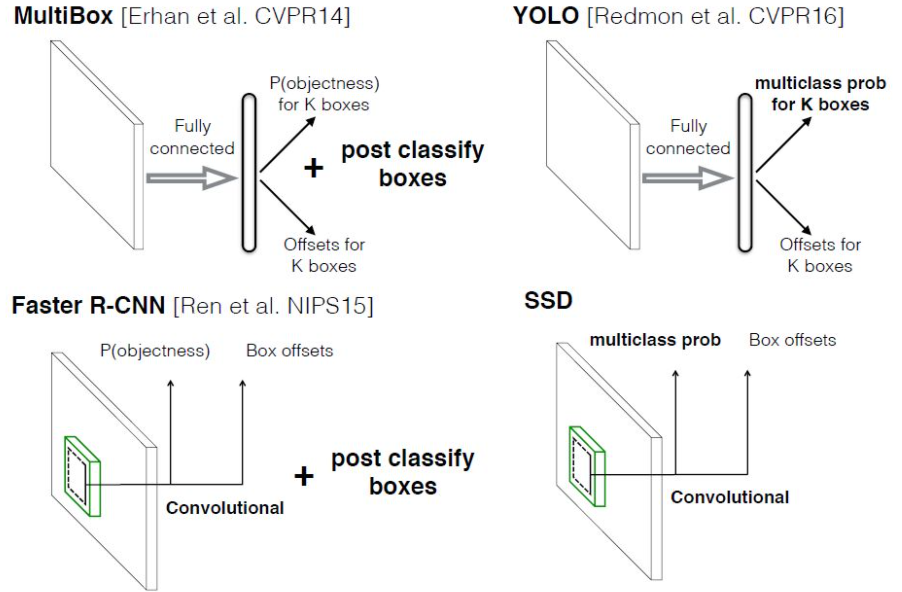
**SSD(Single Shot MultiBox Detector)**

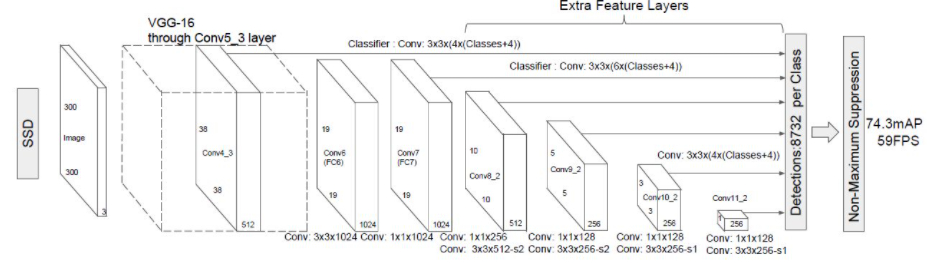
**Refer to:https://zhuanlan.zhihu.com/p/33544892**

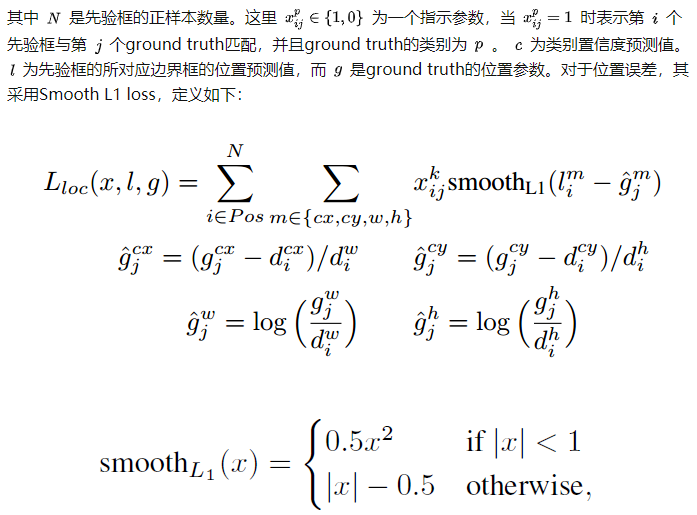
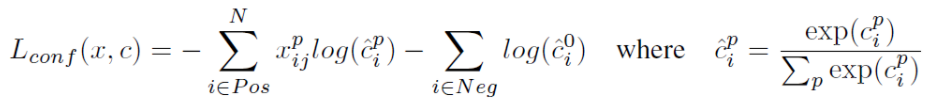
* One-stage: 均勻在圖片不同位置進行密集抽樣，抽樣採用不同尺度和長寬比，利用CNN提取特徵後直接進行分類與回歸，速度快，但訓練困難(正樣本與負樣本(背景)極其不均衡)，導致模型準確度稍低，ex) YOLO, SSD
* Two-stage: 先通過啟發式方法或CNN(RPN)產生一系列稀疏候選框，對其候選框進行分類與回歸，準確度高，ex) Faster R-CNN, Fast R-CNN, R-CNN



1. **採用多尺度特徵圖用於檢測**: 前面feature map較大(檢測小目標ex 8x8 可劃分更多單元，但每個單元先驗框尺度較小)，後面使用stride or pooling 降低feature map(檢測大目標)大小
2. **採用CNN進行檢測**: SSD直接採用CNN對不同feature map進行提取檢測結果，採用3x3 conv.得到檢測值
3. **設置先驗框(prior boxes, default boxes, anchors)**: SSD借鑑Faster R-CNN anchor理念，每個單元設置尺度or長寬比不同的先驗框，預測的BB(BoundingBox)以先驗框為基準，減少訓練難度，一般情況下，每個單元設置多個先驗框，接輸出一套獨立的檢驗值，對應BB，分兩部分: a. 各類別置信度or評分，SSD將背景也做為一個特殊類別，共c+1置信值，c個類別置信度(包含背景)，b. BB location: (cx, cy, w, h)，先驗框d= (d^cx, d^cy, d^w, d^h)，對應的BB b = (b^cx, b^cy, b^w, b^h)，預測的BB =l =( (b^cx - d^cx)/d^w, b^cy- d^cy)/d^h, log(b^w/d^w), log(b^h/d^h))，一個大小為mxn feature map，共 mn個單元，每個單元先驗框為k，每個單元需(c+4)k個預測值，所有單元需mnk(c+4)，由於SSD使用conv. 需要(c+4)k個conv.

