



中国科学技术大学

Focal Loss for Dense Object Detection

Facebook AI Research (FAIR)





论文简介

理论方案

实验结果



论文简介

问题引入:

目前目标检测的框架一般分为两种：基于候选区域的two-stage的检测框架（r-cnn系列），基于回归的one-stage的检测框架（yolo、ssd系列），two-stage效果好但是速度慢，one-stage速度快但是效果差。

为什么one-stage的检测器准确率不高？作者给出的解释是由于正负样本不均衡的问题。样本中会存在大量的easy examples，且都是负样本(属于背景的样本)。这样，en masse easy negative examples会对loss起主要贡献作用，进而主导梯度的更新方向。网络无法学习有用的信息，无法对object进行准确分类。

- (1) training is inefficient as most locations are easy negatives that contribute no useful learning signal;
- (2) en masse, the easy negatives can overwhelm training and lead to degenerate models.



论文简介

负样本数量太大，占总的loss的大部分，而且多是容易分类的，因此使得模型的优化方向并不是我们所希望的那样。先前也有一些算法来处理类别不均衡的问题，比如OHEM（online hard example mining），OHEM算法虽然增加了错分类样本的权重，但是OHEM算法忽略了容易分类的样本。

针对类别不均衡问题，作者提出一种新的损失函数：focal loss，这个损失函数是在标准交叉熵损失基础上修改得到的。这个函数可以通过减少易分类样本的权重，使得模型在训练时更专注于难分类的样本。为了证明focal loss的有效性，作者设计了一个dense detector：RetinaNet，并且在训练时采用focal loss训练。实验证明RetinaNet不仅可以达到one-stage detector的速度，也能有two-stage detector的准确率。



