greatlearning
Learning for Life

Object Detection

Computer Vision

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Objectives

- Go beyond image classification
- Metrics to measure the quality of Object Detection
- Match advanced NN architectures suitable for these
- Different Approaches to solve the problems

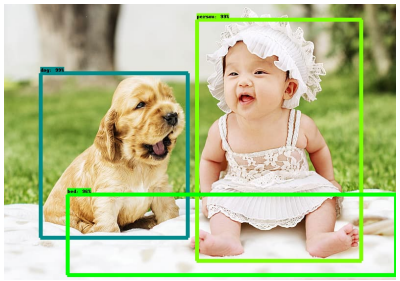
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Outline

- Object localization (detection)
- Evaluation Metrics
- Object detection state-of-the-art
- Object detection with region proposal and Selective search
- R-CNN, Fast R-CNN, Faster R-CNN
- Object detection with Yolo and SSD
- Hands-On - Python notebook

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Computer vision tasks

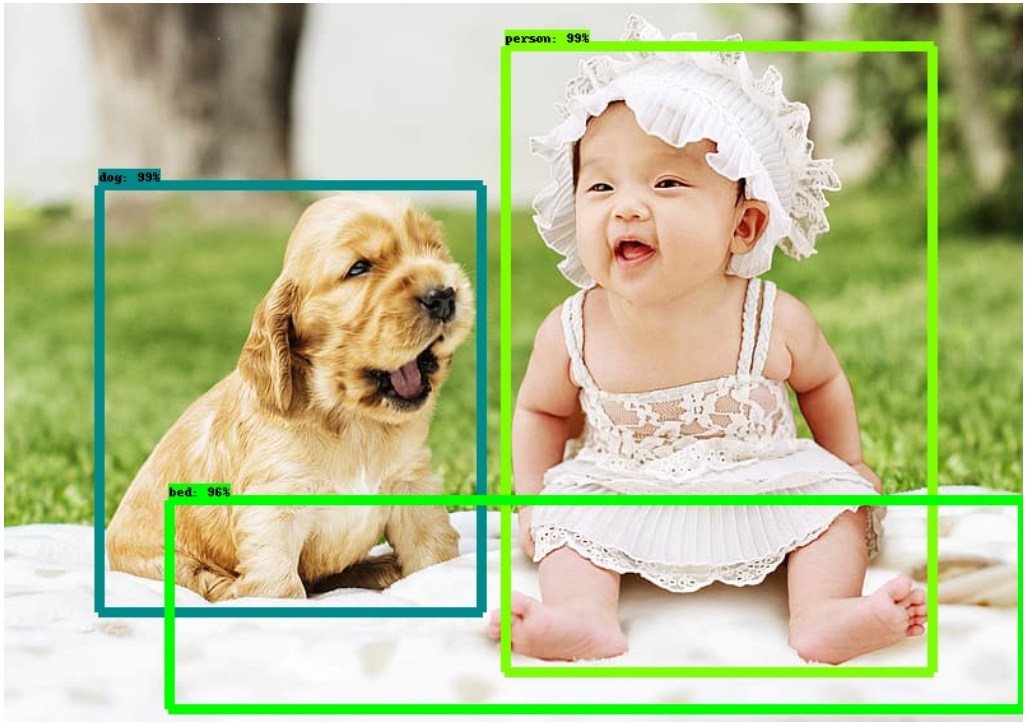
Classification	Classification + localization	Object detection	Semantic segmentation
			
Cat	Cat	Dog, Person, Bed	Dog, Person

Single object

Multiple objects

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection



Given an image we want to detect all the object in the image that belong to a specific classes and give their location. An image can contain more than one object with different classes.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Data: the oil driving object detection research

330K images (>200K labeled) and 80 object categories

14M images and more than 20K object categories

11,530 images and 20 classes

Google Open Images Dataset V5

~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, and visual relationships.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Image classification

Output a class label. The whole image represents one class. We don't want to know exactly where are the object. Usually only one object in the image.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

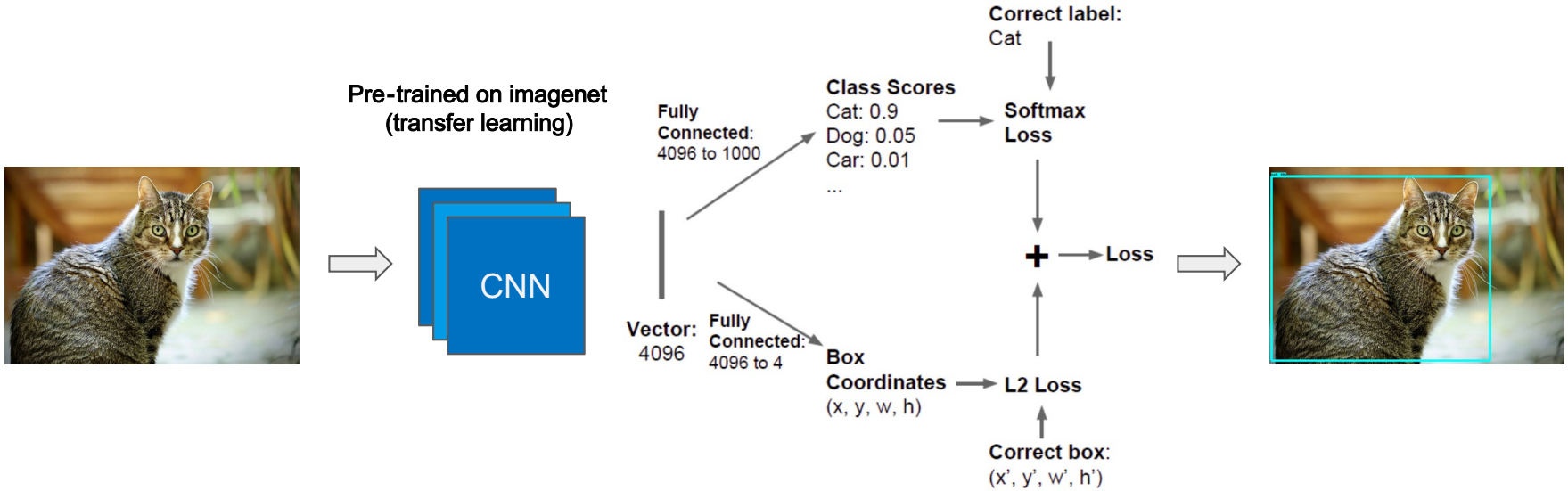
Classification + localization

Given an image we want to learn the class of the image and where are the class location in the image. Usually only one object in the image.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Classification + localization

- Treat localization as a regression problem



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection as regression

Pre-trained on imagenet (transfer learning)

Predicted class and coordinates for bounding box

Each image needs different number of outputs

CNN

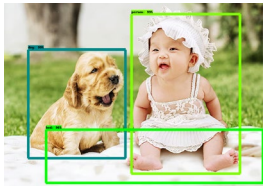
Cat: (x, y, w, h)

4 Numbers

CNN

Dog: (x, y, w, h)
Person: (x, y, w, h)
Bed: (x, y, w, h)

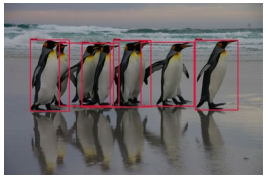
12 Numbers



CNN

Bird: (x, y, w, h)
Bird: (x, y, w, h)
. . .

Many numbers



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Notes

- To make image classification we use a ConvNet with a Softmax attached to the end of it.
- To make classification with localization we use a ConvNet with a softmax attached to the end for the classification part, and another output four numbers bx, by, bh, and bw to tell you the location of the class in the image.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Challenges in Object Detection

1. Two tasks - Classification and Localization
2. Results/prediction take lot of time but we need fast predictions for real-time task
3. Variable number of boxes as output
4. Different scales and aspect ratios
5. Limited data and labelled data
6. Imbalanced data-classes

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Performance metrics

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

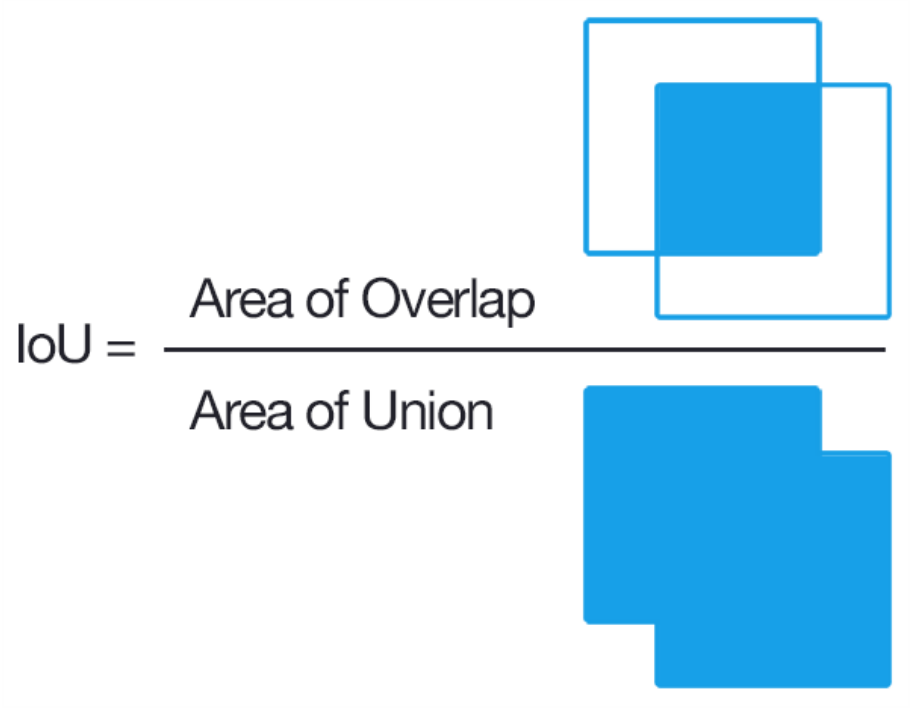
greatlearning
Learning for Life

Performance metrics for object detection

- Intersection over union (IoU)
- Precision and recall
- Mean average precision (mAP)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

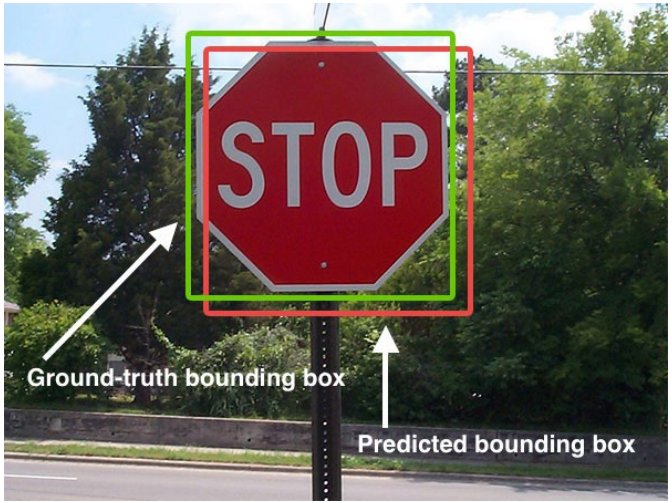
IoU (Intersection over Union)



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

IoU (Intersection over Union)

- IoU is a function used to evaluate the object detection algorithm.
- It computes size of intersection and divide it by the union. More generally, IoU is a measure of the overlap between two bounding boxes.
- For example in the diagram here : The red is the predicted output and green is the labeled output.
- To compute Intersection Over Union we first compute the union area of the two rectangles which is "the first rectangle + second rectangle" Then compute the intersection area between these two rectangles.
- Finally IOU = intersection area / Union area
- If IoU >= 0.5 (threshold) then it's good. The best answer will be 1.



