YOLO V4

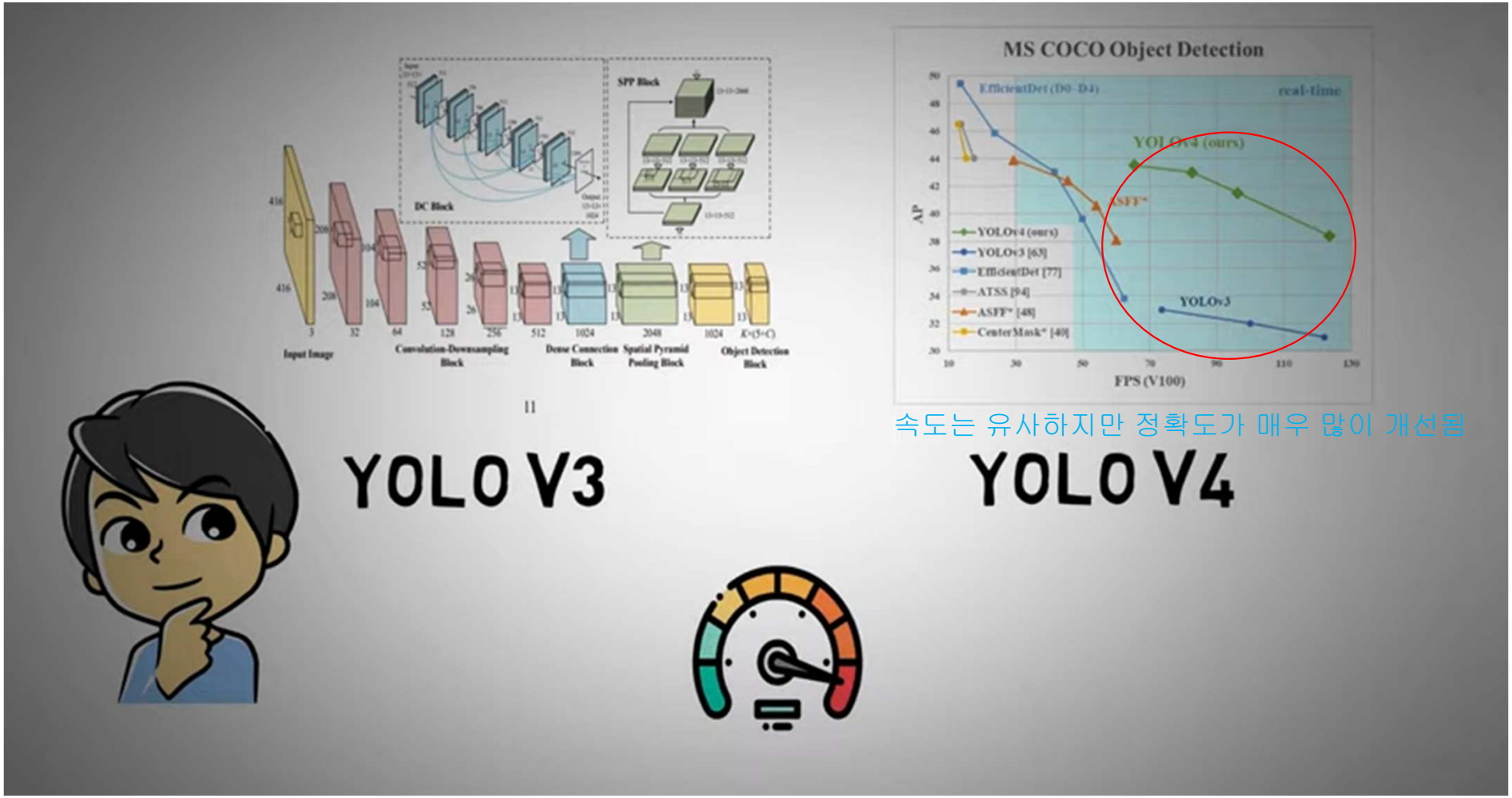
Optimal Speed And Accuracy of Object Detection

박 철

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Introduction to yolo v4 object detection

