۱)

مقایسه منصفانه بین روش‌های متفاوت object detection بسیار سخت است. هیچ پاسخ قطعی در مورد اینکه کدام مدل بهترین است وجود ندارد. برای کاربردهای واقعی، انتخاب‌هایی را برای متعادل کردن دقت و سرعت انجام می‌دهیم. علاوه بر انواع روش‌های object detection، باید از انتخاب‌های دیگری که بر عملکرد تأثیر می گذارد آگاه باشیم:

استخراج کننده‌های ویژگی (VGG16، ResNet، Inception، MobileNet).

strideهای خروجی برای استخراج کننده.

رزولوشن تصویر ورودی

استراتژی تطبیق و آستانه IoU

آستانه حذف مقادیر بیشنه (Non-max suppression) برای IoU.

نسبت استخراج نمونه سخت (نسبت anchor مثبت در برابر منفی).

تعداد پروپوزال‌ها یا پیش‌بینی‌ها.

رمزگذاری Boundary box

داده‌افزایی (Data augmentation)

مجموعه داده‌های آموزشی

استفاده از تصاویر multi-scale در آموزش یا آزمون

و...

اما با توجه به نتایج مقالات ما یک مقایسه کلی انجام می‌دهیم.

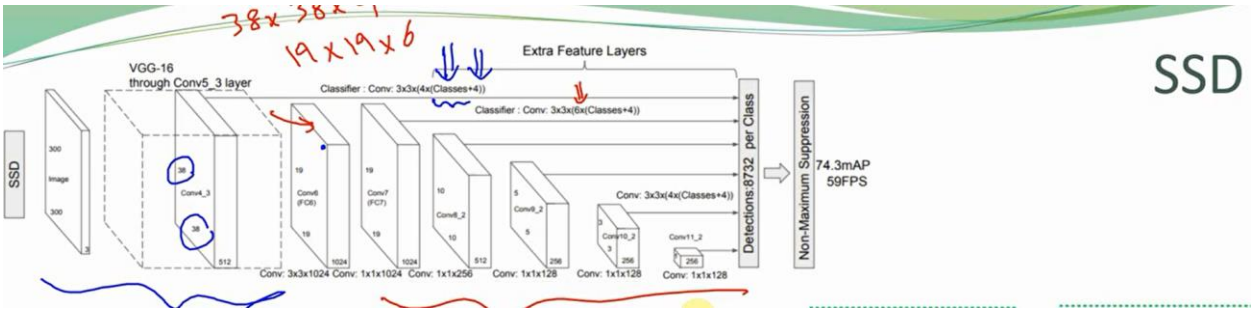
R-CNN: سرعت پایین هست چون برای هر تصویر حدود ۲۰۰۰ پروپوزال استخراج می‌کند و شبکه کانولوشنی را ۲۰۰۰ بار صدا می‌زند، دقت بالاست، چند مقیاسه (multi-scale) نیز هست.

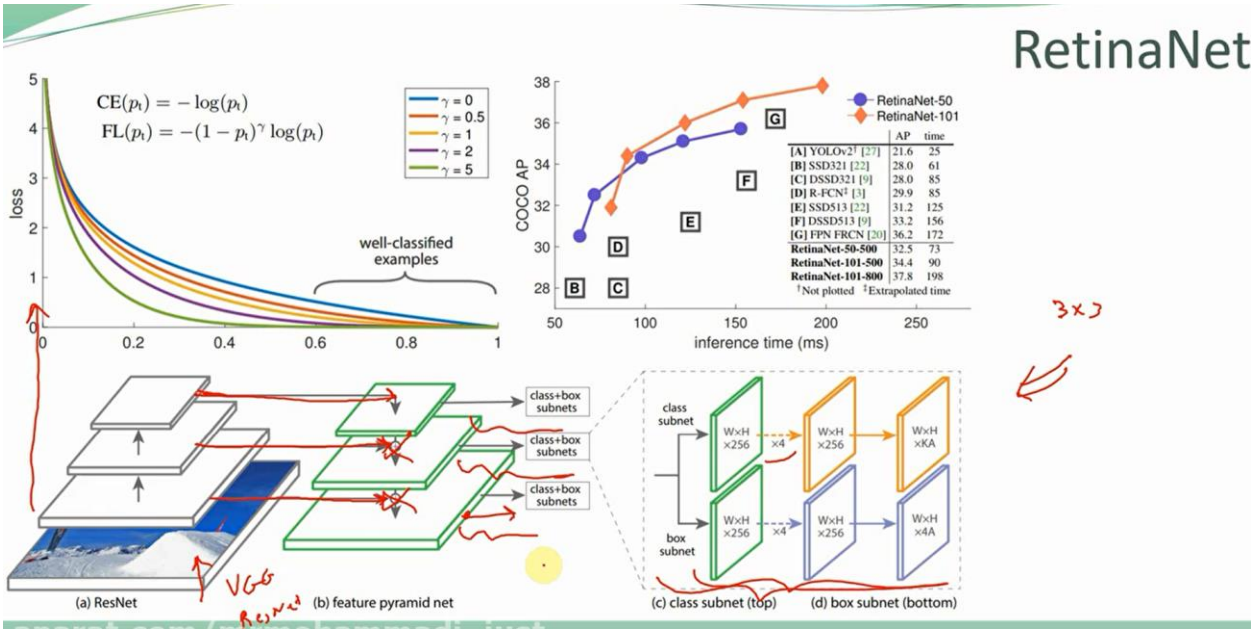
Fast R-CNN سرعت بالاتری نسبت به R-CNN دارد چون شبکه کانولوشنی را ابتدا و فقط یک بار روی تصویر اجرا می‌کند. دقت آن نیز تفاوت چندانی با R-CNN ندارد، اما می‌تواند بهتر هم باشد. چند مقیاسه (multi-scale) نیز هست.