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Why to use IoU score?

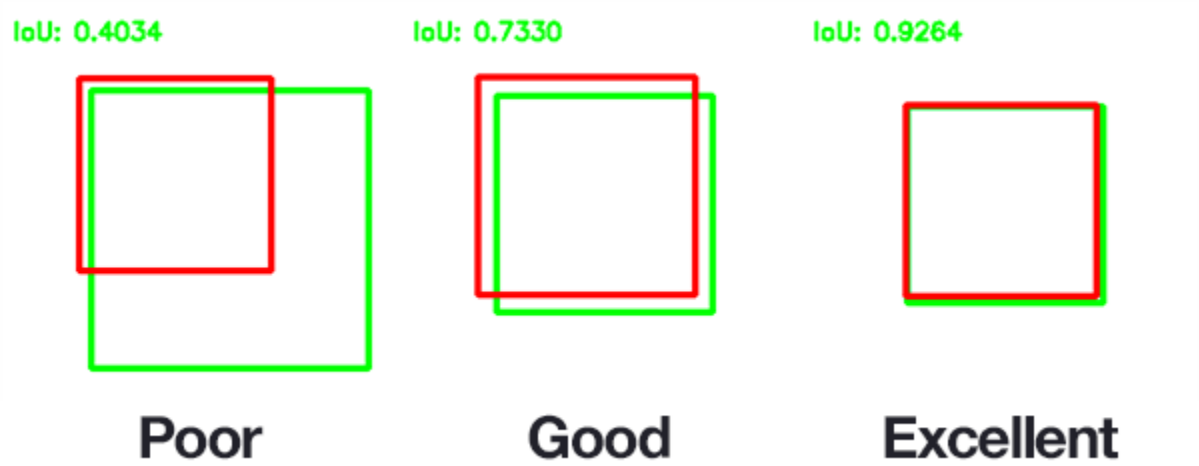
We use Accuracy, precision, recall etc as accuracy metrics for classification tasks but for object detection it's not so straightforward.

It's very unlikely that the (x, y)-coordinates of our predicted bounding box are going to exactly match the (x, y)-coordinates of the ground-truth bounding box.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Why to use IoU score?

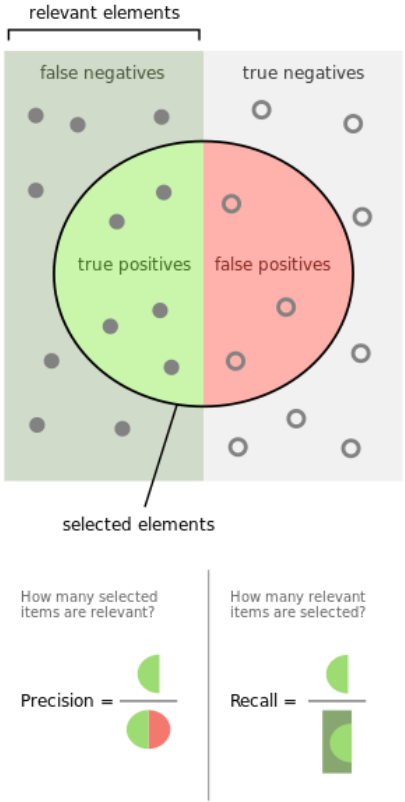
More the overlap predicted bounding boxes have with the ground-truth bounding boxes better (higher) their IoU scores will be.



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Precision and recall

- **Precision:** what percentage of your positive predictions are correct
- **Recall:** what percentage of ground truth objects were found



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

mAP (Mean Average Precision)

Step 1: Sort predictions according to confidence (usually classifier's output after softmax)

Step 2: Calculate IoU of every predicted box with every ground truth box

Step 3: Match predictions to ground truth using IoU, correct predictions are those with IoU > threshold (.5)

Confidence	Rank	Correct
0.91	1	True
0.87	2	True
0.83	3	False
0.81	4	True
0.77	5	False
0.65	6	True
0.56	7	True
0.40	8	False
0.32	9	False
0.31	10	True

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

mAP (Mean Average Precision)

Step 4: Calculate precision and recall at every row

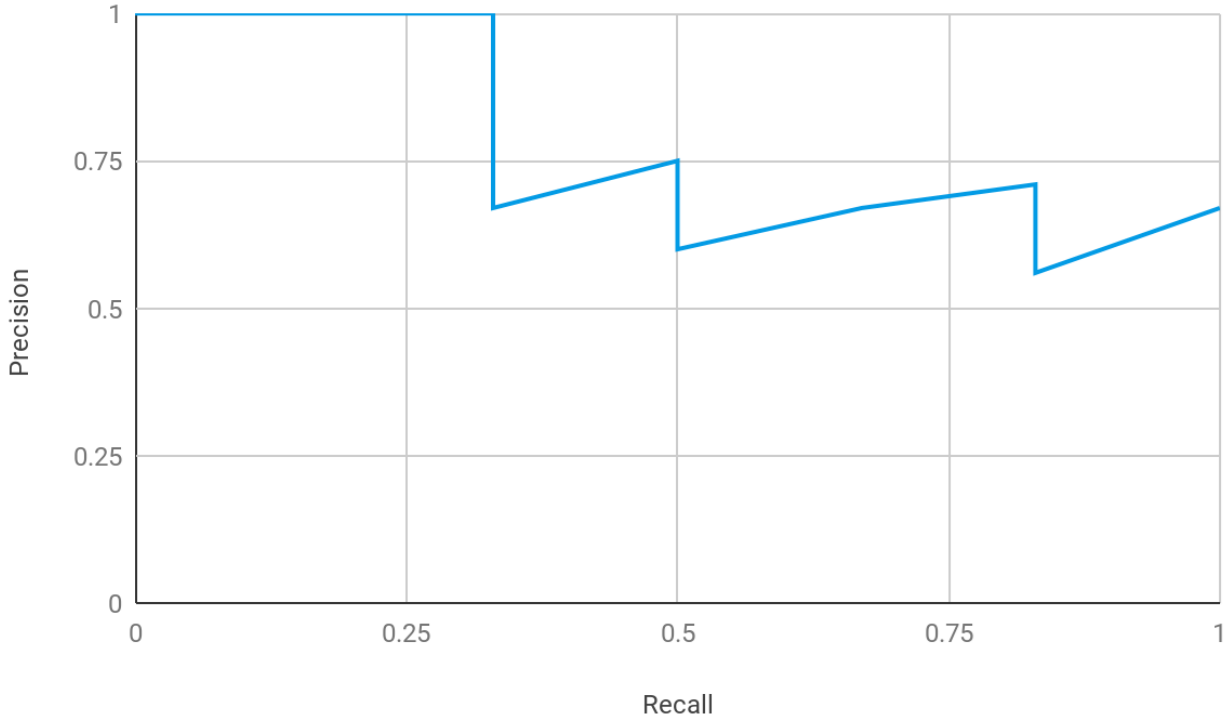
Step 5: Take the mean of maximum precision at 11 recall values (0.0, 0.1, … 1.0) to get AP

Step 6: Average across all classes to get the mAP

Confidence	Rank	Correct	Precision	Recall
0.91	1	True	1.00	0.17
0.87	2	True	1.00	0.33
0.83	3	False	0.67	0.33
0.81	4	True	0.75	0.50
0.77	5	False	0.60	0.50
0.65	6	True	0.67	0.67
0.56	7	True	0.71	0.83
0.40	8	False	0.63	0.83
0.32	9	False	0.56	0.83
0.31	10	True	0.67	1.00

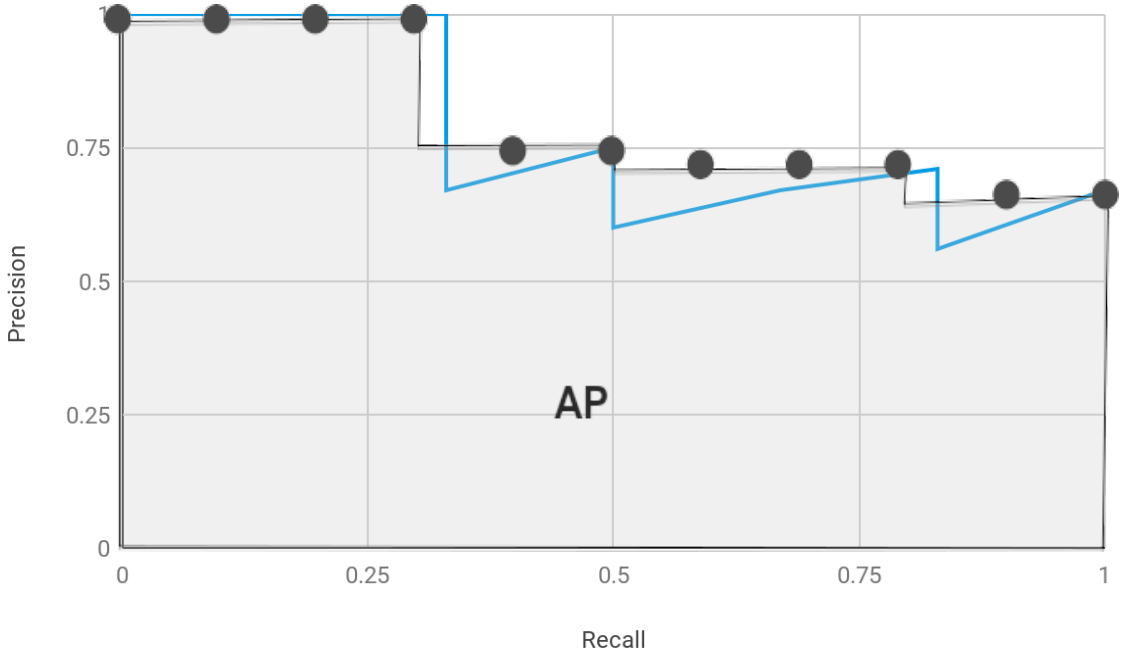
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Precision and recall curve



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Precision and recall curve



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

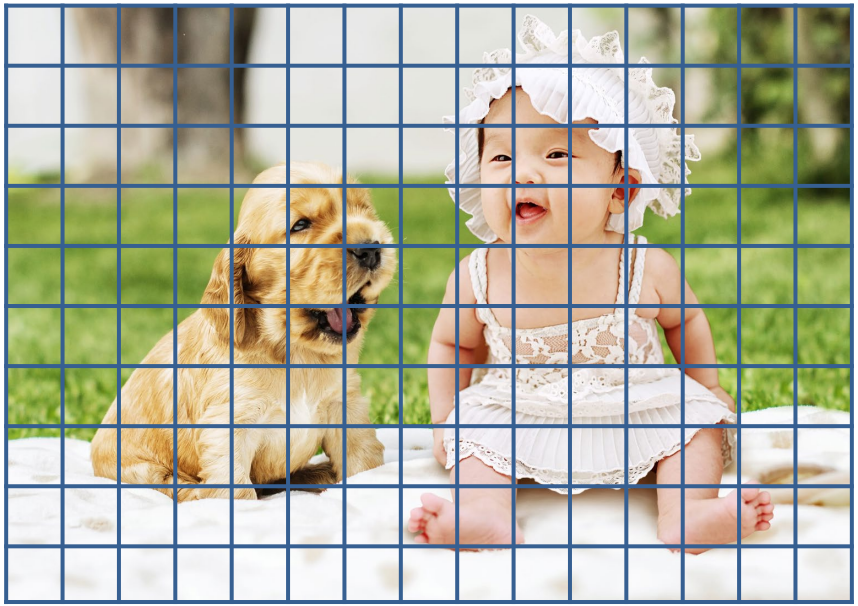
Object detection approaches

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Object detection: 1. Brute force approach

