DS GROUP

Diversifying Portfolio,
Enhanced Value.

## VISION

To be a leading quality & Innovation driven global conglomerate.

## MISSION

We are constantly striving to achieve excellence in all our endeavors to create sustainable value for our stakeholders and the community at large.

# OBJECT DETECTION

An ***Artificial Intelligence (AI) based*** approach
*Advanced Technology Group (DS – ATG)*

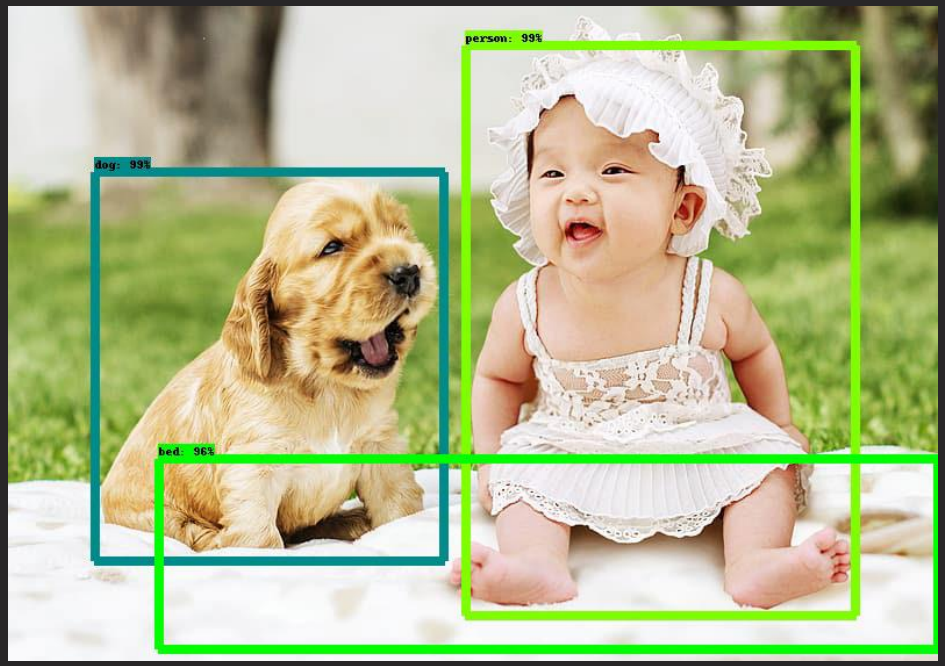
SUBMITTED BY:
SURAT BANERJEE
AI TEAM

# Outline

- Brief of Object Detection
- Performance Metrics
- Object Detection Approaches
- Region Proposal Based Algo.
- Regression-Classification Proposal Based Algo.

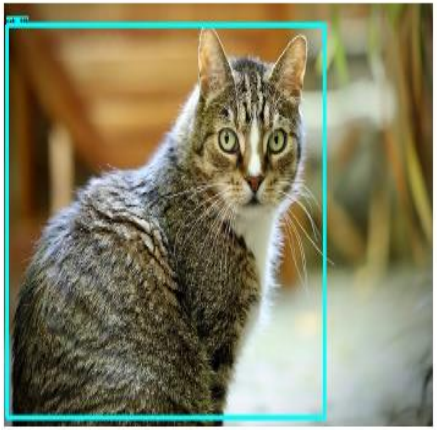
# Brief of Object Detection

# Object Detection




Given an image we want to detect all the object in the image that belong to a specific classes and give their location. An image can contain more than one object with different classes.

# Computer Vision Task

| Classification | Classification + localization | Object detection | Semantic segmentation |
| --- | --- | --- | --- |
|  |  |  |  |
| Cat | Cat | Dog, Person, Bed | Dog, Person |