Faster R-CNN نسبت به R-CNN و Fast R-CNN هم سرعت بالاتر و هم دقت بالاتری دارد. سرعت بالاتری دارد چون با استفاده از همان ویژگی‌ها استخراج شده توسط شبکه کانولوشنی پروپوزال استخراج می‌کند نه به صورت دستی. دقت بهتری دارد چون پروپوزال‌های استخراج شده توسط ایجاد کننده پروپوزال عمومی ساخته نشدند و خاص مسئله ایجاد شده‌اند. چند مقیاسه (multi-scale) نیز هست.

دلیل چند مقیاسه بودن سه روش قبلی این است که تصمیم‌گیری در رابطه با اندازه پروپوزال به عهده ابزار استخراج پروپوزال است.

روش‌های YOLOو SSD و RetinaNet از روش‌های قبلی سریعتر هستند که سرعت SSD و YOLOv3 از بقیه بیشتر است (چون یک مرحله‌ای هستند و مرحله استخراج پروپوزال را ندارند). اگر سرعت مطرح نباشد دقت Faster R-CNN از بقیه بیشتر است اما اگر سرعت را در نظر بگیریم. در کل دقت روش‌هایی که بخش استخراج پروپوزال دارند از این روش‌ها بیشتر است. RetinaNet سرعت کمتری نسبت به YOLOو SSD دارد (همانطور که از معماری آن پیداست نسبت به ssd چند لایه تخصصی‌تر دارد که باعث دقت بالاتر و سرعت پایینتر می‌شود.) اما دقت آن از تمام روش‌های آن به جز YOLOv4 بهتر است. SSD و RetinaNet دارای خاصیت چند مقیاسه (multi-scale) هستند چون receptive field با اندازه‌های مختلفی را در نظر میگیرند. YOLOv1و YOLOv2 خاصیت چند مقیاسه (multi-scale) را نداشتند اما این ویژگی یکی از اهداف در پیاده‌سازی YOLOv3 بود.





منابع

https://jonathan-hui.medium.com/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359

https://towardsdatascience.com/review-retinanet-focal-loss-object-detection-38fba6afabe4

