目标检测

RetinaNet-iccv17

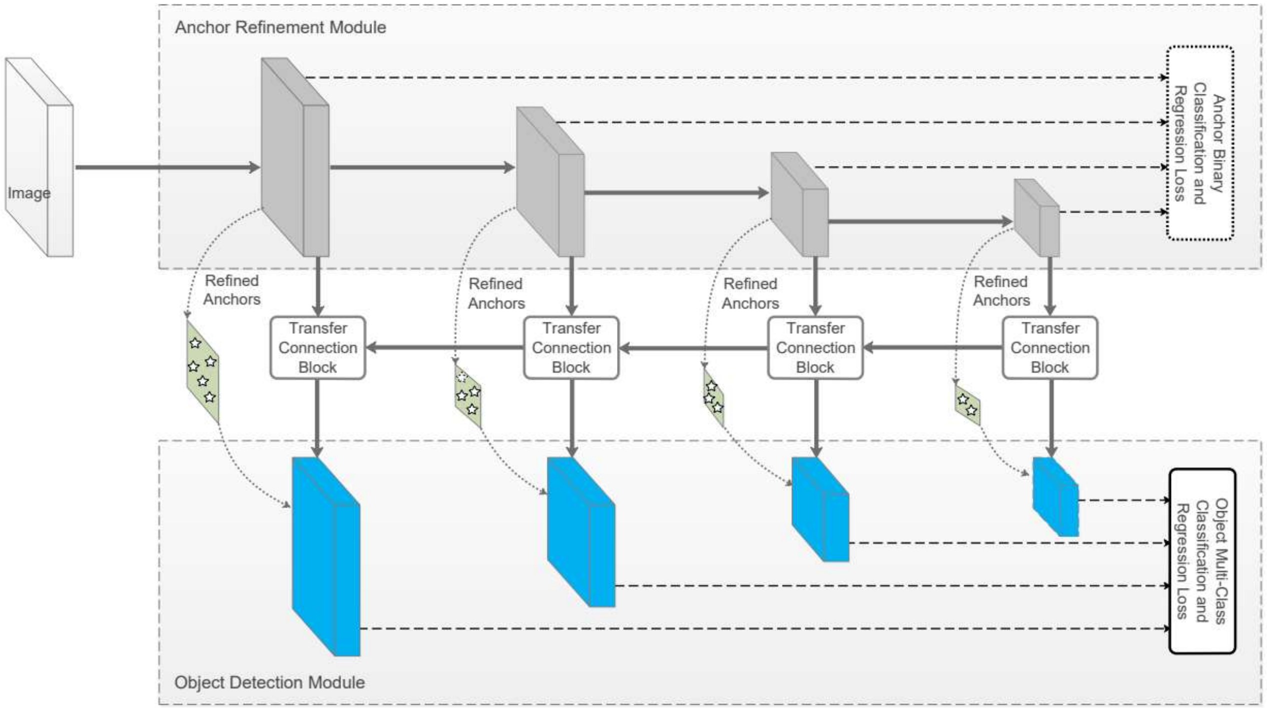
- focal loss

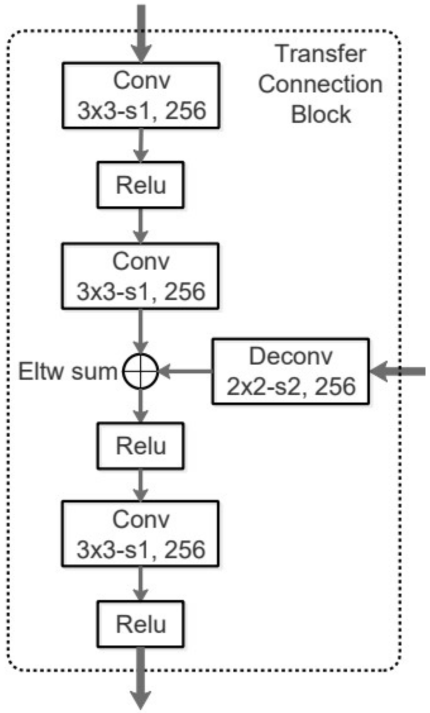
$$FL(p_t) = -\alpha_t(1 - p_t)^{\gamma} \log(p_t).$$

loss的量级	数量大的类别：backgroud	数量少的类别：foregroud
正确分类loss值	大幅度下降	稍微下降
错误分类loss值	稍微下降	基本不变