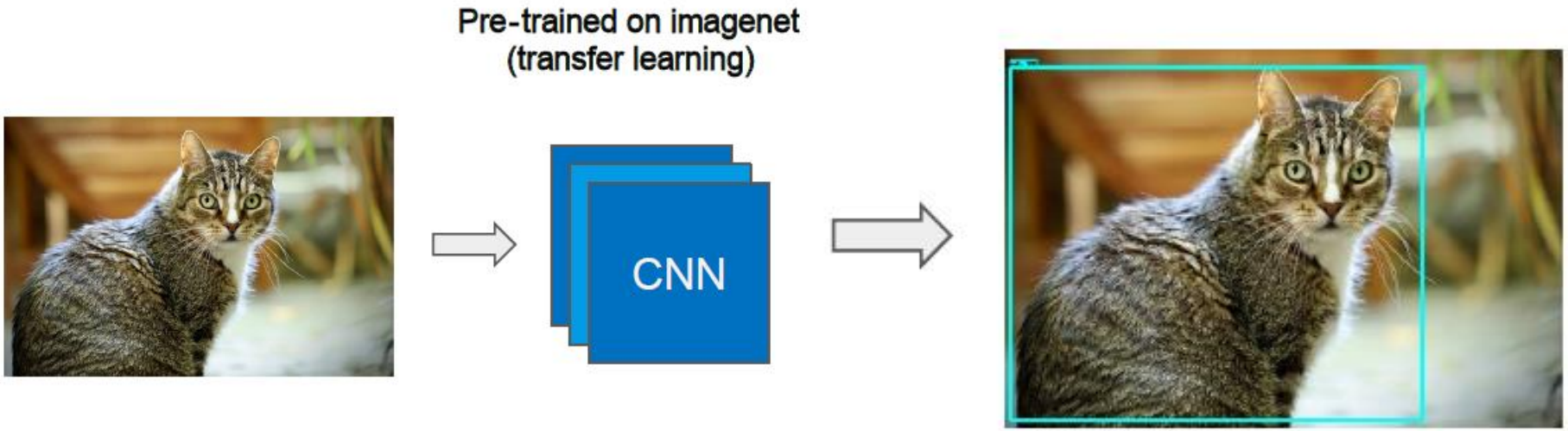
**Single object** **Multiple objects**

# Image Classification

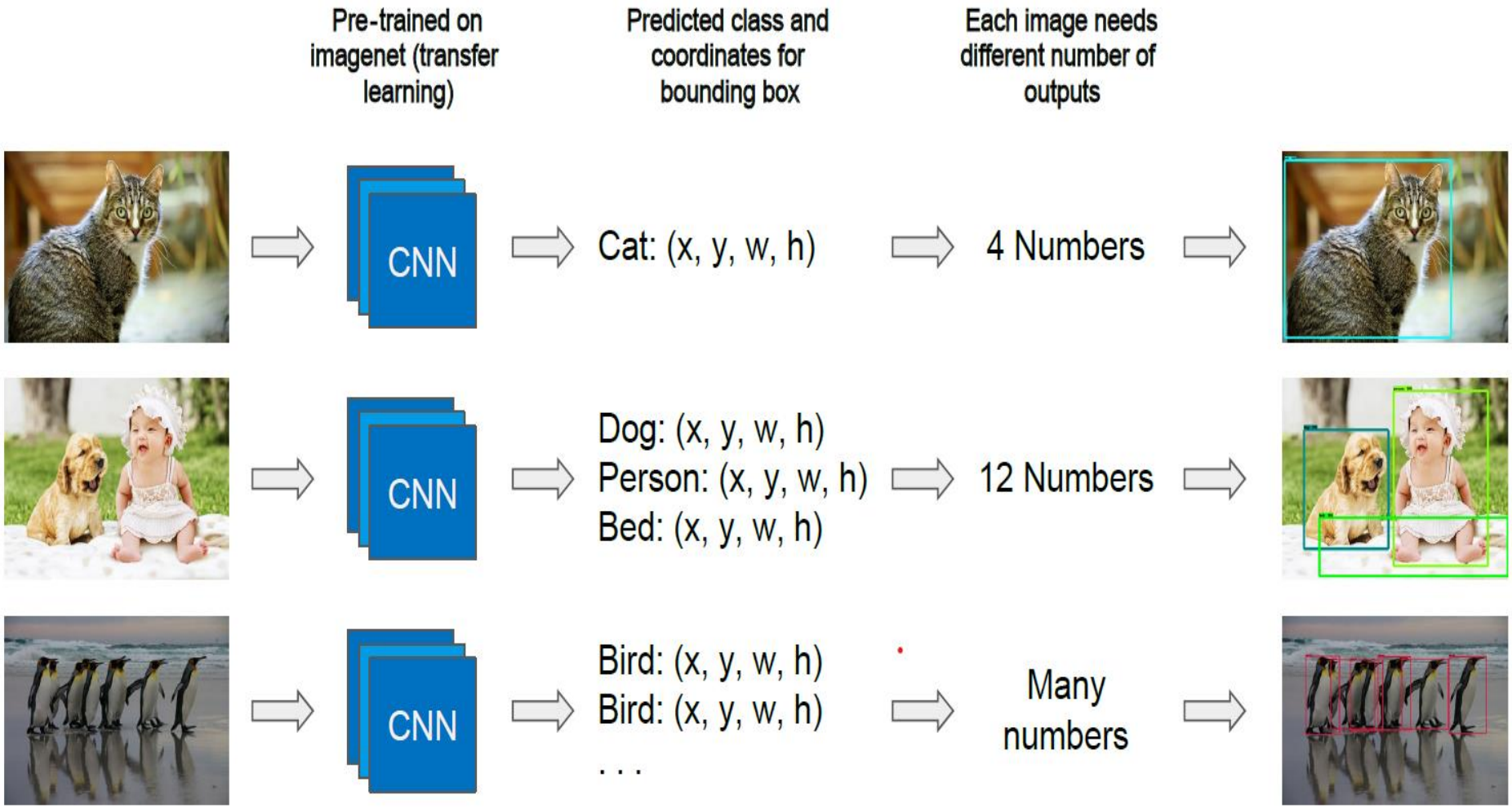
- Output a class label. The whole image represents one class. We don't want to know exactly where are the object. Usually only one object in the image.

