计算机图像实体检测综述

A Survey on Computer Image Entity Detection

郑涵潇 林超超 吴浩凯

汇报目录

01

简单
介绍

Introduce

02

进展
总结

Survey

03

项目
展示

Demo

04

未来
展望

Future

01 简单介绍

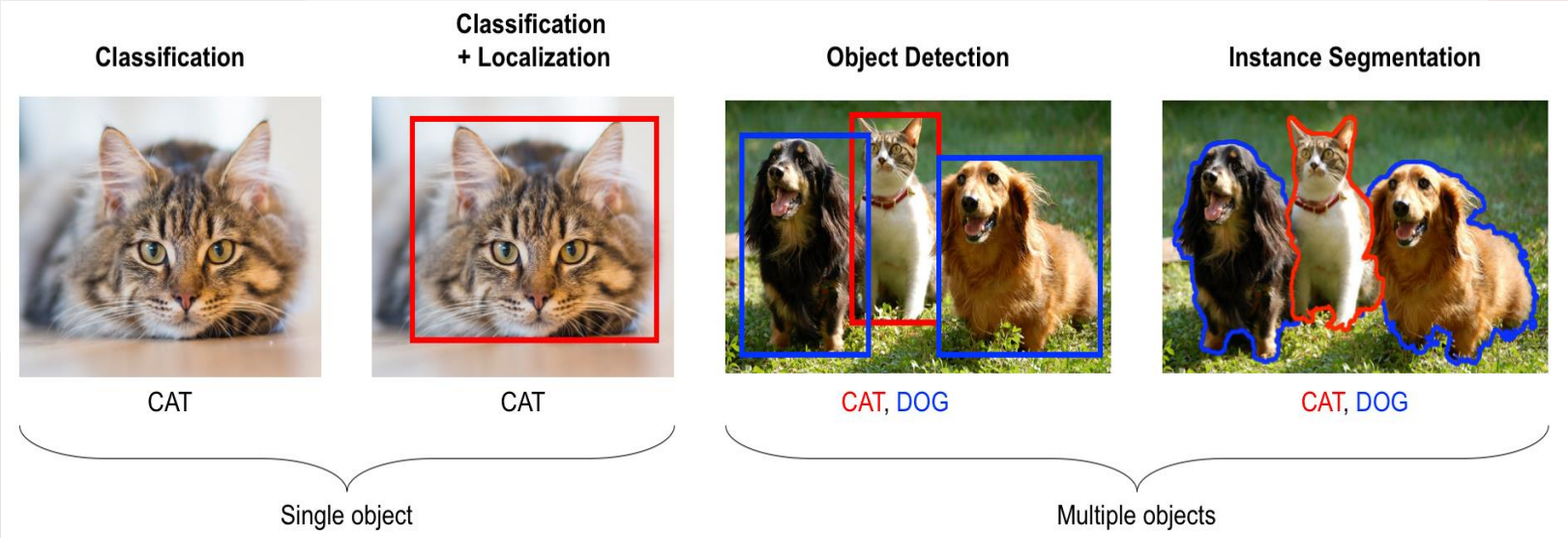
Introduce

What is 计算机图像实体检测？

定位 + 分类

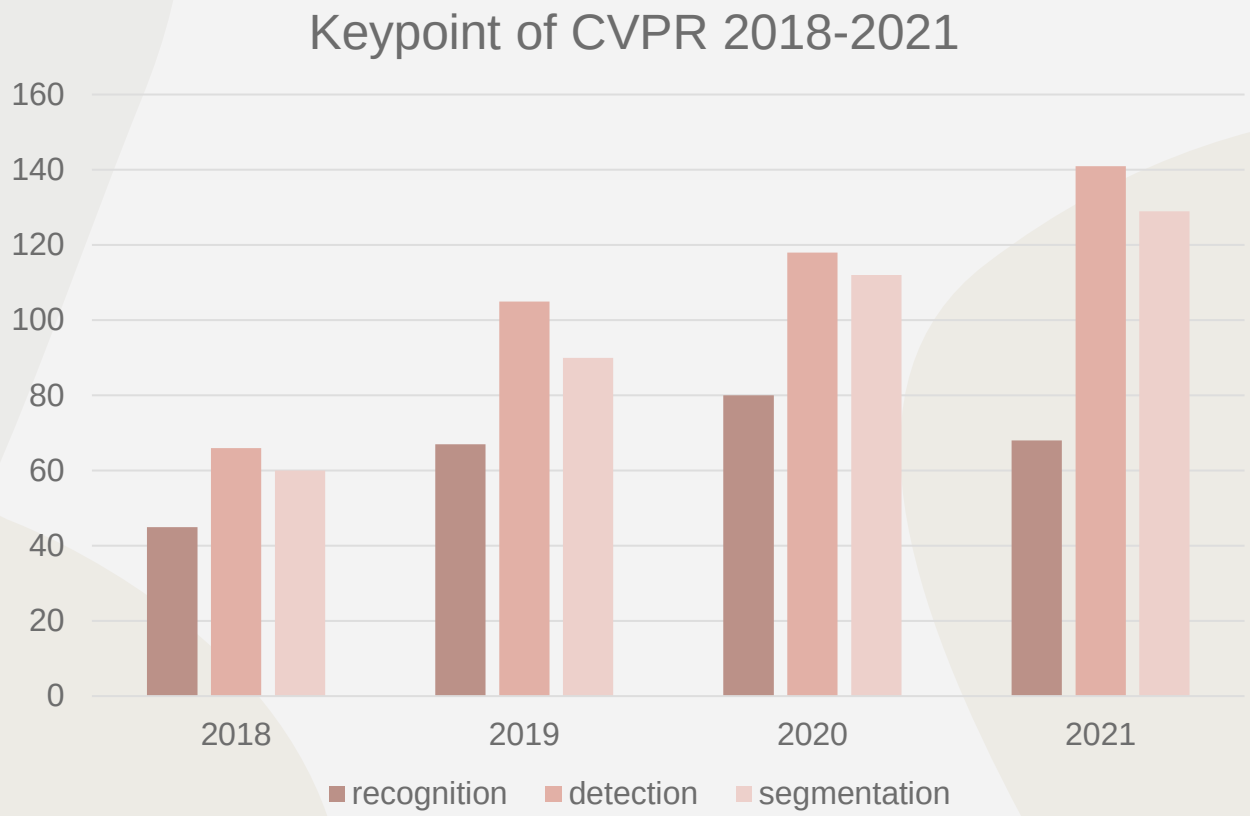
Our Demo1 垃圾分类 By YOLO

What is 计算机图像实体检测？

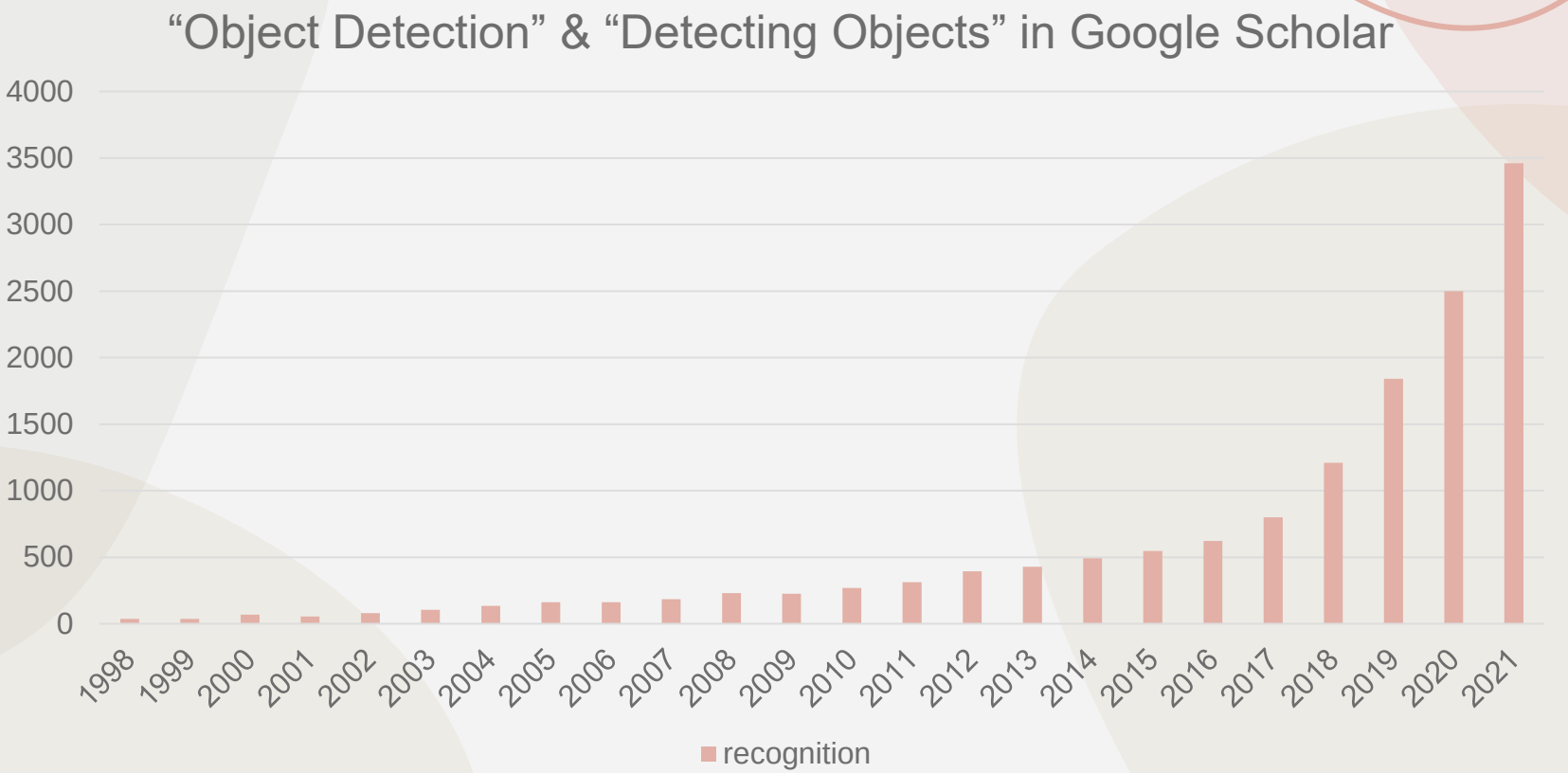


Where is 计算机图像实体检测？

How is 计算机图像实体检测？



How is 计算机图像实体检测？



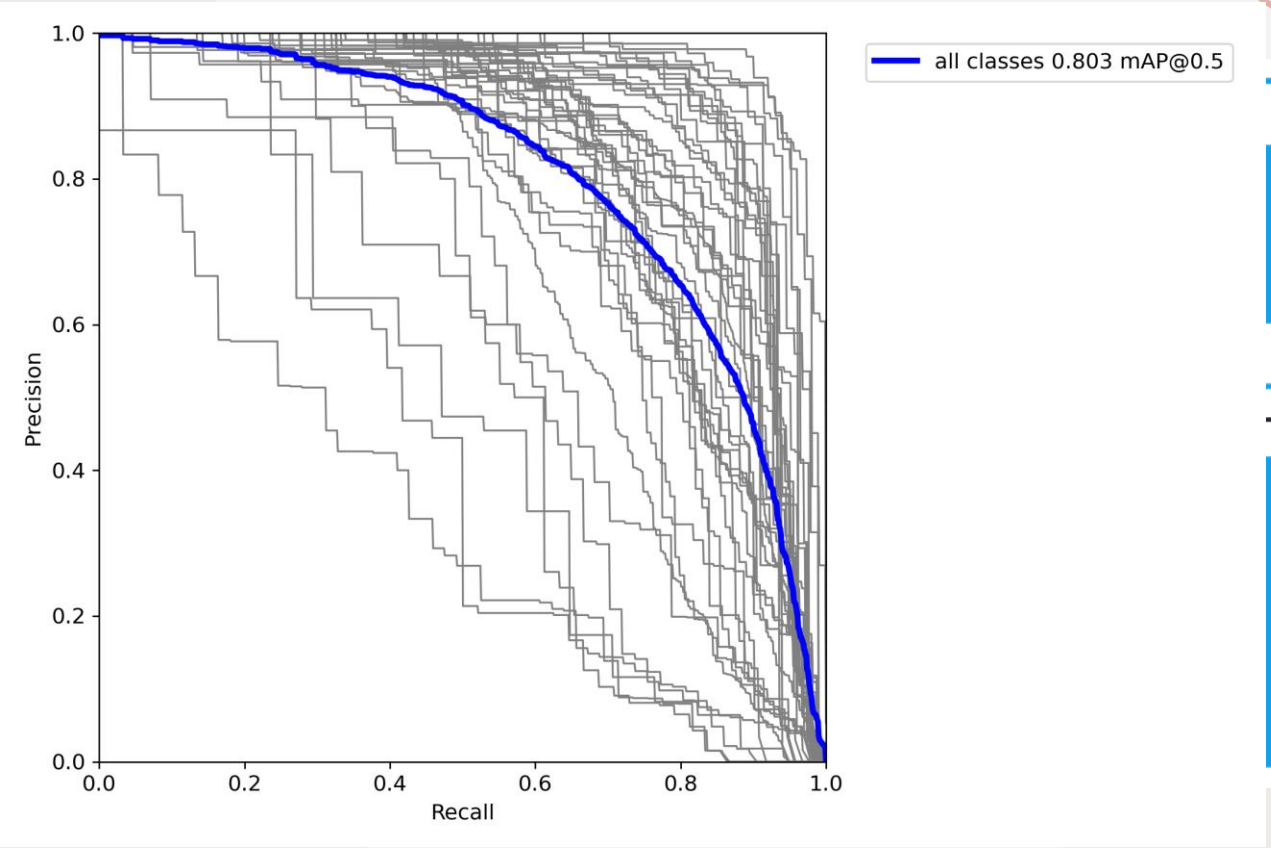
How to 计算机图像实体检测？

$Accur$

$Precis$

$Recal$

$FPR =$



Our Demo1 垃圾分类 By YOLO

How to 计算机图像实体检测？

应用领域	数据集	发布时间	数据集描述
	WebFace	2021	该数据集是最大的百万级公开人脸识别训练集，包含了有噪声的 4M 身份/260M 人脸(WebFace260M) 和清理过的 2M 身份/42M 人脸(WebFace42M) 链接：https://www.face-benchmark.org
文本检测	MSRA-TD500	2012	该数据集包含 500 张自然图像，作为评估文本检测算法的基准。从室内（办公室和商场）和室外（街道）场景中拍摄，由于文本的多样性和图像背景的复杂性，数据集具有挑战性。链接：http://www.iapr-c11.org/mediawiki/index.php/MSRA_Text_Detection_500_Database_(MSRA-TD500)
	COCOText	2016	该数据集基于 MSCOCO 数据集，用于自然图像中的文本检测和识别，最新版包含 63,686 个图像，145,859 个文本实例，包括手写、打印、易读、难读、英文和非英文，数据集多样化。 链接：https://vision.cornell.edu/se3/coco-text-2/
	Google FSNS	2017	该数据集包含超过一百万张从法国 Google 街景图像中裁剪出来的街道名称标志图像。 每个图像包含相同街道名称标志的四个视图， 路牌中的文字最多可以跨越三行。 链接：https://rrc.cvc.uab.es/?ch=6
遥感检测	VeDAI	2015	该数据集用于遥感图像中的多类车辆检测，包含 1210 张图像，总共 3640 个车辆实例与 9 个类别，空间分辨率 12.5 cm。车辆实例具有多方向、光照/阴影变化、镜面反射或遮挡等表现。 链接：https://downloads.greyc.fr/vedai/
	DOTA	2017	该数据集来源 GoogleEarth 和中国卫星，是目前最大的光学遥感图像数据集，包含 2806 张遥感图像（图片尺寸从 800*800 到 4000*4000），一共 188,282 个实例，分为 15 个类别。 链接：https://captain-whu.github.io/DOTA/dataset.html

https://image-net.org/download-images.php https://cocodataset.org/#home http://host.robots.ox.ac.uk:8080/ https://github.com/openimages/dataset

02 进展总结

Survey

Life of 计算机图像实体检测

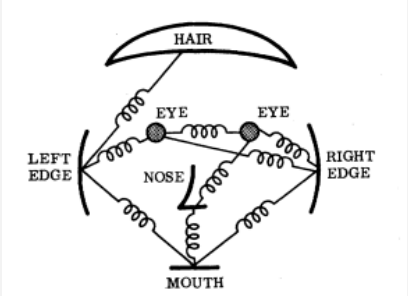
早期探索阶段(1966-1998)