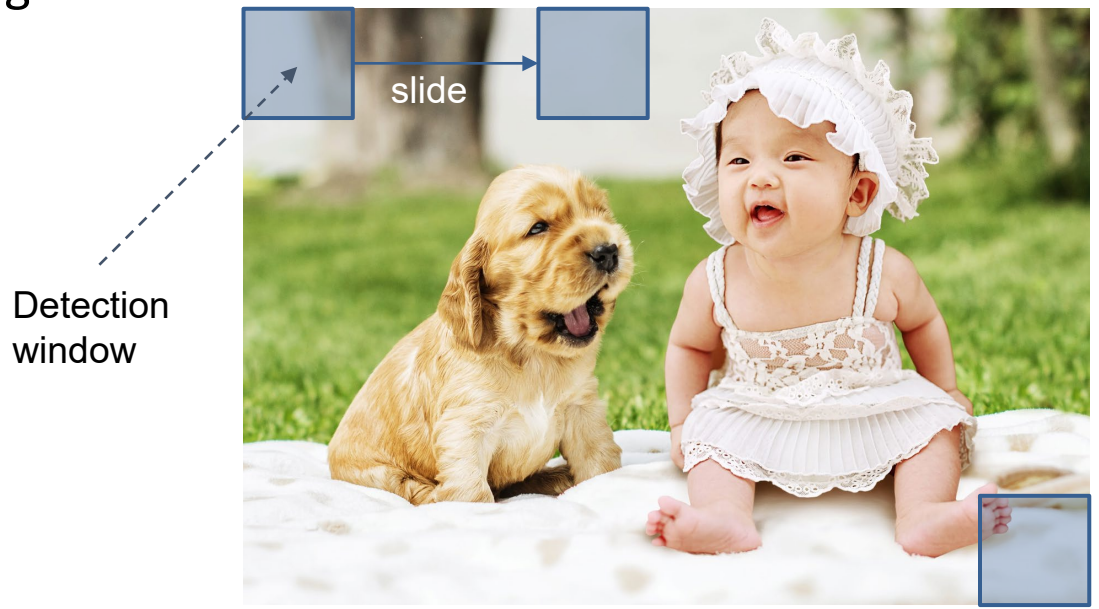
- Run a classifier for every possible box
- This is a 15 x 10 grid, there are 150 small boxes. How many total boxes?
- Computationally expensive



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection: 2. Sliding window approach

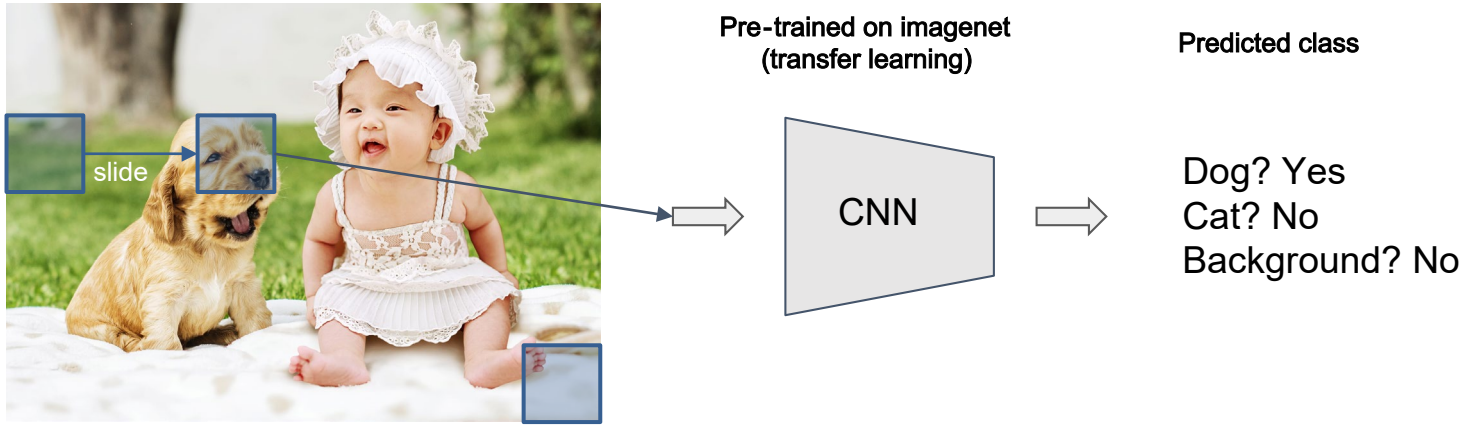
- Run classifier in a sliding window fashion



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection: Sliding window approach

- Apply a CNN to many different crops of the image
- CNN classifies each crop as object or background
- **Problem:** Need to apply CNN to huge number of locations (and scales), very computationally expensive



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection: Better approach

- Ideas how to reduce number of boxes?
  - Find ‘blobby’ image regions which are likely to contain objects
  - Run classifier for region proposals or boxes likely to contain objects
- Later: Class-agnostic object detector - “Region Proposals”

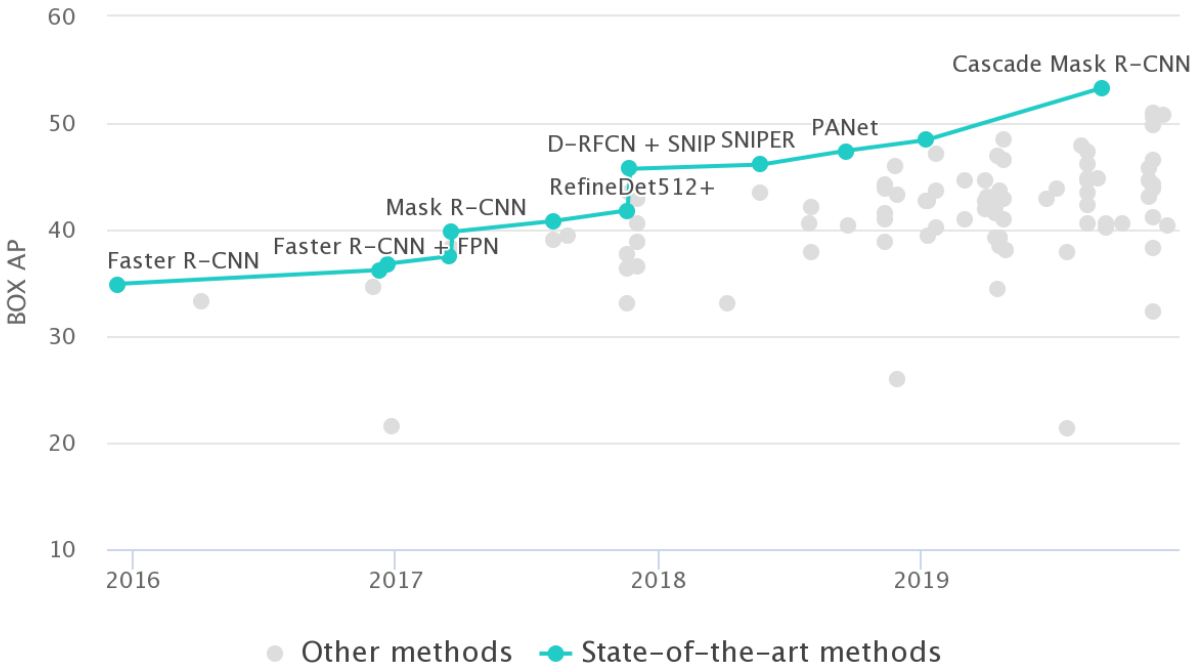
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection - State of the art

- State of the Art methods are generally categorised in two categories
  - One stage methods
  - Two stage methods
- One-stage methods give priority to inference speed, these include SSD, YOLO, RetinaNet. Whereas Faster-RCNN, Mask RCNN and Cascade RCNN are example of two-stage methods where priority is given to detection accuracy.
- Popular benchmark dataset is Microsoft COCO
- Popular metric for evaluation is mAP (Mean Average Precision)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object detection: SOTA on MS-COCO dataset



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Object detection: SOTA on MS-COCO dataset

Method	Backbone	AP-box	AP-50	AP-75
one stage:				
SSD512 (Liu et al. 2016)	VGG16	28.8	48.5	30.3
RetinaNet (Lin et al. 2017b)	ResNeXt101	40.8	61.1	44.1
RefineDet (Zhang et al. 2018)	ResNet101	41.8	62.9	45.7
CornerNet (Zhang et al. 2018)	Hourglass-104	42.2	57.8	45.2
M2Det (Zhao et al. 2018)	VGG16	44.2	64.6	49.3
FSAF (Zhu, He, and Savvides 2019)	ResNext-101	44.6	65.2	48.6
NAS-FPN (Ghiasi, Lin, and Le 2019)	AmoebaNet	48.3	-	-
two stage:				
Faster R-CNN (Ren et al. 2015)	VGG16	21.9	42.7	-
R-FCN (Dai et al. 2016)	ResNet101	29.9	51.9	-
FPN (Lin et al. 2017a)	ResNet101	36.2	59.1	39
Mask R-CNN (He et al. 2017)	ResNet101	39.8	62.3	43.4
Cascade R-CNN (Cai and Vasconcelos 2018)	ResNet101	42.8	62.1	46.3
Libra R-CNN (Pang et al. 2019)	ResNext-101	43	64	47
SNIP (model ensemble) (Singh and Davis 2018)	-	48.3	69.7	53.7
SINPER (Singh, Najibi, and Davis 2018)	ResNet101	47.6	68.5	53.4
Cascade Mask R-CNN (Girshick et al. 2018)	ResNeXt152	50.2	68.2	54.9
MegDet (model ensemble) (Peng et al. 2018)	-	52.5	-	-
CBNet (Liu et al. 2019)	**Dual-ResNeXt152**	**52.8**	**70.6**	**58**
CBNet (Liu et al. 2019)	**Triple-ResNeXt152**	**53.3**	**71.9**	**58.5**