# Classification + localization

- Given an image we want to learn the class of the image and where are the class location in the image. Usually only one object in the image.



# Object Detection



# SEGMENTATION

**Semantic Segmentation**

•**Goal:** Divide an image into different regions, where each region corresponds to a specific object class.

**Instance Segmentation**

•**Goal:** Identify and segment individual instances of objects within an image.

**Panoptic Segmentation**

•**Goal:** Combine both semantic and instance segmentation to provide a complete understanding of the scene.

•**Object Representation:**

- Semantic Segmentation: Regions based on object class.
- Instance Segmentation: Individual instances based on object class.
- Panoptic Segmentation: Regions and instances based on object class.

# Data: object detection research


330K images (>200K labeled)
and 80 object categories

IMAGENET

14M images and more than
20K object categories


11,530 images and 20
classes

Google Open Images
Dataset V5

~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, and visual relationships.

# Challenges in Object Detection

- Two tasks -Classification and Localization
- Results/prediction take lot of time but we need fast predictions for real-time task
- Different scales and aspect ratios
- Limited data and labelled data
- Imbalanced data-classes

# Performance Metrics

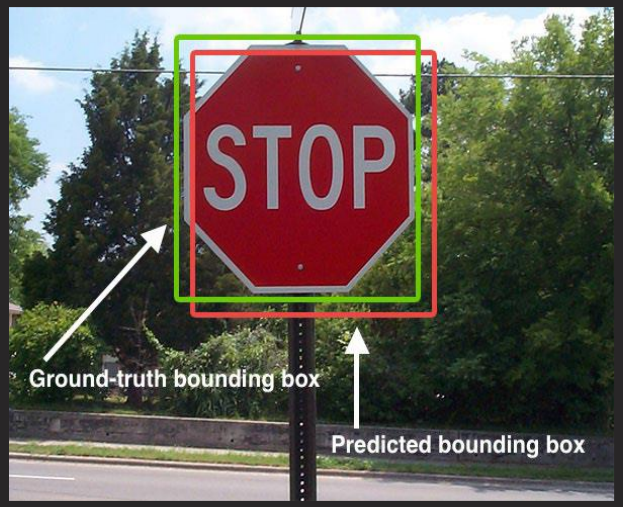
# Performance metrics for object detection

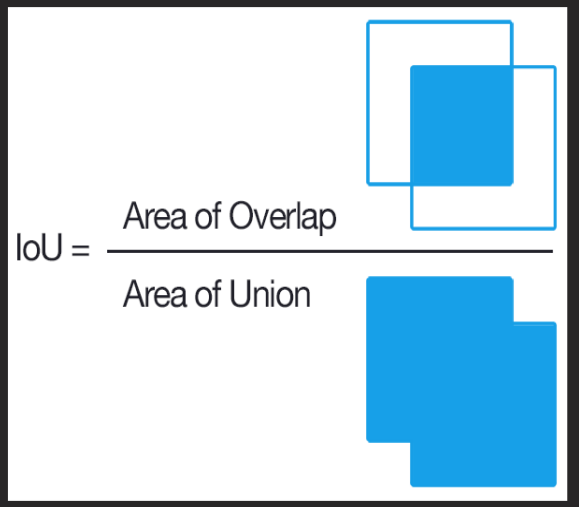
- Intersection over union (IoU)
- Precision and recall
- Mean average precision (mAP)