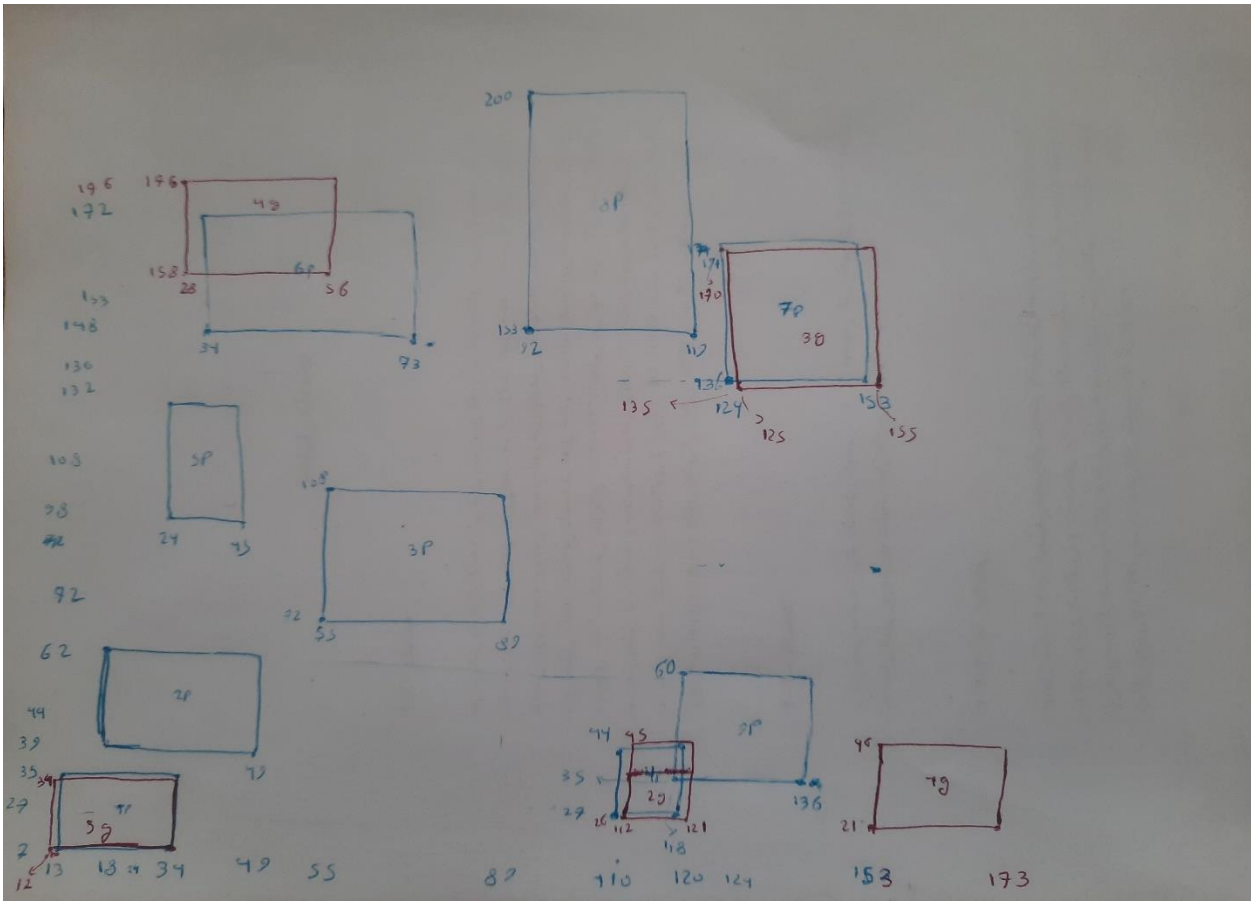
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00434-w

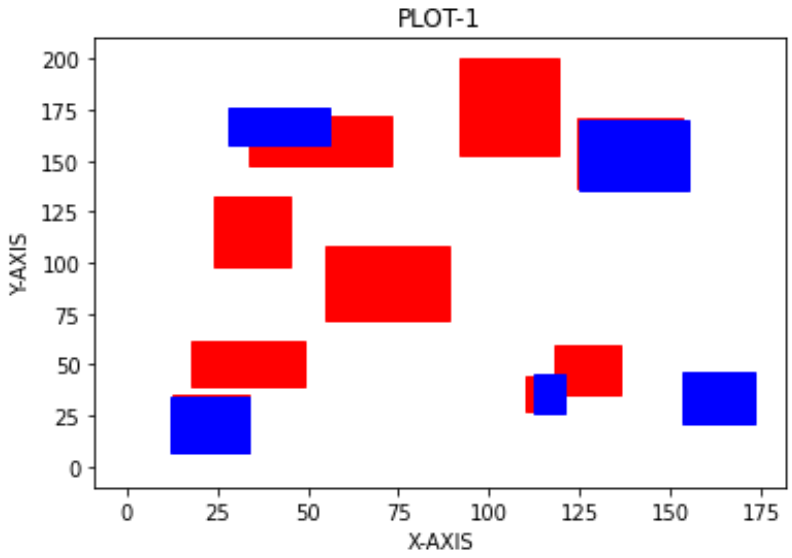
۲)

