

YOLO V1 to YOLO V3

YOLO V3 Object Detection Better, Strong, Faster

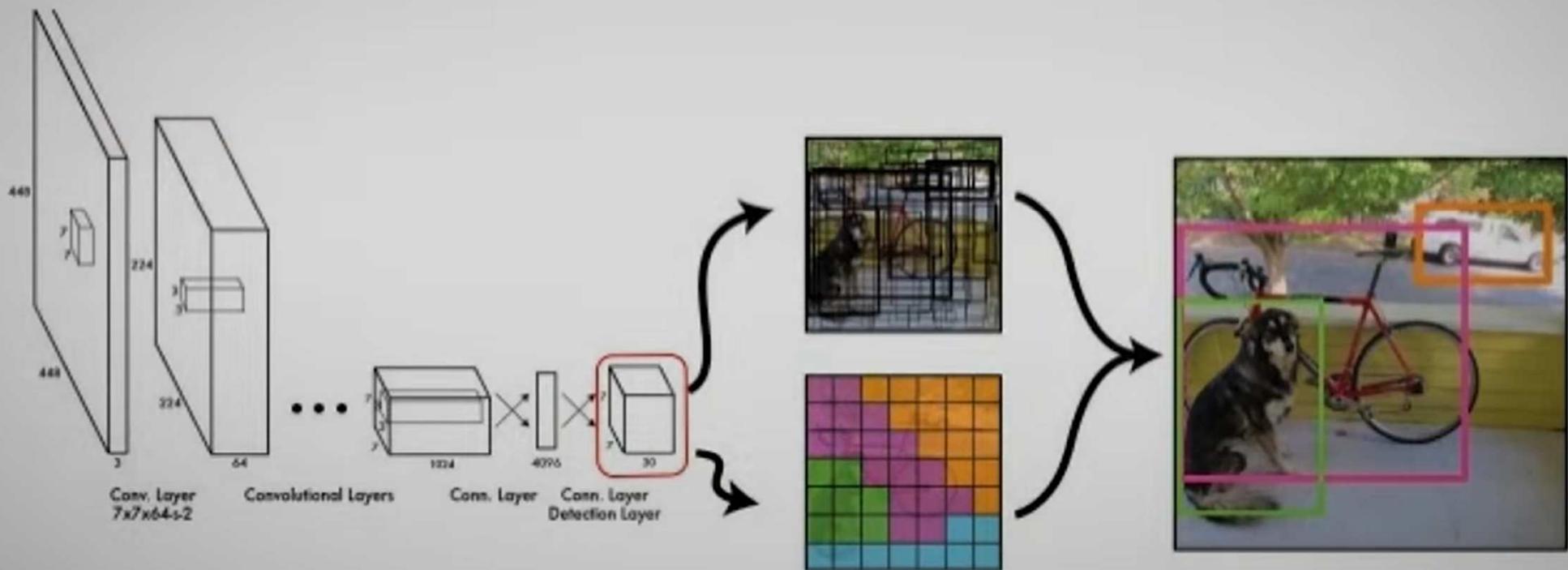
박 철

You Only Look Once



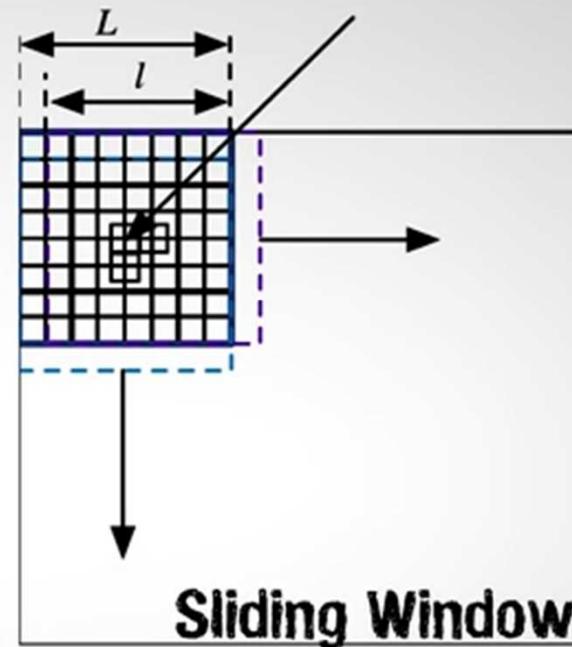
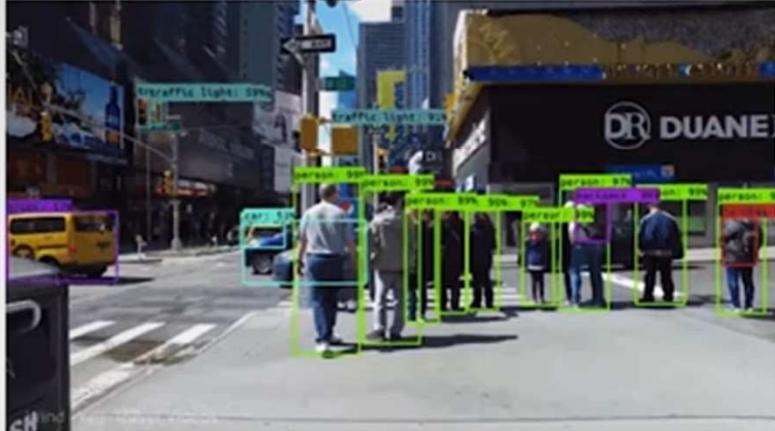
Yolo – V1, V2 and V3

