

fast lidar-based road detection using fully convolutional neural network

lidar-camera fusion for road detection using fully convolutional neural network

两篇论文出自同一个人。

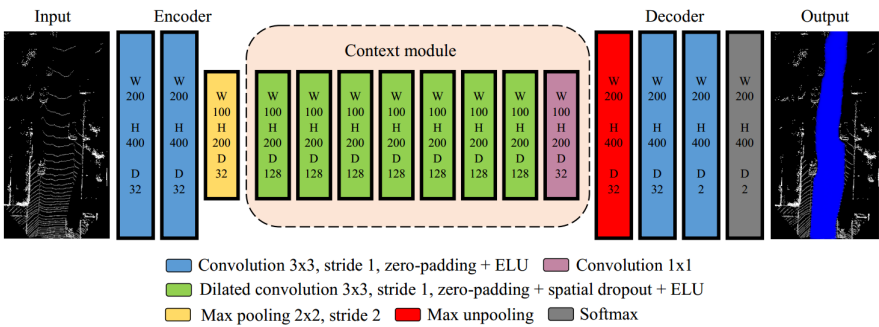
fast

输入：

- $x \in [6, 46], y \in [-10, 10]$, 栅格 0.1*0.1
- 点的个数
- 平均反射率
- 高度均值
- 高度标准差
- 最大高度
- 最小高度

200 × 400 × 6

网络结构：



输入6通道

encoder 提取特征然后maxpooling , 主要作用是下采样减小内存。

context module 聚集多尺度的上下文信息通过dilated convolutions.

decoder 上采样feature map , 通过 max-unpooling layer接两个卷积层

输出层返回一个道路置信图，每个像素代表对应lidar中栅格位置的道路的概

率

Context module：

保持参数和层数尽量少的情况下扩大感受野.

dilated convolution 提供感受野的指数级增长但是不损失分辨率 (特征图大小不变).

Multi-scale context aggregation by dilated convolutions

数据扩增：

绕 lidar z轴 $[-30,30]$, 3度步进。然后关于x轴镜像。数据量扩大了42倍。

label生成：

road从图像到点云投影不准，所以把点云投影到对应的图像视角中，决定那个点属于road.

采用相同的输入处理方式，每个cell中不再是高度等数据，而是 属于road的概率.

为了增加点云的密度，得到一个稠密的标注，点云被线性插值在狭窄的圆形区域，

Loss:

$$L = -\frac{1}{N \times W \times H} \sum_{i=1}^N \sum_{m=1}^W \sum_{n=1}^H \log p_{m,n}^i$$

p代表正确分类的概率

batch 4

在每个dilated convolution layer之后都接了spatial dropout layer (0.25)

lidar-camera

网络结构：

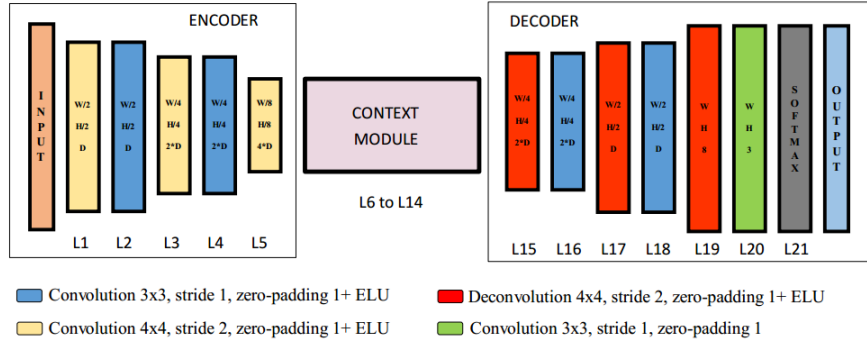
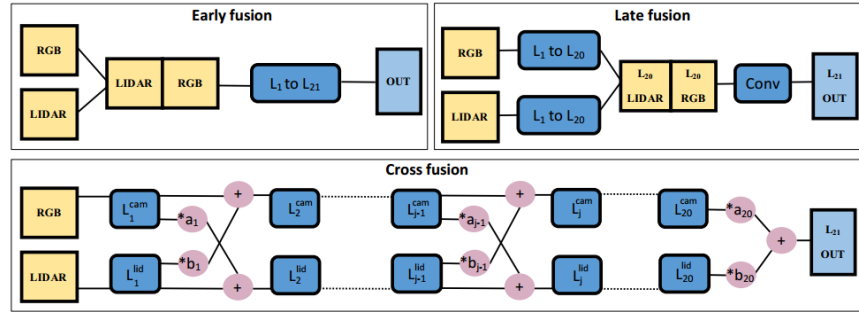


Fig. 1. A schematic illustration of the proposed base FCN architecture which consists of 21 layers. W represents the width, H denotes the height, and D is the number of feature maps in the first layer which was set to 32. The FCN uses the exponential linear unit (ELU) activation function after each convolutional layer. See Table 1 for details about the context module architecture.

主要提出了一种新的融合方法cross fusion。
在每层之后接两个可学习的参数，来融合不同层的特征。



$$I_j^{\text{Lid}} = L_{j-1}^{\text{Lid}} + a_{j-1} L_{j-1}^{\text{Cam}}$$

$$I_j^{\text{Cam}} = L_{j-1}^{\text{Cam}} + b_{j-1} L_{j-1}^{\text{Lid}}$$

点云映射到图像平面，图像区域中很多像素对应不到点云的点，也就是点云映射到图像中是稀疏的，通常的做法是上采样lidar 图像。

上采样：Pedestrian detection combining RGB and dense LIDAR data

效果：