# IoU (Intersection over Union)

- IoU is a function used to evaluate the object detection algorithm.
- It computes size of intersection and divide it by the union. More generally, IoU is a measure of the overlap between two bounding boxes.
- For example in the diagram here : The red is the predicted output and green is the labeled output.
- To compute Intersection Over Union we first compute the union area of the two rectangles which is "the first rectangle + second rectangle" Then compute the intersection area between these two rectangles.

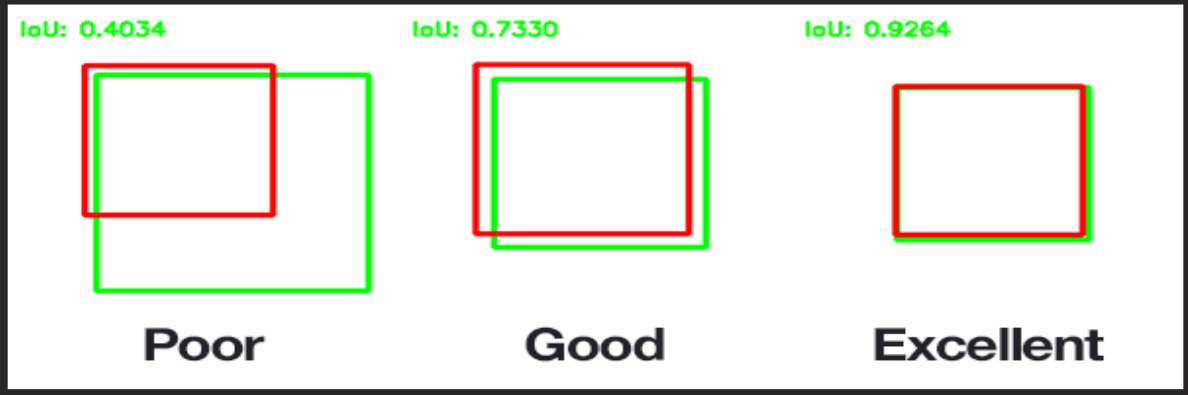
- Finally IOU = intersection area / Union area
- If IoU >= 0.5 (threshold) then it's good. The best answer will be 1



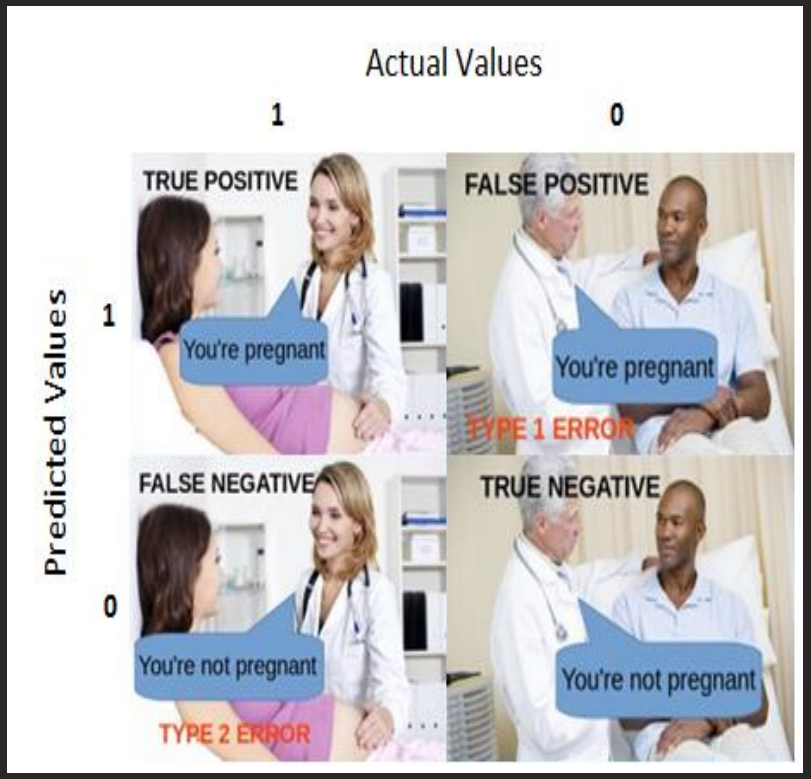



# Why to use IoU score?

- We use Accuracy, precision, recall etc as accuracy metrics for classification tasks but for object detection it's not so straightforward.
- It's very unlikely that the (x, y)-coordinates of our predicted bounding box are going to exactly match the (x, y)-coordinates of the ground-truth bounding box.
- More the overlap predicted bounding boxes have with the ground-truth bounding boxes better (higher) their IoU scores will be.

