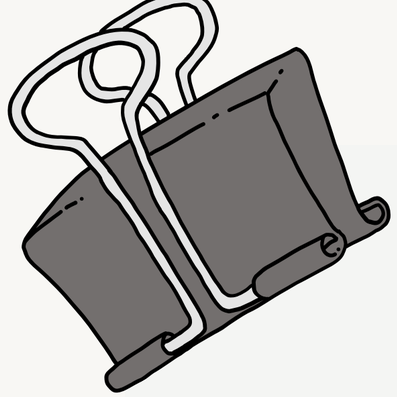


Object Detection with R-CNN, Fast R-CNN, and Faster R-CNN

Teams: Neural Nets
Team Members:YiJie Cao, Jiaqi Liu, Hao Niu, Yue Zhang

Overview

Introduction

Technical Details

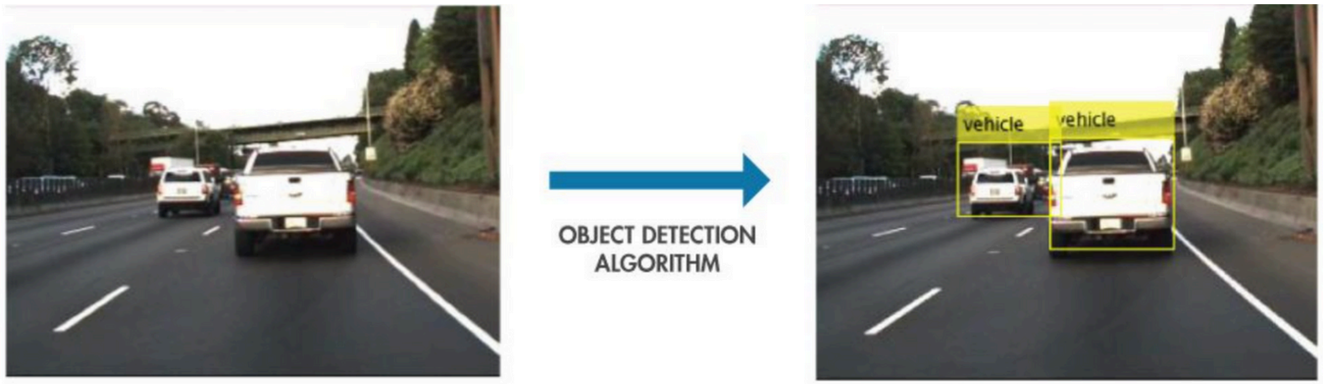
Applications

Strengths and Limitations

Introduction

What is Object Detection?

- Object detection is a computer vision technique for locating instances of objects in images or videos. Object detection algorithms typically leverage machine learning or deep learning to produce meaningful results. When humans look at images or videos, we can recognize and locate objects of interest within a matter of moments. The goal of object detection is to replicate this intelligence using a computer.



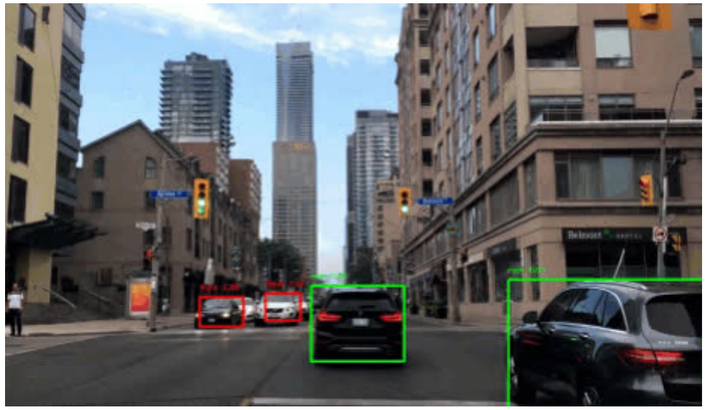
Using object detection to identify and locate vehicles.

[1]"What Is Object Detection?," Mathworks.com, 2019. https://www.mathworks.com/discovery/object-detection.html

Evolution of Object Detection Techniques

The traditional methods before the advent of deep learning:

HOG(Histogram of Oriented Gradients) is a feature extraction method based on the histogram of image gradient directions, commonly used to detect edges and shapes in images. It has been widely applied in object detection, such as pedestrian detection.



SVM (Support Vector Machine):
Brief Introduction: SVM is a classification algorithm commonly used for binary classification tasks in object detection. It is often combined with feature extraction methods like HOG.

Evolution of Object Detection Techniques

With the development of deep learning:

In traditional methods, feature extraction and classification are done separately. Deep learning changed this by using Convolutional Neural Networks (CNNs) to do both at the same time, which makes detection much more accurate.

