DNN_8.1

Object Detection

Shahrukh Khan

Table of Content

- Object Detection
- History of Object Detection
- RCNN

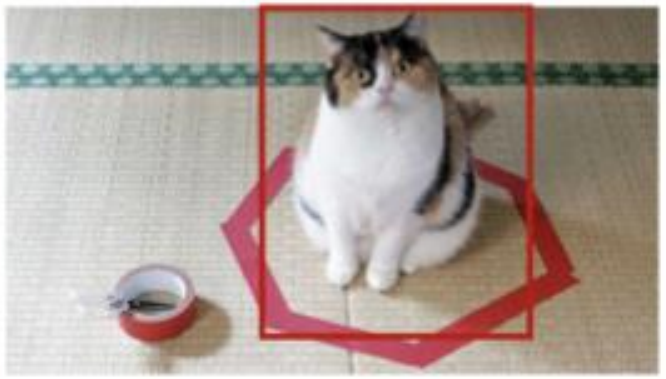
Computer Vision Tasks

1

Is this image of Cat or not?

Image classification problem

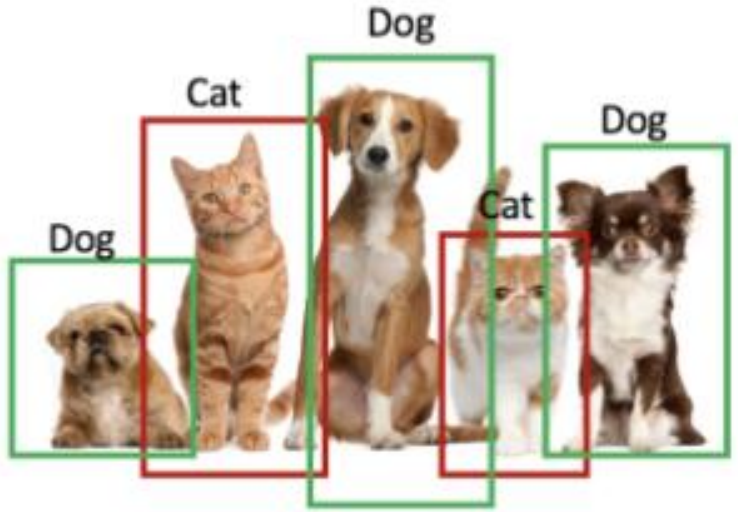
2



Where is Cat?