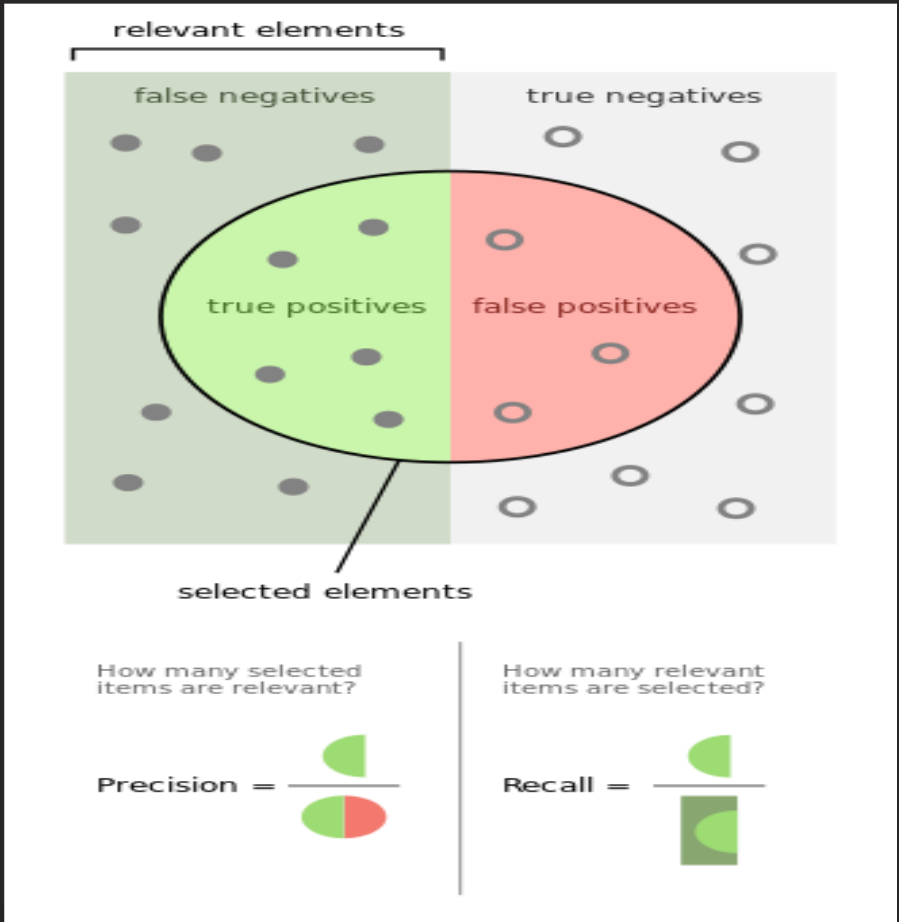


# Confusion Matrix




- **True Positives (TP):** These are cases in which we predicted yes and they do have the pregnency.
- **True Negatives (TN):** We predicted no, and they don't have the pregnency.
- **False Positives (FP):** We predicted yes, but they don't actually have the pregnency.
- **False Negatives (FN):** We predicted no, but they actually do have the pregnency.

# Precision and Recall




- **Precision:** what percentage of your positive predictions are correct

  (TP/TP+FP)

- **Recall:** what percentage of ground truth objects were found

  (TP/TP+FN)