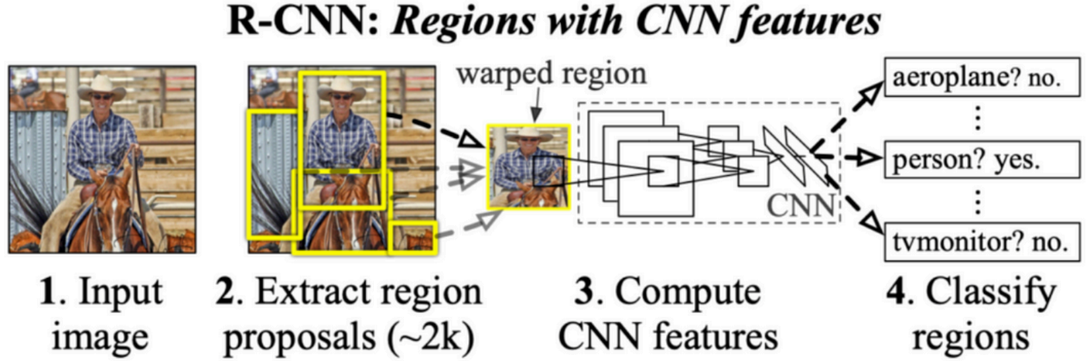
R-CNN – The first object detection method that combined deep learning.

Fast R-CNN – Improved the speed issues of R-CNN.

They created a more useful way to suggest areas in the image to focus on, which made the detection process much faster.

What is R-CNN?

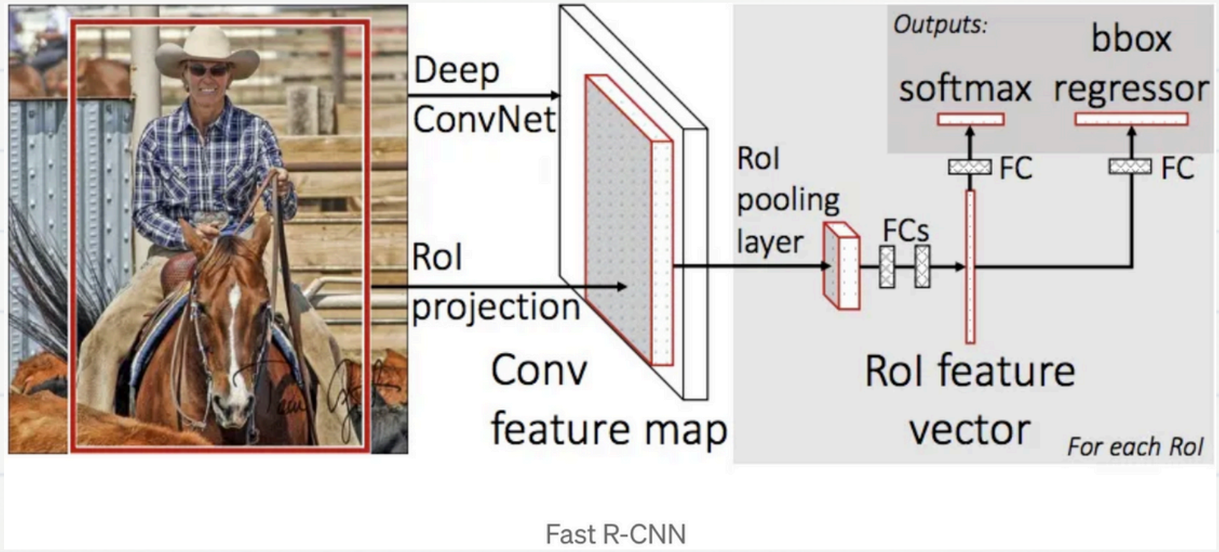
R-CNN (Region-based Convolutional Neural Network) was an epoch-making model in 2013 which successfully combined CNN with classical computer vision techniques for object detection and broke the previous record[2].



1. Input Image → 2. The model looks at the image and selects about 2000 possible areas that might contain objects. → 3. Each of these areas is resized and sent into a neural network (CNN) to find important features. → 4. The model uses a classifier to check each area and decide what object is in it, like "person," or "TV."

[2]"KiKaBeN - R-CNN: Region-based Convolutional Neural Network," KiKaBeN, Jun. 11, 2022. https://kikaben.com/r-cnn-original/

Improving Efficiency: Fast R-CNN



Fast R-CNN

Input image to CNN:

The entire image is input to the convolutional neural network (CNN) to create a feature map.

Find region proposals:

Identify possible object regions from the feature map.

Resize regions:

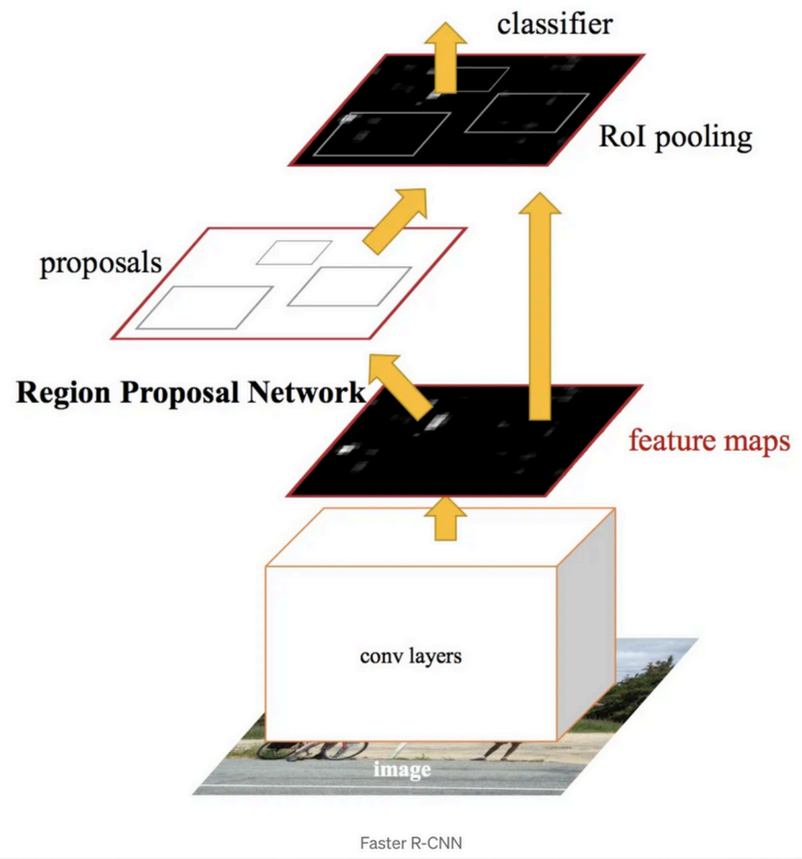
Use the RoI pooling layer to resize all regions to the same size.

