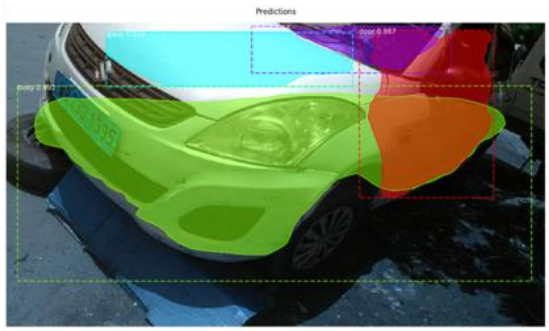
The Four Wheeler & Two Wheeler valuation engine will be able to do the following:

1. Identify car make & model
2. Basis the Smart Card, ascertain the year of manufacture
3. Calculate the normal depreciation rules
4. Calculate the KMs travelled and arrive at the wear & tear of the engine, suspension etc.
5. On top of the normal depreciation, the engine will process the images and arrive at
    a. Damages like scratch or dent
    b. Cost of repair, paint etc.
    c. Do a comparison between repair vs replacement
    d. Arrive at the final probable cost for sale (post necessary adjustment of the repair/repaint/replacement on the depreciated value)

Below are some actual images where the valuation engine has identified the damage sections and have arrived at a repair/replacement/repaint cost.



Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Scratch
(on closer images)

➔ Severity = Area of masked damage *divided by* area of masked part (Area ~ No. of pixels)
➔ Severity lies between 0 to 1
➔ Example:

27195 Pixels

68893 Pixels

Severity of bumper dent
= 27195 / 68893
= 0.395

Example criteria for cost estimation:

- **Scratch** -- *always repairable* -- repaint cost
- **Glass Shatter** -- *always replaceable* -- market cost of the part
- **Bumper dent and door dent** --
  if: severity > maximum severity limit
  => *replace* -- market cost of the part
  else:
  => *repair* -- (rate of repair * severity * 100)

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

The valuation engine will be trained on historical data (images and valuations) as are available from portals like Cars24 etc. the engine. This will help the engine to

- Identify the various make & model
- Identify the various components – car parts – metal and non-metal
- Understand the various cost structures for repair/replacement etc.
- Superimpose KMs run for wear & tear assessment of engine, suspensions etc.

The engine will use highly sophisticated AI & ML techniques like

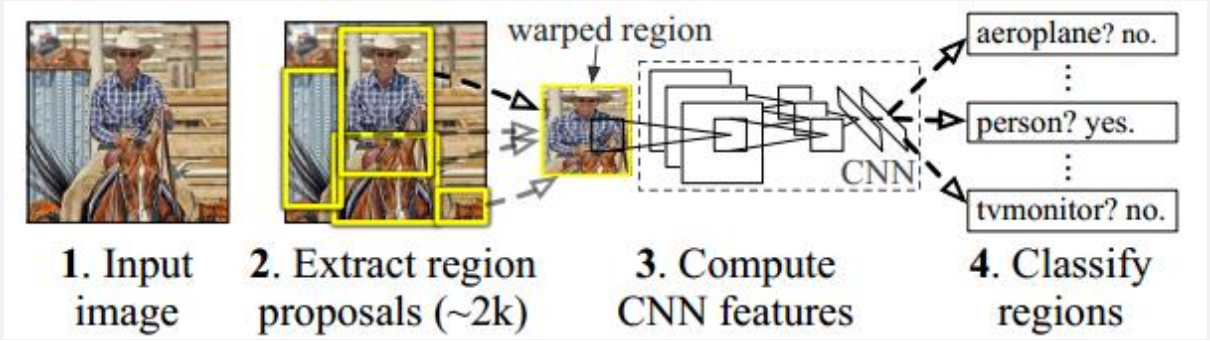
- Convolutional Neural Network (CNN)
- Recursive CNN
- SVM

The appendix section covers how the algorithm will work and the efficacy/efficiency of the same.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Appendix

This algorithm does object detection in the following way:

