YOLO V4

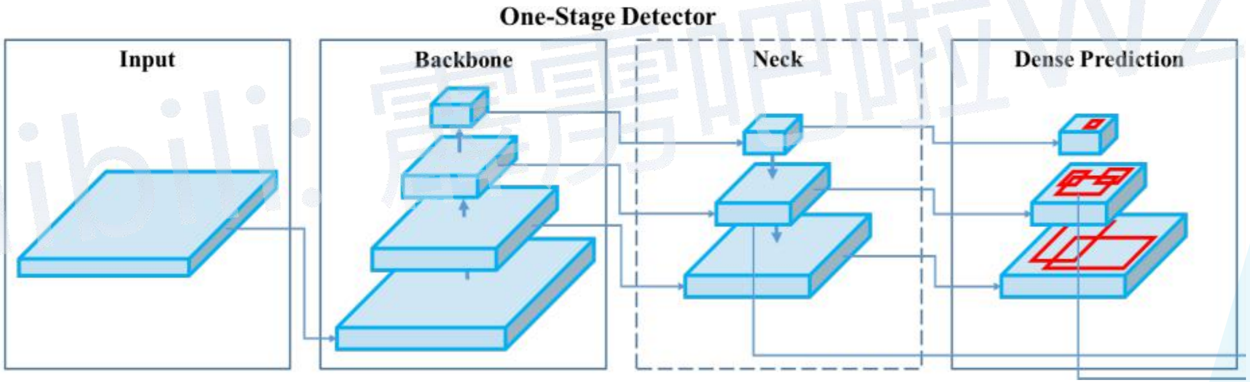
YOLOv4: Optimal Speed and Accuracy of Object Detection

Alexey Bochkovskiy*
alexeyab84@gmail.com

Chien-Yao Wang*
Institute of Information Science
Academia Sinica, Taiwan
kinyiu@iis.sinica.edu.tw

Hong-Yuan Mark Liao
Institute of Information Science
Academia Sinica, Taiwan
liao@iis.sinica.edu.tw

2020 CVPR



论文地址：https://arxiv.org/abs/2004.10934
推荐博文：https://blog.csdn.net/qq_37541097/article/details/123229946

YOLO V4

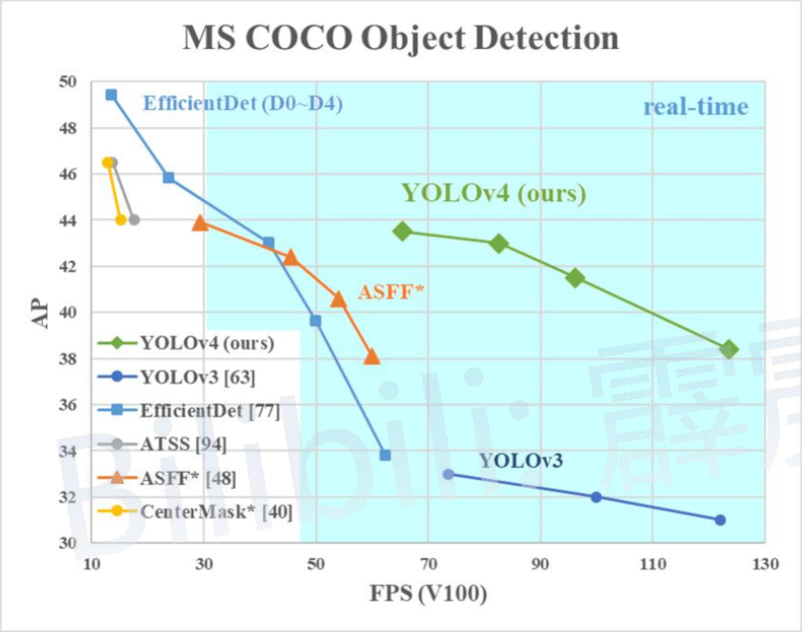


Figure 1: Comparison of the proposed YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10% and 12%, respectively.

