1.For object detection, the two-stage approach (e.g.,Faster R-CNN) has been achieving the highest accuracy, whereas the one-stage approach (e.g., SSD) has the advantage of high efﬁciency. To inherit the merits of both while overcoming their disadvantages, in this paper, we propose a novel single-shot based detector, called ReﬁneDet.

结合two-stage精度优势和one-stage的速度优势。提出了新的检测器，命名为RefineDet。

2.two module（**两个module的作用**）:

(1) the anchor reﬁnement module**(ARM)**

①ﬁlter out negative anchors to reduce search space for the classiﬁer②coarsely adjust the locations and sizes of anchors to provide better initialization for the subsequent regressor

(2) object detection module**(ODM)**

①aims to regress accurate object locations and predict multi-class labels based on the reﬁned anchors

3. In our opinion, the current state-of-the-art two-stage Methods have three advantages over the one-stage methods as follows:

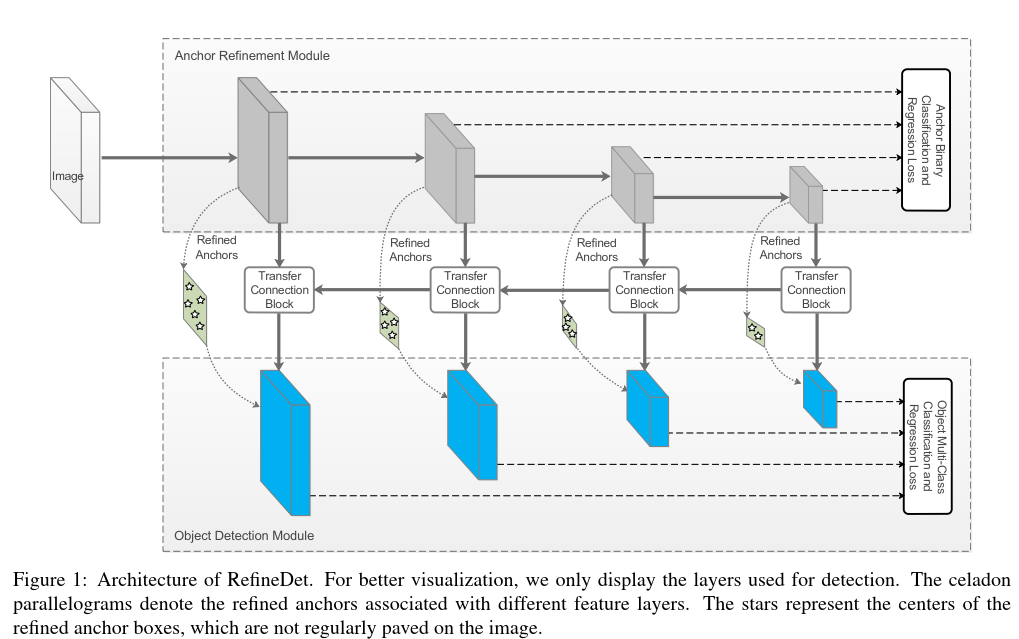
(1)using two-stage structure with sampling heuristics to handle class imbalance.

(2)using two-step cascade to regress the object box parameters

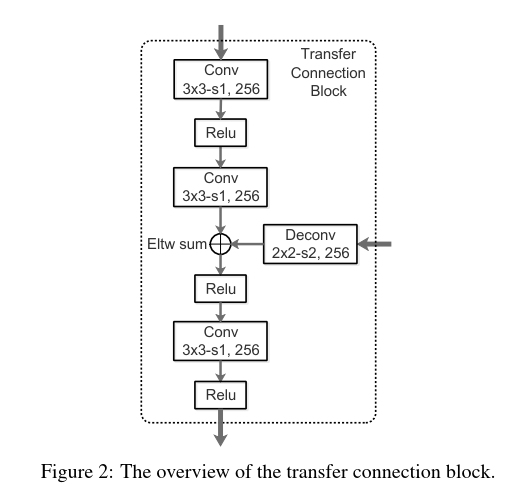
(3)using two-stage features to describe the objects

关于（3）的解读：

In case of Faster R-CNN, the features (excluding shared features) in the ﬁrst stage (i.e., RPN) are trained for binary classiﬁcation (being an object or not), while the features (excluding shared features) in the second stage(i.e., Fast R-CNN) are trained for multi-class classiﬁcation (background or object classes).



