

# Focal Loss for Dense Object Detection

- 作者：Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar. [Facebook AI Research (FAIR)]

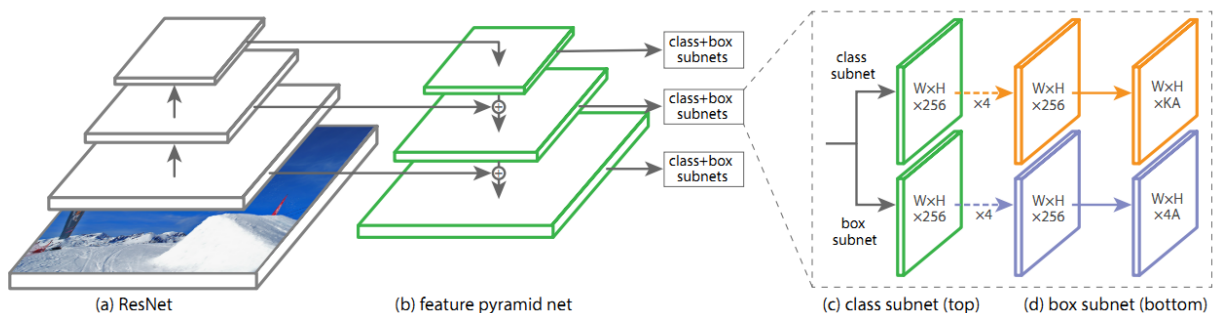
- 本文先分析了当前流行的检测框架，主要分为两类：
  - Faster RCNN (two stage): 先提取Proposals，再进行分类和回归；检测效果好，实时性较差；
  - Yolo、SSD (one stage): 直接出结果，不用Proposals；速度快，效果较two stage差。
- 分析如下：
  - Two stage方法速度慢主要是因为提取proposals后，每个ROI都需要进行一次ROI pooling+分类+回归，速度较慢；而One stage方法将proposals提取固定在预先设定的网格中，直接一次前向对proposals进行分类、回归，所以速度较快；
  - One stage方法虽简单粗暴，但是会造成proposals正负样本比差距较大，因为训练时没有对负样本进行抑制；而Two stage方法训练时，对于提取到的proposals会进行一定比例的筛选抑制，保持一定的正负样本比例，使得训练较为稳定，效果较好；

- 文章提出了新的损失函数：Focal loss，可用于缓解正负样本比例差距较大的问题；提出一个新的网络结构：Retina Net，借助Focal loss，Retina Net能够兼顾One stage方法的速度和Two stage方法的效果。

- Focal loss:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

- $p_t$ 是softmax输出的置信度，当置信度越高，该sample对应的loss就越低，使网络偏向于拟合困难样本；
  - 对于简单样本， $p_t$ 较大loss较小；对于困难、易错样本， $p_t$ 较小loss较大（指数级的差距），使得网络更加focus困难样本；不过，对于网络中已识别错误但 $p_t$ 较大的样本，纠错能力有限；
  - $\gamma$ 是可调参数，当 $\gamma = 0$ ，Focal loss就成为交叉熵，随着 $\gamma$ 增大，调节效果越明显，文章中取2效果最好。
- Retina Net:



- 使用“ResNet” + “Feature Pyramid Network”【Feature pyramid networks for object detection. In CVPR, 2017.】作为特征提取网络；
- “FPN”可以提取多尺度特征，对于不同尺度的特征设计不同大小、比例的anchor；
- 不通过ROI pooling，直接对多尺度特征进行分类+回归，网络同“Faster RCNN”。