网络结构：

- **Backbone**: CSPDarknet53
- **Neck**: SPP，PAN
- **Head**: YOLOv3

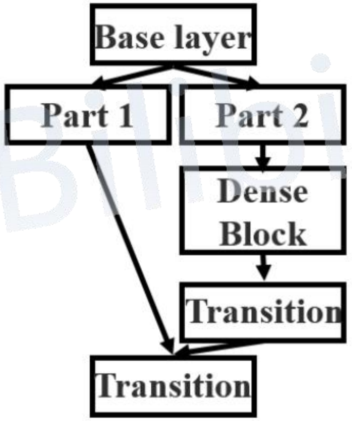
优化策略：

- **Eliminate grid sensitivity**
- **Mosaic data augmentation**
- **IoU threshold(match positive samples)**
- **Optimizered Anchors**
- **CIOU**

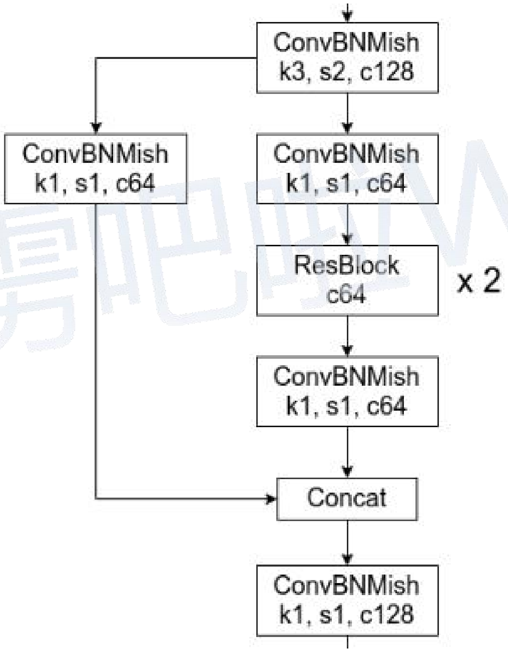
YOLO V4

网络结构 - CSPDarknet53

- Strengthening learning ability of a CNN
- Removing computational bottlenecks
- Reducing memory costs