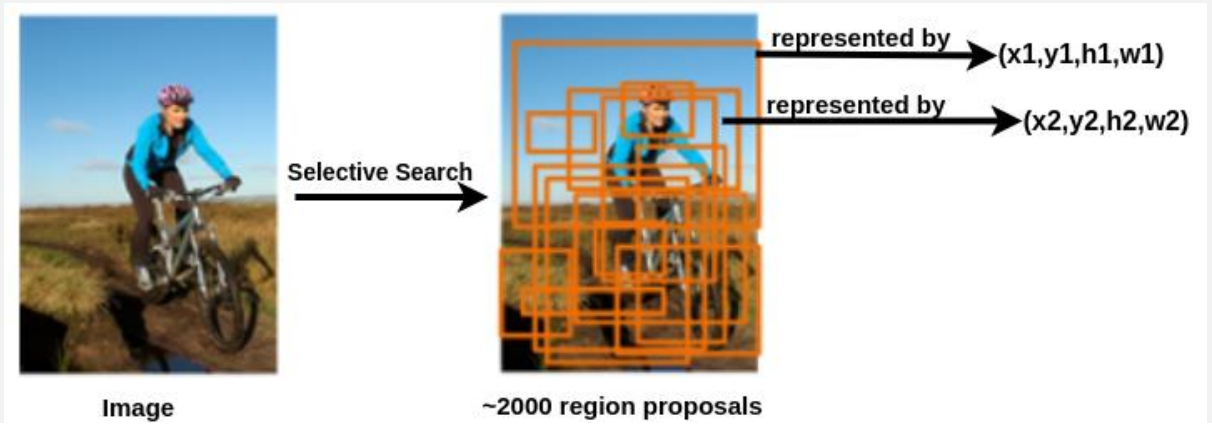


1. The method takes an image as input and extracts around 2000 region proposals from the image(Step 2 in the above image).
2. Each region proposal is then warped(reshaped) to a fixed size to be passed on as an input to a CNN.
3. The CNN extracts a fixed-length feature vector for each region proposal(Step 3 in the above image).
4. These features are used to classify region proposals using category-specific linear SVM(Step 4 in the above image).
5. The bounding boxes are refined using bounding box regression so that the object is properly captured by the box.

Now the post will dive into details explaining how the model is trained and how it predicts the bounding boxes.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Calculating region Proposals

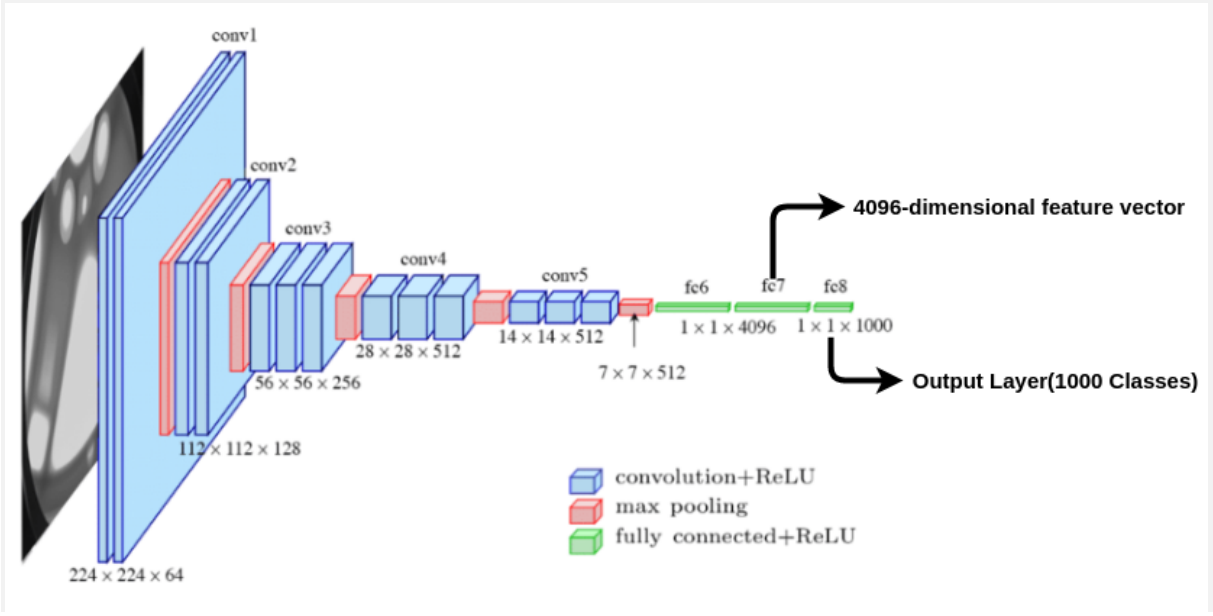


Region Proposals | Source: Image by author

Region proposals are the bounding boxes that may contain an object. These are represented by a tuple of 4 numbers(x,y,h,w). The (x,y) are the coordinates of the centre of the bounding box and (h,w) are the height and width of the bounding box respectively. These region proposals are calculated by an algorithm called selective search. For an image, approximately 2000 region proposals are extracted.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Training CNN feature extractor:



Pretrained network: To train the CNN for feature extraction, an architecture like VGG-16 is initialized with the pre-trained weights from imagenet data. The output layer having 1000 classes is chopped off. So when a region proposal image(warped to size 224x224) is passed to the network we get a 4096-dimensional feature vector as shown in the above image. In this way, each region proposal is represented by a 4096-dimensional feature vector.

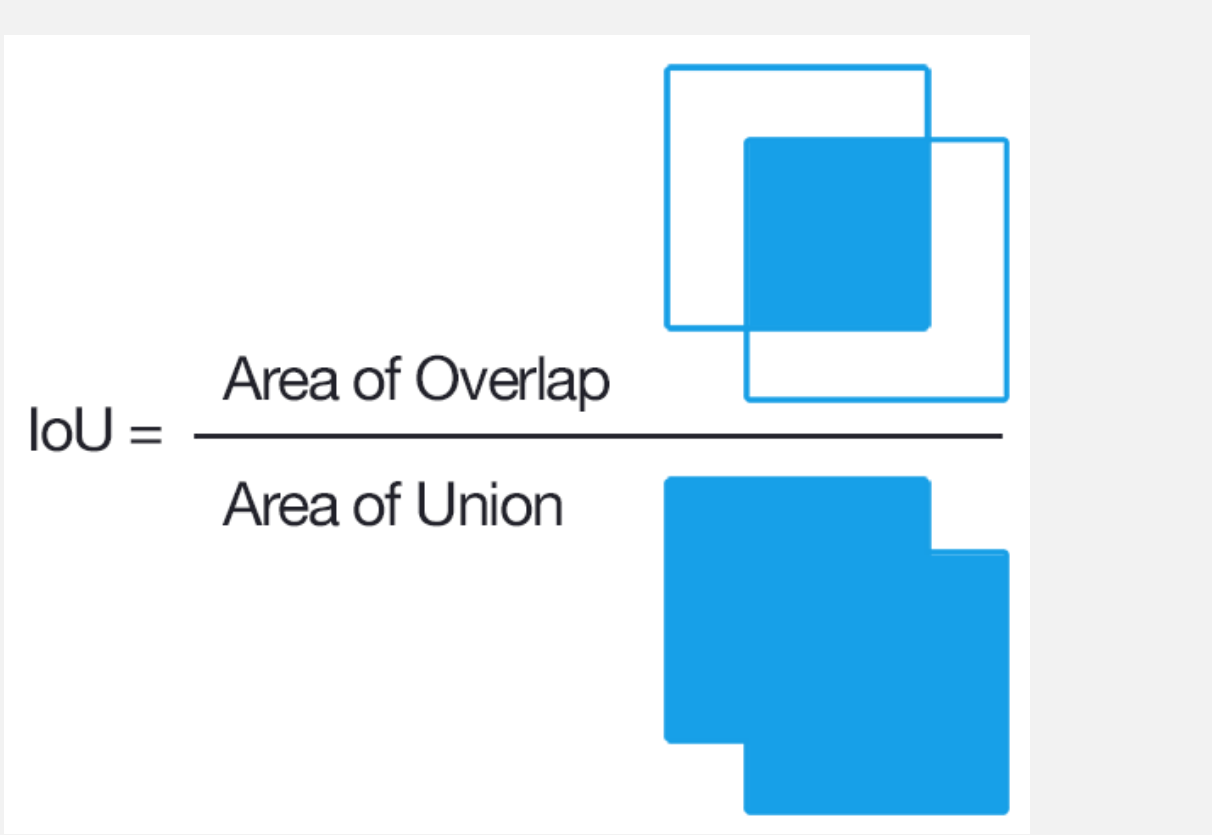
The next step is to fine-tune the weights of the network with the region proposal images. To understand this a new metric called intersection-over-union or IoU score will be introduced.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Intersection-Over-Union(IoU)

