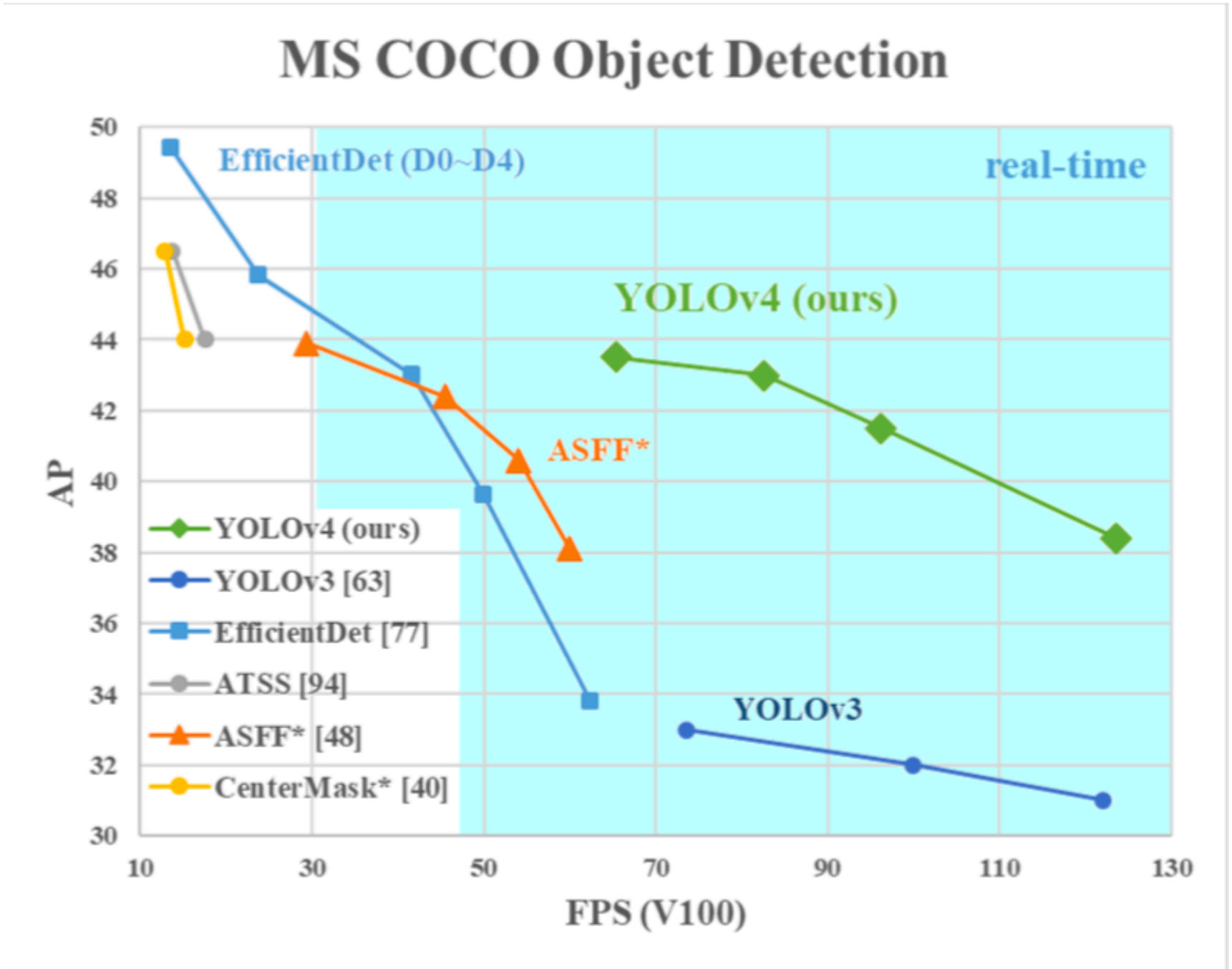
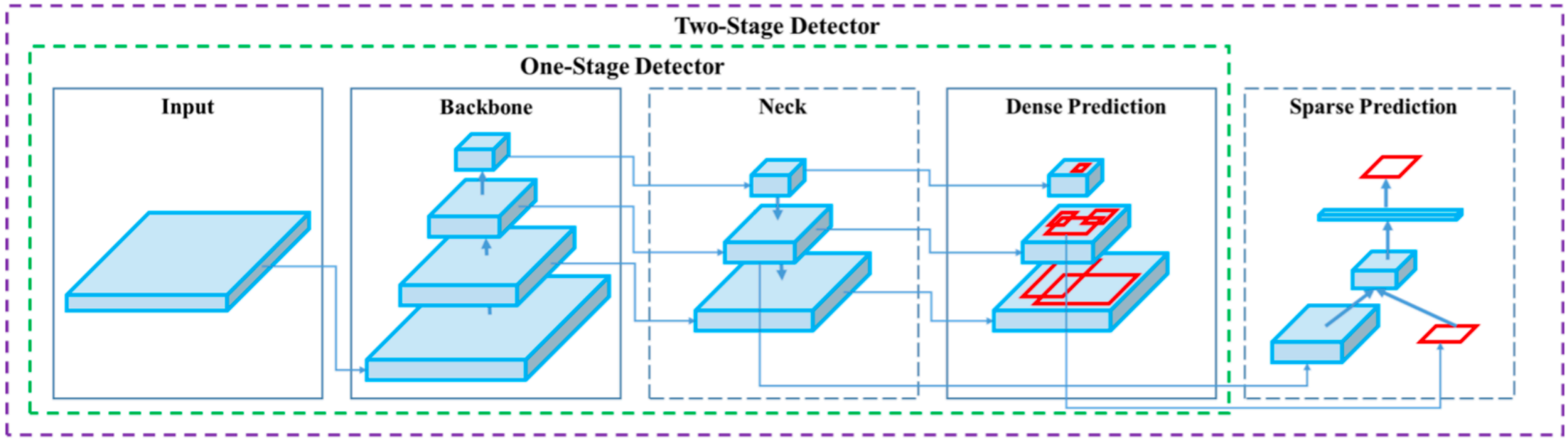