解决方案: Focal loss

(1) 常用的交叉熵损失

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases}$$

其中, y 表示实际的类别概率值, p 为分类所得到的类别概率。为方便表示, 使用 p_t 代替。

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise.} \end{cases}$$

(2) 对于正负样本不均衡

$$CE(p_t) = -\alpha_t \log(p_t)$$

(3) 对于难分类与易分类样本不均衡

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

(4) 得到最终的focal loss表达式

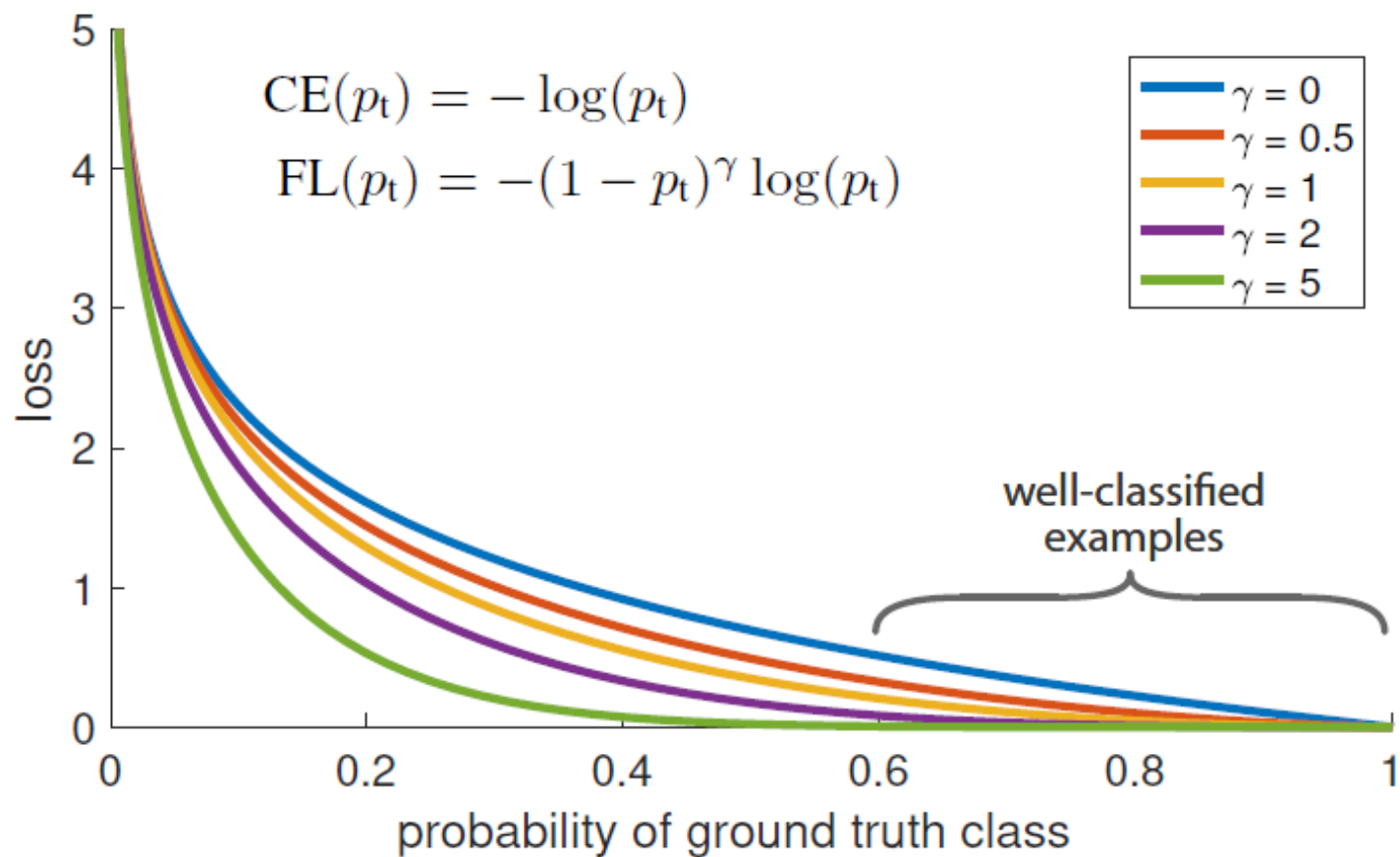
$$FL(p_t) = -\alpha^t (1-p_t)^\gamma \log(p_t)$$



误差性能曲线:

γ : 加权系数

随着系数的增加, 在易分类的区域 (分类概率为0.6-1.0的区域), 其loss减小。





实验框架：Resnet + FPN

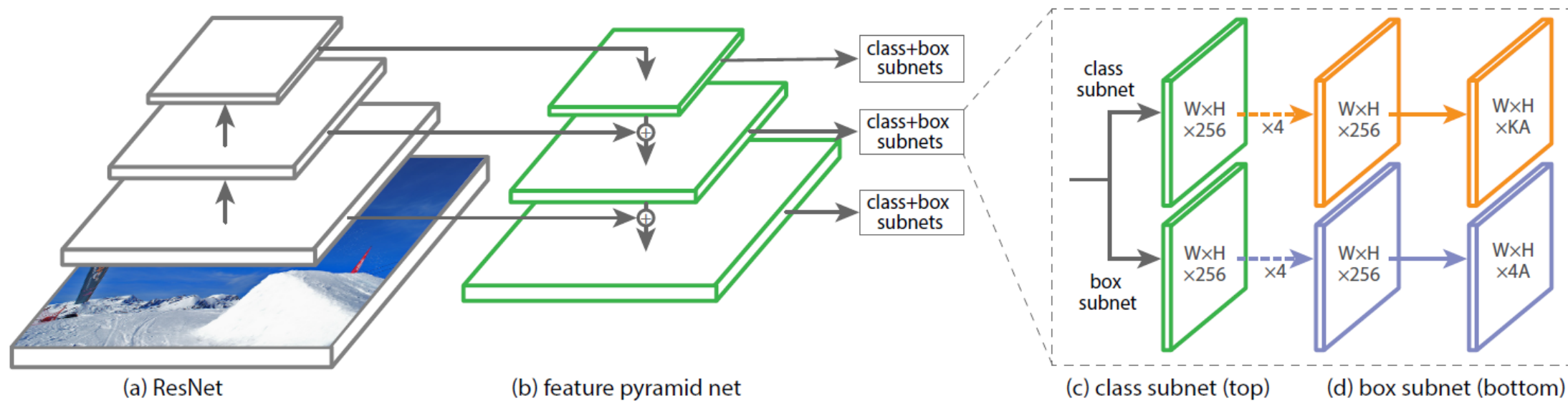


Figure 3. The one-stage **RetinaNet** network architecture uses a Feature Pyramid Network (FPN) [20] backbone on top of a feedforward ResNet architecture [16] (a) to generate a rich, multi-scale convolutional feature pyramid (b). To this backbone RetinaNet attaches two subnetworks, one for classifying anchor boxes (c) and one for regressing from anchor boxes to ground-truth object boxes (d). The network design is intentionally simple, which enables this work to focus on a novel focal loss function that eliminates the accuracy gap between our one-stage detector and state-of-the-art two-stage detectors like Faster R-CNN with FPN [20] while running at faster speeds.



实验框架: Resnet + FPN

作者为了测试所提出的损失函数的性能，在网络结构上没有做过多的设计。检测所利用的网络结构是 Resnet + FPN，设计了两路分支分别用来得到检测框以及检测结果的置信度，并将此结构命名为RetinaNet。

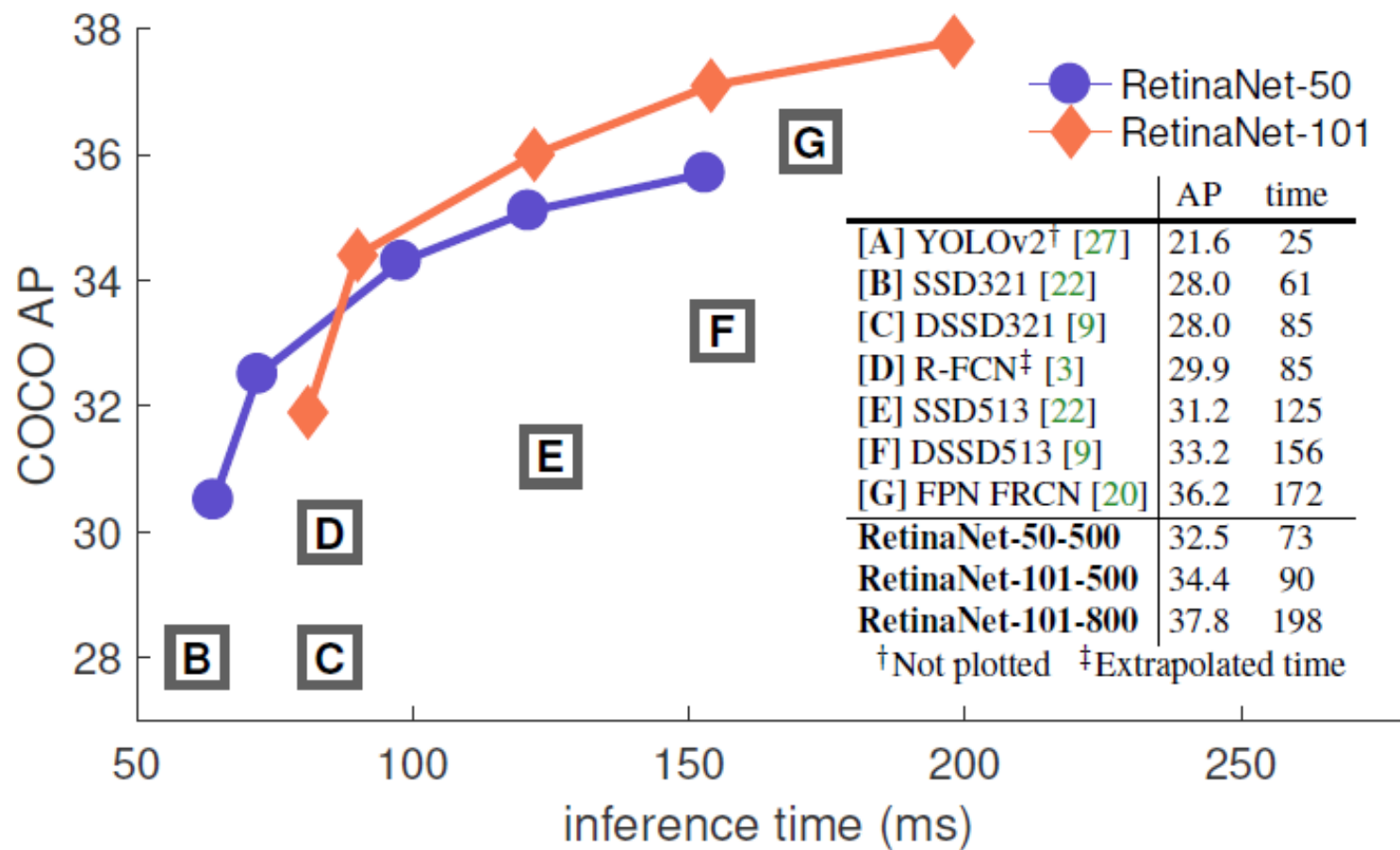
Anchors:作者用了translation-invariant anchor boxes 平移不变anchor，在每个金字塔层，作者用的长宽比是{ 1:2, 1:1, 2:1 }。在每层，对于三个长宽比的anchor，加了anchor的形状的{ 2^0 , $2^{1/3}$, $2^{2/3}$ }的anchor。对于每层，有A=9个anchor。

