

YOLOv3

: An Incremental Improvement (2018)

2023.01.18

The Deal : Bounding Box Prediction

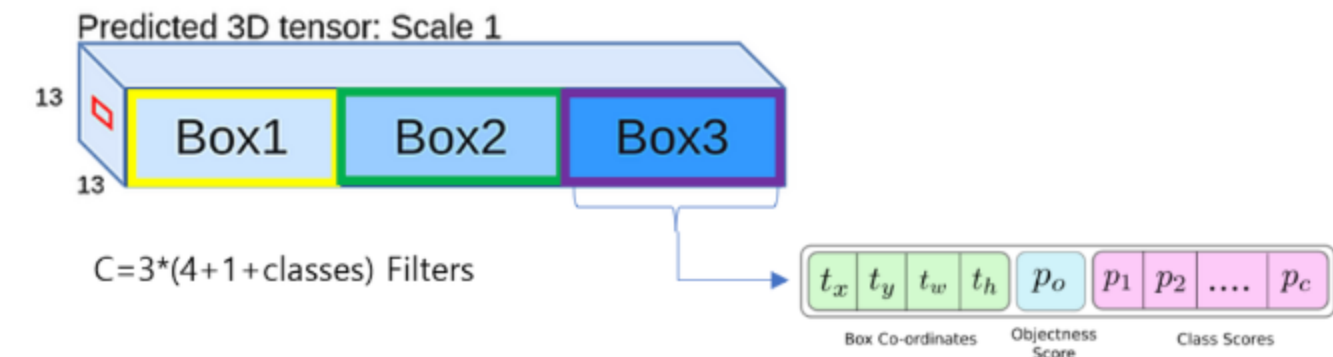
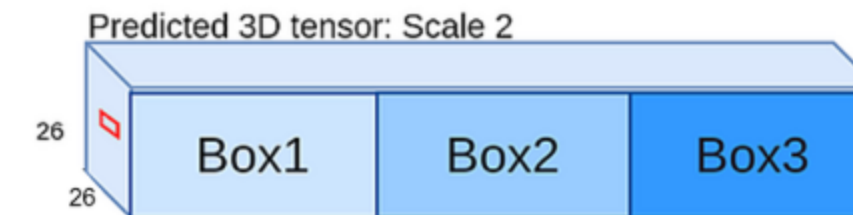
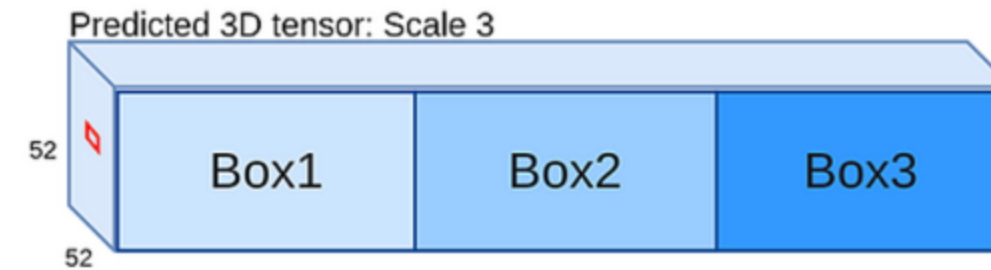
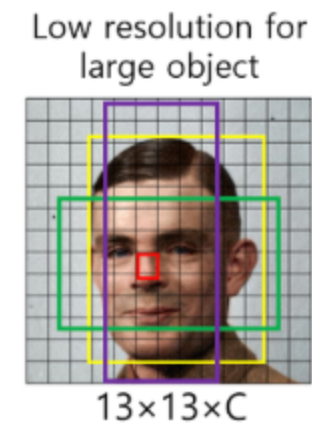
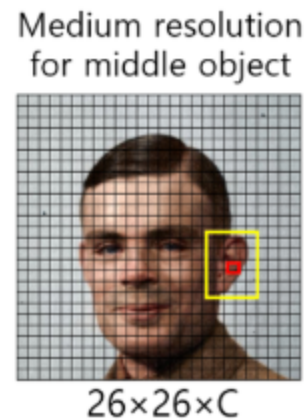
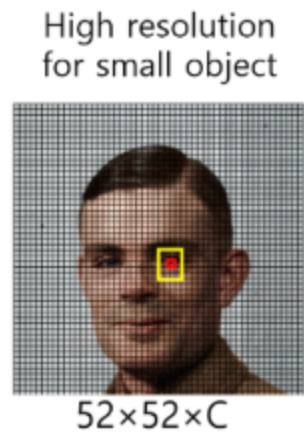
- anchor box 역할을 하는 dimension cluster들을 이용해 bounding box들을 예측
- 기존식의 역함수를 구하여 L1 loss를 구함.
 - ⇒ Why? v2와는 달리 각 bounding box에 대한 score를 logistic regression으로 측정
- v3에서는 threshold를 0.5로 설정하고 하나의 bounding box에 하나의 GT box만 할당
 - ⇒ 해당 bounding box에 GT가 할당되지 않는다면 Objectness에 대한 손실 값만 적용

The Deal : Class Prediction

- v3의 bounding box에 대한 class값 예측은 multi label classification로 class값 예측
 - multiclass classification이 되도록 class별로 각각 확률을 sigmoid를 취해 binary classification을 수행
- ⇒ Why? Person의 경우 man처럼 hierarchical한 클래스들이 존재하는 경우, 여러 클래스를 동시에 고려

The Deal : Predictions Across Scales

- 3가지 scale에 대해 3개의 bounding box(aspect ratio)를 사용
 - Tensor size : $N \times N \times [3 * (4 + 1 + 80)]$
 - k-means를 이용하여 anchor box의 size 지정
 - scale별로 총 9개의 cluster를 정함
 - Small: 10x13, 16x30, 33x23
 - Medium: 30x61, 62x45, 59x119
 - Large: 116x90, 156x198, 373x326
 - 사용한 bounding box의 수
 - 10,647 boxes
- $\Rightarrow (13 \times 13 \times 3) + (26 \times 26 \times 3) + (52 \times 52 \times 3)$



The Deal : Feature Extractor

- DarkNet 53 사용
- Darknet-53이 속도와 연산량 측면에서 ResNet-152보다 1.5배정도 나은 성능
- GPU를 가장 잘 활용하는 것을 의미하며 빠르고 효율적으로 동작

Backbone	Top-1	Top-5	Bn Ops	BFLOP/s	FPS
Darknet-19 [15]	74.1	91.8	7.29	1246	171
ResNet-101[5]	77.1	93.7	19.7	1039	53
ResNet-152 [5]	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.

	Type	Filters	Size	Output
1x	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	128 × 128
	Convolutional	64	3 × 3	
	Residual			
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	64 × 64
	Convolutional	128	3 × 3	
	Residual			
	Convolutional	256	3 × 3 / 2	32 × 32
8x	Convolutional	128	1 × 1	32 × 32
	Convolutional	256	3 × 3	
	Residual			
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	16 × 16
	Convolutional	512	3 × 3	
	Residual			
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	8 × 8
	Convolutional	1024	3 × 3	
	Residual			
4x	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. **Darknet-53.**

Training

- hard negative mining의 적용 없이 전체 이미지를 학습
- detection시 뽑힌 bounding box중 실제 제대로 탐지된 객체의 갯수가 매우 적어 background class에만 데이터량이 매우 많아 네트워크가 학습되지 않는 현상을 방지하기 위해 적용
- Multi-scale training, data augmentation, batch normalization 등과 같은 일반적인 방법들을 적용시켜 네트워크 학습

How We Do

- COCO의 mAP metric을 기준으로 측정할때 SSD와 비슷하며, 속도적인 측면에서는 3배 더 빠름
- AP50을 사용하게 된다면 RetinaNet과 성능이 비슷해지면서 SSD보다는 훨씬 좋은 성능을 보임
- 새로운 multi-scale 예측으로 인해 굉장히 높은 APs 성능을 보임

	backbone	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
<i>Two-stage methods</i>							
Faster R-CNN+++ [5]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [8]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [6]	Inception-ResNet-v2 [21]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [20]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
<i>One-stage methods</i>							
YOLOv2 [15]	DarkNet-19 [15]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [11, 3]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [3]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [9]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [9]	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

Things We Tried That Didn't Work

- 성능 향상을 위한 다양한 시도
 - 1) Anchor box x, y offsets predictions
 - 2) Linear x, y, predictions instead of logistic regression
 - 3) Focal Loss
 - 4) Dual IoU threshold and truth assignment

What This All Means

- 빠르고 정확하다는 관점에서 좋은 detector
- YOLOv3가 mAP와 AP50 간의 성능차가 심하게 나는 것을 강조
- mAP를 평가 지표로 사용하는 것에 대한 불만 제기

⇒ Why? 사람도 20% 차이나는 IoU gap을 구별X, 50%, 55%, ..., 90% 나눠 평가하고 mAP를 구하는 게 무슨 의미?

Reference

<https://www.youtube.com/watch?v=HMgcvgRrDcA&t=1627s>
(PR-207: YOLOv3: An Incremental Improvement)

<https://www.youtube.com/watch?v=jqykPH3jbic>
([Paper Review] YOLOv3: An Incremental Improvement)

<https://89douner.tistory.com/109>
(13. YOLO V3)

<https://ys-cs17.tistory.com/23>
(YOLOv3: An Incremental Improvement)

<https://ropiens.tistory.com/43>
(YOLOv3 : An Incremental Improvement 리뷰)

감사합니다 :)

2023.01.18