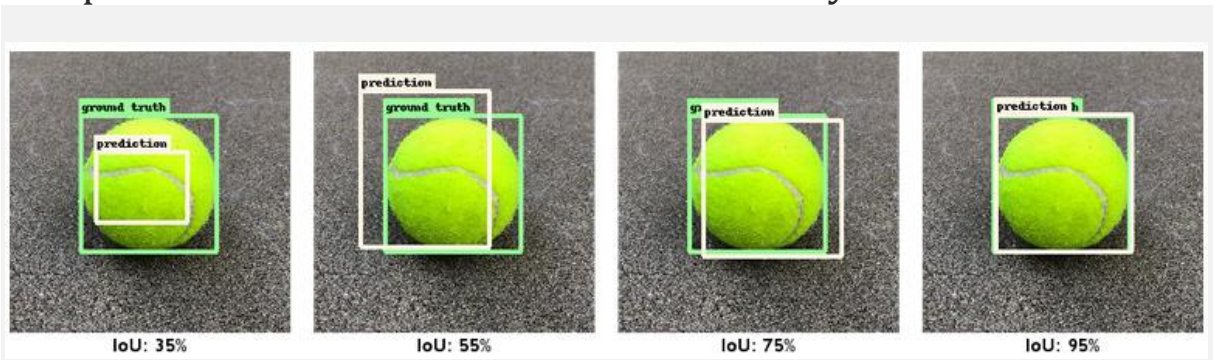
To measure the performance of a classification model, we generally use metrics like accuracy, recall, precision etc. But how to measure the performance of object detection. In object detection we have to evaluate two things:

1. How well the bounding box can locate the object in the image. In other words, how close the predicted bounding box is to the ground truth.
2. Whether the bounding box is classifying the enclosed object correctly



Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

The IoU score measures how close the predicted box is to the ground truth. It is the ratio of the area common to ground truth and predicted box to the total area enclosed by both the boxes.



In the left-most image, it can be seen that the predicted box is not close to the ground truth so the IoU score is only 35% while in the right-most image the predicted box completely overlaps the ground truth box, hence a very high value of 95% is obtained. The IoU value varies from 0 to 1.

Fine-tuning the network: To fine-tune the model the output layer having 1000 classes is replaced with N+1 classes(softmax layers) and the rest of the model is kept unchanged. N is the number of the distinct classes the objects be classified and plus 1 for the background class.

Next, data is required for fine-tuning. The region proposals whose IoU > 50% are considered a positive class for that object and rest are considered as background. In the above image of balls, the

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

region proposal with 35% IoU score will be labelled as background while the rest of the boxes will be labelled as ball. These images(region proposals) are warped(resized) to the size compatible with CNN, in case of VGG 16 it is 224x224. Using these images the weights of the network are fine-tuned.