MS COCO Object Detection
AP
FPS (V100)
real-time
EfficientDet (D0~D4)
YOLOv4 (ours)
ASFF*
YOLOv3
YOLOv4 (ours)
YOLOv3 [63]
EfficientDet [77]
ATSS [94]
ASFF* [48]
CenterMask* [40]
50
48
46
44
42
40
38
36
34
32
30
10
30
50
70
90
110
130



Input: { Image, Patches, Image Pyramid, … }

Backbone: { VGG16 [68], ResNet-50 [26], ResNeXt-101 [86], Darknet53 [63], … }

Neck: { FPN [44], PANet [49], Bi-FPN [77], … }

Head:

Dense Prediction: { RPN [64], YOLO [61, 62, 63], SSD [50], RetinaNet [45], FCOS [78], … }

Sparse Prediction: { Faster R-CNN [64], R-FCN [9], … }

Figure 2: Object detector.

Table 2: Influence of BoF and Mish on the CSPResNeXt-50 classifier accuracy.

MixUp	CutMix	Mosaic	Bluring	Label Smoothing	Swish	Mish	Top-1	Top-5
							77.9%	94.0%
✓							77.2%	**94.0%**
	✓						**78.0%**	**94.3%**
		✓					**78.1%**	**94.5%**
			✓				77.5%	93.8%
				✓			**78.1%**	**94.4%**
					✓		64.5%	86.0%
						✓	**78.9%**	**94.5%**
	✓	✓		✓			**78.5%**	**94.8%**
	✓	✓		✓		✓	**79.8%**	**95.2%**

Table 3: Influence of BoF and Mish on the CSPDarknet-53 classifier accuracy.

MixUp	CutMix	Mosaic	Bluring	Label Smoothing	Swish	Mish	Top-1	Top-5
							77.2%	93.6%
	✓	✓		✓			**77.8%**	**94.4%**
	✓	✓		✓		✓	**78.7%**	**94.8%**

Table 4: Ablation Studies of Bag-of-Freebies. (CSPResNeXt50-PANet-SPP, 512x512).

S	M	IT	GA	LS	CBN	CA	DM	OA	loss	AP	AP_{50}	AP_{75}
									MSE	38.0%	60.0%	40.8%
✓									MSE	37.7%	59.9%	40.5%
	✓								MSE	**39.1%**	**61.8%**	**42.0%**
		✓							MSE	36.9%	59.7%	39.4%
			✓						MSE	**38.9%**	**61.7%**	**41.9%**
				✓					MSE	33.0%	55.4%	35.4%
					✓				MSE	**38.4%**	**60.7%**	**41.3%**
						✓			MSE	**38.7%**	**60.7%**	**41.9%**
							✓		MSE	35.3%	57.2%	38.0%
✓									GIoU	**39.4%**	59.4%	**42.5%**
✓									DIoU	**39.1%**	58.8%	**42.1%**
✓									CIoU	**39.6%**	59.2%	**42.6%**
✓	✓	✓	✓						CIoU	**41.5%**	**64.0%**	**44.8%**
	✓		✓					✓	CIoU	36.1%	56.5%	38.4%
✓	✓	✓	✓					✓	MSE	**40.3%**	**64.0%**	**43.1%**
✓	✓	✓	✓					✓	GIoU	**42.4%**	**64.4%**	**45.9%**
✓	✓	✓	✓					✓	CIoU	**42.4%**	**64.4%**	**45.9%**

Table 5: Ablation Studies of Bag-of-Specials. (Size 512x512).

