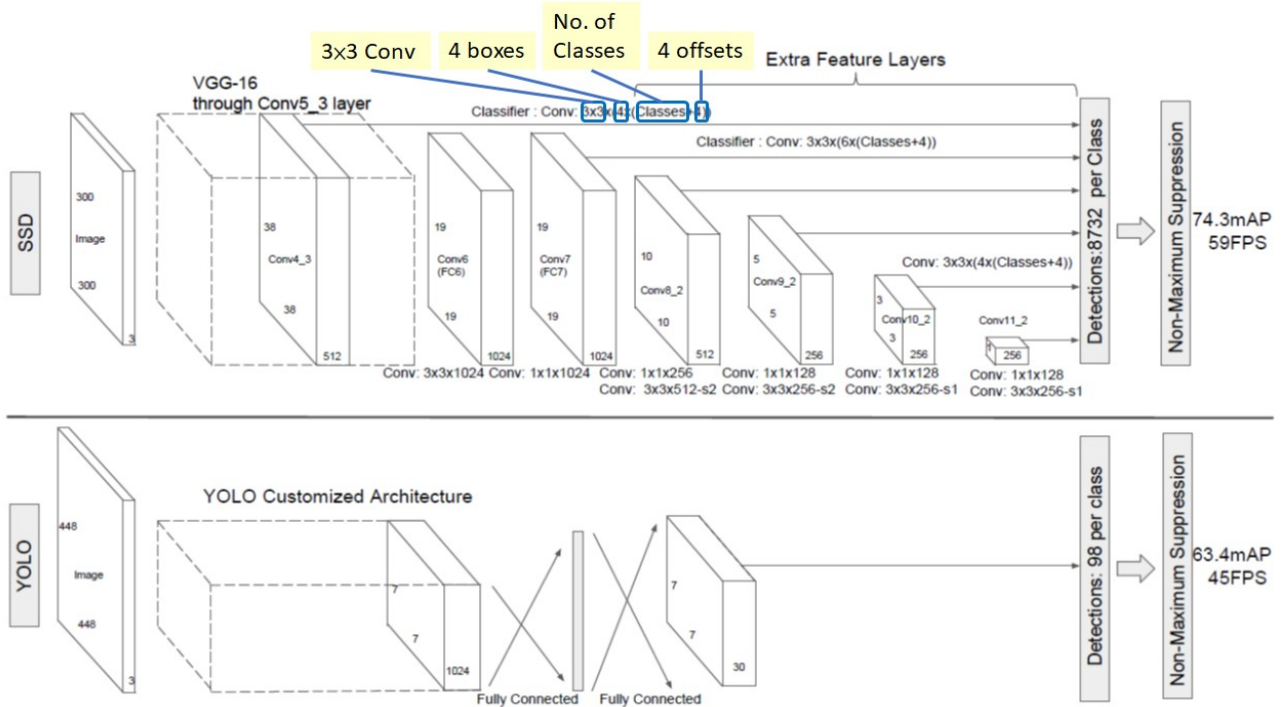


Báo cáo tuần 4: One-Shot Object Detection.

Thực tập sinh: Trần Văn Trung

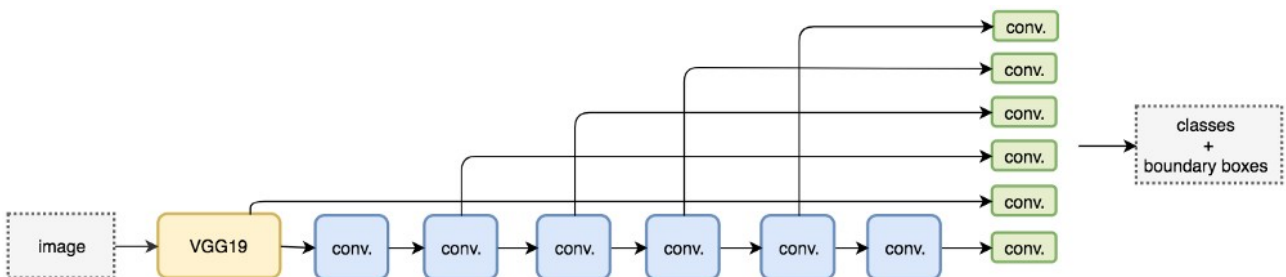
1. SSD vs YOLO.