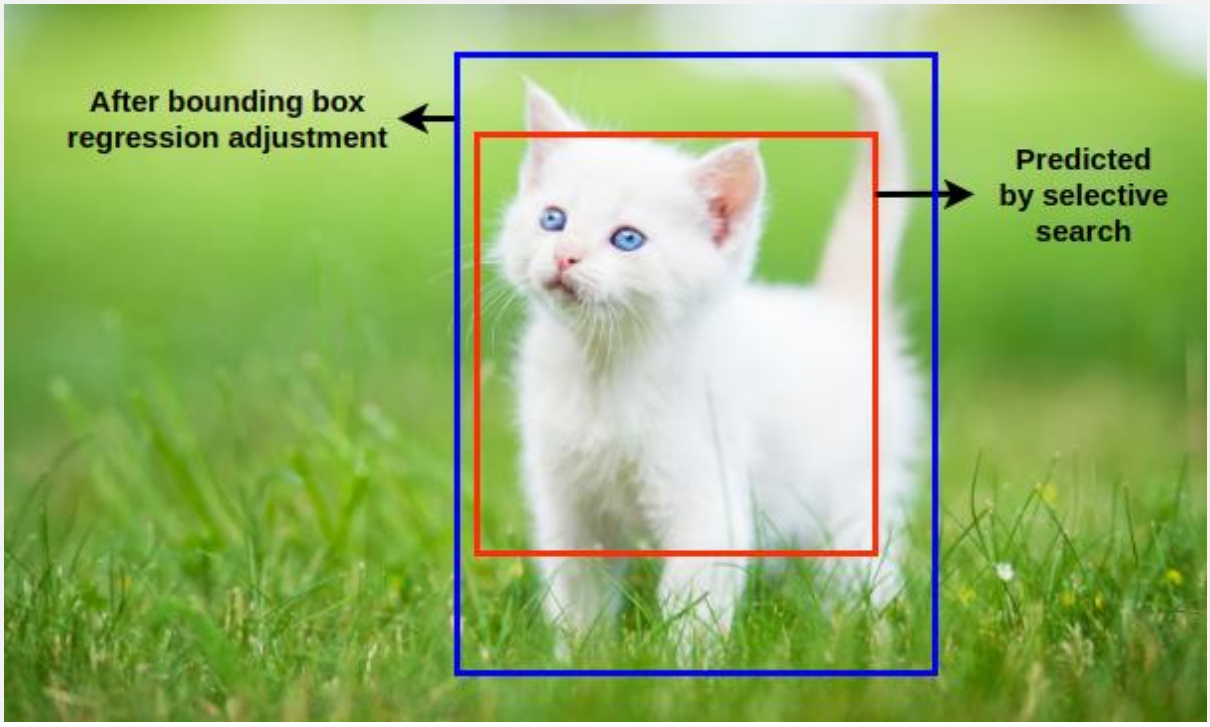
Training class-specific SVM

Once the 4096-dimensional feature is obtained for each region proposal, the next task is to train a binary SVM for each class. For example, if the detection model is to detect three distinct objects — cat, dog and people, then three SVMs need to be trained for each of the class.

Data preparation: All the object proposals belonging to a particular class is segregated. The manually labelled ground truth image for the class is considered as a positive class while object proposals having IoU < 30% for that class is considered as negative class. The same thing is done for each class and N SVM models are trained for classifying each region proposal into N classes.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

Bounding box regression



The region proposal bounding box predicted by selective search[2] algorithm might not be able to capture the entire object. To fine-tune the predicted bounding box, bounding box regression is used.

Consider the ground truth region proposal G and the predicted region proposal P by the selective search[2] algorithm.