Classification Subnet:分类子网络在每个空间位置，为A个anchor和K个类别，预测目标存在的概率。子网络的参数在整个金字塔的层间共享。

Box Regression Subnet:与object classification子网络平行，作者在金字塔每个层都接到一个的FCN上，意图回归每个anchor box对邻近ground truth object的偏移量。



实验结果:





实验结果:

α	AP	AP ₅₀	AP ₇₅
.10	0.0	0.0	0.0
.25	10.8	16.0	11.7
.50	30.2	46.7	32.8
.75	31.1	49.4	33.0
.90	30.8	49.7	32.3
.99	28.7	47.4	29.9
.999	25.1	41.7	26.1

(a) Varying α for CE loss ($\gamma = 0$)

γ	α	AP	AP ₅₀	AP ₇₅
0	.75	31.1	49.4	33.0
0.1	.75	31.4	49.9	33.1
0.2	.75	31.9	50.7	33.4
0.5	.50	32.9	51.7	35.2
1.0	.25	33.7	52.0	36.2
2.0	.25	34.0	52.5	36.5
5.0	.25	32.2	49.6	34.8

(b) Varying γ for FL (w. optimal α)

#sc	#ar	AP	AP ₅₀	AP ₇₅
1	1	30.3	49.0	31.8
2	1	31.9	50.0	34.0
3	1	31.8	49.4	33.7
1	3	32.4	52.3	33.9
2	3	34.2	53.1	36.5
3	3	34.0	52.5	36.5
4	3	33.8	52.1	36.2

(c) Varying anchor scales and aspects

method	batch size	nms thr	AP	AP ₅₀	AP ₇₅
OHEM	128	.7	31.1	47.2	33.2
OHEM	256	.7	31.8	48.8	33.9
OHEM	512	.7	30.6	47.0	32.6
OHEM	128	.5	32.8	50.3	35.1
OHEM	256	.5	31.0	47.4	33.0
OHEM	512	.5	27.6	42.0	29.2
OHEM 1:3	128	.5	31.1	47.2	33.2
OHEM 1:3	256	.5	28.3	42.4	30.3
OHEM 1:3	512	.5	24.0	35.5	25.8
FL	n/a	n/a	36.0	54.9	38.7