1966

Marvin Minsky
&
Gerald Sussman

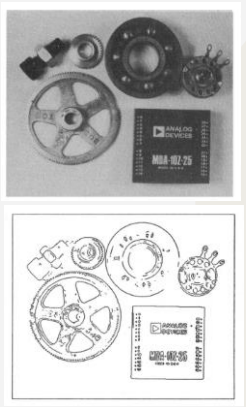
1973

图画结构
Pictorial Structure



1980s

人工智能寒冬
各种滤波器诞生



1998

神经网络
人脸检测

Life of 计算机图像实体检测

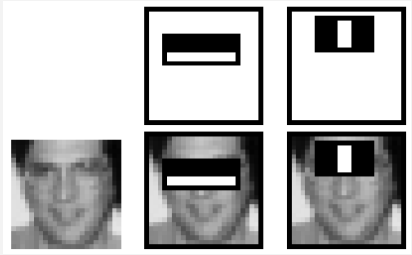
传统计算阶段(1999-2011)

1999

SIFT
尺度不变特征变换

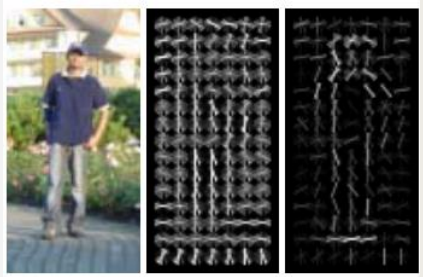
2001

VJ检测器
Harr-like特征



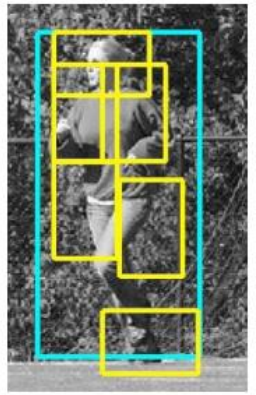
2005

HOG检测器
方向梯度直方图



2008

DPM
可变形部件模型



Life of 计算机图像实体检测

传统计算阶段(1999-2011)

- 选取候选区域

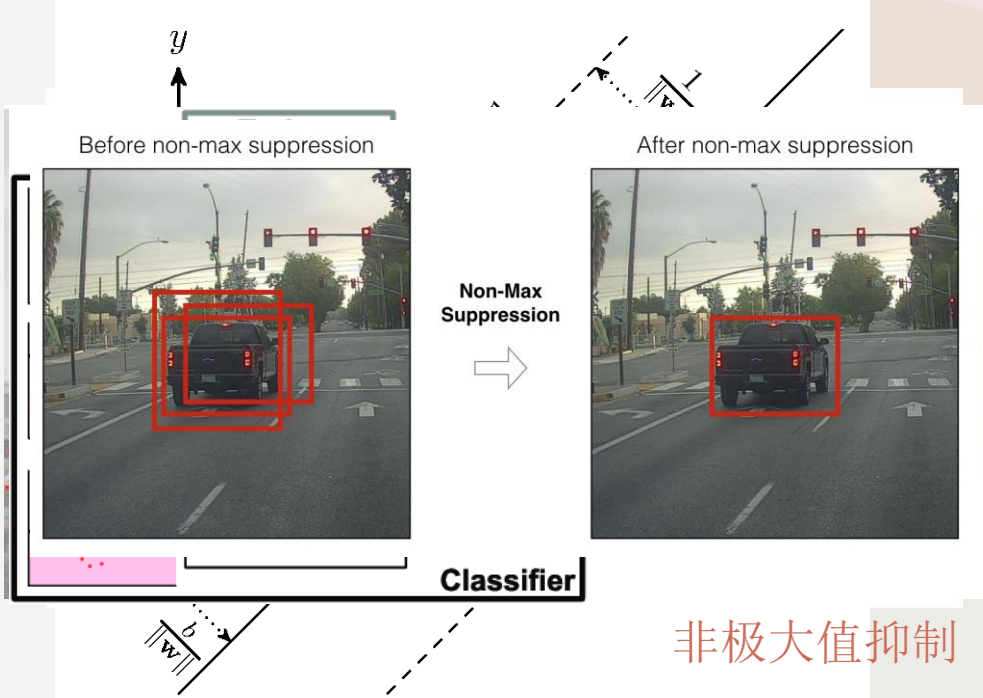
  滑动窗口法　选择性搜索法　…

- 特征提取

  SIFT　HOG　…

- 特征向量分类

  SVM　AdaBoost　…



非极大值抑制

Life of 计算机图像实体检测

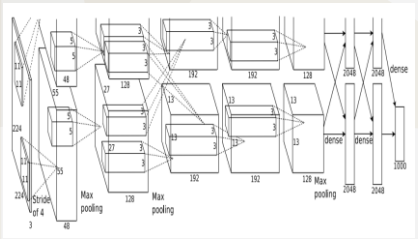
深度学习阶段(2012-至今)

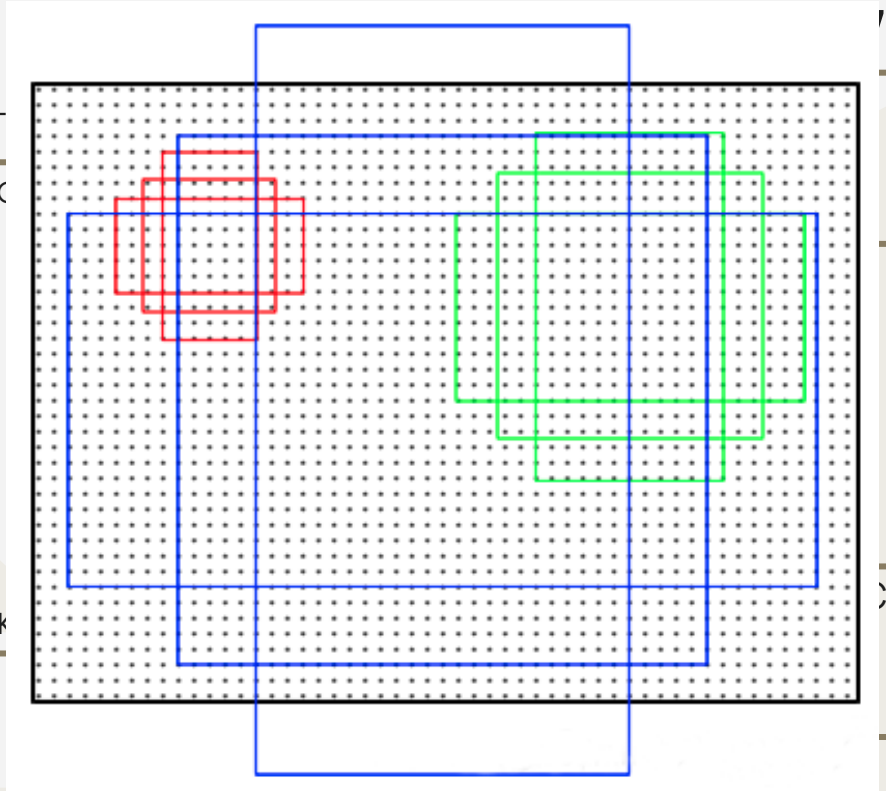
2012

深度卷积神经网络
AlexNet

Anchor-based

Anchor-free





2018

Cascade RCNN

2019

Grid RCNN

2020

YOLOv4

2021

YOLOv5

2019

CornerNet-Lite
ExtremeNet

ResPoint

2021

SAPD

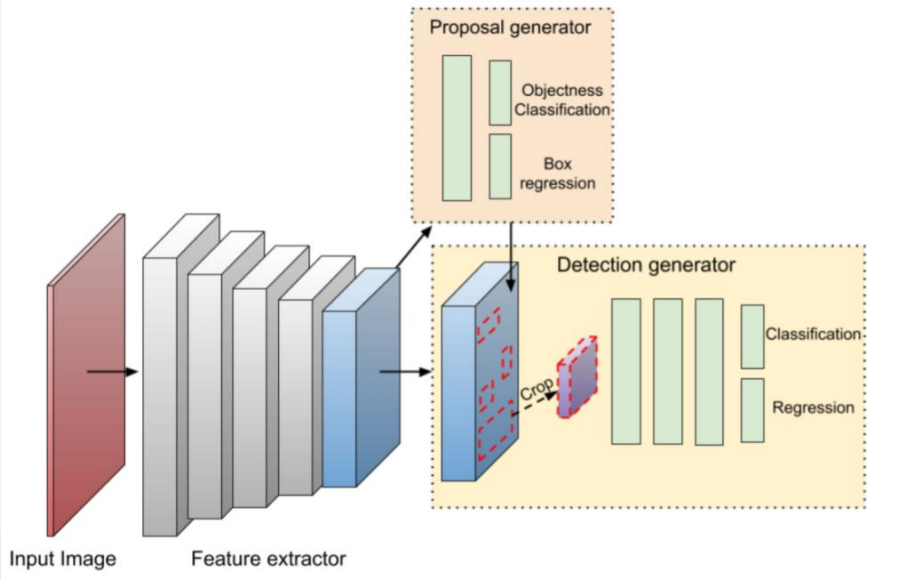
YOLOX

Anchor-based

Two-Stage

Life of 计算机图像实体检测

Anchor-based Two-Stage



2014	2015	2016	2017	2018	2019
RCNN SPPNet	Fast RCNN	Faster RCNN	FPN	Cascade RCNN	Grid RCNN

Life of 计算机图像实体检测

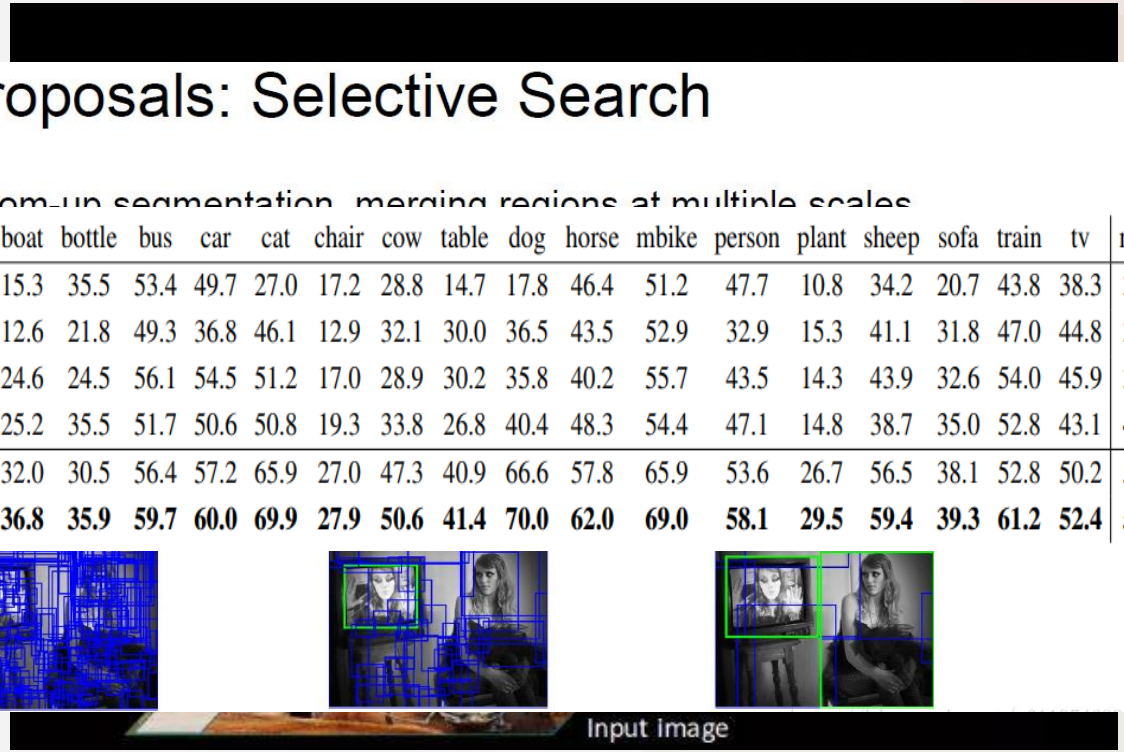
Anchor-based Two-Stage

RCNN：Region CNN

BBox Regressor

AlexNet

NMS



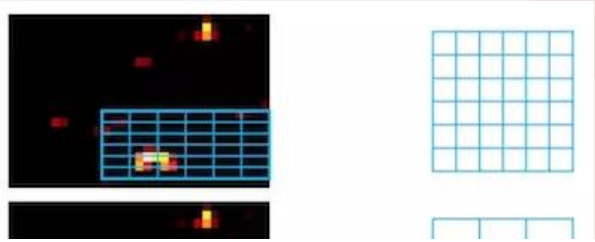
VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5 [17]†	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
UVA [32]	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.1	30.0	36.5	43.5	52.9	32.9	15.3	41.1	31.8	47.0	44.8	35.1
Regionlets [35]	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM [15]†	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	**71.8**	**65.8**	**53.0**	**36.8**	**35.9**	**59.7**	**60.0**	**69.9**	**27.9**	**50.6**	**41.4**	**70.0**	**62.0**	**69.0**	**58.1**	**29.5**	**59.4**	**39.3**	**61.2**	**52.4**	**53.7**

https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html

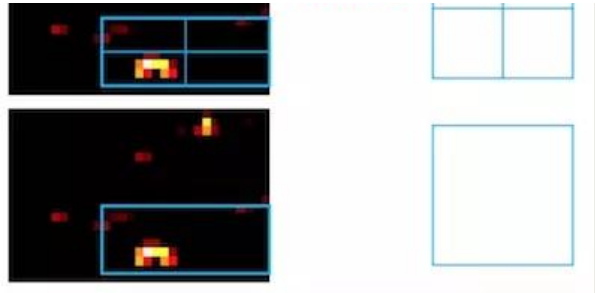
Life of 计算机图像实体检测

Anchor-based Two-Stage

SPPNet：Spatial Pyramid Pooling



method	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
SPP-net (1)	59.2	**68.6**	69.7	57.1	41.2	40.5	66.3	71.3	72.5	34.4	**67.3**	61.7	63.1	71.0	69.8	57.6	29.7	59.0	50.2	65.2	68.0
SPP-net (2)	59.1	65.7	71.4	57.4	**42.4**	39.9	67.0	71.4	70.6	32.4	66.7	61.7	64.8	71.7	70.4	56.5	30.8	59.9	53.2	63.9	64.6
combination	**60.9**	68.5	**71.7**	**58.7**	41.9	**42.5**	**67.7**	**72.1**	**73.8**	**34.7**	67.0	**63.4**	**66.0**	**72.5**	**71.3**	**58.9**	**32.8**	**60.9**	**56.1**	**67.9**	**68.8**



https://ieeexplore.ieee.org/abstract/document/7005506/

Life of 计算机图像实体检测

Anchor-based Two-Stage

Fast RCNN

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
SPPnet BB [11]†	07 \ diff	73.9	72.3	62.5	51.5	44.4	74.4	73.0	74.4	42.3	73.6	57.7	70.3	74.6	74.3	54.2	34.0	56.4	56.4	67.9	73.5	63.1
R-CNN BB [10]	07	73.4	77.0	63.4	45.4	**44.6**	75.1	78.1	79.8	40.5	73.7	62.2	79.4	78.1	73.1	64.2	**35.6**	66.8	67.2	70.4	**71.1**	66.0
FRCN [ours]	07	74.5	78.3	69.2	53.2	36.6	77.3	78.2	82.0	40.7	72.7	67.9	79.6	79.2	73.0	69.0	30.1	65.4	70.2	75.8	65.8	66.9
FRCN [ours]	07 \ diff	74.6	**79.0**	68.6	57.0	39.3	79.5	**78.6**	81.9	**48.0**	74.0	67.4	80.5	80.7	74.1	69.6	31.8	67.1	68.4	75.3	65.5	68.1
FRCN [ours]	07+12	**77.0**	78.1	**69.3**	**59.4**	38.3	**81.6**	**78.6**	**86.7**	42.8	**78.8**	**68.9**	**84.7**	**82.0**	**76.6**	**69.9**	31.8	**70.1**	**74.8**	**80.4**	70.4	**70.0**

Table 1. **VOC 2007 test** detection average precision (%). All methods use VGG16. Training set key: **07**: VOC07 trainval, **07 \ diff**: **07** without "difficult" examples, **07+12**: union of **07** and VOC12 trainval. †SPPnet results were prepared by the authors of [11].