4. three core components in **ReﬁneDet**

(1) transfer connection block (TCB), converting the features from the ARM to the ODM for detection; 

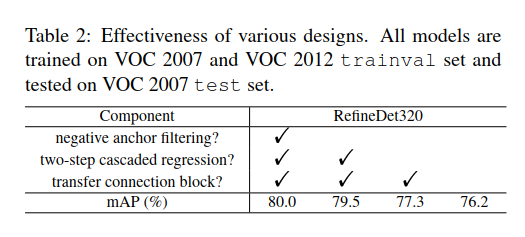
(2) two-step cascaded regression,accurately regressing the locations and sizes of objects;

第一步对ARM进行回归，对每一个feature map的grid学习四个offset，这样就可以大致预测出每个grid的anchor box。

第二步通过TCB得到ODM的feature map，对其每一个grid，学习**相对于上层得到的achor box 的offset**。和类别C。然后得到新的achor box和类型。

(3)negative anchor ﬁltering, early rejecting well-classiﬁed negative anchors and mitigate the imbalance issue.

**对于ARM得到的AnchorBox**，如果属于背景的置信度太高，比如0.99，则将其**抛弃**，在**ODM层不再回归**。。if its negative conﬁdence is larger than a preset threshold θ (i.e., set θ = 0.99empirically), we will discard it in training the ODM。



5.训练实现细节

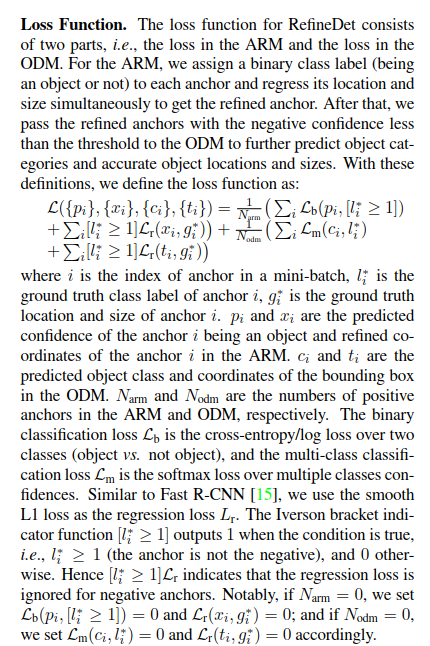
**（1）Data Augmentation**：参考了SSD的图像增强处理

**（2）Backbone Network**：VGG-16 and ResNet-101

**（3）Anchors Design and Matching**：we select four feature layers with the total stride sizes 8, 16, 32, and 64 pixels for both VGG-16 and ResNet-101. three aspect ratios (i.e., 0.5, 1.0, and 2.0)