Classification with localization problem

3

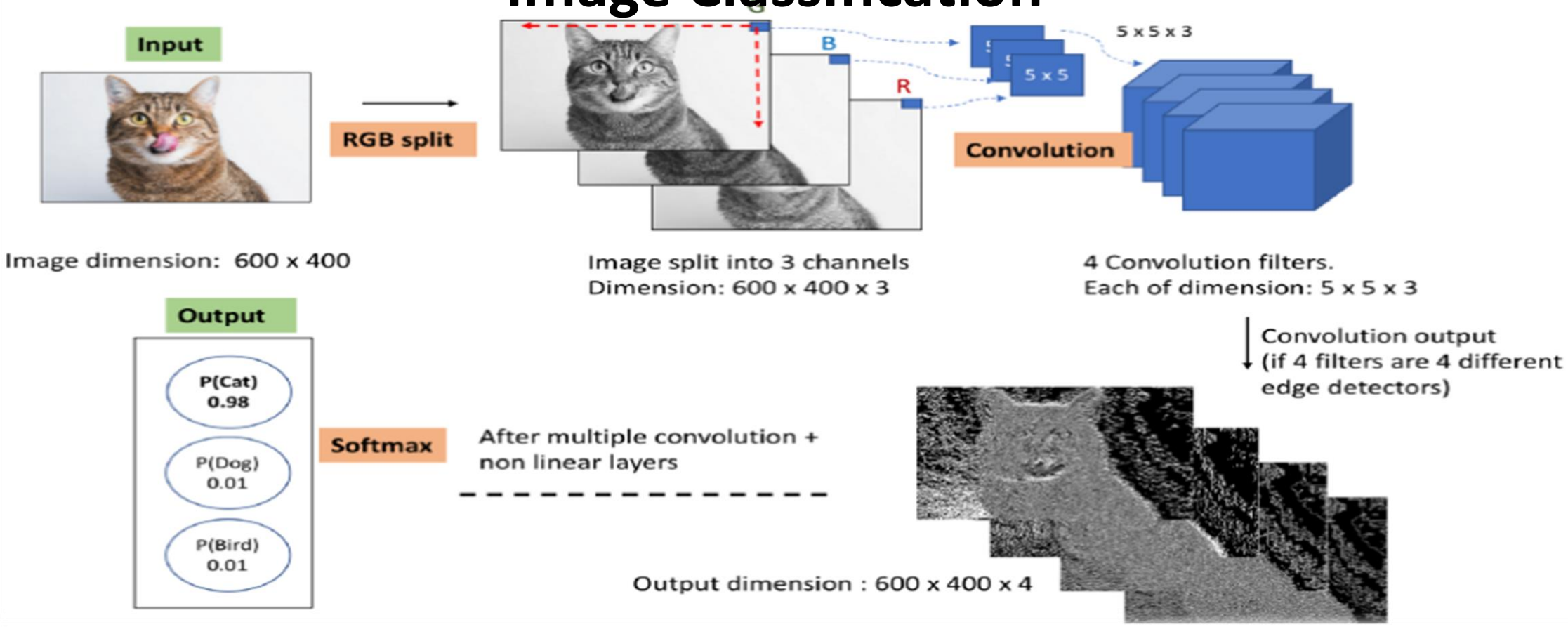


Which animals are there in image and where?