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17

RefineDet CVPR'18





ARM：1.kernel_size = 3x3, stride = 1 ,channel = num_anchor✕4（坐标回归）
2.kernel_size = 3x3, stride = 1 ,channel = num_anchor✕2（判断前后景）

ODM：1.kernel_size = 3x3, stride = 1 ,channel = num_anchor✕4（坐标回归）
2.kernel_size = 3x3, stride = 1 ,channel = num_anchor✕num_cls（分类）

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18

Soft-NMS ICCV' 17

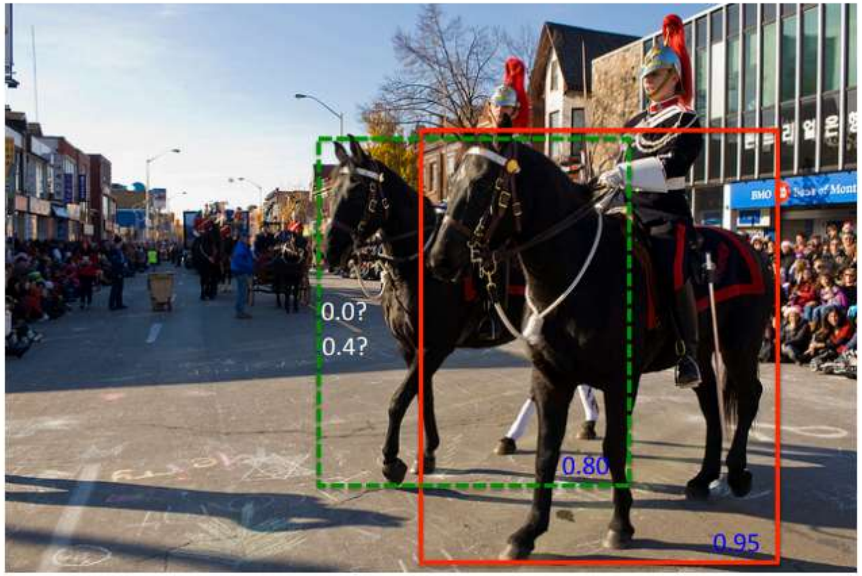


Figure 1. This image has two confident horse detections (shown in red and green) which have a score of 0.95 and 0.8 respectively. The green detection box has a significant overlap with the red one. Is it better to suppress the green box altogether and assign it a score of 0 or a slightly lower score of 0.4?

Input : $\mathcal{B} = \{b_1, .., b_N\}, \mathcal{S} = \{s_1, .., s_N\}, N_t$
$\mathcal{B}$ is the list of initial detection boxes
$\mathcal{S}$ contains corresponding detection scores
N_t is the NMS threshold

begin
$\mathcal{D} \leftarrow \{\}$
while $\mathcal{B} \neq empty$ **do**
$m \leftarrow \text{argmax } \mathcal{S}$
$\mathcal{M} \leftarrow b_m$
$\mathcal{D} \leftarrow \mathcal{D} \bigcup \mathcal{M}; \mathcal{B} \leftarrow \mathcal{B} - \mathcal{M}$
for b_i *in* $\mathcal{B}$ **do**
if $iou(\mathcal{M}, b_i) \geq N_t$ **then**
$\mathcal{B} \leftarrow \mathcal{B} - b_i; \mathcal{S} \leftarrow \mathcal{S} - s_i$
end
NMS
$s_i \leftarrow s_i f(iou(\mathcal{M}, b_i))$
Soft-NMS
end
end
return $\mathcal{D}, \mathcal{S}$
end

$$s_i = s_i e^{-\frac{\mathrm{iou}(\mathcal{M}, b_i)^2}{\sigma}}, \forall b_i \notin \mathcal{D}$$

不要粗鲁地删除所有IOU大于阈值的框，而是降低其置信度。
对密集目标提升较大，非密集目标基本没提升。计算量增加。

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18
Soft-NMS	比普通NMS提升1点	ICCV' 17