تعداد کانتورهایی که به طور کامل توسط یک box محصور شده‌اند، نشان دهنده احتمال وجود یک شی در box است. اگر تمام پیکسل‌های لبه متعلق به کانتور در داخل box قرار داشته باشند، کانتور به طور کامل توسط box محصور می‌شود. لبه‌ها معمولاً با مرزهای شی مطابقت دارند و به همین دلیل boxهایی که مجموعه‌ای از لبه‌ها را کاملا در خود قرار می‌دهند احتمالاً حاوی یک شی هستند. با این وجود، برخی از لبه‌هایی که در bounding box یک شی قرار دارند، ممکن است بخشی از شی موجود نباشند. به طور خاص، پیکسل‌های لبه‌ای که متعلق به کانتورهایی هستند که از مرزهای box خارج می‌شوند ، احتمالاً با اشیاء یا ساختارهایی مطابقت دارند که خارج از جعبه قرار دارند. امتیاز یک box متناسب است با جمع تعداد

پیکسل‌های لبه‌های منهای تعداد پیکسل‌های لبه‌هایی که بخشی از یک کانتور هستند که از مرز box خارج می‌شوند.

۳)





5g-1p

Intr: (34-13) * (34 -7) = 21 * 27 = 567

1p: 21 * 28 = 588

5g: 22 * 27 = 594

IOU = 567 / (588 + 594 – 567) = 567 / 615 = 0.92

2g-4p

Intr: (120-112) * (44 -27) = 8 * 17 = 136

4p: 10 * 17 = 170

2g: 9 * 19 = 171

IOU = 136 / (170 + 171 – 136) = 136 / 205 = 0.66

2g-9p

Intr: (121-118) * (45 -35) = 3 * 10 = 30

9p: 18 * 25 = 450

2g: 9 * 19 = 171

IOU = 54 / (450 + 171 – 30) = 30 / 591 = 0.05

3g-7p

Intr: (153-125) * (170 -136) = 28 * 34 = 952

7p: 29 * 35 = 1015

3g: 9 * 19 = 1050

IOU = 952 / (1015 + 1050 – 952) = 952 / 1113 = 0.85

4g-6p

Intr: (56-34) * (172 -158) = 22 * 14 = 308

6p: 21 * 34 = 714

4g: 28 * 18 = 504

IOU = 308 / (714 + 504 – 308) = 308 / 910 = 0.33

برای predictionهای باقیمانده IOU برابر 0 است.

IOU_th = 0.75

x	y	w	h	IOU	Correct?	score	Precision	Recall
110	27	10	17	0.66	FALSE	0.96	0	0
55	72	34	36	0	FALSE	0.89	0	0
13	7	21	28	0.92	TRUE	0.84	0.33	0.5
18	39	31	23	0	FALSE	0.79	0.25	0.5
124	136	29	35	0.85	TRUE	0.74	0.4	1
118	35	18	25	0.05	FALSE	0.62	0.33	1
24	98	21	34	0	FALSE	0.47	0.28	1
34	148	39	24	0.33	FALSE	0.39	0.25	1
92	153	27	47	0	FALSE	0.29	0.22	1