Yolo - V1, V2 and V3

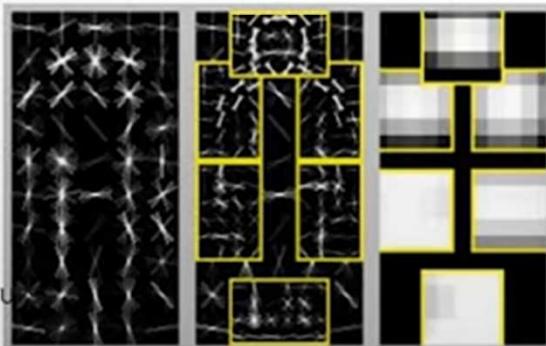


Back ground

Background



Deformable Parts Model
(DPM)



RCNN and Faster RCNN

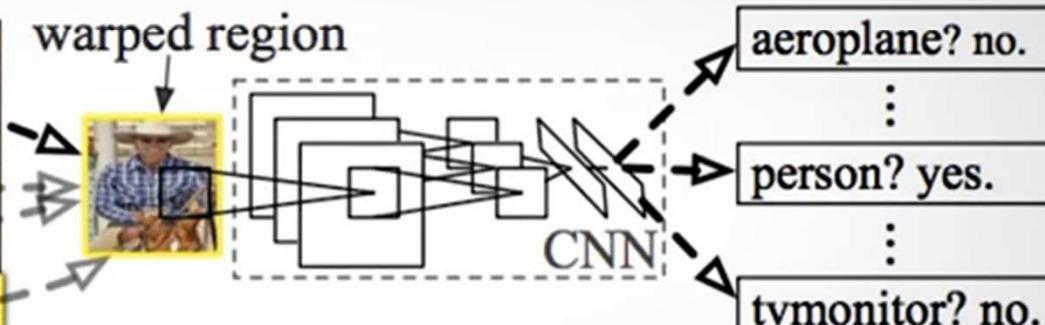
RCNN and Faster RCNN



1. Input
image



2. Extract region
proposals (~2k)

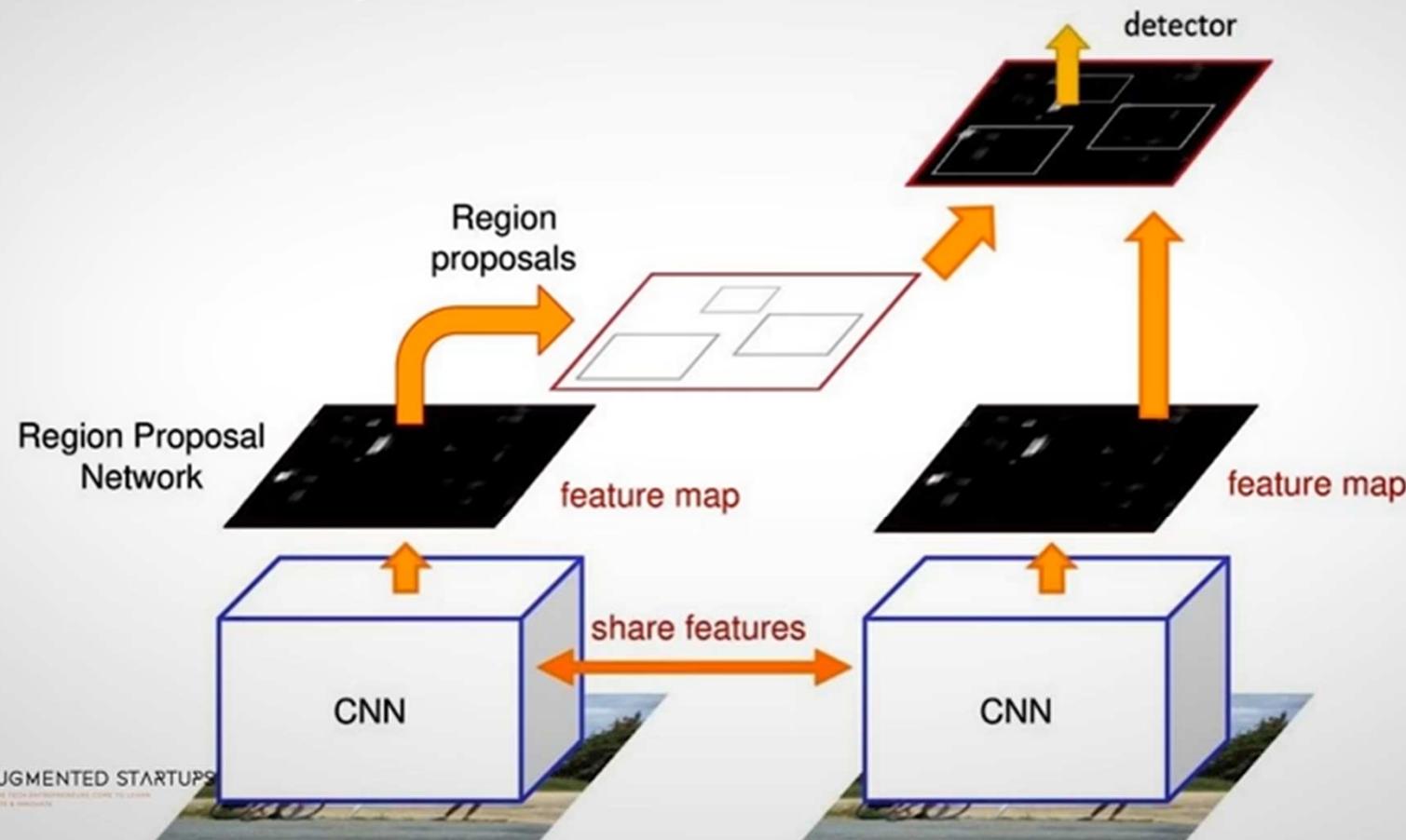


3. Compute
CNN features

4. Classify
regions

Region Proposal Network

Region Proposal Network



RCNN Performance

Accurate

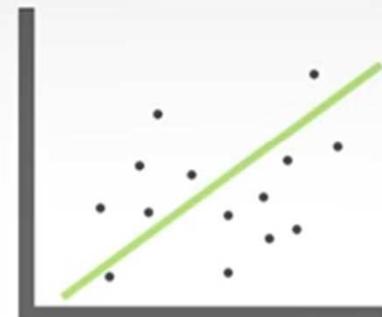


Slow

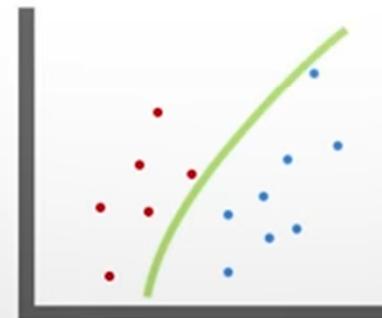


So How is YOLO different

So How is YOLO different?