Softer-NMS CVPR' 19

$$P_\Theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_e)^2}{2\sigma^2}}$$

- 预测的概率分布

$$P_D(x) = \delta(x - x_g)$$

- 真值，概率分布

- 两个分布采用KL损失

- 当$|x_g-x_e| <=1$

$$L_{reg} \propto \frac{e^{-\alpha}}{2}(x_g - x_e)^2 + \frac{1}{2}\alpha$$

- 当$|x_g-x_e| >1$

$$L_{reg} = e^{-\alpha}(|x_g - x_e| - \frac{1}{2}) + \frac{1}{2}\alpha$$

Softer-NMS CVPR' 19

Algorithm 1 var voting

$\mathcal{B}$ is $N \times 4$ matrix of initial detection boxes. $\mathcal{S}$ contains corresponding detection scores. $\mathcal{C}$ is $N \times 4$ matrix of corresponding variances. $\mathcal{D}$ is the final set of detections. σ_t is a tunable parameter of var voting. The lines in blue and in green are soft-NMS and var voting respectively.

$\mathcal{B} = \{b_1, .., b_N\}, \mathcal{S} = \{s_1, .., s_N\}, \mathcal{C} = \{\sigma_1^2, .., \sigma_N^2\}$
$\mathcal{D} \leftarrow \{\}$
$\mathcal{T} \leftarrow \mathcal{B}$
while $\mathcal{T} \neq$ empty **do**
  $m \leftarrow \text{argmax } \mathcal{S}$
  $\mathcal{T} \leftarrow \mathcal{T} - b_m$
  $\mathcal{S} \leftarrow \mathcal{S} f(IoU(b_m, T))$ ▷ soft-NMS
  $idx \leftarrow IoU(b_m, B) > 0$ ▷ var voting
  $p \leftarrow exp(-(1 - IoU(b_m, \mathcal{B}[idx]))^2/\sigma_t)$
  $b_m \leftarrow p(\mathcal{B}[idx]/\mathcal{C}[idx])/p(1/\mathcal{C}[idx])$
  $\mathcal{D} \leftarrow \mathcal{D} \bigcup b_m$
end while
return $\mathcal{D}, \mathcal{S}$

网络的预测值

$$\{x_1, y_1, x_2, y_2, s, \sigma_{x_1}, \sigma_{y_1}, \sigma_{x_2}, \sigma_{y_2}\}$$

$$p_i = e^{-(1-IoU(b_i,b))^2/\sigma_t}$$

$$x = \frac{\sum_i p_i x_i/\sigma_{x,i}^2}{\sum_i p_i/\sigma_{x,i}^2}$$

subject to $IoU(b_i, b) > 0$

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18
Soft-NMS	比普通NMS提升1点	ICCV' 17
Softer-NMS	使用KL-loss提升2点，使用Softer-NMS比soft-NMS再提升2点	CVPR' 19

此外关于NMS的改进还有

[CVPR' 18] [Fitness NMS]

Improving Object Localization with Fitness NMS and Bounded IoU Loss

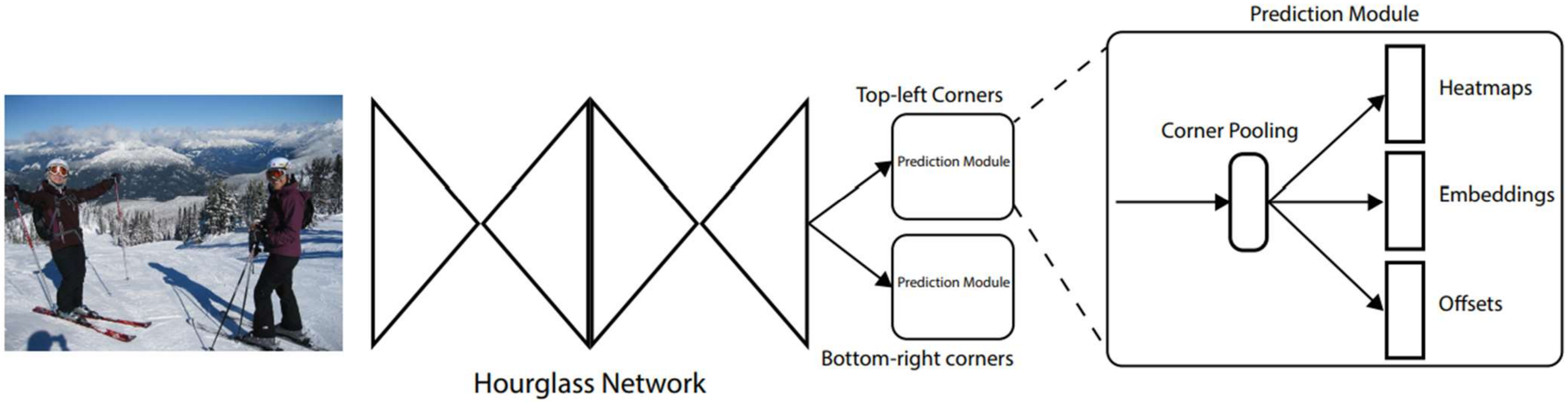
[CVPR’ 19] [Adaptive NMS]