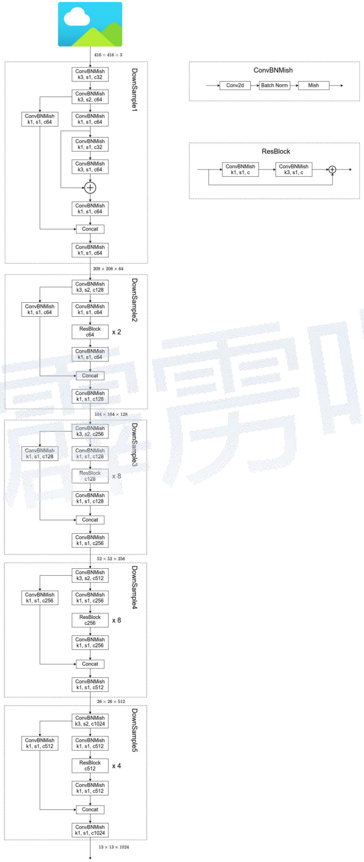


(b) CSPDenseNet



YOLO V4

网络结构 - CSPDarknet53

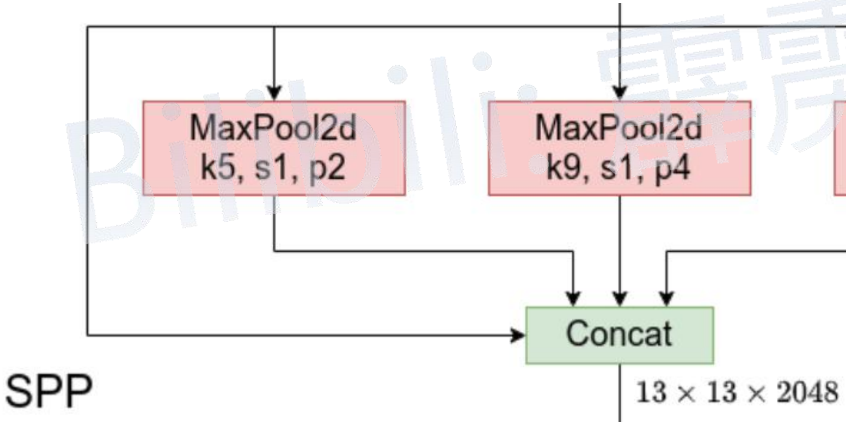


YOLO V4

网络结构 - SPP

Spatial Pyramid Pooling

解决多尺度问题



YOLO V4

网络结构 - PAN

Path Aggregation Network

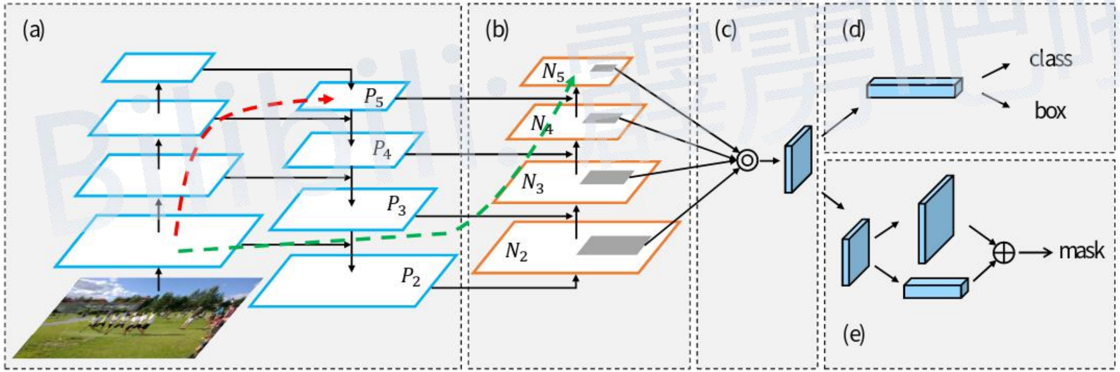
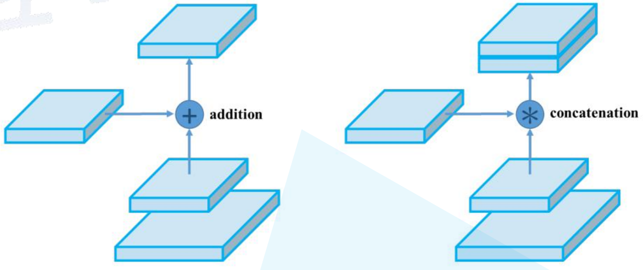


Figure 1. Illustration of our framework. (a) FPN backbone. (b) Bottom-up path augmentation. (c) Adaptive feature pooling. (d) I
branch. (e) Fully-connected fusion. Note that we omit channel dimension of feature maps in (a) and (b) for brevity.

