

# Computer Vision: Object Detection using Faster R-CNN, Mask R-CNN, and YOLOv8

Hector R. Gavilanes, Project Lead

Jacob Knight, Deep Learning Architect

Kyle Knuth, Data Scientist

Sai Devarashetty, Software Engineer



# Agenda

Introduction

Architecture

Overview

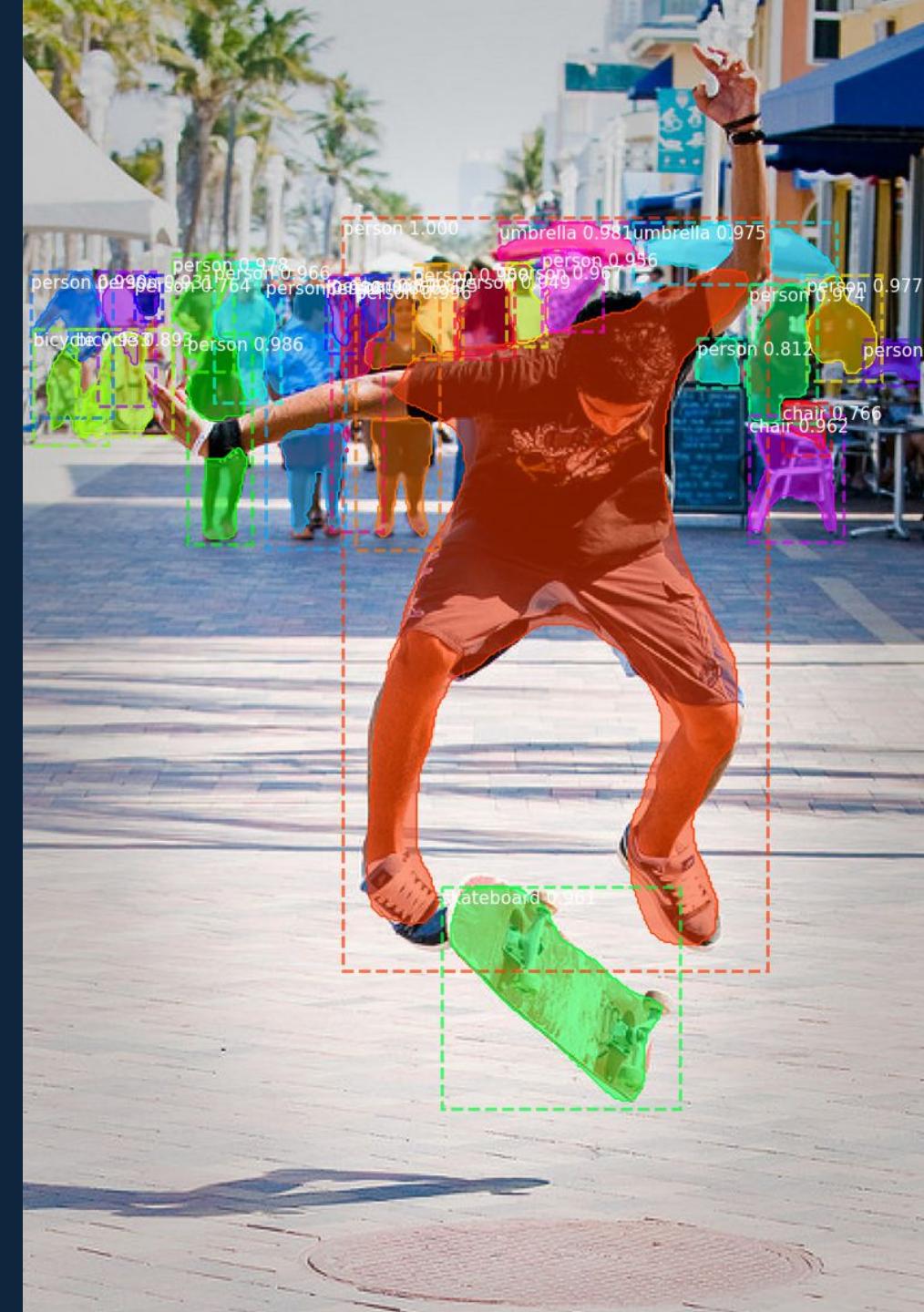