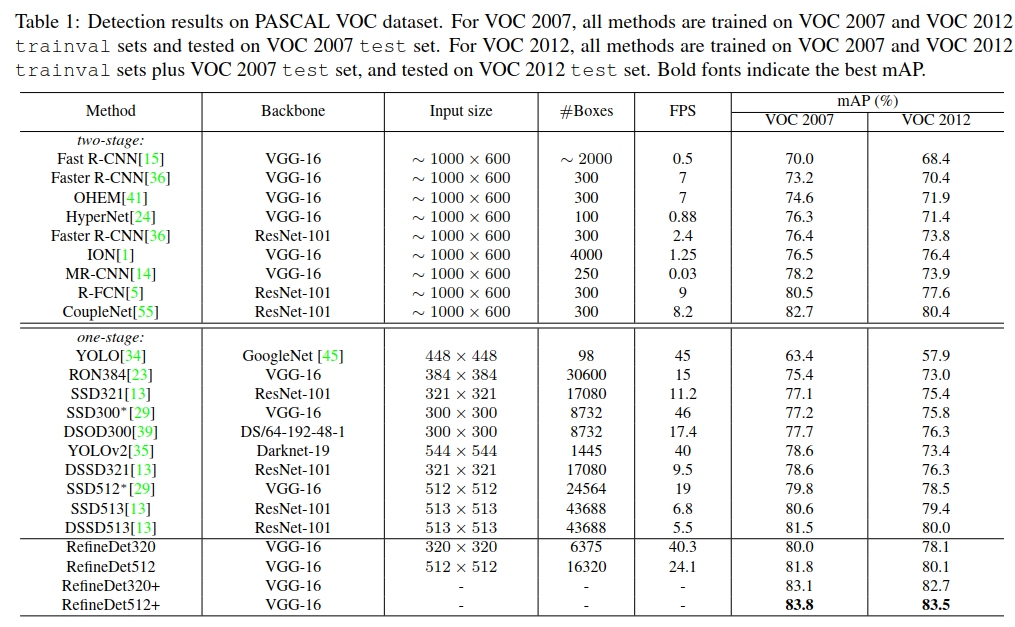
**（4）Hard Negative Mining**:和SSD类似

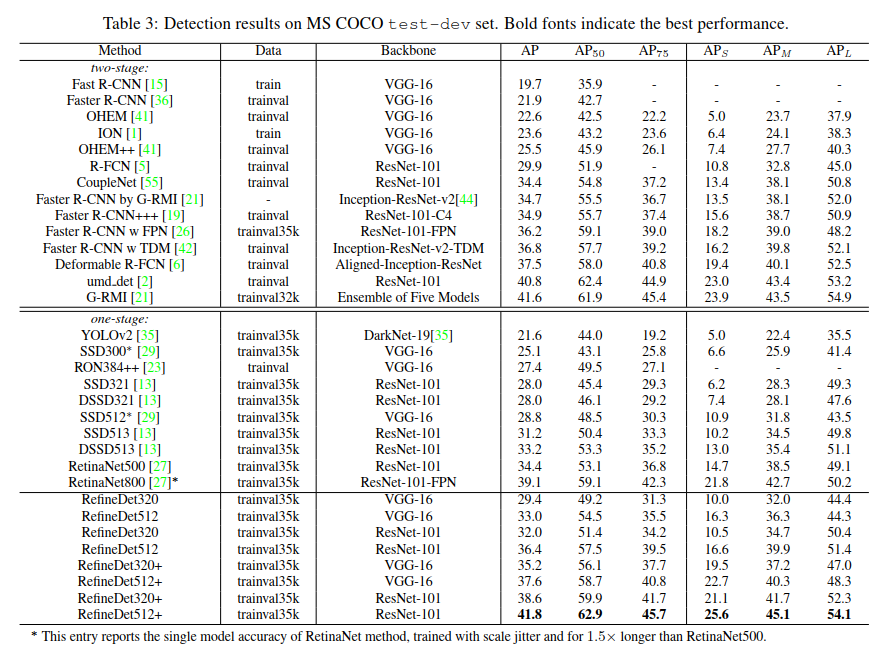
**（5）Loss Function：相当于2次SSD，一次模糊，一次精校准**



**（6）Optimization**：the “xavier” method进行初始化，VGG-16，a zero-mean Gaussian distribution with standard deviation 0.01初始化ResNet-101，batch：32，SGD with 0.9 momentum and 0.0005 weight decay，initial learning rate to 10-3，不同数据集上用不同的方法进行衰减。

**Run Time Performance**

****

****