Model	AP	AP_{50}	AP_{75}
CSPResNeXt50-PANet-SPP	42.4%	64.4%	45.9%
CSPResNeXt50-PANet-SPP-RFB	41.8%	62.7%	45.1%
CSPResNeXt50-PANet-SPP-SAM	**42.7%**	**64.6%**	**46.3%**
CSPResNeXt50-PANet-SPP-SAM-G	41.6%	62.7%	45.0%
CSPResNeXt50-PANet-SPP-ASFF-RFB	41.1%	62.6%	44.4%

Table 6: Using different classifier pre-trained weightings for detector training (all other training parameters are similar in all models) .

Model (with optimal setting)	Size	AP	AP_{50}	AP_{75}
CSPResNeXt50-PANet-SPP	512x512	42.4	64.4	45.9
CSPResNeXt50-PANet-SPP (BoF-backbone)	512x512	42.3	64.3	45.7
CSPResNeXt50-PANet-SPP (BoF-backbone + Mish)	512x512	42.3	64.2	45.8
CSPDarknet53-PANet-SPP (BoF-backbone)	512x512	42.4	64.5	46.0
CSPDarknet53-PANet-SPP (BoF-backbone + Mish)	512x512	43.0	64.9	46.5

Table 7: Using different mini-batch size for detector training.

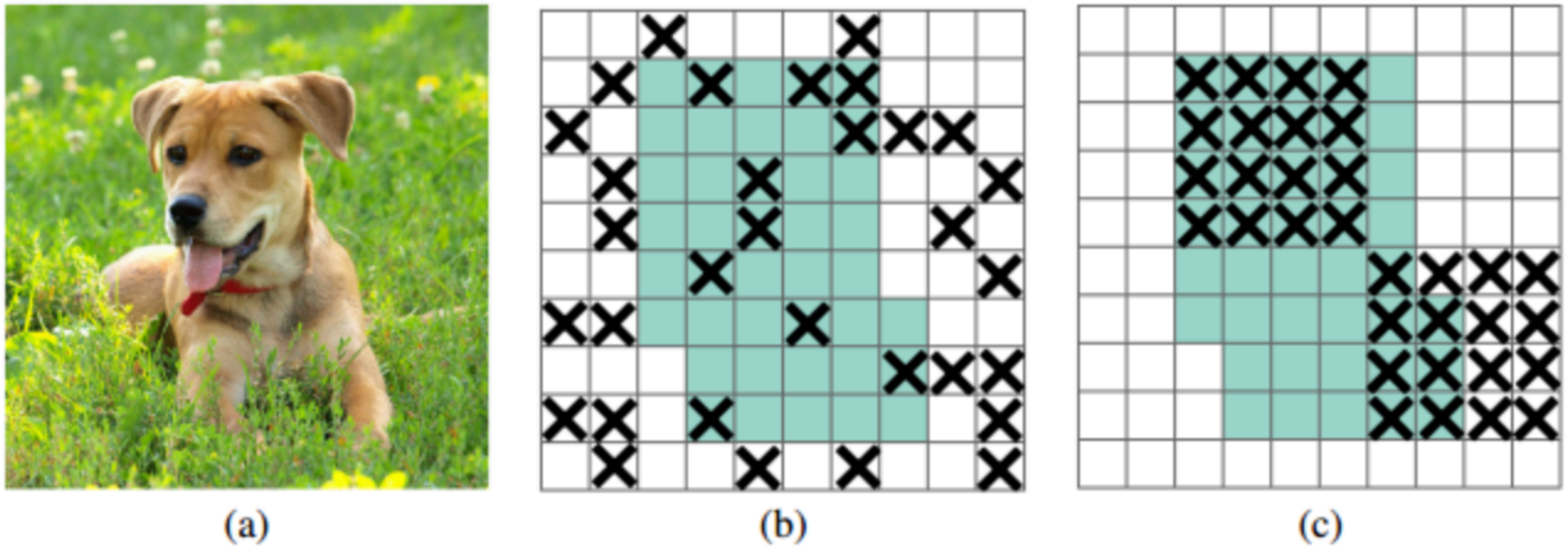
Model (without OA)	Size	AP	AP_{50}	AP_{75}
CSPResNeXt50-PANet-SPP (without BoF/BoS, mini-batch 4)	608	37.1	59.2	39.9
CSPResNeXt50-PANet-SPP (without BoF/BoS, mini-batch 8)	608	38.4	60.6	41.6
CSPDarknet53-PANet-SPP (with BoF/BoS, mini-batch 4)	512	41.6	64.1	45.0
CSPDarknet53-PANet-SPP (with BoF/BoS, mini-batch 8)	512	41.7	64.2	45.2

Mosaic data augmentation
Mosaic数据增强

	ResNet-50	Mixup [48]	Cutout [3]	CutMix
Image				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4

Figure 3: Mosaic represents a new method of data augmentation.