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
BabyLearning	Prop.	77.7	73.8	62.3	48.8	45.4	67.3	67.0	80.3	41.3	70.8	49.7	79.5	74.7	78.6	64.5	36.0	69.9	55.7	70.4	61.7	63.8
R-CNN BB [10]	12	79.3	72.4	63.1	44.0	44.4	64.6	66.3	84.9	38.8	67.3	48.4	82.3	75.0	76.7	65.7	35.8	66.2	54.8	69.1	58.8	62.9
SegDeepM	12+seg	**82.3**	75.2	67.1	50.7	**49.8**	71.1	69.6	88.2	42.5	71.2	50.0	85.7	76.6	81.8	69.3	**41.5**	**71.9**	62.2	73.2	**64.6**	67.2
FRCN [ours]	12	80.1	74.4	67.7	49.4	41.4	74.2	68.8	87.8	41.9	70.1	50.2	86.1	77.3	81.1	70.4	33.3	67.0	63.3	77.2	60.0	66.1
FRCN [ours]	07++12	82.0	**77.8**	**71.6**	**55.3**	42.4	**77.3**	**71.7**	**89.3**	**44.5**	**72.1**	**53.7**	**87.7**	**80.0**	**82.5**	**72.7**	36.6	68.7	**65.4**	**81.1**	62.7	**68.8**

Table 2. **VOC 2010 test** detection average precision (%). BabyLearning uses a network based on [17]. All other methods use VGG16. Training set key: **12**: VOC12 trainval, **Prop.**: proprietary dataset, **12+seg**: **12** with segmentation annotations, **07++12**: union of VOC07 trainval, VOC07 test, and VOC12 trainval.

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
BabyLearning	Prop.	78.0	74.2	61.3	45.7	42.7	68.2	66.8	80.2	40.6	70.0	49.8	79.0	74.5	77.9	64.0	35.3	67.9	55.7	68.7	62.6	63.2
NUS_NIN_c2000	Unk.	80.2	73.8	61.9	43.7	**43.0**	70.3	67.6	80.7	41.9	69.7	51.7	78.2	75.2	76.9	65.1	**38.6**	**68.3**	58.0	68.7	63.3	63.8
R-CNN BB [10]	12	79.6	72.7	61.9	41.2	41.9	65.9	66.4	84.6	38.5	67.2	46.7	82.0	74.8	76.0	65.2	35.6	65.4	54.2	67.4	60.3	62.4
FRCN [ours]	12	80.3	74.7	66.9	46.9	37.7	73.9	68.6	87.7	41.7	71.1	51.1	86.0	77.8	79.8	69.8	32.1	65.5	63.8	76.4	61.7	65.7
FRCN [ours]	07++12	**82.3**	**78.4**	**70.8**	**52.3**	38.7	**77.8**	**71.6**	**89.3**	**44.2**	**73.0**	**55.0**	**87.5**	**80.5**	**80.8**	**72.0**	35.1	**68.3**	**65.7**	**80.4**	**64.2**	**68.4**

Table 3. **VOC 2012 test** detection average precision (%). BabyLearning and NUS_NIN_c2000 use networks based on [17]. All other methods use VGG16. Training set key: see Table 2, **Unk.**: unknown.

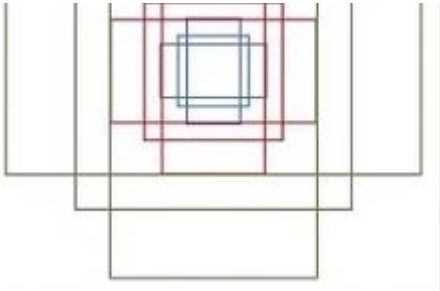
http://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html

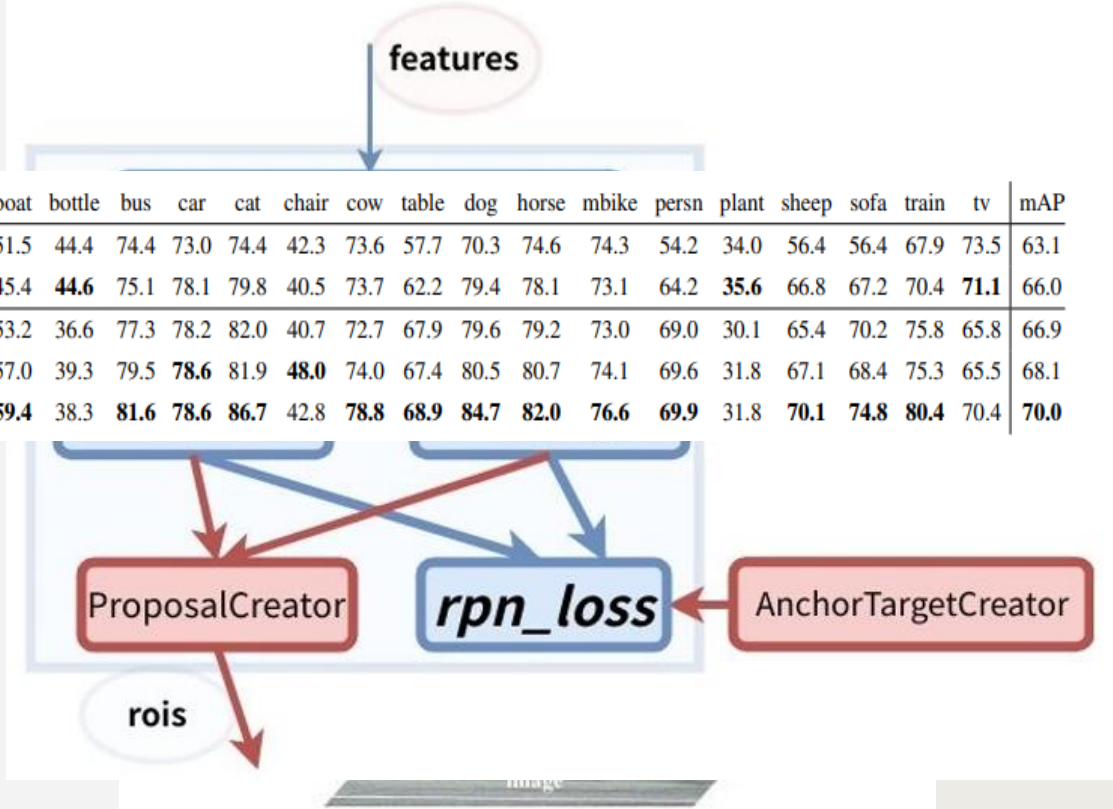
Life of 计算机图像实体检测

Anchor-based Two-Stage

Faster RCNN

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
SPPnet BB [11]†	07 \ diff	73.9	72.3	62.5	51.5	44.4	74.4	73.0	74.4	42.3	73.6	57.7	70.3	74.6	74.3	54.2	34.0	56.4	56.4	67.9	73.5	63.1
R-CNN BB [10]	07	73.4	77.0	63.4	45.4	**44.6**	75.1	78.1	79.8	40.5	73.7	62.2	79.4	78.1	73.1	64.2	**35.6**	66.8	67.2	70.4	**71.1**	66.0
FRCN [ours]	07	74.5	78.3	69.2	53.2	36.6	77.3	78.2	82.0	40.7	72.7	67.9	79.6	79.2	73.0	69.0	30.1	65.4	70.2	75.8	65.8	66.9
FRCN [ours]	07 \ diff	74.6	**79.0**	68.6	57.0	39.3	79.5	**78.6**	81.9	**48.0**	74.0	67.4	80.5	80.7	74.1	69.6	31.8	67.1	68.4	75.3	65.5	68.1
FRCN [ours]	07+12	**77.0**	78.1	**69.3**	**59.4**	38.3	**81.6**	**78.6**	**86.7**	42.8	**78.8**	**68.9**	**84.7**	**82.0**	**76.6**	**69.9**	31.8	**70.1**	**74.8**	**80.4**	70.4	**70.0**



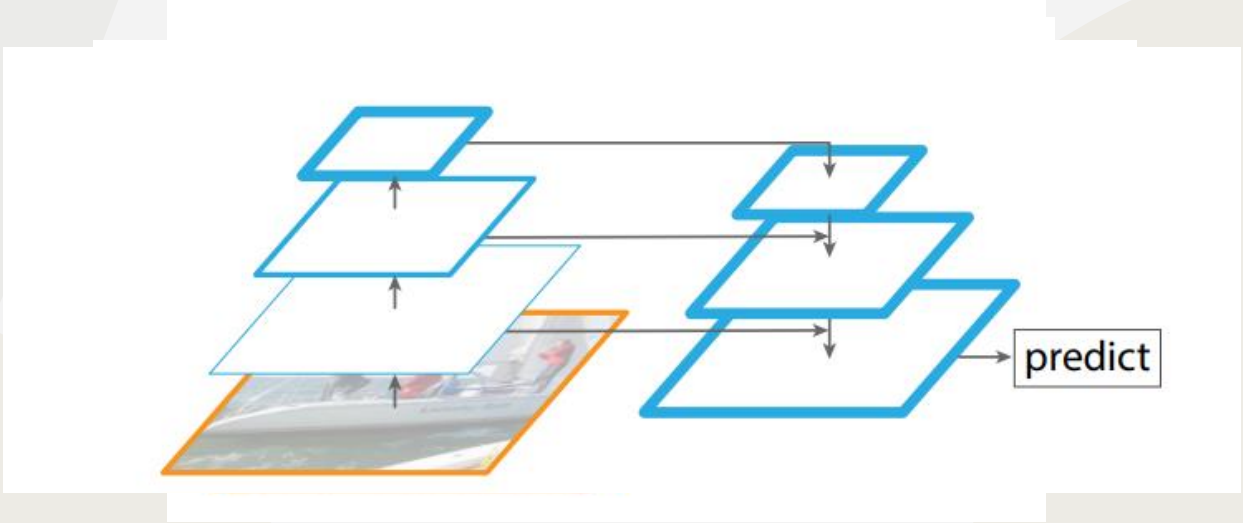


https://arxiv.org/pdf/1506.01497.pdf

Life of 计算机图像实体检测

Anchor-based Two-Stage

基于金字塔的实体检测



单层特征金字塔

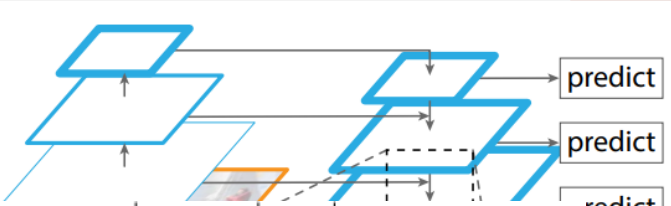
多层特征金字塔

http://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.html

Life of 计算机图像实体检测

Anchor-based Two-Stage

FPN：Feature Pyramid Networks



Fast R-CNN	proposals	feature	head	lateral?	top-down?	AP@0.5	AP	AP_s	AP_m	AP_l
(a) baseline on conv4	RPN, $\{P_k\}$	C_4	conv5			54.7	31.9	15.7	36.5	45.5
(b) baseline on conv5	RPN, $\{P_k\}$	C_5	2*fc*			52.9	28.8	11.9	32.4	43.4
(c) **FPN**	RPN, $\{P_k\}$	$\{P_k\}$	2*fc*	✓	✓	**56.9**	**33.9**	**17.8**	**37.7**	**45.8**
Ablation experiments follow:										
(d) bottom-up pyramid	RPN, $\{P_k\}$	$\{P_k\}$	2*fc*	✓		44.9	24.9	10.9	24.4	38.5
(e) top-down pyramid, w/o lateral	RPN, $\{P_k\}$	$\{P_k\}$	2*fc*		✓	54.0	31.3	13.3	35.2	45.3
(f) only finest level	RPN, $\{P_k\}$	P_2	2*fc*	✓	✓	56.3	33.4	17.3	37.3	45.6

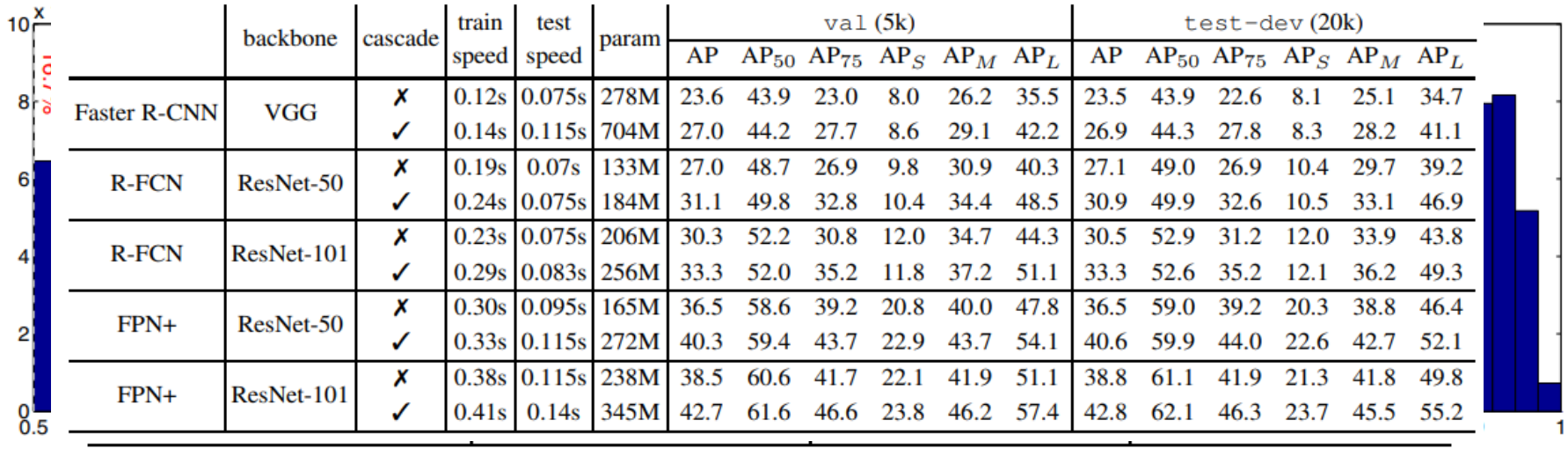
多层特征融合金字塔

http://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.html

Life of 计算机图像实体检测

Anchor-based Two-Stage

Cascade RCNN



	backbone	cascade	train speed	test speed	param	val (5k) AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$	test-dev (20k) AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$
Faster R-CNN	VGG	✗	0.12s	0.075s	278M	23.6	43.9	23.0	8.0	26.2	35.5	23.5	43.9	22.6	8.1	25.1	34.7
		✓	0.14s	0.115s	704M	27.0	44.2	27.7	8.6	29.1	42.2	26.9	44.3	27.8	8.3	28.2	41.1
R-FCN	ResNet-50	✗	0.19s	0.07s	133M	27.0	48.7	26.9	9.8	30.9	40.3	27.1	49.0	26.9	10.4	29.7	39.2
		✓	0.24s	0.075s	184M	31.1	49.8	32.8	10.4	34.4	48.5	30.9	49.9	32.6	10.5	33.1	46.9
R-FCN	ResNet-101	✗	0.23s	0.075s	206M	30.3	52.2	30.8	12.0	34.7	44.3	30.5	52.9	31.2	12.0	33.9	43.8
		✓	0.29s	0.083s	256M	33.3	52.0	35.2	11.8	37.2	51.1	33.3	52.6	35.2	12.1	36.2	49.3
FPN+	ResNet-50	✗	0.30s	0.095s	165M	36.5	58.6	39.2	20.8	40.0	47.8	36.5	59.0	39.2	20.3	38.8	46.4
		✓	0.33s	0.115s	272M	40.3	59.4	43.7	22.9	43.7	54.1	40.6	59.9	44.0	22.6	42.7	52.1
FPN+	ResNet-101	✗	0.38s	0.115s	238M	38.5	60.6	41.7	22.1	41.9	51.1	38.8	61.1	41.9	21.3	41.8	49.8
		✓	0.41s	0.14s	345M	42.7	61.6	46.6	23.8	46.2	57.4	42.8	62.1	46.3	23.7	45.5	55.2

http://openaccess.thecvf.com/content_cvpr_2018/html/Cai_Cascade_R-CNN_Delving_CVPR_2018_paper.html