Analysis

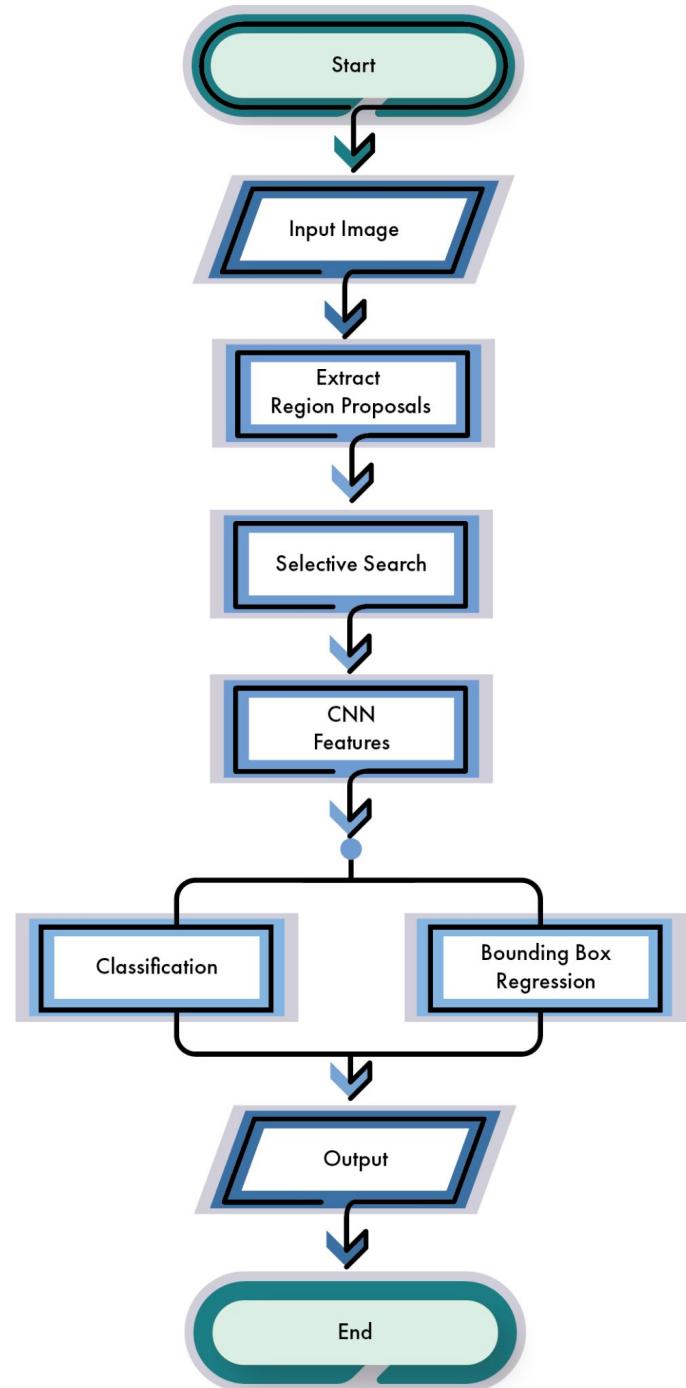
Results

# Introduction

Our study compares  
Faster R-CNN, Mask R-CNN, and YOLOv8  
object detection and segmentation evaluating  
accuracy and speed.

Our findings help choose the best method based on  
performance and practical deployment.