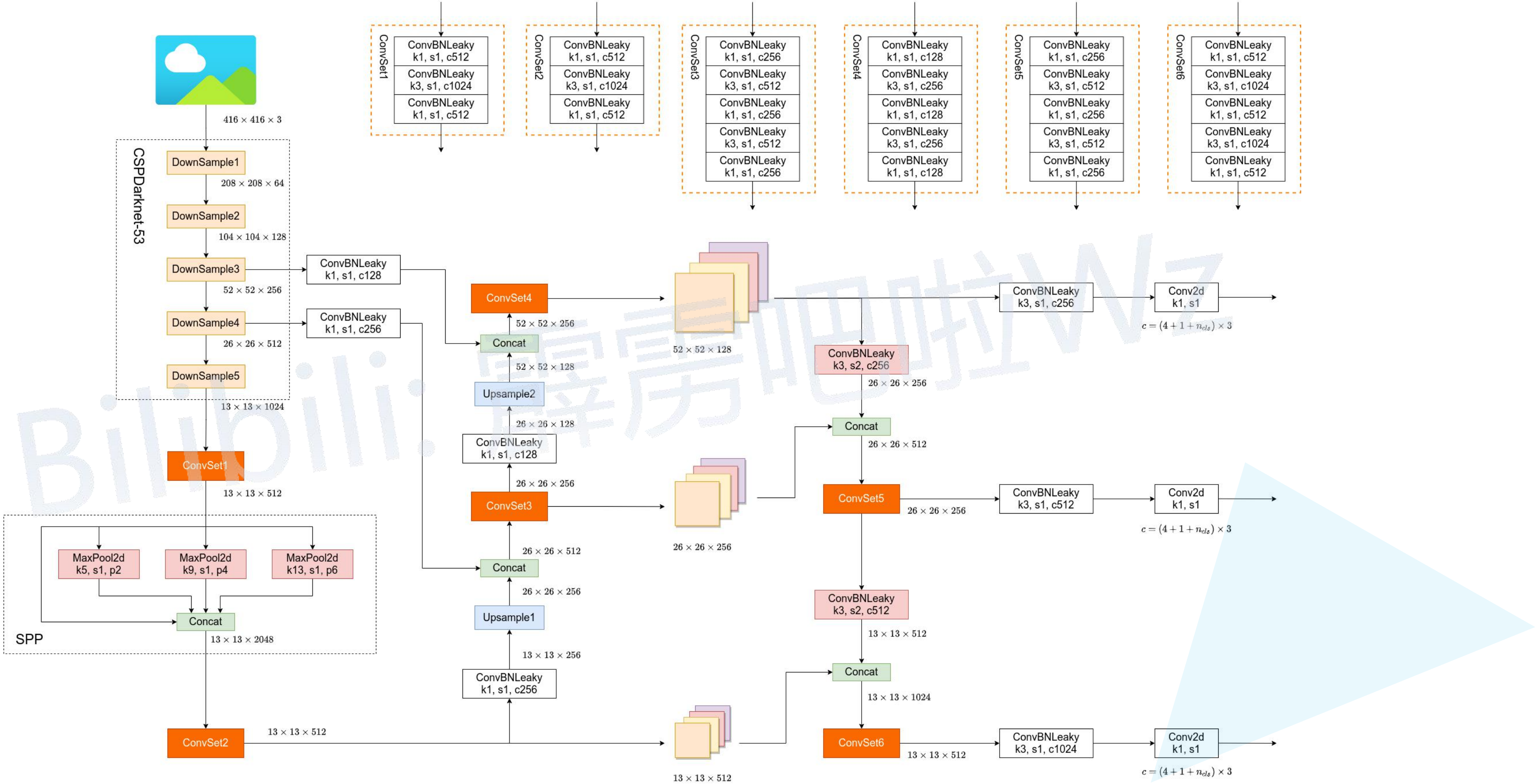


(a) PAN [49]

(a) Our modified PAN

Figure 6: Modified PAN.

YOLO V4



ConvSet1
ConvBNLeaky k1, s1, c512
ConvBNLeaky k3, s1, c1024
ConvBNLeaky k1, s1, c512
ConvSet2
ConvBNLeaky k1, s1, c512
ConvBNLeaky k3, s1, c1024
ConvBNLeaky k1, s1, c512
ConvSet3
ConvBNLeaky k1, s1, c256
ConvBNLeaky k3, s1, c512
ConvBNLeaky k1, s1, c256
ConvBNLeaky k3, s1, c512
ConvBNLeaky k1, s1, c256
ConvSet4
ConvBNLeaky k1, s1, c128
ConvBNLeaky k3, s1, c256
ConvBNLeaky k1, s1, c128
ConvBNLeaky k3, s1, c256
ConvBNLeaky k1, s1, c128
ConvSet5
ConvBNLeaky k1, s1, c256
ConvBNLeaky k3, s1, c512
ConvBNLeaky k1, s1, c256
ConvBNLeaky k3, s1, c512
ConvBNLeaky k1, s1, c256
ConvSet6
ConvBNLeaky k1, s1, c512
ConvBNLeaky k3, s1, c1024
ConvBNLeaky k1, s1, c512
ConvBNLeaky k3, s1, c1024
ConvBNLeaky k1, s1, c512
416 × 416 × 3
CSPDarknet-53
DownSample1
208 × 208 × 64
DownSample2
104 × 104 × 128
DownSample3
52 × 52 × 256
DownSample4
26 × 26 × 512
DownSample5
13 × 13 × 1024
ConvBNLeaky k1, s1, c128
ConvBNLeaky k1, s1, c256
ConvSet1
13 × 13 × 512
SPP
MaxPool2d k5, s1, p2
MaxPool2d k9, s1, p4
MaxPool2d k13, s1, p6
Concat
13 × 13 × 2048
ConvSet2
13 × 13 × 512
ConvBNLeaky k1, s1, c256
13 × 13 × 256
Upsample1
26 × 26 × 256
Concat
26 × 26 × 512
ConvSet3
26 × 26 × 256
ConvBNLeaky k1, s1, c128
26 × 26 × 128
Upsample2
52 × 52 × 128
Concat
52 × 52 × 256
ConvSet4
52 × 52 × 128
26 × 26 × 256
13 × 13 × 512
ConvBNLeaky k3, s2, c256
26 × 26 × 256
Concat
26 × 26 × 512
ConvSet5
26 × 26 × 256
ConvBNLeaky k3, s2, c512
13 × 13 × 512
Concat
13 × 13 × 1024
ConvSet6
13 × 13 × 512
ConvBNLeaky k1, s1, c256
ConvBNLeaky k1, s1, c512
ConvBNLeaky k3, s1, c1024
Conv2d k1, s1
$c = (4 + 1 + n_{cls}) \times 3$
Bilibili: 霹雳吧啦Wz

