Object detection

# Overview

**Key Concepts**

Sliding Windows

Bounding Box

Bounding Box Pipeline

Score

**Score Image Classification**

* SIC predicts the class of an object in an image.
* Object Localization locates the presence of and object, indicating the location with a bounding box.
* Generally, to classify an object, all possible objects in the image must be found. To do this an algorithm called **Sliding Window** **Detection** is used. The window can be considered a sub-image of the main image.
* The window starts in one region of the image, the region is classified and then the process repeats shifted a few pixels (horizontally then vertically).
* The SWD provides a systematic approach to object detection.
* Any objects not of the class that is sought are considered background.
* In addition to challenges with classification, there are additional challenges specific to object detection namely, the object detectors often identify multiple overlapping detections.
* Object sizes also provide issues. How big should the box be?
* Reshaping and resizing an image is a solution.
* Objects within pictures may also overlap.
* **Bounding boxes** are another method of object detection.
* It can be used on its own, in conjunction with sliding windows or with more complex object detection methods.
* A bounding box is rectangular with coordinates x 0, y 0 and of dimensions m \* n
* The goal of object detection is to predict the coordinates of the upper left and lower right of the bounding box.
* The trained data will therefore be a set of classes and their bounding boxes.
* Models will normally offer a confidence score for its prediction.
* Usually models will only output objects above a confidence threshold.

# Object Detection with Haar Cascade Classifier

* A machine learning method, trained on a large number of positive images, using the idea of Haar Wavelets in the feature classifier.
* Haar Wavelets are convolutional kernels used to extract features.
* They extract information about edges, lines and diagonals.
* The algorithm selects a few important features and is highly efficient, using an Adaboost
* The idea of an Adaboost is that weights are set to both classifiers and samples in a way that forces classifiers to concentrate on things that are difficult to correctly classify.
* In this way, only features that help improve classification accuracy by creating a strong classifier out of a series of weaker classifiers.

# Object Detection with Deep Learning

Object detection with CNNs.

**Object Detection Prediction**

A sliding window and a CNN can be used for object detection. The image is classified as background or part of another class. Usually popular CNN Architectures pretrained on ImageNet are used (image below)

A diagram of a diagram of a diagram

Description automatically generated

Figure : CNN Architecture with Softmax output.

A diagram of a diagram of a structure

Description automatically generated

Figure : CNN Architectures with the SoftMax and box prediction on the output layer.

Each neuron outputs a different component of the box. Fig 3 depicts the relationship between the bounding box and the object. This output is collectively referred to as ‘box hat’.

A dog sitting on a white surface

Description automatically generated

Figure : Relationship with the corresponding bounding box and neurons, we use oversized pixel indexes for clarity.

Unlike classification, the output values of the neuron are real numbers aka. The coordinates.

A dog sitting on a grid

Description automatically generated

Figure : Real numbers output values of box neurons

To predict the class of the bounding box, softmax layers are used, utilizing probability or output for activation.

A dog with a frisbee on its head

Description automatically generated

Figure : softmax layers used to predict the class of bounding box for three classes

In fig 6, the example shows a bounding box. To find the class of the bounding box, the softmax layer is used. As dog is the highest probability, dog is the classification.

A dog with a bird on its head

Description automatically generated

**Training for object detection**

|Training in object detection has 2 objectives:

1. Determine the learnable parameters for the box
2. Determine the bounding boxes class.

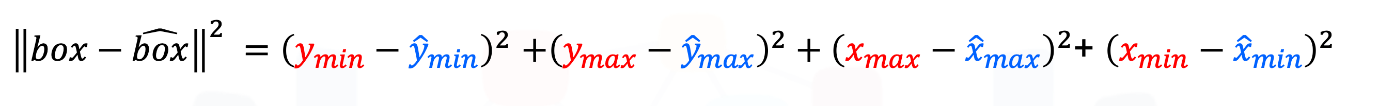
To determine the learnable parameters for the bounding box, L2 or squared loss is used. This calculates the squared differences between the actual box value and predicted.

A dog sitting on the floor

Description automatically generated

Figure : the L2 Loss Function calculates squared differences between the actual box value and the predicted box.

Total loss for each box is given as



Finally for determining classification, the L2 cost is combined with the cross entropy loss in a weighted sum, called the Multitask loss. Gradient descent is then used to minimize the cost.

**Types of object detection**

Sliding windows are however slow.

There are two main types of object detection that sped up the process.

1. Region-based object detection. Breaks the image up into regions and performs a prediction.
2. Single-stage uses the entire image.

Region based CNNs (R-CNN) are more accurate but slower.

Single-Stage methods are faster but less accurate, such as Single-Shot Detction (SSD) or You only look once (YOLO)