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



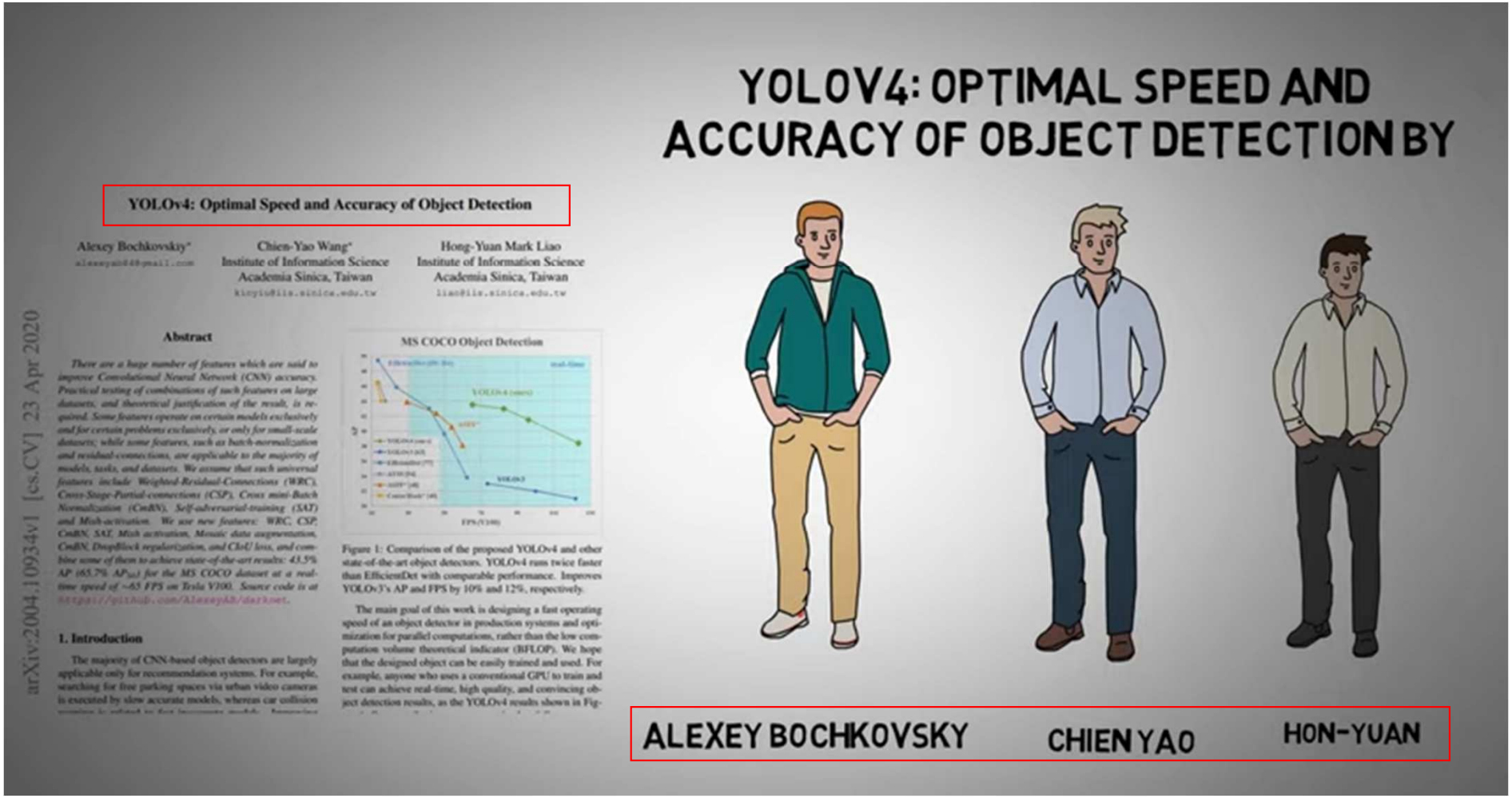
속도는 유사하지만 정확도가 매우 많이 개선됨

quite popular, robust and quick

speed and performance : 매우 향상됨

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

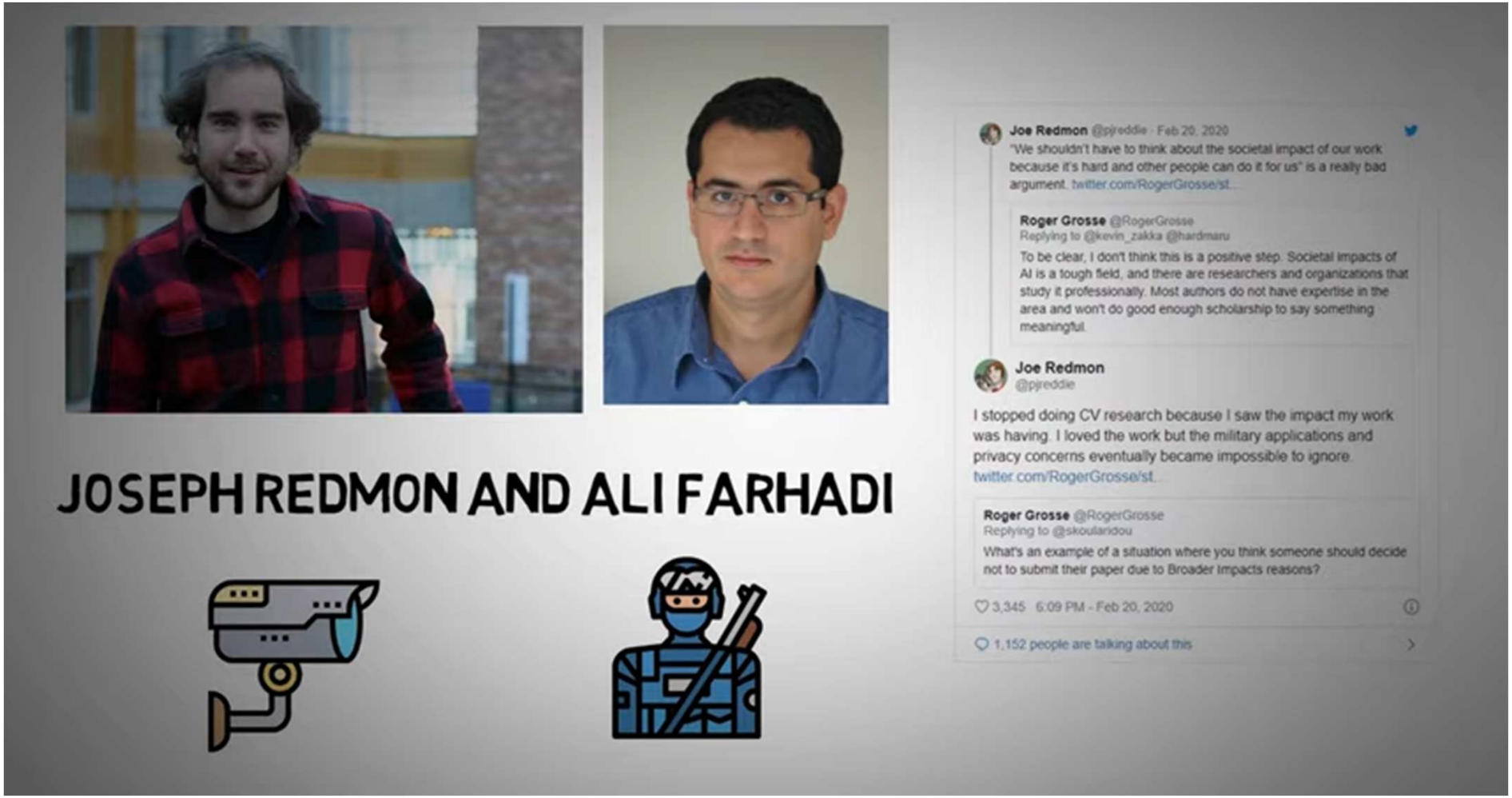
the paper YOLOv4



출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Joseph Redmon and Ali Farhadi