YOLO V4

优化策略 - Eliminate grid sensitivity

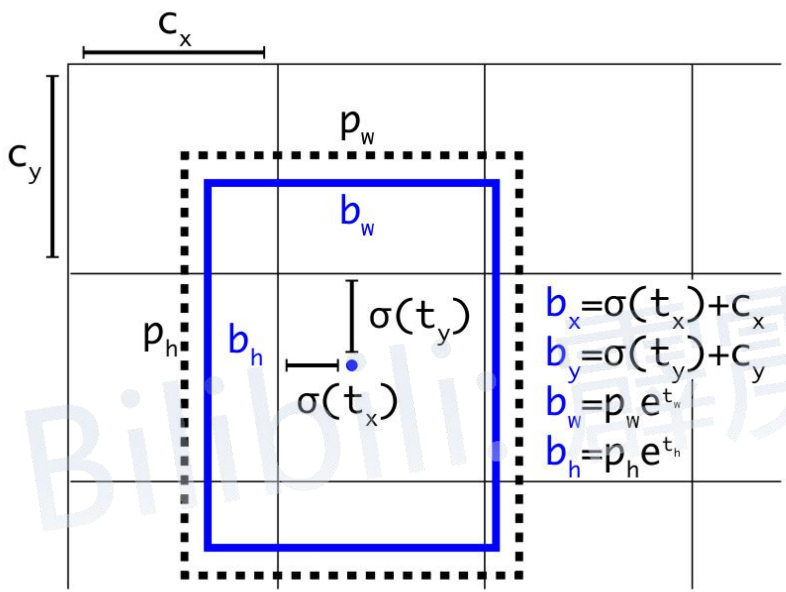
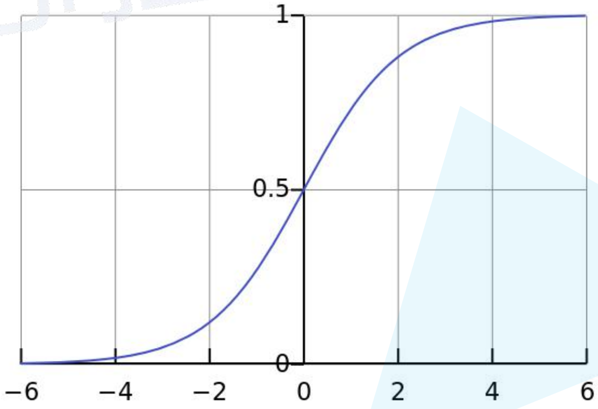


Figure 3: Bounding boxes with dimension priors and location prediction. We predict the width and height of the box as offsets from cluster centroids. We predict the center coordinates of the box relative to the location of filter application using a sigmoid function.

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



YOLO V4

优化策略 - Eliminate grid sensitivity

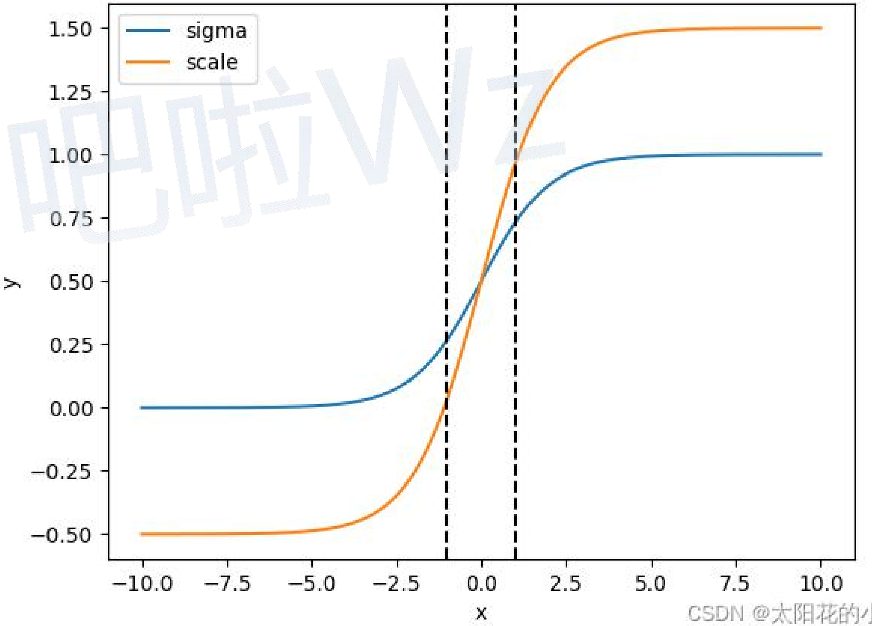
$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$

$$\Downarrow$$

$$b_x = \left(\sigma(t_x) \cdot scale_{xy} - \frac{scale_{xy} - 1}{2}\right) + c_x$$
$$b_y = \left(\sigma(t_y) \cdot scale_{xy} - \frac{scale_{xy} - 1}{2}\right) + c_y$$

$$\Downarrow$$

$$b_x = (2 \cdot \sigma(t_x) - 0.5) + c_x$$
$$b_y = (2 \cdot \sigma(t_y) - 0.5) + c_y$$