REGRESSION



CLASSIFICATION

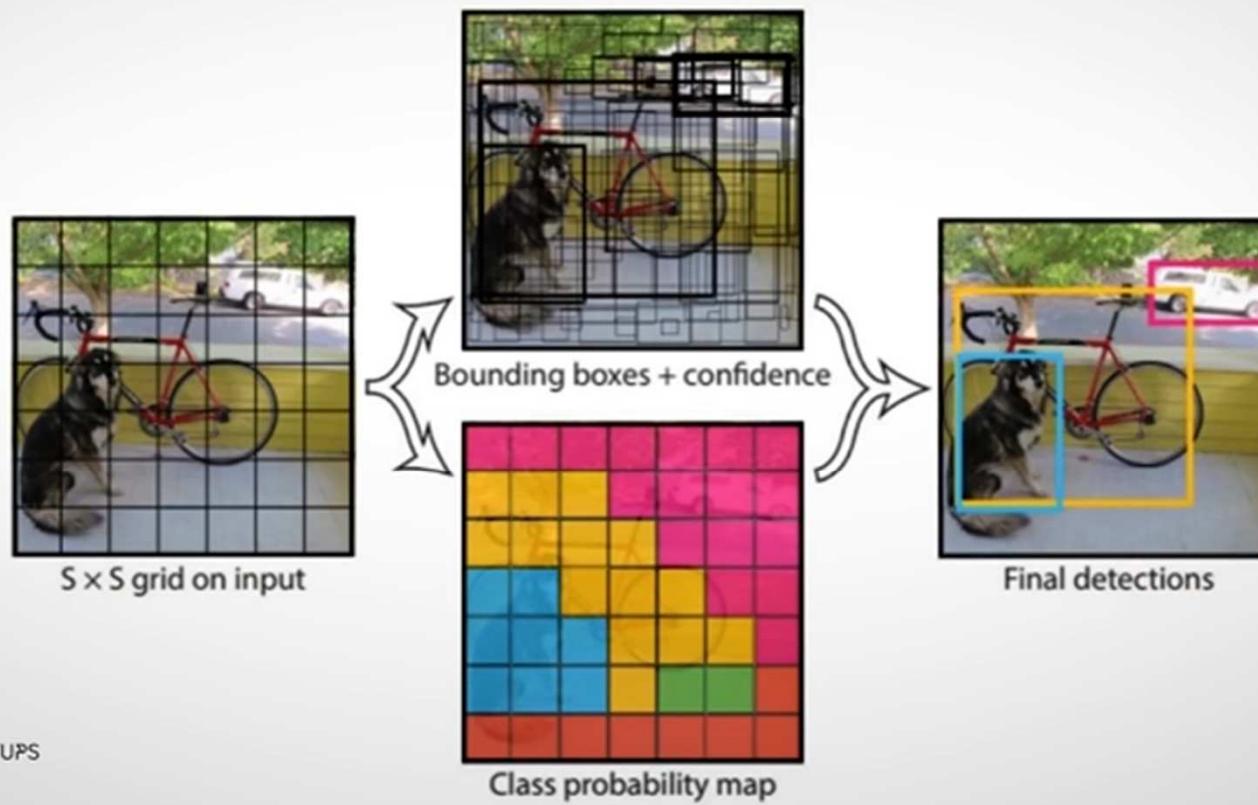


AUGMENTED STARTUPS
WHERE ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE



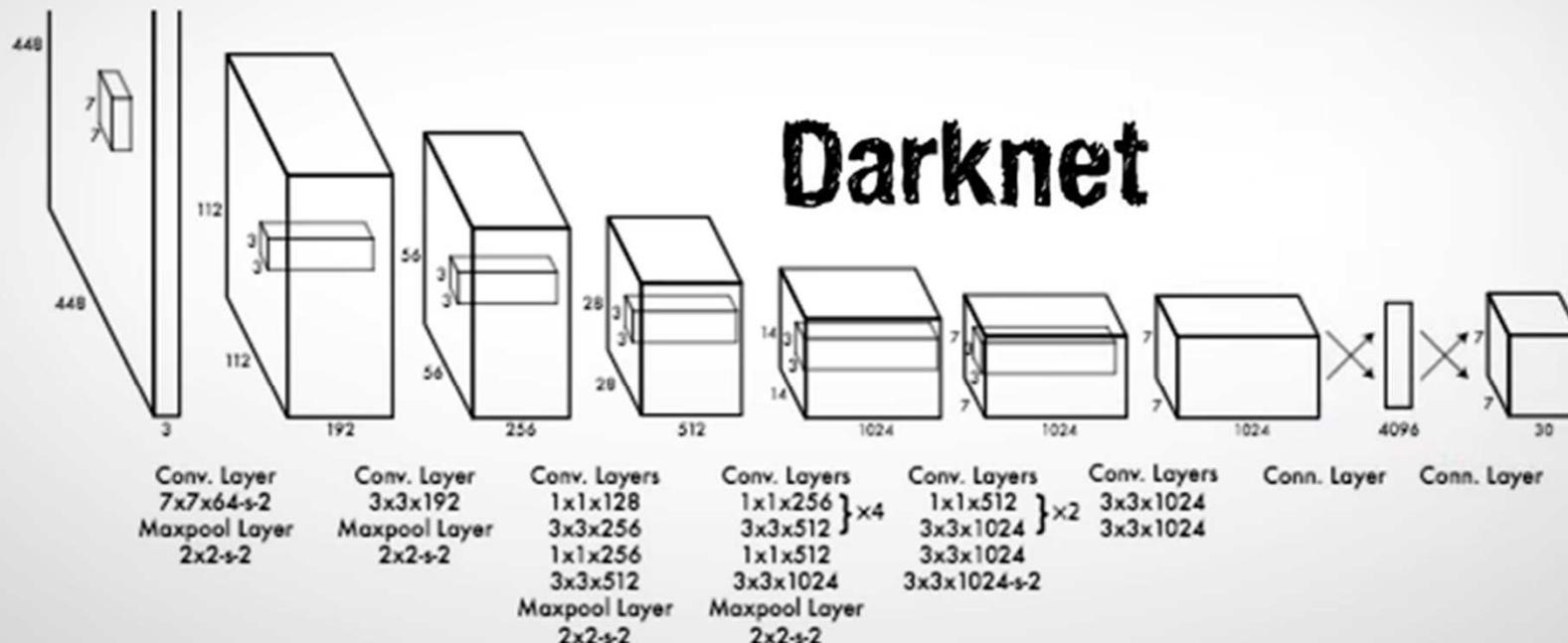
How Does it Work

How Does it Work



YOLO Architecture

Yolo Architecture



AUGMENTED STARTUPS
INSPIRE TECH ENTREPRENEURS, CODE TO LEARN
CREATE & INNOVATE

Limitations

- i - The model imposes strong spatial constraints on the bounding box predictions.
- ii - The model samples down the input image to an SxS grid.
- iii - Difficulty localizing small objects .



