# Erfnet

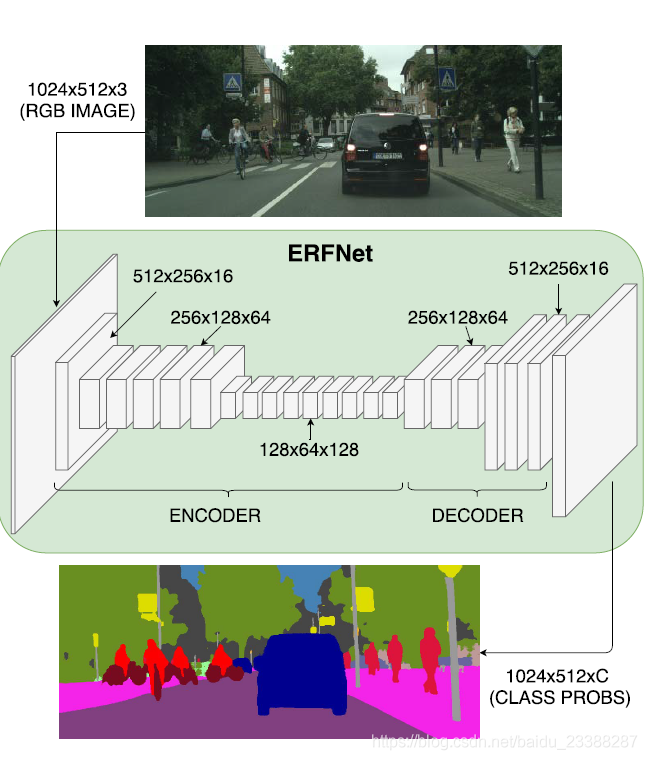
作者：马小莉

**1.背景**

语义分割能够解决自动驾驶领域很多的感知任务需求，自动驾驶需要在感知任务中采用复杂的解决方案，如识别道路，车辆，行人，交通标志等。但是当模型结构复杂，计算开销较大会存在无法部署在边缘计算核上。近些年在分类任务上的精度不断增加，但是很难应用于真实应用（在真实环境中受设备算力限制）。此外，能够满足实时分割任务的网络，却牺牲了过多的精度。erfnet在效率和准确性之间的实现了平衡。

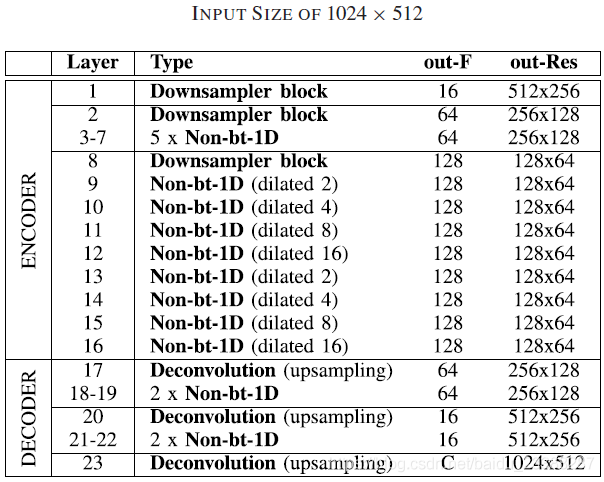
**2.模型结构**

模型依然采用了语义分割模型最经典的结构：**Encoder-Decoder**



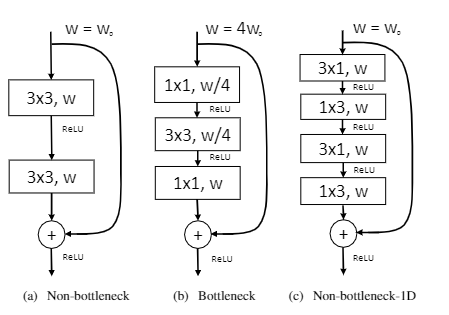
**详细的网络结构如下：**

整个网络包含**23**层，其中**1-16**层为Encoder, **17-23**层为Decoder。



**实现参数量减少的trick:**

如图所示： 下图a和b均为ResNet提出的模块，两者有着相似的参数和接近的精度。但是，bottleneck需要更少的计算资源，随着深度增加这个特点更加划算，因此更加通用。但是，non-bottleneck模块能够获得更好的精度，并且bottleneck仍存在退化问题。 于是erfnet作者提出了一个新的版本，如下图c所示。新版本完全使用了1D 卷积。



作者引入Non-bottleneck-1D概念，并能再不影响模型精度的情况下减少参数量是因为：如下图所示,由N个神经元到M个神经元，需要N\*1\*M=W个权重，当再M和N之间增加一层，如下图右边所示，权重则变成了U和V，我们可以得出此时的权重参数为：V=N\*1\*K,U=K\*1\*M,则这两层一共需要：K\*(M+N)个参数，只要K较小，则可以得到：K\*(M+N)<N\*M，同时作者在2层1D卷积中增加了非线性函数，理论上也增加了模型的学习能力