Adaptive NMS: Refining Pedestrian Detection in a Crowd

[CVPR' 19] [MaxpoolNMS]

MaxpoolNMS: Getting Rid of NMS Bottlenecks in Two-Stage Object Detectors

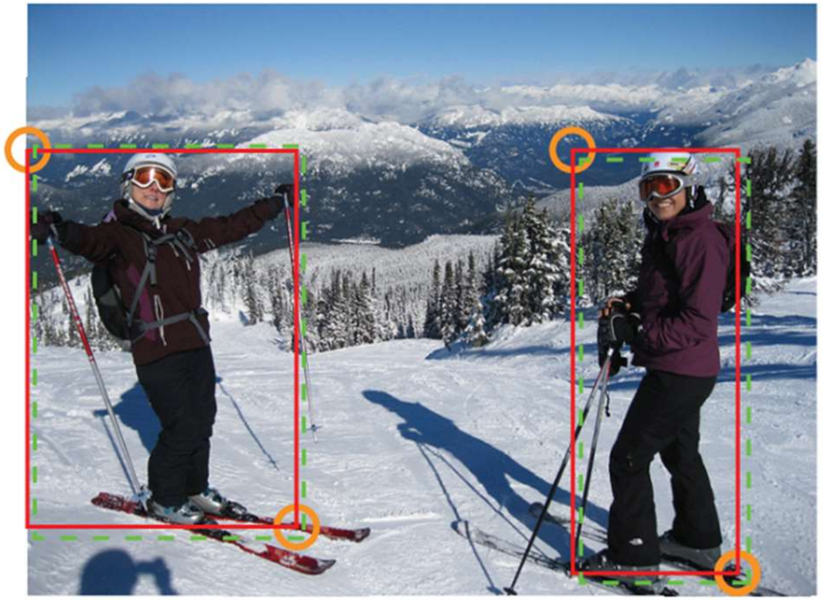
CornerNet ECCV'18



1. 边界框的左上角和右下角
2. corner的embedding vector，同一目标的两个corner的嵌入之间的距离应很小
3. corner的offset

CornerNet-Corner

- Each set of heatmaps has C channels, where C is the number of categories, and is of size H × W. There is **no background channel**.
- Each channel **is a binary mask** indicating the locations of the corners for a class.



$$e^{-\frac{x^2+y^2}{2\sigma^2}}$$

σ is 1/3 of the radius.

$$L_{det} = \frac{-1}{N} \sum_{c=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} \begin{cases} (1-p_{cij})^{\alpha} \log(p_{cij}) & \text{if } y_{cij}=1 \\ (1-y_{cij})^{\beta} (p_{cij})^{\alpha} \log(1-p_{cij}) & \text{otherwise} \end{cases} \quad (1)$$

we set α to 2 and β to 4 in all experiments

CornerNet-offset

- a location (x, y) in the image is mapped to the location

$$\left(\left\lfloor \frac{x}{n} \right\rfloor, \left\lfloor \frac{y}{n} \right\rfloor\right)$$

in the heatmaps, where n is the down sampling factor.