0.89 < th < 0.96

TP = 0 FN = 2

FP = 1 TN = 6

Precision = 0 / (0 + 1) = 0 / 1 = 0

Recall = 0 / (0 + 2) = 0 / 2 = 0

0.84 < th < 0.89

TP = 0 FN = 2

FP = 2 TN = 5

Precision = 0 / (0 + 2) = 0 / 2 = 0

Recall = 0 / (0 + 2) = 0 / 2 = 0

0.79 < th < 0.84

TP = 1 FN = 1

FP = 2 TN = 5

Precision = 1 / (1 + 2) = 1 / 3 = 0.33

Recall = 1 / (1 + 1) = 1 / 2 = 0.5

0.74 < th < 0.79

TP = 1 FN = 1

FP = 3 TN = 4

Precision = 1 / (1 + 3) = 1 / 4 = 0.25

Recall = 1 / (1 + 1) = 1 / 2 = 0.5

0.62 < th < 0.74

TP = 2 FN = 0

FP = 3 TN = 4

Precision = 2 / (2 + 3) = 2 / 5 = 0.4

Recall = 2 / (2 + 0) = 2 / 2 = 1

0.47 < th < 0.62

TP = 2 FN = 0

FP = 4 TN = 3

Precision = 2 / (2 + 4) = 2 / 6 = 0.33

Recall = 2 / (2 + 0) = 2 / 2 = 1

$0.39 < th < 0.47$

TP = 2 FN = 0

FP = 5 TN = 2

Precision = 2 / (2 + 5) = 2 / 7 = 0.28

Recall = 2 / (2 + 0) = 2 / 2 = 1

$0.29 < th < 0.39$

TP = 2 FN = 0

FP = 6 TN = 1

Precision = 2 / (2 + 6) = 2 / 8 = 0.25

Recall = 2 / (2 + 0) = 2 / 2 = 1

$th < 0.29$

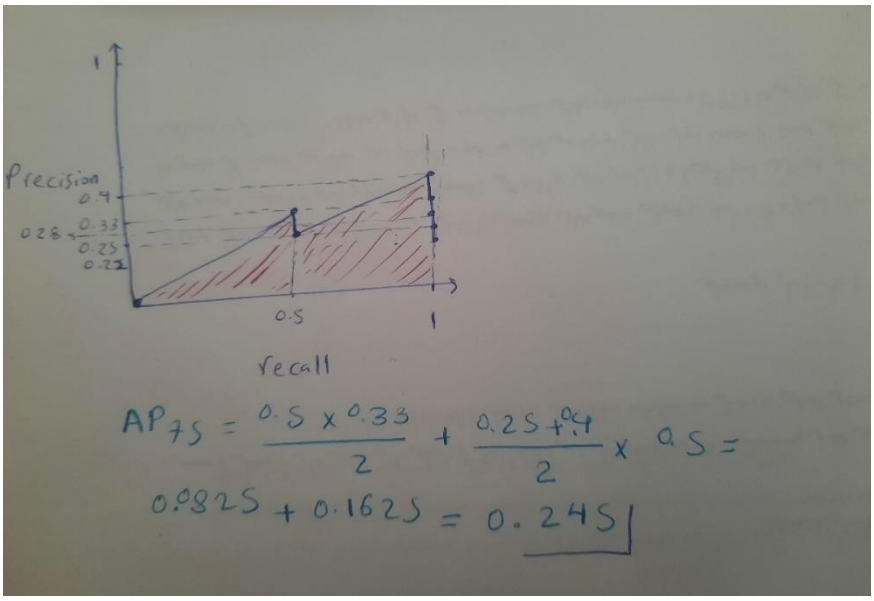
TP = 2 FN = 0

FP = 7 TN = 0

Precision = 2 / (2 + 7) = 2 / 9 = 0.22

Recall = 2 / (2 + 0) = 2 / 2 = 1

.



IOU_th = 0.5

x	y	w	h	IOU	Correct?	score	Precision	Recall
110	27	10	17	0.66	TRUE	0.96	1	0.33
55	72	34	36	0	FALSE	0.89	0.5	0.33
13	7	21	28	0.92	TRUE	0.84	0.66	0.66
18	39	31	23	0	FALSE	0.79	0.5	0.66
124	136	29	35	0.85	TRUE	0.74	0.6	1
118	35	18	25	0.05	FALSE	0.62	0.5	1
24	98	21	34	0	FALSE	0.47	0.43	1
34	148	39	24	0.33	FALSE	0.39	0.375	1
92	153	27	47	0	FALSE	0.29	0.33	1

