**Deep Snake for Real-Time Instance Segmentation**

**Mask R-CNN的局限性**

基本的pipeline是先检测物体矩形框，然后在矩形框中做逐像素分割，但是这样的pipeline其实有一些limitation:

(1)在矩形框中做逐像素分割会受限于矩形框的准确度。如果矩形框本来就不准，比如没有完全覆盖物体，那就算框中的分割做的再好，也无法得到正确的instance mask。

(2)逐像素分割计算量其实很大，后续处理计算量也大。因为逐像素分割计算量大，所以网络一般将矩形框区域downsample为28x28的网格，然后进行分割，之后再把分割结果upsample到原图大小。这个upsample根据Mask R-CNN论文里的统计是有15ms的，比较费时。还有个问题是，在28x28网格上做分割会损失精度。即使28x28的网格上的分割结果完全正确，upsample到原图的mask仍然是很粗糙的。