Object detection problem

Computer Vision Tasks

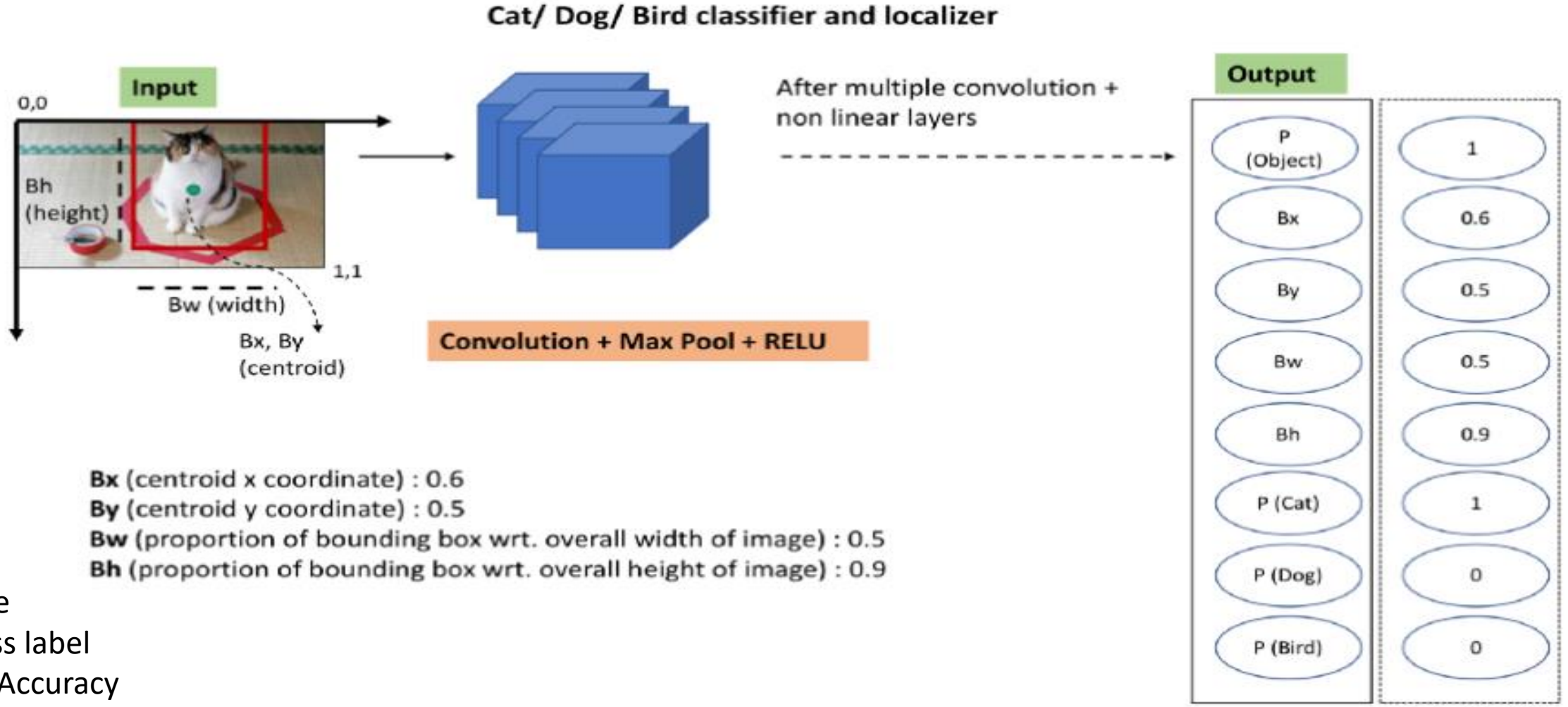
Image Classification



Input : Image
Output: class label
Evaluation: Accuracy

Computer Vision Tasks

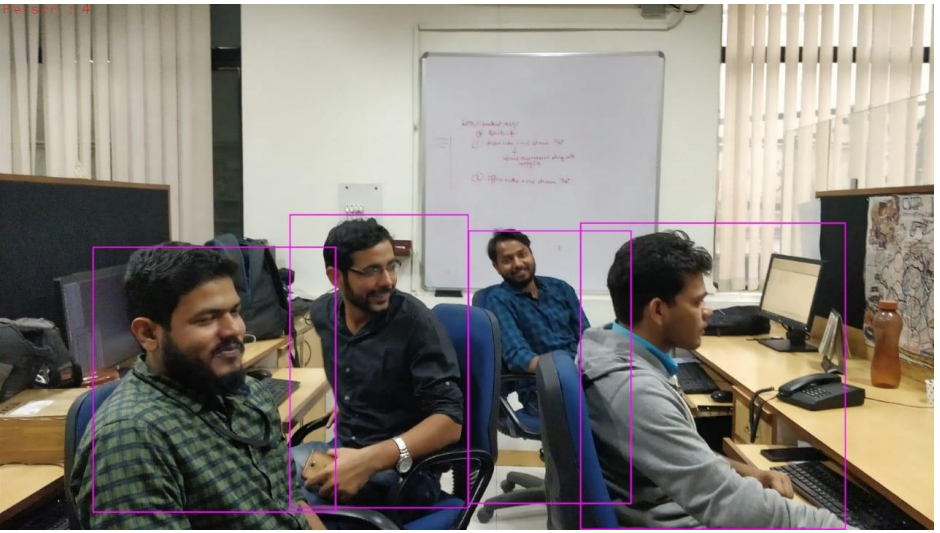
Object Classification and Localization



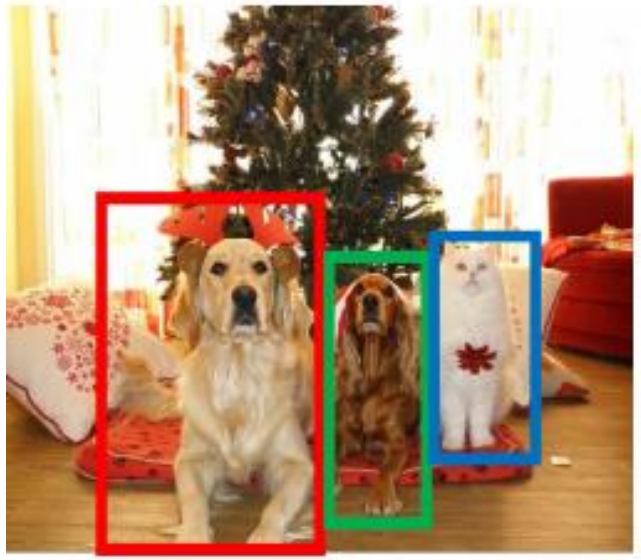
Bx (centroid x coordinate) : 0.6
By (centroid y coordinate) : 0.5
Bw (proportion of bounding box wrt. overall width of image) : 0.5
Bh (proportion of bounding box wrt. overall height of image) : 0.9

Input :image
Output: class label
Evaluation: Accuracy

Multiple objects detection and localization



Person, (x, y, w, h)
Person, (x, y, w, h)
Person, (x, y, w, h)
Person, (x, y, w, h)



Dog, (x, y, w, h)
Dog, (x, y, w, h)
Cat , (x, y, w, h)