Data Source: https://arxiv.org/pdf/1909.03625v1.pdf

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

List of Algorithms -

Region Proposal Based Algorithms -

- R-CNN
- Fast R-CNN
- Faster R-CNN

Algorithms without Region Proposal -

- YOLO
- SSD

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

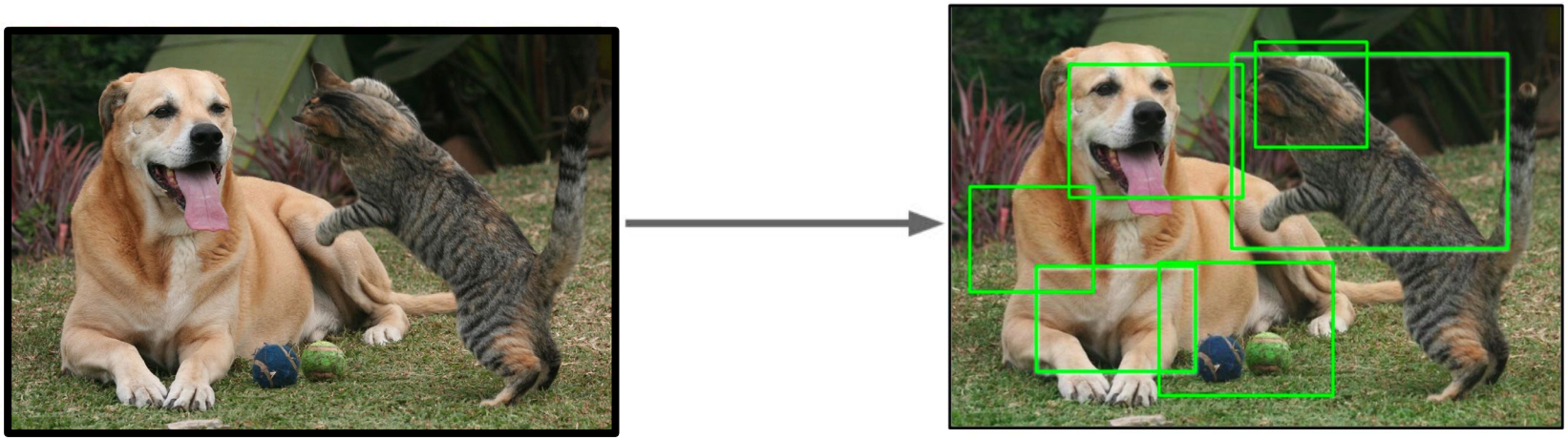
greatlearning
Learning for Life

Region proposal

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

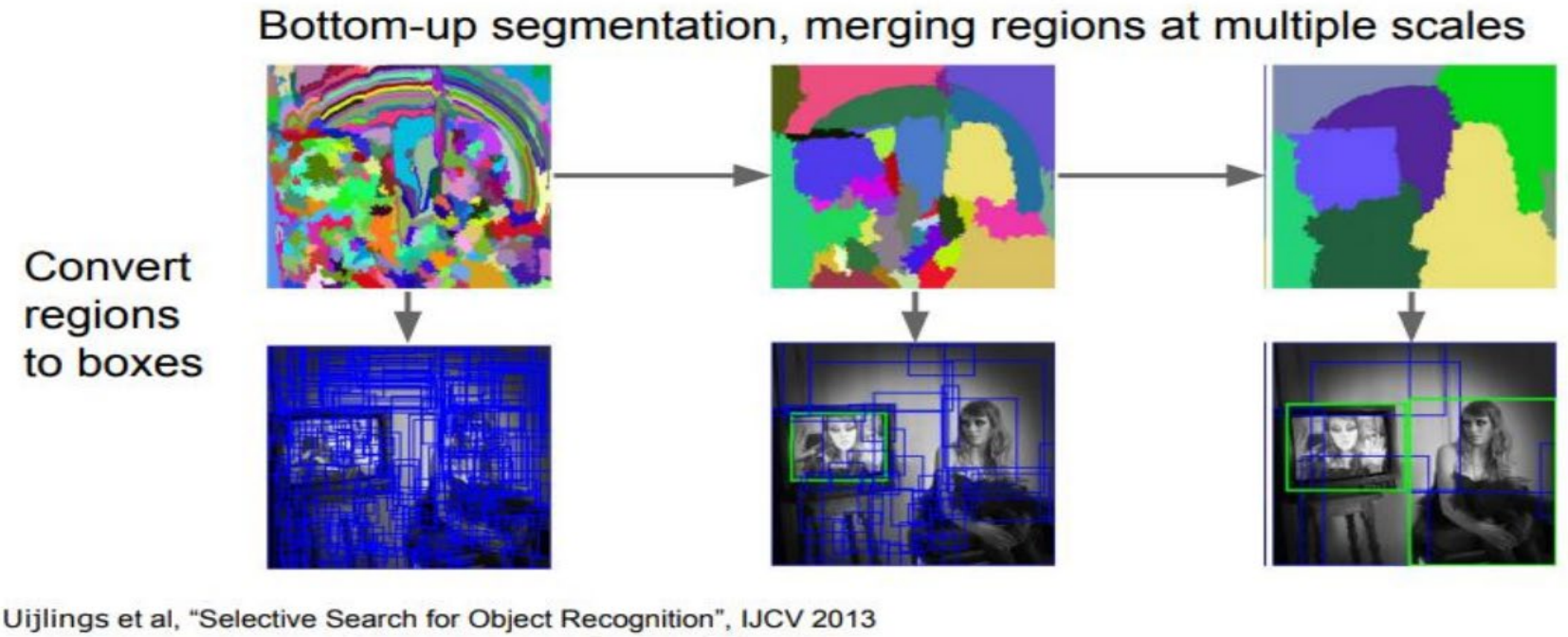
Region Proposal

- Find “blobby” image regions that are likely to contain objects
- Relatively fast to run; e.g. Selective Search gives 1000 region proposals in a few seconds on CPU



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Region Proposal - Selective Search



Uijlings et al, "Selective Search for Object Recognition", IJCV 2013

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Selective Search for Objective Recognition

- **Selective Search** uses the best of both worlds: Exhaustive search and segmentation.
- Segmentation improve the sampling process of different boxes - **reduces considerably the search space.**
- To improve the algorithm's robustness a variety of strategies are used during the bottom-up boxes' merging.
- Selective Search produces boxes that are good proposals for objects, it handles well different image conditions, but more important it is fast enough to be used in a prediction pipeline (like Fast-RCNN) to do real-time object detection.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Selective Search Advantages

- Capture All Scales - Objects can occur at any scale within the image. Furthermore, some objects may not have clear boundaries. This is achieved by using an hierarchical algorithm.
- Diversification - Regions may form an object because of only colour, only texture, or lighting conditions etc. Therefore instead of a single strategy which works well in most cases, we prefer to have a diverse set of strategies to deal with all cases.
- Fast to compute

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Selective Search

Figure 2: Two examples of our selective search showing the necessity of different scales. On the left we find many objects at different scales. On the right we necessarily find the objects at different scales as the girl is contained by the tv.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Region Proposal - Many other choices

Method	Approach	Outputs Segments	Outputs Score	Control #proposals	Time (sec.)	Repea-tability	Recall Results	Detection Results
Bing [18]	Window scoring		✓	✓	0.2	⋆⋆⋆	⋆	·
CPMC [19]	Grouping	✓	✓	✓	250	-	⋆⋆	⋆
EdgeBoxes [20]	Window scoring		✓	✓	0.3	⋆⋆	⋆⋆⋆	⋆⋆⋆
Endres [21]	Grouping	✓	✓	✓	100	-	⋆⋆⋆	⋆⋆
Geodesic [22]	Grouping	✓		✓	1	⋆	⋆⋆⋆	⋆⋆
MCG [23]	Grouping	✓	✓	✓	30	⋆	⋆⋆⋆	⋆⋆⋆
Objectness [24]	Window scoring		✓	✓	3	·	⋆	·
Rahtu [25]	Window scoring		✓	✓	3	·	·	⋆
RandomizedPrim's [26]	Grouping	✓		✓	1	⋆	⋆	⋆⋆
Rantalankila [27]	Grouping	✓		✓	10	⋆⋆	·	⋆⋆
Rigor [28]	Grouping	✓		✓	10	⋆	⋆⋆	⋆⋆
SelectiveSearch [29]	Grouping	✓	✓	✓	10	⋆⋆	⋆⋆⋆	⋆⋆⋆
Gaussian				✓	0	·	·	⋆
SlidingWindow				✓	0	⋆⋆⋆	·	·
Superpixels		✓			1	⋆	·	·
Uniform				✓	0	·	·	·

Grey check-marks indicate that the number of proposals is controlled by indirectly adjusting parameters. Repeatability, quality, and detection rankings are provided as rough summary of the experimental results: "-" indicates no data, ".", "*", "**", "***" indicate progressively better results.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

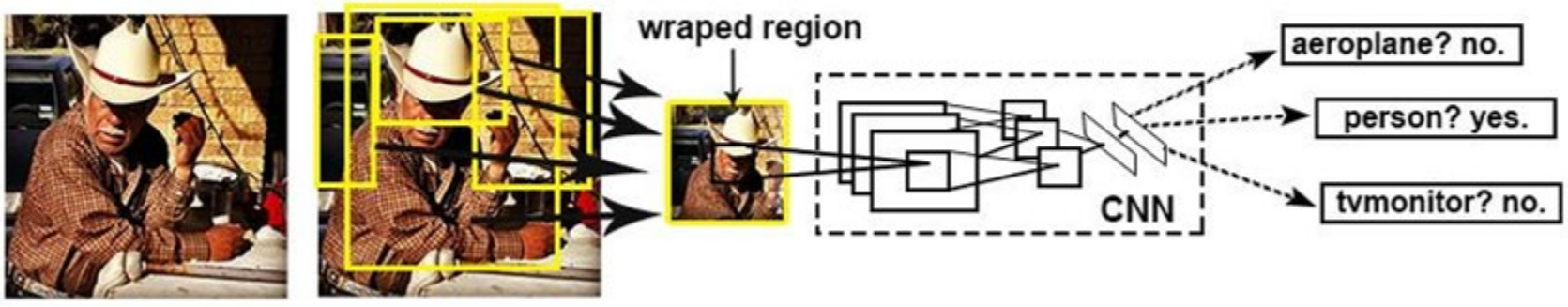
Region CNN (R-CNN)

- R-CNN is an algorithm which can also be used for object detection.
- R-CNN stands for regions with Conv Nets.
- R-CNN tries to pick a few windows and run a Conv net (your confident classifier) on top of them.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

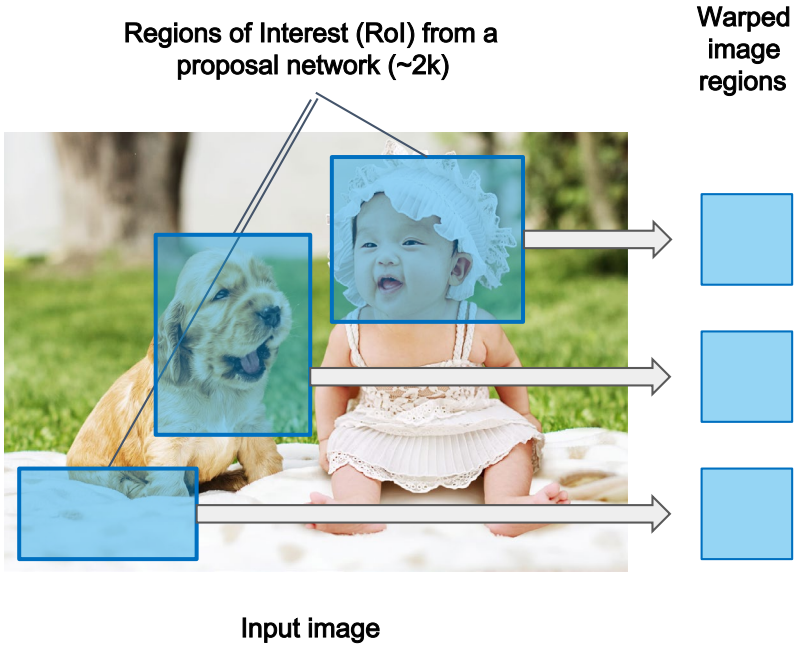
Region CNN (R-CNN)

R-CNN, first generates 2K region proposals (bounding box candidates), then detect object within the each region proposal as below:



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

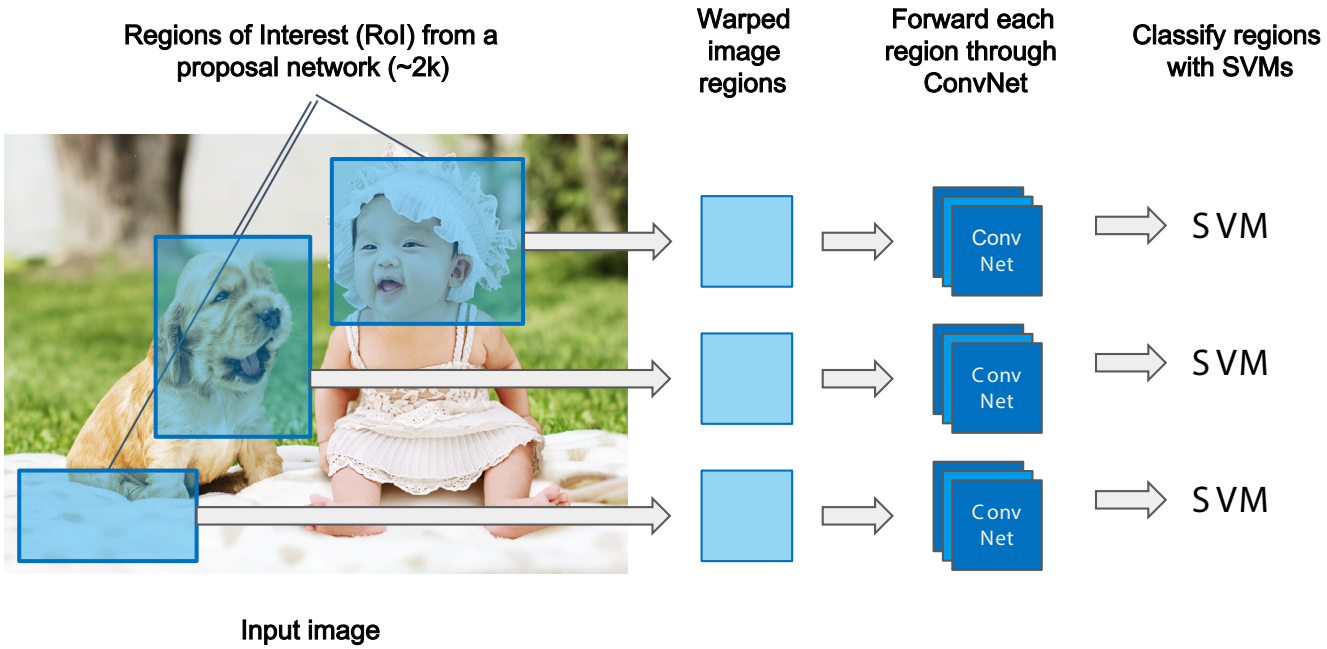
How R-CNN works?



Source: arXiv:1311.2524v5 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

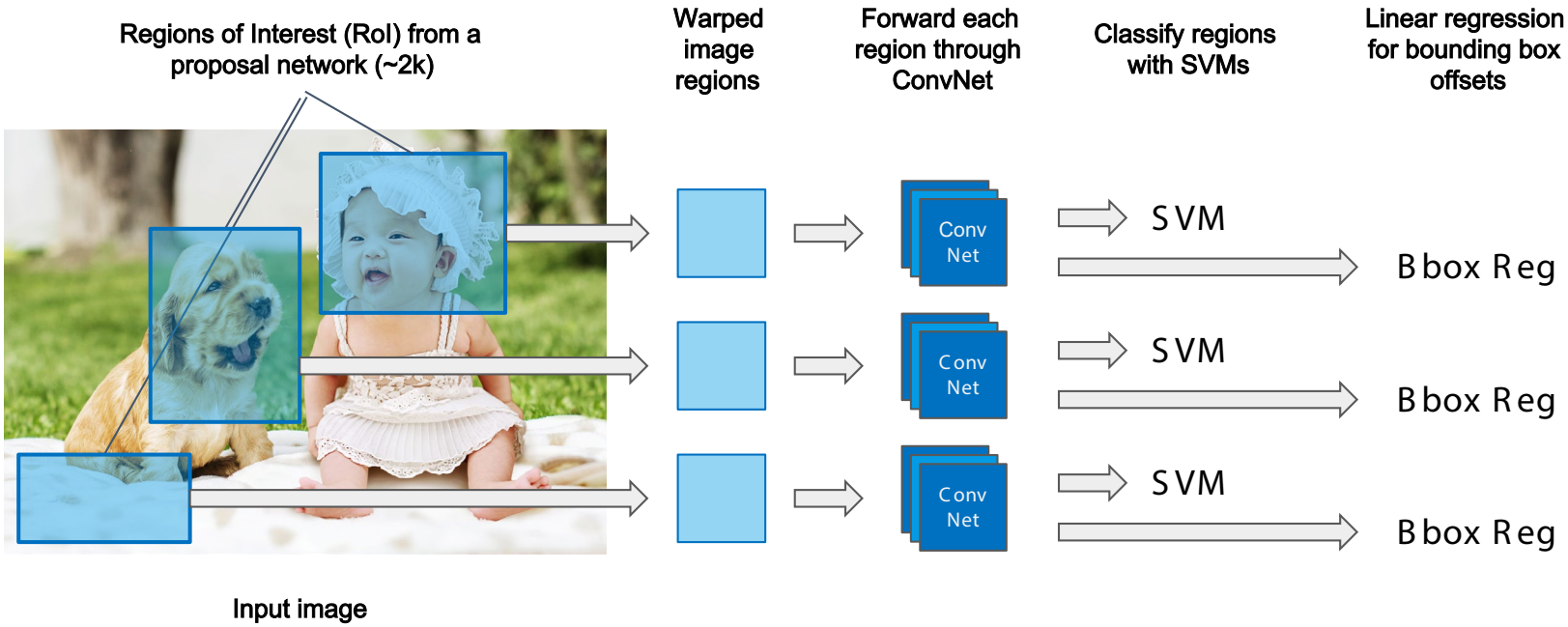
How R-CNN works?



Source: arXiv:1311.2524v5 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

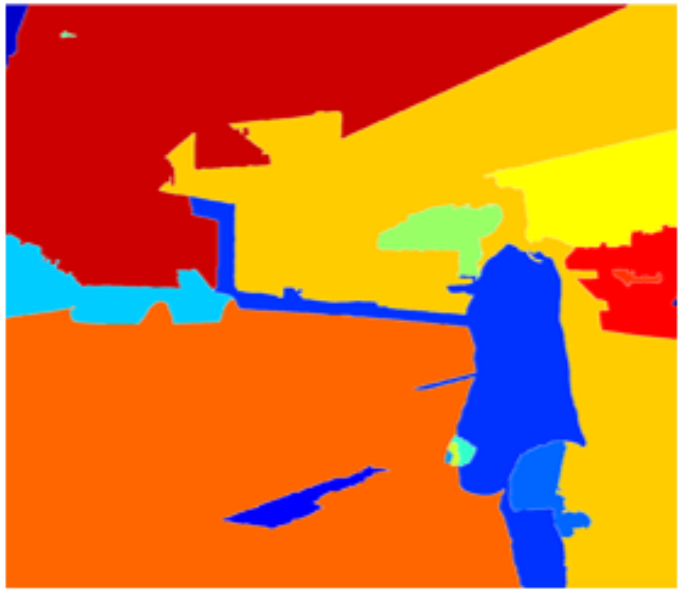
How R-CNN works?



Source: arXiv:1311.2524v5 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

R-CNN



The algorithm R-CNN uses to pick windows is called a segmentation algorithm.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

