M2Det: Overview

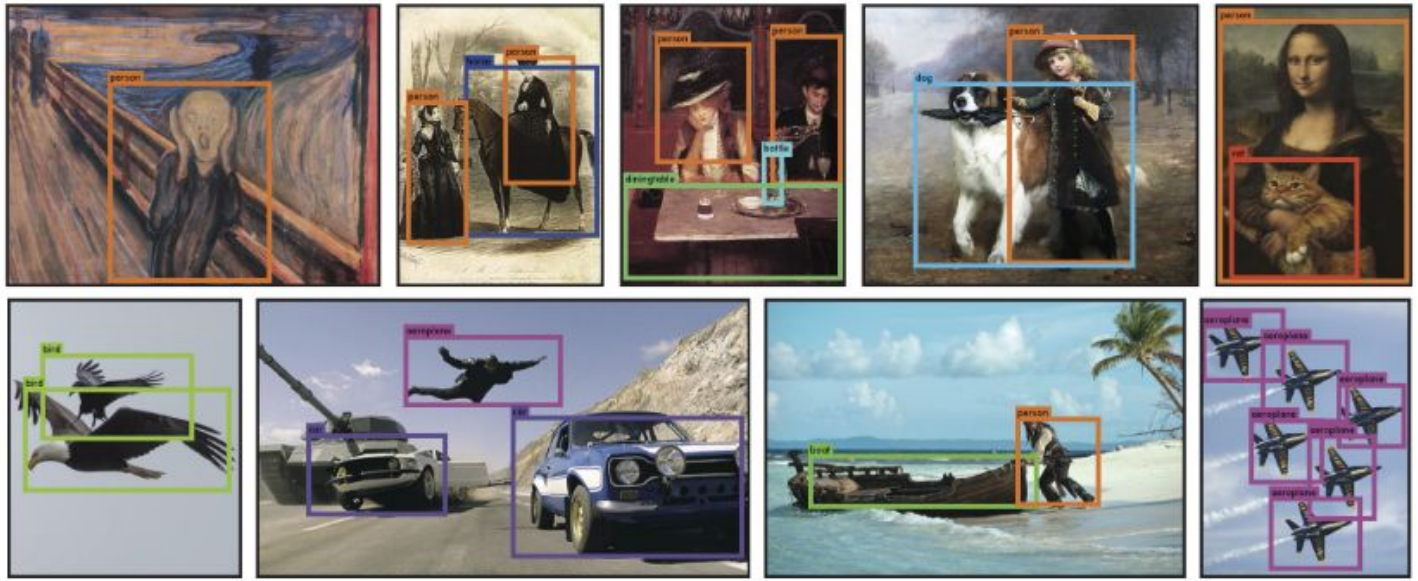
Nikita Toroptsev,
Evocargo, MIPT PSAMCS
Moscow, Jan 2021

telegram: nikitatoropcev

EVOCARGO

girafe ai

Detection problem



girafe ai

my Ache and Pain: mAP

$$Precision = \frac{TP}{\text{total positive results}}$$

$$Recall = \frac{TP}{\text{total cancer cases}}$$

1

2

Rank	Correct?	Precision	Recall
1	True	1.0	0.2
2	True	1.0	0.4
3	False	0.67	0.4
4	False	0.5	0.4
5	False	0.4	0.4
6	True	0.5	0.6
7	True	0.57	0.8
8	False	0.5	0.8
9	False	0.44	0.8
10	True	0.5	1.0