- predict location offsets to slightly adjust the corner locations before remapping them to the input resolution

$$\boldsymbol{o}_k = \left(\frac{x_k}{n} - \left\lfloor \frac{x_k}{n} \right\rfloor, \frac{y_k}{n} - \left\lfloor \frac{y_k}{n} \right\rfloor\right)$$

$$L_{off} = \frac{1}{N} \sum_{k=1}^{N} \mathrm{SmoothL1Loss}\left(\boldsymbol{o}_k, \hat{\boldsymbol{o}}_k\right)$$

CornerNet-Grouping Corner

- Our approach is inspired by the Associative Embedding method proposed by Newell et al.

- 网络预测每个检测到的角点的嵌入向量，使得如果左上角和右下角属于同一个边界框，则它们的嵌入之间的距离应该小。然后，我们可以根据左上角和右下角嵌入之间的距离对角点进行分组。

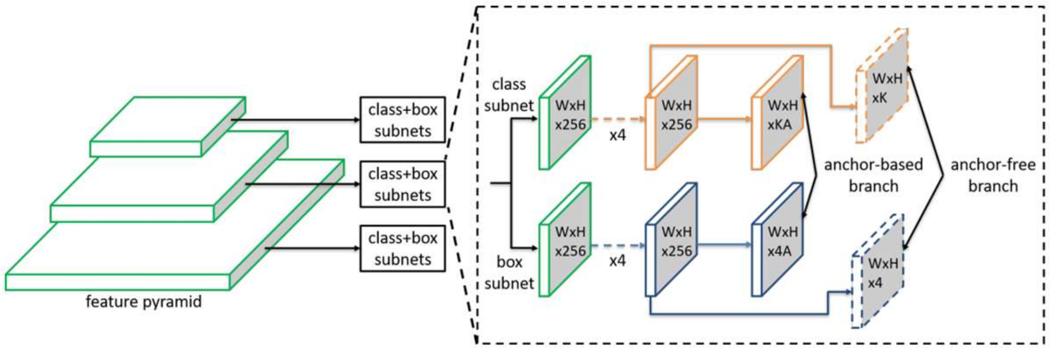
$$L_{pull} = \frac{1}{N}\sum_{k=1}^{N}\left[(e_{tk} - e_k)^2 + (e_{bk} - e_k)^2\right] \quad (4)$$

$$L_{push} = \frac{1}{N-1}\sum_{k=1}^{N}\sum_{j=1, j \neq k}^{N} max(0, \Delta - |e_k - e_j|) \quad (5)$$

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18
Soft-NMS	比普通NMS提升1点	ICCV' 17
Softer-NMS	使用KL-loss提升2点，使用Softer-NMS比soft-NMS再提升2点	CVPR' 19
CornerNet	42.1	ECCV'18

Feature Selective Anchor-Free Module for Single-Shot Object Detection CVPR'19

- feature select anchor-free module（FSAF）



定义监督信号，也就是要定义好groundtruth box和loss函数在介绍这一部分之前，需要先定义几个概念，

（1）ground truth box的类别：k；

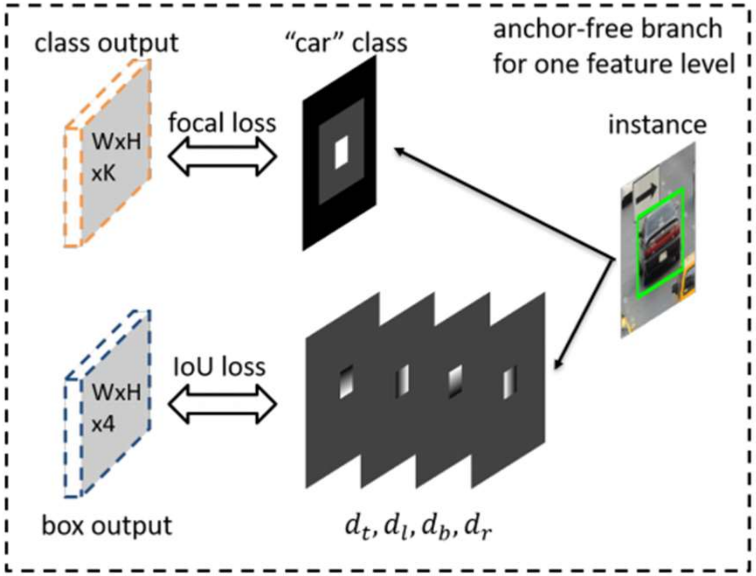
（2）ground truth box的坐标：$b = [x, y, w, h]$，其中，(x, y)表示box的center坐标；