$$G = (G_x, G_y, G_w, G_h)$$

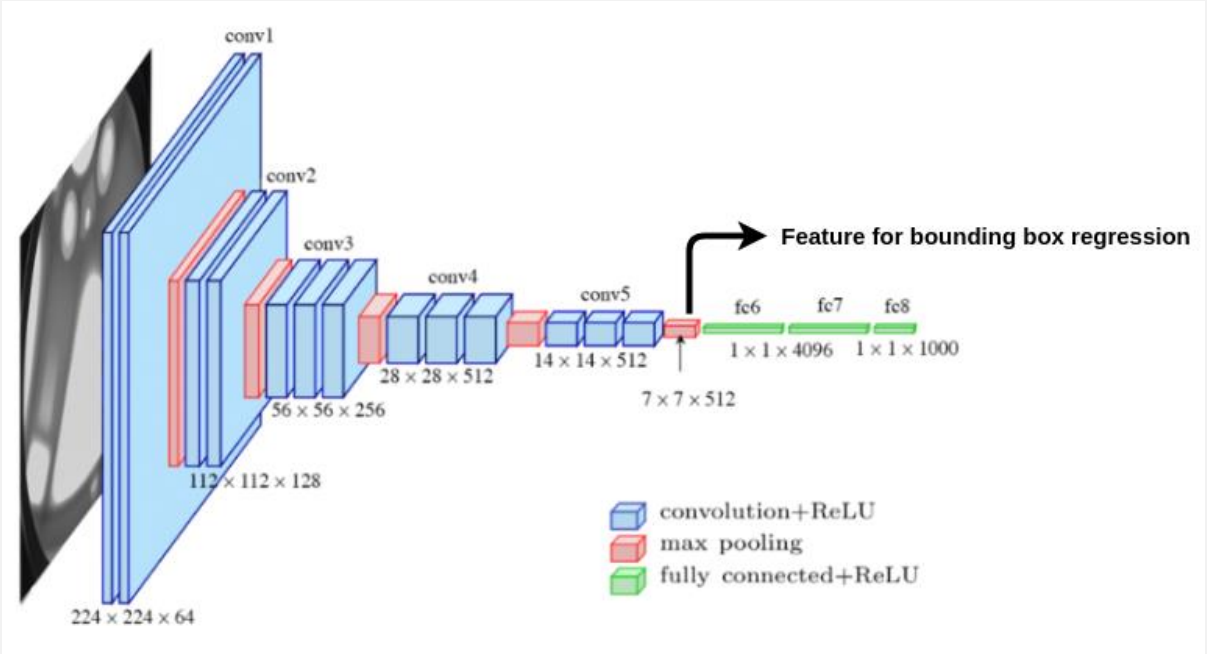
$$P = (P_x, P_y, P_w, P_h)$$

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

To make the prediction of G scale-invariant, the following transformations are done such that the target for the regression is *t*.

$$t_x = (G_x - P_x)/P_w$$
$$t_y = (G_y - P_y)/P_h$$
$$t_w = \log(G_w/P_w)$$
$$t_h = \log(G_h/P_h).$$

Transformation equation



The input to the regression model is the feature from the last pooling layer of the CNN. We train 4 regression models per class with the target as *t* and input features as last pooling layer feature from CNN to learn regression parameter w.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

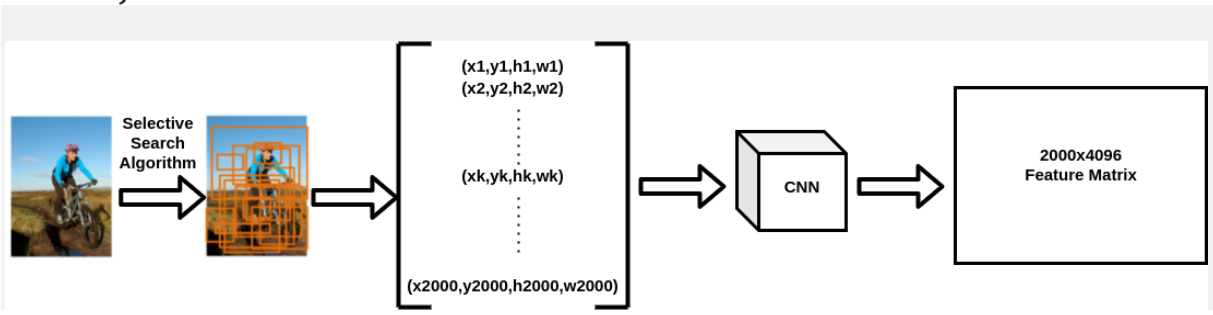
$$t_* = w_*^T \phi(P)$$

Regression equation

Here * is the placeholder for (x,y,w,h) and phi(P) is the last pooling layer feature corresponding to proposal P. So to predict ground truth G we can use the regression equation to calculate t from the region proposal P and then substitute the value of t and P in transformation equation to obtain G.

Prediction

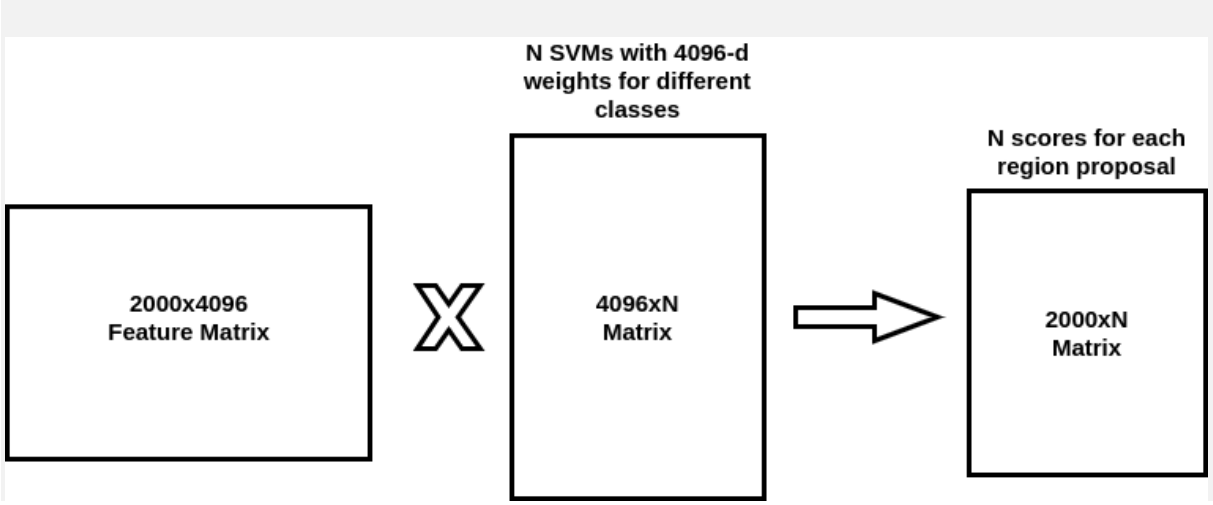
Once different parts of R-CNN are trained, the next part is to do the object detection.



