

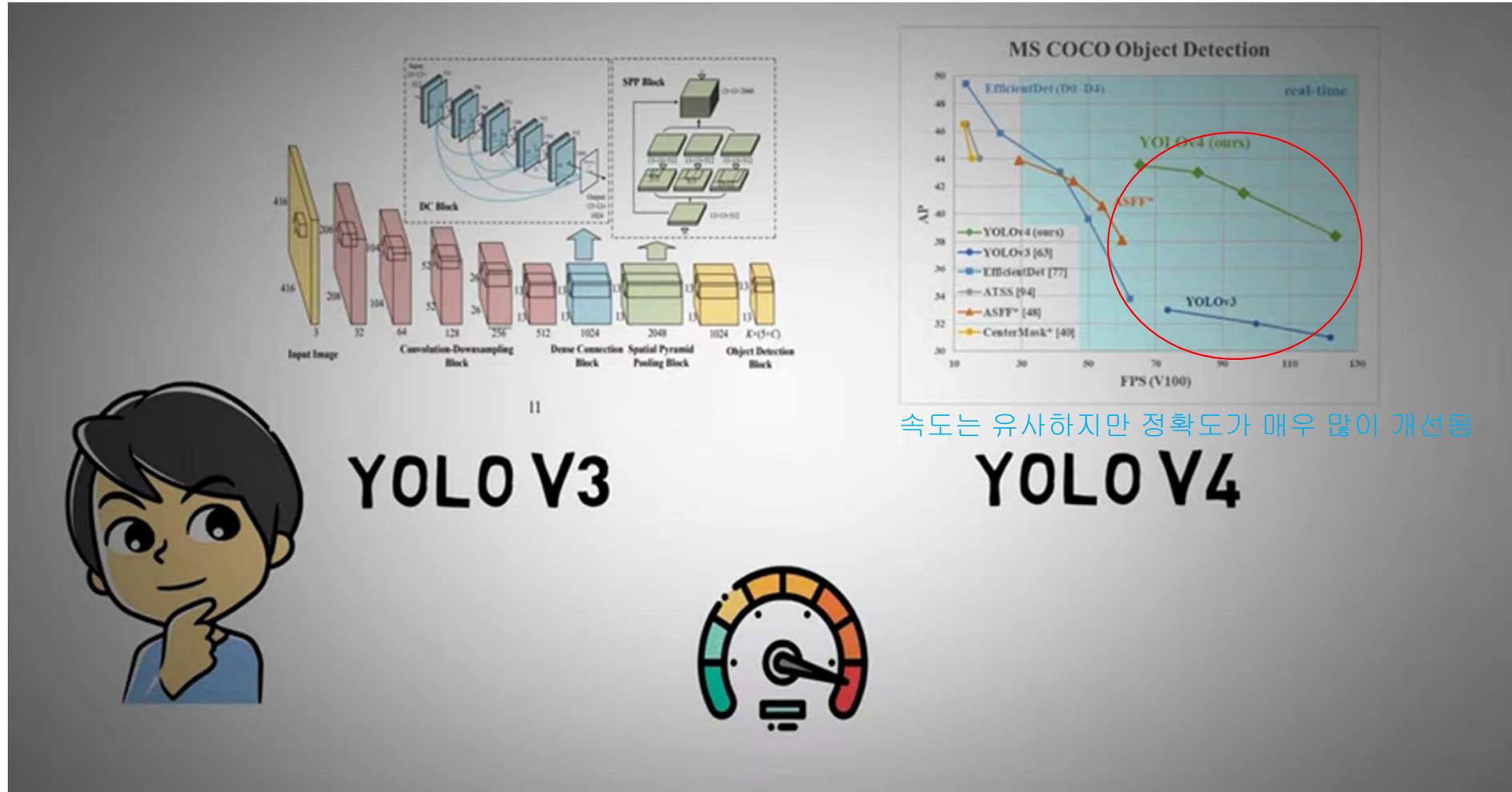
YOLO V4

Optimal Speed And Accuracy of Object Detection

박 철

Introduction to yolo v4 object detection

YOLOv4는 YOLOv3 이후에 나온 딥러닝의 정확도를 개선하는 다양한 방법을 적용해 YOLO의 성능을 극대화 하는 방법을 설명함



the paper YOLOv4

arXivv2004.10934v1 [cs.CV] 23 Apr 2020

YOLOv4: OPTIMAL SPEED AND ACCURACY OF OBJECT DETECTION BY



YOLOv4: Optimal Speed and Accuracy of Object Detection

Alexey Bochkovsky*
alexeyab4@gmail.com

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

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

Abstract

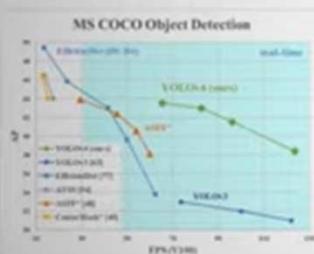
There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the results, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, DropBlock regularization, and ChU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (65.7% AP₅₀) for the MS COCO dataset at a real-time speed of ~65 FPS on Tesla V100. Source code is at <https://github.com/AlexeyAB/darknet>.

1. Introduction

The majority of CNN-based object detectors are largely applicable only for recommendation systems. For example, searching for free parking spaces via urban video cameras is executed by view accurate models, whereas car collision avoidance is solved by deep reinforcement methods. However,

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.