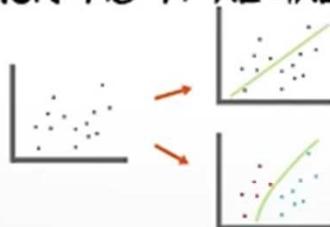
RECAP

RECAP

1 YOLO CONSISTS OF A SINGLE CNN WHICH MAKES IT VERY FAST DURING INFERENCE



2 IT TREATS OBJECT DETECTION AS A REGRESSION PROBLEM INSTEAD OF A CLASSIFICATION PROBLEM



3 IT LOOKS AT THE ENTIRE INPUT IMAGE TO LEARN GLOBALLY RATHER THAN LOCALLY



4 INPUT IMAGE NEEDS TO HAVE 448X448 DIMENSION

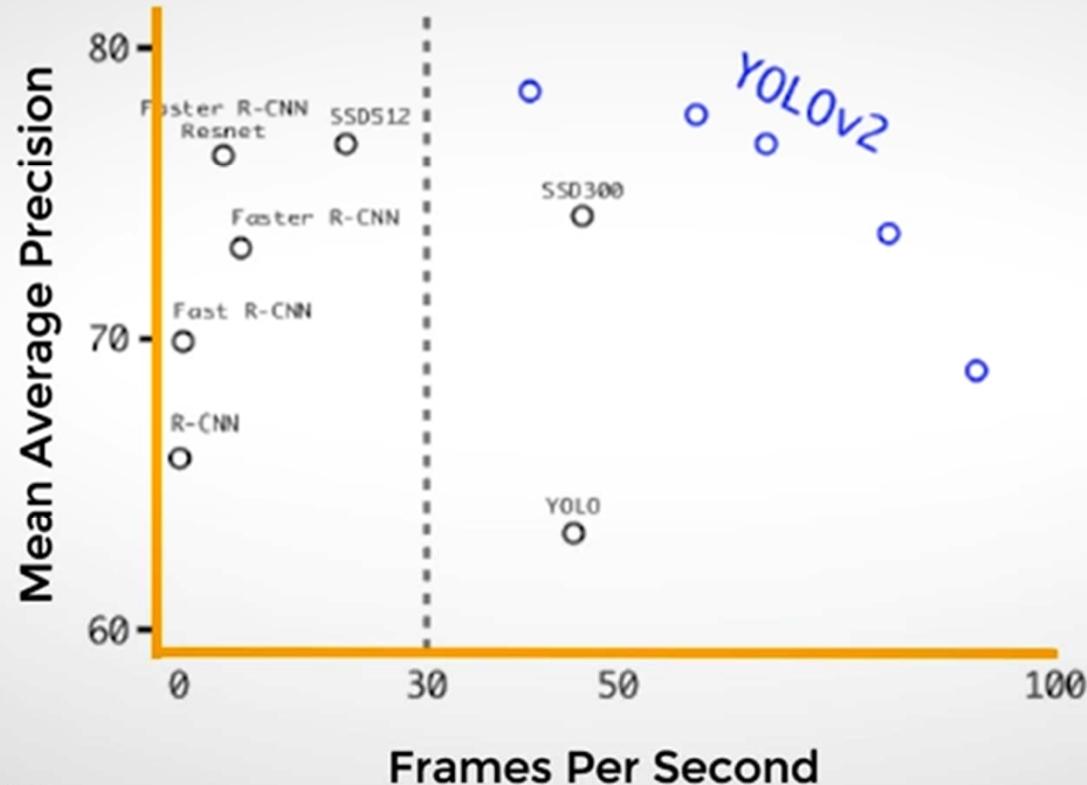


YOLO V2

The slide has a light gray background with a dark blue header bar at the top containing the title 'YOLO V2'. Below the title is a large, stylized black font 'YOLO V2' with the subtitle 'AKA YOLO 9000' in smaller black font to its right. In the center is a black and white portrait of a smiling man with a beard, identified as 'Joseph Redmon'. Below his name is the logo for 'AUGMENTED STARTUPS' which features a stylized lightbulb icon and the text 'AUGMENTED STARTUPS' with the tagline 'WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE' underneath. To the right of the portrait is a dark blue square graphic with the words 'BETTER', 'FASTER', and 'STRONGER' stacked vertically in red, hand-drawn style font. Each word is partially obscured by three horizontal yellow bars. A small red logo consisting of two interlocking shapes is located at the bottom right corner of the dark blue square.

Accuracy and Speed on VOC 2007

Accuracy and Speed on VOC 2007



YOLO V2 DESIGN DESCISIONS

YOLO V2 DESIGN DESCISIONS

	YOLO								YOLOv2
batch norm?	✓	✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?		✓	✓	✓	✓	✓	✓	✓	✓
convolutional?			✓	✓	✓	✓	✓	✓	✓
anchor boxes?				✓	✓				
new network?					✓	✓	✓	✓	✓
dimension priors?						✓	✓	✓	✓
location prediction?						✓	✓	✓	✓
passthrough?							✓	✓	✓
multi-scale?								✓	✓
hi-res detector?									✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	78.6

YOLO V2 Architecture

Yolo V2 Architecture

Darknet 19

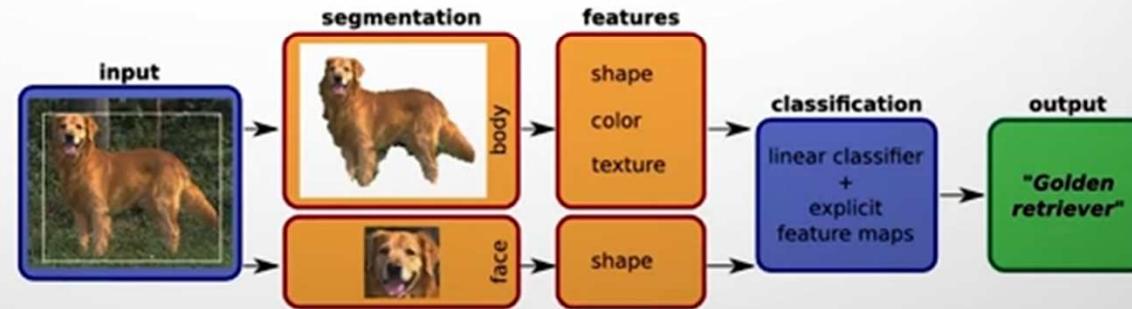
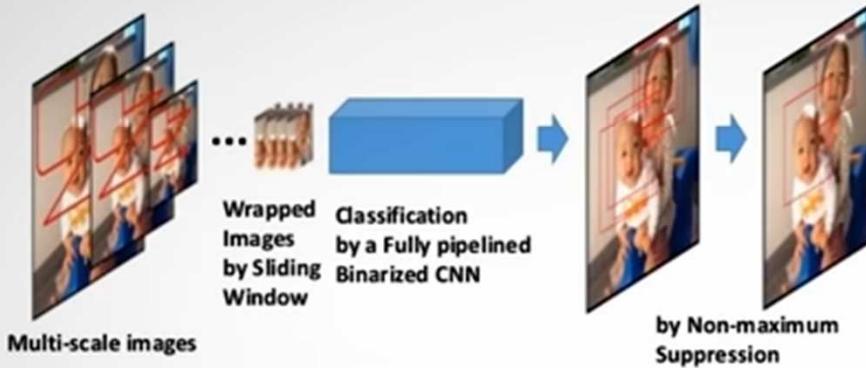