3

4

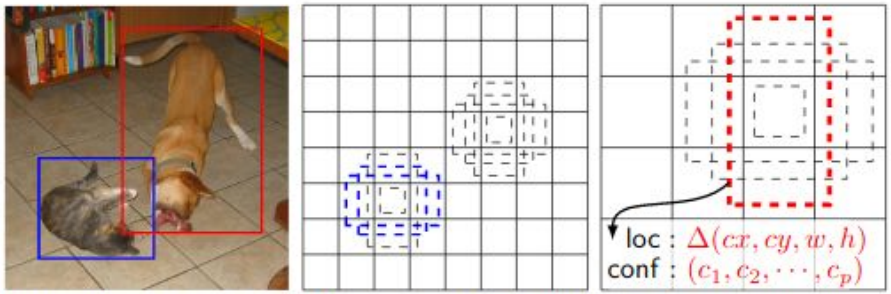
girafe ai

Quick overview

- 2019
- One-pass algorithm
- Exploits FPN in a backbone and uses SSD head
- 41.0 mAP (YOLOv3 - 33)
- 11.8 FPS (YOLOv3 - 20)

SSD - Single Shot Detection

1. Backbone (fully conv)
2. N multi-level detection maps: S×S×N anchors×N classes
3. NMS



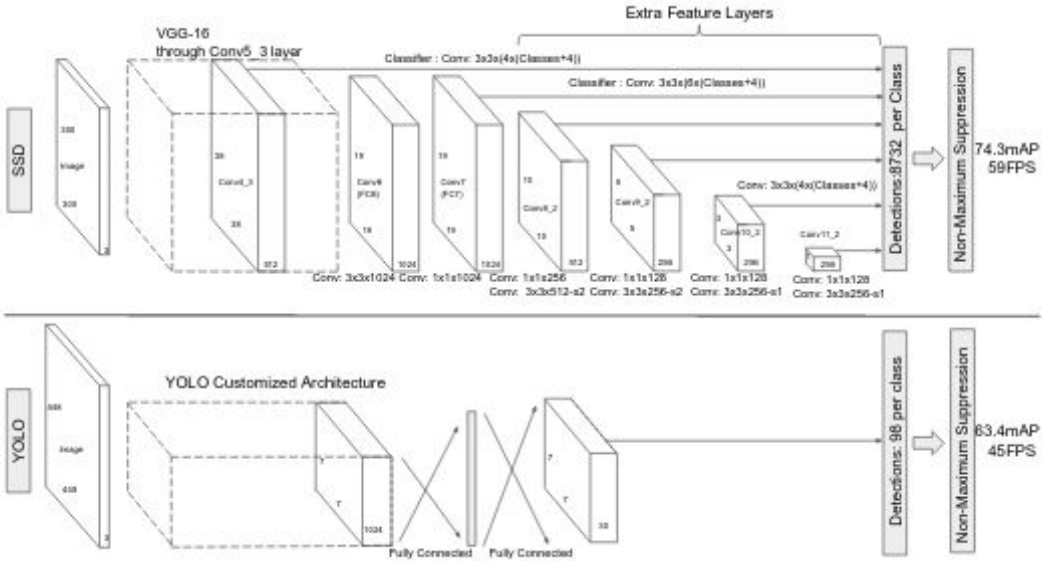
(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

girafe ai

NMS - Non Maximum Suppression

1. Select the max-prob class, c_i - confidence
2. Rank the confidence
3. Confidence threshold
4. For those with the same class label:
   if the IoU is higher than some constant threshold, remove the lower confidence one
5. Continue

SSD vs YOLO (v1)