original creators of Yolo V1-3

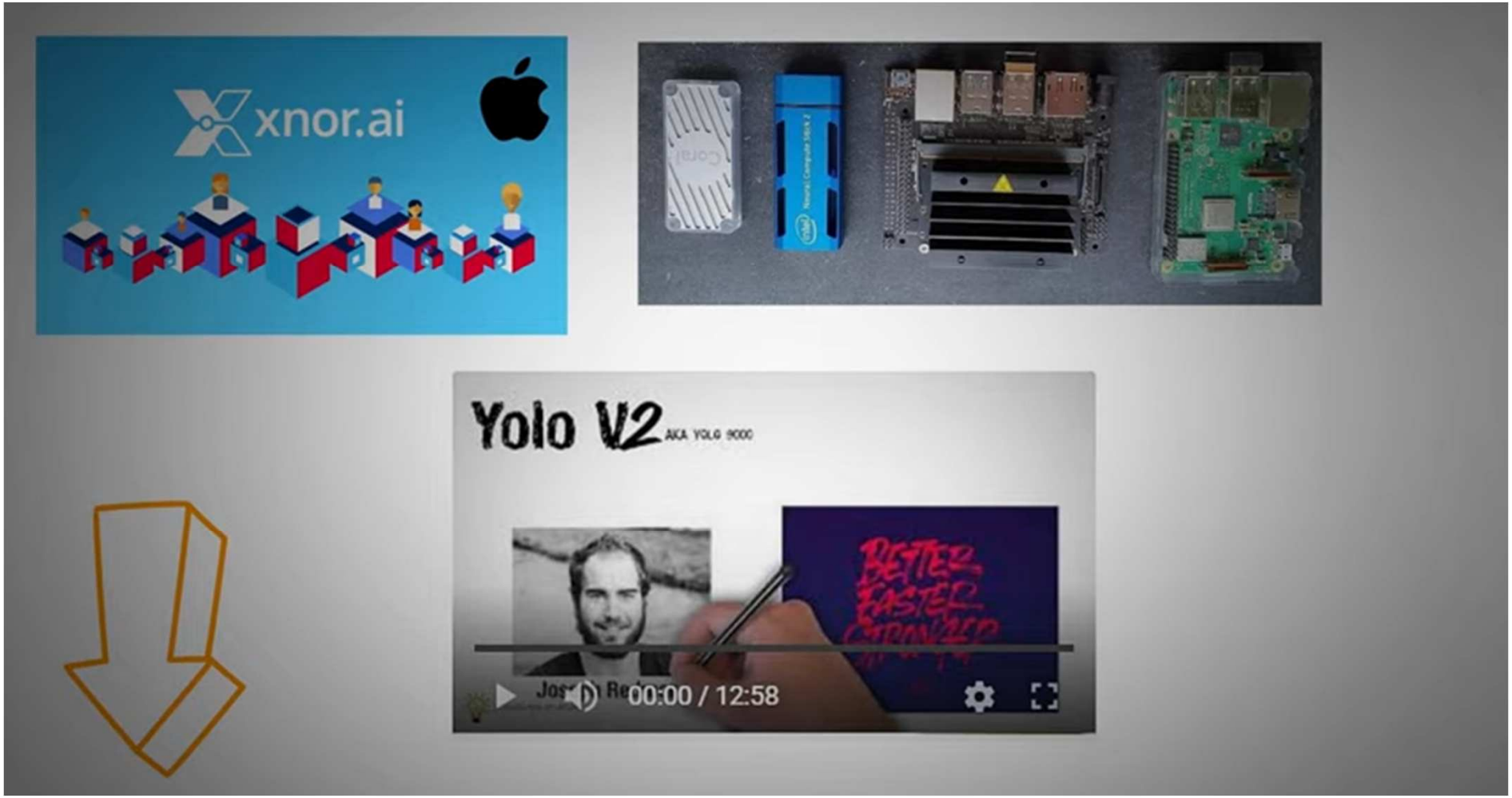


in Feb 2020 that Joseph stopped Computer vision research

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

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

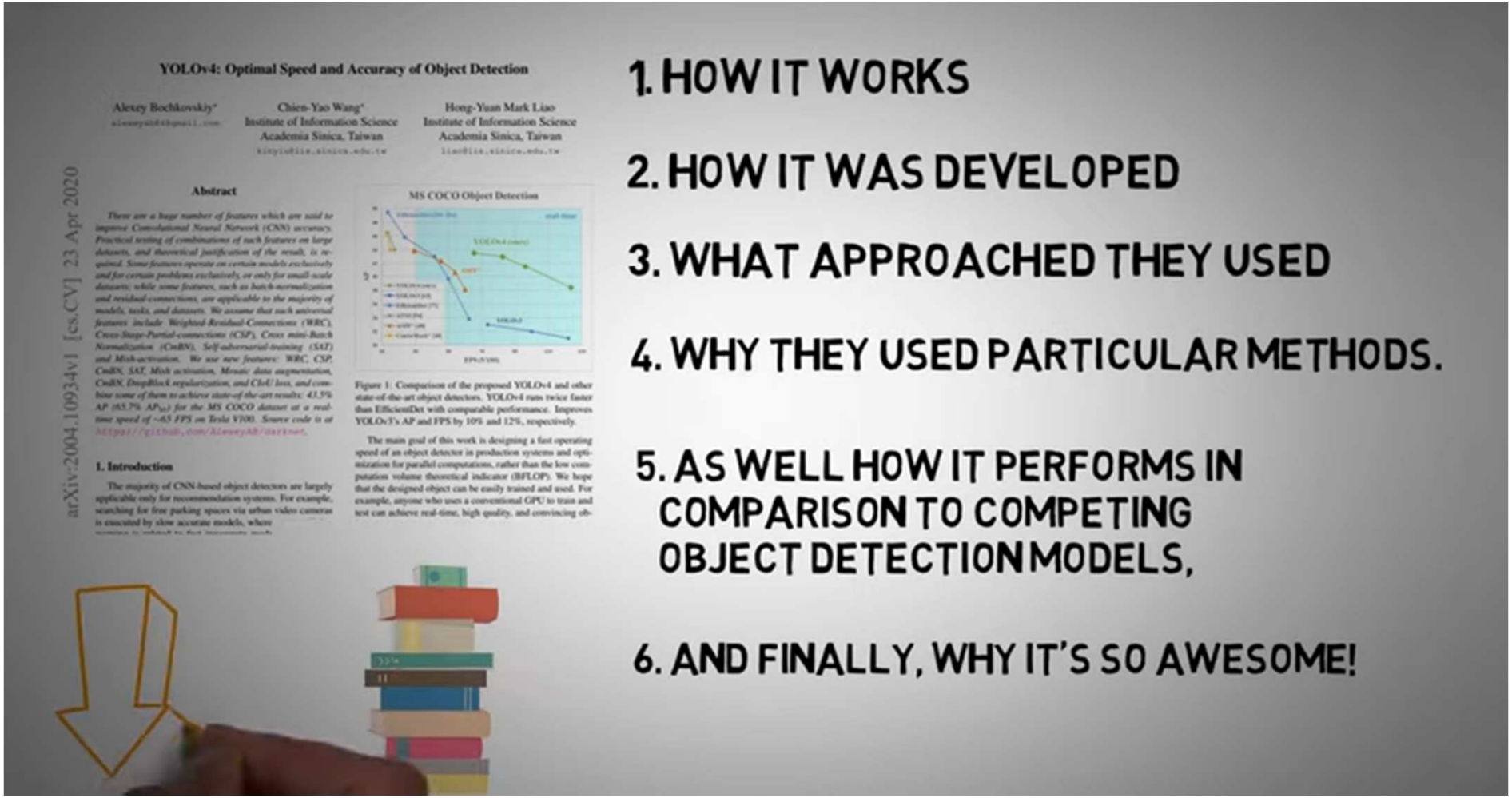
Ali Farhadi



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

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

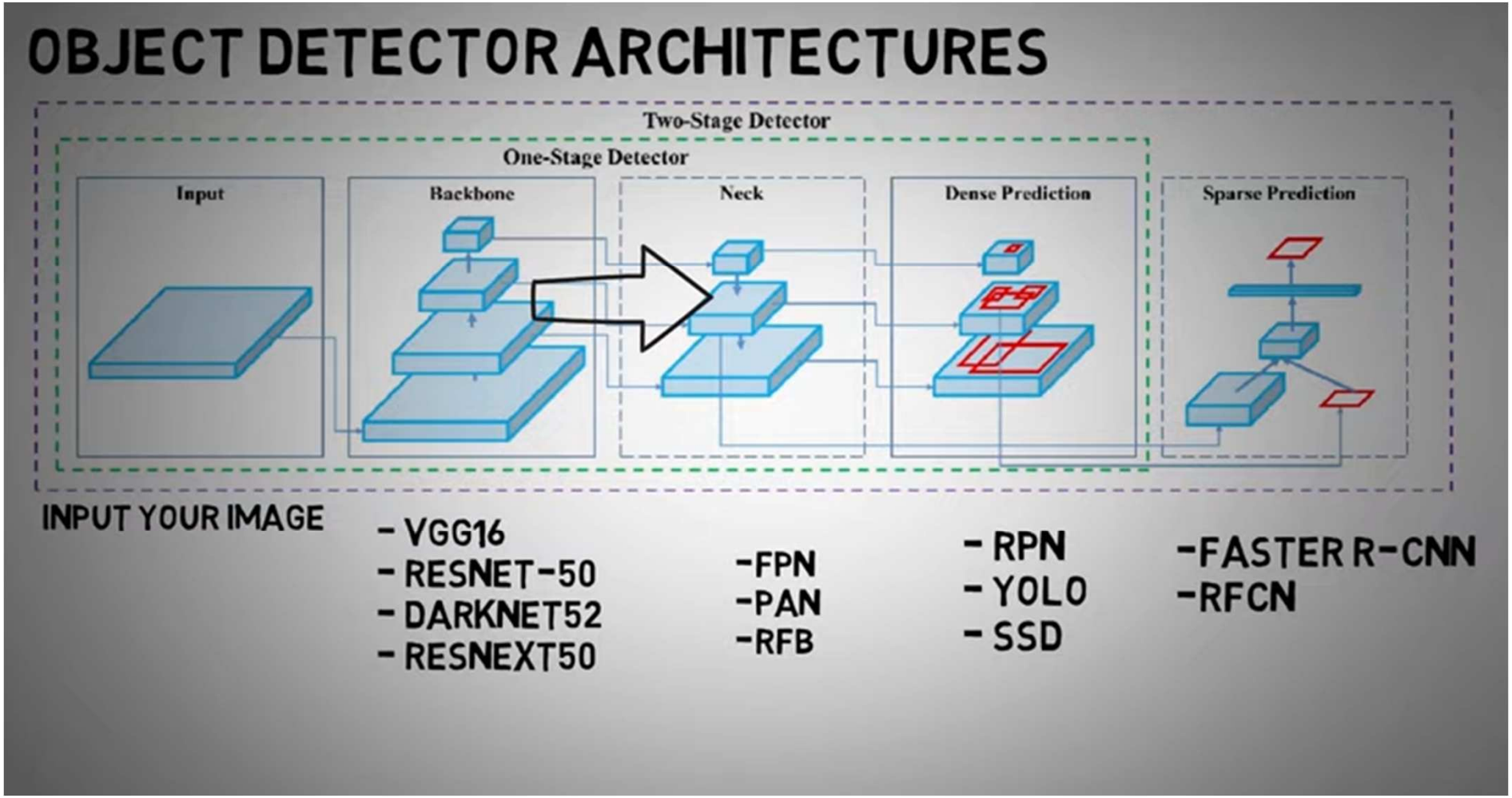
dissecting the YOLO v4 paper