MS COCO Object Detection



Model	AP (mAP)	FPS
YOLOv4	~0.43	~65
EfficientDet-D4	~0.38	~100
YOLOv3	~0.35	~100
EfficientDet-D0	~0.32	~100
YOLOv2	~0.30	~100
YOLOv1	~0.28	~100
SSD300	~0.25	~100
SSD512	~0.22	~100
RetinaNet	~0.20	~100
FCOS	~0.18	~100
ATSS	~0.15	~100
SSD300	~0.12	~100
CenterNet	~0.10	~100

ALEXEY BOCHKOVSKY CHIEN YAO HON-YUAN

Joseph Redmon and Ali Farhadi

original creators of Yolo V1-3



JOSEPH REDMON AND ALI FARHADI



in Feb 2020 that Joseph stopped Computer vision research

Ali Farhadi went on to found a company called xnor.ai which specialized in on edge-centric AI.

 Joe Redmon @pjreddie · Feb 20, 2020

"We shouldn't have to think about the societal impact of our work because it's hard and other people can do it for us" is a really bad argument. twitter.com/RogerGrosse/st...

 Roger Grosse @RogerGrosse
Replying to @kevin_zakka @hardmaru

To be clear, I don't think this is a positive step. Societal impacts of AI is a tough field, and there are researchers and organizations that study it professionally. Most authors do not have expertise in the area and won't do good enough scholarship to say something meaningful.

 Joe Redmon
@pjreddie

I stopped doing CV research because I saw the impact my work was having. I loved the work but the military applications and privacy concerns eventually became impossible to ignore.
twitter.com/RogerGrosse/st...

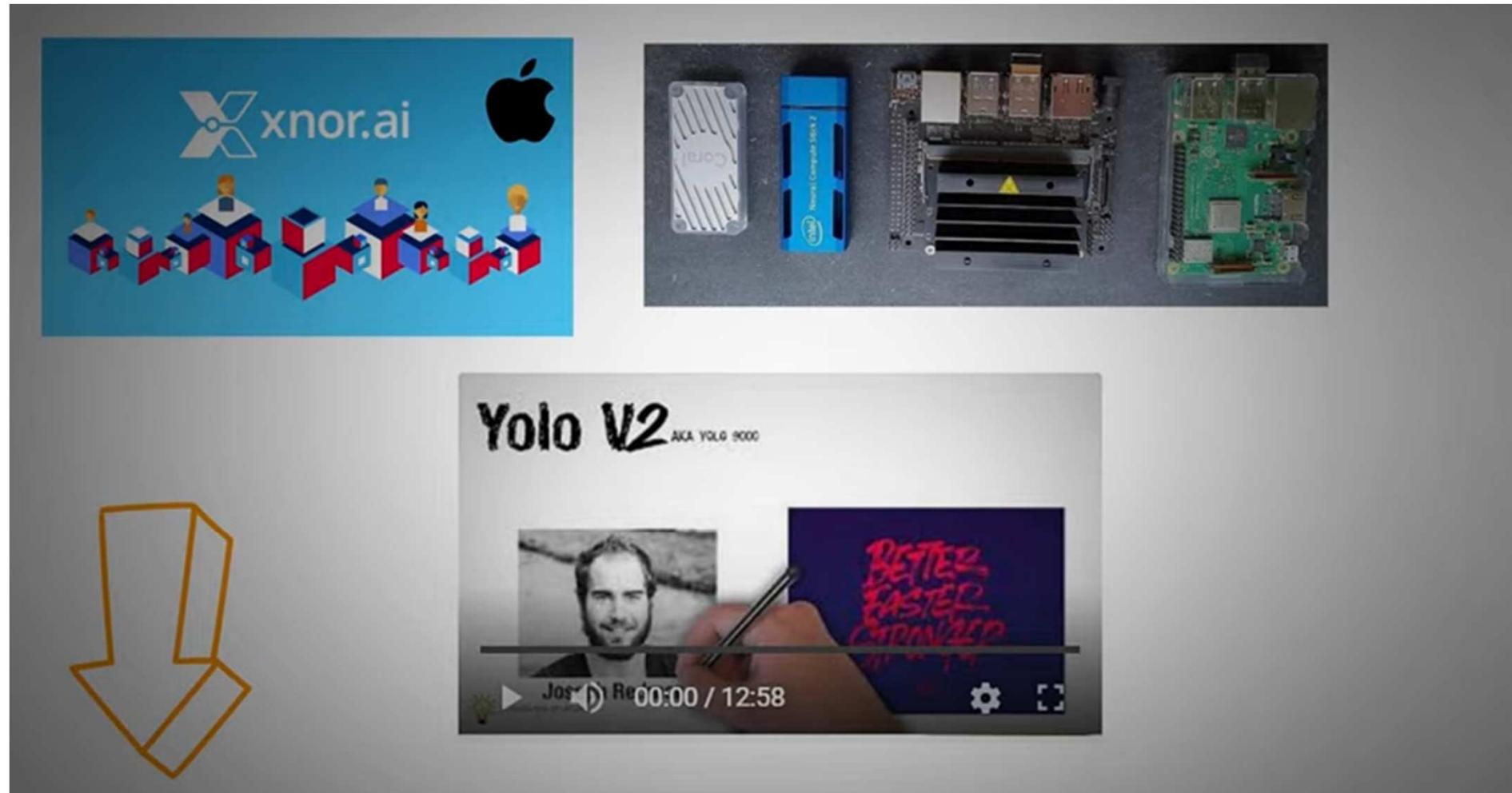
 Roger Grosse @RogerGrosse
Replying to @skouaridou

What's an example of a situation where you think someone should decide not to submit their paper due to Broader Impacts reasons?