# R-CNN: Architecture

Multiple Region Proposal

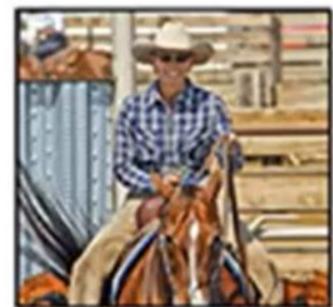
Feature Extraction

Selective Search

Classification

Bounding Box regression

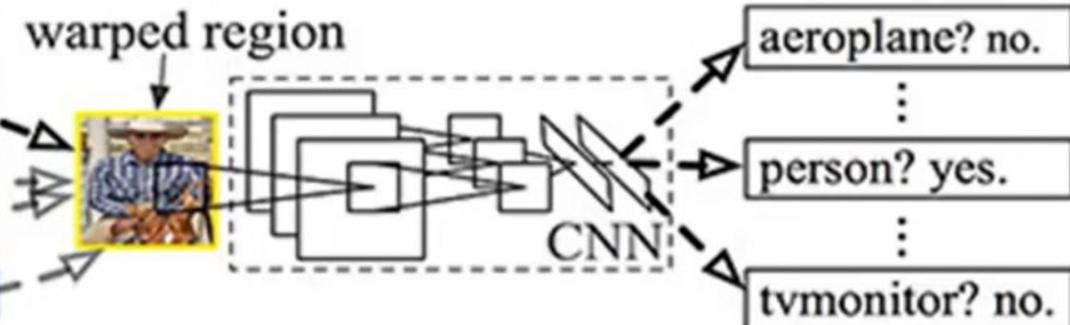
## R-CNN: *Regions with CNN features*



1. Input image

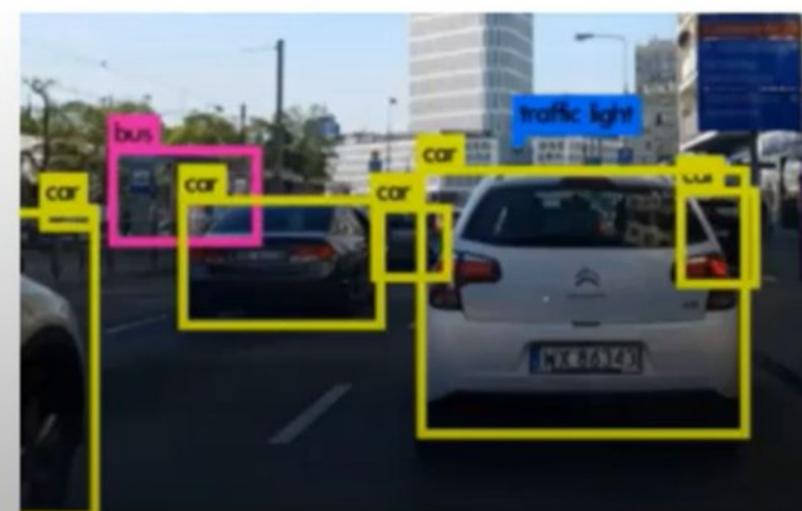
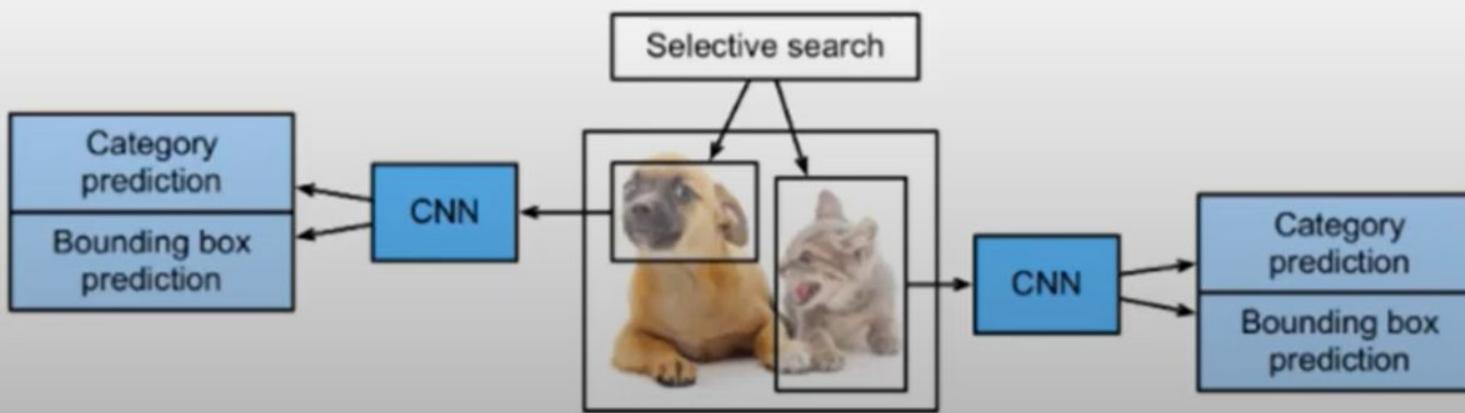