AUGMENTED STARTUPS
MAKING TECH ENTREPRENEURS COME TO LIFE
CREATE & INNOVATE

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

Comparison

Comparison

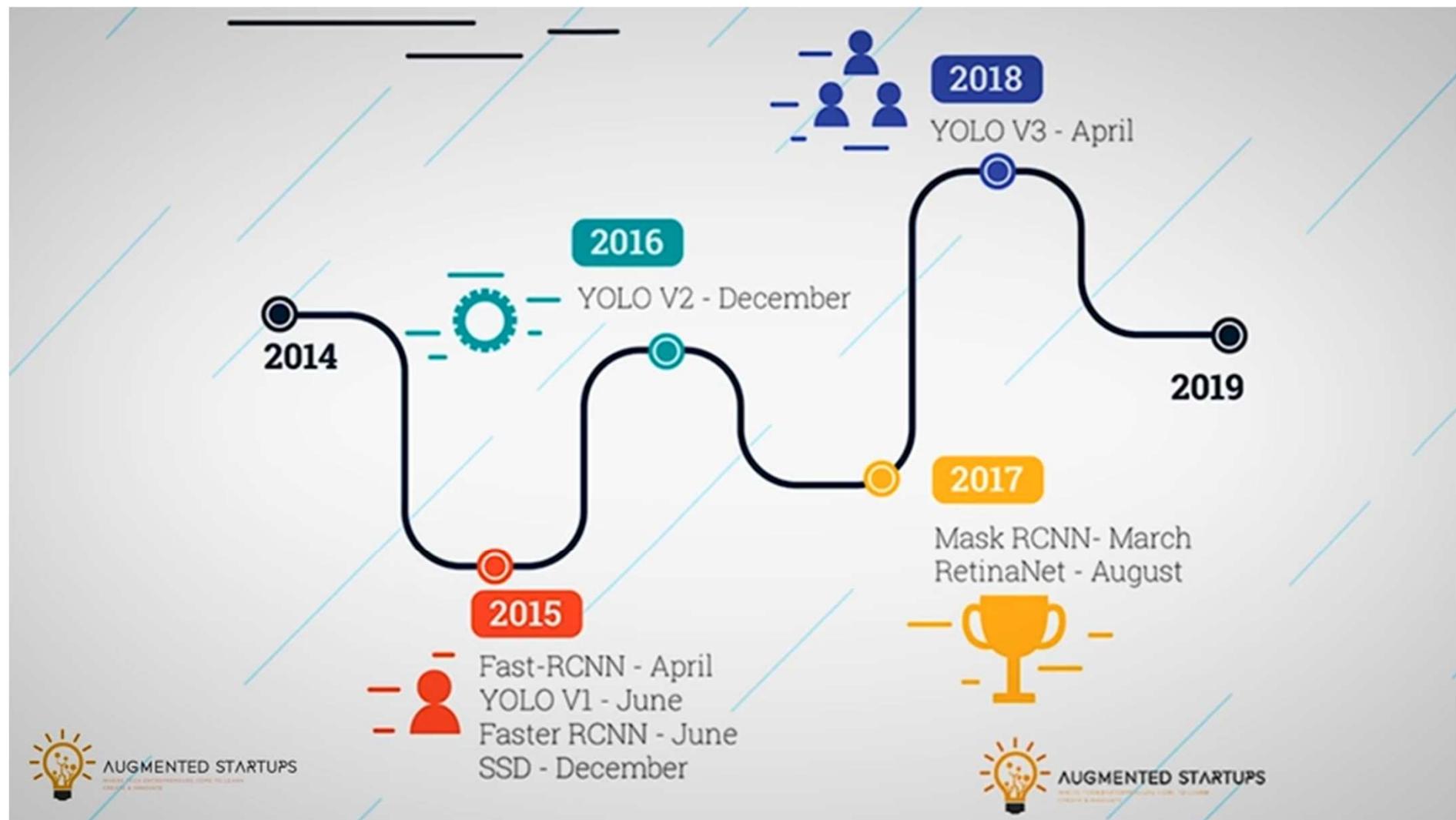


AUGMENTED STARTUPS
INSPIRE | TEACH | ENTREPRENEURS | LEARN TO LEARN
CREATE & INNOVATE

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



AUGMENTED STARTUPS
INSPIRE, EDUCATE, ENTREPRENEURSHIP, INNOVATION, LEADERSHIP,
CREATE & INNOVATE



comparison – table

	backbone	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
<i>Two-stage methods</i>							
Faster R-CNN+++ [3]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [6]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [4]	Inception-ResNet-v2 [19]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [18]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
<i>One-stage methods</i>							
YOLOv2 [13]	DarkNet-19 [13]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [9, 2]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [2]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [7]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [7]	ResNeXt-101-FPN	40.8	61.1	44.1	24.1	44.2	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

What are the so-called Incremental Improvements?