Classify and locate:

Use a classifier to determine the object class in each region and a regressor to calculate the exact location of the bounding box.

[3]R. Gandhi, “R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms,” Towards Data Science, Jul. 09, 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

Improving Efficiency: Faster R-CNN



Faster R-CNN

Input image:

Both Fast R-CNN and Faster R-CNN first input the image into a convolutional network to create feature maps.

Region proposal network (RPN) in Faster R-CNN:

Faster R-CNN uses a region proposal network (RPN) to automatically create region proposals, making it faster without the need for the slow selective search used in Fast R-CNN.

RoI pooling for resizing:

Both Fast R-CNN and Faster R-CNN use RoI pooling to resize each region to the same size.

Classify and locate:

Both models use a classifier and regressor to predict the object class and the bounding box location in the final step.

[4]"KiKaBeN - Faster R-CNN: Real-Time Object Detection with RPN," KiKaBeN, Jul. 11, 2022. https://kikaben.com/faster-r-cnn-rpn/ (accessed Oct. 18, 2024).

Why is R-CNN Important?

First to use deep learning for object detection:

R-CNN was the first model to use deep learning (CNN) for object detection, improving accuracy.

Combining region proposals with CNN:

R-CNN creates region proposals and uses Convolutional Neural Networks (CNN) for feature extraction, allowing for precise object detection and recognition.

Led to faster models:

Although R-CNN is slow, it laid the foundation for faster and better models like Fast R-CNN and Faster R-CNN. It also opened up more possibilities for real-world applications, like self-driving cars and security monitoring.

Technical Details

“High-level” Comparison

Traditional CNN
e.g., ResNet

Input Image

CNN Layers

Classification

RCNN

Input Image

Region Proposals
(Selective Search)

CNN for each Region

SVM

Bounding Box

Fast RCNN

Input Image

CNN Feature map

Region Proposals
(Selective Search)

ROI poling

FC Layers

Classification

Bounding Box

Faster RCNN

Input Image

CNN Feature map

Region Proposal Network

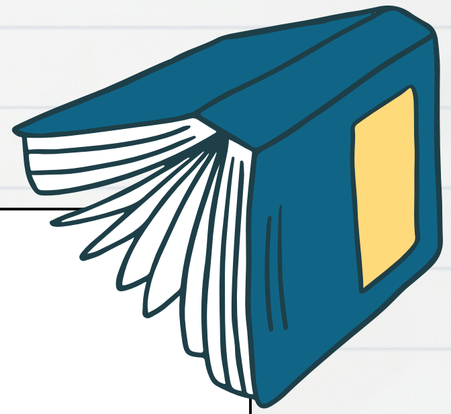
ROI poling

FC Layers

Classification

Bounding Box

RCNN Pipeline



Pre-CNN

Input Image

↓

Selective Search

↓

2000 Region Proposals

CNN Feature Extraction

Warp each region to fixed size

↓

CNN (AlexNet etc.)