Life of 计算机图像实体检测

Anchor-based Two-Stage

Grid RCNN

method	backbone	AP	$AP_{.5}$	$AP_{.75}$	AP_S	AP_M	AP_L
Faster R-CNN	ResNet-50	33.8	55.4	35.9	17.4	37.9	45.3
Grid R-CNN	ResNet-50	**35.9**	54.0	38.0	18.6	40.2	47.8
Faster R-CNN w FPN	ResNet-50	37.4	59.3	40.3	21.8	40.9	47.9
Grid R-CNN w FPN	ResNet-50	**39.6**	58.3	42.4	22.6	43.8	51.5
Faster R-CNN w FPN	ResNet-101	39.5	61.2	43.1	22.7	43.7	50.8
Grid R-CNN w FPN	ResNet-101	**41.3**	60.3	44.4	23.4	45.8	54.1
Grid R-CNN w FPN (ours)	ResNeXt-101		43.2	63.0	46.6	25.1	46.5	55.2

http://openaccess.thecvf.com/content_CVPR_2019/html/Lu_Grid_R-CNN_CVPR_2019_paper.html

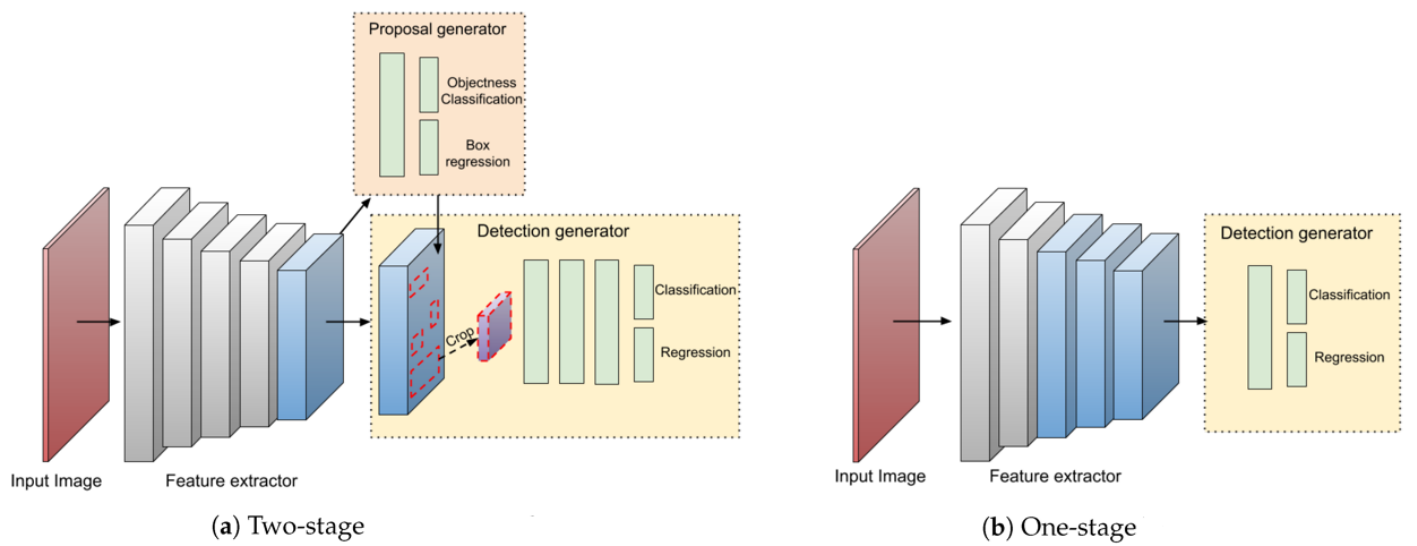
Anchor-based

One-Stage

Life of 计算机图像实体检测

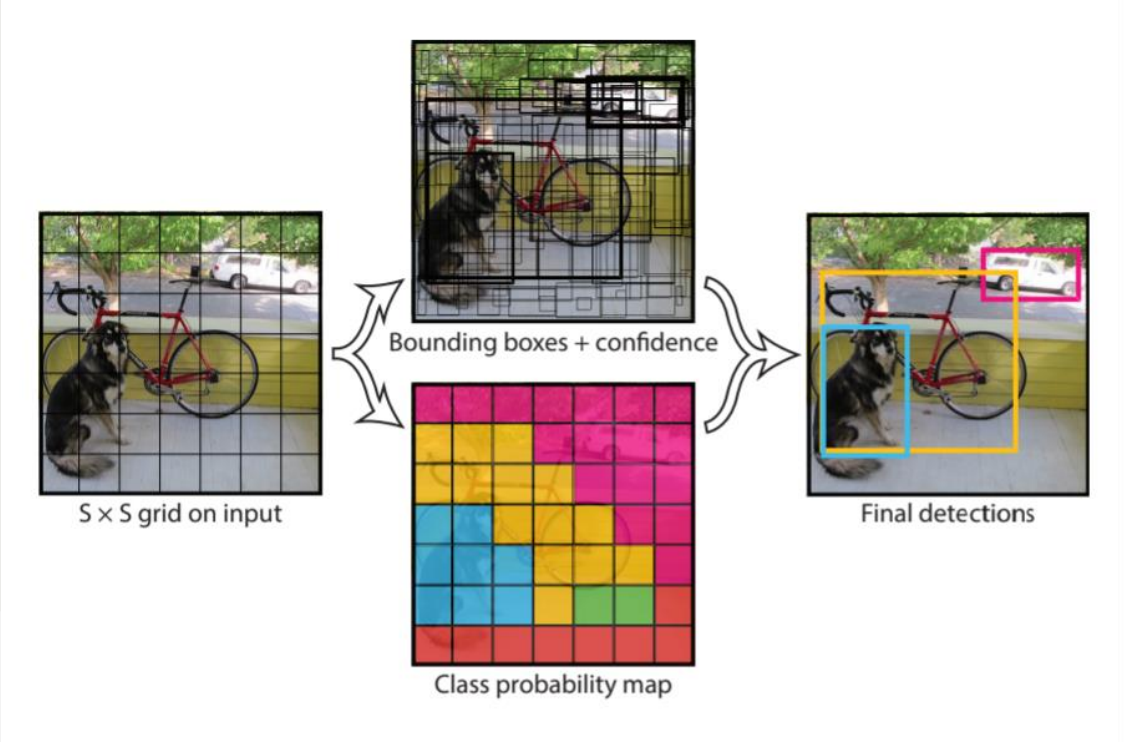
One Stage

2016	2017	2018	2020	2021
YOLOv1 SSD	YOLOv2	RetinaNet YOLOv3	YOLOv4	YOLOv5



Life of 计算机图像实体检测

YOU ONLY LOOK ONCE! 2016



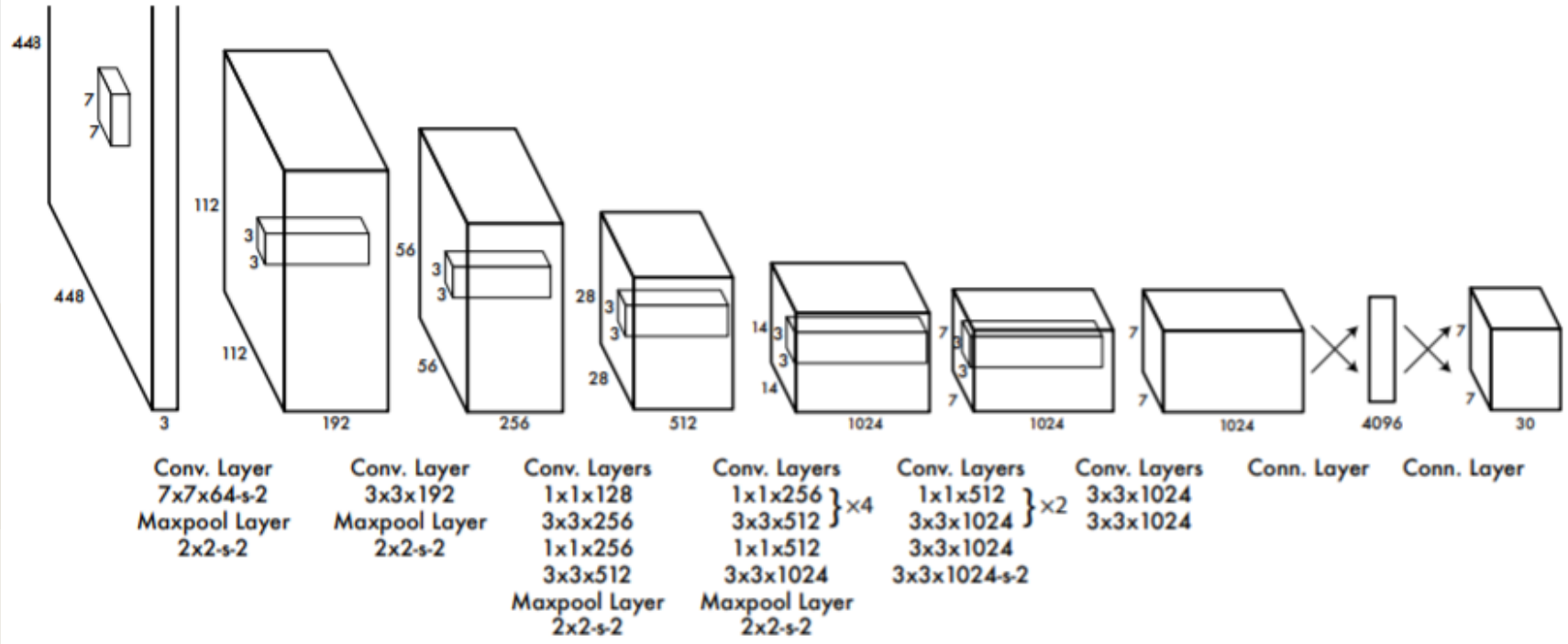
Grid Detection

Anchor-free?

https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Life of 计算机图像实体检测

YOU ONLY LOOK ONCE! 2016



https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Life of 计算机图像实体检测

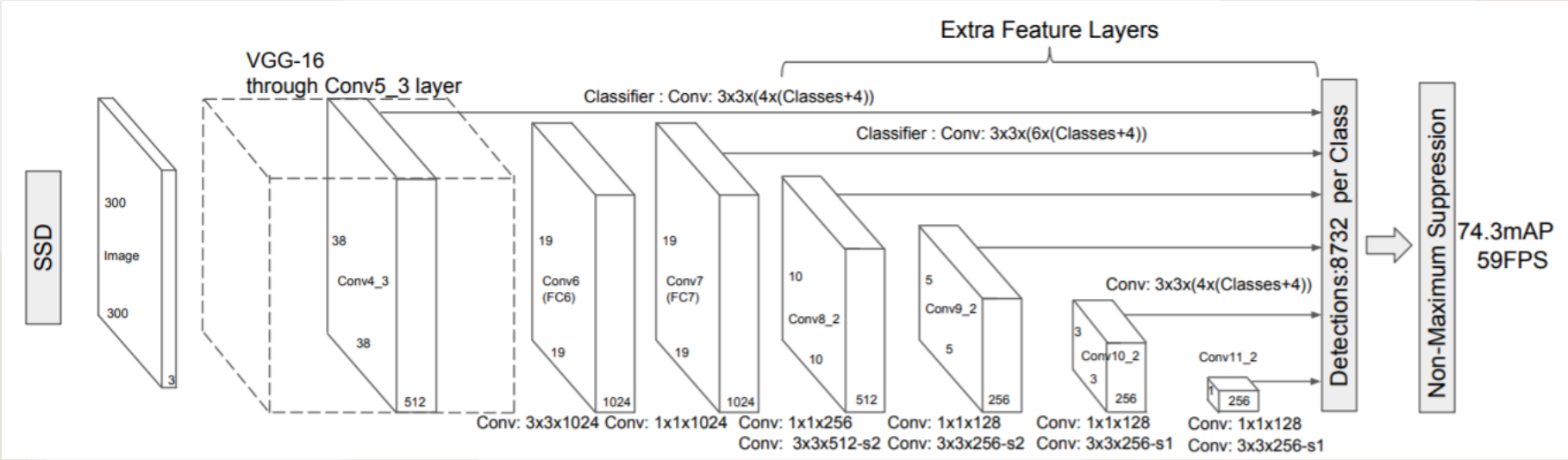
YOU ONLY LOOK ONCE! 2016

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	**155**
YOLO	2007+2012	**63.4**	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Life of 计算机图像实体检测

SSD: Single Shot MultiBox Detector 2016



https://arxiv.org/pdf/1512.02325

Life of 计算机图像实体检测

SSD: Single Shot MultiBox Detector 2016

Method	mAP	FPS	batch size	# Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	1	~ 6000	~ 1000 × 600
Fast YOLO	52.7	155	1	98	448 × 448
YOLO (VGG16)	66.4	21	1	98	448 × 448
SSD300	74.3	46	1	8732	300 × 300
SSD512	76.8	19	1	24564	512 × 512
SSD300	74.3	59	8	8732	300 × 300
SSD512	76.8	22	8	24564	512 × 512

https://arxiv.org/pdf/1512.02325

Life of 计算机图像实体检测

YOLO9000: Better, Faster, Stronger 2017

batch r
hi-res class
convoluti
anchor b
new netv
dimension p
location predi
passthr
multi-s
hi-res det
VOC2007

Detection Frameworks	Train	mAP	FPS
Fast R-CNN [5]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[15]	2007+2012	73.2	7
Faster R-CNN ResNet[6]	2007+2012	76.4	5
YOLO [14]	2007+2012	63.4	45
SSD300 [11]	2007+2012	74.3	46
SSD500 [11]	2007+2012	76.8	19
YOLOv2 288 × 288	2007+2012	69.0	91
YOLOv2 352 × 352	2007+2012	73.7	81
YOLOv2 416 × 416	2007+2012	76.8	67
YOLOv2 480 × 480	2007+2012	77.8	59
YOLOv2 544 × 544	2007+2012	**78.6**	40

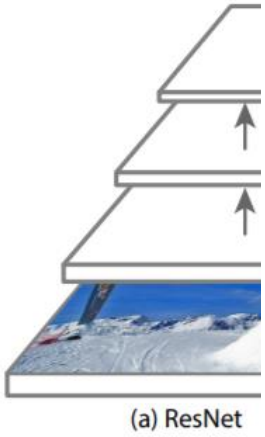
Hi-resclassifier...

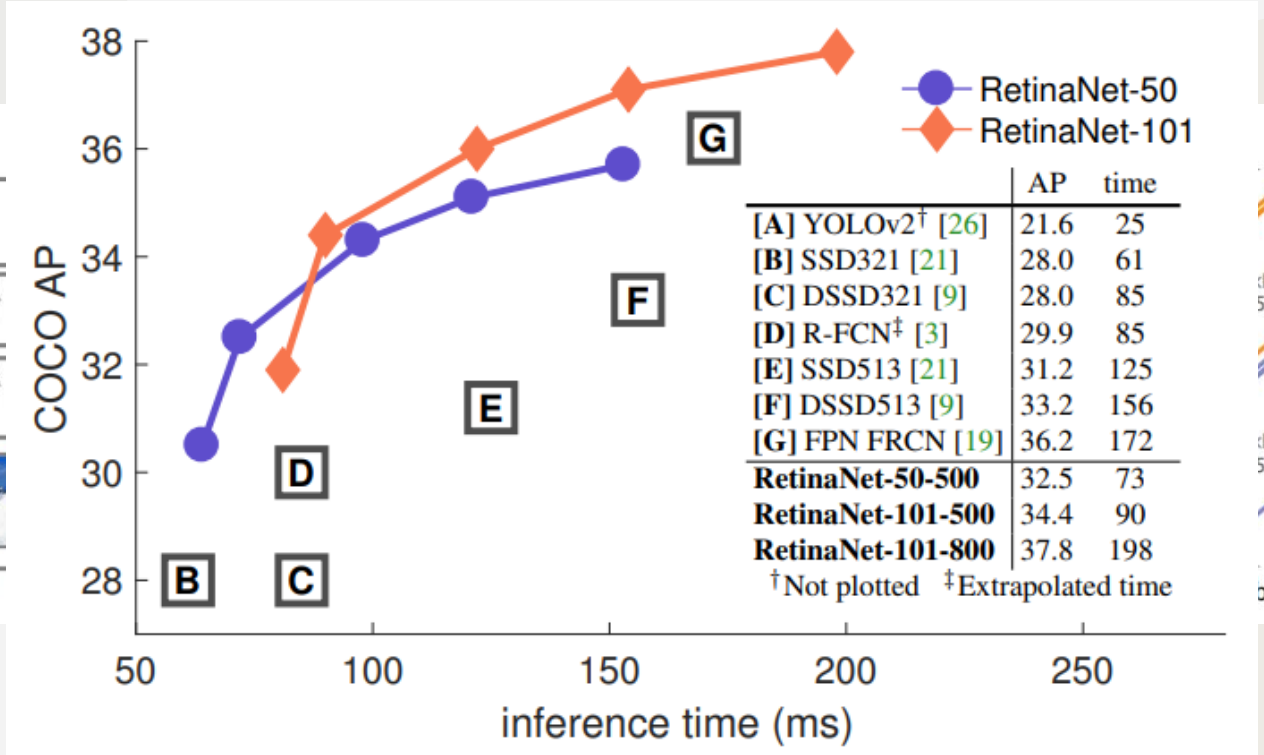
Darknet19

http://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html

Life of 计算机图像实体检测

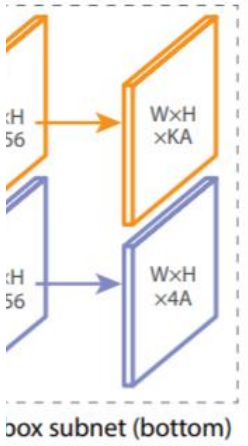
Focal Loss for Dense Object Detection 2018





