

## 引言

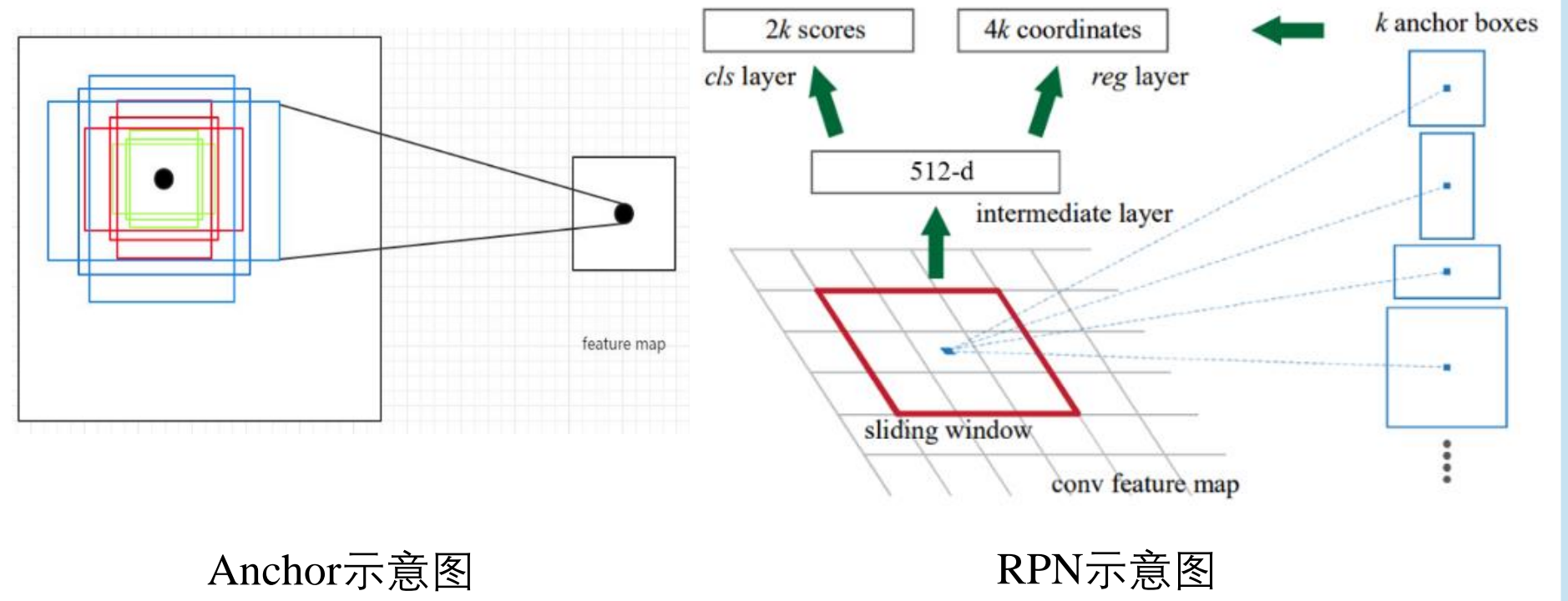
目前主流二阶段检测算法使用选择性搜索算法生成区域建议，然而这种方法单步处理时间很长，这大大增加了算法的推理时间。同时，分阶段的训练给训练过程带来了不必要的繁琐。

单阶段检测算法在速度上有一定的优势，但由于类别失衡等问题，它的检测精度通常落后于两阶段的方法。

## 贡献

- 使用新颖的RPN网络，RPN将区域提案提取集成到深度卷积神经网络中，计算proposal的成本很小，检测速度提升。
- RPN在每个位置同时预测目标边界框和类别置信度分数，实现了目标检测端到端的训练。
- 修改和扩充了VOC 2007数据集，得到07 (“pizza”) 和07 (partial COCO) 数据集，并在这些数据集上验证了我们模型在速度和精度上的高效性。

## RPN网络



### Anchor生成

- 选取 $k$ 个不同尺度、纵横比的anchors，默认选取3个尺度( $128^2$ ,  $256^2$ ,  $512^2$ )和3个纵横比(1:1, 1:2, 2:1)，共9个anchors；
- 对于一个大小为 $W \times H$ 的特征图来说，总共有 $W \times H \times k$ 个anchors；

### Proposal生成

- cls层输出的 $2k$ 个目标分数，reg层输出 $4k$ 个回归参数；
- 根据回归参数 $[t_x, t_y, t_w, t_h]$ 对所有的anchors做bounding box回归；
- 限定超出图像边界的anchors至最大到图像边界；
- 提取目标分数前 $N_{pre}$  (e.g. 6000)个位置回归后的anchors；
- 剔除尺寸非常小的anchors；
- 对剩余的anchors进行非极大抑制（NMS）；