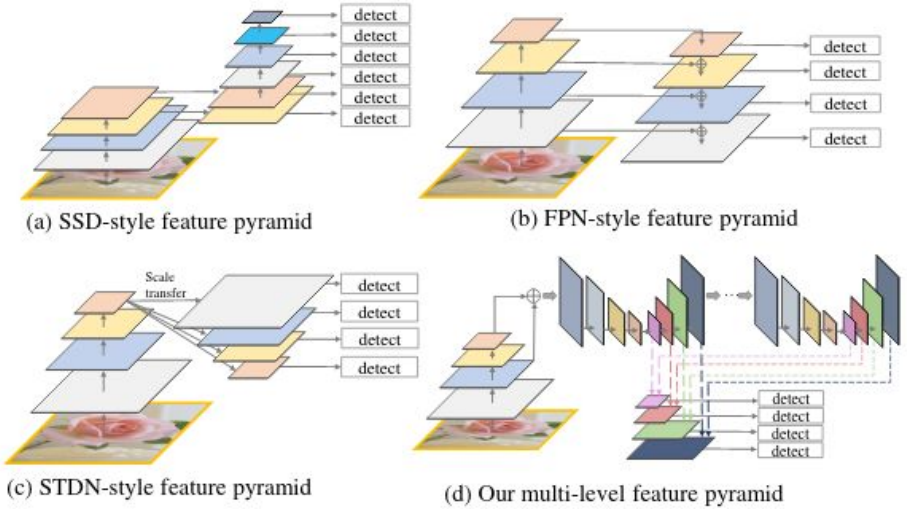


Main Part - Feature Pyramids

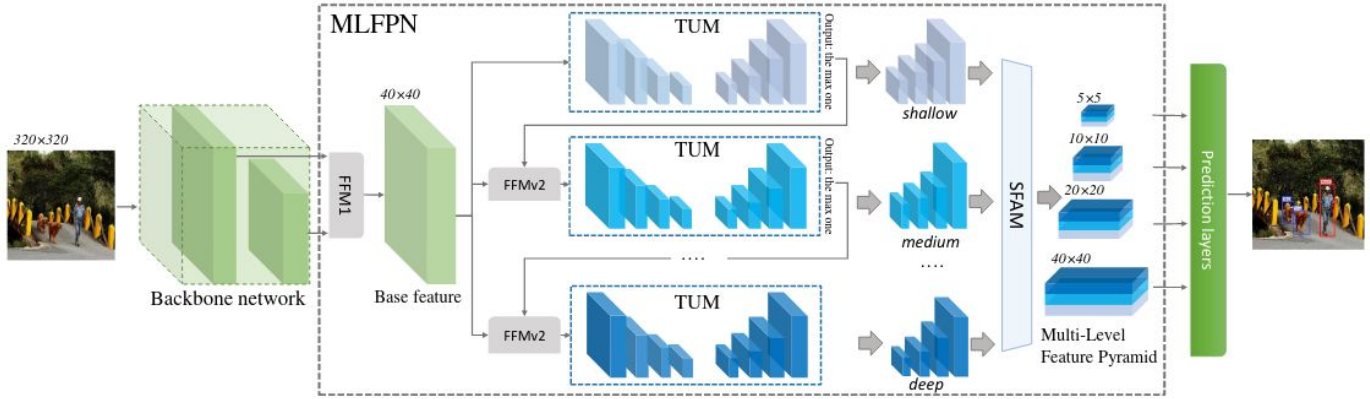
Idea:

- shallow layers have large resolution but less informative representation
- deeper layers are the opposite

Where is the optima? We can let the network choose for us!



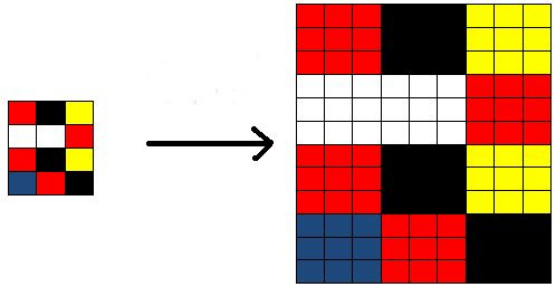
Architecture

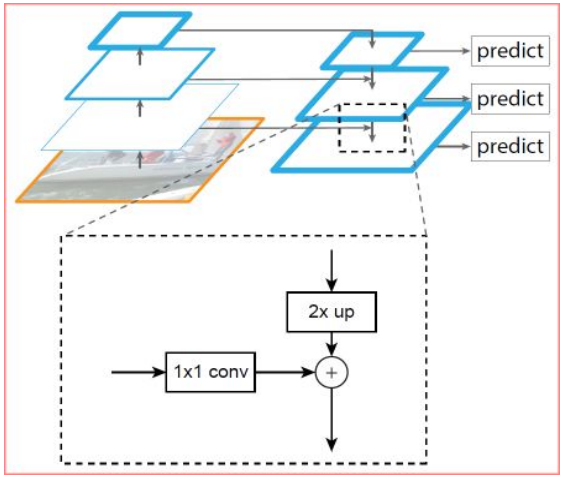


- MLFPN - Multi-Level Feature-Pyramid Network
- FFM - Feature Fusion Module
- TUM - Thinned U-shape Module
- SFAM - Scale-wise Feature Aggregation Module

