

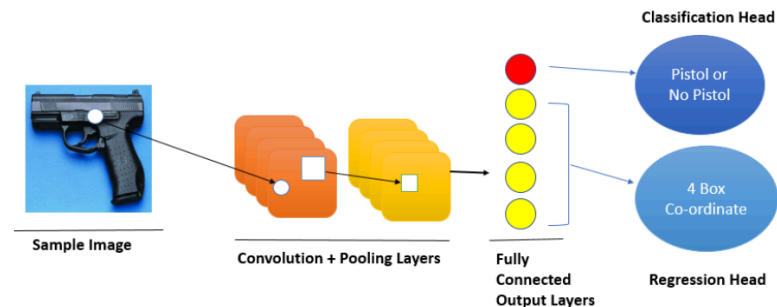
## What is Object Detection?

Object detection is a computer vision task to locate and identify objects in an image or video.

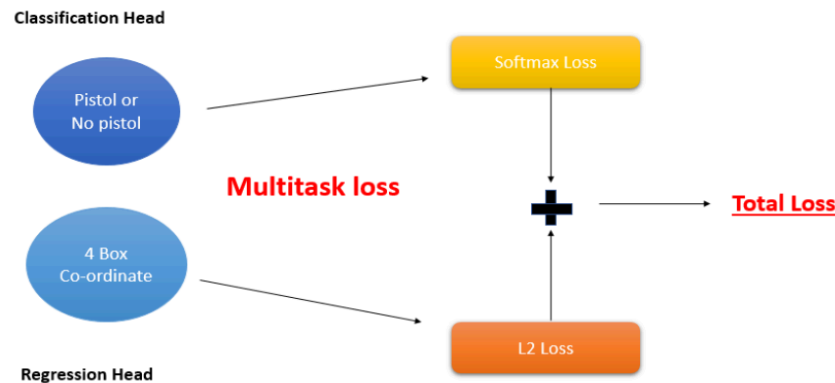
It answers the following question:

- Where is the object of interest present in the Image?
- How many objects are present in the Image?

## Classification + Localization

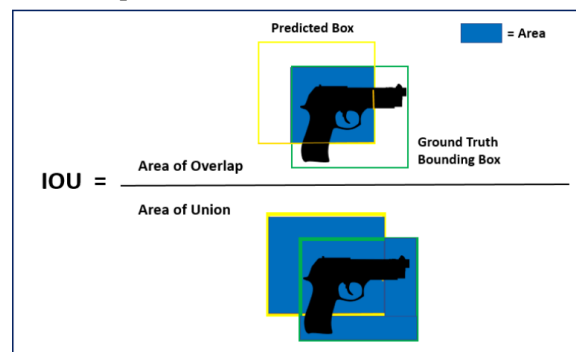


## Model Configurations and Loss function mappings



## Introducing IoU as BBox Evaluation Metric

- IoU stands for Intersection over Union, which is a commonly used evaluation metric for object detection tasks.
- Typically, an IoU threshold is set, such as 0.5, and a prediction is considered a true positive if the IoU score is greater than or equal to the threshold.



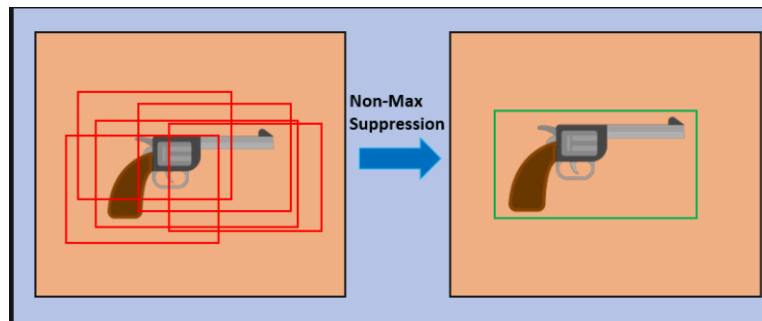
## What is Sliding Window Detection?



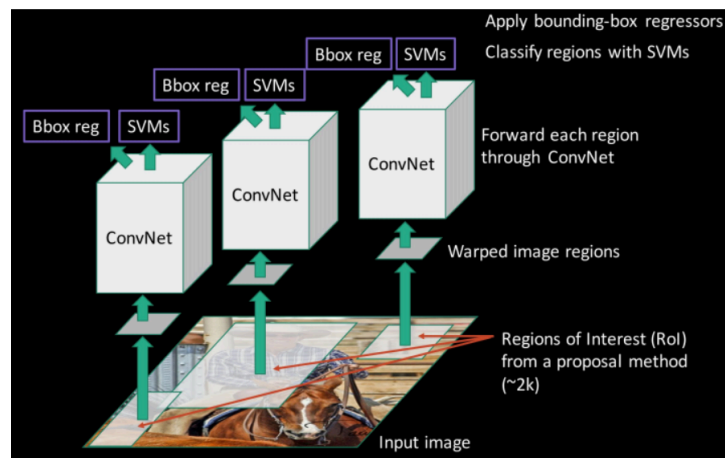
The sliding window algorithm in CNN is a technique where a small window of fixed size is slid over the image to extract features and make predictions on a region of interest.