출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

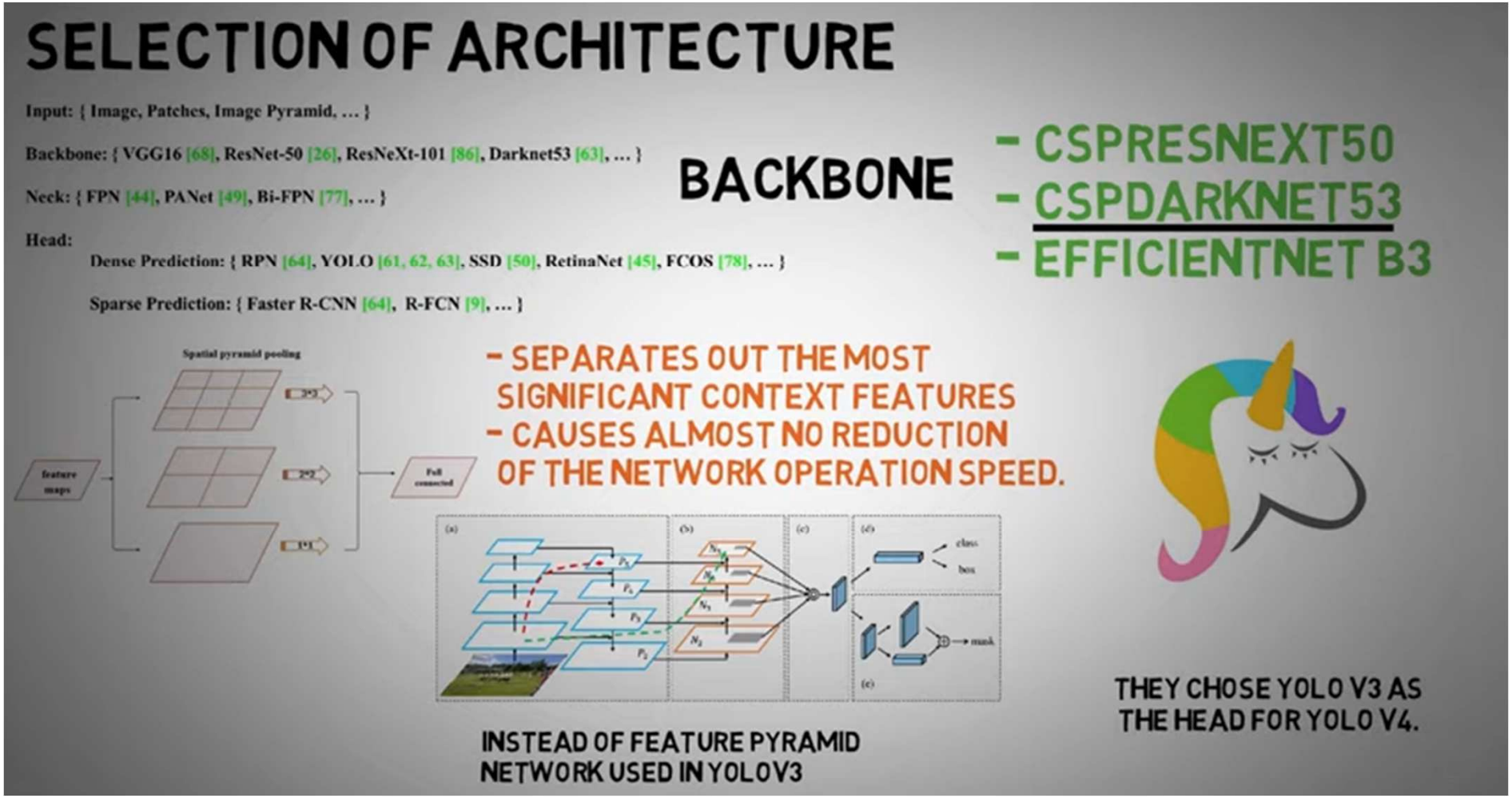
Object Detector Architectures



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,

출처 : https://www.youtube.com/watch?v=_JZOPWx1VZg

Selection of Architecture



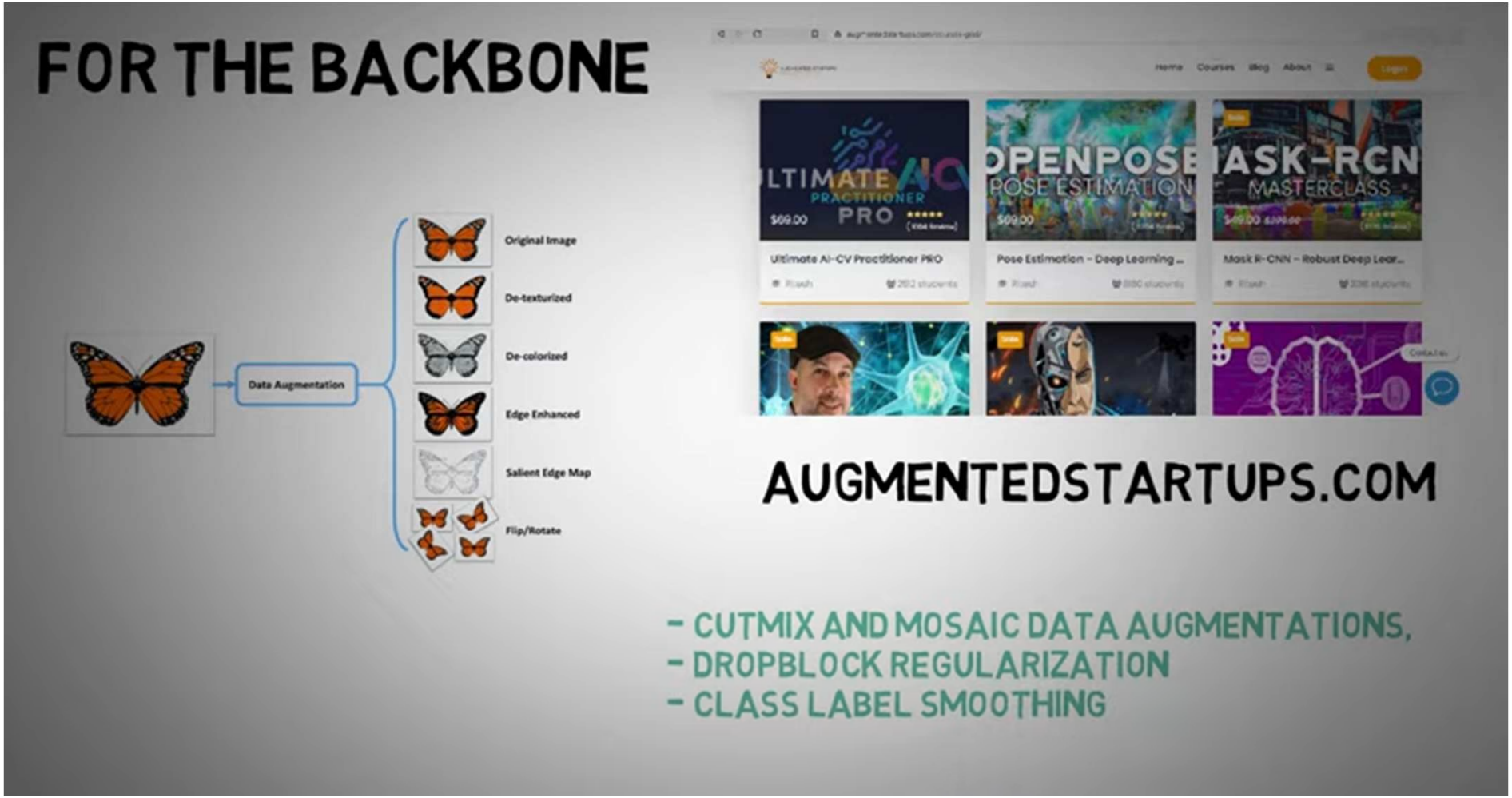
Yolo V4 For the backbone – There was a choice Between CSPResNeXt50, CSPDarknet53, and EfficientNetB3 – based on theoretical justification and several experiments CSPDarknet53 neural network was shown to be the most optimal model.