## RoI Head

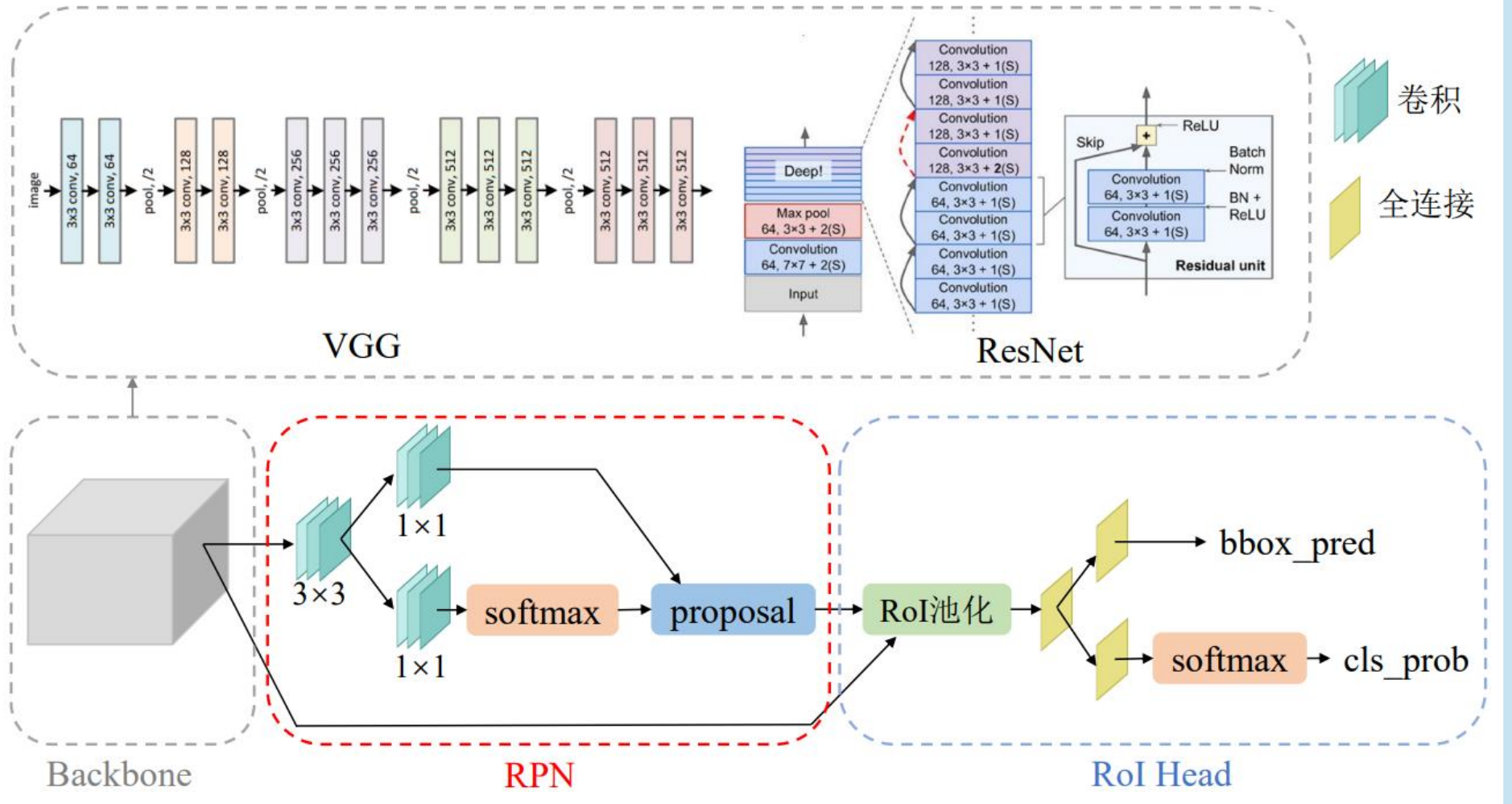
### RoI池化

- 将每个proposal对应的特征图区域水平分为 $w_p \times h_p$ 的网格；
- 实现尺寸的固定（红色区域和绿色区域所示）

### 类别概率和偏移量生成

- 两个独立的全连接层分别负责类别概率输出以及预测的位置偏移量输出；

## 网络结构



## 网络训练

对于RPN标签分配：正标签：(1) 与GT真实框具有最高IoU的anchor (2)与GT真实框的IoU高于0.7的anchors；负标签：与GT真实框的IoU均低于0.3的anchor 。损失函数定义为：

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*),$$
$$L_{reg}(t_i, t_i^*) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L1}(t_i - t_i^*)$$

对于边界框回归4个坐标的参数化，

$$\begin{aligned} t_x &= (x - x_a)/w_a, & t_y &= (y - y_a)/h_a, \\ t_w &= \log(w/w_a), & t_h &= \log(h/h_a), \\ t_x^* &= (x^* - x_a)/w_a, & t_y^* &= (y^* - y_a)/h_a, \\ t_w^* &= \log(w^*/w_a), & t_h^* &= \log(h^*/h_a), \end{aligned}$$

## 实验结果

表1. VOC 2007测试mAP			
方法	训练集	fps	mAP
SPPnet [18]	07\diff	-	63.1%
R-CNN [6]	07	-	66.0%
Fast R-CNN [7]	07	0.5	66.9%
<b>Faster R-CNN (ours)</b>	<b>07</b>	<b>5.0</b>	<b>69.9%</b>

- 我们的模型在精度和速度上实现了高效性。

表2. Faster R-CNN (VGG-16)在VOC 2007、07 (“pizza”)和07(partial COCO)测试集上的结果

数据集	mAP (%)	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	pizza	sheep	sofa	train	tv
07 (“pizza”)	70.0	68.2	78.5	68.5	56.9	51.7	80.4	79.1	84.4	44.5	76.5	57.5	78.5	79.5	76.2	74.6	72.5	64.7	62.7	74.1	69.5
07 (partial COCO)	71.6	68.9	78.4	68.6	63.9	56.1	83.1	78.7	84.7	48.5	75.8	58.4	79.6	84.9	78.0	75.4	74.0	64.3	62.0	74.4	72.5

- 07 (“pizza”) 数据集将VOC 2007中 “potted plant”类别替换成MS COCO数据集中的部分 “pizza”类别。07 (partial COCO) 数据集在07 (“pizza”) 数据集基础上对训练集中的 “bottle”， “boat”， “chair”类别做了数据扩充。
- 可以看出我们的模型在修改和扩充后的数据集上表现出依然表现出高效性和鲁棒性。