$0.89 < th < 0.96$

TP = 1 FN = 2

FP = 0 TN = 6

Precision = 1 / (1 + 0) = 1 / 1 = 1

Recall = 1 / (1 + 2) = 1 / 3 = 0.33

0.84 < th < 0.89

TP = 1 FN = 2

FP = 1 TN = 5

Precision = 1 / (1 + 1) = 1 / 2 = 0.5

Recall = 1 / (1 + 2) = 1 / 3 = 0.33

0.79 < th < 0.84

TP = 2 FN = 1

FP = 1 TN = 5

Precision = 2 / (2 + 1) = 2 / 3 = 0.66

Recall = 2 / (2 + 1) = 2 / 3 = 0.66

0.74 < th < 0.79

TP = 2 FN = 1

FP = 2 TN = 4

Precision = 2 / (2 + 2) = 2 / 4 = 0.5

Recall = 2 / (2 + 1) = 2 / 3 = 0.66

0.62 < th < 0.74

TP = 3 FN = 0

FP = 2 TN = 4

Precision = 3 / (3 + 2) = 3 / 5 = 0.6

Recall = 3 / (3 + 0) = 3 / 3 = 1

0.47 < th < 0.62

TP = 3 FN = 0

FP = 3 TN = 3

Precision = 3 / (3 + 3) = 3 / 6 = 0.5

Recall = 3 / (3 + 0) = 3 / 3 = 1

0.39 < th < 0.47

TP = 3 FN = 0

FP = 4 TN = 2

Precision = 3 / (3 + 4) = 3 / 7 = 0.43

Recall = 3 / (3 + 0) = 3 / 3 = 1

0.29 < th < 0.39

TP = 3 FN = 0

FP = 5 TN = 1

Precision = 3 / (3 + 5) = 3 / 8 = 0.375

Recall = 3 / (3 + 0) = 3 / 3 = 1

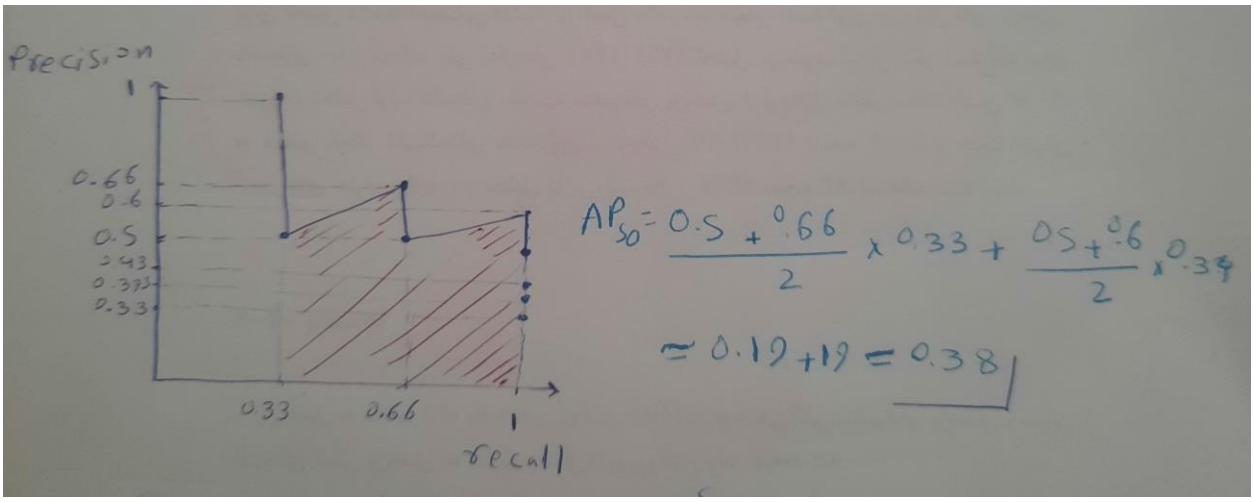
th < 0.29

TP = 3 FN = 0

FP = 6 TN = 0

Precision = 3 / (3 + 6) = 3 / 9 = 0.33

Recall = 3 / (3 + 0) = 3 / 3 = 1



IOU_th = 0.25

x	y	w	h	IOU	Correct?	score	Precision	Recall
110	27	10	17	0.66	TRUE	0.96	1	0.25
55	72	34	36	0	FALSE	0.89	0.5	0.25
13	7	21	28	0.92	TRUE	0.84	0.66	0.5
18	39	31	23	0	FALSE	0.79	0.5	0.5
124	136	29	35	0.85	TRUE	0.74	0.6	0.75
118	35	18	25	0.05	FALSE	0.62	0.5	0.75
24	98	21	34	0	FALSE	0.47	0.43	0.75
34	148	39	24	0.33	TRUE	0.39	0.5	1
92	153	27	47	0	FALSE	0.29	0.44	1