（3）ground truth box在第l个特征层上的投影：$b_p^l = [x_p^l, y_p^l, w_p^l, h_p^l]$；

（4）effective box：$b_e^l = [x_e^l, y_e^l, w_e^l, h_e^l]$，它表示$b_p^l$的一部分，缩放比例系数$\epsilon_e = 0.2$；

（5）ignoring box：$b_i^l = [x_i^l, y_i^l, w_i^l, h_i^l]$，它也表示$b_p^l$的一部分，缩放比例系数为$\epsilon_i = 0.5$；

$x_e^l = x_p^l, y_e^l = y_p^l, w_e^l = \epsilon_e w_p^l, h_e^l = \epsilon_e h_p^l, x_i^l = x_p^l, y_i^l = y_p^l, w_i^l = \epsilon_i w_p^l, h_i^l = \epsilon_i h_p^l$. We set $\epsilon_e = 0.2$ and $\epsilon_i = 0.5$.



3.2.1 Classification Output

effective box b_e^l表示positive区域，如图中白色部分所示。$b_i^l - b_e^l$这部分ignoring区域信息不参与分类任务，如图中灰色部分所示。ground truth map的剩余区域表示negative区域，如图中黑色部分所示。那么分类任务就是对每一个像素值做分支，考虑到正负样本的不均衡，作者采用了Focal loss损失函数。