The Plus

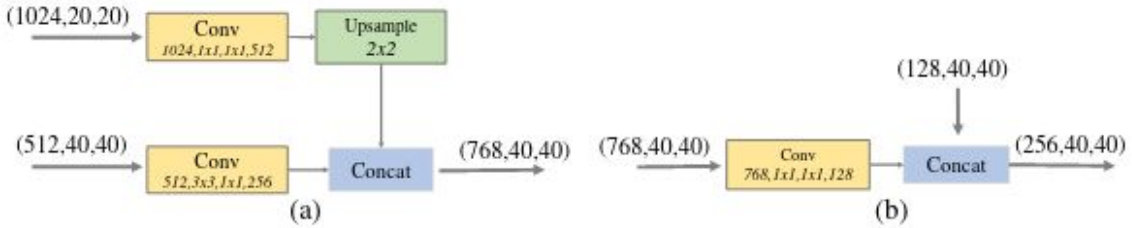
- Upsample
- Element-wise sum
- 1x1 kernel convolution





Feature Fusion Module

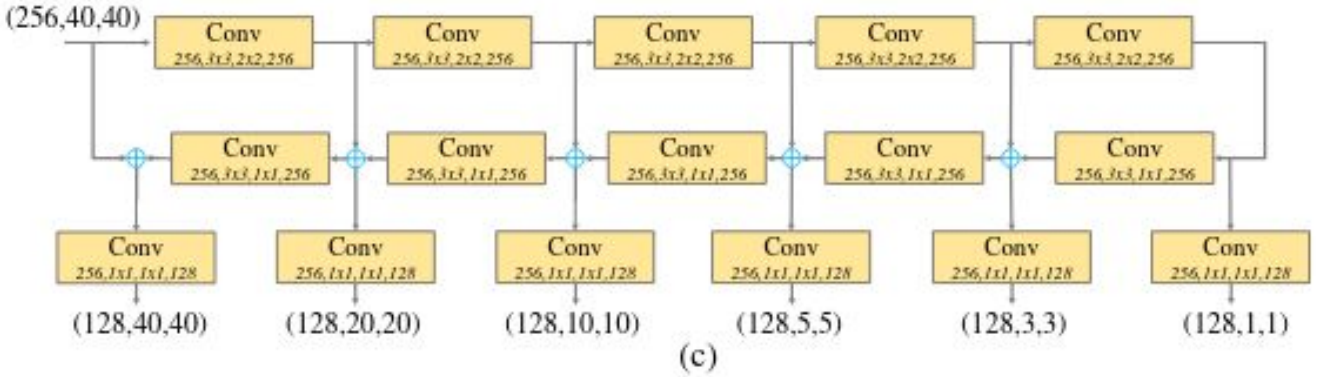
Simple feature fusion module for two maps. Two versions are combined in the network