	AP	time
[A] YOLOv2† [26]	21.6	25
[B] SSD321 [21]	28.0	61
[C] DSSD321 [9]	28.0	85
[D] R-FCN‡ [3]	29.9	85
[E] SSD513 [21]	31.2	125
[F] DSSD513 [9]	33.2	156
[G] FPN FRCN [19]	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	37.8	198

†Not plotted ‡Extrapolated time



http://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

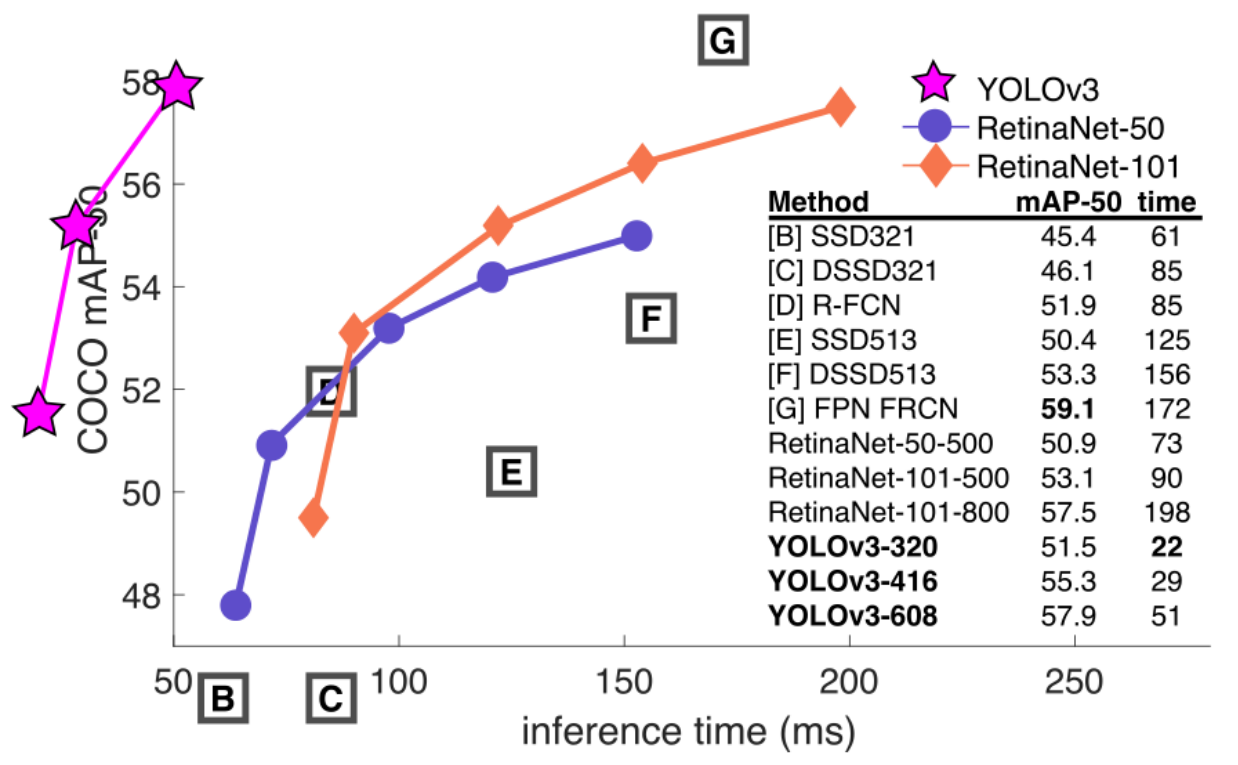
Life of 计算机图像实体检测

YOLOv3: An Incremental Improvement 2018

Darknet53

多尺度

忽略样本



Method	mAP-50	time
[B] SSD321	45.4	61
[C] DSSD321	46.1	85
[D] R-FCN	51.9	85
[E] SSD513	50.4	125
[F] DSSD513	53.3	156
[G] FPN FRCN	**59.1**	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	**22**
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

https://arxiv.org/abs/1804.02767

Life of 计算机图像实体检测

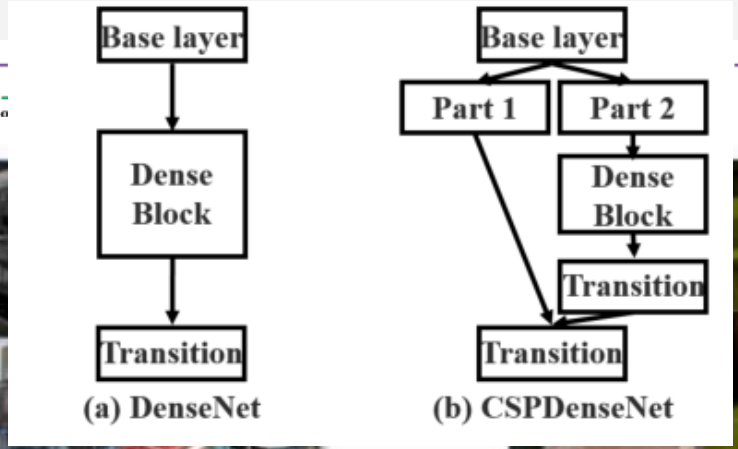
YOLOv4: Optimal Speed and Accuracy of Object Detection 2020

Mosaic数据增强

CSPDarknet53

Bag-of-Freebies

Bag-of-Special

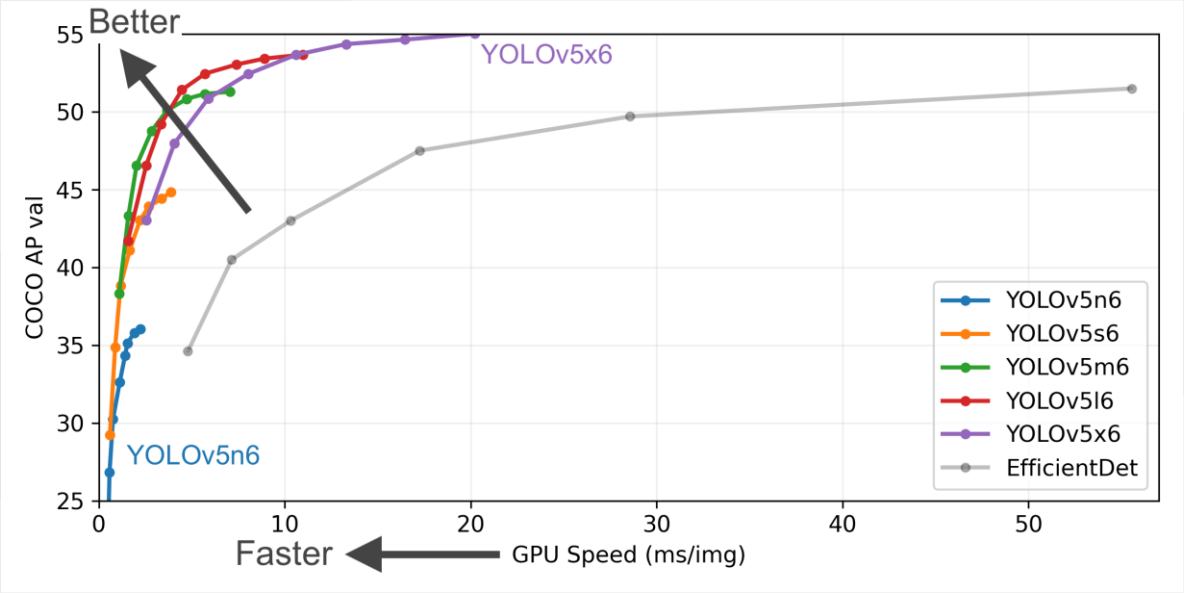


Backbone	Top-1	Top-5	Bn Ops	BFLOP/s	FPS
Darknet-19 [15]	74.1	91.8	7.29	1246	**171**
ResNet-101[5]	77.1	93.7	19.7	1039	53
ResNet-152 [5]	**77.6**	**93.8**	29.4	1090	37
Darknet-53	77.2	**93.8**	18.7	**1457**	78

https://arxiv.org/abs/2004.10934

Life of 计算机图像实体检测

YOLOv5



Mosaic数据增强

自适应锚框计算

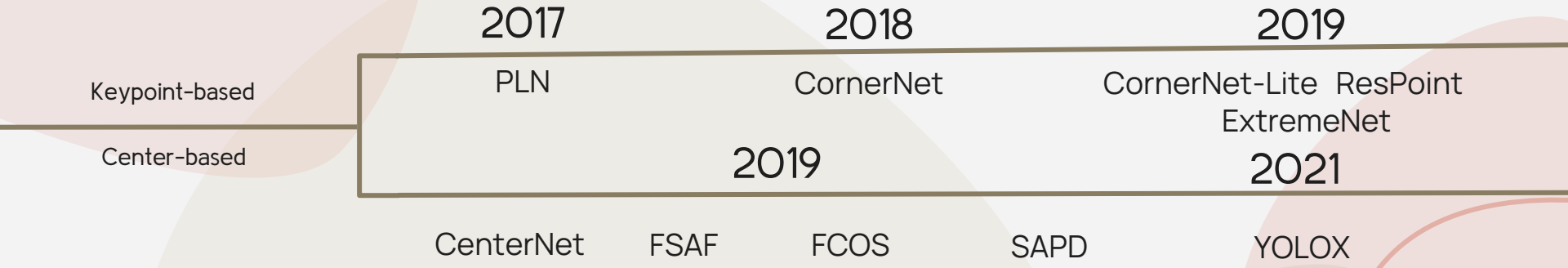
自适应图片缩放

Focus结构

CSP结构.....

https://github.com/ultralytics/yolov5

Anchor-free


2017
2018
2019
Keypoint-based
PLN
CornerNet
CornerNet-Lite
ResPoint
ExtremeNet
Center-based
2019
2021
CenterNet
FSAF
FCOS
SAPD
YOLOX

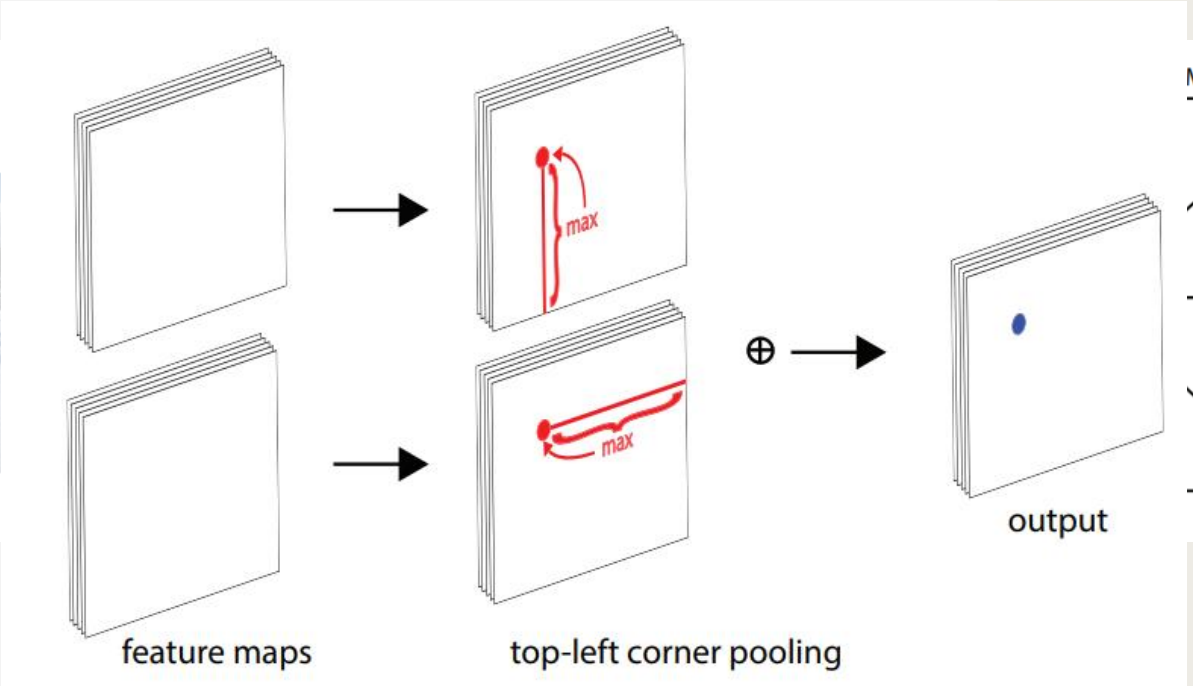
Anchor-free

Keypoint-based

Life of 计算机图像实体检测

Anchor-free Keypoint-based

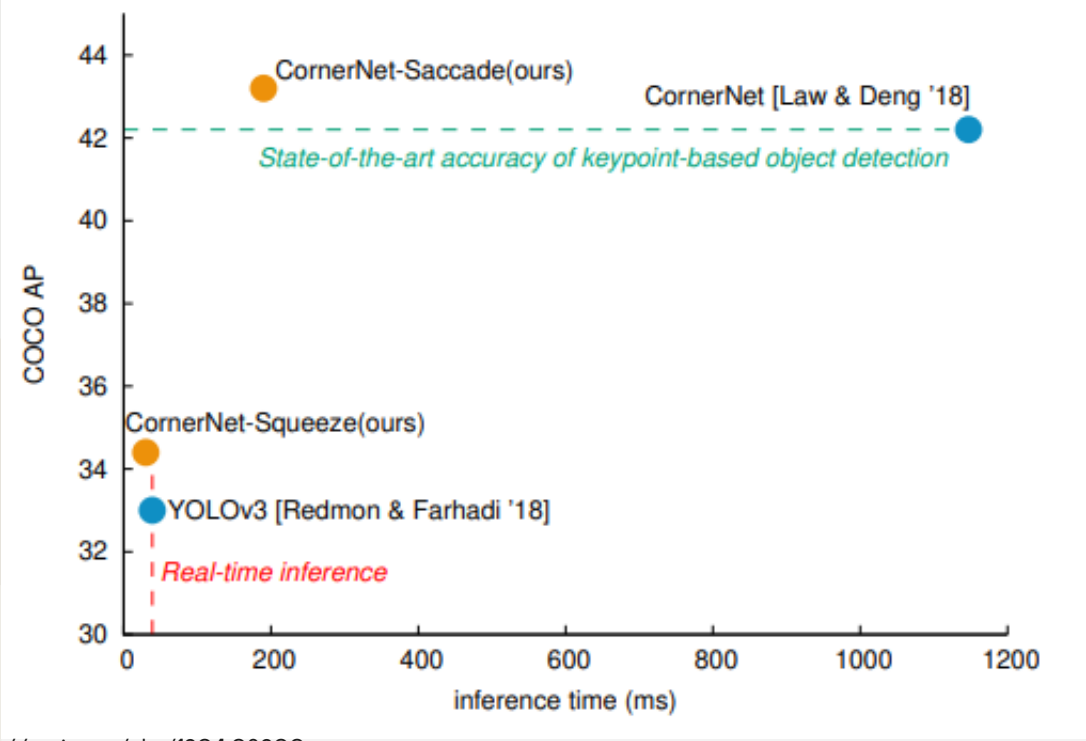
CornerNet: Detecting Objects as Paired Keypoints 2018



http://openaccess.thecvf.com/content_ECCV_2018/html/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.html