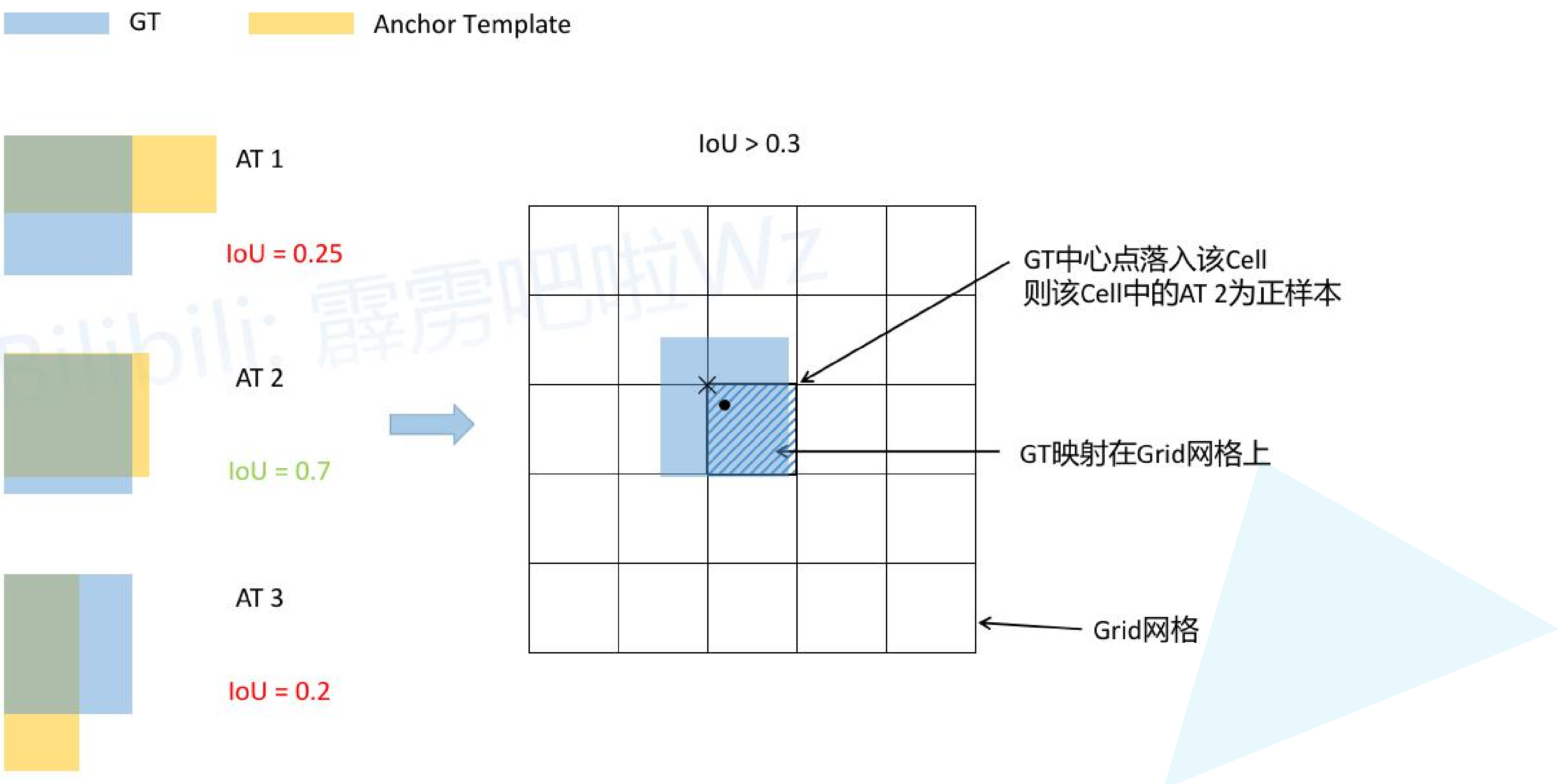


YOLO V4

优化策略 - Mosaic data augmentation

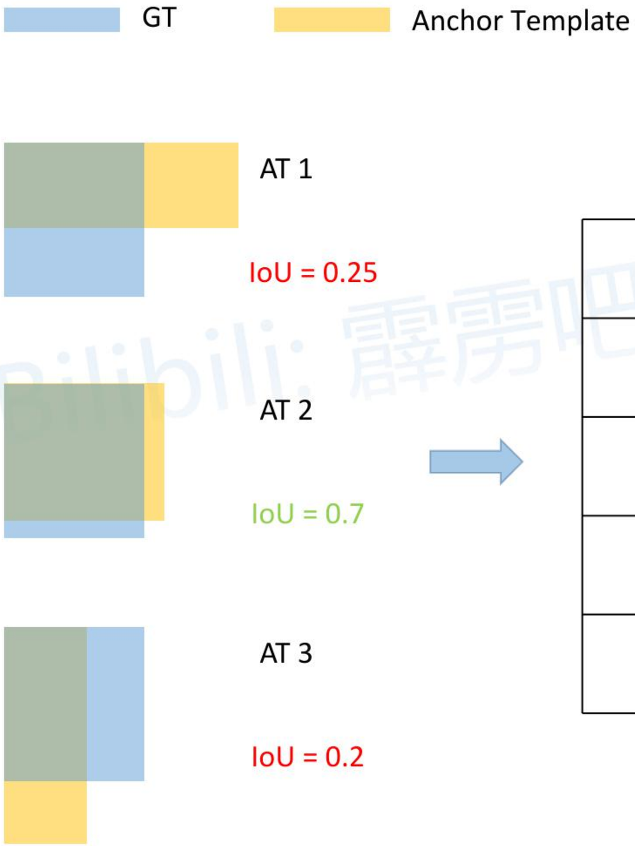
YOLO V4

