RCNN

RCNN (region-based CNN) is used for object detection where we might have different instances of different classes in an image. They were invented after CNN had gained popularity to improve upon it, and is a milestone for building other advanced and improved models like Fast RCNN, Mask RCNN, each that improved on the original RCNN architecture.

Traditional Convolutional Neural Networks (CNNs) struggle to handle the frequency of object occurrences and multi-object scenarios. A brute-force approach using a sliding window to select regions and apply CNNs is computationally expensive

# How RCNNs work?

1. It divides the image into region candidates with selective search (2000 regions) using Selective Search:

* The algorithm starts by performing an initial sub-segmentation of the input image to generate the Initial Segmentation.
* It then recursively combines similar bounding boxes into larger ones. Similarities are evaluated based on factors such as color, texture, and region size.
* Finally, these larger bounding boxes are used to create region proposals for object detection

1. It warps these candidates into a fixed size the CNN expects. Because, just like any neural network layers, the CNN layers take inputs of fixed dimensions, so we need to warp the dimensions of those regions into a dimension the CNN expects.
2. It passes these warped regions into a CNN to extract features. The CNN originally used was Alexnet, for image classification, but any trained CNN can do the job to extract features vectors
3. These extracted features vectors are fed into another machine learning model for classification (like svm) which will give us our labels for each class on the image.
4. Then we perform bounding box regression, to refine the location and size of the bounding box. One bounding box regression algorithm typically used is Non-Maximum Suppression (NMS):
   * Removes proposals with confidence scores below a threshold (e.g., 0.5).
   * Selects the highest-probability region among candidates for each object
   * Discards overlapping regions with an IoU (Intersection over Union) above 0.5 to eliminate duplicate detections
5. After we apply NMS, we should be left with one bounding box at a more precise location and size to where the object is, instead of many overlapping bounding boxes with imprecise sizes and locations

# Pros and Cons

The main advantages of using R-CNN is that it can handle many object classes, and put labels on many objects in each image, compared to a normal CNN where it gives out only one label per image, and it is a faster approach than using sliding windows, as the fewer regions require less computational power.

The main drawback is, RCNNs are still computationally expensive, for dividing the image into many regions, and then infering the feature vectors extracted from these regions. It also is prone to errors, as the whole process is not end-to-end. We have the region candidates generation, feature extraction by trained model, and then inference of those features into another model, where each step is prone to errors and inaccuracies.