1. An input image is taken and using selective search[2] algorithm, around 2000 region proposals are obtained for an image.
2. Each region proposal image is warped to a fixed size of 224x224.
3. These region proposal images are then passed to the trained CNN to obtain 4096-dimensional feature vector for all the

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

2000 region proposals which result in 2000x4096 dimensional matrix.



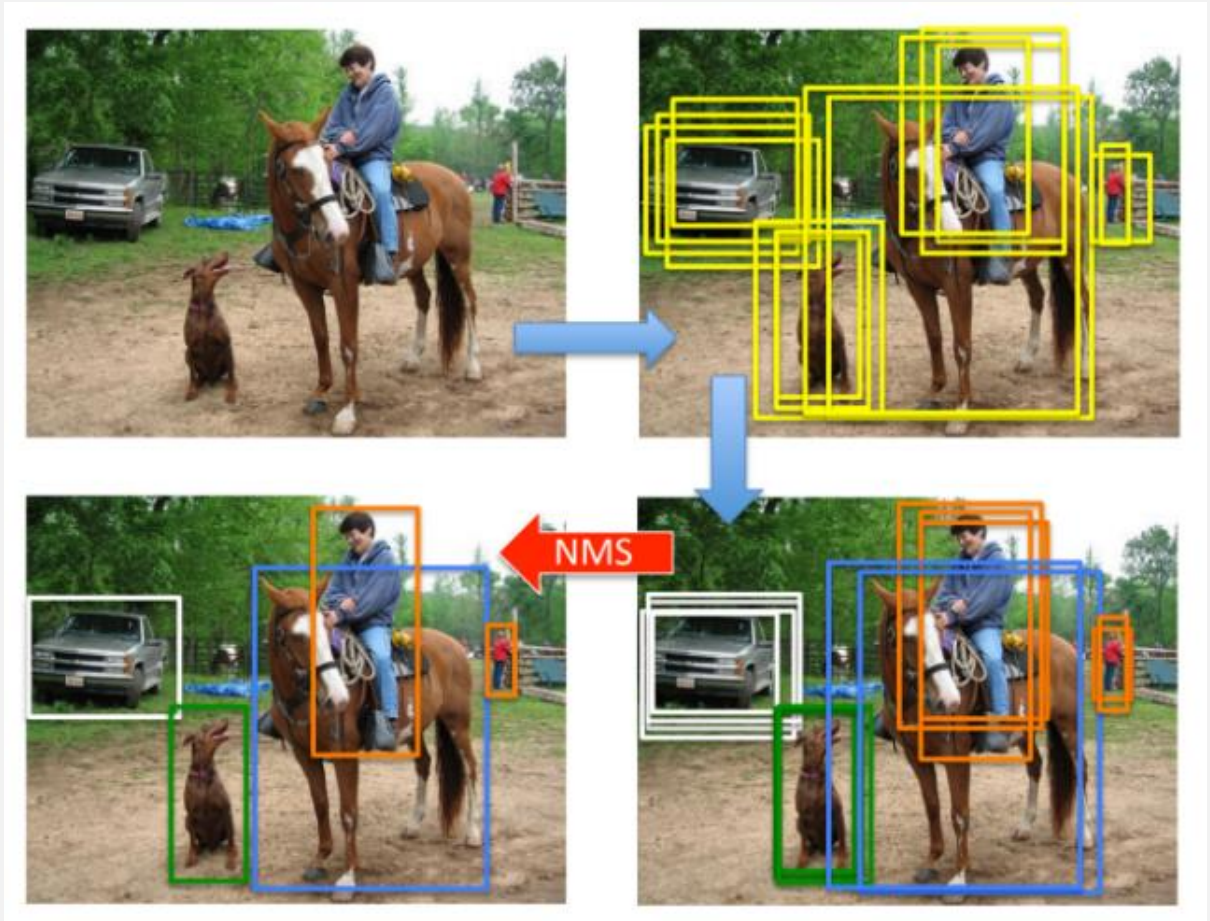
3. Each region proposal is classified using SVM for each class. generally for N classes, the SVM weights(4096-dimensional) are stacked in the form of a matrix and multiplied with the feature matrix. This results in a matrix which assigns a score to each class to which a region proposal can belong to.