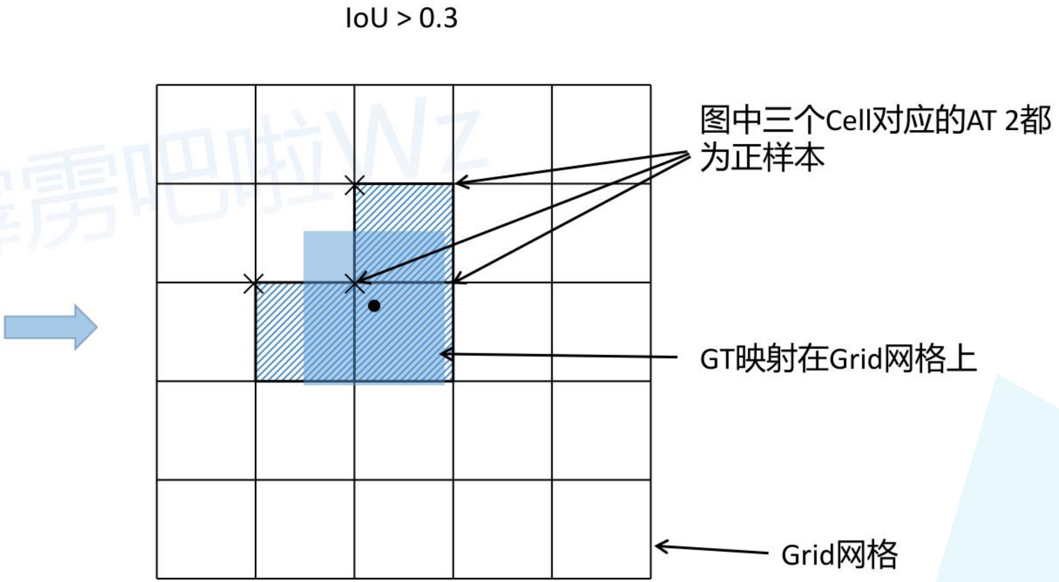
优化策略 - IoU threshold(match positive samples)



YOLO V4

优化策略 - IoU threshold(match posotive samples)