↓

4096-dim Feature Vector

Post-CNN

Class SVMs → Class scores

Bounding Box Regressor → Adjust Coordinates

↓

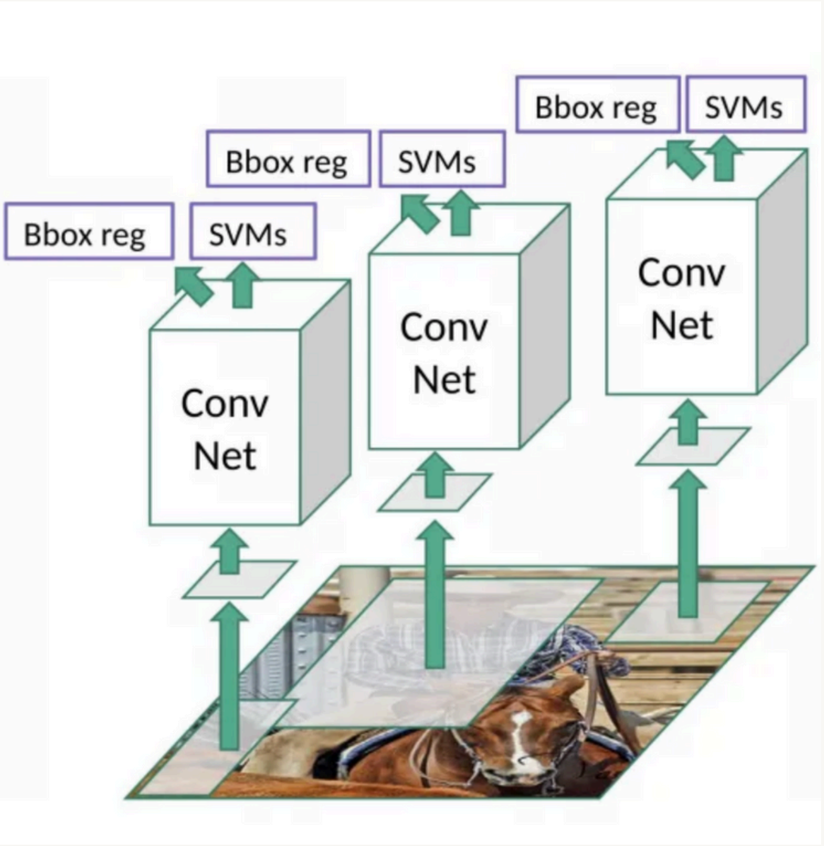
Non-max Suppression

↓

Final Detection

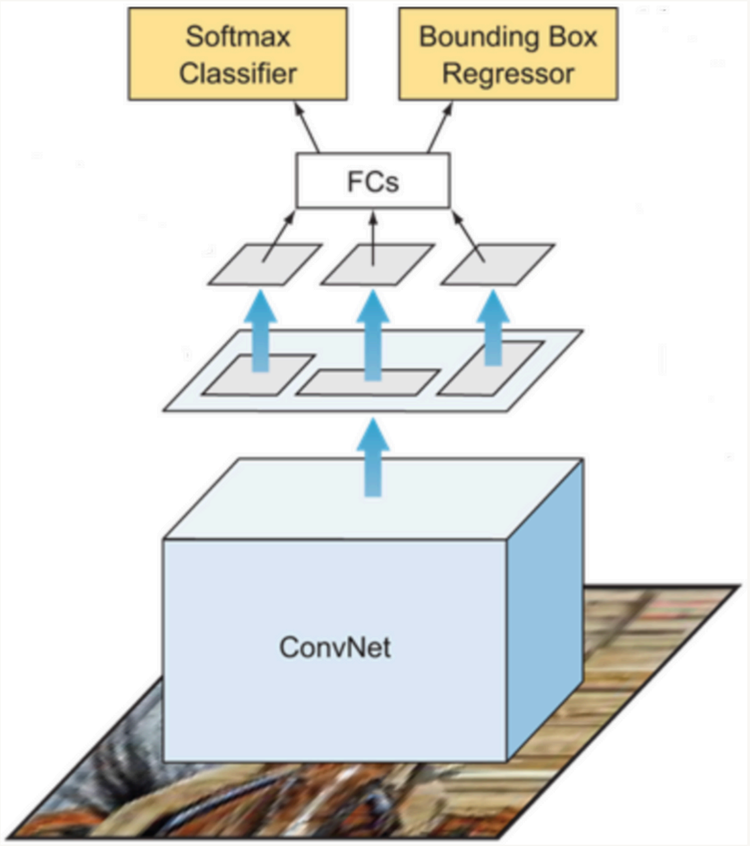
RCNN – Fast RCNN – Faster RCNN

RCNN



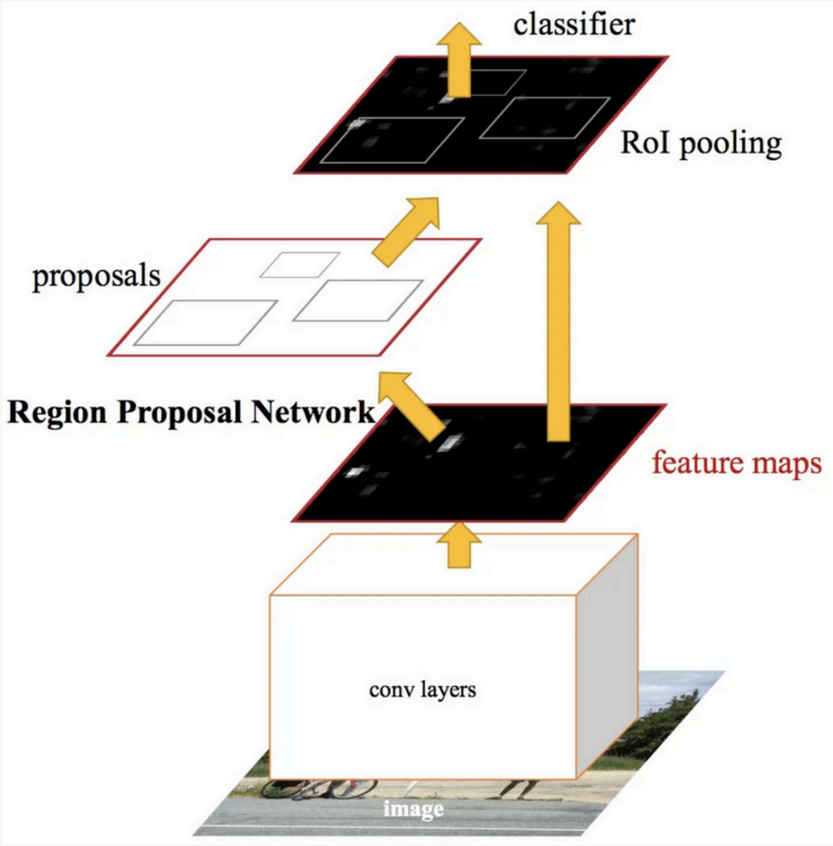
50 sec / img

Fast RCNN



2 sec / img

Faster RCNN (actually 2 NNs)



