CV - CLASS - 2

Image classification / object - classification

< Object - Localization >

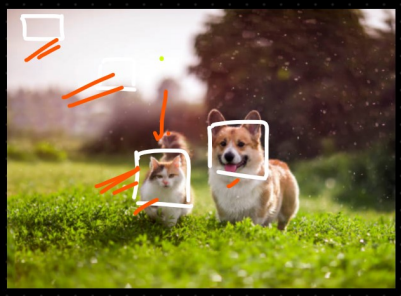
object - Detection (classification + obj localization) (Regression | classification)

1. Sliding window method →

2014-2015

2. R-CNN → Region based Convolution
3. Fast-RCNN
4. Faster-RCNN

	Pascal 2007 mAP	Speed	
DPM v5	33.7	.07 FPS	14 s/img
R-CNN	66.0	.05 FPS	20 s/img
Fast R-CNN	70.0	.5 FPS	2 s/img
Faster R-CNN	73.2	7 FPS	140 ms/img
YOLO	63.4	45 FPS	22 ms/img



5. YOLO ($v_1, v_2, v_3, v_4 \dots v_8$)
   - YOLO tiny
   - YOLO Nano
   - PP-YOLO
   - Scaled-YOLO

6. SSD (Single Shot - multibox Detector)
   - SSD-1, 2, 3
   - SSD - inception
   - SSD - mobilenet
   - DSSD.
   - RefineDet . Efficient-DET

R-CNN , YOLO , SSD

1. FPS → Frame Per Second

Base Arch

Video → it is a collection of frame (img)

model → FPS

- FSR-CNN ⇒ 7 FPS
- YOLO ⇒ 45 FPS
- SSD ⇒ 59 FPS

=  t

SOTA ⇒ Object - Detection (Supervised - Learning)

① training ⇒ calculation, Loss, optimize
② Prediction

Model + Hardware ⇒ Speed

GPU, Edge

(Domain)

① application req.
② Hardware availability
③ trade off BW Accuracy / Speed.

YOLO Speed ↑ Accuracy ↓

fsRCNN Speed ↓ Accuracy ↑

SSD Speed ↑ Accuracy ↓

(Balanced)

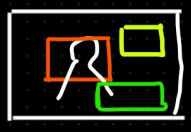
{Medical Domain}
{Self Driving car}
{traffic Surveillance}

① Region Proposal ⇒ DL Model ⇒ locate a region of interest (ROI)

② Feature extraction or network Prediction ⇒

③ NMS (Non - maximum Suppression)

i) Region - Proposal ⇒ NN



$ROI_1 >> ROI_2 >> ROI_3$

?

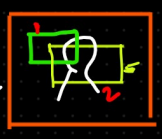
RoI

Region the system believes have a high likeyhood of containing an object

Object score → Object
→ background

Region

Green, Yellow ⇒ Region Proposal → NN → ROI



Green box ⇒ Background. (not RoI)

Second box ⇒ Object. (RoI ②)

② Feature extracting & Network Prediction

Ⓐ RoI. => Con_Nets & NN => classification

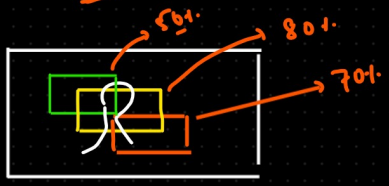
Ⓑ Bounding Box

③ NMS => Non-maximum suppression

- We solve the issue of overlapping Box

White

threshold < confidence threshold



Training

IOU

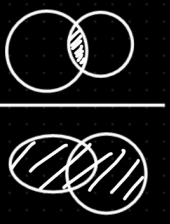
Prediction

{ Annotated.
Ground truth.
Person, {x, y, w, h} }

↓

XML → Rcnn, Yolo, SSD

IOU = area of intersection / area of union =



IOU = =

Confidence threshold

= = 90%

IOU

RCNN, YOLO, SSD ⇒ TFOD2 Detectron2

Masking ⇒ Detectron2

{ tagging, OCR } , (2,3) ⇒ (2,3)

MAP ⇒ (Mean Average Precision)

(Mean of all the Avg. Precision)

for evaluating ⇒ MAP

	Faster-RCNN	YOLO	SSD
① Architecture type	Two-Stage	One-Stage	one-Stage
② Region Proposals	Region Proposal Network (RPN)	NO (grid) region Proposals	No RP
③ Speed	Slower	Faster	Balanced
④ Handling Diff Scales	effective	Limited for the small Obj	Effective
⑤ Object Proposals Bbox.	Variable No. of obj. Proposal Bbox.	Fixed Number per grid cell	Fixed number of Bbox per grid cell
	NMS, IOU	Bbox, NMS, IOU	NMS ⇒ IOU
⑥ Complexity	More Complex.	Relatively Simple	Relatively Simple