0.89 < th < 0.96

TP = 1 FN = 3

FP = 0 TN = 5

Precision = 1 / (1 + 0) = 1 / 1 = 1

Recall = 1 / (1 + 3) = 1 / 4 = 0.25

0.84 < th < 0.89

TP = 1 FN = 3

FP = 1 TN = 4

Precision = 1 / (1 + 1) = 1 / 2 = 0.5

Recall = 1 / (1 + 3) = 1 / 4 = 0.25

0.79 < th < 0.84

TP = 2 FN = 2

FP = 1 TN = 4

Precision = 2 / (2 + 1) = 2 / 3 = 0.66

Recall = 2 / (2 + 2) = 2 / 4 = 0.5

0.74 < th < 0.79

TP = 2 FN = 2

FP = 2 TN = 3

Precision = 2 / (2 + 2) = 2 / 4 = 0.5

Recall = 2 / (2 + 2) = 2 / 4 = 0.5

0.62 < th < 0.74

TP = 3 FN = 1

FP = 2 TN = 3

Precision = 3 / (3 + 2) = 3 / 5 = 0.6

Recall = 3 / (3 + 1) = 3 / 4 = 0.75

0.47 < th < 0.62

TP = 3 FN = 1

FP = 3 TN = 2

Precision = 3 / (3 + 3) = 3 / 6 = 0.5

Recall = 3 / (3 + 1) = 3 / 4 = 0.75

0.39 < th < 0.47

TP = 3 FN = 1

FP = 4 TN = 1

Precision = 3 / (3 + 4) = 3 / 7 = 0.43

Recall = 3 / (3 + 1) = 3 / 4 = 0.75

0.29 < th < 0.39

TP = 4 FN = 0

FP = 4 TN = 1

Precision = 4 / (4 + 4) = 4 / 8 = 0.5

Recall = 4 / (4 + 0) = 4 / 4 = 1

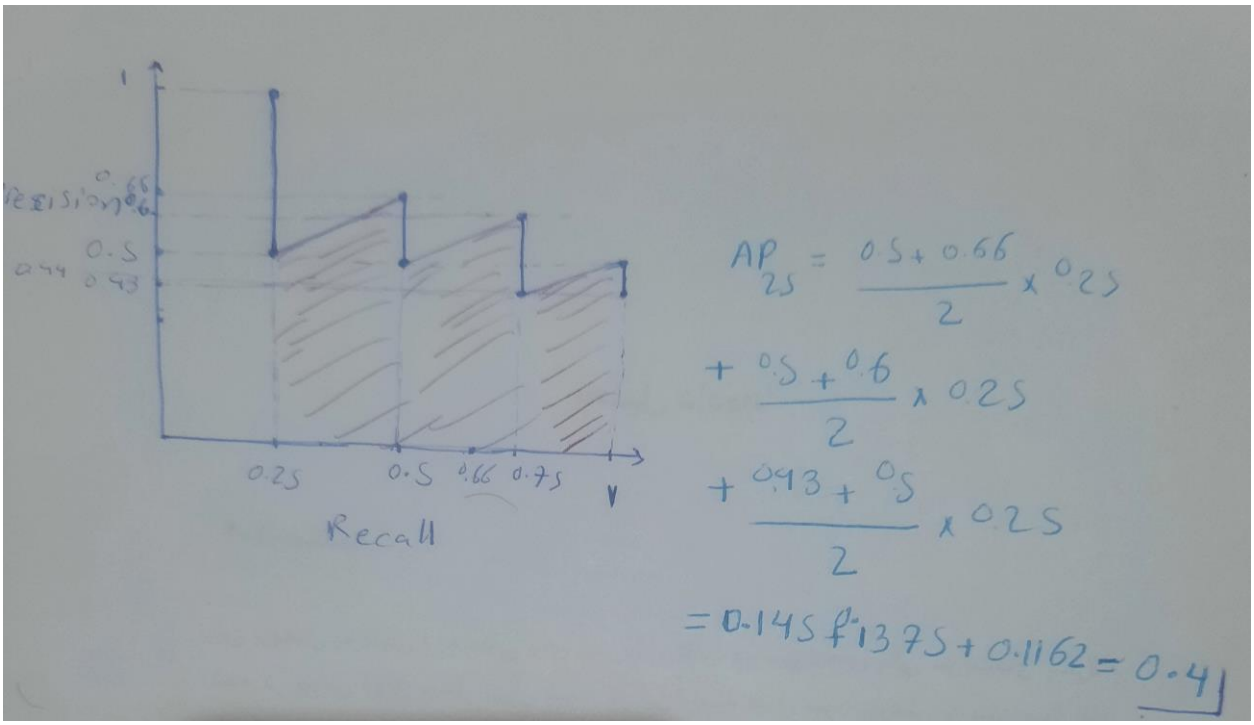
th < 0.29

TP = 4 FN = 0

FP = 5 TN = 0

Precision = 4 / (4 + 5) = 4 / 9 = 0.44

Recall = 4 / (4 + 0) = 4 / 4 = 1



۵)

ب)

Bag of freebies

معمولاً یک object detector معمولی به صورت آفلاین آموزش داده می‌شود. بنابراین، محققان همیشه دوست دارند از این مزیت استفاده کنند و روش‌های آموزشی بهتری را توسعه دهند که می‌تواند باعث شود object detector بدون افزایش هزینه استنتاج، دقت بهتری داشته باشد. ما به این روش‌ها که فقط استراتژی آموزشی را تغییر می‌دهند یا فقط هزینه آموزش را افزایش می‌دهند، می‌گوییم «Bag of freebies ».