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Training Optimizations

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

For the backbone



출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

For the Detector

FOR THE DETECTOR.

THE AUTHORS USE:

- CIOU-LOSS,
- CMBN,
- DROPBLOCK 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.

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Bag of Specials

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

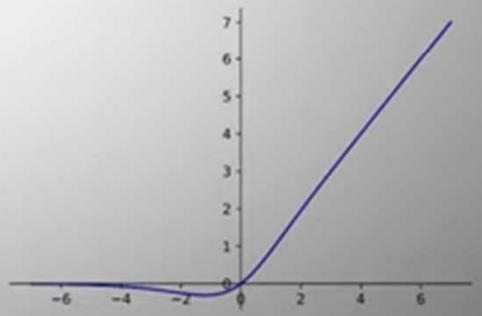
BoS for the Backbone/BoS for the Detector

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,
- DIOU-NMS

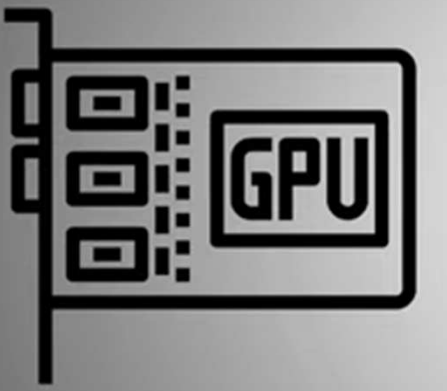


MISH ACTIVATION

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

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

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Experimental Setup

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

IMAGENET

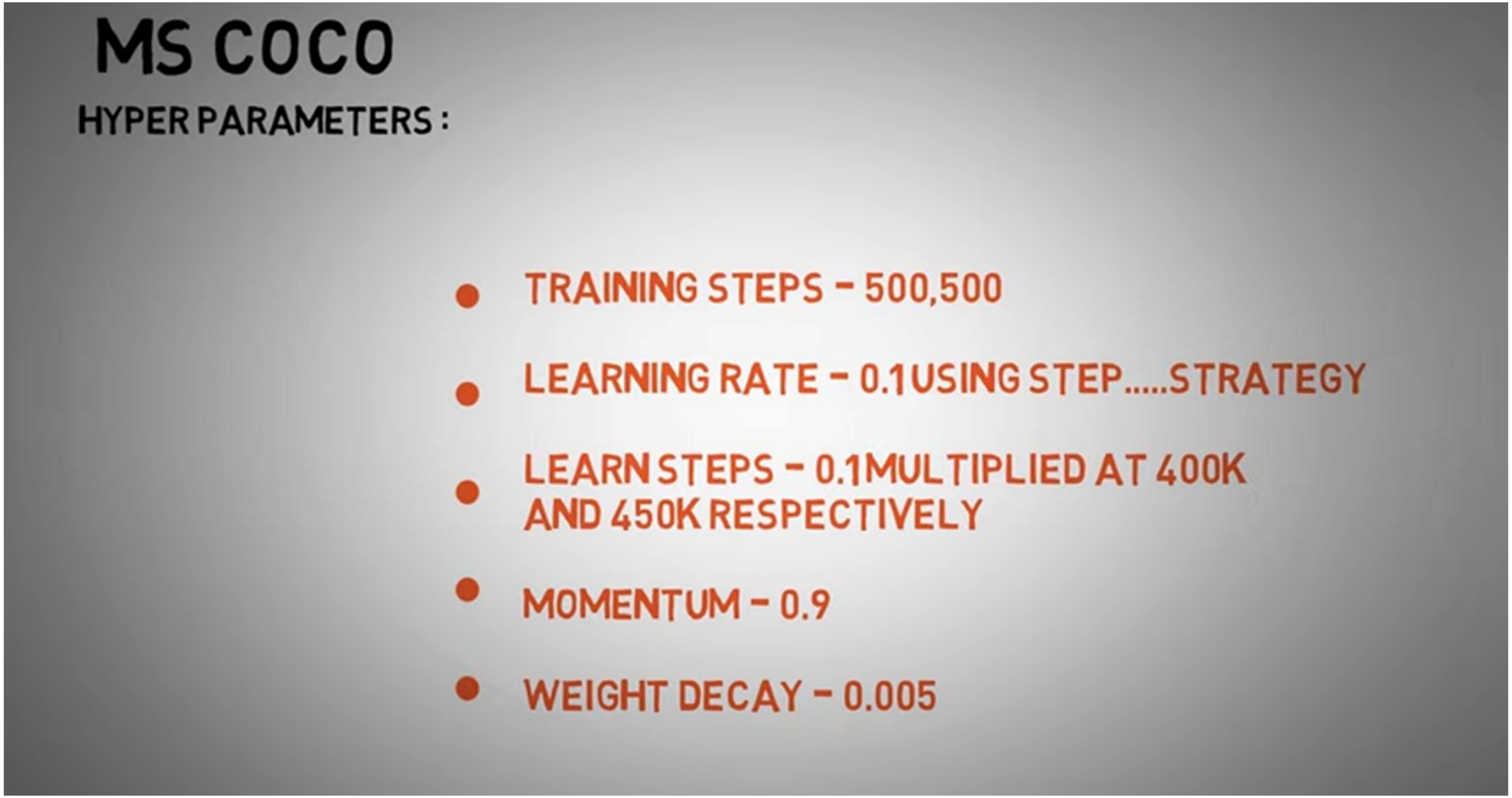
IMAGENET

HYPER PARAMETERS :