3,345 · 6:09 PM - Feb 20, 2020

1,152 people are talking about this

Ali Farhadi



According to Forbes [1], They have since been acquired by Apple, surprise surprise.

dissecting the YOLO v4 paper



YOLOv4: Optimal Speed and Accuracy of Object Detection
Alexey Bochkovskiy*, Chien-Yao Wang*, Hong-Yuan Mark Liao
Institute of Information Science
Academia Sinica, Taiwan
kityo@iis.sinica.edu.tw
LiaoH1@iis.sinica.edu.tw

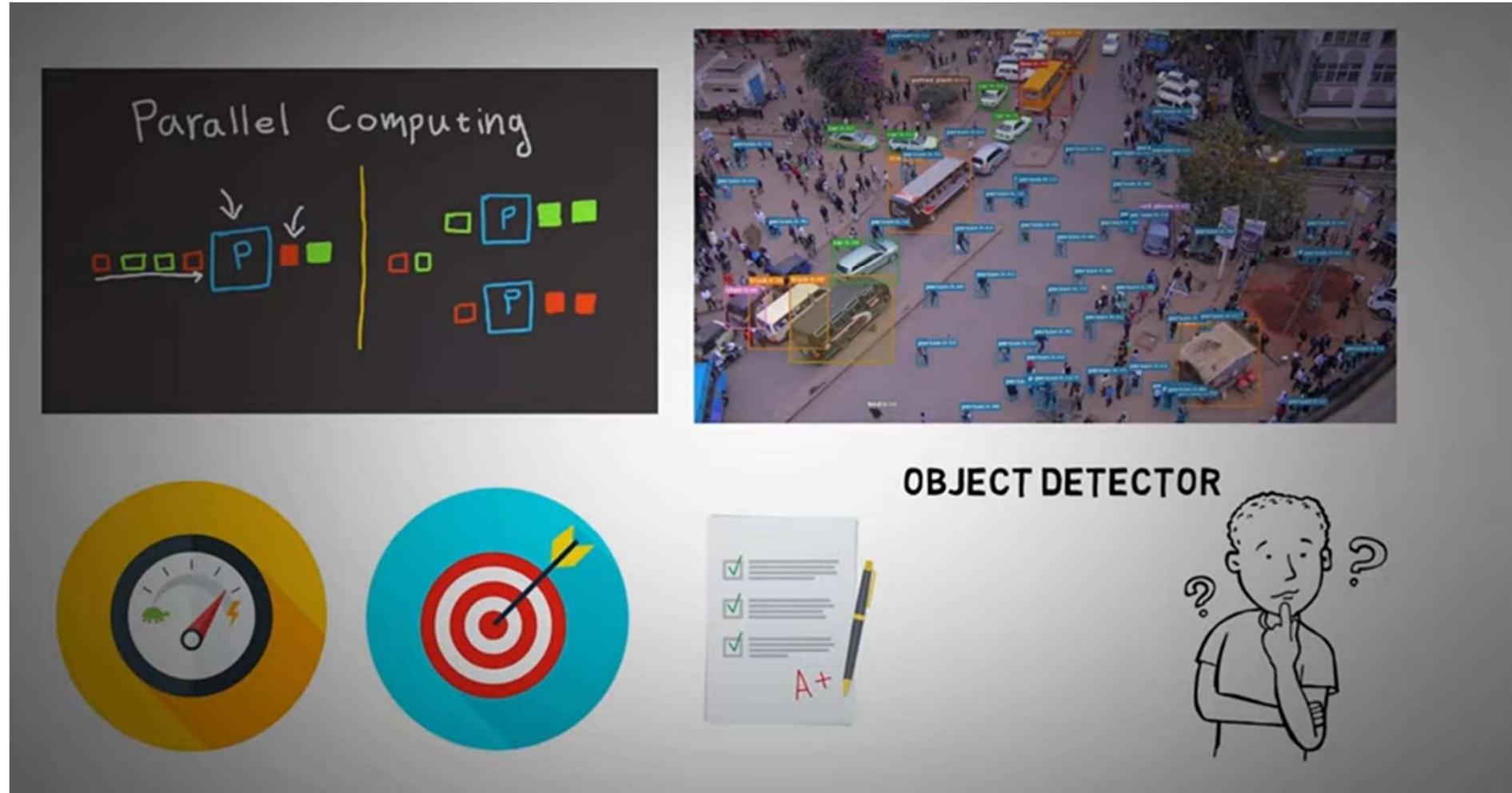
Abstract
There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the results, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets, while some features, such as batch-normalization and residual connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CnBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CnBN, SAT, Mish activation, Mosaic data augmentation, Dropblock regularization, and CIoU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (85.7% AP_m) for the MS COCO dataset at a real-time speed of ~65 FPS on Tesla V100. Source code is at <https://github.com/AlexeyAB/darknet>.