DropBlock regularization



(a) (b) (c)

网络还会从drouout掉的激活单元附近学习到同样的信息。
通过dropout掉一部分相邻的整片的区域（比如头和脚），网络就会去注重学习狗的别的部位的特征，来实现正确分类，从而表现出更好的泛化。

Class label smoothing

[0, 0, 1] —> [0.01, 0.01, 0.98]

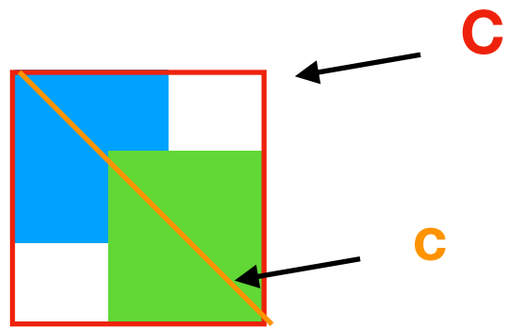
[…]·(1-a) + a/n·[1,1…]

CIoU-loss

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}, \qquad \mathcal{L}_{IoU} = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}.$$

$$\mathcal{L}_{GIoU} = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|},$$

解决IoU梯度消失问题



$$\mathcal{L}_{DIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2}.$$

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v.$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2.$$

$$\alpha = \frac{v}{(1 - IoU) + v},$$

v是用来度量长宽比的相似性

alpha是权重函数

DIoU-NMS

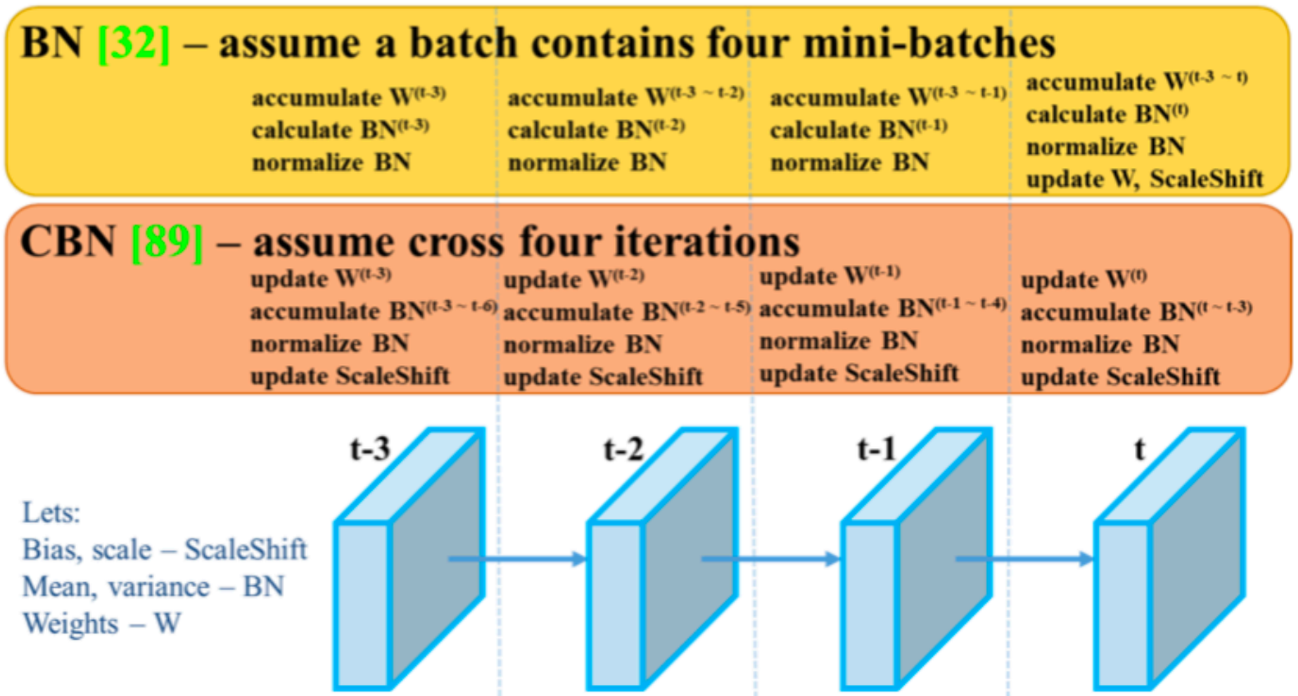
$$s_i = \begin{cases} s_i, & IoU - \mathcal{R}_{DIoU}(\mathcal{M}, B_i) < \varepsilon, \\ 0, & IoU - \mathcal{R}_{DIoU}(\mathcal{M}, B_i) \geq \varepsilon, \end{cases}$$