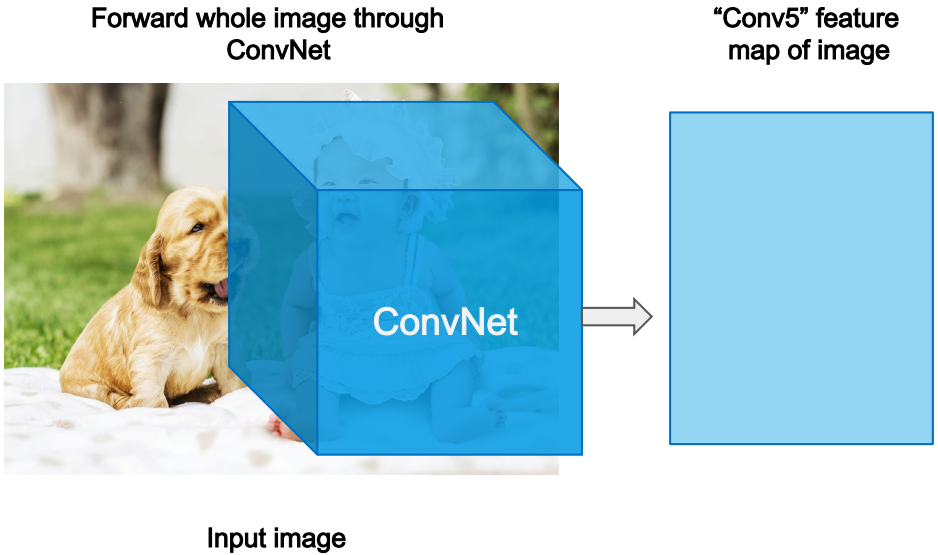
Disadvantages of R-CNN

Separate convnet for each box

(slow)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

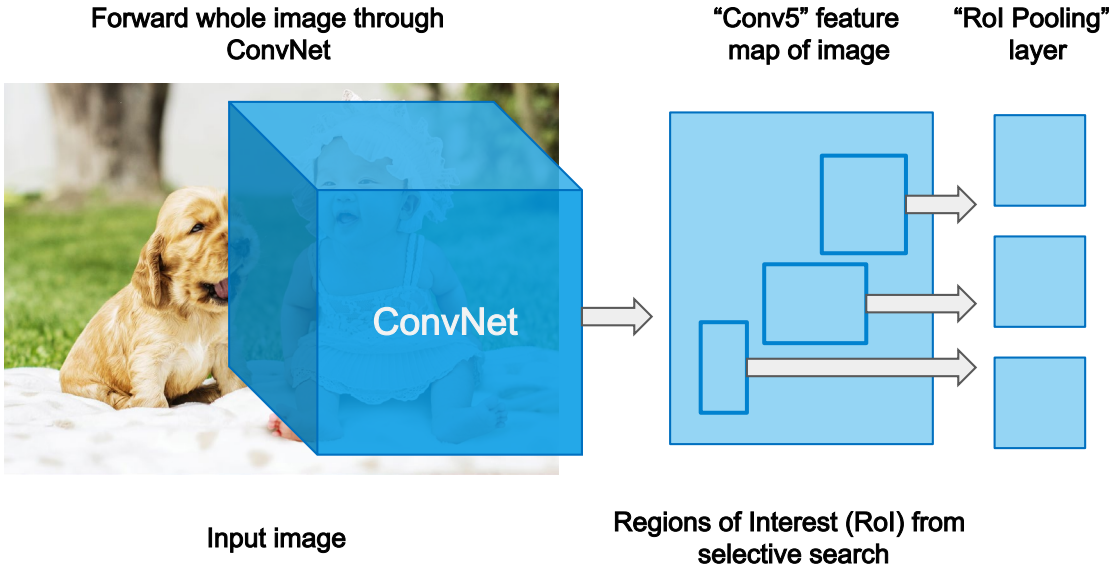
Fast R-CNN



Source: arXiv:1504.08083v2 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

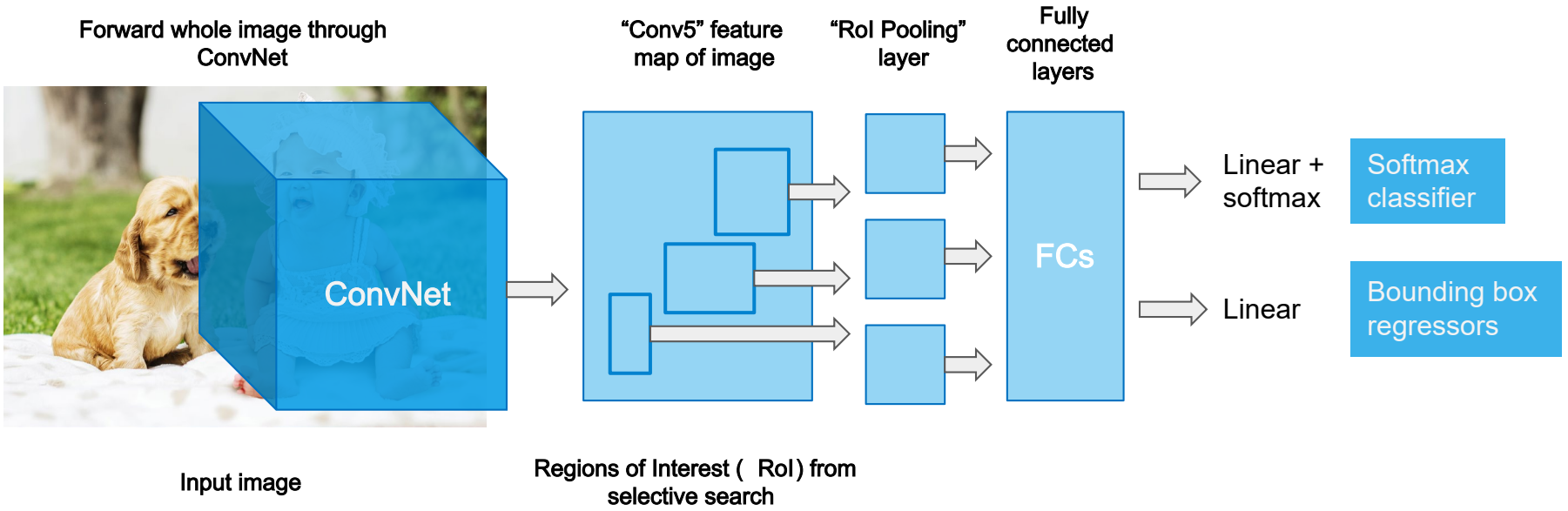
Fast R-CNN



Source: arXiv:1504.08083v2 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

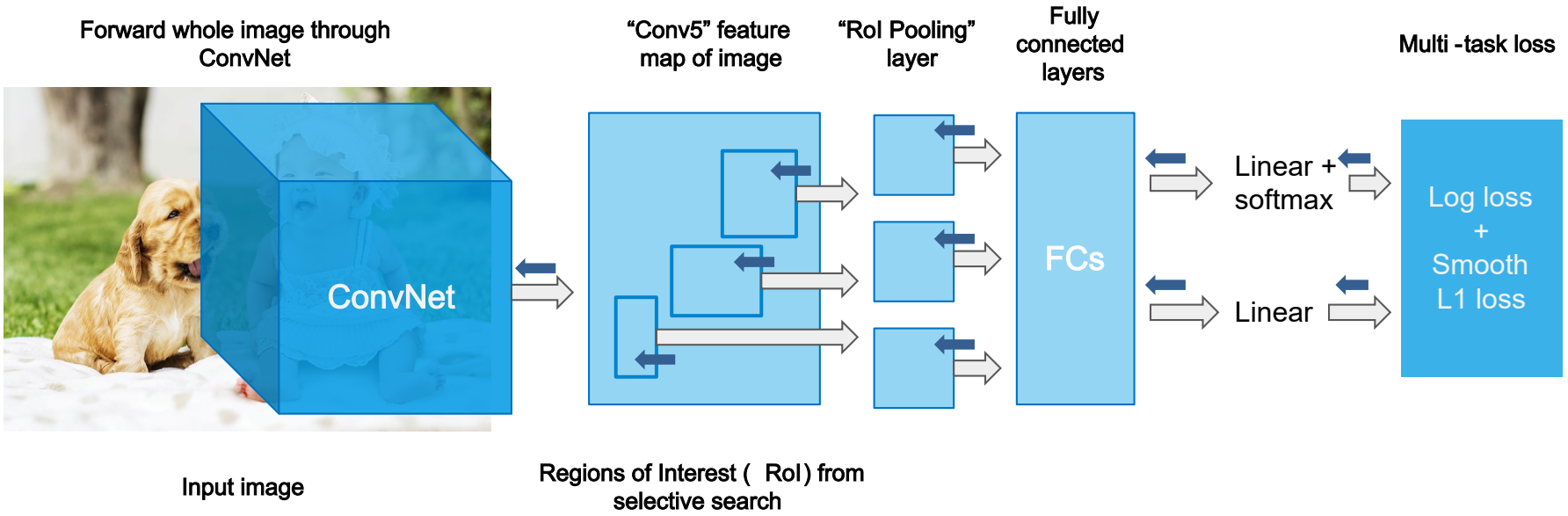
Fast R-CNN



Source: arXiv:1504.08083v2 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Fast R-CNN training

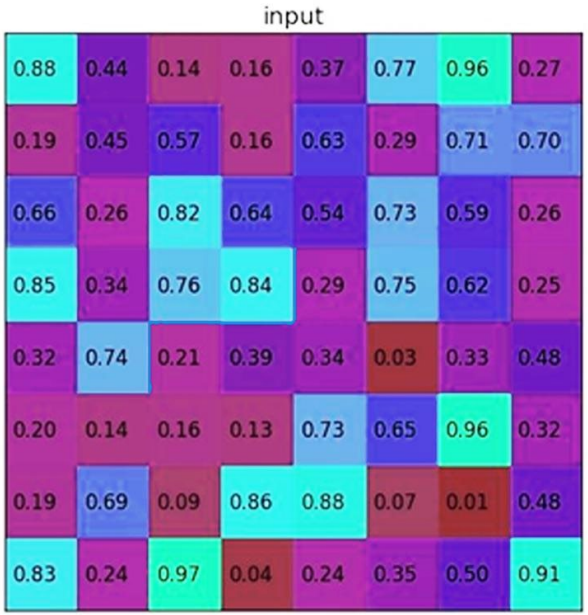


Source: arXiv:1504.08083v2 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Fast R-CNN: RoI Pooling

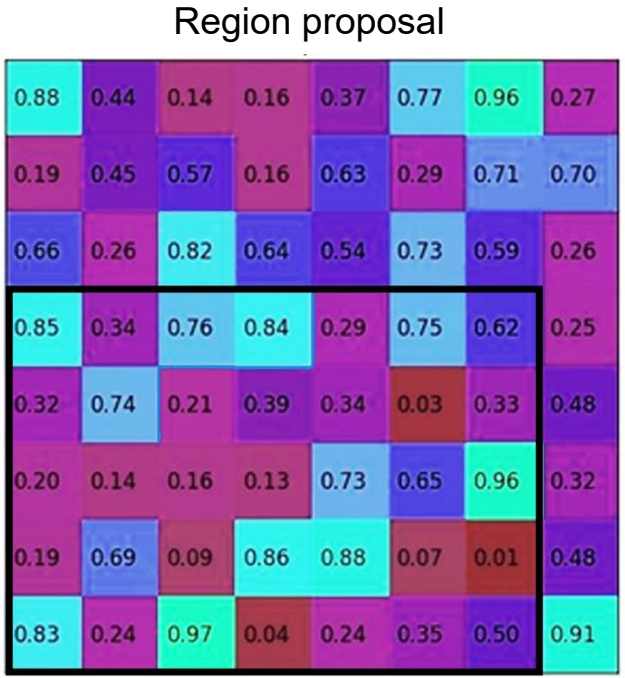
RoI pooling on a single 8×8 feature map, one region of interest and an output size of 2x2. Our input feature map looks like this:



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Fast R-CNN: RoI Pooling

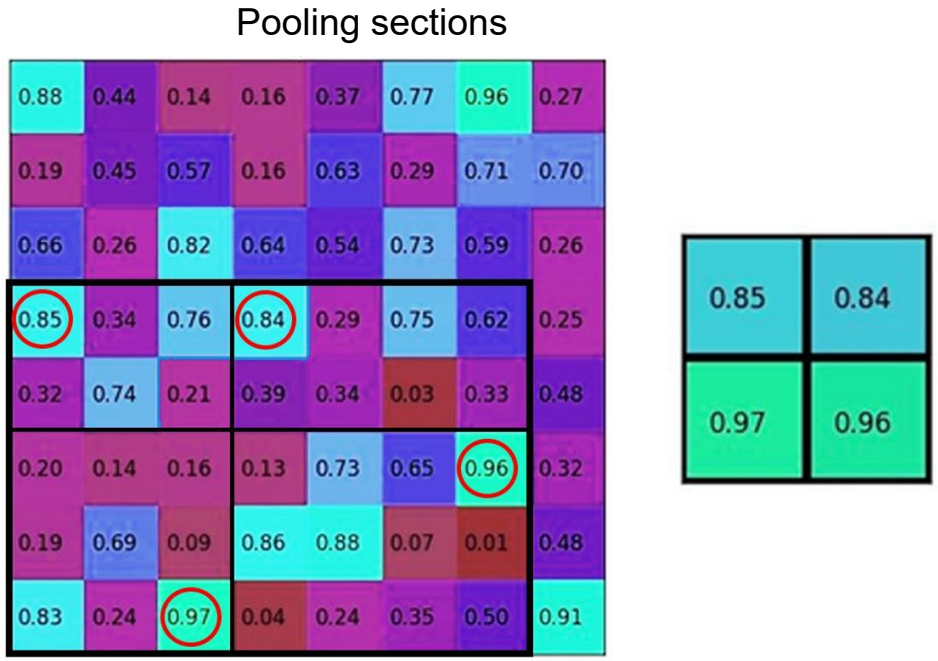
Let’s say we also have a region proposal (0, 3) - (7, 8)



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

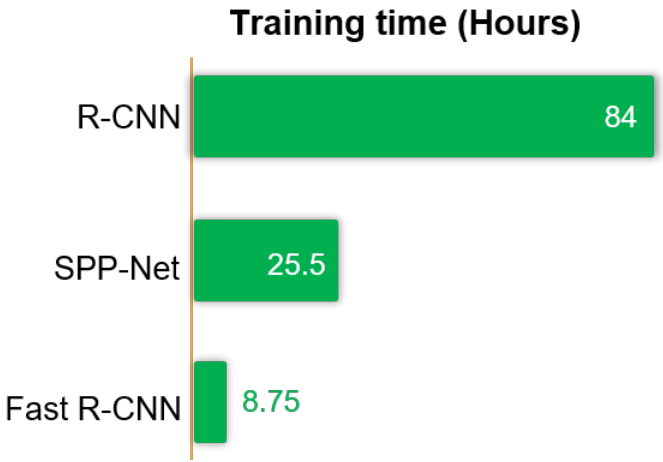
Fast R-CNN: RoI Pooling

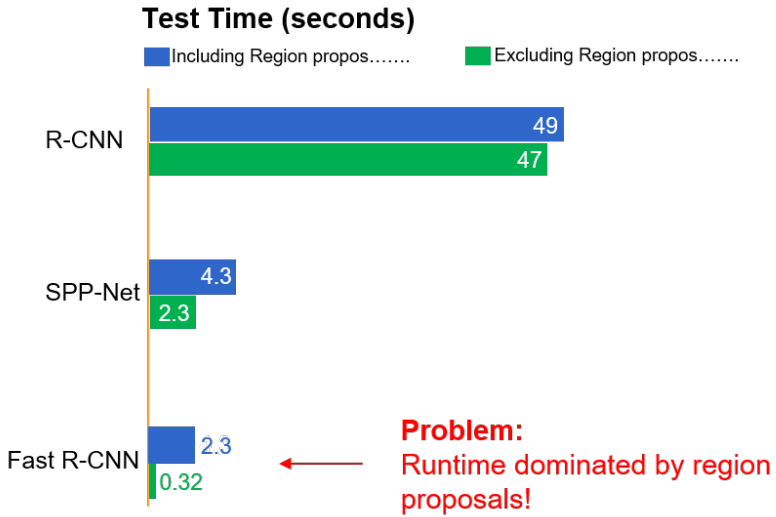
By dividing it into 2x2 sections (because the output size is 2x2) we get:



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

R-CNN vs SPP vs Fast R-CNN

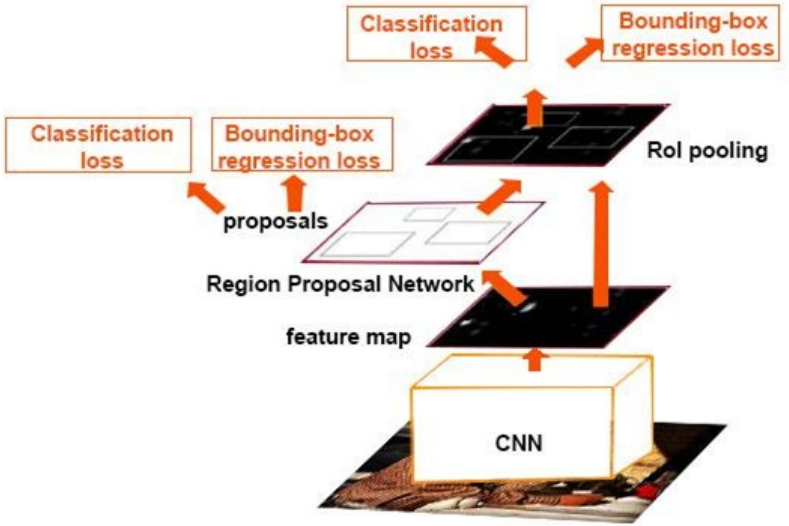




Data Source: arXiv:1311.2524v5 [cs.CV], arXiv:1406.4729 [cs.CV], and arXiv:1504.08083 [cs.CV]
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Faster R-CNN

- Make CNNs do proposals!
- Insert Region Proposal Network (RPN) to predict proposals from features.
- Jointly train with 4 losses
  - RPN classify object/not object
  - RPN regress box coordinates
  - Final classification score (object classes)
  - Final box coordinates



Source: arXiv:1506.01497v3 [cs.CV]

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Faster R-CNN Performance

In the paper “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, speed of various models have been compared on Pascal Voc dataset.

It was found that Faster R-CNN is much faster than R-CNN.

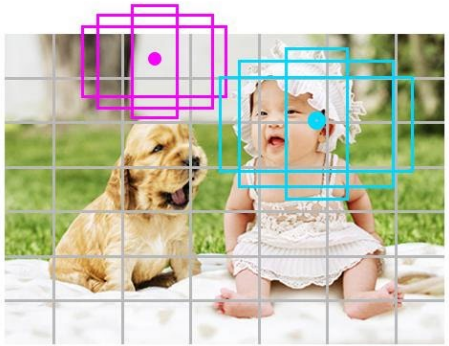
https://arxiv.org/pdf/1506.01497.pdf

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Detection without proposals: YOLO/SSD

Go from input image to tensor of scores with one big convolutional network!

Input image
3 x H x W



Divide image into grid
7 x 7

Image a set of **base boxes** centered at each grid cell
Here B = 3

Within each grid cell:

- Regress from each of the B base boxes to a final box with 5 numbers: (dx, dy, dh, dw, confidence)
- Predict scores for each of C classes (including background as a class)

Output:
7 x 7 x (5 * B + C)

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

YOLO

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

YOLO (You only look once)

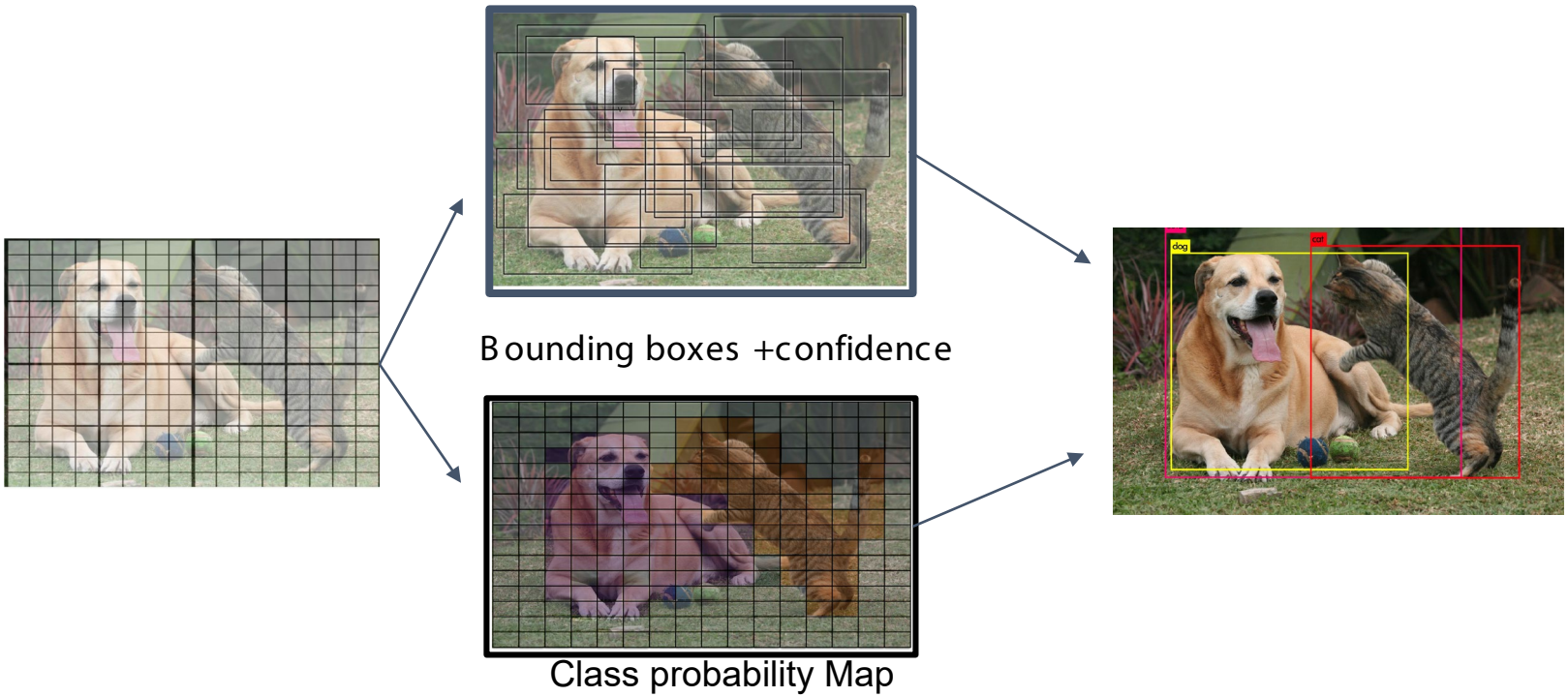
- YOLO – You only look once, looks at the image just once but in a clever way

  “We reframe the object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities”

- Algorithmic flow :
  - Actually Divides the image into a grid of say, 13*13 cells (S=13)
  - Each of these cells is responsible for predicting 5 bounding boxes (B=5) (A bounding box describes the rectangle that encloses an object)
  - YOLO for each bounding box
    - outputs a confidence score that tells us how good is the shape of the box
    - the cell also predicts a class
  - The confidence score of bounding box and class prediction are combined into final score -> probability that this bounding box contains a specific object

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

YOLO



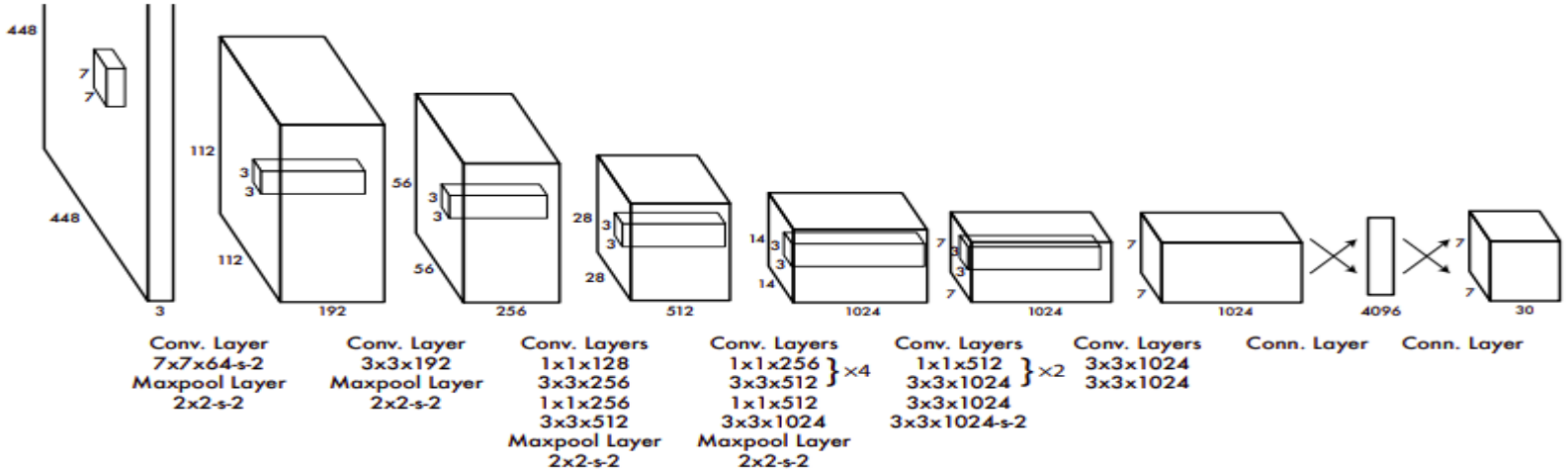
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Example - YOLO Details

- 13*13 = 169 grid cells
- Each cell predicts 5 bounding boxes
- 169*5 = 845 bounding boxes
- Most of the boxes have low confidence score
- Threshold of 30% or more
- Input image: 416*416 pixels
- 13*13*125 tensor describing the bounding boxes for grid cells
- YoLO v2 vs v1 – faster version

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

YOLO Network Architecture



This detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. the convolutional layers are pretrained on the ImageNet classification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