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

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

MS COCO



출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Experiments

EXPERIMENTS

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

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

ABLATION STUDY

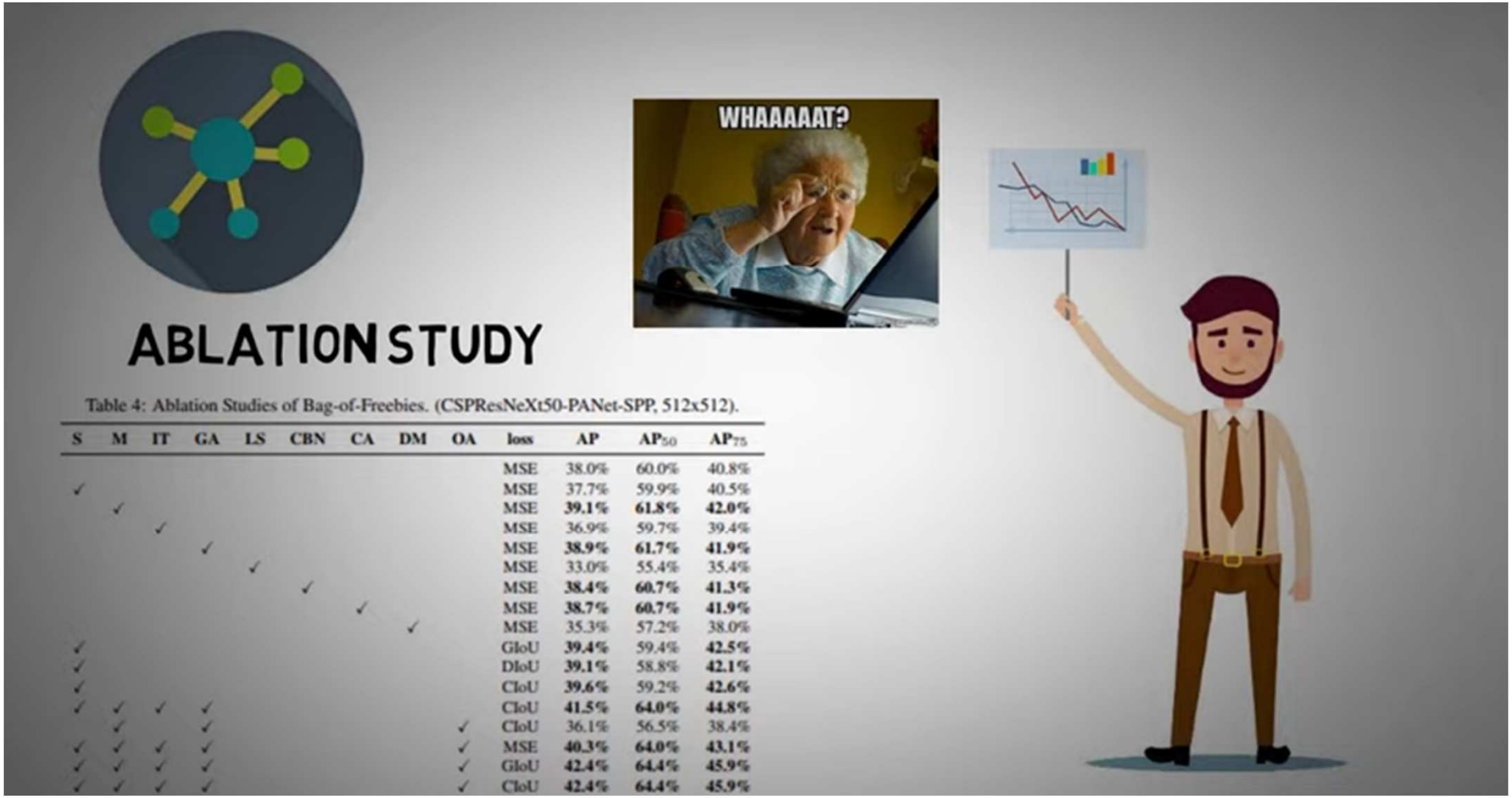
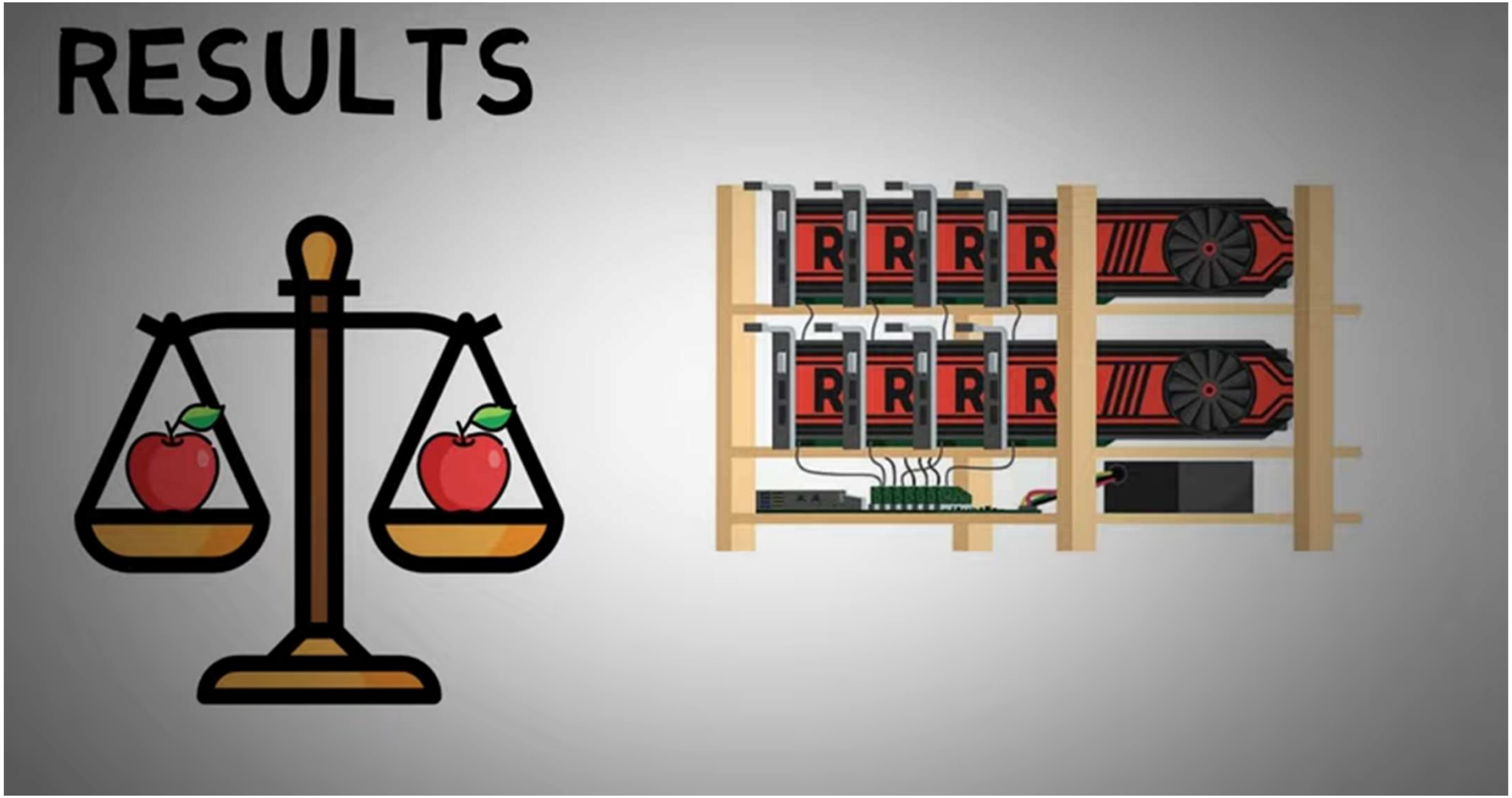


Table 4: Ablation Studies of Bag-of-Freebies. (CSPResNeXt50-PANet-SPP, 512x512).

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%**

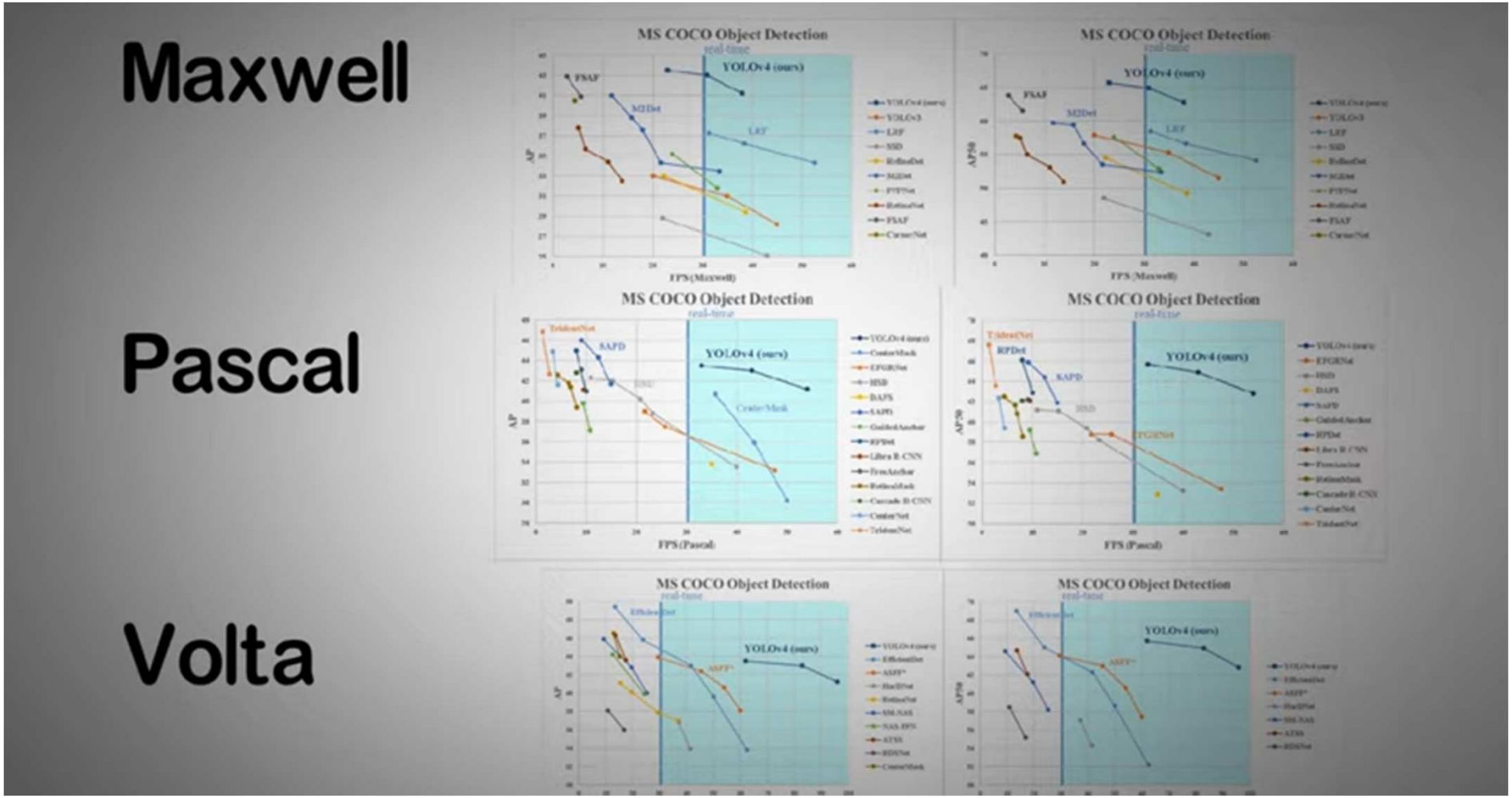
출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

RESULTS



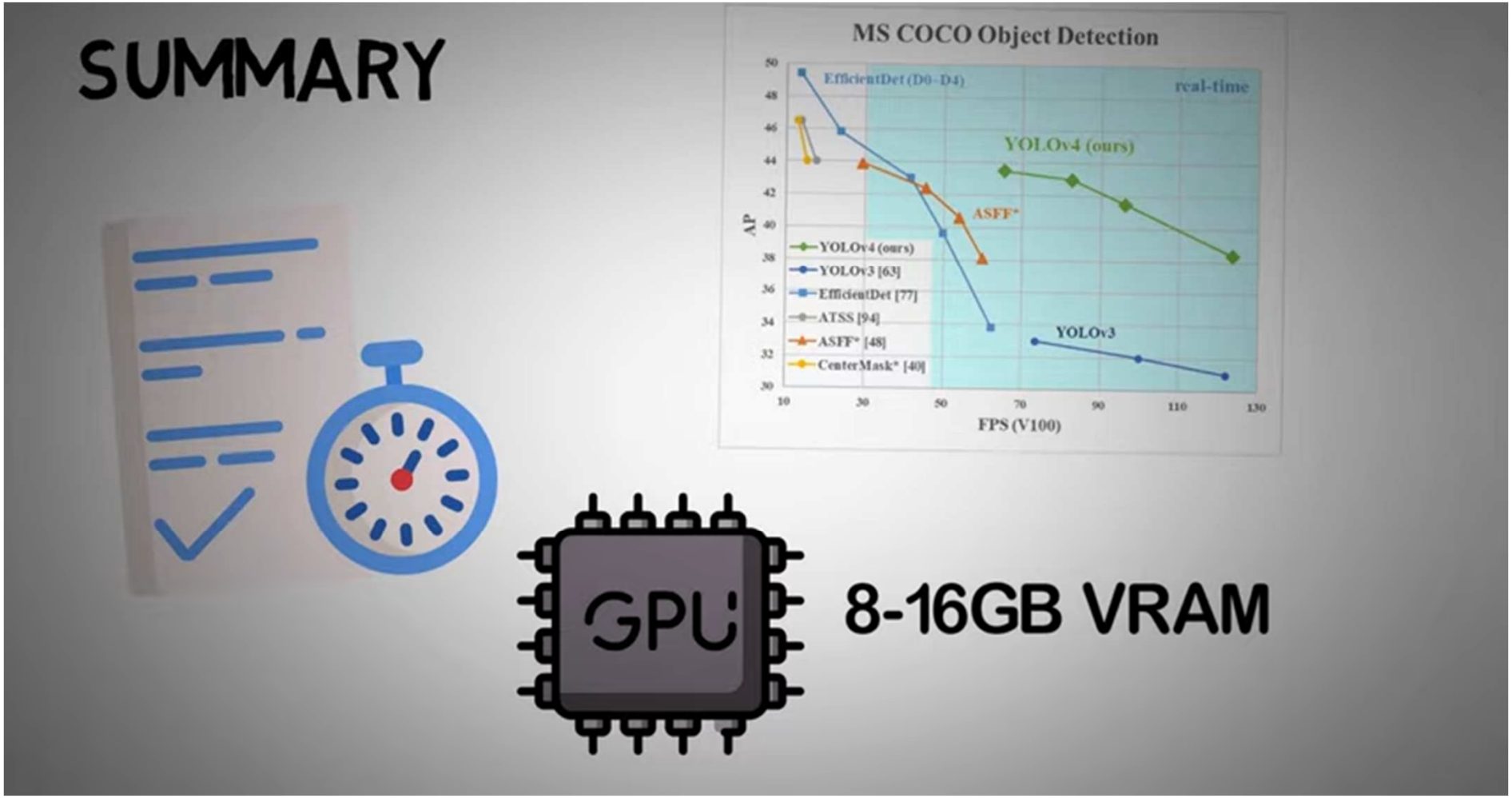
출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

