CS F425: Deep Learning

21

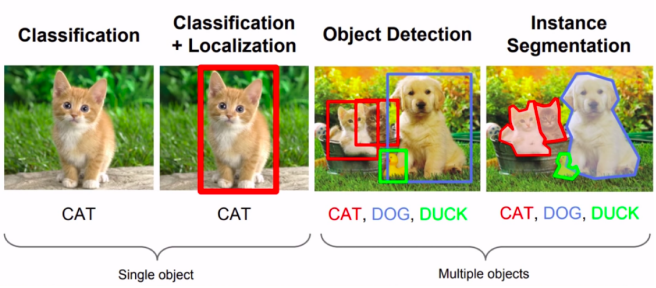
R-CNN Family Yolo, SSD

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

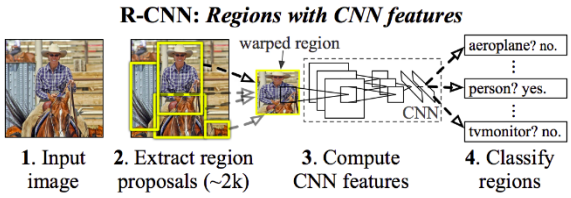
Mar 09, 2023 ON-CAMPUS Campus @ BITS-Pilani [Jan-May 2023]

http://ktiwari.in/dl

Object Detection and Localization



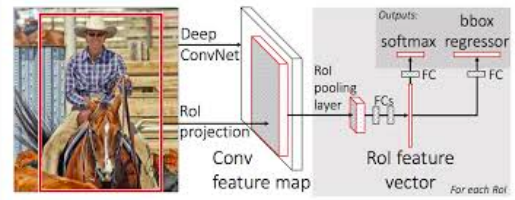
R-CNN [1]



- Region proposals ~20K (from external selective search)
- Warped image (resizing)
- SVM for classification (one for each class)

[1] **Cite 11049** Girshick, Ross and Donahue, Jeff and Darrell, Trevor and Malik, Jitendra, *Rich feature hierarchies for accurate object detection and semantic segmentation*, Conference on computer vision and pattern recognition, pages 580–587, IEEE-2014

Fast R-CNN [2]

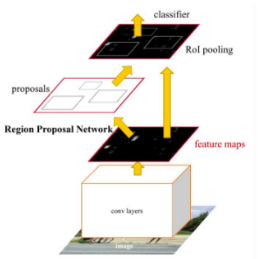


- RoI pooling
- Multi-task loss

[2] **Cite 7885** Girshick, Ross *Fast R-CNN*, International Conference on computer vision, pages 1440–1448, IEEE-2015

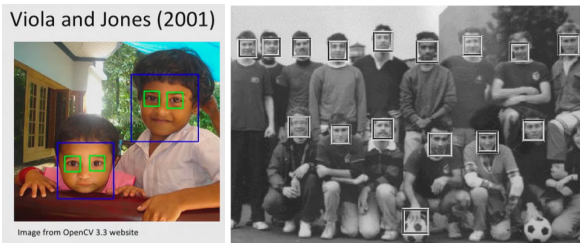
Faster R-CNN [3]

- Region Proposal Network (RPN)
- Four loss: RPN classification loss, RPN regress loss, Final classification loss, Final box coordinate loss
- 250 time faster than R-CNN. (Fast R-CNN is 25 times fast)



[3] **Cite 14669** Ren, Shaoqing and He, Kaiming and Girshick, Ross and Sun, Jian *Faster R-CNN: Towards real-time object detection with region proposal networks*, Advances in neural information processing systems, pages 91–99, IEEE-2015

Face Detection [4]



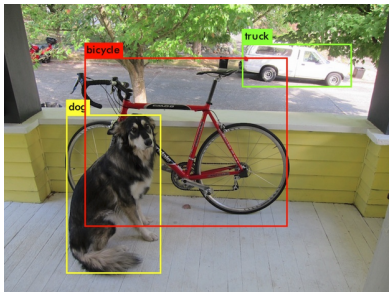
Visual object detection capable of processing images extremely rapidly and achieving high detection rates.

