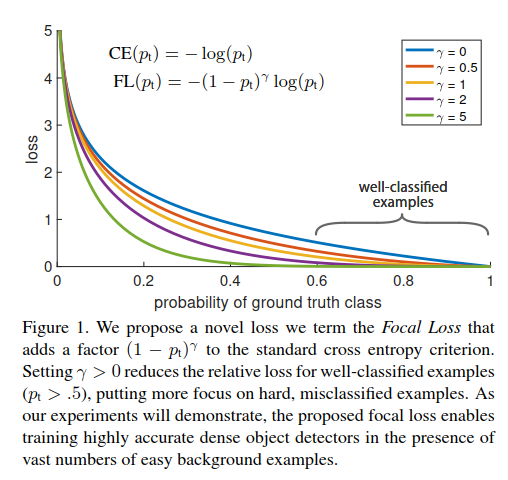
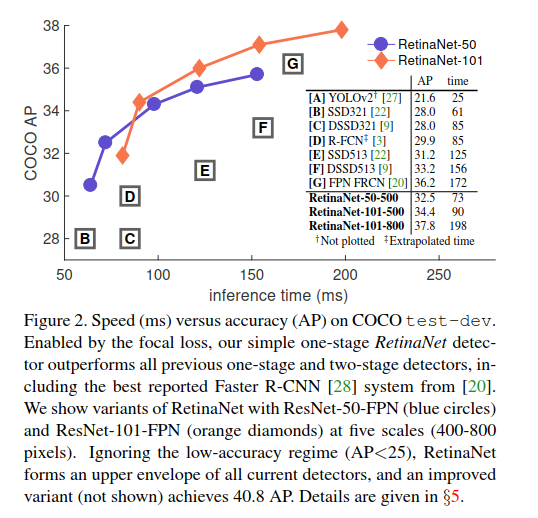
1.成果展示



改进了交叉熵损失函数，使得损失函数的能够更好的收敛，并且，给出了改进方案的多参数值比较。



Retina网络，比以往的one-stage和two-stage都还要好，有更高的识别准确率。同时也比较了多组Retina网络结构。

2．We discover that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause。

we identify class imbalance during training as the main obstacle impeding one-stage detector from achieving state-of-the-art accuracy and propose a new loss function that eliminates this barrier。

发现one-stage的方法，准确率低的问题在于，前景背景提取的失衡，当在训练一个致密的检测器时。

3. Our novel Focal Loss focuses **training on a sparse set of hard examples** and **prevents the vast number of easy negatives** from overwhelming the detector during training

4. In this paper, we propose a new loss function that acts as a more effective alternative to previous approaches for dealing with class imbalance. The loss function is a dynamically scaled cross entropy loss, where the scaling factor decays to zero as conﬁdence in the correct class increases.Intuitively, this scaling factor can automatically down-weight the contribution of easy examples during training and rapidly focus the model on hard examples.

十分直观，在优化过程中，降低了一些易于分类的样本的权重。

5. RetinaNet：named for its dense sampling of object locations in an input image.Its design features an efﬁcient in-network feature pyramid and use of anchor boxes

RetinaNet使用了特征金子塔结构和锚点盒技术。

6. These detectors have been tuned for speed but their accuracy trails that of two-stage methods

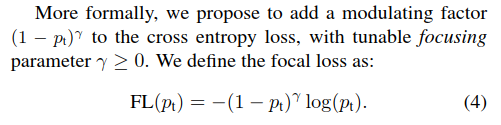
说明one-stage的特点。

7. In contrast, the aim of this work is to understand if one-stage detectors can match or surpass the accuracy of two-stage detectors while running at similar or faster speeds. 本文目的

8. class imbalance case two problem：(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.

One-stage中的分类不平衡，会产生的**两个问题**。

9.**Focal Loss**



We note two properties of the focal loss：

(1) When an example is misclassiﬁed and pt is small, the modulating factor is near 1 and the loss is unaffected. As pt → 1, the factor goes to 0 and the loss for well-classiﬁed examples is down-weighted.

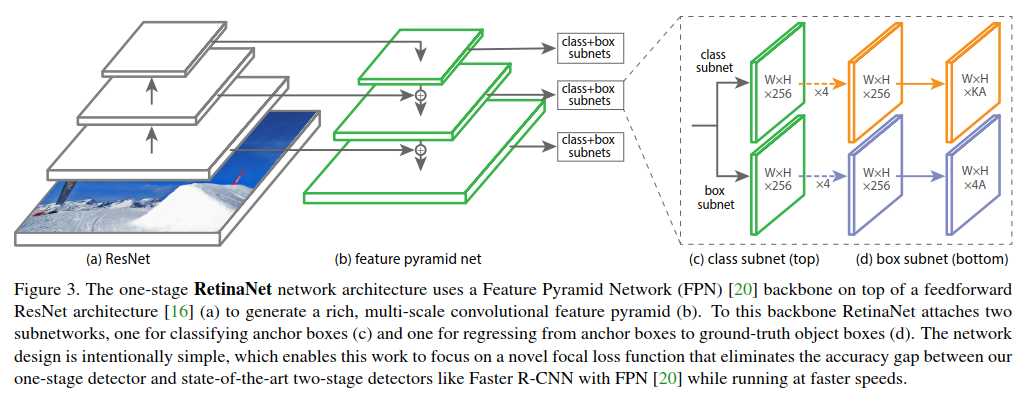
(2) The focusing parameter γsmoothly adjusts the rate at which easy examples are downweighted. When γ = 0, FL is equivalent to CE, and as γ is increased the effect of the modulating factor is likewise increased (we found γ = 2 to work best in our experiments).