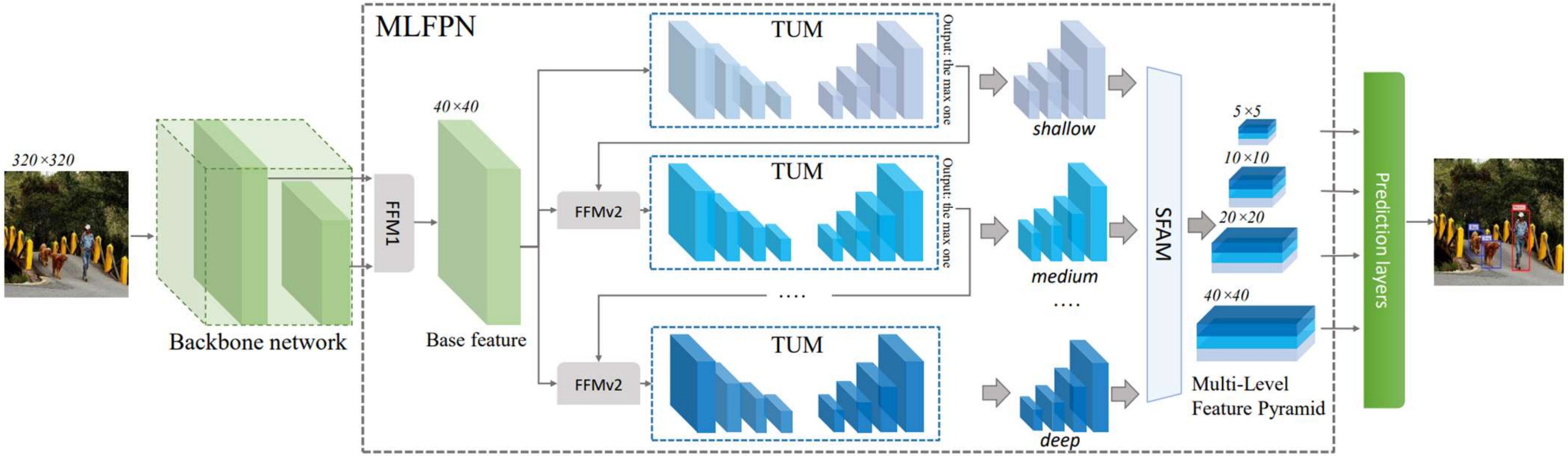
3.2.2 Box Regression Output

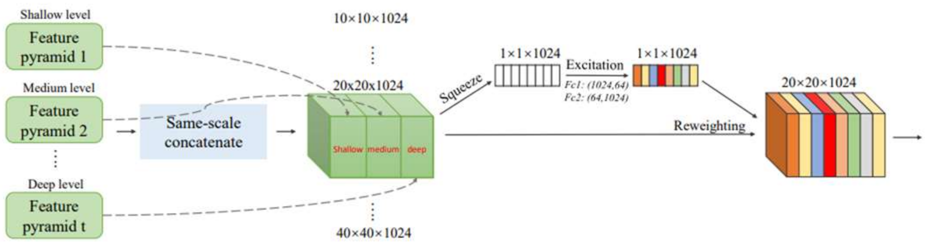
对于回归任务分支，它有4个输出offset map，从channel维度来看，每一个像素点对应了预测box的四个坐标，只不过作者取了相对偏移，即当前像素(i, j)与b_p^l的四条边的距离。而且，因为ground truth box只影响了b_e^l区域，所以这里的(i, j)是该区域内的所有像素。从上图中也可以看出，回归分支的groundtruth offset map中的有效区域尺寸和分类分支中的白色区域相同。回归分支作者采用了IoU损失函数。

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18
Soft-NMS	比普通NMS提升1点	ICCV' 17
Softer-NMS	使用KL-loss提升2点，使用Softer-NMS比soft-NMS再提升2点	CVPR' 19
CornerNet	42.1	ECCV'18
FSAF	44.6	CVPR'19