(d) FL vs. OHEM baselines (with ResNet-101-FPN)

depth	scale	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	time
50	400	30.5	47.8	32.7	11.2	33.8	46.1	64
50	500	32.5	50.9	34.8	13.9	35.8	46.7	72
50	600	34.3	53.2	36.9	16.2	37.4	47.4	98
50	700	35.1	54.2	37.7	18.0	39.3	46.4	121
50	800	35.7	55.0	38.5	18.9	38.9	46.3	153
101	400	31.9	49.5	34.1	11.6	35.8	48.5	81
101	500	34.4	53.1	36.8	14.7	38.5	49.1	90
101	600	36.0	55.2	38.7	17.4	39.6	49.7	122
101	700	37.1	56.6	39.8	19.1	40.6	49.4	154
101	800	37.8	57.5	40.8	20.2	41.1	49.2	198

(e) Accuracy/speed trade-off RetinaNet (on test-dev)

Table 1. Ablation experiments for RetinaNet and Focal Loss (FL). All models are trained on trainval35k and tested on minival unless noted. If not specified, default values are: $\gamma = 2$; anchors for 3 scales and 3 aspect ratios; ResNet-50-FPN backbone; and a 600 pixel train and test image scale. (a) RetinaNet with α -balanced CE achieves at most 31.1 AP. (b) In contrast, using FL with the same exact network gives a 2.9 AP gain and is fairly robust to exact γ/α settings. (c) Using 2-3 scale and 3 aspect ratio anchors yields good results after which point performance saturates. (d) FL outperforms the best variants of online hard example mining (OHEM) [31, 22] by over 3 points AP. (e) Accuracy/Speed trade-off of RetinaNet on test-dev for various network depths and image scales (see also Figure 2).



实验结果:

为了更好的观察focal loss在reweighting example的效果，作者随机选取了 10^7 个负样本框和 10^5 个正样本框，然后通过网络之后分别计算这些正负样本的loss，最后，分别对于正样本和负样本，把所有框的loss进行归一化(softmax)，画出累计loss随样本数目的增长曲线。

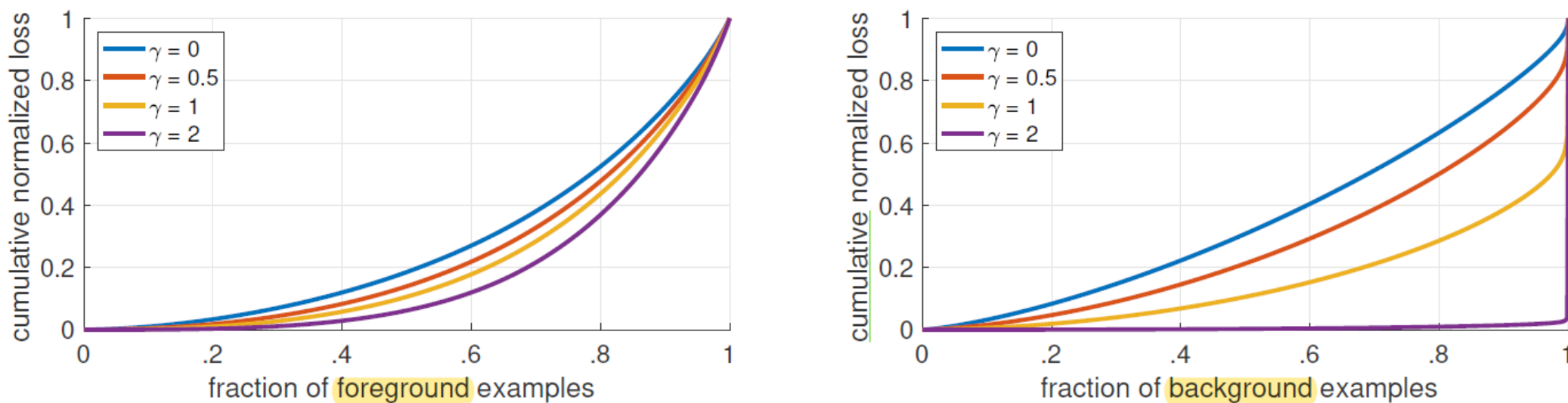


Figure 4. Cumulative distribution functions of the normalized loss for positive and negative samples for different values of γ for a *converged* model. The effect of changing γ on the distribution of the loss for positive examples is minor. For negatives, however, increasing γ heavily concentrates the loss on hard examples, focusing nearly all attention away from easy negatives.