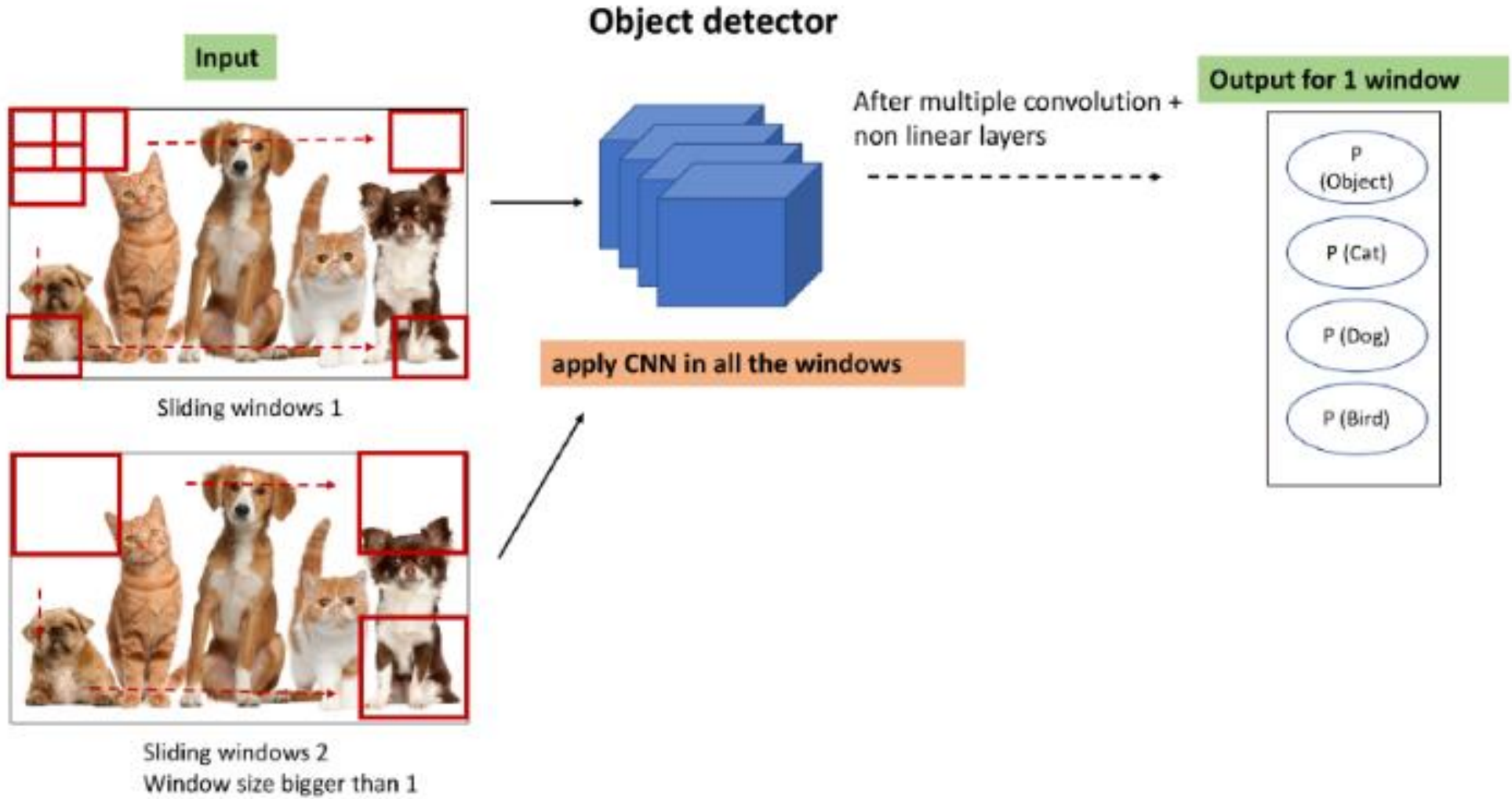
Each image needs a different
number of outputs!

Multiple objects detection and localization

Problems

- Object in image can have different scale.
- Different Aspect ratio.
- Object can be present at any spatial location.

Object detection as Classification : Sliding Window



Object detection as Classification : Sliding Window

- Feature can also be extracted using HOG, LBP, Haar Cascade and can be feed to any M.L model (ex – Face Detection , Pedestrian detection).
- A neural network can be used as a feature extractor and SVM as a classifier.
- Need to apply CNN to huge number of scales and location , Computationally very expensive.
- Object can have different aspect ratio, therefore it generate inaccurate bounding box.