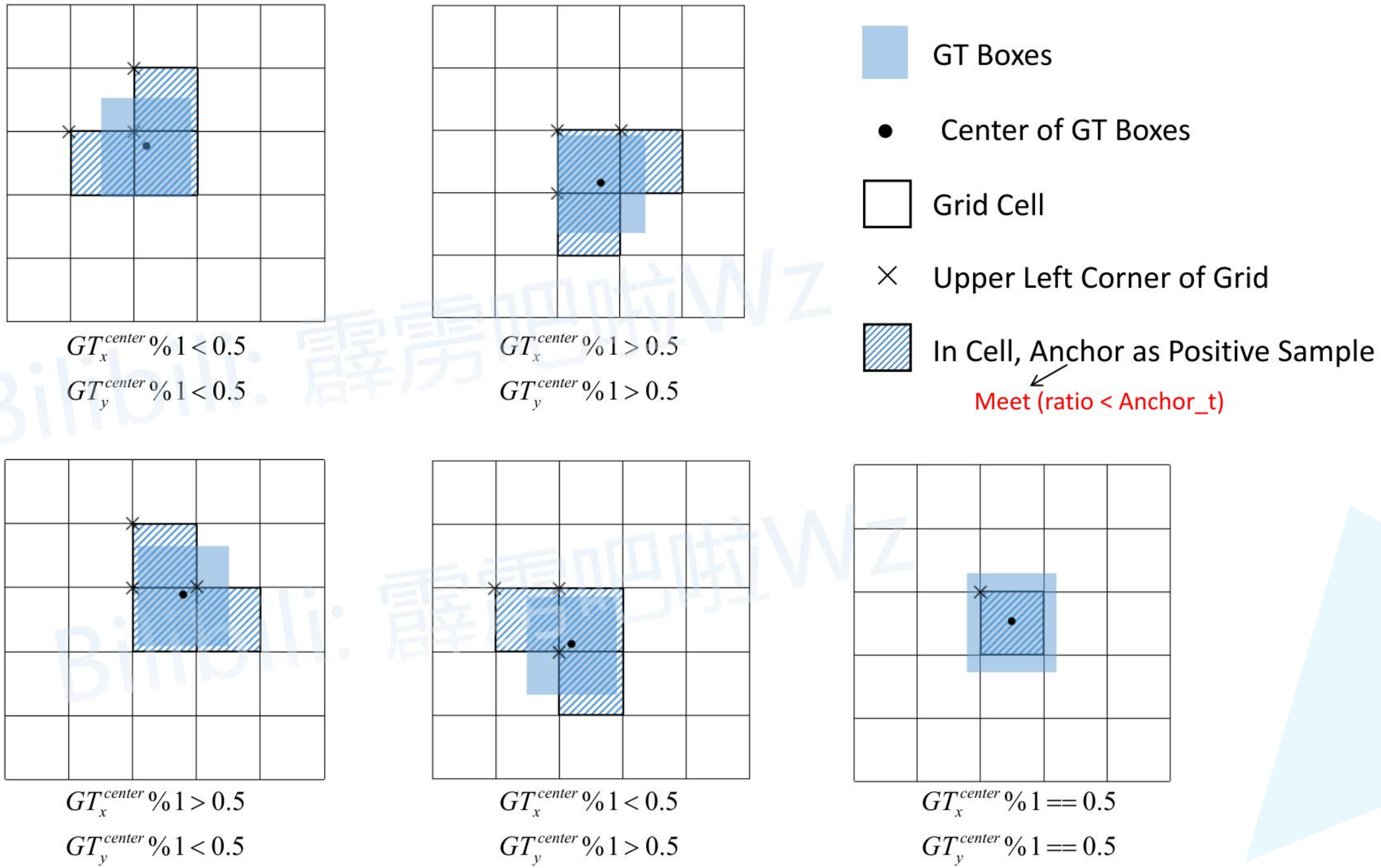


$$b_x = (2 \cdot \sigma(t_x) - 0.5) + c_x$$

$$b_y = (2 \cdot \sigma(t_y) - 0.5) + c_y$$

YOLO V4

优化策略 - IoU threshold(match posotive samples)



YOLO V4

优化策略 - Optimizered Anchors

YOLO V3

目标类型	Anchors模板
小尺度	(10x13), (16x30), (33x23)
中尺度	(30x61), (62x45), (59x119)
大尺度	(116x90), (156x198), (373x326)

YOLO V4 (512x512)

目标类型	Anchors模板
小尺度	(12x16), (19x36), (40x28)
中尺度	(36x75), (76x55), (72x146)
大尺度	(142x110), (192x243), (459x401)

YOLO V4

优化策略 - CIOU

YOLOv3 SPP

CIoU Loss

Complete-IoU

$$CIoU = IoU - \left(\frac{\rho^2(b, b^{gt})}{c^2} + \alpha\upsilon\right)$$

$$\upsilon = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2$$

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

一个优秀的回归定位损失应该考虑到3种几何参数：

重叠面积 中心点距离 长宽比

$$L_{CIoU} = 1 - CIoU$$

Loss / Evaluation	AP		AP75	
	IoU	GIoU	IoU	GIoU
$\mathcal{L}_{IoU}$	46.57	45.82	49.82	48.76
$\mathcal{L}_{GIoU}$	47.73	46.88	52.20	51.05
Relative improv. %	2.49%	2.31%	4.78%	4.70%
$\mathcal{L}_{DIoU}$	48.10	47.38	52.82	51.88
Relative improv. %	3.29%	3.40%	6.02%	6.40%
$\mathcal{L}_{CIoU}$	**49.21**	**48.42**	**54.28**	**52.87**
Relative improv. %	**5.67%**	**5.67%**	**8.95%**	**8.43%**
$\mathcal{L}_{CIoU}(D)$	**49.32**	**48.54**	**54.74**	**53.30**
Relative improv. %	**5.91%**	**5.94%**	**9.88%**	**9.31%**