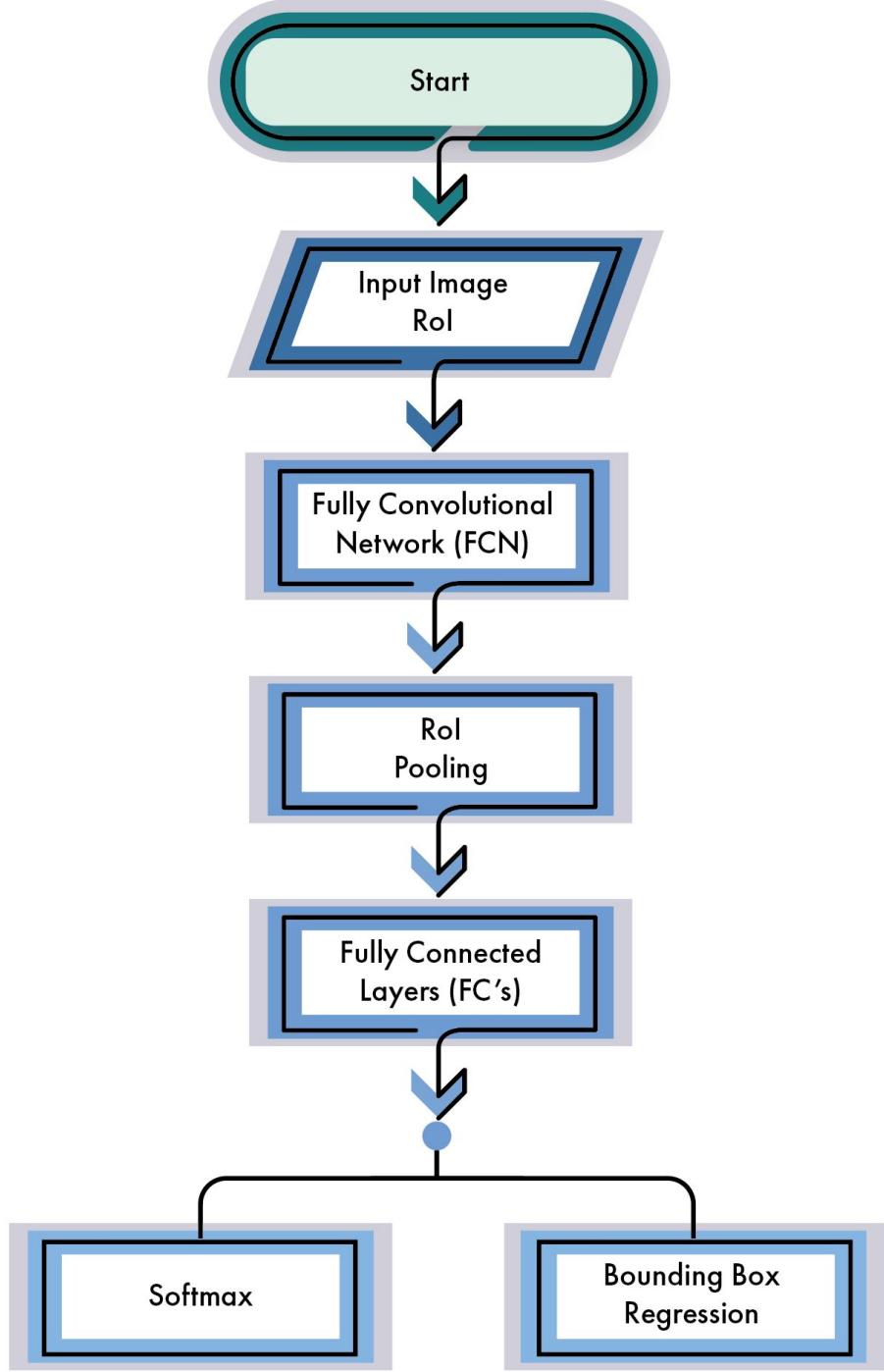
2. Extract region proposals (~2k)