1 - Bounding Box Predictions

2 - Class Predictions

3 - Predictions across scales

4 - Feature Extractor



Backbone Comparison

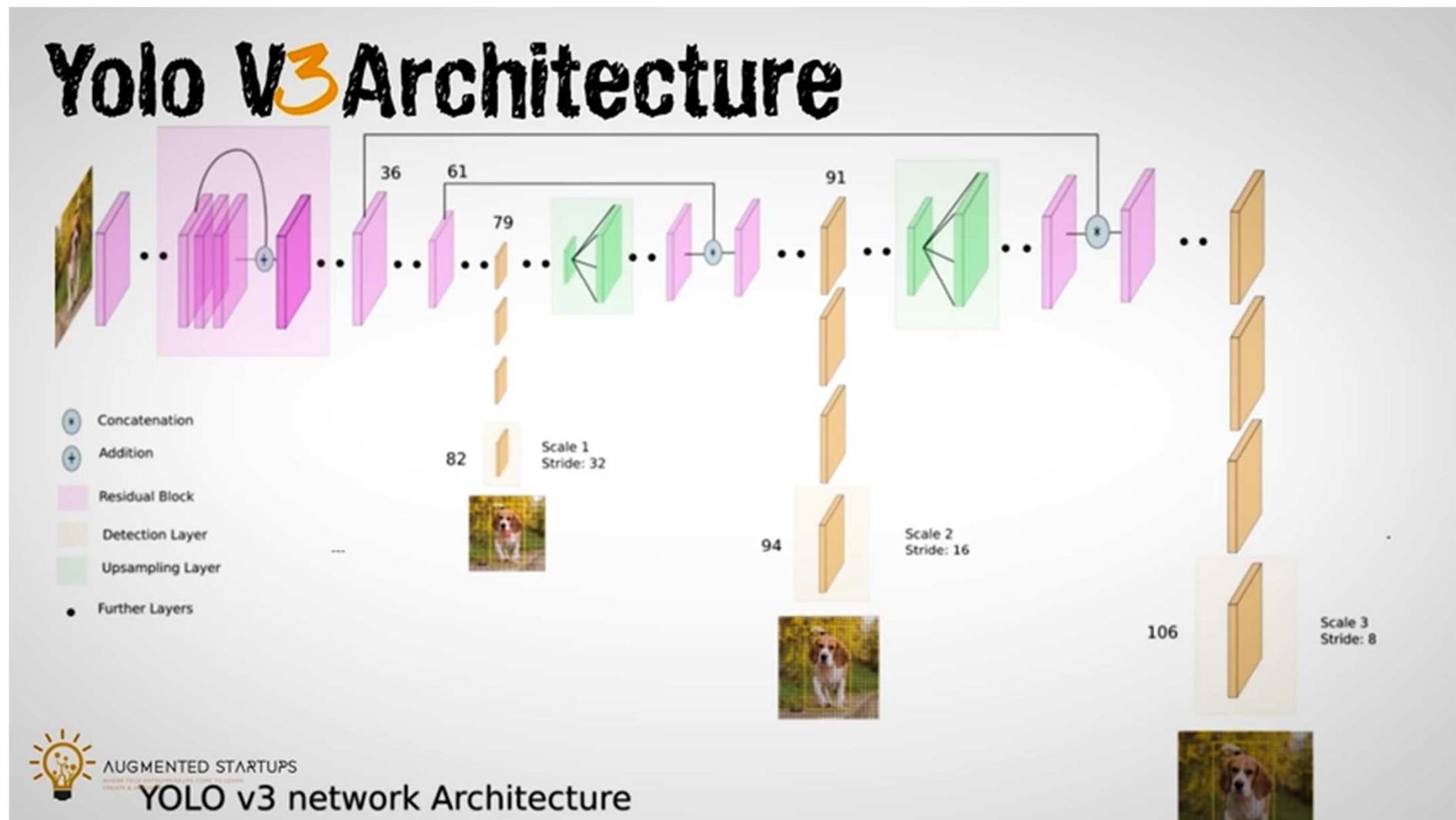
Backbone Comparison

Backbone	Top-1	Top-5	Bn Ops	BFLOP/s	FPS
Darknet-19 [13]	74.1	91.8	7.29	1246	171
ResNet-101[3]	77.1	93.7	19.7	1039	53
ResNet-152 [3]	77.6	93.8	29.4	1090	37
Darknet-53	77.2	93.8	18.7	1457	78

Table 2. **Comparison of backbones.** Accuracy, billions of operations, billion floating point operations per second, and FPS for various networks.



YOLO V3 Architecture

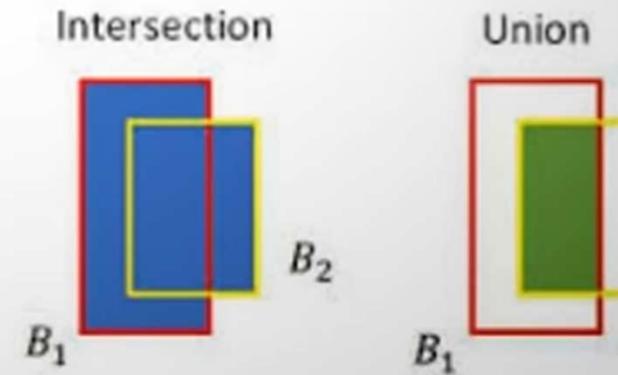


What Improved

What Improved?