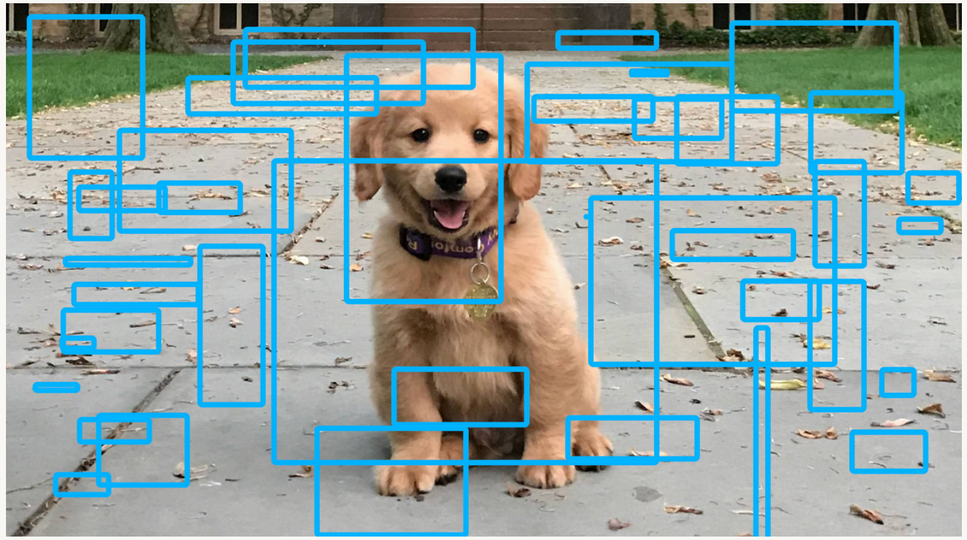
0.2 sec / img

Very similar MAP (Mean Average Precision) on PASCAL VOC 2007 - 66.% ~ 66.9%

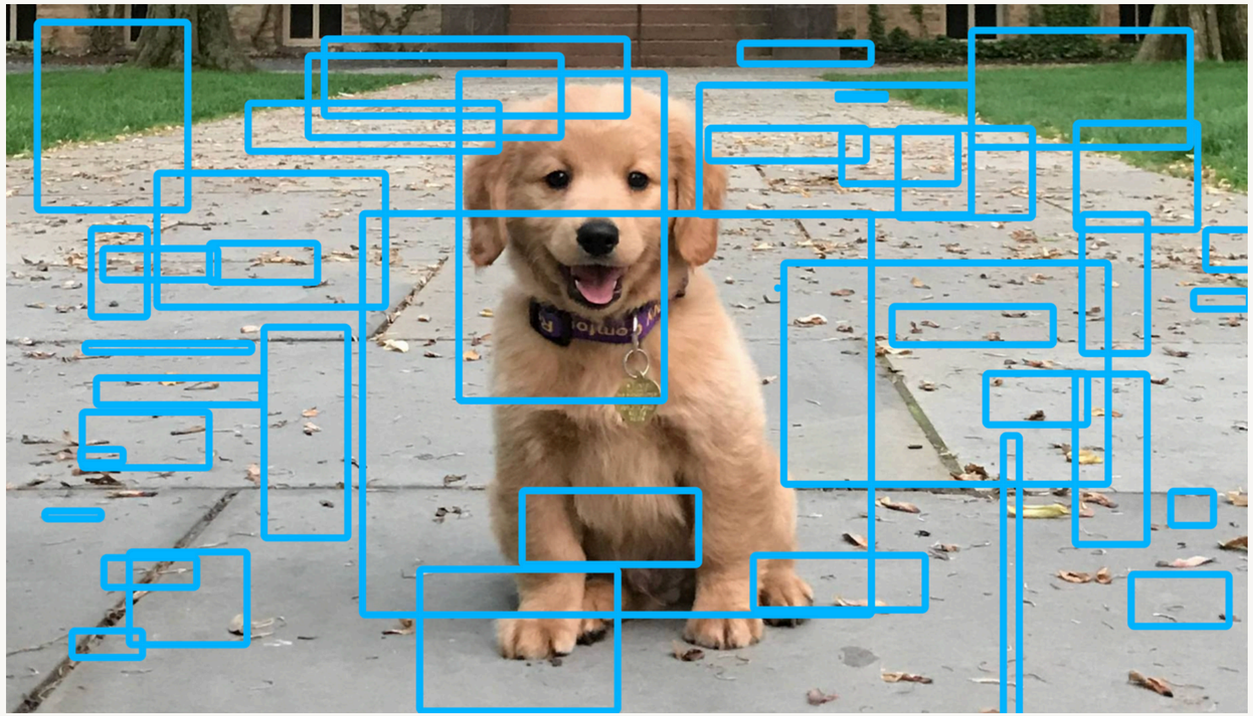
In Human Language

Puppy Ready

Draw 2000 boxes

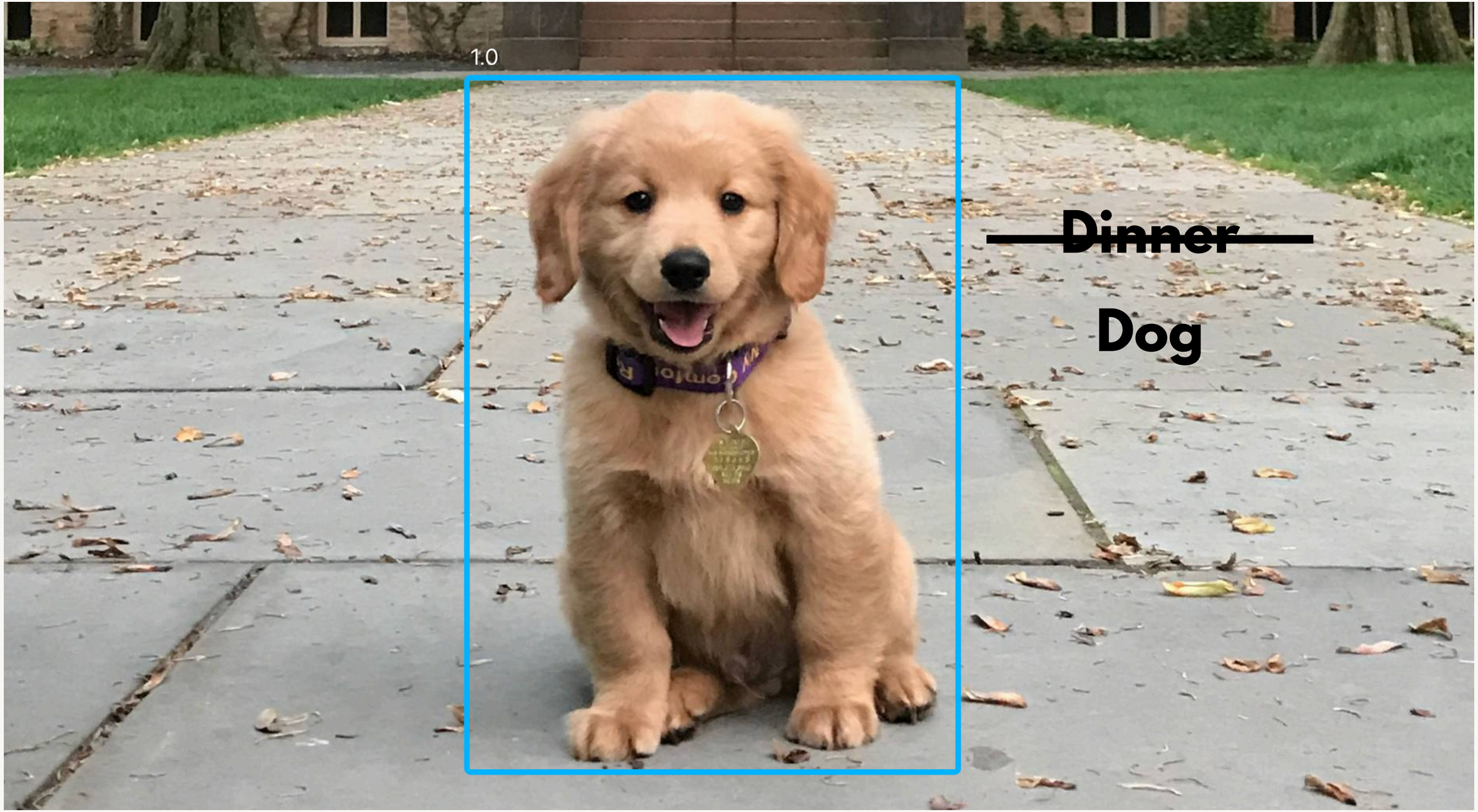


Be smart, “guess” what is in the box



After a series of very sophisticated manipulations & calculations

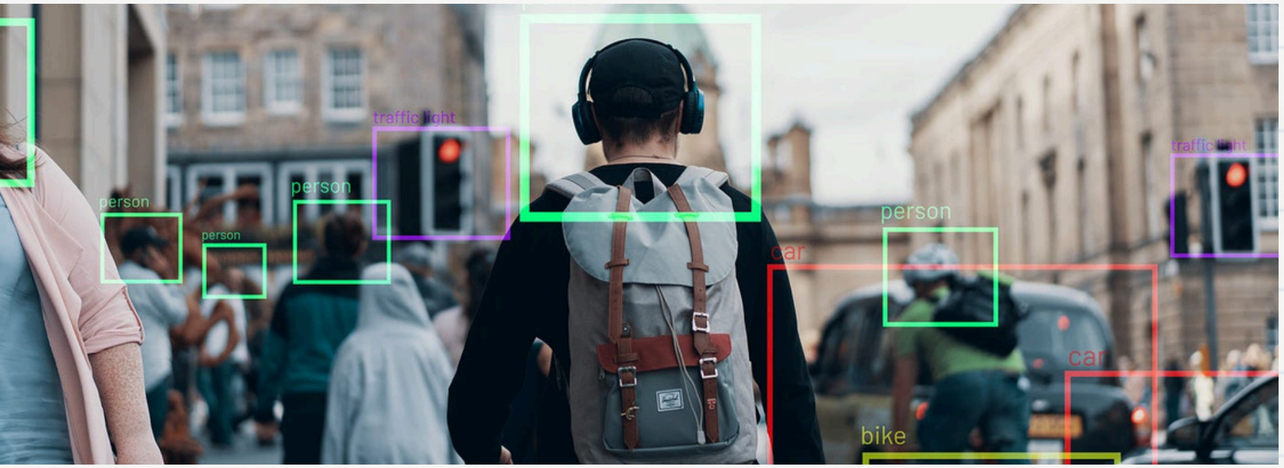
Prediction



Applications

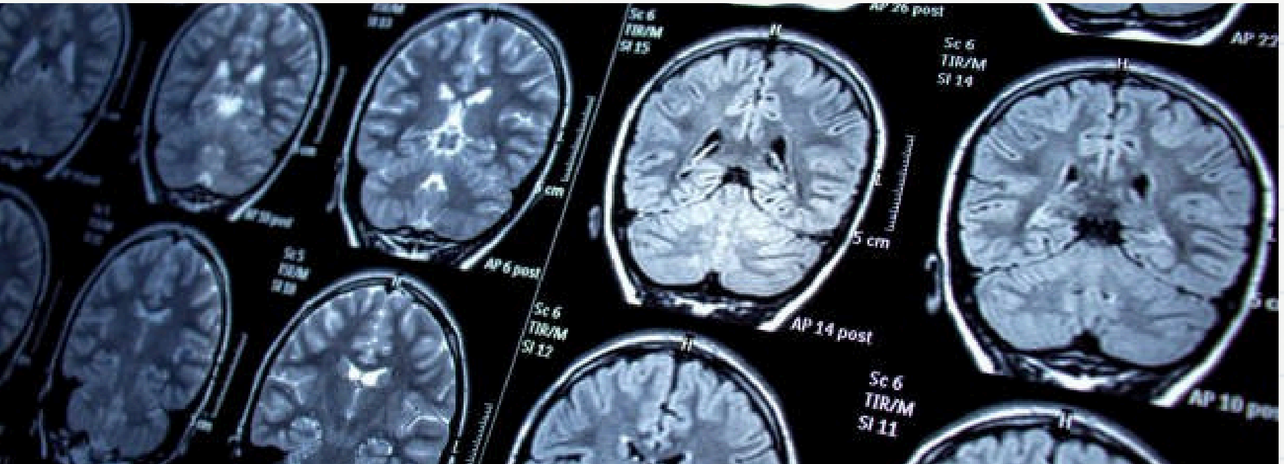