Life of 计算机图像实体检测

CornerNet-Lite: Efficient Keypoint Based Object Detection 2019



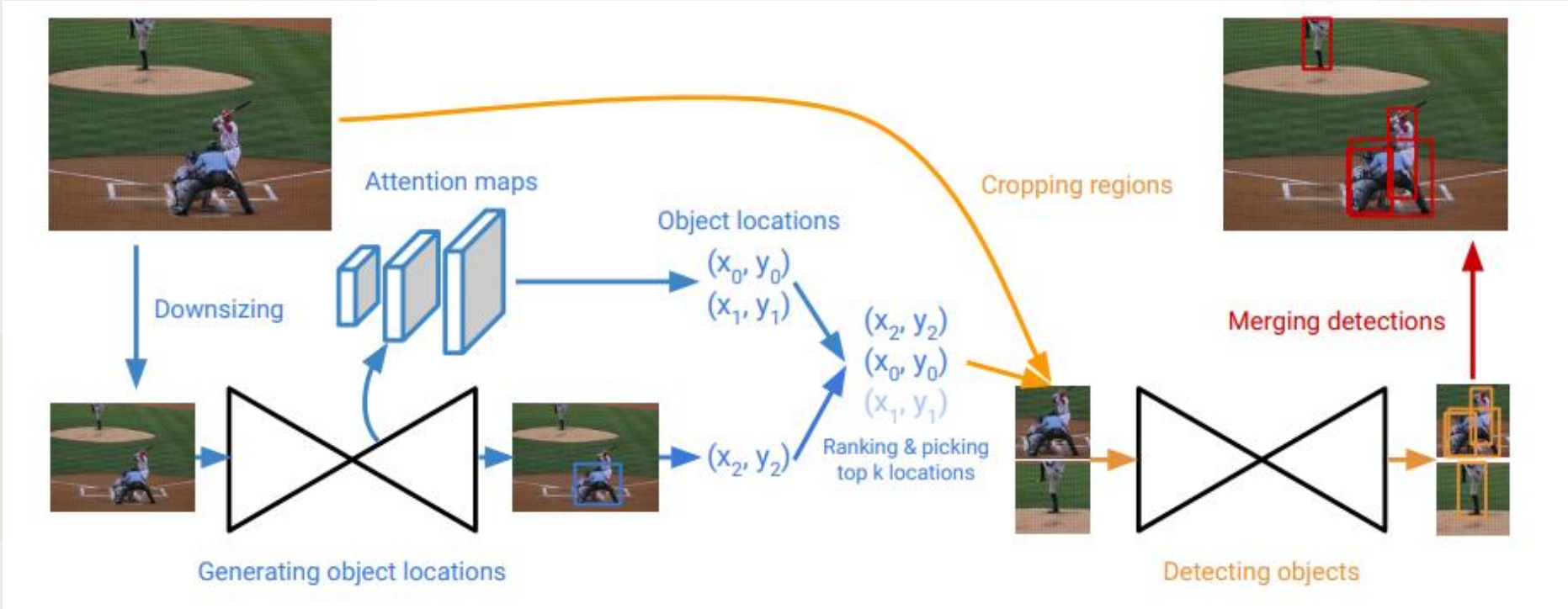
CornerNet-Saccade

CornerNet-Squeeze

https://arxiv.org/abs/1904.08900

Life of 计算机图像实体检测

CornerNet-Saccade



https://arxiv.org/abs/1904.08900

Life of 计算机图像实体检测

CornerNet-Squeeze

SqueezeNet

MobileNet

Input	Operator	Output
Residual block in CornerNet		
$h \times w \times k$	3×3 Conv, ReLU	$h \times w \times k'$
$h \times w \times k'$	3×3 Conv, ReLU	$h \times w \times k'$
Fire module in CornerNet-Squeeze		
$h \times w \times k$	1×1 Conv	$h \times w \times \frac{k'}{2}$
$h \times w \times \frac{k'}{2}$	1×1 Conv + 3×3 Dwise, ReLU	$h \times w \times k'$

Comparison between the residual block and the new fire module.

https://arxiv.org/abs/1904.08900

Life of 计算机图像实体检测

Anchor-free Keypoint-based

Point

Botto

Cent

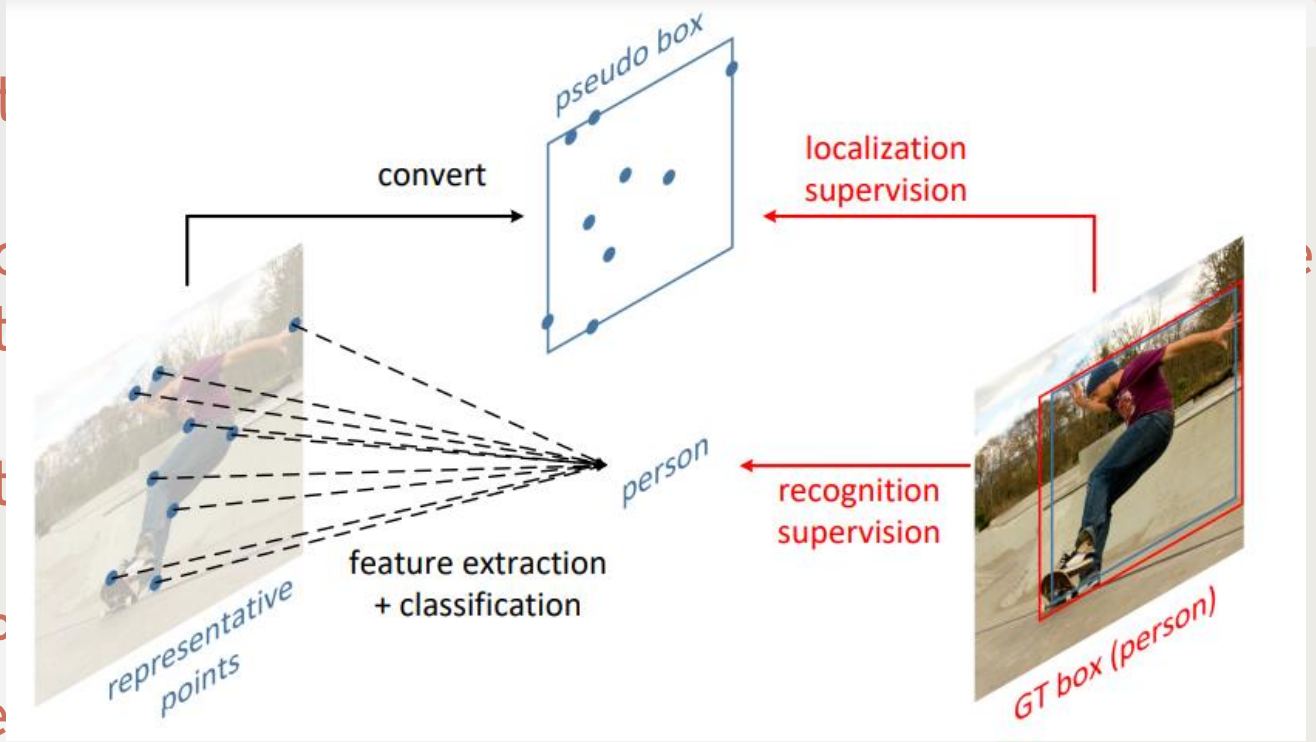
Cent

RepP

Dete

and

2019



Reppoints: http://openaccess.thecvf.com/content_ICCV_2019/html/Yang_RepPoints_Point_Set_Representation_for_Object_Detection_ICCV_2019_paper.html

Anchor-free

Center-based

Life of 计算机图像实体检测

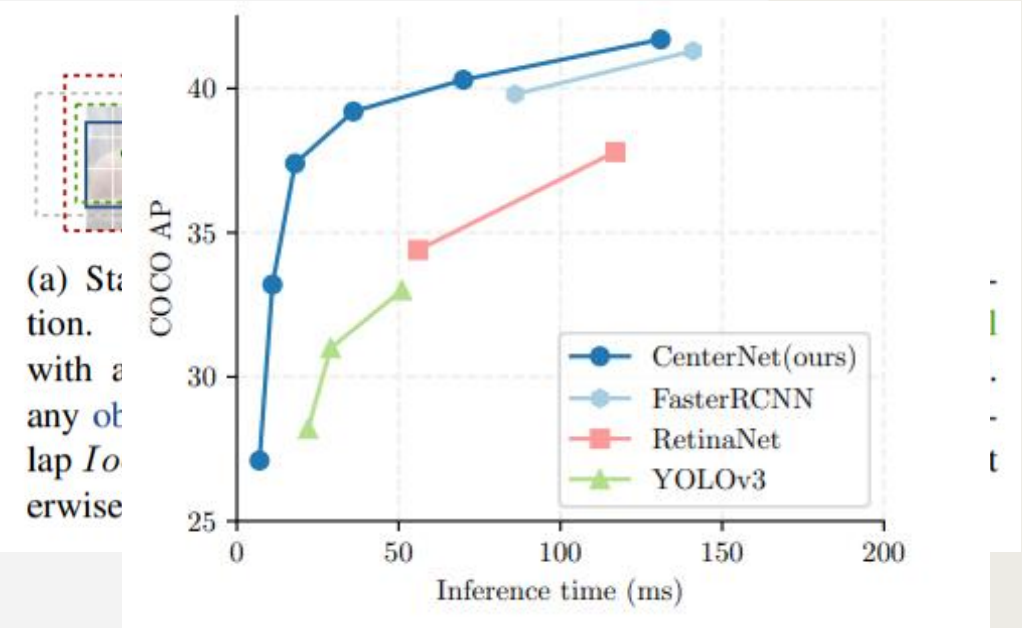
Anchor-free Center-based

CenterNet：Objects as Points 2019

前后景分类 ×

NMS ×

FPN ×

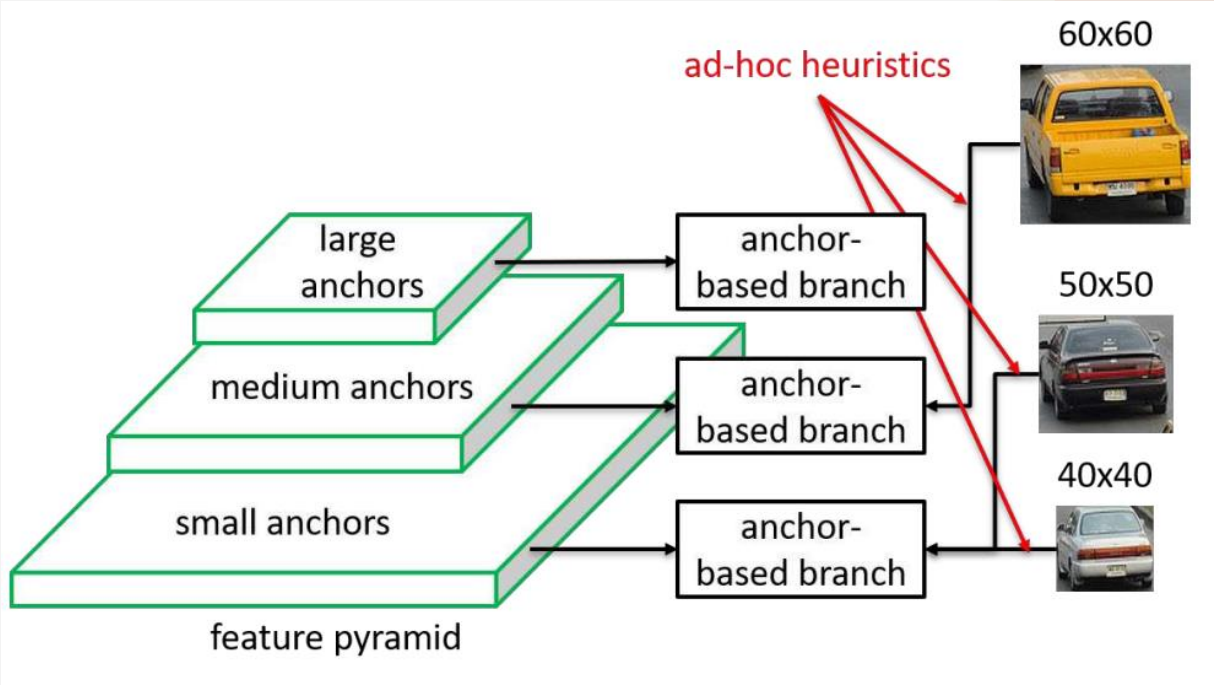


https://arxiv.org/abs/1904.07850

Life of 计算机图像实体检测

Feature Selective Anchor-Free Module for Single-Shot Object Detection 2019

人为设定的特征水平？

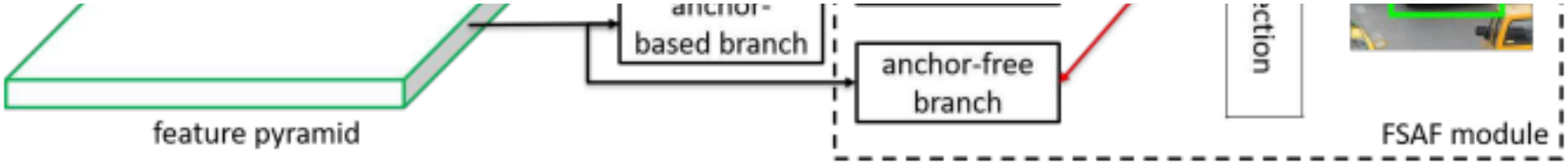


http://openaccess.thecvf.com/content_CVPR_2019/html/Zhu_Feature_Selective_Anchor-Free_Module_for_Single-Shot_Object_Detection_CVPR_2019_paper.html

Life of 计算机图像实体检测

Feature Selective Anchor-Free Module for Single-Shot Object Detection 2019

CornerNet511 [17] (single-scale)	Hourglass-104	anchor-	40.5	56.5	43.1	19.4	42.7	53.9
CornerNet [17] (multi-scale)			42.1	57.8	45.3	20.8	44.8	56.7
GHM800 [18]	ResNeXt-101		41.6	62.8	44.2	22.3	45.1	**55.3**
Ours800 (single-scale)			42.9	63.8	46.3	26.6	46.2	52.7
Ours (multi-scale)			**44.6**	**65.2**	**48.6**	**29.7**	**47.1**	54.6



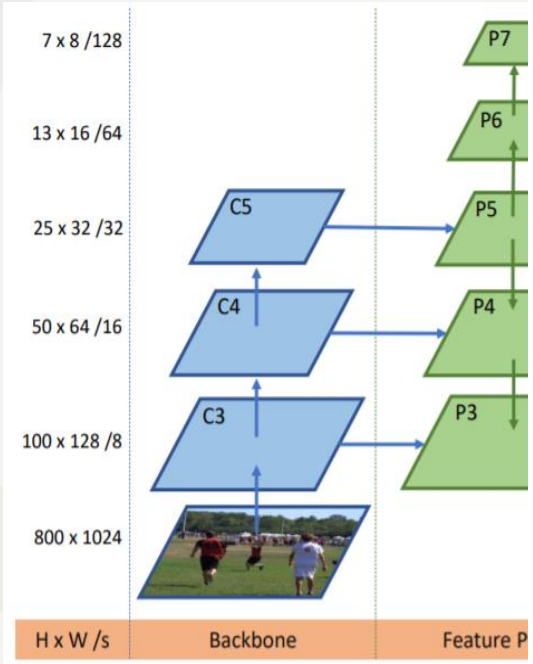
http://openaccess.thecvf.com/content_CVPR_2019/html/Zhu_Feature_Selective_Anchor-Free_Module_for_Single-Shot_Object_Detection_CVPR_2019_paper.html

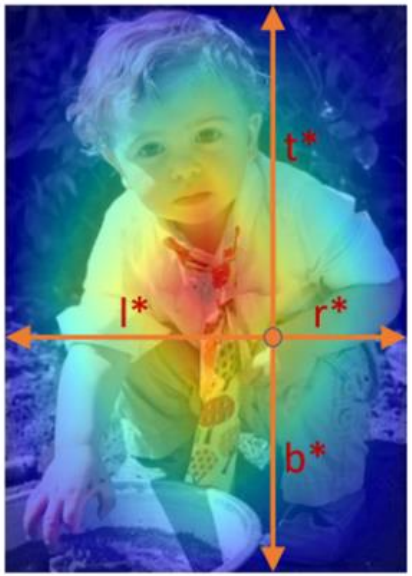
Life of 计算机图像实体检测

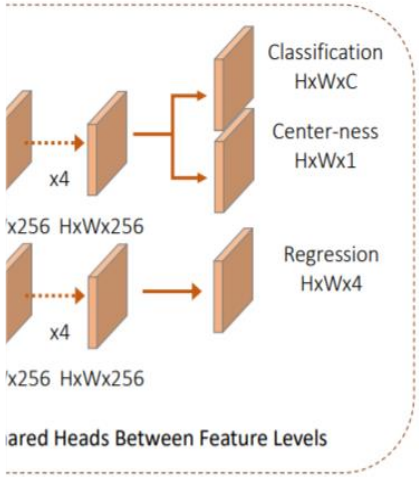
FCOS: Fully Convolutional One-Stage Object Detection 2019