**NN arch.**



* 以VGG16為基礎網絡，新增conv.獲得更多feature map，將VGG16的fc6 & fc7轉換成3x3 conv6 & 1x1 conv7，pool5 2x2 stride = 2 --> 3x3 stride = 1，conv6採用dilation conv. 3x3 and dilation=6帶孔卷積，不增加參數與模型複雜度擴大卷積視野， conv4\_3 feature map 38x38，增加L2 Normalization(僅對每個像素點在channel維度做歸一化)保證後面檢測層差異不大，(BN: 在三維度做歸一化)
* Conv4\_3(38x38), conv7(19x19), conv8\_2(10x10), conv9\_2(5x5), conv10\_2(3x3), conv11\_2(1x1)提取6個feature map，不同feature map設置先驗框數目不同(一個單元的先驗數目)，隨feature map大小降低，先驗框尺度線性增加，計算後各feature map尺度為 30, 60, 111, 162, 213, 264，每個feature map有6個先驗框{1,2, 3, 1/2, 1/3, 1’}，SSD300共可預測(38x38x4+19x19x6+10x10x6+5x5x6+3x3x4+1x1x4=8732 BB
* Training -- 先驗框匹配
  + 第一順序: 每個ground truth找到與其IOU最大的先驗框，保證每個ground truth一定與某個先驗框匹配，通常稱與GT匹配的先驗框為正樣本，反之，一個 先驗框沒有與任何GT進行匹配，此先驗框為負樣本，而GT很少，對應一個先驗框，而先驗框很多，導致正負樣本不平衡
  + 第二順序: 對於剩餘未匹配先驗框，if 某個GT的IOU >threshold(0.5)，此先驗框也與GT匹配，故一個GT可匹配多個先驗框，而一個先驗框只能匹配一個GT，先驗框與IOU最大GT匹配
* 一個GT可以與多個先驗框匹配，但GT太少，負樣本>>正樣本，為平衡樣本，SSD採用hard negative mining，對負樣本進行抽樣，按照置信度誤差(預測背景置信度越小，誤差越大)，進行降序排列，選取誤差較大的top-k作為訓練負樣本，保證正負樣本在1:3
* Loss function
  + 定義為位置誤差(locatization loss, loc)與置信度誤差(confidence loss, conf)的加權和:
  + 
  + Lloc位置誤差，僅針對正樣本進行計算，針對GT編碼，預測值l 也進行編碼
  + 
  + Lconf 置信度誤差採用 softmax loss
  + 
* **Data Augmentation**: 提升SSD性能，採用horizontal flip(水平翻轉)，random crop&color distortion(隨機裁剪加顏色扭曲)，randomly sample a patch(隨機採集塊域獲取小目標訓練樣本)透過zoom out來創造小的訓練樣本
* **Predict processing**: 對於每個predict box，

1. 根據類別置信度確定其類別(置信度大)與置信度值
2. 過濾屬於背景預測框，與過濾threshold(0.5)更低的預測框，留下後的預測框進行解碼，根據先驗框得到真實位置參數(解碼後須做clip，防止預測框位置超出圖片)
3. 排序僅保留top-k(如400)個預測框
4. 最後進行**NMS**，過濾重疊較大的預測框，最後就是檢測結果。

\*\*\*\*\*\*\*NMS\*\*\*\*\*\*\*\*

NMS(Non-Maximum Suppression)非極大值抑制

抑制不是極大值元素，在目標檢測中，**提取置信度高的目標檢測框，抑制置信度低的誤檢框**，目標框數量由anchor數量決定，nms用來去除重複的框(重複框定位到同一目標)，

以下為平行於座標軸:

1. **標準nms**
2. **Soft-nms**: 將相鄰低智信度檢測框分數均強制歸零，如同類目標比較密集，存在遮擋，如密集人群，用此方式處理容易產生漏洞，soft-nms吸取nms教訓，算法執行過程中，不是簡單對IoU > threshold檢測框刪除，而是降低分數
3. **Adaptive nms**: 與soft-nms解決問題一樣，而soft-nms智信度threshold = Nt主要依賴人工設定，不能適應目標壅擠程度不同的情況，adaptive nms根據目標密度來自動設定置信度threshold

以下為傾斜框(rotate box, rbox)的nms:

1. **INMS**: 當前遍歷rbox與剩餘rbox進行交集運算得到相應的相交點集合，並根據判斷相交點集合組成凸邊形面積
2. **PNMS(Polygon nms)**: 物體無法用矩形表達，而是多邊形，計算IoU方式相同，得到兩個多邊形相交點集，判斷相交點集合組成的凸邊形面積，計算兩多邊形IoU，可使用boost酷的geometry的polygon相關函數實現，也可使用shapely.geometry的polygen
3. **MNMS(mask nms)**: 分割區域

\*\*\*\*\*\*\*NMS\*\*\*\*\*\*\*\*