1. Autonomous Vehicles and Driver Assistance Systems

- Detects vehicles, pedestrians, traffic signs, and lane markers
- Supports real-time decisions in self-driving cars
- Used in collision warnings and driver assistance features



2. Surveillance and Security Systems

- Monitors public and restricted areas
- Detects intruders, weapons, and suspicious activities
- Provides real-time alerts for faster threat responses



3. Medical Imaging

- Identifies tumors, lesions, and abnormalities in X-rays, MRIs, and CT scans
- Assists in early diagnosis and medical assessments
- Improves patient outcomes through accurate detection

4. Retail and Inventory Management

- Automates detection of empty shelves and misplaced products
- Ensures optimal stock levels
- Enhances workflow efficiency and customer satisfaction

5. Agriculture and Precision Farming

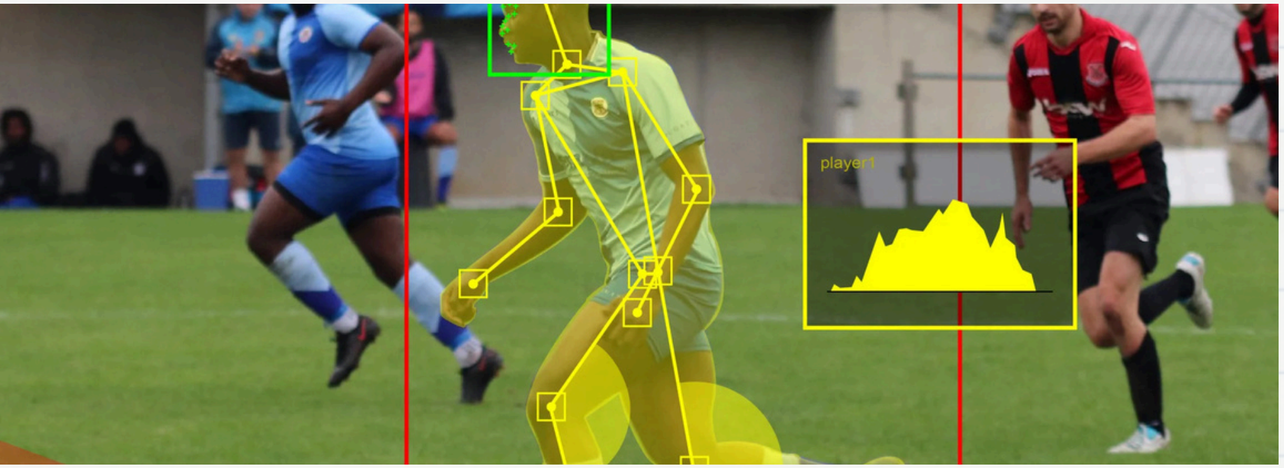
- Analyzes crop health through aerial images
- Detects weeds, pests, and diseases
- Supports sustainable farming with timely interventions