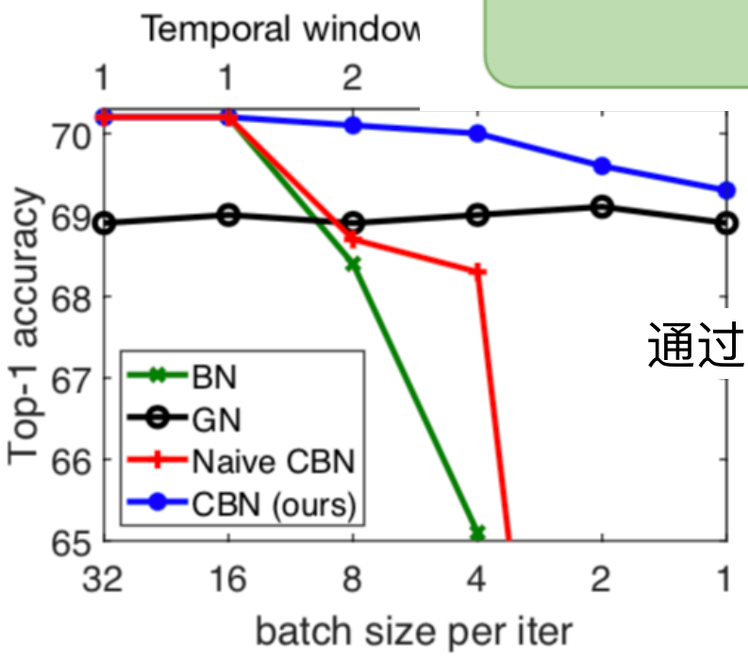
hard-nms

在原始的NMS中，IoU度量被用来抑制冗余检测盒，其中重叠区域是唯一的因素，在有遮挡的情况下常常产生错误的抑制。DIoU-NMS
不仅考虑了检测区域的重叠，而且还考虑了检测区域间的中心点距离。