L. Introduction
The majority of CNN-based object detectors are largely applicable only for recommendation systems. For example, searching for free parking spaces via urban video cameras is executed by slow accurate models, where

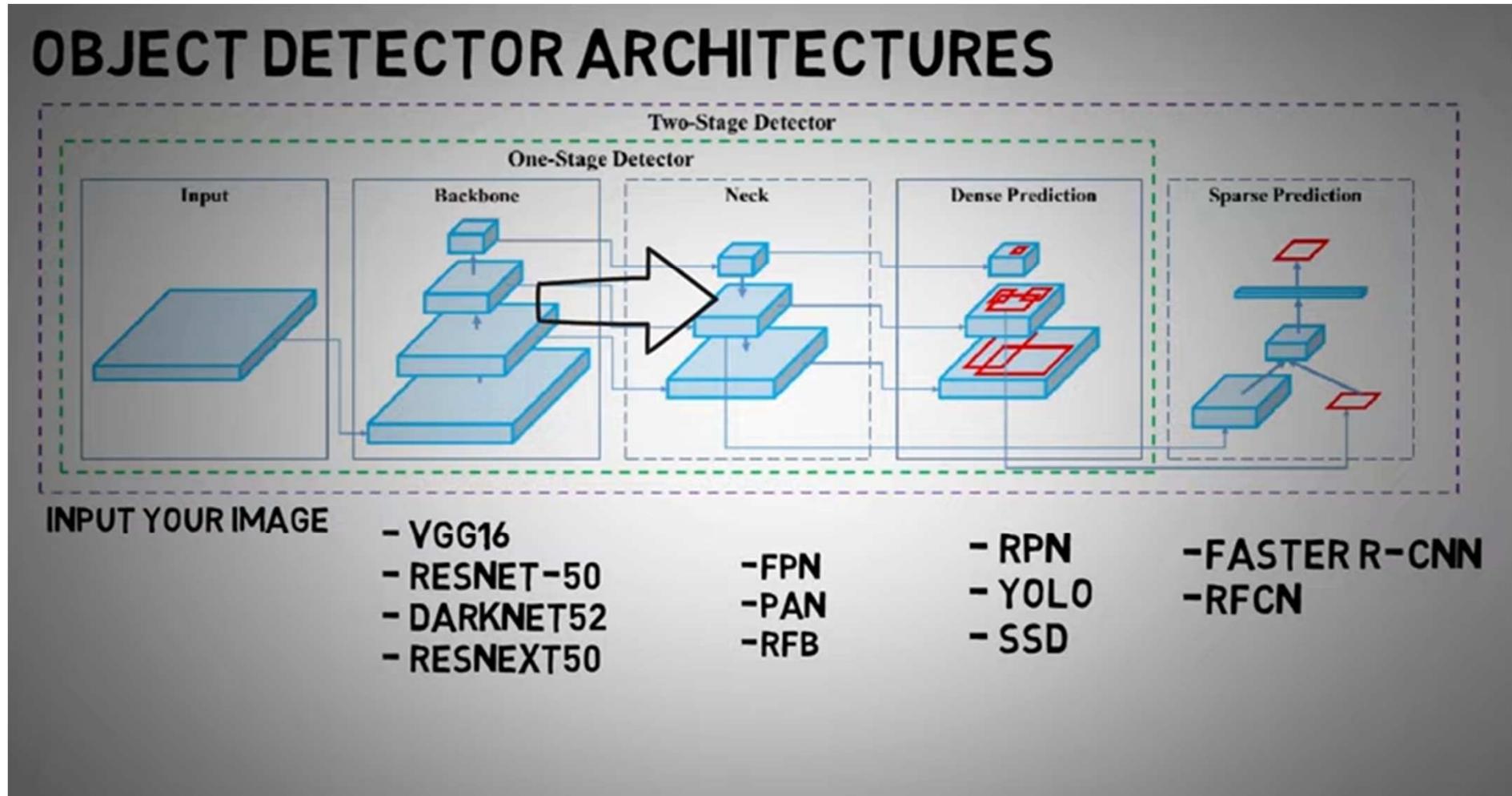
arXiv:2004.10934v1 [cs.CV] 23 Apr 2020



1. HOW IT WORKS
2. HOW IT WAS DEVELOPED
3. WHAT APPROACHED THEY USED
4. WHY THEY USED PARTICULAR METHODS.
5. AS WELL HOW IT PERFORMS IN COMPARISON TO COMPETING OBJECT DETECTION MODELS,
6. AND FINALLY, WHY IT'S SO AWESOME!



Object Detector Architectures



object detectors : compose of several components:

Input – This is where you input your image.,

Backbone – which refers to the network which takes as input the image and extracts the feature map – this may be either VGG16, Resnet-50, Darknet52 or ResNext50 variants.

neck and head are sub-sets of the backbone,

and this serves to enhance the feature discriminability and robustness using the likes of FPN, PAN, RFB etc,

Selection of Architecture

SELECTION OF ARCHITECTURE

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

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

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

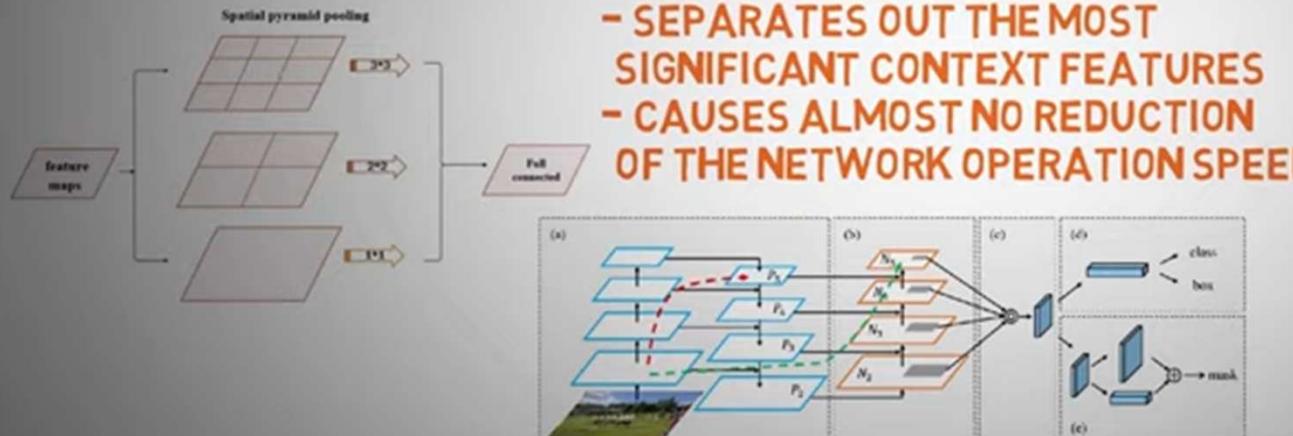
Head:

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

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

BACKBONE

- CSPResNEXT50
- CSPDarknet53
- EfficientNet B3



INSTEAD OF FEATURE PYRAMID NETWORK USED IN YOLOV3



THEY CHOSE YOLO V3 AS THE HEAD FOR YOLO V4.

Yolo V4 For the backbone – There was a choice Between CSPResNeXt50, CSPDarknet53, and EfficientNetB3 – based on theoretical justification and several experiments CSPDarknet53 neural network was shown to be the most optimal model.

Training Optimizations

TRAINING OPTIMIZATIONS



BAG OF FREEBIES



For the backbone

The collage consists of three main sections:

- Left Section:** A diagram titled "FOR THE BACKBONE" showing a butterfly image being processed by "Data Augmentation" into various forms: Original Image, De-texturized, De-colored, Edge Enhanced, Salient Edge Map, and Flip/Rotate.
- Middle Section:** A screenshot of the AugmentedStartups.com website showing three AI courses:
 - ULTIMATE AI-CV PRACTITIONER PRO** (\$69.00) with 1004 reviews and 2012 students.
 - OPENPOSE POSE ESTIMATION** (\$0.00) with 200 reviews and 8880 students.
 - MASK-RCNN MASTERCLASS** (\$49.00) with 6394 reviews and 3080 students.
- Bottom Section:** The text "AUGMENTEDSTARTUPS.COM" in large, bold, black letters.

Bottom Text:

- CUTMIX AND MOSAIC DATA AUGMENTATIONS,
- DROPOBLOCK REGULARIZATION
- CLASS LABEL SMOOTHING

For the Detector

FOR THE DETECTOR.

THE AUTHORS USE:

- CIOU-LOSS,
- CMBN,
- DROPOUT REGULARIZATION,
- MOSAIC DATA AUGMENTATION,
- SELF-ADVERSARIAL TRAINING,
- ELIMINATE GRID SENSITIVITY,
- USING MULTIPLE ANCHORS FOR A SINGLE GROUND TRUTH,
- COSINE ANNEALING SCHEDULER
- OPTIMAL HYPERPARAMETERS, RANDOM TRAINING SHAPES.

Bag of Specials



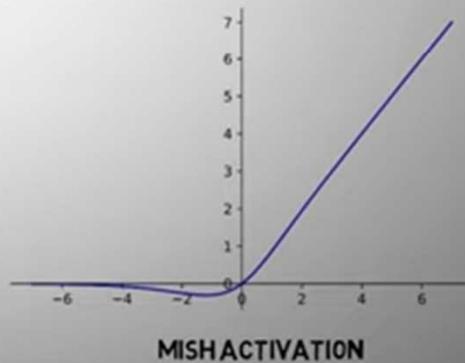
BoS for the Backbone/BoS for the Detector

BAG OF SPECIALS (BOS) FOR THE BACKBONE:

- MISH ACTIVATION,
- CROSS-STAGE PARTIAL CONNECTIONS (CSP), AND
- MULTIINPUT WEIGHTED RESIDUAL CONNECTIONS (MIWRC)

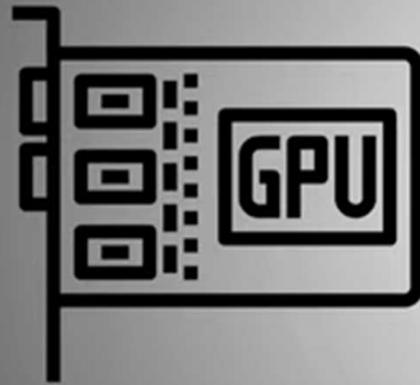
BAG OF SPECIALS (BOS) FOR THE DETECTOR:

- MISH ACTIVATION,
- SPP-BLOCK, SAM-BLOCK,
- PAN PATH-AGGREGATION BLOCK,
- DI0U-NMS

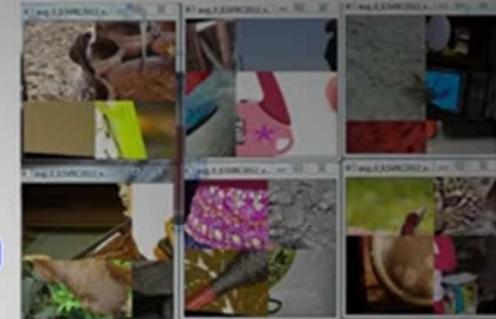


Additional Improvements

ADDITIONAL IMPROVEMENTS



- THEY INTRODUCE A NEW METHOD OF DATA AUGMENTATION:
-MOSAIC,
-SELF-ADVERSARIAL TRAINING (SAT)
- THEY SELECT OPTIMAL HYPER-PARAMETERS WHILE APPLYING GENETIC ALGORITHMS
- THEY ALSO MODIFY SOME EXISTING METHODS TO MAKE THEIR DESIGN SUITABLE FOR EFFICIENT TRAINING AND DETECTION