Bag of specials

برای آن دسته از ماژول‌های پلاگین و روش‌های post-processing که فقط هزینه استنتاج را به میزان کمی افزایش می‌دهند، اما می‌توانند دقت object detection را به میزان قابل توجهی بهبود بخشند، ما آنها را " Bag of specials " می‌نامیم. به طور کلی، این ماژول‌های پلاگین برای افزایش ویژگی‌های خاص در یک

مدل مانند بزرگ‌کردن receptive field ، معرفی مکانیسم توجه، یا تقویت قابلیت یکپارچه‌سازی ویژگی و غیره هستند و post-processing روشی برای غربالگری نتایج پیش‌بینی مدل است.

YOLOv4 از موارد زیر استفاده می‌کند:

Bag of Freebies (BoF) برای ستون فقرات (کلاسیفایر): روش‌های داده‌افزایی CutMix و Mosaic، رگولاریزیشن DropBlock، smooth کردن برچسب کلاس

Bag of specials (BoS) برای ستون فقرات (کلاسیفایر): تابع فعال سازی Mish، اتصالات جزئی بین مرحله‌ای (CSP)، اتصالات residual وزن‌دار با چند ورودی (MiWRC)

Bag of Freebies (BoF) برای تشخیص‌دهنده (detector): خطای رگرسیون جعبه مرزی CIoU ، Cross mini-Batch Normalization (CmBN) ، رگولاریزیشن DropBlock، روش داده‌افزایی Mosaic ، آموزش خود خصمانه (Self-Adversarial) ، حذف حساسیت grid، استفاده از anchorهای متعدد برای یک ground truth واحد، زمان بندی ذوب کسینوسی ، هایپرپارامترهای بهینه، اشکال آموزشی تصادفی

Bag of specials (BoS) برای تشخیص‌دهنده (detector): فعال سازی Mish، SPP-block، SAM-block، PAN path-Aggregation block، خطای رگرسیون جعبه مرزی DIoU-NMS

YOLOv4 = CSPDarknet53+SPP+PAN+YOLOv3

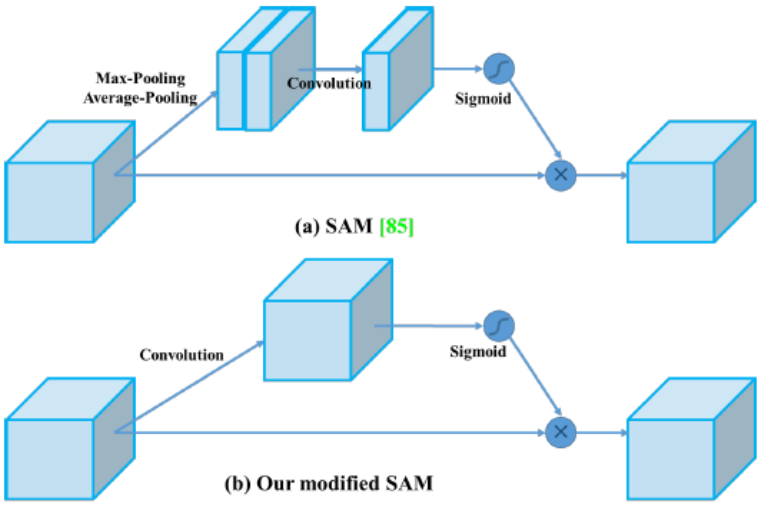


Figure 5: Modified SAM.

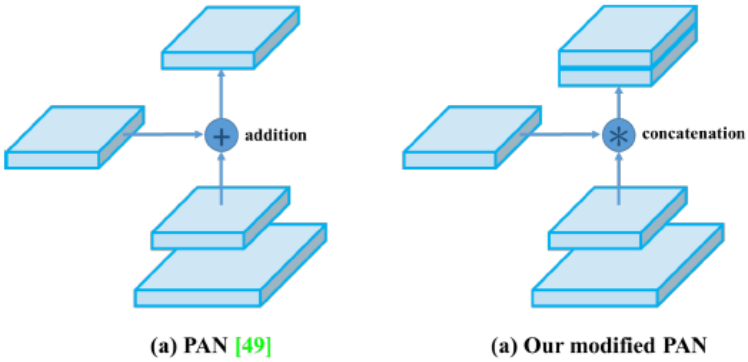


Figure 6: Modified PAN.