逐像素回归预测

center-ness







$$\text{centerness}^* = \sqrt{\frac{\min(l^*, r^*)}{\max(l^*, r^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)}}.$$

http://openaccess.thecvf.com/content_ICCV_2019/html/Tian_FCOS_Fully_Convolutional_One-Stage_Object_Detection_ICCV_2019_paper.html

Life of 计算机图像实体检测

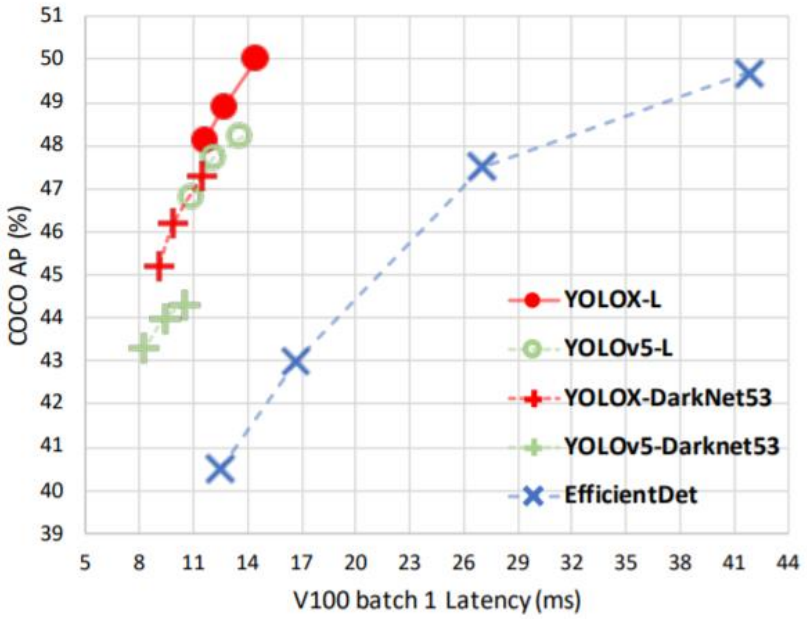
FCOS: Fully Convolutional One-Stage Object Detection 2019

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods:							
Faster R-CNN w/ FPN [14]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [11]	Inception-ResNet-v2 [27]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w/ TDM [25]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
One-stage methods:							
YOLOv2 [22]	DarkNet-19 [22]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [18]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [5]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [15]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
CornerNet [13]	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9
FSAF [34]	ResNeXt-64x4d-101-FPN	42.9	63.8	46.3	26.6	46.2	52.7
FCOS	ResNet-101-FPN	41.5	60.7	45.0	24.4	44.8	51.6
FCOS	HRNet-W32-5l [26]	42.0	60.4	45.3	25.4	45.0	51.0
FCOS	ResNeXt-32x8d-101-FPN	42.7	62.2	46.1	26.0	45.6	52.6
FCOS	ResNeXt-64x4d-101-FPN	43.2	62.8	46.6	26.5	46.2	53.3
FCOS w/ improvements	ResNeXt-64x4d-101-FPN	**44.7**	**64.1**	**48.4**	**27.6**	**47.5**	**55.6**

http://openaccess.thecvf.com/content_ICCV_2019/html/Tian_FCOS_Fully_Convolutional_One-Stage_Object_Detection_ICCV_2019_paper.html

Life of 计算机图像实体检测

YOLOX: Exceeding YOLO Series in 2021



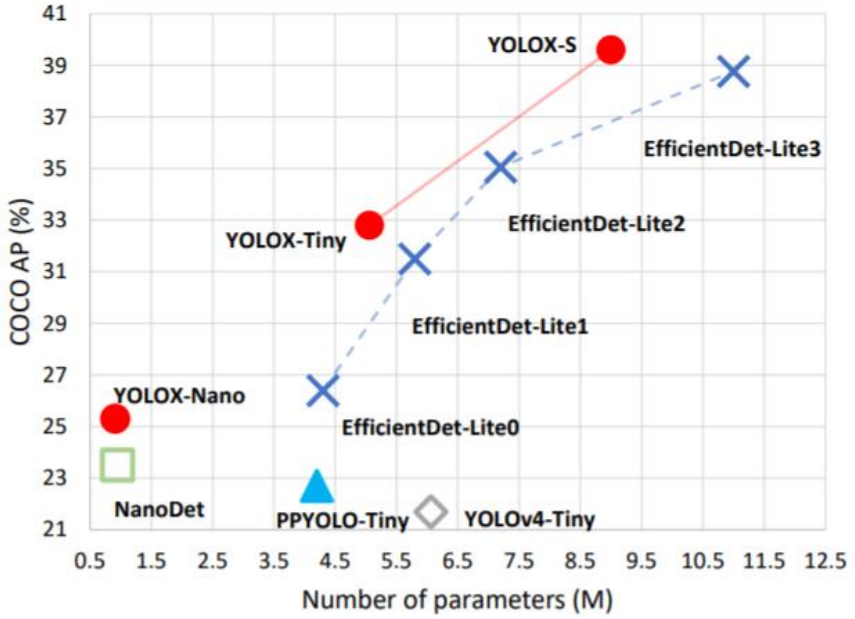


Figure 1: Speed-accuracy trade-off of accurate models (top) and Size-accuracy curve of lite models on mobile devices (bottom) for YOLOX and other state-of-the-art object detectors.

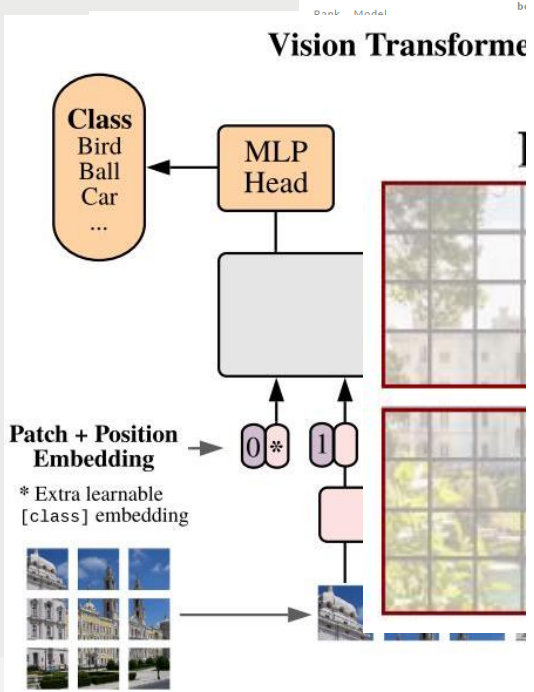
https://arxiv.org/abs/2107.08430

新进展

Transformer

Life of 计算机图像实体检测

新进展——Vision Transformer、Swin Transformer



(a) Various frameworks

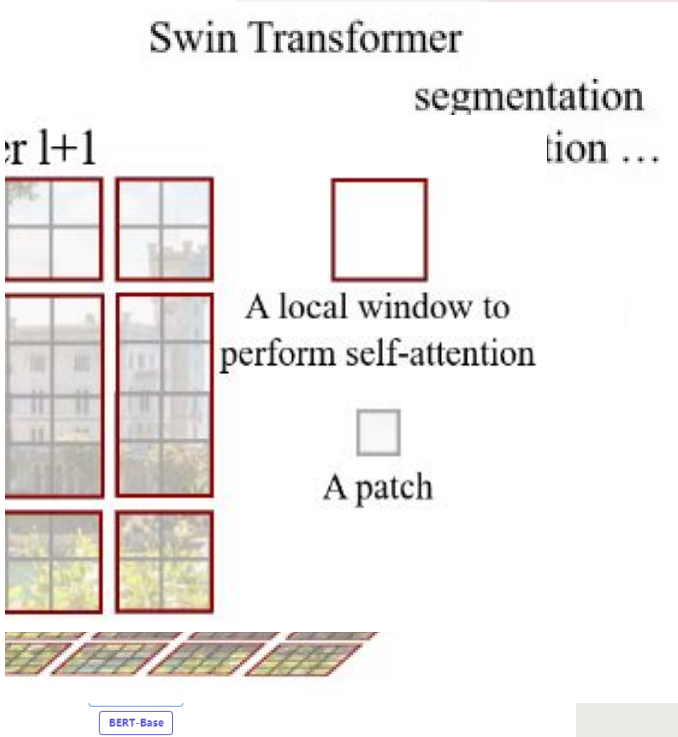
Method	Backbone	AP^{box}	AP^{box}_{50}	AP^{box}_{75}	#param.	FLOPs	FPS
Cascade Mask R-CNN	R-50	46.3	64.3	50.5	82M	739G	18.0
	Swin-T	**50.5**	**69.3**	**54.9**	86M	745G	15.3
ATSS	R-50	43.5	61.9	47.0	32M	205G	28.3
	Swin-T	**47.2**	**66.5**	**51.3**	36M	215G	22.3
RepPointsV2	R-50	46.5	64.6	50.3	42M	274G	13.6
	Swin-T	**50.0**	**68.5**	**54.2**	45M	283G	12.0
Sparse R-CNN	R-50	44.5	63.4	48.2	106M	166G	21.0
	Swin-T	**47.9**	**67.3**	**52.3**	110M	172G	18.4

(b) Various backbones w. Cascade Mask R-CNN

	AP^{box}	AP^{box}_{50}	AP^{box}_{75}	AP^{mask}	AP^{mask}_{50}	AP^{mask}_{75}	param	FLOPs	FPS
DeiT-S†	48.0	67.2	51.7	41.4	64.2	44.3	80M	889G	10.4
R50	46.3	64.3	50.5	40.1	61.7	43.4	82M	739G	18.0
Swin-T	**50.5**	**69.3**	**54.9**	**43.7**	**66.6**	**47.1**	86M	745G	15.3
X101-32	48.1	66.5	52.4	41.6	63.9	45.2	101M	819G	12.8
Swin-S	**51.8**	**70.4**	**56.3**	**44.7**	**67.9**	**48.5**	107M	838G	12.0
X101-64	48.3	66.4	52.3	41.7	64.0	45.1	140M	972G	10.4
Swin-B	**51.9**	**70.9**	**56.5**	**45.0**	**68.4**	**48.7**	145M	982G	11.6

(c) System-level Comparison

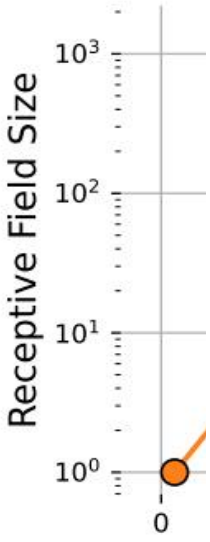
Method	mini-val		test-dev		#param.	FLOPs
	AP^{box}	AP^{mask}	AP^{box}	AP^{mask}		
RepPointsV2* [12]	-	-	52.1	-	-	-
GCNet* [7]	51.8	44.7	52.3	45.4	-	1041G
RelationNet++* [13]	-	-	52.7	-	-	-
DetectoRS* [42]	-	-	55.7	48.5	-	-
YOLOv4 P7* [4]	-	-	55.8	-	-	-
Copy-paste [23]	55.9	47.2	56.0	47.4	185M	1440G
X101-64 (HTC++)	52.3	46.0	-	-	155M	1033G
Swin-B (HTC++)	56.4	49.1	-	-	160M	1043G
Swin-L (HTC++)	57.1	49.5	57.7	50.2	284M	1470G
Swin-L (HTC++)*	**58.0**	**50.4**	**58.7**	**51.1**	284M	-



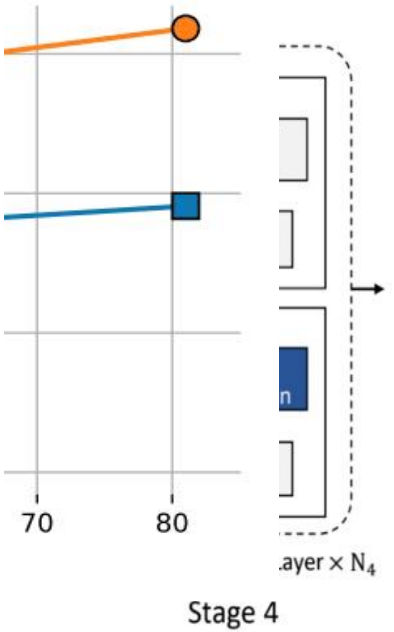
https://openaccess.thecvf.com/content/ICCV2021/html/Arnab_ViViT_A_Video_Vision_Transformer_ICCV_2021_paper.html?ref=https://githubhelp.com
http://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html

Life of 计算机图像实体检测

新进展——Focal Transformer



Backbone	RetinaNet AP^b	Mask R-CNN AP^b	Mask R-CNN AP^m
ResNet-50 [34]	36.3	38.0	34.4
PVT-Small	40.4	40.4	37.8
ViL-Small [80]	41.6	41.8	38.5
Swin-Tiny [44]	42.0	43.7	39.8
Focal-Tiny (Ours)	**43.7** (+1.7)	**44.8** (+1.1)	**41.0** (+1.3)
ResNet-101 [34]	38.5	40.4	36.4
ResNeXt101-32x4d [70]	39.9	41.9	37.5
PVT-Medium [63]	41.9	42.0	39.0
ViL-Medium [80]	42.9	43.4	39.7
Swin-Small [44]	45.0	46.5	42.1
Focal-Small (Ours)	**45.6** (+0.6)	**47.4** (+0.9)	**42.8** (+0.7)
ResNeXt101-64x4d [70]	41.0	42.8	38.4
PVT-Large [63]	42.6	42.9	39.5
ViL-Base [80]	44.3	45.1	41.0
Swin-Base [44]	45.0	46.9	42.3
Focal-Basc (Ours)	**46.3** (+1.3)	**47.8** (+0.9)	**43.2** (+0.9)