Experimental Setup

EXPERIMENTAL SETUP



NVIDIA 1080 TI AND 2080 TI



Car Motorbike Chair Sofa Bottle



IMAGENET

HYPER PARAMETERS :

- TRAINING STEPS - 8 MILLION
- BATCH SIZE - 128
- MINI BATCH SIZE - 32
- LEARNING RATE - 0.1 USING POLYNOMIAL DECAY..... STRATEGY
- WARMUP STEPS - 1000
- MOMENTUM - 0.9
- WEIGHT DECAY - 0.005

MS COCO

HYPER PARAMETERS :

- TRAINING STEPS - 500,500
- LEARNING RATE - 0.1 USING STEP.....STRATEGY
- LEARN STEPS - 0.1 MULTIPLIED AT 400K
AND 450K RESPECTIVELY
- MOMENTUM - 0.9
- WEIGHT DECAY - 0.005

Experiments

EXPERIMENTS

YOU GET A TECHNICAL!

AND YOU GET A TECHNICAL!

EVERYONE GETS A TECHNICAL!

AIYO-Bachokonyi*, Chen-Yue Wang[#], Hung-Yuan Lin^{**}
Institute of Information Science
Academia Sinica, Taiwan
RESEARCH CENTER FOR
CLOUD COMPUTING, TAIWAN

YOLOv4: Optimal Speed and Accuracy of Object Detection

Abstract

There are a huge number of datasets which are used in different “Object detection” Model training (YOLO series). However, there are many cases that the performance of the model is not good enough. In this paper, we propose a new dataset named “Technical Dataset” which is composed of technical jargon and technical terms. This dataset contains 1000 images and 1000 labels. Some features of this dataset are: 1) It is a technical dataset, so it is suitable for technical applications. 2) It is a dataset with some features, such as hard-to-discriminate and low-quality images, are applicable to the design of real-world applications. 3) The objects in this dataset are not static, and they are dynamic. 4) The objects in this dataset include: Biological Object Detection (BOD), Environmental Object Detection (EOD), Industrial Object Detection (IOD), Office Object Detection (OOD), and Multitask Detection. We use two datasets: WBC-CFD and YOLOv4. The WBC-CFD dataset is composed of CFD, Fluid flow simulation, and CFD box, and contains some types of objects in a strict way. The size of our dataset is 1000 images and 1000 labels. The accuracy of our dataset is about 40% F1P@10. The YOLOv4 dataset is composed of 1000 images and 1000 labels. The accuracy of our dataset is about 40% F1P@10. The YOLOv4 dataset is composed of 1000 images and 1000 labels. The accuracy of our dataset is about 40% F1P@10.

Figure 1: Comparison of the proposed YOLOv4 and other state-of-the-art object detector. Without fine-tune YOLOv4, the mAP is 39.7%, AP and AP50 by 39.1% and 12.5%, respectively. The main goal of this work is designing a fast operating speed of an object detector in production systems and optimizing the performance of the detector. The most important positive value of our indicator (MAP@50). We hope that the designed object can be easily learned and used. For example, searching for the first parking space via office video camera is monitored by slow learning models, whereas our software is learned in few seconds. Therefore, we can use our system to detect objects in real-time, high-quality, and fast per detection results, as the YOLOv4 results shown in Fig. 1.

- DIFFERENT FEATURES ON CLASSIFIER TRAINING
- DIFFERENT FEATURES ON DETECTOR TRAINING
- DIFFERENT BACKBONES AND PRETRAINED WEIGHTINGS ON DETECTOR TRAINING
- DIFFERENT MINIBATCH SIZE ON DETECTOR TRAINING

ABLATION STUDY

The collage consists of three images. On the left is a dark circular icon containing a stylized molecular or neural network structure with green and blue spheres connected by yellow lines. In the center is a photograph of an elderly woman with glasses, looking surprised or confused, with the text "WHAAAAAT?" overlaid. On the right is a cartoon illustration of a man with a beard, wearing a white shirt, brown trousers, and a tie, holding a pointer stick and standing next to a graph showing a fluctuating line with a bar chart at the top.