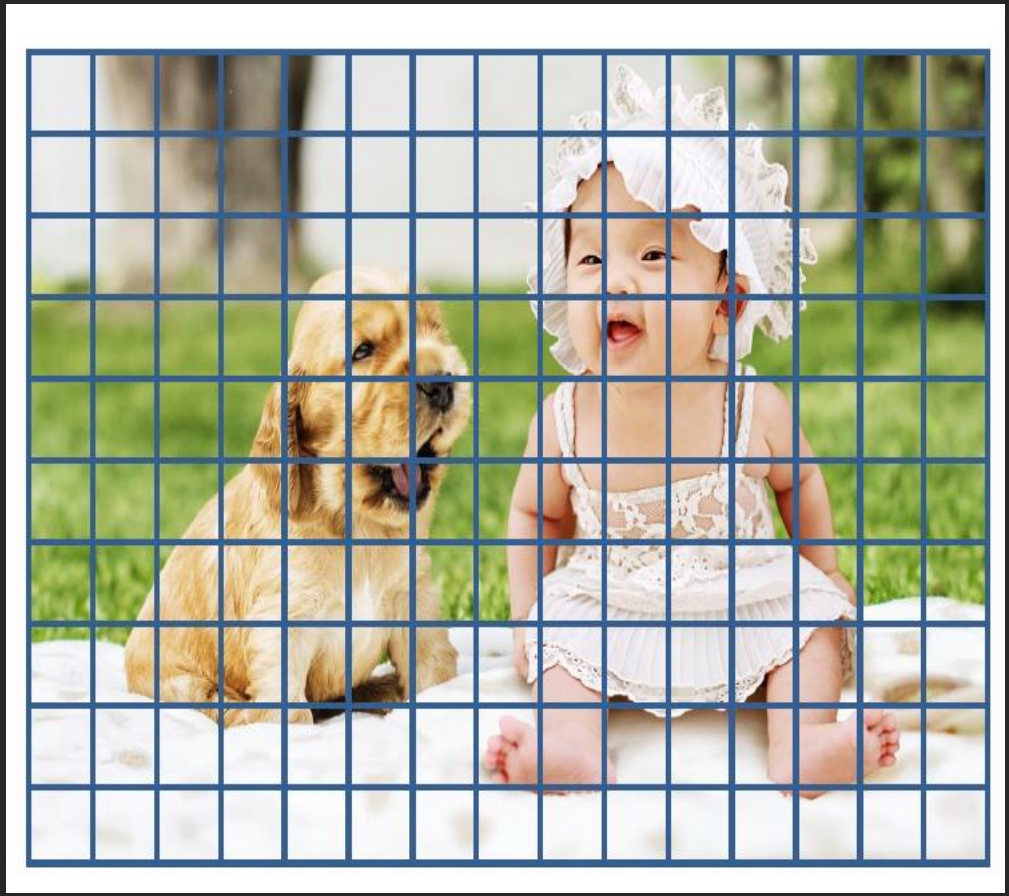
# mAP (Mean Average Precision)

- **Step 1:**Sort predictions according to confidence
- **Step 2:**Calculate IoU of every predicted box with every ground truth box
- **Step 3:**Match predictions to ground truth using IoU, correct predictions are those with IoU > threshold (.5)
- **Step 4:**Calculate precision and recall at every row
- **Step 5:**Take the mean of maximum precision at 11 recall values (0.0, 0.1, … 1.0) to get Average Precision
- **Step 6:**Average across all classes to get the mAP

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i$$

# Object Detection Approaches

# Object detection: 1. Brute force approach



- Run a classifier for every possible box
- This is a 15 x 10 grid, there are 150 small boxes.

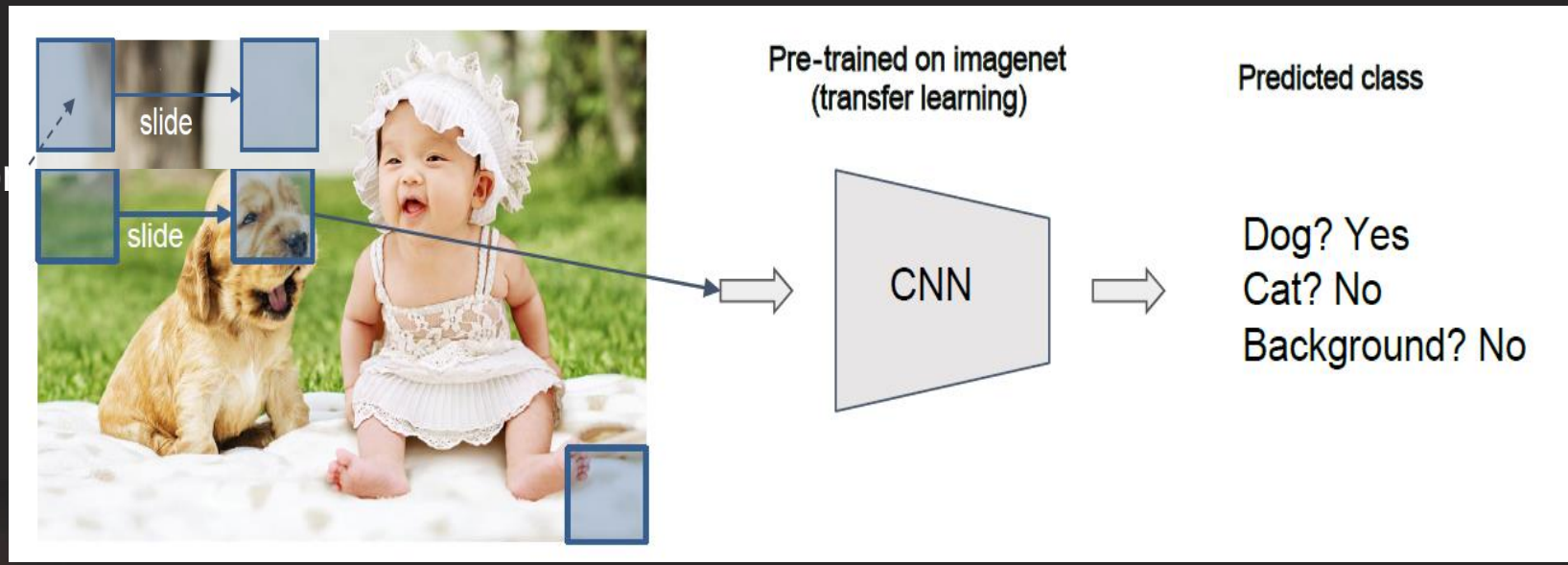
How many total boxes?

- Computationally expensive

# Object detection: 2. Sliding window approach

- Run classifier in a sliding window fashion
- Apply a CNN to many different crops of the image
- CNN classifies each crop as object or background

Detection window

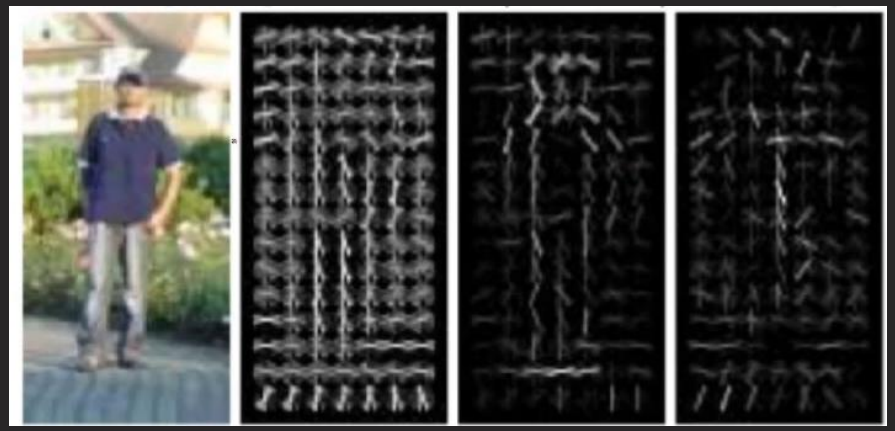


# Traditional Object Detection

Three stages:

- Informative Region selection,
- Feature Extraction
- Classification.

Ex: Haar Features & Histogram of Oriented Gradients



# Object Detection State

State of the Art methods are generally categorised in two categories

- One stage methods / Regression-Classification Based Frameworks
- Three stage methods / Region Proposal Based Frameworks

**Region Proposal based frameworks**

It is composed of 3 correlated stages, including region proposal generation, feature extraction with CNN, classification and bounding box regression, which are usually trained separately.

**Regression/Classification Based Framework**

One-step frameworks based on global regression/classification, mapping straightly from image pixels to bounding box coordinates and class probabilities, can reduce time expense.

two significant frameworks,

You only look once (YOLO) and Single Shot MultiBox Detector (SSD)

# Object Detection State - Algorithms

Region Proposal Based Algorithms -

- R-CNN
- Fast R-CNN
- Faster R-CNN

Regression-Classification Based Algorithm -

- YOLO
- SSD

# Region Proposal Based Algo

# Regression-Classification Proposal Based Algo.


Thank
you !!!
...