- Về kiến trúc:



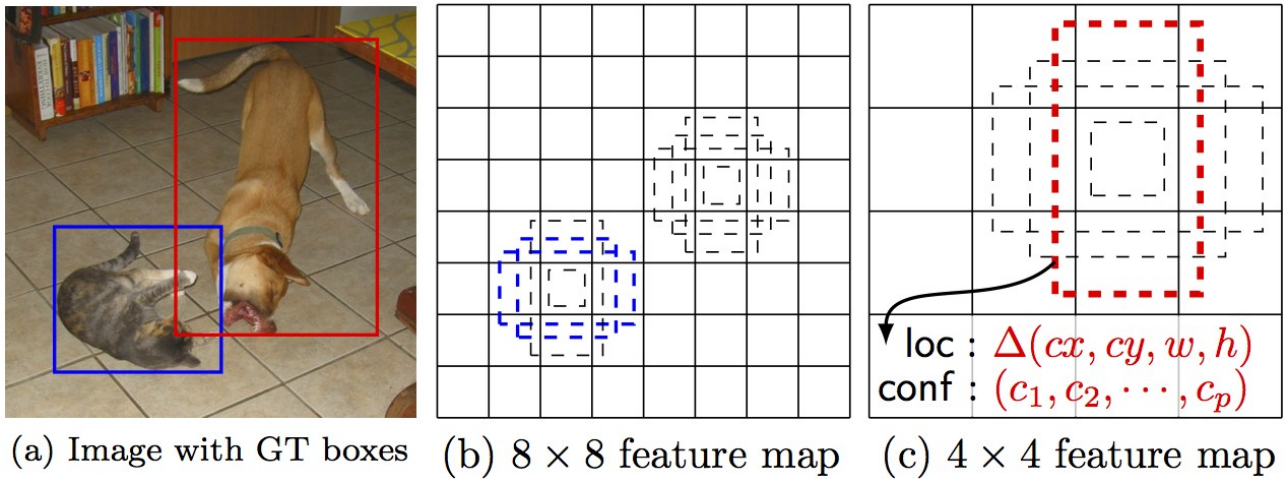
Hình 1.1: Kiến trúc mạng của SSD và YOLO (YOLOv1).

- SSD sử dụng kiến trúc mạng VGG-16 đã lược bỏ tầng FC và thay bằng 6 tầng Conv, giảm dần kích thước đầu vào qua các mạng, nhận biết được các tính chất của ảnh ở nhiều tỉ lệ khác nhau, có thể detect được những đối tượng với nhiều kích thước khác nhau, có hiệu quả với ảnh có độ phân giải cao.



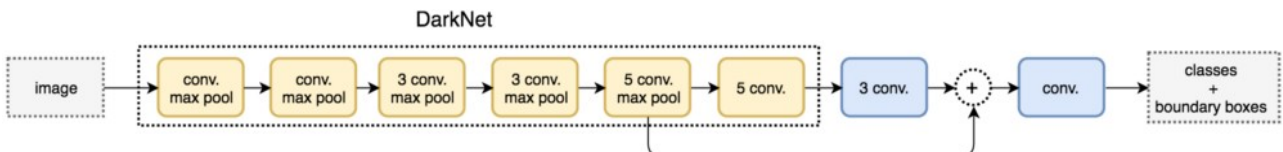
Hình 1.2: Kiến trúc SSD.

- Cải tiến của SSD là sử dụng MultiBox để xây dựng các features map, sử dụng nhiều Priors (các bounding box có kích thước cố định) cho một cell trong features map. Features map có tỉ lệ thay đổi theo các lớp mạng, để có thể nhận dạng được các đối tượng có kích thước khác nhau. Dựa trên IoU để xác định bounding box nào tốt nhất.



Hình 1.3: Sử dụng các features map có tỉ lệ khác nhau để nhận dạng các đối tượng có kích thước khác nhau

- YOLO sử dụng kiến trúc mạng DarkNet, không sử dụng Multiple Scale Features Map như SSD. YOLO thực hiện phẳng hóa output của Conv Layer và kết hợp với Conv Layer trong DarkNet để tạo thành features map. Features map này được dùng để dự đoán.



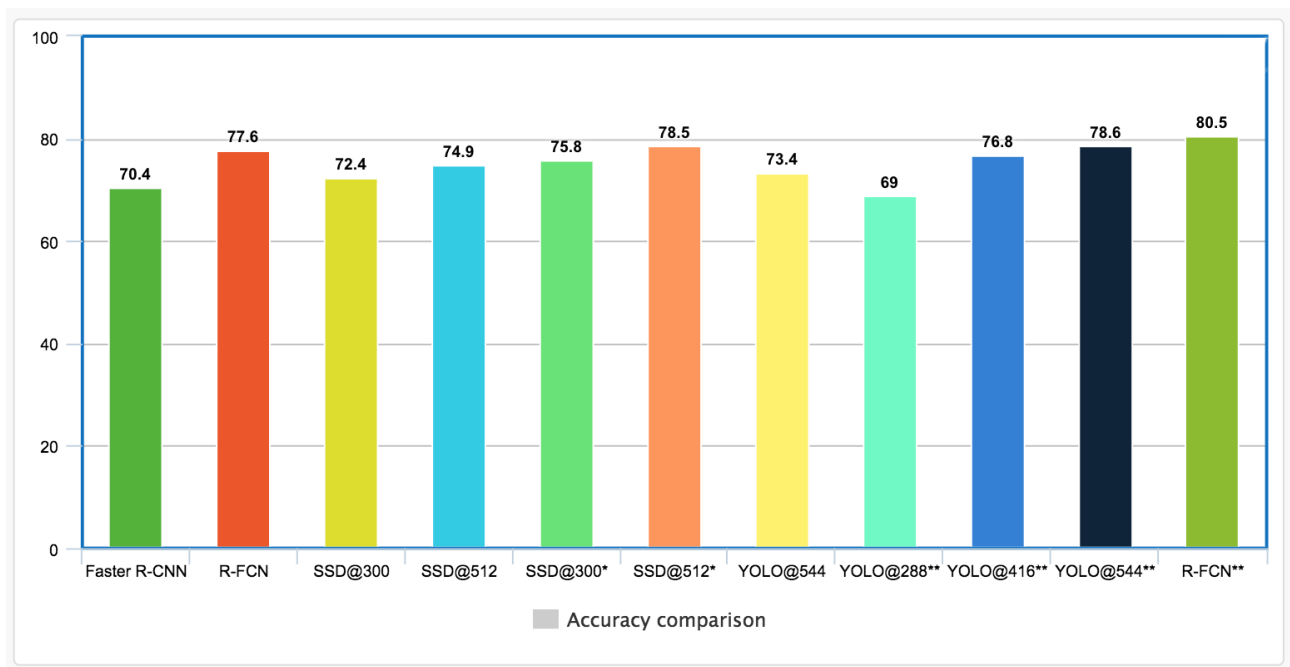
Hình 1.4: Kiến trúc YOLO.

- Về hiệu quả:

	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

Hình 1.4: So sánh trên tập dữ liệu COCO.

- SSD513 và DSSD513 có hiệu quả cao hơn so với YOLOv2, và kết quả khá tương đồng với YOLOv3.



Hình 1.5: So sánh về các phiên bản cải tiến của SSD và YOLO.

2. Implement bài toán nhận dạng hình khối.

- Epochs = 10.
- Lưu các kết quả training để sử dụng lại do Colab bị kết thúc runtime khi training.