Region-Based Convolutional Neural Network

- RCNN uses selective search to extract bounding boxes(regions of Interest) where objects may be present.
- Selective search produces (~ 2K propsoals) on which we apply CNN for classification.

Region-Based Convolutional Neural Network

Selective Search

There are basically four regions that can form and differentiate between objects.

- Varying Scales
- Varying Colors
- Varying textures
- Enclosures(edges).

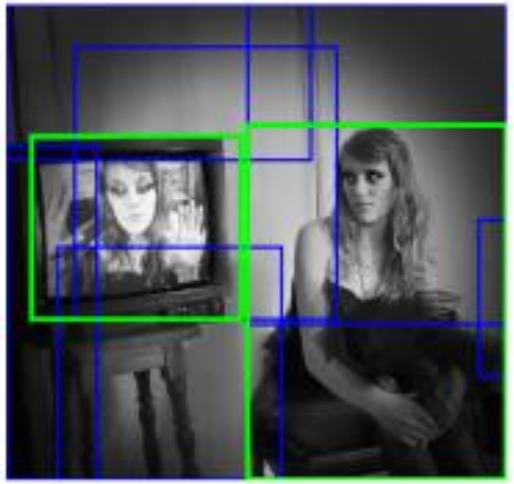
Selective Search proposes various regions based on above properties.

Region-Based Convolutional Neural Network

Selective Search

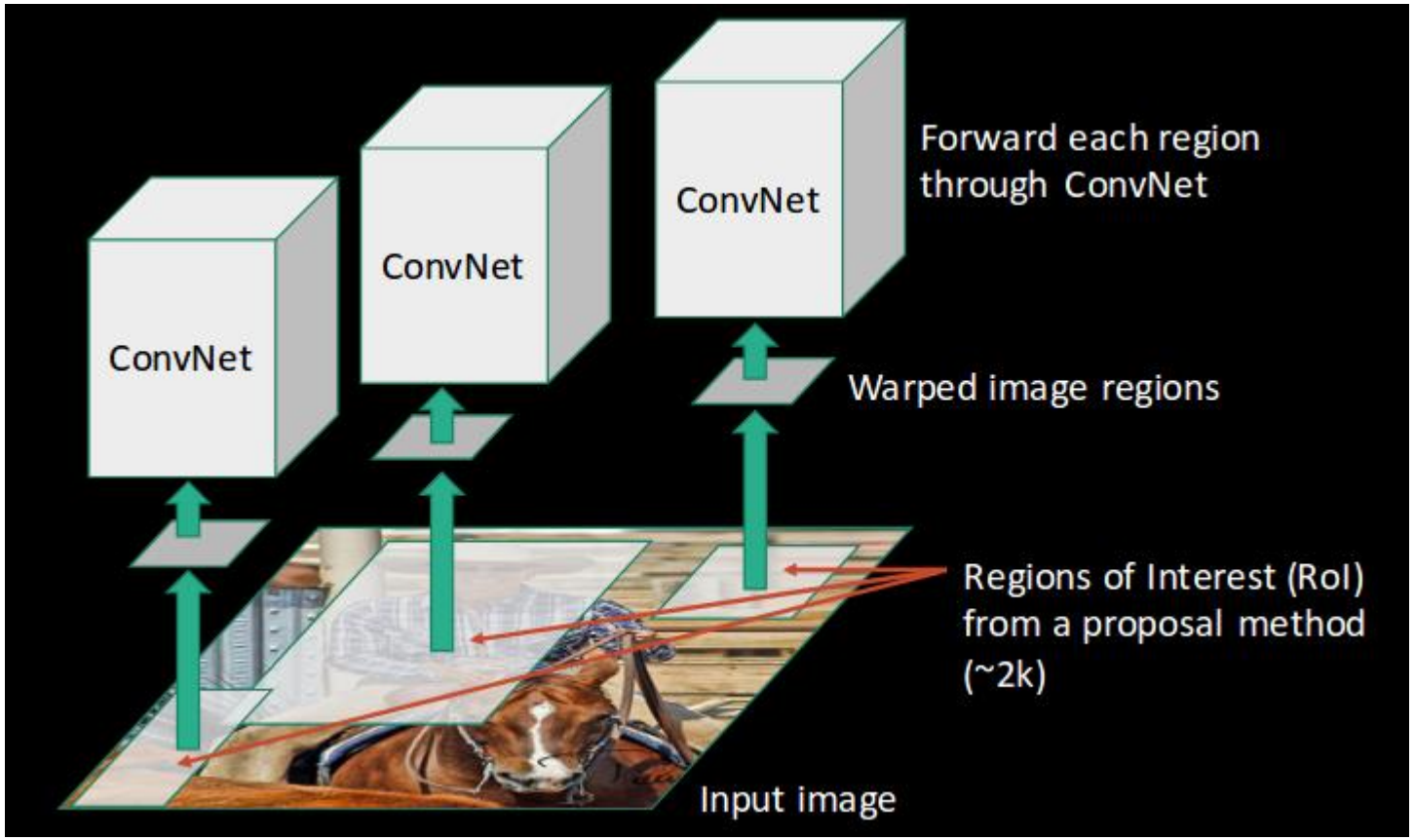
- Generates initial sub-segmentations so that we have multiple regions from this image
- Combines the similar regions to form a larger region (based on color similarity, texture similarity, size similarity, and shape compatibility)
- Finally, these regions then produce the final object locations (Region of Interest).