## Introduce NMS to Suppress BBOXES

- NMS (Non-Maxima Suppression) is a technique used to suppress redundant bounding boxes (BBOXES) in object detection algorithms in computer vision.
- It works by keeping only the bounding box with the highest confidence score and removing other overlapping boxes, reducing the number of false positive detections.



## RCNN

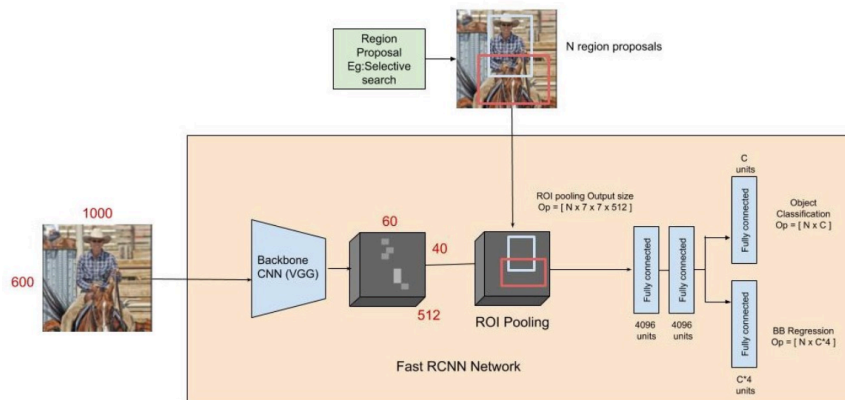


## What is ROI Pooling?

- ROI (Region of Interest) pooling is a technique used in computer vision and deep learning for object detection and classification.
- It works by dividing the input image into regions of interest and then resizing each region to a fixed size, allowing the network to process each region independently and efficiently.
- ROI pooling reduces the spatial dimensionality of the features, making it easier for the network to extract meaningful information from each region and make predictions.
- This technique helps improve the performance of object detection algorithms and makes them computationally more efficient.

## Fast R-CNN

Instead of extracting features from each region proposal independently, Fast RCNN uses a shared convolutional layer to extract features from the entire image, reducing the number of computations needed.



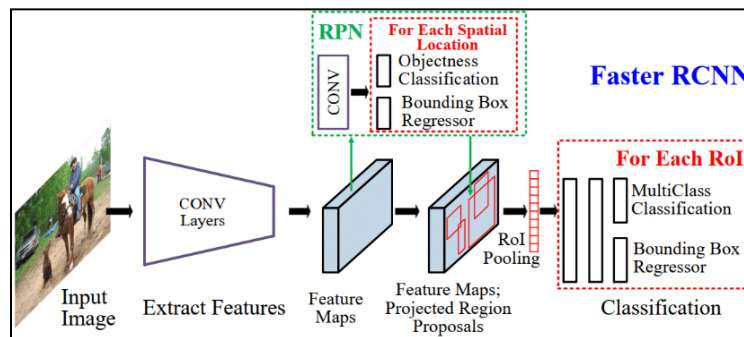
## What is RPN?

- RPN in R-CNN is a component that generates candidate object regions in an image by using a sliding window approach and outputting anchor boxes with objectness scores.
- The RPN helps speed up object detection by reducing the number of region proposals processed by the classifier and by providing high-quality proposals for objects of different shapes

## Faster RCNN

Major components of Faster RCNN includes:

1. **Region Proposal Network (RPN):** Faster R-CNN introduces RPN to efficiently generate region proposals for objects in an image.
  - a. This network proposes candidate object bounding boxes at various locations and scales across the image, allowing for more accurate detection.
2. **Two-stage Detection Framework:** Faster R-CNN follows a two-stage detection framework.
  - a. In the first stage, the RPN generates region proposals, which are then refined and classified in the second stage.
  - b. This two-stage architecture enables Faster R-CNN to achieve better detection accuracy compared to its predecessors while maintaining speed and efficiency.
3. **Anchor Boxes:** Faster R-CNN employs anchor boxes to predict object bounding boxes.
  - a. Anchor boxes are predefined boxes of various sizes and aspect ratios placed at different positions within each sliding window of the feature map.
  - b. These anchor boxes serve as reference templates for predicting object locations and sizes.



Comparison of Speed and mAP metric

External region proposals method (Selective Search Algorithm)		✓	✓	✗
	R-CNN		Fast R-CNN	Faster R-CNN
Test time per image	50 seconds		2 seconds	0.2 seconds
Speed-up	1x		25x	250x
mAP (VOC 2007)	66.0%		66.9%	66.9%