https://arxiv.org/abs/2107.00641

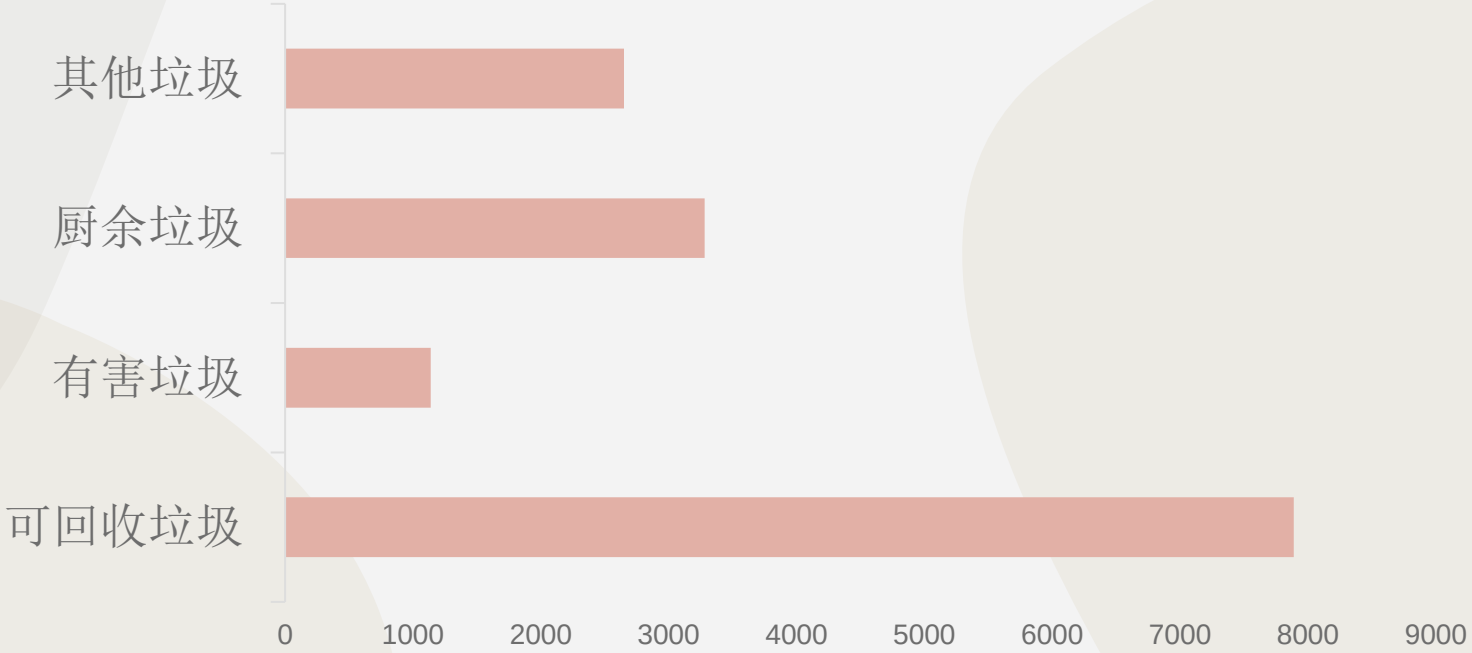
项目展示

垃圾分类检测Demo

Demo of 计算机图像实体检测

The Trash Dataset of Our Demo

四类垃圾分布

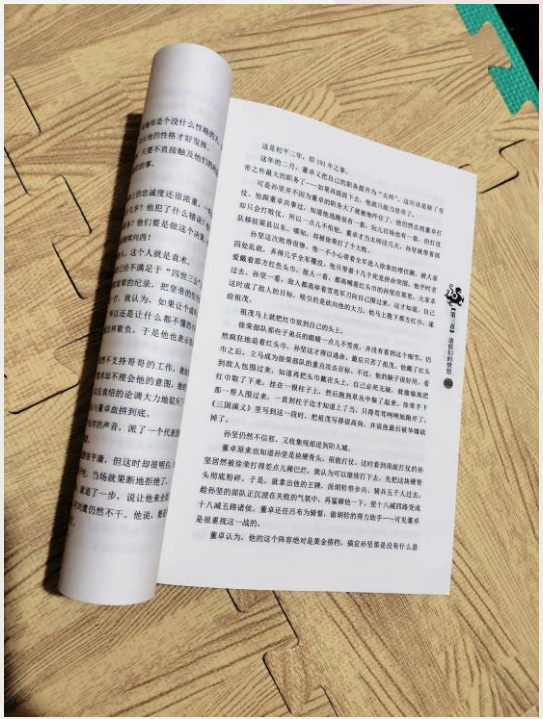


Demo of 计算机图像实体检测

The Trash Dataset of Our Demo

Demo of 计算机图像实体检测

The Trash Dataset of Our Demo

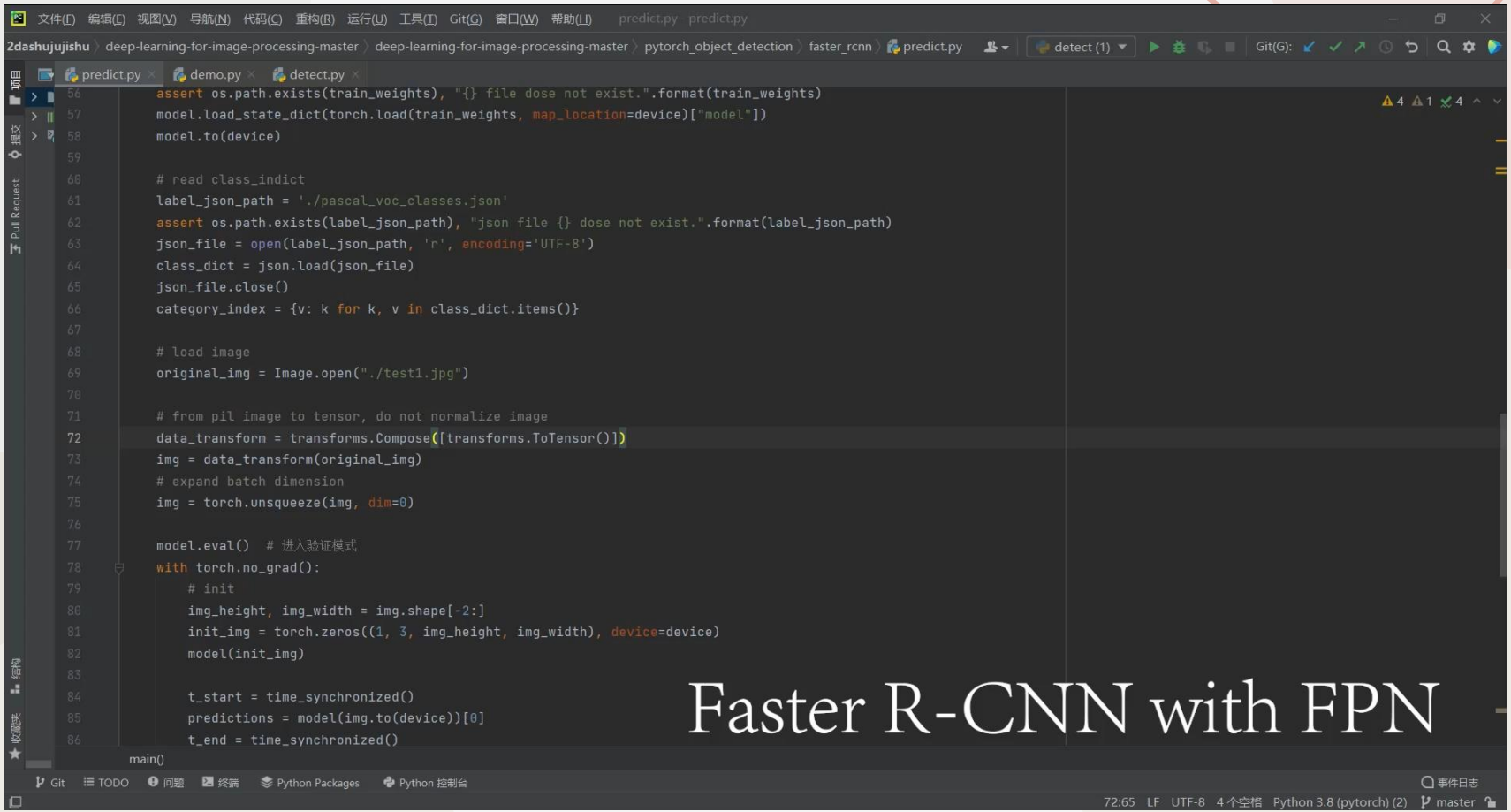


Demo of 计算机图像实体检测

The Models of Our Demo

	Val Datasets	AP(%)
One-Stage YOLOv5-s	COCO	36.7
Two-Stage Faster R-CNN + FPN		35.9
Anchor-free YOLOX-s		39.6

Demo of 计算机图像实体检测



Demo of 计算机图像实体检测

The Models of Our Demo

YOLOX

可回收垃圾63%

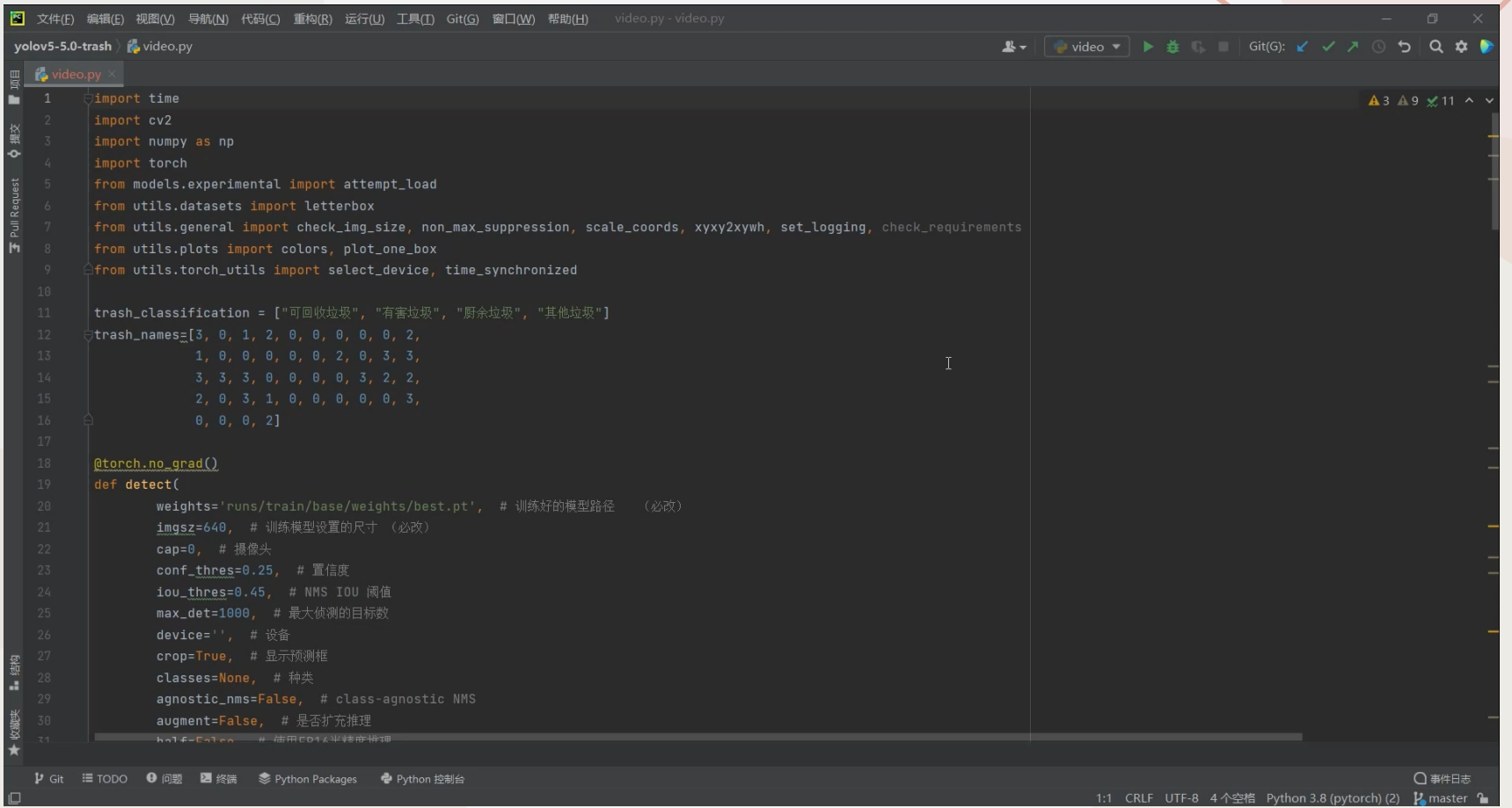
Faster R-CNN

可回收垃圾86.2%

YOLOv5

可回收垃圾73%

Demo of 计算机图像实体检测



Demo of 计算机图像实体检测

The Models of Our Demo

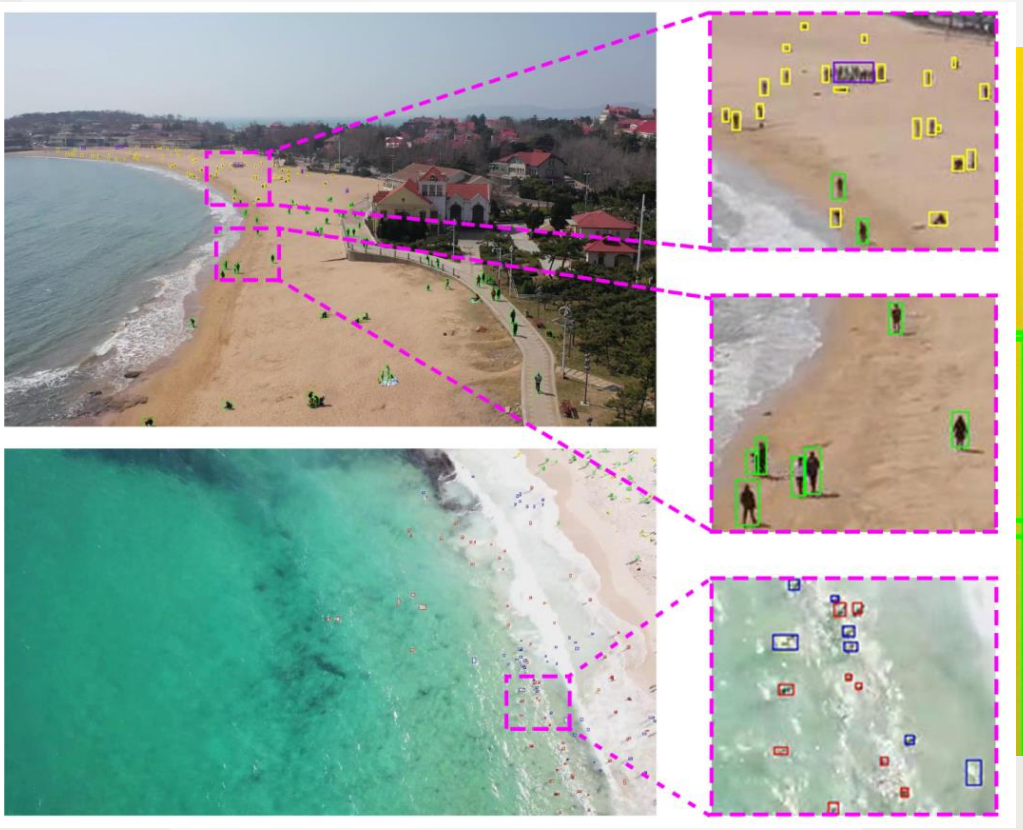
	mAP	Model	Inference Time
One-Stage YOLOv5-s	80.3%	13.9MB	27ms
Two-Stage Faster R-CNN + FPN	53.9%	316.7MB	108ms
Anchor-free YOLOX-s	71.4%	68.6MB	205ms

Tested on RTX1080

04 未来展望

Future

Challenge of 计算机图像实体检测



· 实体多样

光照、视角、类间 …

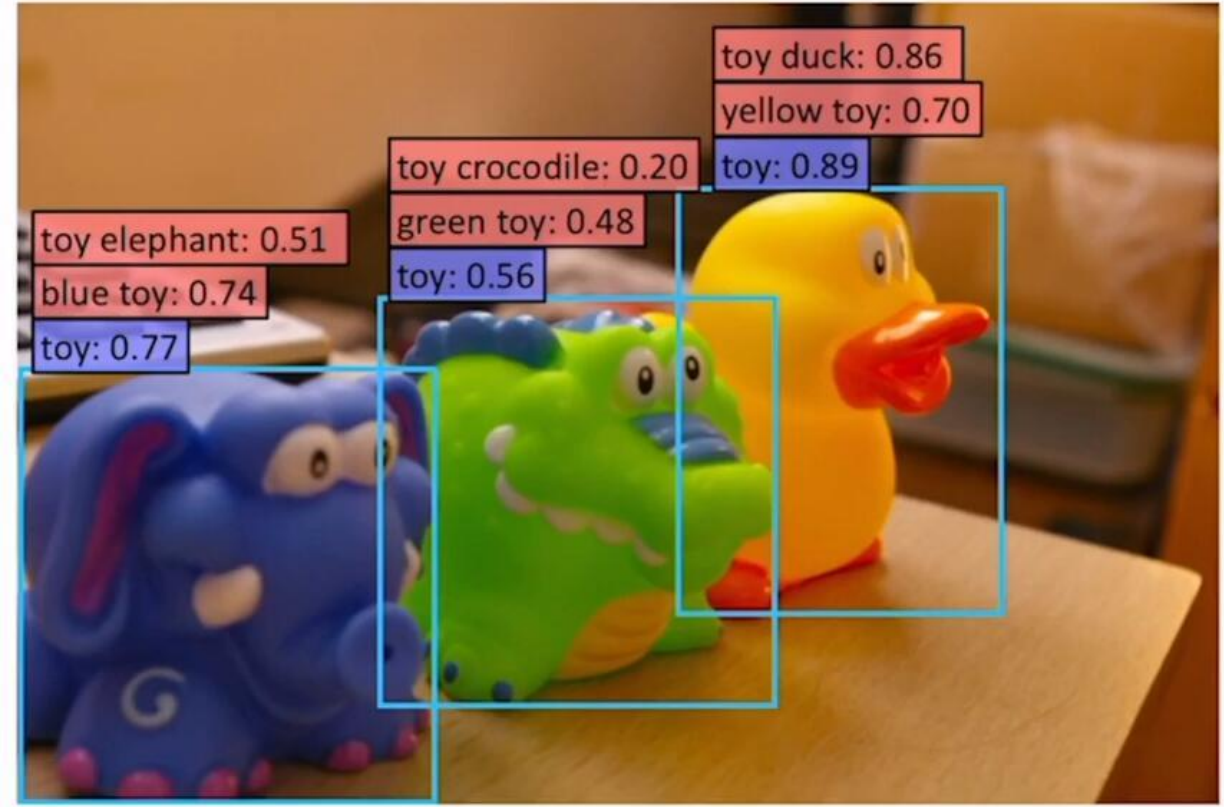
实体

湖图像

Our Demo1 垃圾分类 By YOLO

Future of 计算机图像实体检测

- 精
- 自
- 弱



https://arxiv.org/abs/2103.00020

THANKS!