girafe ai

~~U-Net~~ Thinned U-shaped Module

U-Shaped feature pyramid

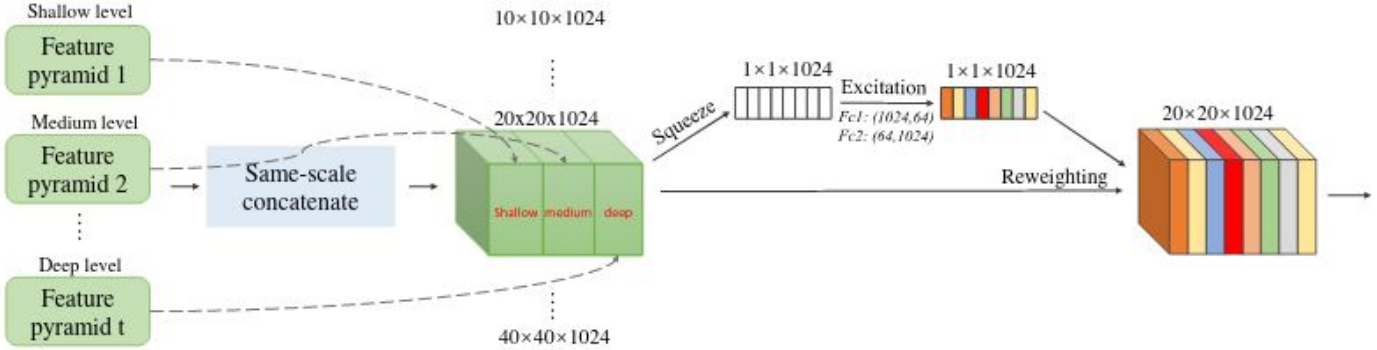


(c)

~~Attention~~

Scale-wise Feature Aggregation Module

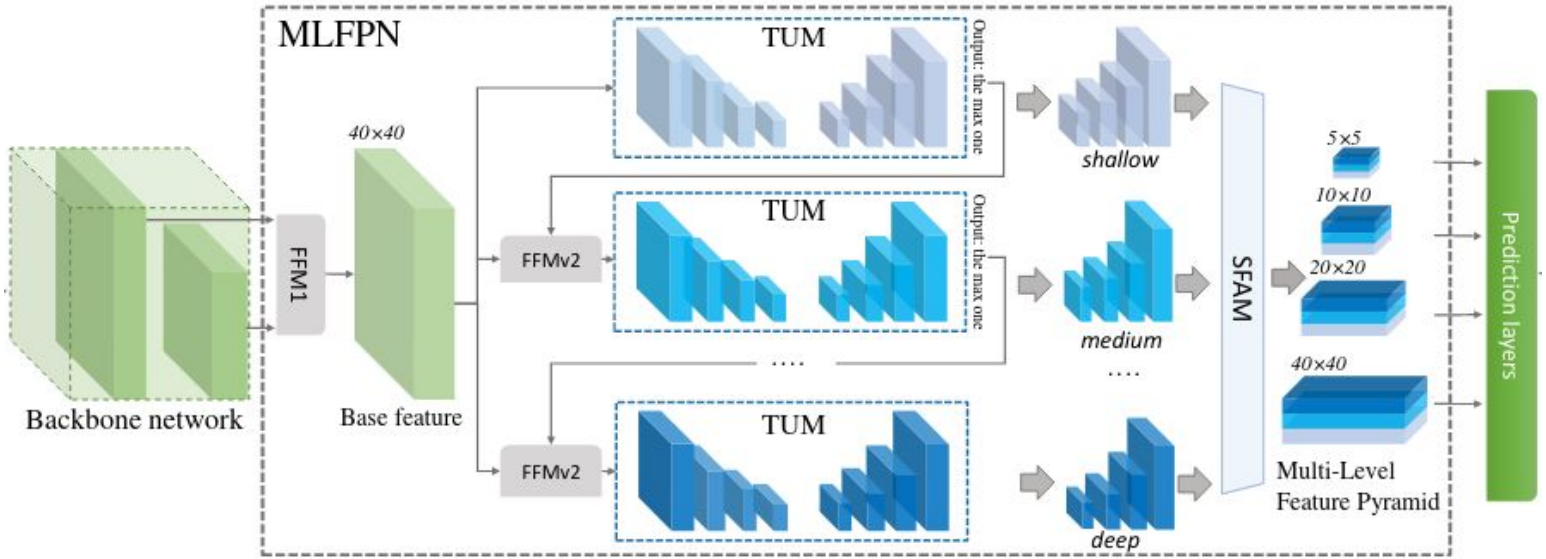
Excitation: learn the weights to enhance or weaken features