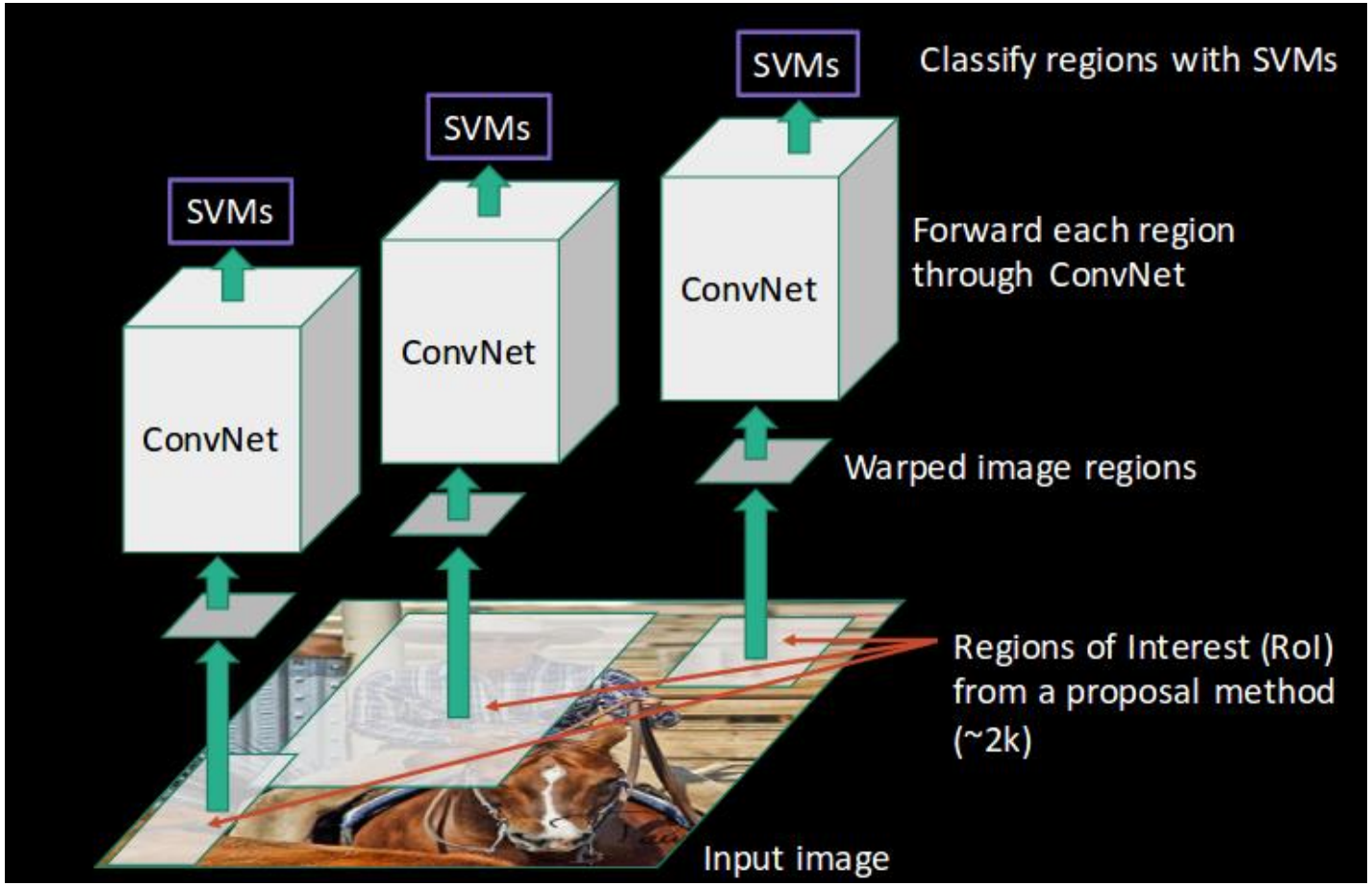



Region-Based Convolutional Neural Network

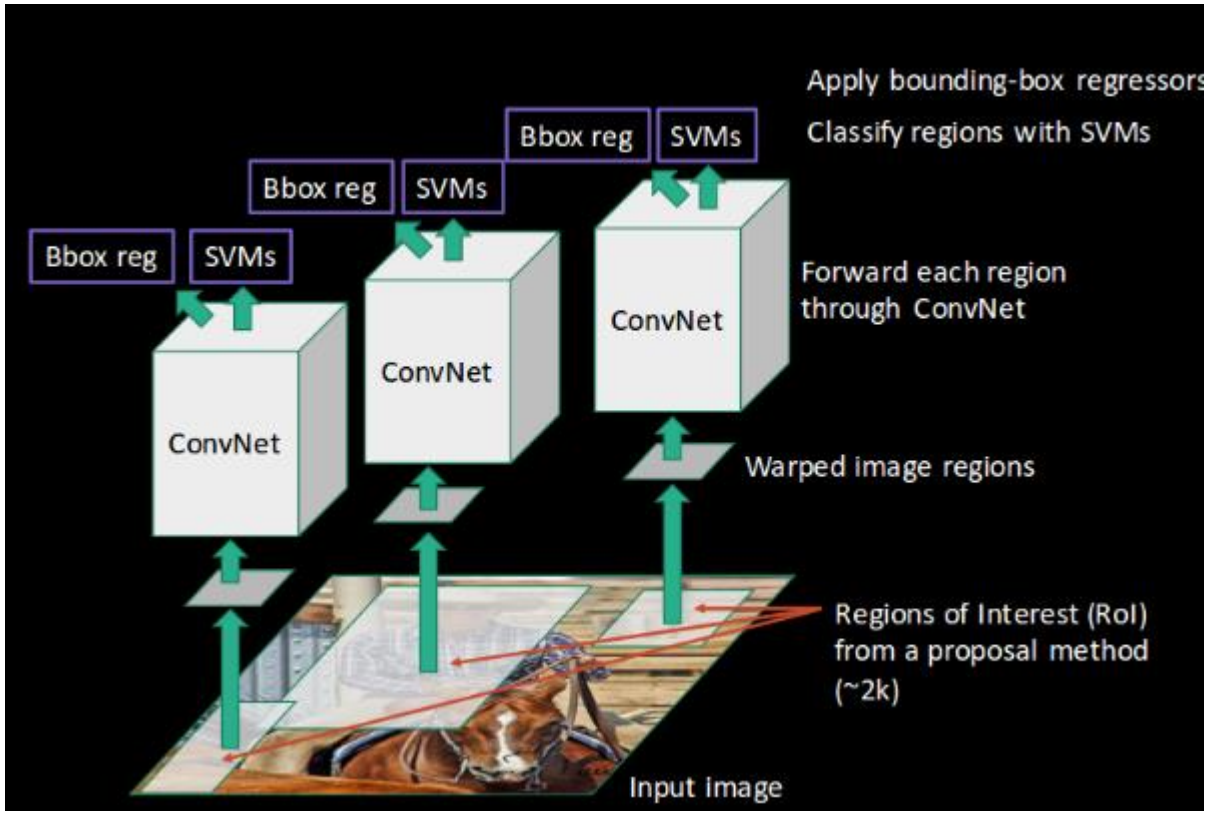
Region-Based Convolutional Neural Network



Region-Based Convolutional Neural Network



Region-Based Convolutional Neural Network

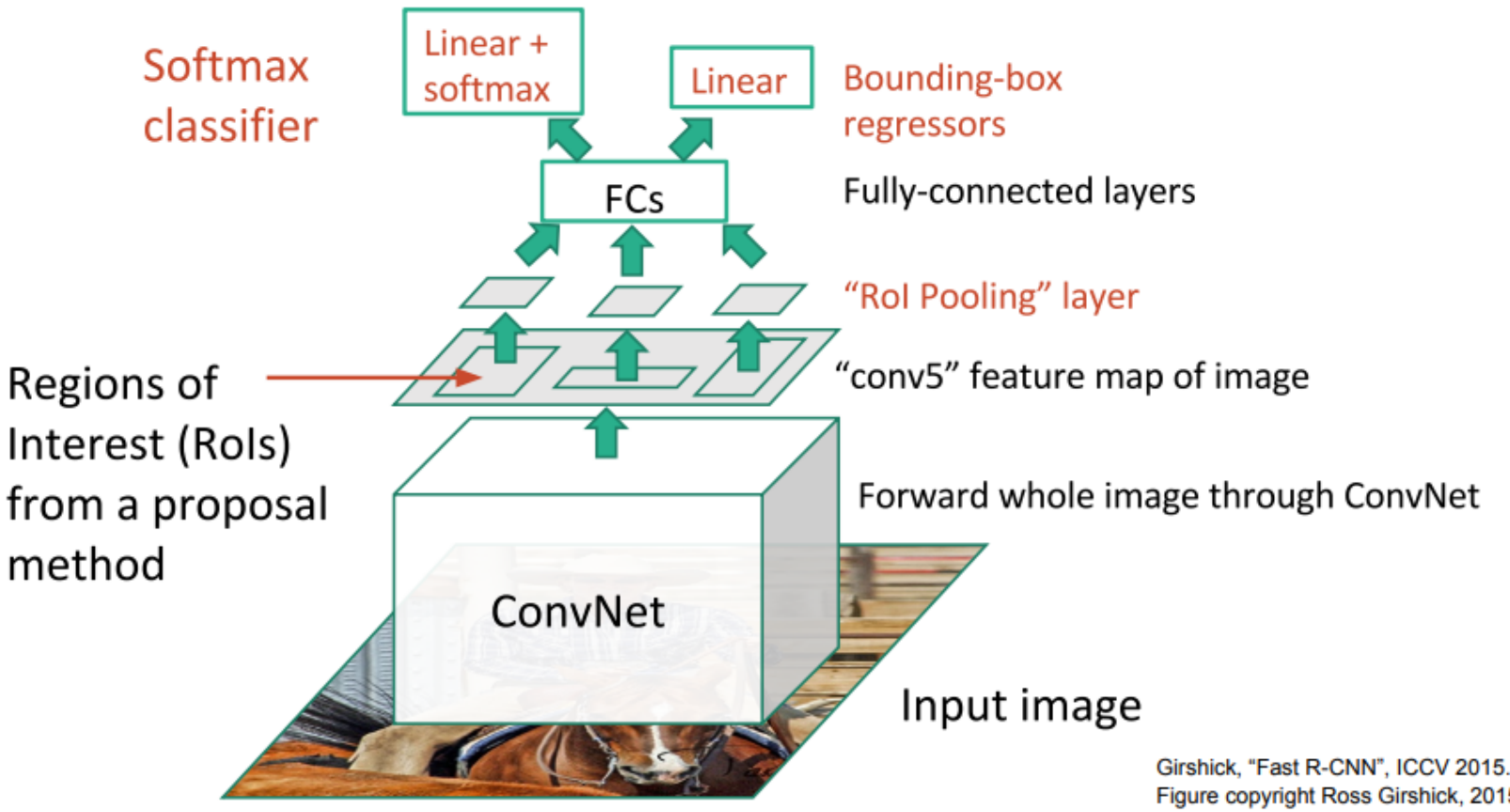


Region-Based Convolutional Neural Network

Problems

- Extracting 2,000 regions for each image based on selective search
- Extracting features using CNN for every image region. Suppose we have N images, then the number of CNN features will be N*2,000
- Training is slo
- Inference (detection) is slow • 47s / image with VGG16

Fast R-CNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; source. Reproduced with permission.