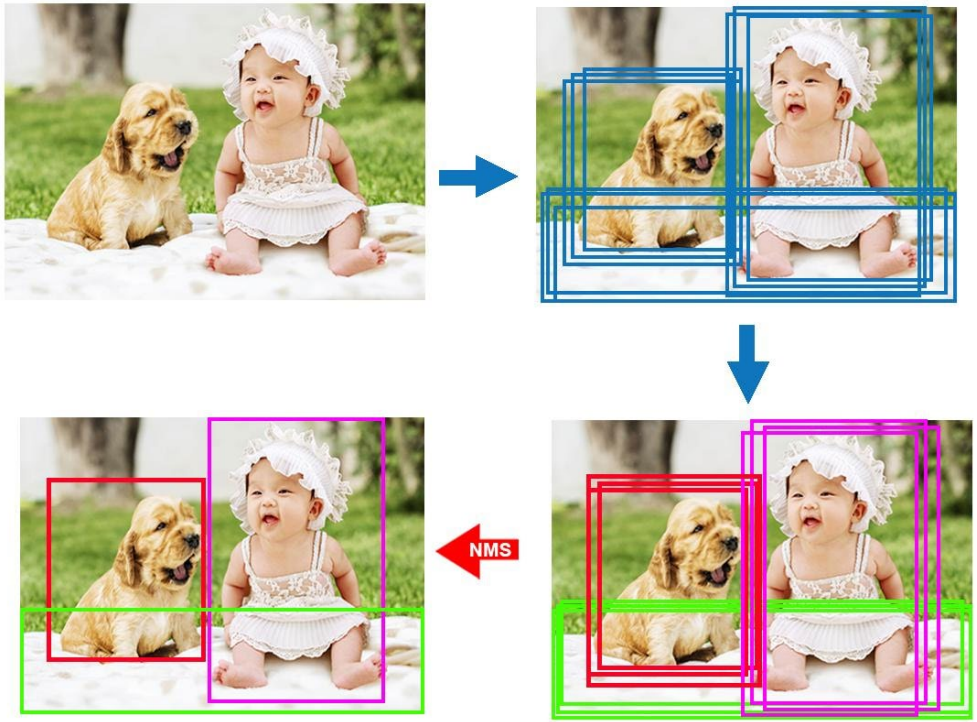
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Non-maximal suppression

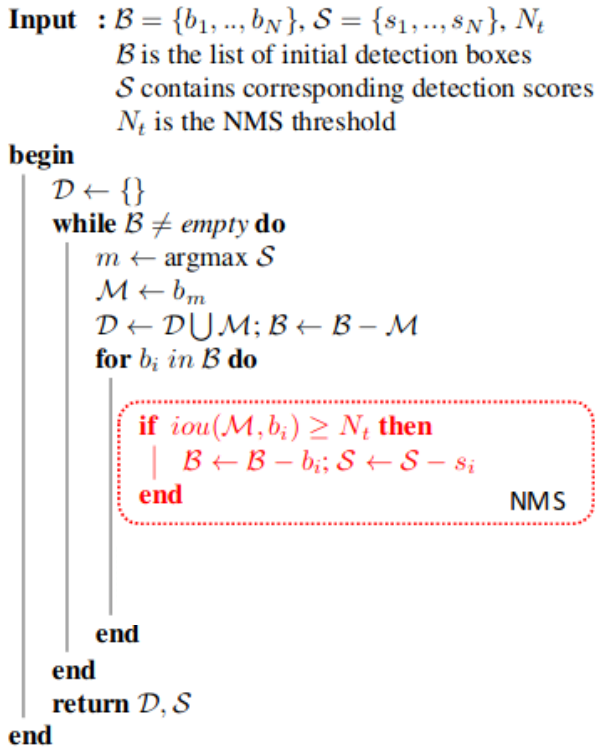
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Non-maximal suppression



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Non-maximal suppression



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

greatlearning
Learning for Life

Single-shot detector

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

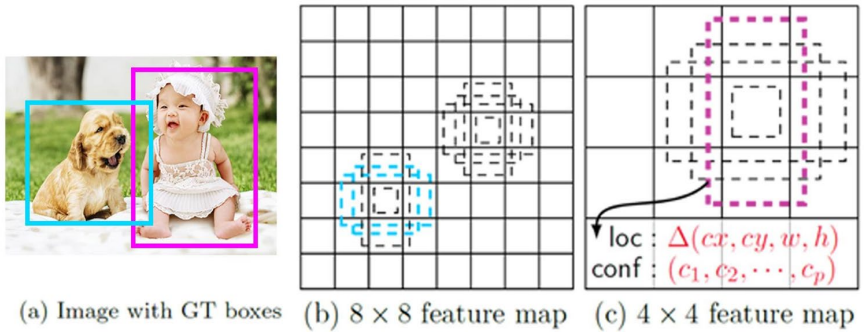
Single-shot detection

- In SSD, like YOLO, only one single pass is needed to detect multiple objects within the image.
- Two passes are needed in Regional proposal network (RPN) based approaches such as R-CNN, Fast R-CNN series. One pass for generating region proposals and another pass to loop over the proposals for detecting the object of each proposal.
- SSD is much faster compared to two-shot RPN-based approaches.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Single-shot detection

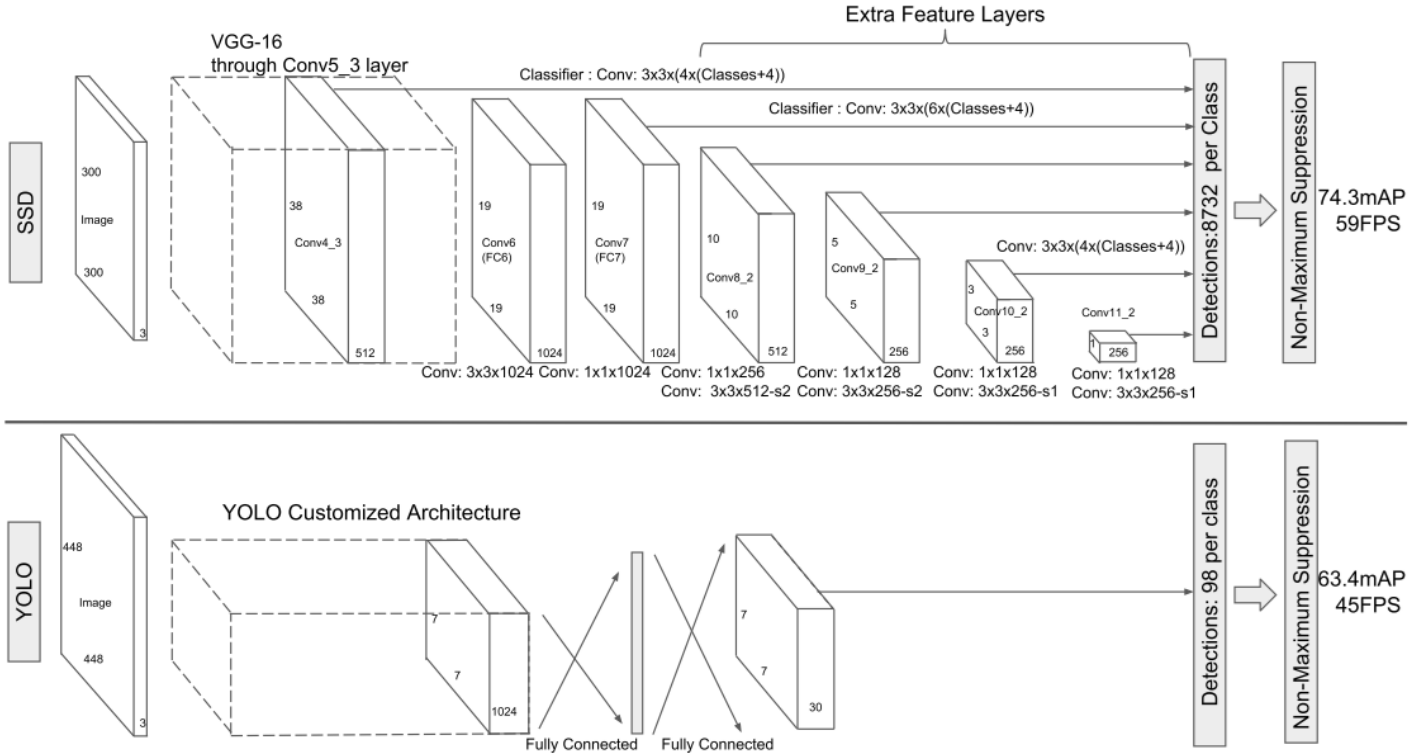
- A feature layer of size m×n (number of locations) with p channels
- For each location, we got k bounding boxes
- For each of the bounding box, we will compute c class scores and 4 offsets relative to the original default bounding box shape.
- Thus, we got (c+4) x k x m x n outputs.



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

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

SSD and YOLO



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

SSD vs YOLO

In the diagram given in previous slide,

SSD model adds several feature layers to the end of a base network, which predict the offsets to default boxes of different scales and aspect ratios and their associated confidences.

SSD with a 300 × 300 input size significantly outperforms its 448 × 448.

YOLO counterpart in accuracy on VOC2007 test while also improving the run-time speed, albeit YOLO customized network is faster than VGG16.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Object Detection - Options

greatlearning
Learning for Life

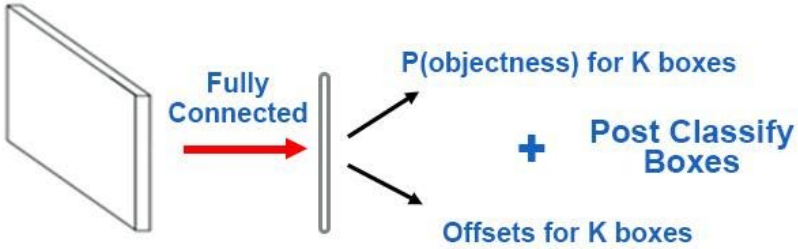
Base Network	Object Detection Architecture
VGG16	Faster R-CNN
ResNet-101	R-FCN
Inception V2	SSD
Inception V3	
Inception	
ResNet	Image Size
MobileNet	# Region Proposals

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

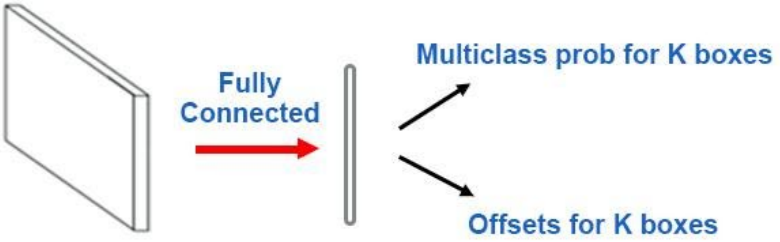
Object Detection - Different Algorithms

greatlearning
Learning for Life

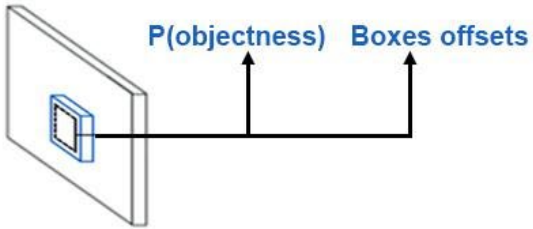
MultiBox [Erhan et al. CVPR14]



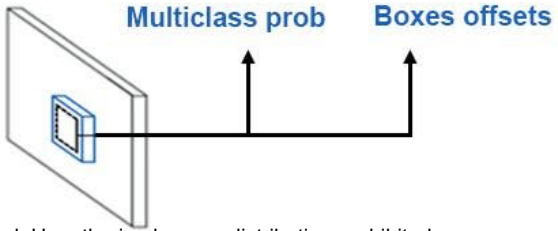
YOLO [Redmon et al. CVPR16]



Faster R-CNN [Ren et al. NIPS15]



SSD



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

References

- Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks
- Object Detection with Deep Learning: A Review
- What makes for effective detection proposals?
- SSD: Single Shot MultiBox Detector
- Selective Search for Object Recognition
- You Only Look Once: Unified, Real-Time Object Detection
- Fast R-CNN
- Rich feature hierarchies for accurate object detection and semantic segmentation

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited