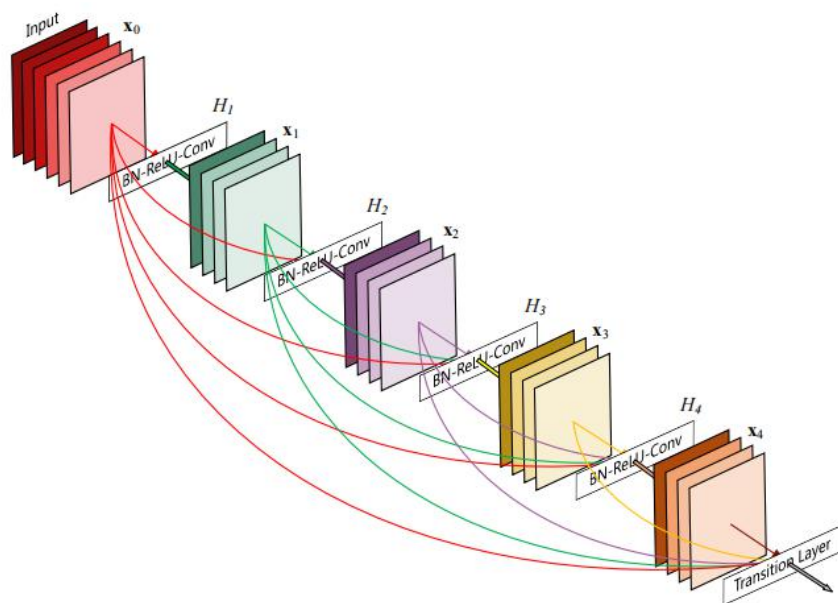


Figure 1: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

DenseNet

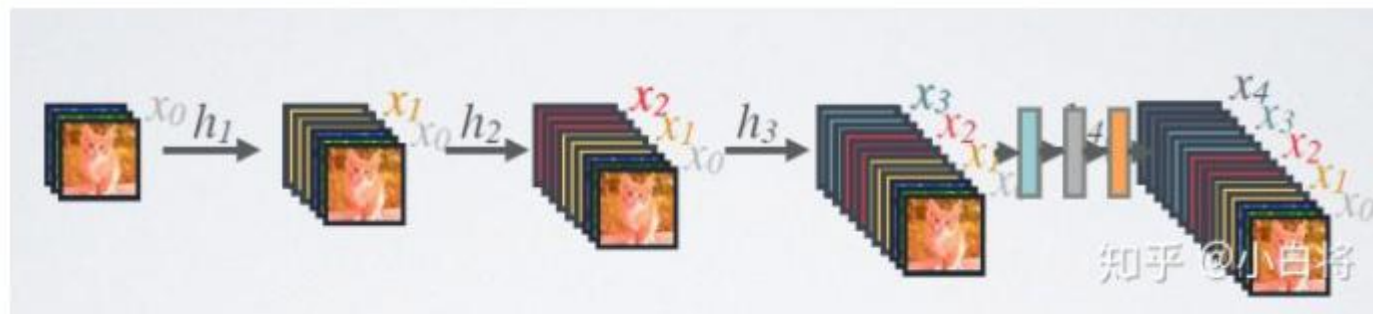
其类似于ResNet但是又有很多区别:

1 由于特征需要反复使用, 在跨层连接时使用
的是在特征维度上的 Concat操作, 而不是
Addition 操作。

2 也因此, 在每个单元模块的最后不需要一个
1X1 的卷积来将特征层数升维到和输入的特征
维度一致。

3 其特征量与运行时间均只有ResNet约一半左
右

Method



一层有两个部分：

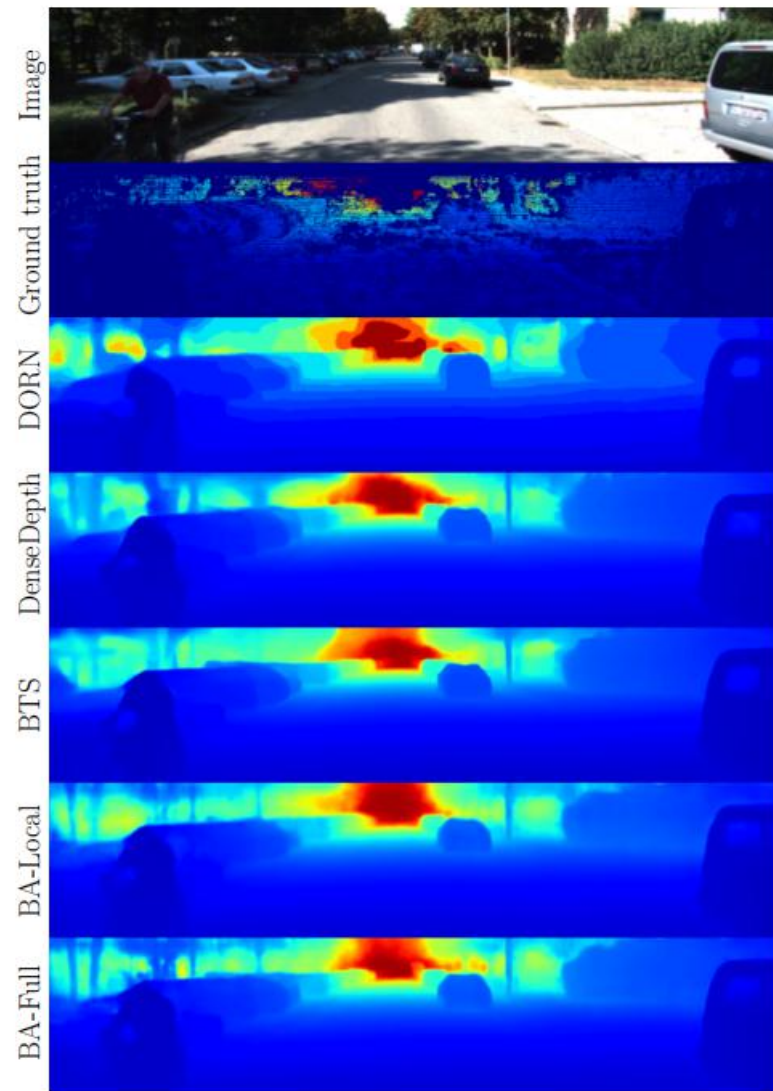
1 DenseBlock 利用所有的输入，通过卷积层得到k个通道的特征传送至下一层

2 Transition 将新得到的特征与原有特征Concat并一起降采样
其中k为增长率，意为每一层都可以产生k个通道，并与之前的特征堆叠来进入下一层，所以每一层都会比前一层多k个通道。

Result

Model	#Params (10 ⁶)	Time (ms)	Lower is better					
			SiLog	SqRel	AbsRel	MAE	RMSE	iRMSE
DORN [1]	89.23	36	12.22	3.03	11.78	2.04	3.80	11.68
DenseDepth [2]	44.61	25	10.66	1.76	8.01	1.58	3.31	8.24
BTS [3]	47.00	29	10.67	1.59	7.51	1.56	3.37	8.10
BANet-Vanilla	28.73	20	10.88	1.65	7.74	1.63	3.51	8.34
BANet-Forward	32.34	25	10.61	1.67	9.02	1.77	3.54	9.99
BANet-Backward	32.34	25	10.54	1.52	7.67	1.59	3.42	8.47
BANet-Markov	35.28	27	10.72	1.85	8.24	1.62	3.38	8.28
BANet-Local	35.08	27	10.53	2.21	9.92	1.77	3.34	9.68
BANet-Full	35.28	28	10.64	1.81	8.25	1.60	3.30	8.47

Method	SiLog	SqRel	AbsRel	iRMSE
DORN [1]	11.77	2.23	8.78	12.98
BTS [3]	11.67	2.21	9.04	12.23
BANet-Full	11.61	2.29	9.38	12.23



From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation

Jin Han Lee, Myung-Kyu Han, Dong Wook Ko and Il Hong Suh
Department of Electronics and Computer Engineering, Hanyang University
`{jinhanlee, mkhan91, pumpblack, ihsuh}@hanyang.ac.kr`

Introduction

- 提出了一种在编码阶段利用新的局部平面引导层的网络结构。
- 不同与一般方法只对特征进行升降采样，该方法求出不同分辨率的深度图进行融合。
- 实验表明，提出的方法优于目前的最先进的工作。

Method

- 不同分辨率深度图

低分辨率下每个深度并非用一个点来计算，而是用四个参数表示一个平面面片的位置和方向，即局部平面假设。

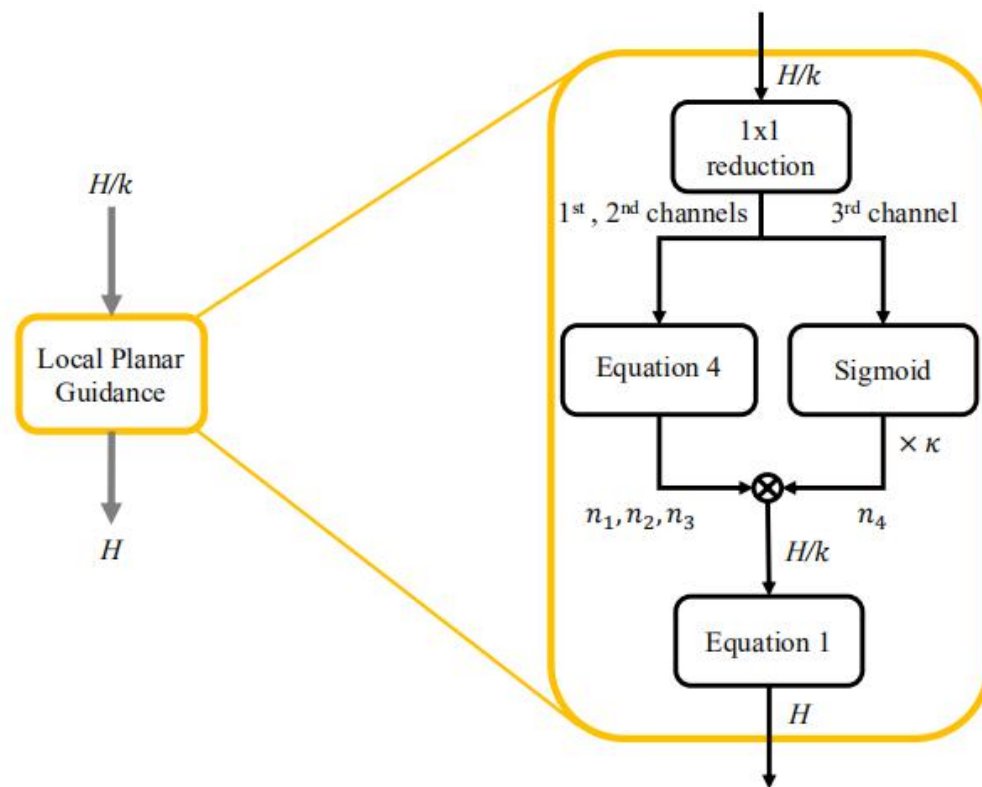


Method

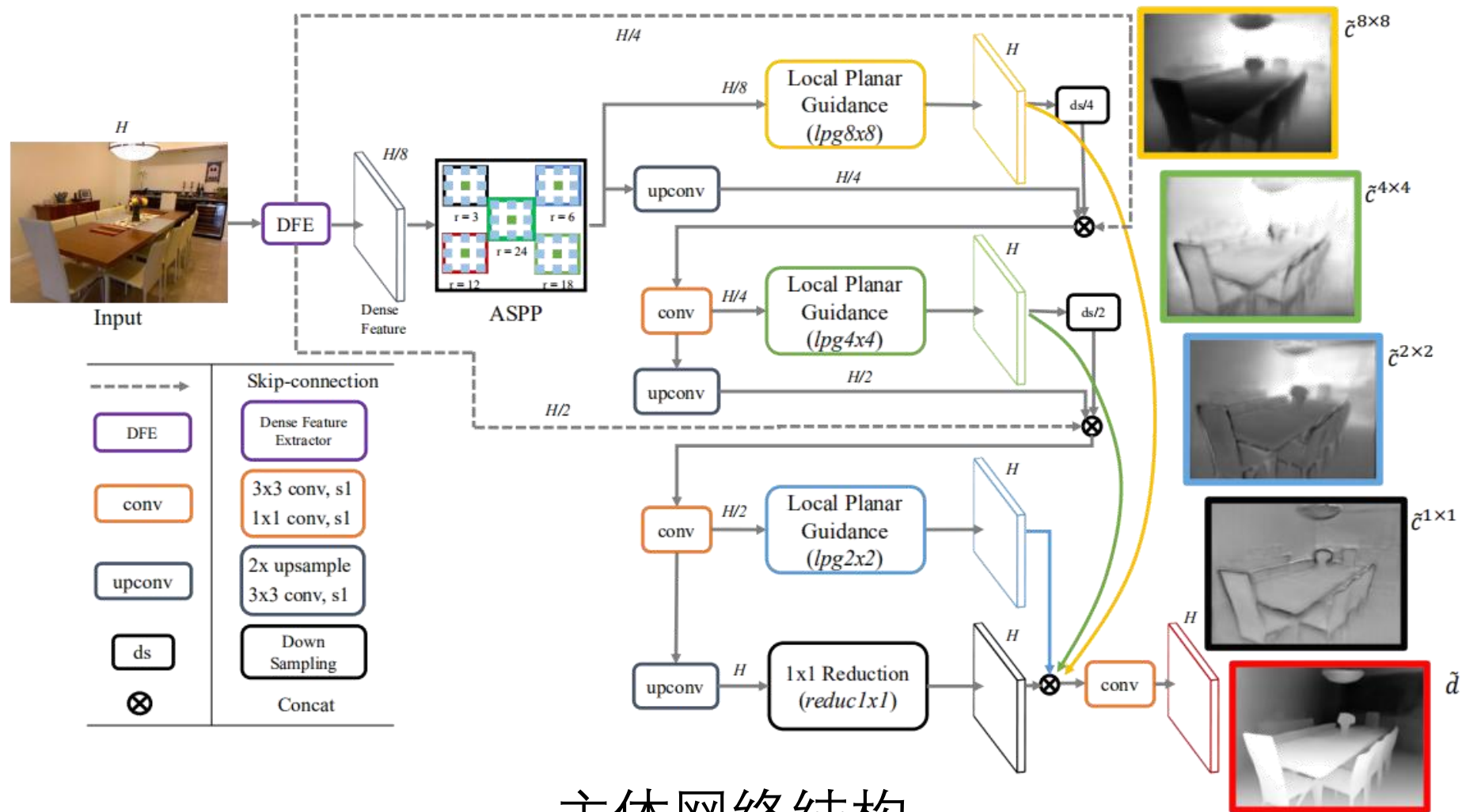
- 局部平面引导图

先将输入变成 H/k ， k 为放缩比例，如8/4/2。

然后三个通道分成两路，一路计算位置处的法向量，另一处计算位置处平面与点的距离。以达到用4个参数描述 $k*k$ 平面的效果。



Method

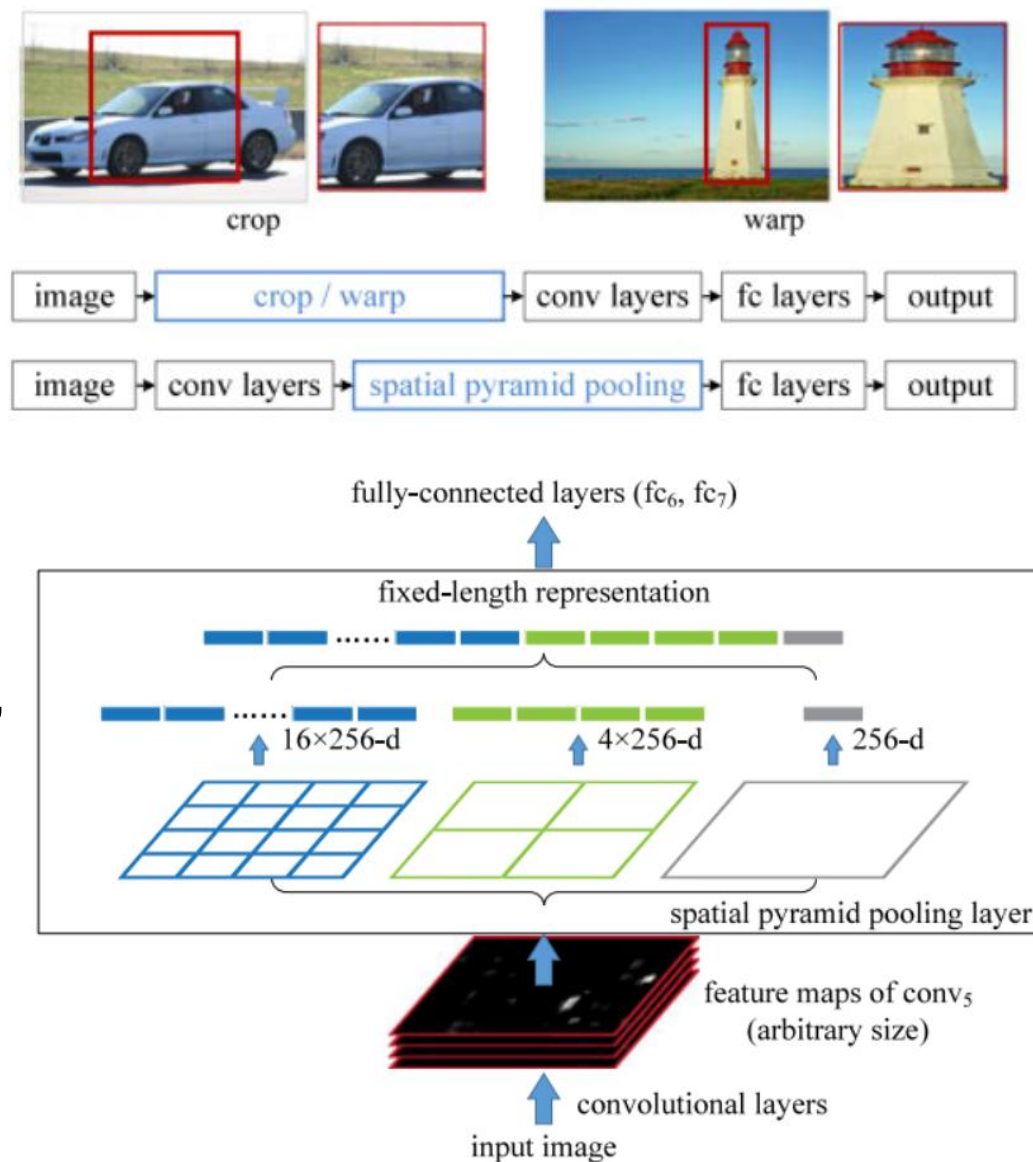


主体网络结构

Method

- SPP(Spatial Pyramid Pooling)

相较于普通卷积层需要切割成固定大小，SPP则从图像分割角度入手：
给定不同分割块，每一个分割块给出一个特征，这样无论输入图像的像素，输出都是固定个数的特征。

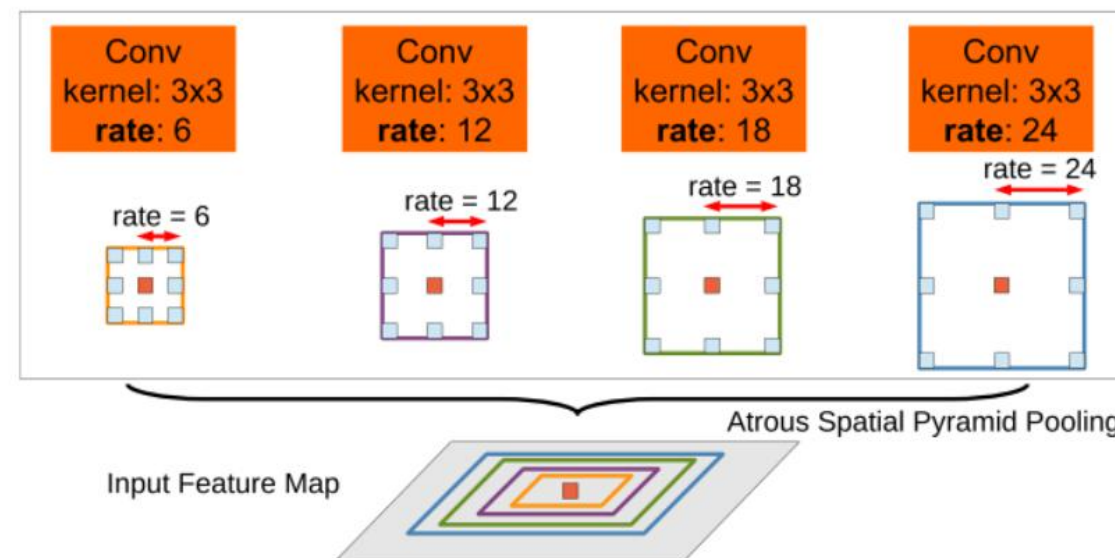
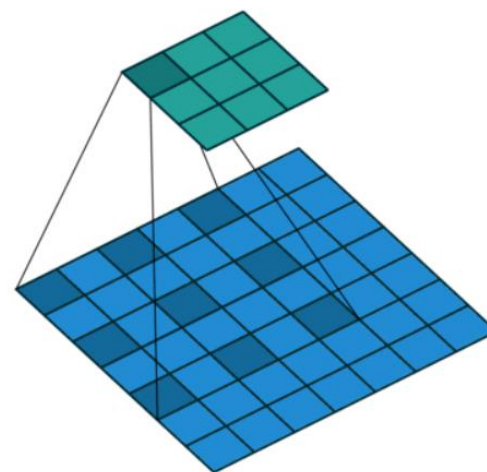


Method

- ASPP(Atrous SPP)

利用不同rate的空洞卷积来进行SPP的提取，被称之为ASPP，其中rate的定义是指的采样点之间间隔rate-1个像素，即：

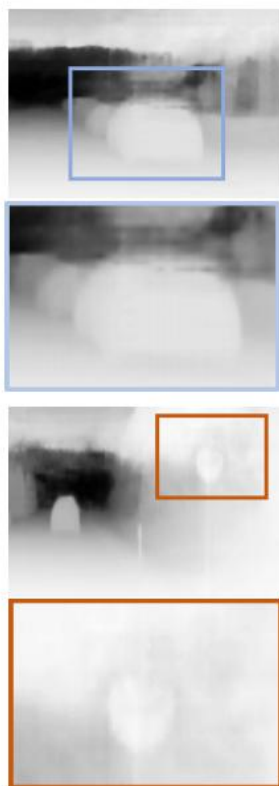
3×3 , rate=6, 则实际大小为 11×11



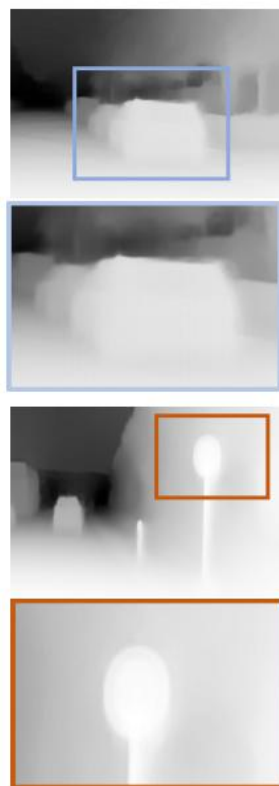
Result



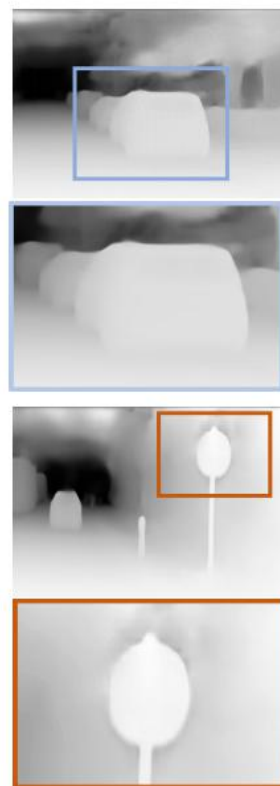
(a) Input



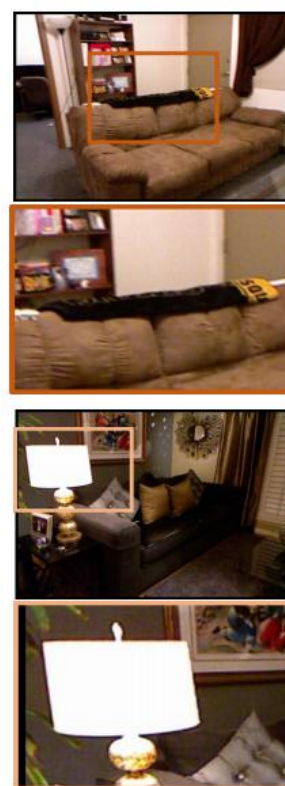
(b) Fu et al. [12]



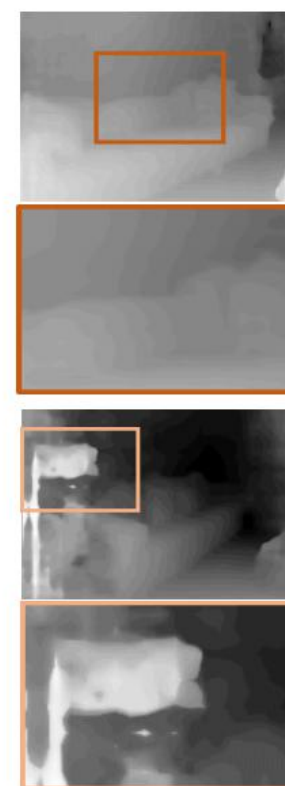
(c) Yin et al. [51]



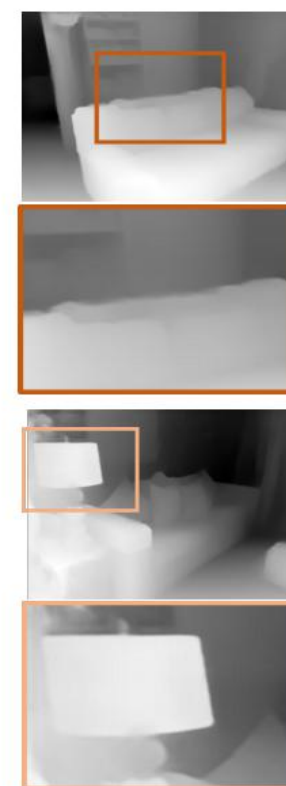
(d) Ours



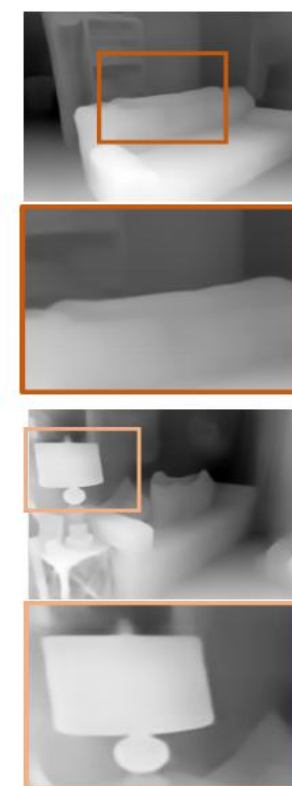
(a) Input



(b) Fu et al. [12]



(c) Yin et al. [51]



(d) Ours

Result

Method	cap	<i>higher is better</i>			<i>lower is better</i>			
		$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	Abs Rel	Sq Rel	RMSE	RMSE <i>log</i>
Saxena et al. [39]	0-80m	0.601	0.820	0.926	0.280	3.012	8.734	0.361
Eigen et al. [11]	0-80m	0.702	0.898	0.967	0.203	1.548	6.307	0.282
Liu et al. [29]	0-80m	0.680	0.898	0.967	0.201	1.584	6.471	0.273
Godard et al. (CS+K) [16]	0-80m	0.861	0.949	0.976	0.114	0.898	4.935	0.206
Kuznetsov et al. [24]	0-80m	0.862	0.960	0.986	0.113	0.741	4.621	0.189
Godard et al. (CS+K) [16]	0-80m	0.861	0.949	0.976	0.114	0.898	4.935	0.206
Gan et al. [13]	0-80m	0.890	0.964	0.985	0.098	0.666	3.933	0.173
Fu et al. [12]	0-80m	0.932	0.984	0.994	0.072	0.307	2.727	0.120
Yin et al. [51]	0-80m	0.938	0.990	0.998	0.072	-	3.258	0.117
Ours-ResNet	0-80m	0.954	0.992	0.998	0.061	0.261	2.834	0.099
Ours-DenseNet	0-80m	0.955	0.993	0.998	0.060	0.249	2.798	0.096
Ours-ResNext	0-80m	0.956	0.993	0.998	0.059	0.245	2.756	0.096

Reference

- S. Aich, J. Vianney, M. Islam, M. Kaur and B. Liu: Bidirectional Attention Network for Monocular Depth Estimation. 2020.
- Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, Il Hong Suh: From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation. 2019
- Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger: Densely Connected Convolutional Networks. 2017
- Wang Y, Liang B, Ding M, et al. Dense Semantic Labeling with Atrous Spatial Pyramid Pooling and Decoder for High-Resolution Remote Sensing Imagery[J]. Remote Sensing. 2018.