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%

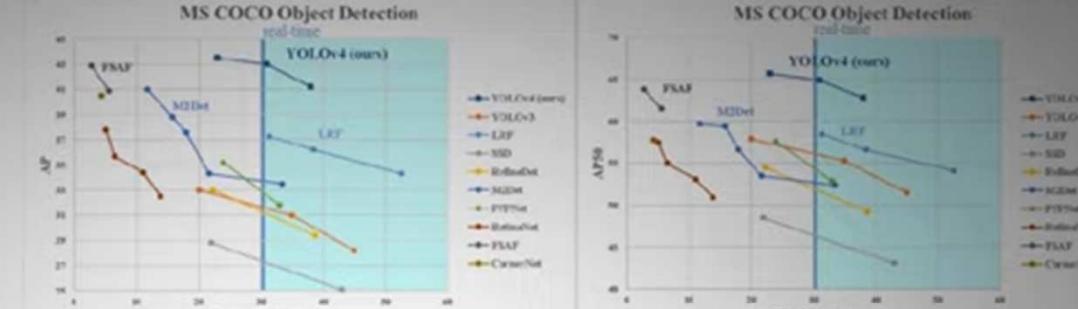
RESULTS

RESULTS

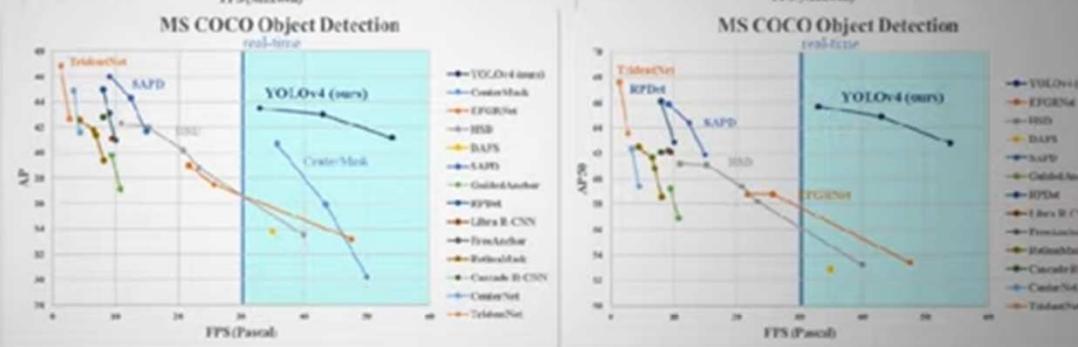


Speed and Accuracy

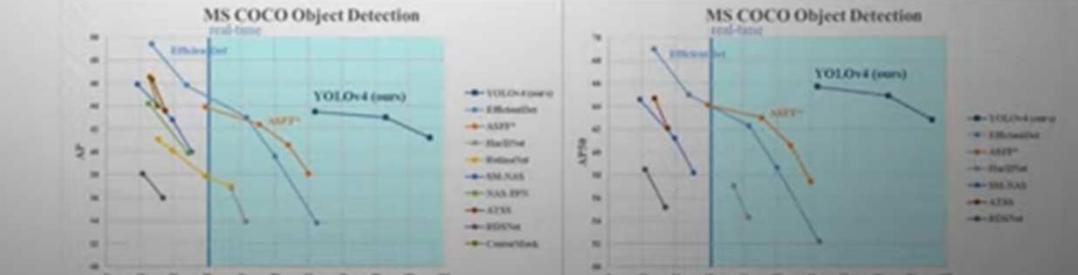
Maxwell



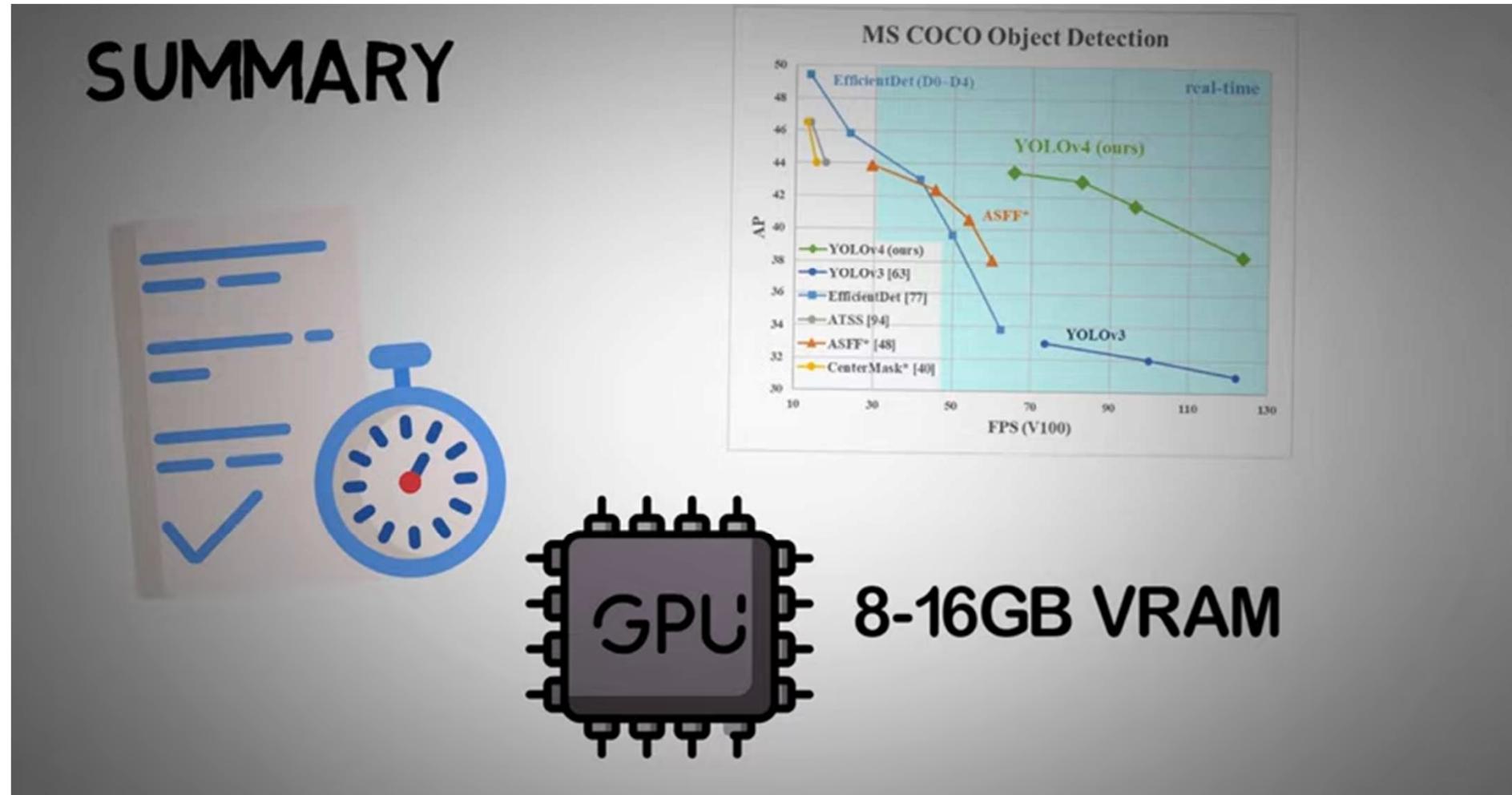
Pascal



Volta



SUMMARY



SUMMARY



YOLOv4: Optimal Speed and Accuracy of Object Detection

Alexey Bochkovskiy*
slexeyab4@gmail.com

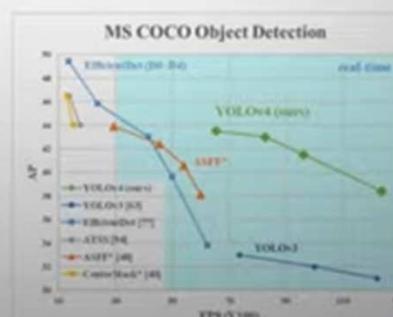
Chien-Yao Wang*
Institute of Information Science
Academia Sinica, Taiwan
kinyiu@iis.sinica.edu.tw

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

Abstract

There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the result, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP,

v1 [cs.CV] 23 Apr 2020





Okay, so earlier I mentioned that if you are interested in winning one of 20 free enrollments to any Augmented Startups courses, then all you have to do is like and comment on this video, and sign up for the webinar at the link below.