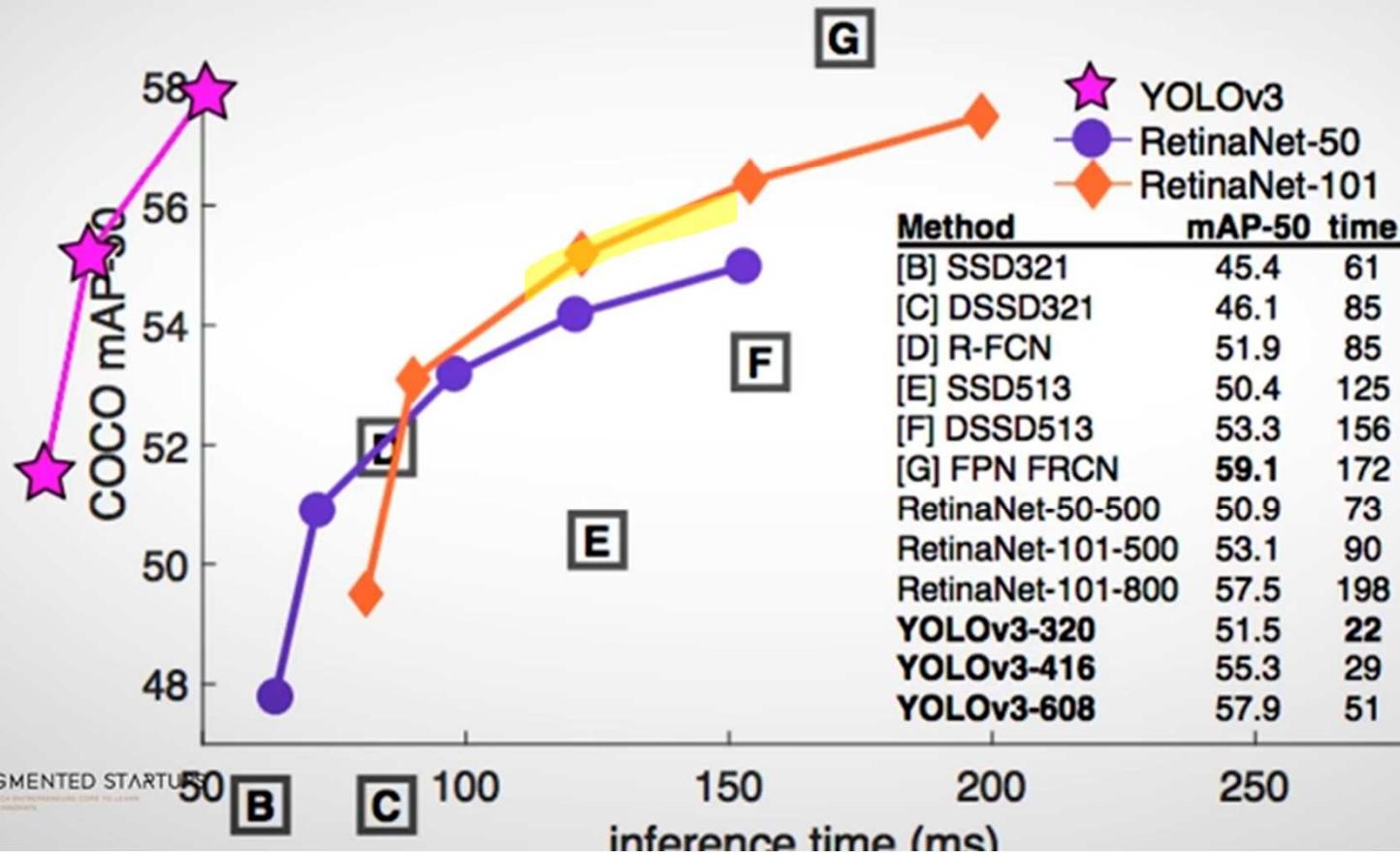


1. The average precision for small objects improved
2. As MAP increased localization errors decreased
3. Predictions at different scales
4. MAP increased significantly



What can be improved?

What can be improved?



AUGMENTED STARTUP
INNOVATE FROM ENTREPRENEURS, LEARN & INNOVATE

References

YOLOv3: An Incremental Improvement

- Joseph Redmon, Ali Farhadi

YOLO v1: Part 2 - Divakar Kapil

YOLOv3: A Huge Improvement - Anand Sonawane



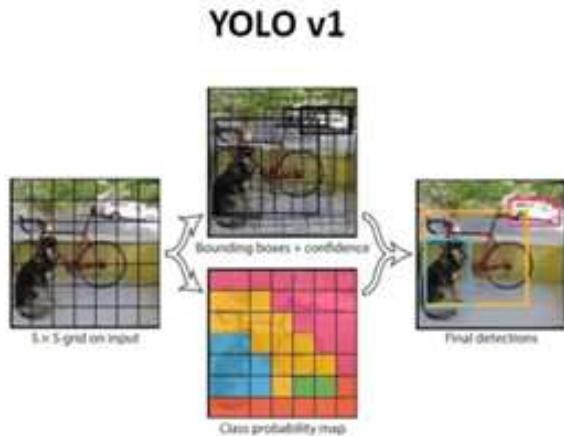
What's new in YOLO v3? - Ayoosh Kathuria



YOLO v1 ~ v3 quick review : YOLO v2

YOLO v1 ~ v3 quick review: YOLO v2

- YOLO v1 + many algorithms



Better

Batch Normalization