$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z})), \qquad \tilde{\mathbf{X}}_i^c = \mathbf{F}_{scale}(\mathbf{X}_i^c, s_c) = s_c \cdot \mathbf{X}_i^c,$$

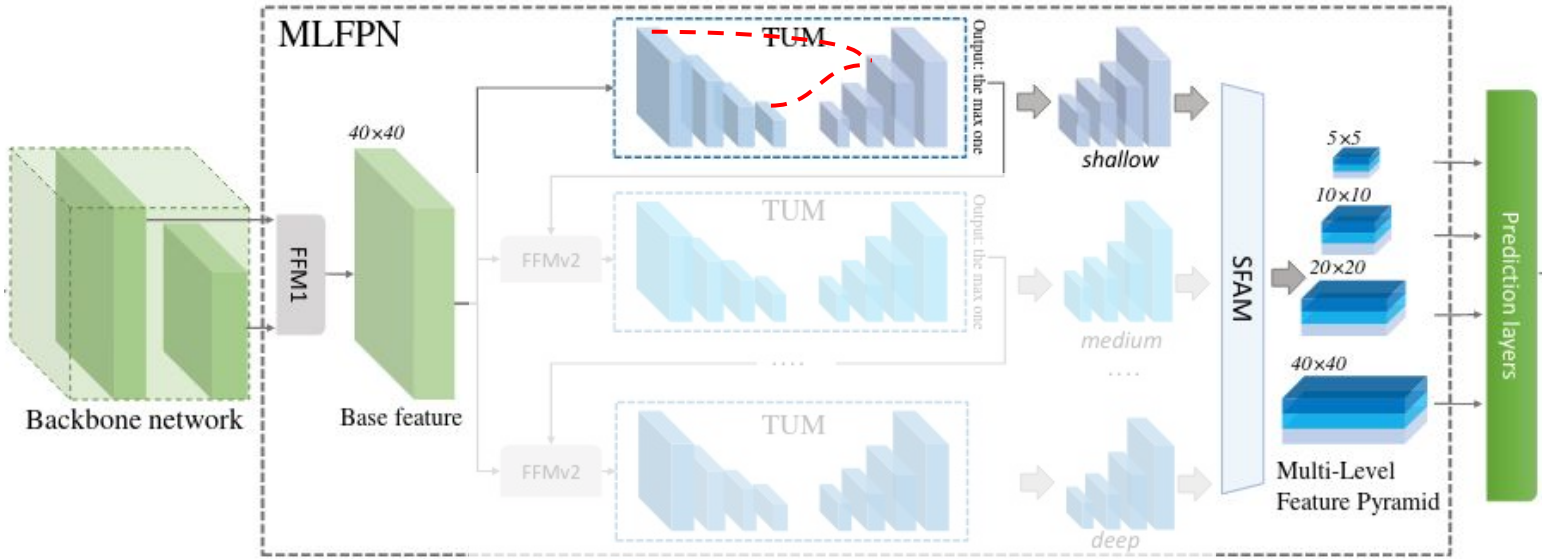
girafe ai

The Final Glance



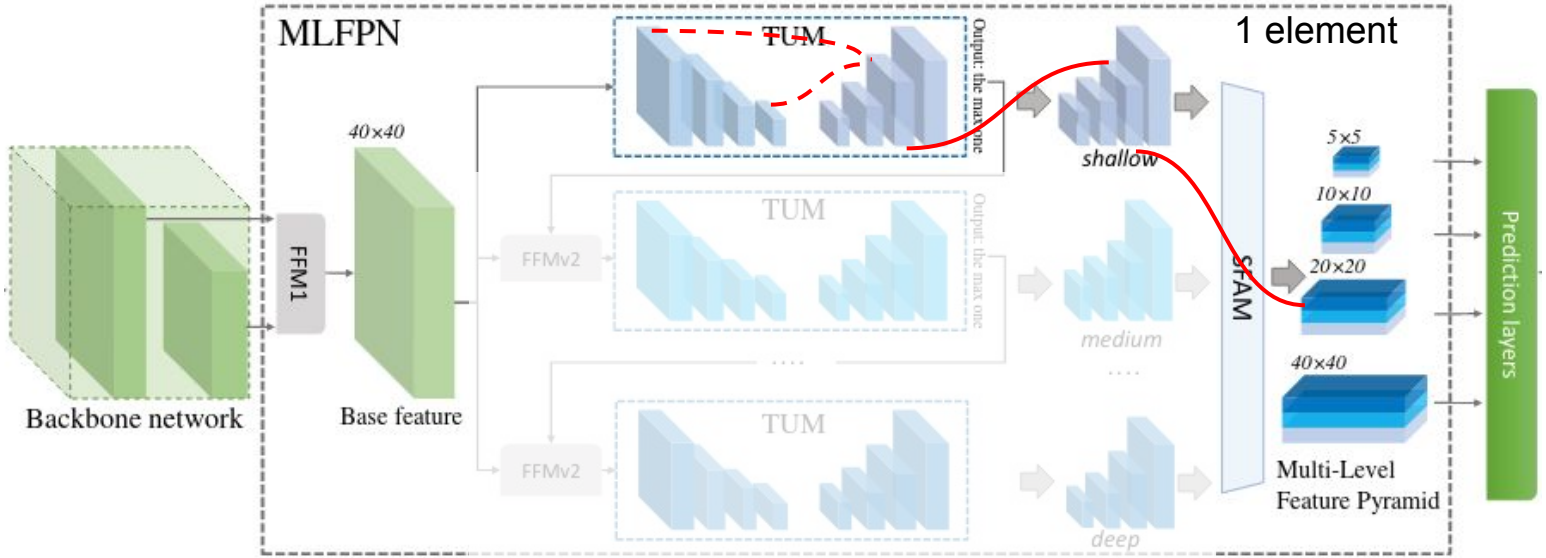
girafe ai

The Final Glance