0 -> 1— **(** $IoU - \mathcal{R}_{DIoU}(\mathcal{M}, B_i)$ **)**

CmBN



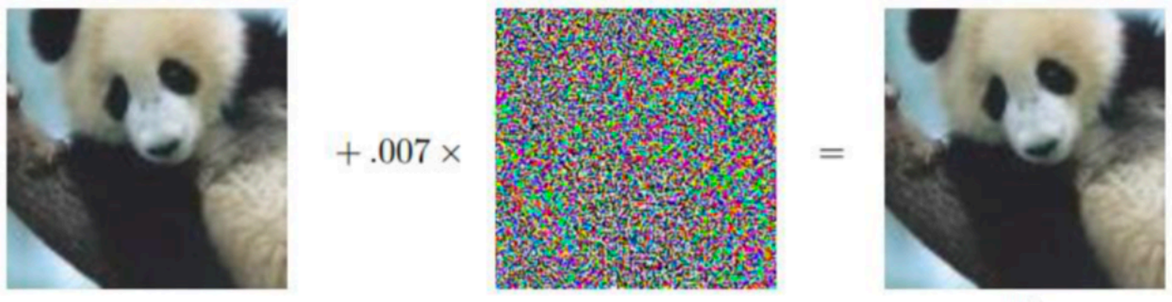


通过泰勒多项式去估计几个连续batch的统计参数

Self-Adversarial Training
自对抗训练

对抗样本的定义

以图像样本为例，在原样本上加入一些轻微的扰动，使得在人眼分辨不出差别的情况下，诱导模型进行错误分类。

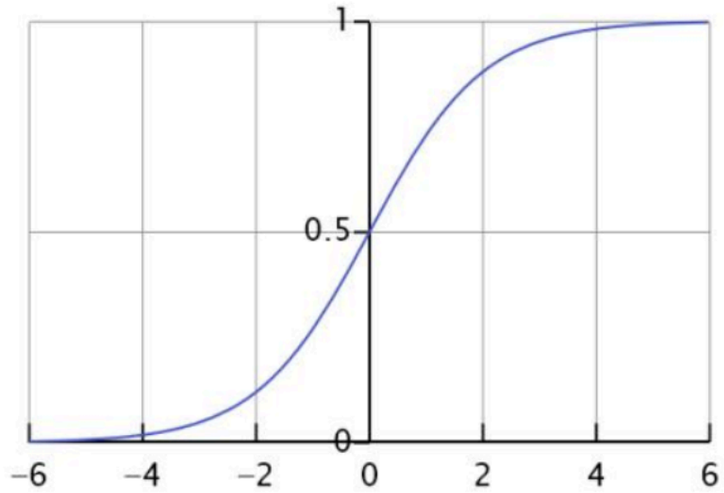


（如图，分类器错误地把加上扰动的“pandas”分类为“gibbon”）

Eliminate grid sensitivity
网格消除敏感

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h}$$

sigmoid



通过将sigmoid乘以一个超过1.0的因子来解决这个问题

Cosine annealing scheduler
模拟余弦退火

例子：

self.learn_rate_init = 1e-4
self.learn_rate_end =1e-6

```
                self.learn_rate_end + 0.5 * (self.learn_rate_init - self.learn_rate_end) *
                    (1 + tf.cos(
                        (self.global_step                    ) / (train_steps                    ) * np.pi))
```

self.global_step += 1

Mish activation

Mish=x * tanh(ln(1+e^x))

理论上对负值的轻微允许允许更好的梯度流，而不是像ReLU中那样的硬零边界。
平滑的激活函数允许更好的信息深入神经网络，从而得到更好的准确性和泛化。

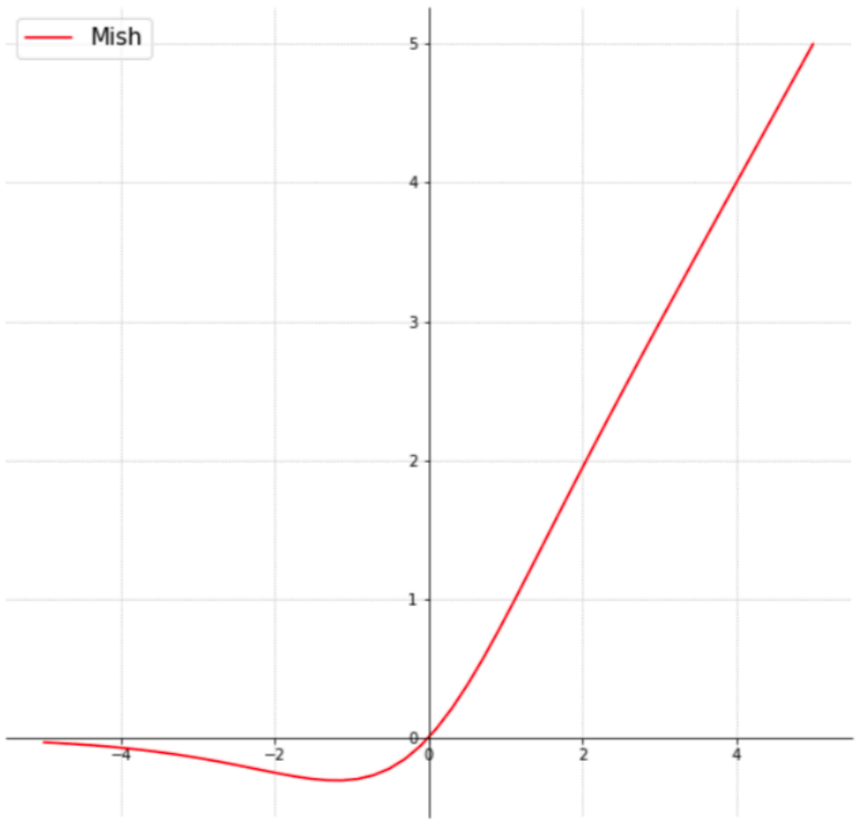
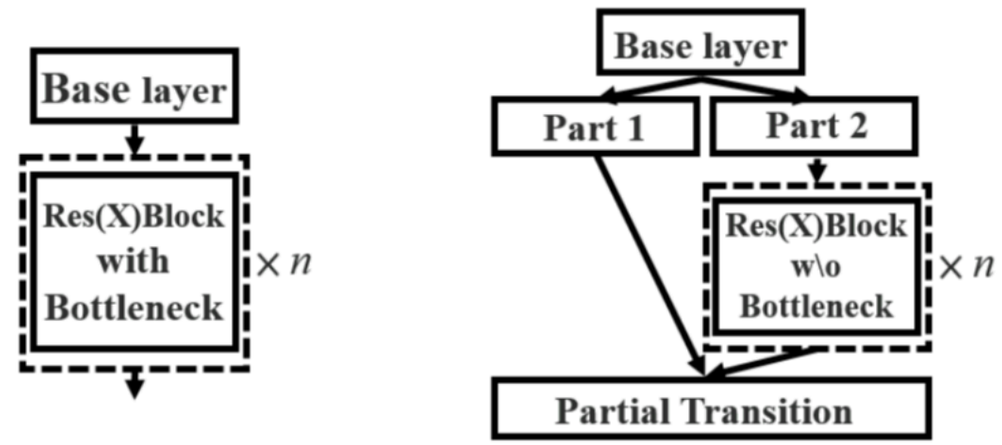


Figure 1. Mish Activation Function

CSP

CSPNet可以大大减少计算量，提高推理速度和准确性



(a) ResNe(X)t **(b) CSPResNe(X)t**

Figure 5: Applying CSPNet to ResNe(X)t.

SPP

不管输入尺寸是怎样，SPP层 可以产生固定大小的输出 ， 用于多尺度训练

```
[pool3x3]          [pool2x2]          [pool1x1]
type=pool          type=pool          type=pool
pool=max           pool=max           pool=max
inputs=conv5       inputs=conv5       inputs=conv5
sizeX=5            sizeX=7            sizeX=13
stride=4           stride=6           stride=13

[fc6]
type=fc
outputs=4096
inputs=pool3x3,pool2x2,pool1x1
```

SAM-block

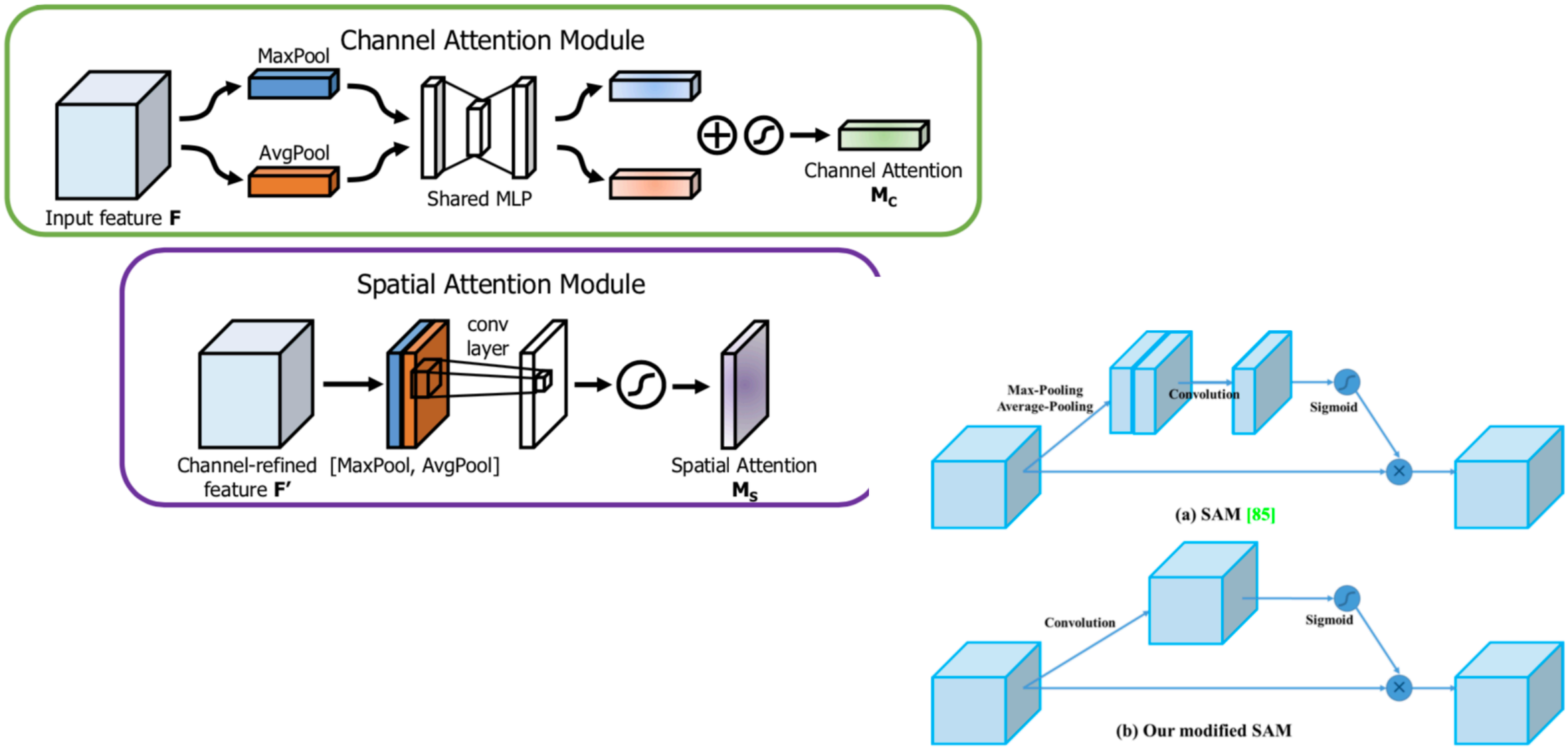


Figure 5: Modified SAM.

PAN

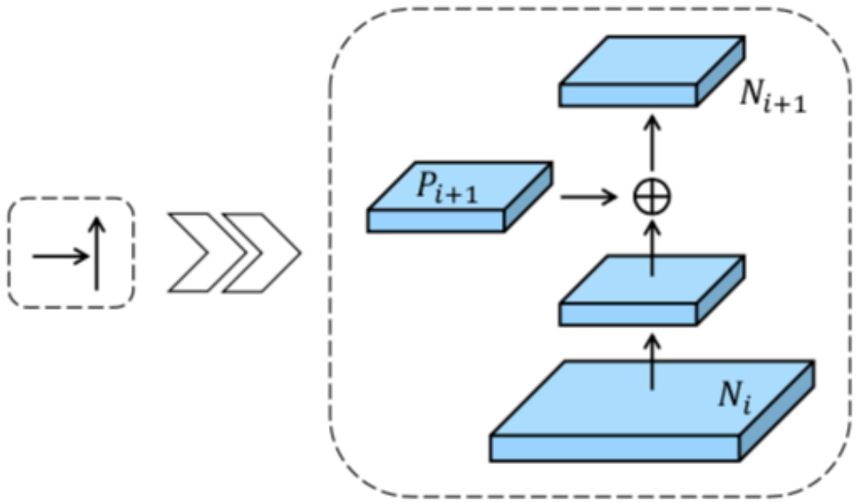
PAN基于FPN和Mask RCNN模型之上提出了三点创新：

1、PANet改进了主干网络结构，加强了特征金字塔的结构，缩短了高低层特征融合的路径

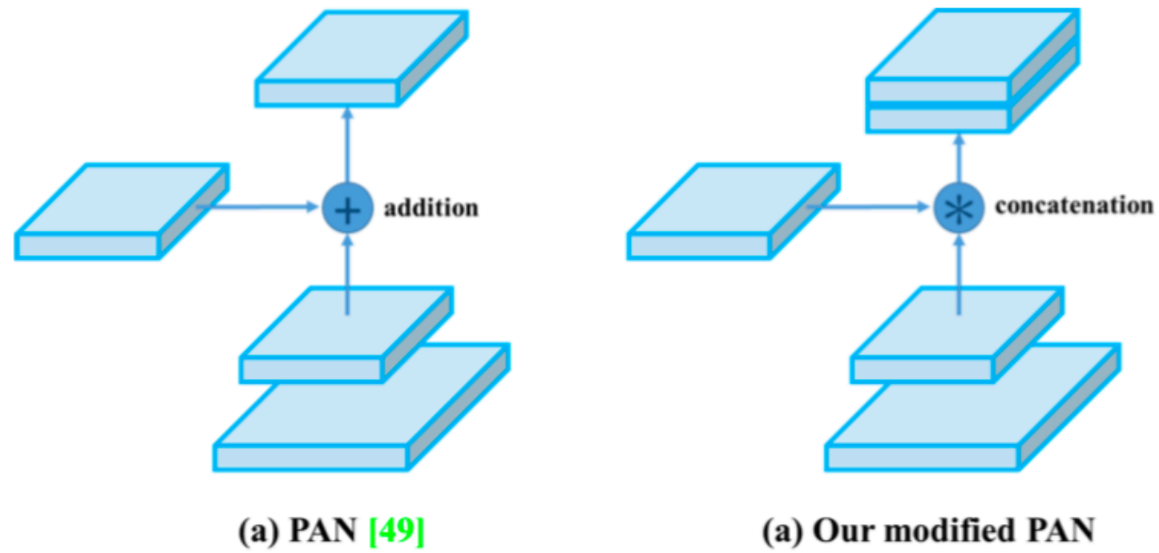
2、提出了更灵活的RoI池化。之前FPN的RoI池化只从高层特征取值，现在则在各个尺度上的特征里操作；

3、预测mask的时候使用一个额外的fc支路来辅助全卷积分割支路的结果。

PANet在COCO 17实例分割竞赛中取得了第一名的成绩，在检测任务中取得了第二的成绩。



$\{N_2, N_3, N_4, N_5\}$.



(a) PAN [49] **(a) Our modified PAN**

Figure 6: Modified PAN.