Speed and Accuracy



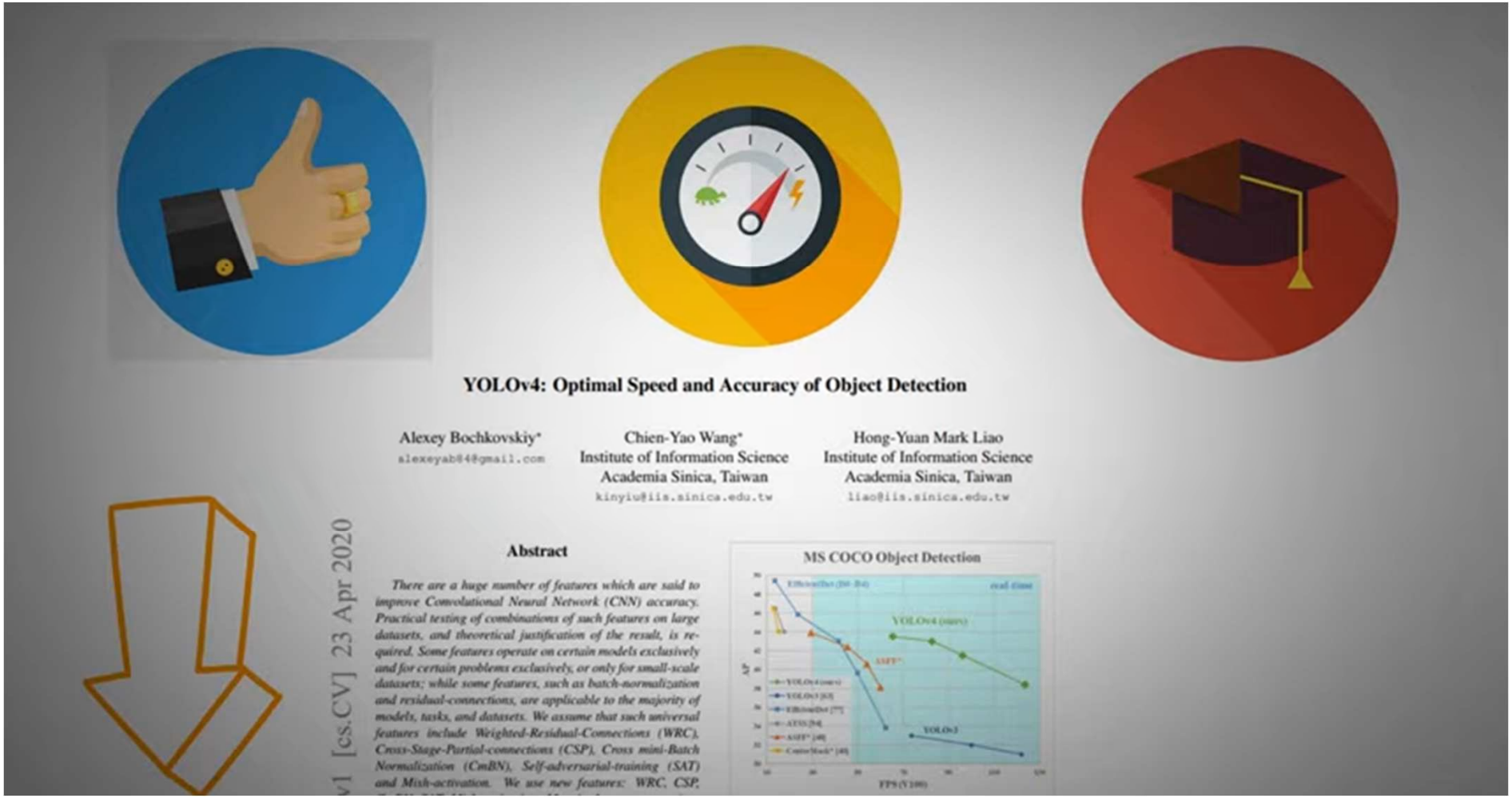
출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

SUMMARY



출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

SUMMARY



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

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,

출처 : https://www.youtube.com/watch?v=_JzOFWx1vZg

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 **20**개의 무료 등록 중 하나를 원하신다면, 이 영상을 좋아요를 누르고 댓글을 달고 아래 링크에서 웨비나에 등록하기만 하면 됩니다.

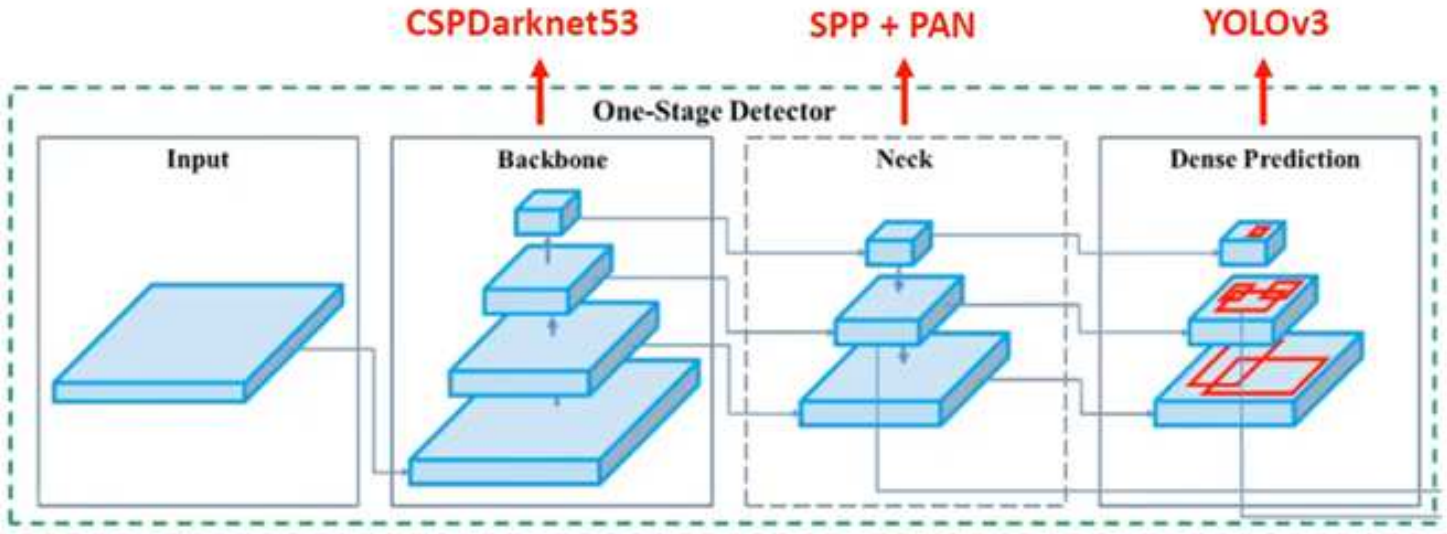
Winners are announced every 3 months.
당첨자는 **3**개월마다 발표됩니다.

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