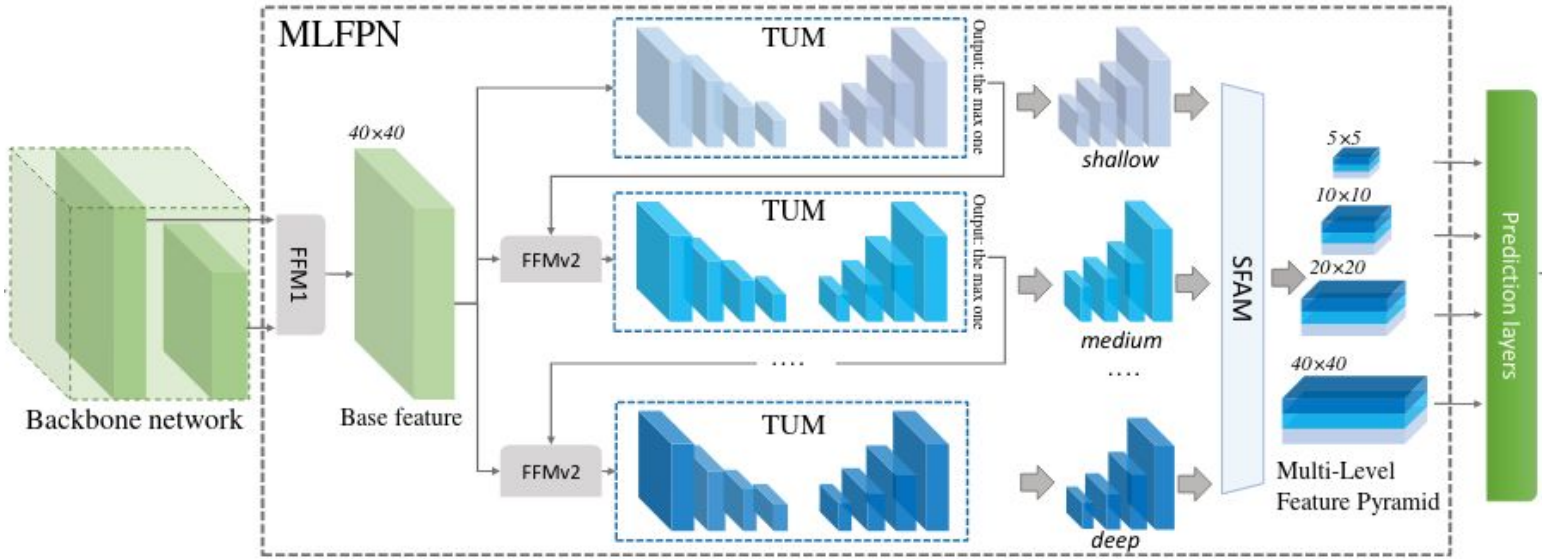


The Final Glance

in this case SFAM is an attention for 1 element

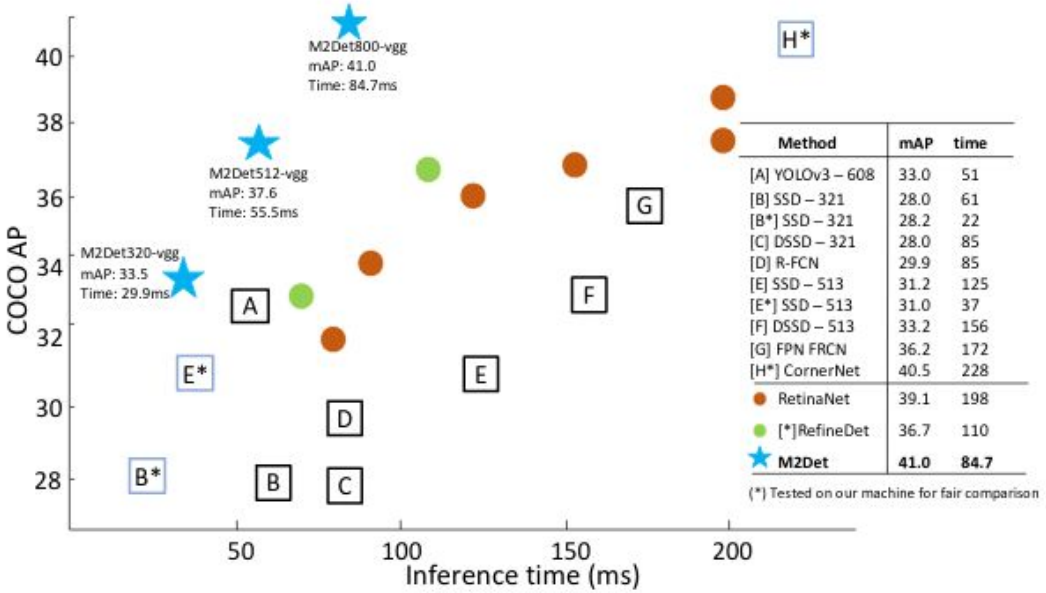


The Final Glance



girafe ai

Summary



Method	mAP	time
[A] YOLOv3 – 608	33.0	51
[B] SSD – 321	28.0	61
[B*] SSD – 321	28.2	22
[C] DSSD – 321	28.0	85
[D] R-FCN	29.9	85
[E] SSD – 513	31.2	125
[E*] SSD – 513	31.0	37
[F] DSSD – 513	33.2	156
[G] FPN FRCN	36.2	172
[H*] CornerNet	40.5	228
RetinaNet	39.1	198
[*]RefineDet	36.7	110
M2Det	**41.0**	**84.7**

(*) Tested on our machine for fair comparison

END
OF
PRESENTATION
SLIDE:67
THANK YOU FOR YOUR ATTENTION