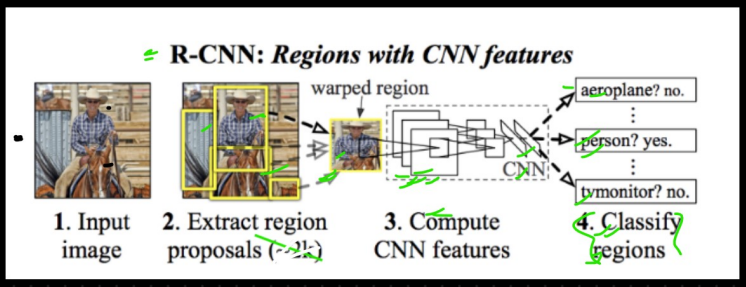
Scale of the obj ⇒

1. Faster-RCNN
2. Yolo
3. SSD

⇒ MAP

Sliding window method.

Image Classification ⇒ ⇒ ConvBase ⇒ FE ⇒ NN ⇒ $C_1/C_2/C_3$ ...?



XML → class, coordinate

C

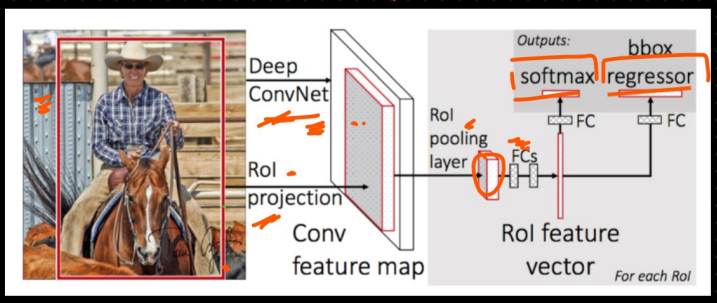
theory

RCNN ⇒ Regions → ConvBase → NN ⇒ C_1 C_2 C_3 C_4 ?

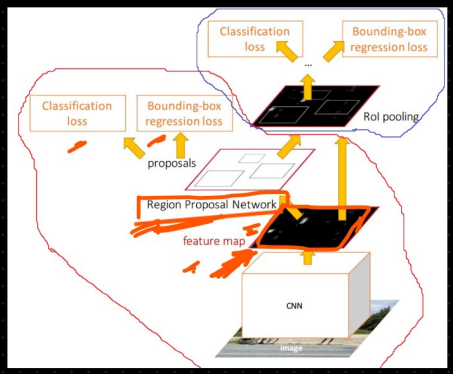
~2K

RPN

Faster-RCNN

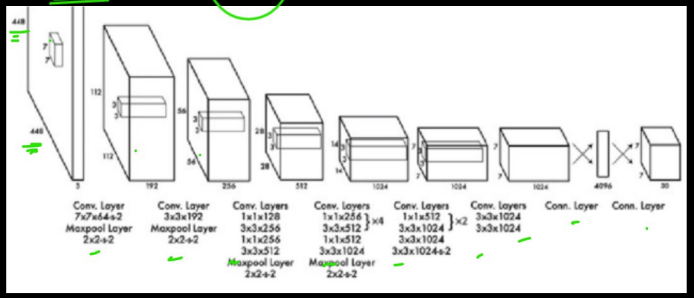


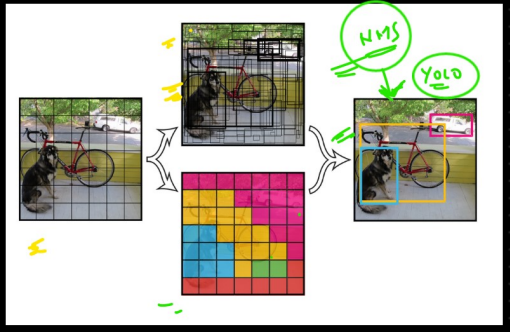
Fast-RCNN

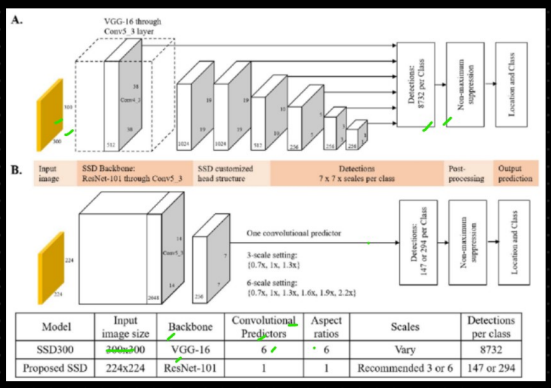


Faster-RCNN

448×448 ⇒ 7×7







Model	Input image size	Backbone	Convolutional Predictors	Aspect ratios	Scales	Detections per class
SSD300	~~300x300~~	VGG-16	6	6	Vary	8732
Proposed SSD	224x224	ResNet-101	1	1	Recommended 3 or 6	147 or 294