4. The proposal is assigned the class which receives the maximum score. So all the 2000 region proposals or bounding boxes in the image are labelled with a class label.

5. Out of those many bounding boxes, a lot of them would be redundant and overlapping bounding boxes which need to be removed. To accomplish that Non-maximum suppression algorithm is used.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

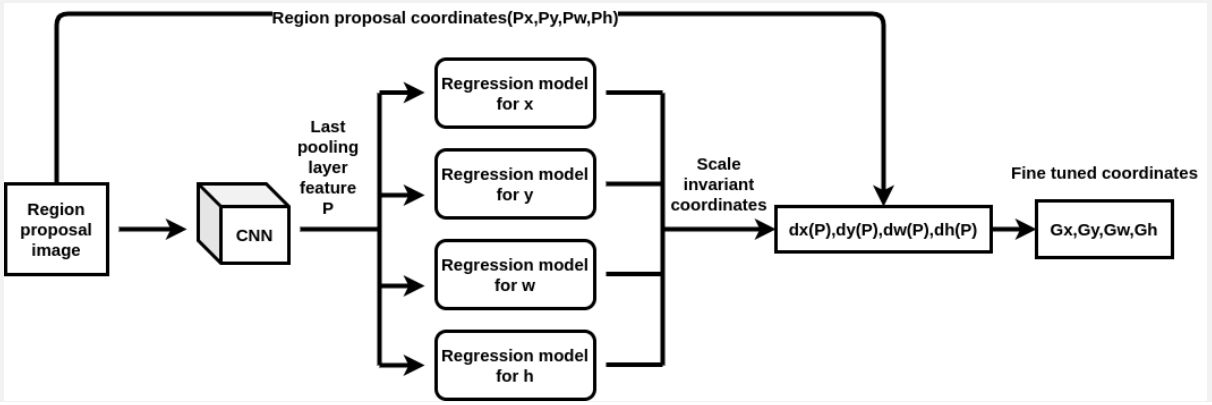
Non-maximum suppression algorithm:



Non-maximum suppression is a greedy algorithm. It works on one class at a time. For a particular class, it picks the box with the maximum score obtained using SVM. Then it calculates IoU score with all other bounding boxes belonging to that class. The boxes having IoU score greater than 70% are removed. In other words, the bounding boxes which have very high overlap are removed. Then the next highest score box is chosen and so on until all the overlapping bounding boxes are removed for that class. This is done for all classes to obtain the result as shown above.

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.

6. Once the labelled bounding boxes are obtained the next task is to fine-tune the location of the boxes using regression.



Fine-tuning region proposals through regression| Source: Image by author

During training,4 regression models were trained for each class. So for a particular bounding box, region proposal image is passed through CNN to obtain features P to be passed to the regression model. The regression model outputs the scale-invariant coordinates (dx(P), dy(P), dw(P), dh(P)). These coordinates are combined with the region proposal coordinates (Px, Py, Pw, Ph) to obtain the adjusted final coordinates (Gx, Gy, Gw, Gh) using the formula below.

$$\hat{G}_x = P_w d_x(P) + P_x$$
$$\hat{G}_y = P_h d_y(P) + P_y$$
$$\hat{G}_w = P_w \exp(d_w(P))$$
$$\hat{G}_h = P_h \exp(d_h(P)).$$

Private & Confidential – All rights reserved with Pentation Analytics Pvt Ltd.