6. Industrial Automation and Robotics

- Detects components for assembly lines
- Ensures accurate assembly and quality control
- Improves production efficiency and reduces errors

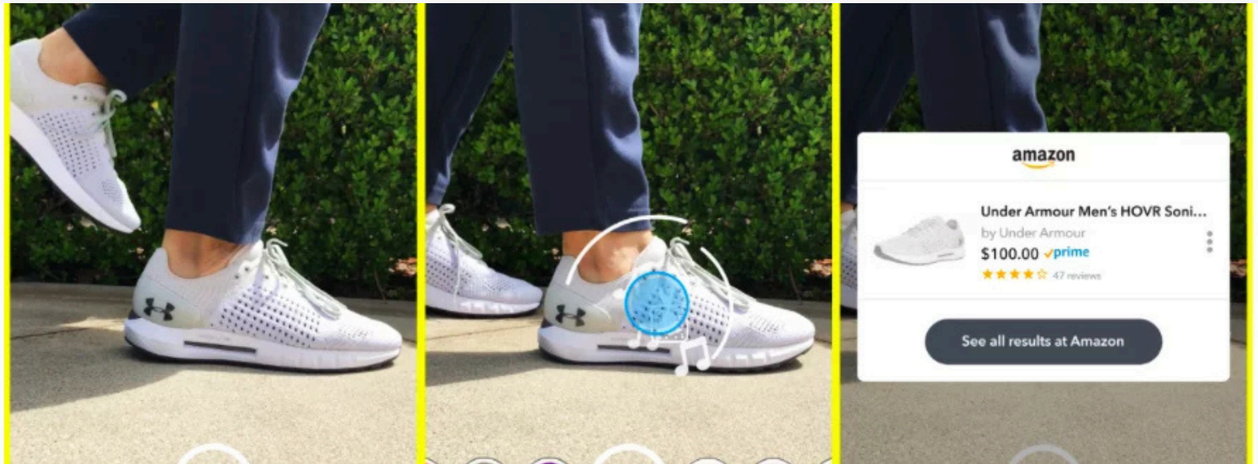
7. Wildlife Monitoring and Conservation

- Tracks animal populations and endangered species
- Detects illegal activities like poaching
- Assists conservationists in biodiversity protection



8. Sports Analytics

- Tracks players and objects (like balls) in real-time
- Provides insights for coaches and analysts
- Powers broadcast features with live statistics



9. E-commerce and Visual Search Engines

- Enables “search by image” functionality
- Matches products from user-uploaded photos
- Enhances the shopping experience and user engagement

Strengths and Limitations

R-CNN

Strengths

-First to combine deep learning (CNN) with object detection.
-High detection accuracy.

Limitations

-Very slow due to processing each region separately.
-Complex and memory-intensive.

Fast R-CNN

Strengths

-Of Course, Fast! Faster by using shared CNN features.
-Simpler end-to-end training.

Limitations

-Relies on selective search, slowing down the process..
-Region proposal generation is not optimized.

Faster R-CNN

Strengths

-Introduces RPN for faster region proposals.
-End-to-end optimization for speed and accuracy.

Limitations

-More complex architecture.
-Still not as fast as real-time models

References

[1]“What Is Object Detection?,” Mathworks.com, 2019. https://www.mathworks.com/discovery/object-detection.html
[2]“KiKaBeN - R-CNN: Region-based Convolutional Neural Network,” KiKaBeN, Jun. 11, 2022. https://kikaben.com/r-cnn-original/
[3]R. Gandhi, “R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms,” Towards Data Science, Jul. 09, 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e
[4]“KiKaBeN - Faster R-CNN: Real-Time Object Detection with RPN,” KiKaBeN, Jul. 11, 2022. https://kikaben.com/faster-r-cnn-rpn/ (accessed Oct. 18, 2024).
[5]OpenAI, “ChatGPT,” chat.openai.com, 2024. https://chat.openai.com/

Potrimba, P. (2024, April 9). What is R-CNN?. Roboflow Blog. https://blog.roboflow.com/what-is-r-cnn/
Faster RCNN-CSDNblog. (n.d.-a). https://blog.csdn.net/weixin_45564943/article/details/121877359
What are the limitations of fast R-CNN? - scispace. (n.d.-b). https://typeset.io/questions/what-are-the-limitations-of-fast-r-cnn-1pbihsx2db

Thanks!

Q & A