10.Two-Stage如何**消除Class Imbalance**

(1) a two-stage cascade and

(2) biased minibatch sampling.（for instance, a 1:3 ratio of positive to negative examples）

11.**RetinaNet**的网络结构

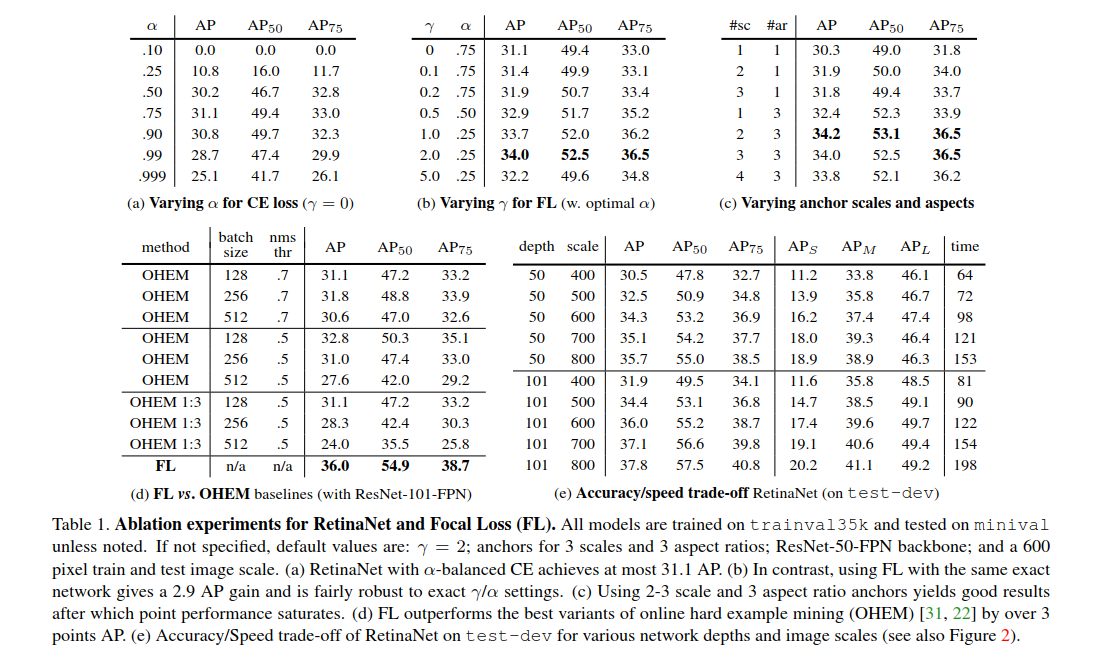


12.To improve speed, we only decode box predictions from at most 1k top-scoring predictions per FPN level, after thresholding detector conﬁdence at 0.05. The top predictions from all levels are merged and non-maximum suppression with a threshold of 0.5 is applied to yield the ﬁnal detections.

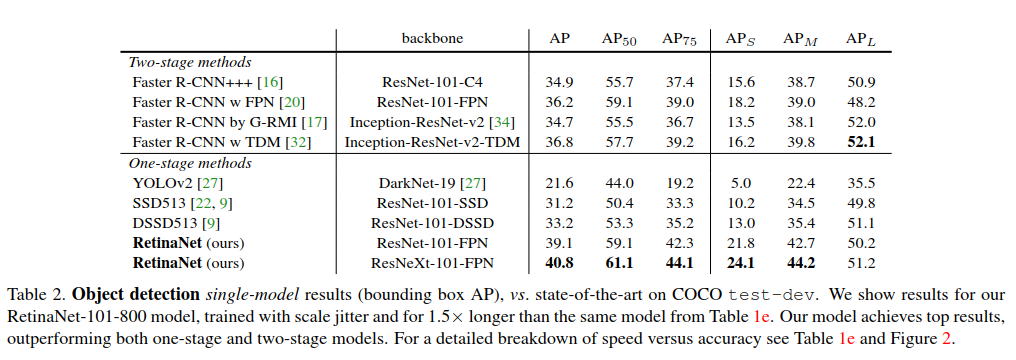
13.**权值初始化（Initialization）**：All new conv layers except the ﬁnal one in the RetinaNet subnets are initialized with bias b = 0 and a Gaussian weight ﬁll with σ = 0.01. For the ﬁnal conv layer of the classiﬁcation subnet, we set the bias initialization to b = − log((1 − π)/π).

14.**优化器训练模型（Optimization）**: RetinaNet is trained with stochastic gradient descent (SGD). We use synchronized SGD over 8 GPUs with a total of 16 images per minibatch (2 images per GPU).Unless otherwise speciﬁed, all models are trained for 90k iterations with an initial learning rate of 0.01, which is then divided by 10 at 60k and again at 80k iterations。

15.实验结果对比



16.跟其他模型的比较



17.**Focal Loss 的效果**：If we observe the positive samples, we see that the CDF looks fairly similar for different values of γ.For example, approximately 20% of the hardest positive samples account for roughly half of the positive loss, as γ increases more of the loss gets concentrated in the top 20% of examples, but the effect is minor. The effect of γ on negative samples is dramatically different. For γ = 0, the positive and negative CDFs are quite similar. However, as γ increases, substantially more weight becomes concentrated on the hard negative examples