이전에 언급한 바와 같이, **Augmented Startups**의 어떤 코스든 20개의 무료 등록 중 하나를 원하신다면, 이 영상을 좋아요를 누르고 댓글을 달고 아래 링크에서 웨비나에 등록하기만 하면 됩니다.

Winners are announced every 3 months.

당첨자는 3개월마다 발표됩니다.

And you can comment about anything or provide suggestions for upcoming video ideas, as I've mentioned earlier.
그리고 앞서 언급한 바와 같이, 무엇이든 댓글을 달거나 다음 영상 아이디어에 대한 제안을 제공하실 수 있습니다.

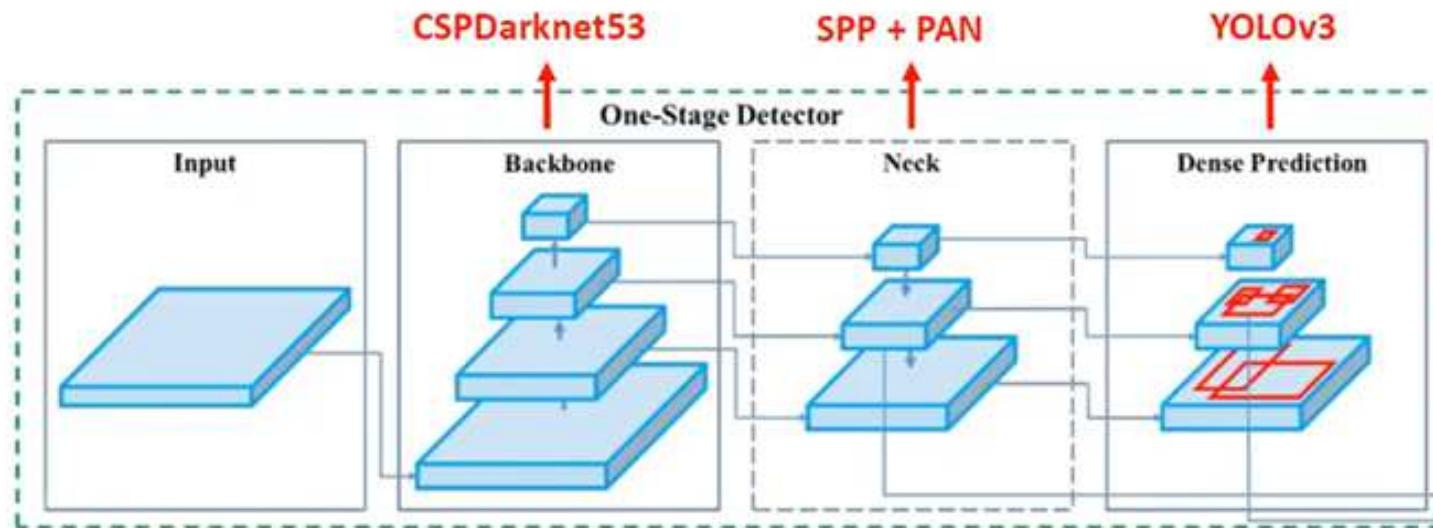
If you are interested in Artificial Intelligence in Computer Vision, we have a course that teaches you AI object detection, object segmentation, pose estimation, Android AI app development, along with the complete Project EDITH and Project Smart Glass tutorial series.

컴퓨터 비전에서 인공지능에 관심이 있으시다면, AI 객체 탐지, 객체 분할, 포즈 추정, 안드로이드 AI 앱 개발, 그리고 **Project EDITH** 및 **Project Smart Glass** 튜토리얼 시리즈 전체를 가르치는 코스가 있습니다.

Alright, thank you for watching, and see you in the next lecture.

자, 시청해 주셔서 감사합니다. 다음 강의에서 뵙겠습니다.

Selection of architecture



YOLOv4 = YOLOv3 + CSPDarknet53 + SPP + PAN + BoF + BoS

↓
Path Aggregation Network