M2Det AAAI'19

多级特征金字塔网络**MLFPN,** 基于提出的MLFPN，结合SSD，提出一种新的Single-shot目标检测模型**M2Det**





M2Det

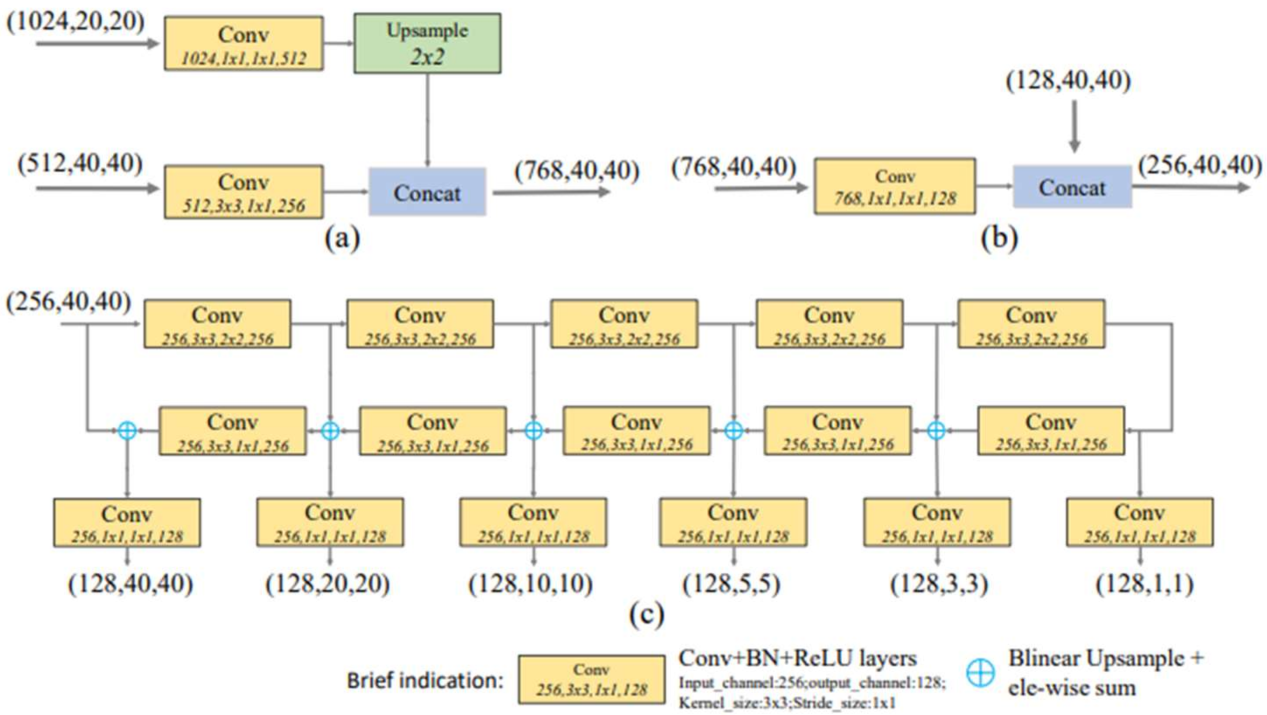


Figure 4: Structural details of some modules. (a) FFMv1, (b) FFMv2, (c) TUM. The inside numbers of each block denote: input channels, Conv kernel size, stride size, output channels.

- 在检测阶段，为6组金字塔特征每组后面添加两个卷积层，以分别实现位置回归和分类。
- 后处理阶段，使用soft-NMS来过滤无用的包围框。

Detector	COCO (mAP@IoU=0.5:0.95)	Published In
yolov3	33.0	arXiv'18
RetinaNet	39.1	ICCV'17
RefineDet	41.8	CVPR'18
Soft-NMS	比普通NMS提升1点	ICCV' 17
Softer-NMS	使用KL-loss提升2点，使用Softer-NMS比soft-NMS再提升2点	CVPR' 19
CornerNet	42.1	ECCV'18
FSAF	44.6	CVPR’19
M2Det	44.2	AAAI'19

Tips

- 使用focal loss减少类别不平衡的问题
- 使用soft-NMS代替NMS
- 多尺度金字塔

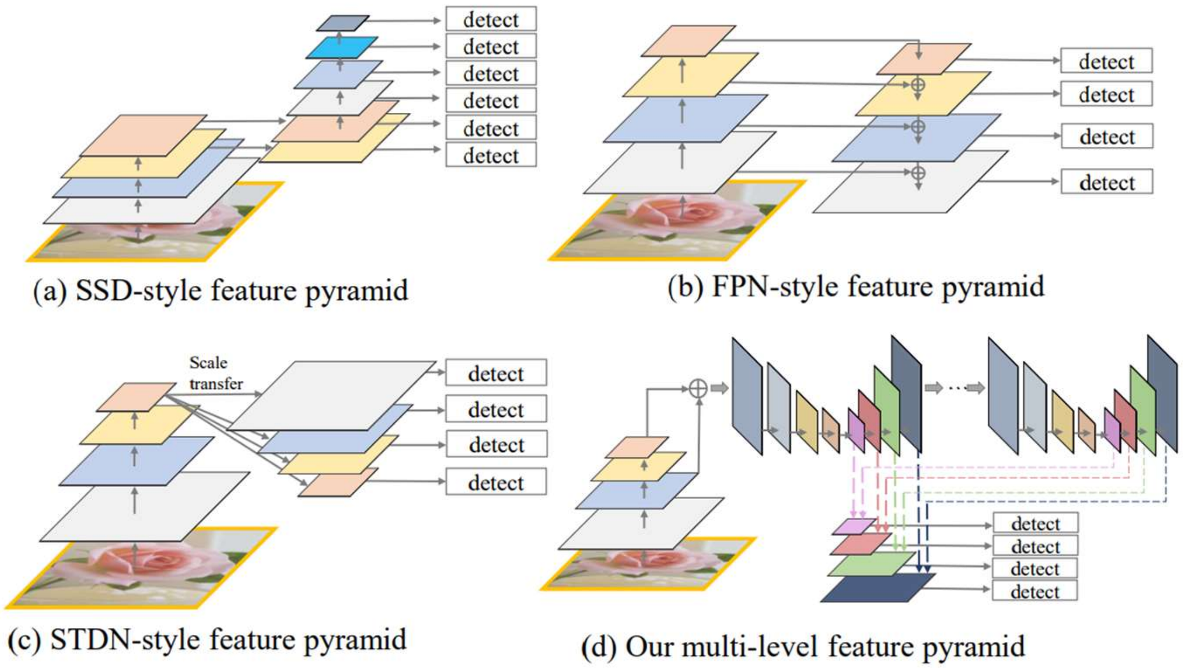


Figure 1: Illustrations of four kinds of feature pyramids.