High resolution classifier
Anchor boxes
Dimension clusters
Direct location prediction
Fine-grained features
Multi-scale training

Faster

Darknet-19
Transfer learning
○.

Stronger

Hierarchical classification
Dataset combination with Word-tree
Joint classification and detection

YOLO v1 ~ v3 quick review : YOLO v3

YOLO v1 ~ v3 quick review: YOLO v3

- YOLO v2 + many algorithms ([YOLOv3: An Incremental Improvement](#))

YOLO v2



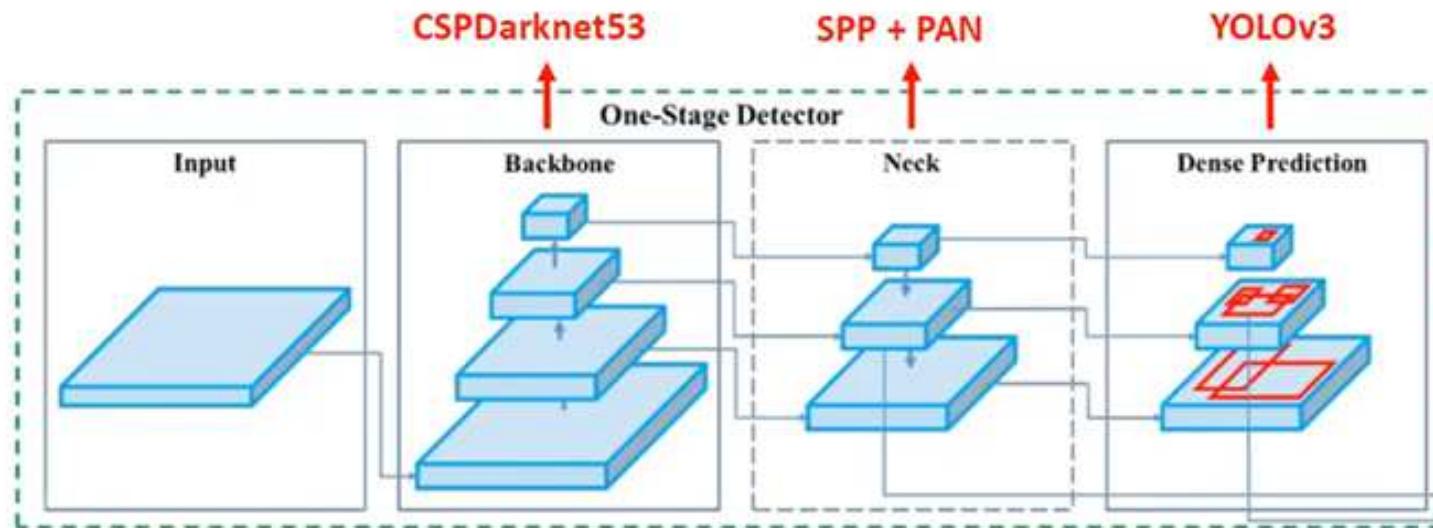
Bounding box prediction → sum of squared loss

Class prediction → Multilabel classification

Predictions across scales

Darknet-53

Selection of architecture



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

↓
Path Aggregation Network