[4] **Cite 24549:** Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." IEEE computer society conference on computer vision and pattern recognition. CVPR 2001.

Object Detection with Yolo

What is there and where?

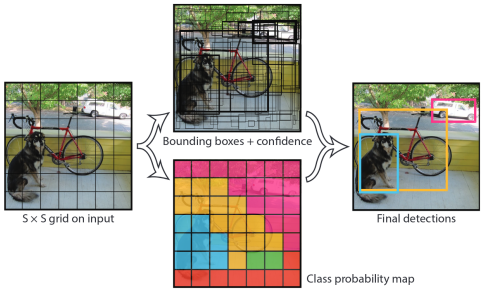
Deformative Part Model, and F-RCNN



- Apply the model to an image at multiple locations and scales. High scoring regions are considered detections.

Yolo: apply a single neural network to the full image that divides it into regions and predicts bounding boxes and probabilities for each region

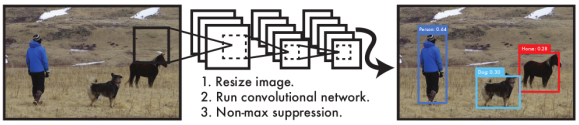
Yolo [5]

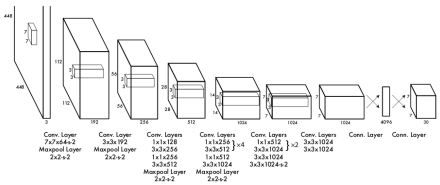


- Conditional probability map
- See https://pjreddie.com/darknet/yolo/

[5] **Cite 6375** Redmon, Joseph and Divvala, Santosh and Girshick, Ross and Farhadi, Ali, *You only look once: Unified, real-time object detection*, IEEE conference on computer vision and pattern recognition (CVPR), pages 779–788, 2016

Yolo





- $S \times S$ segments, gives B bounding boxes with confidence, and C class probabilities. So $S \times S \times (B \times 5 + C)$ values. S:7, B:2, C:20

Yolo

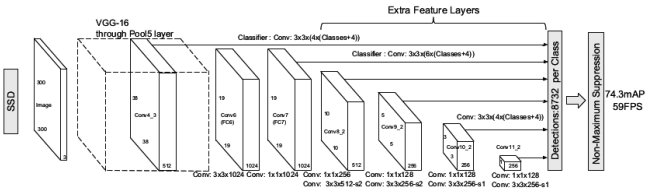
	Accuracy on Pascal	Speed
DPM V5	33.7%	0.07 FPS
R-CNN	66.0%	0.05 FPS
Fast R-CNN	70.0%	0.5 FPS
Faster R-CNN (ZF)	62.1%	17 FPS
Faster R-CNN (VGG16)	73.2%	7 FPS
Yolo	63.4%	45 FPS

- It is fast
- Speed comes at the price of accuracy. Improved to 69%
- Generalizes well
- Latest version YOLOv3 2018

Single Shot Object Detectors

	Accuracy on Pascal	Speed
DPM V5	33.7%	0.07 FPS
R-CNN	66.0%	0.05 FPS
Fast R-CNN	70.0%	0.5 FPS
Faster R-CNN (ZF)	62.1%	17 FPS
Faster R-CNN (VGG16)	73.2%	7 FPS
Yolo	63.4%	45 FPS
SSD-300	74.3%	59 FPS

Single Shot Multibox Detector (SSD) [6]



- Predicts *category score* and *box offsets*
- Uses detection at various scales
- Agregated features at later layers help detect large global objects
- Being single shot, avoids region proposal or selective search

[6] **Citations: 17739** Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, *Single Shot Multibox Detector*, in European conference on computer vision, pp 21–37, Springer 2016

SSD

- Multiple heads (8732)
- Default box having large IoU with ground truth box are trained
- Loss combines localization loss (L2) & confidence loss (softmax)
- Fast non maximal supression
- Parameters: $(p + 4) \times m \times n$

Thank You!

Thank you very much for your attention!