3. Compute CNN features

4. Classify regions





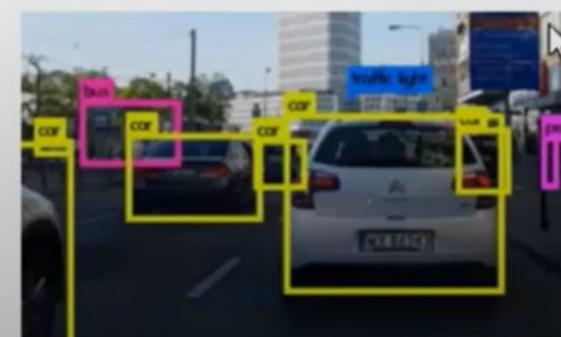
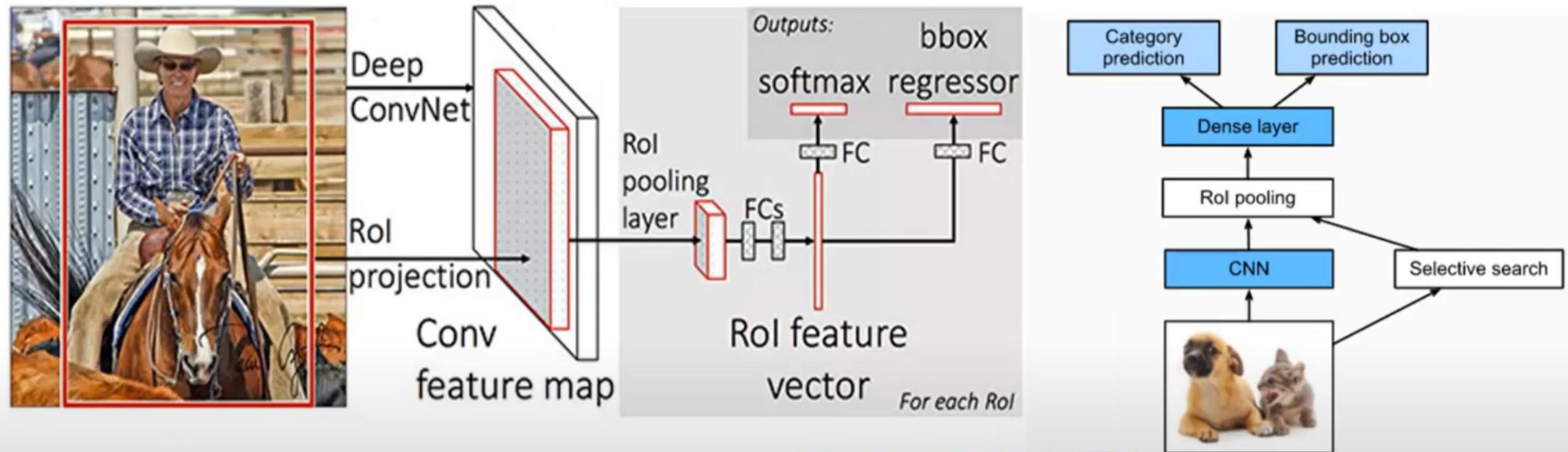
# Fast R-CNN: Architecture

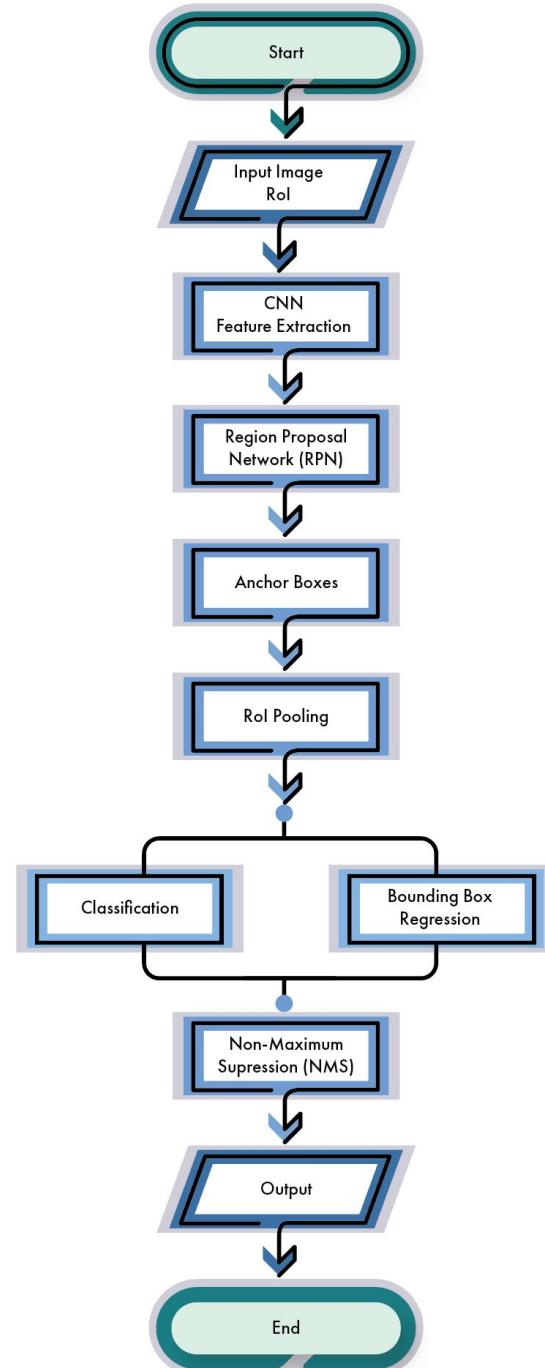
## Region Proposal Generation

- Fully Convolutional Network (FCN)
- RoI Pooling
- RoI Feature Vector

## Feature Extraction

- Entire image process





# Faster R-CNN: Architecture

Feature Map

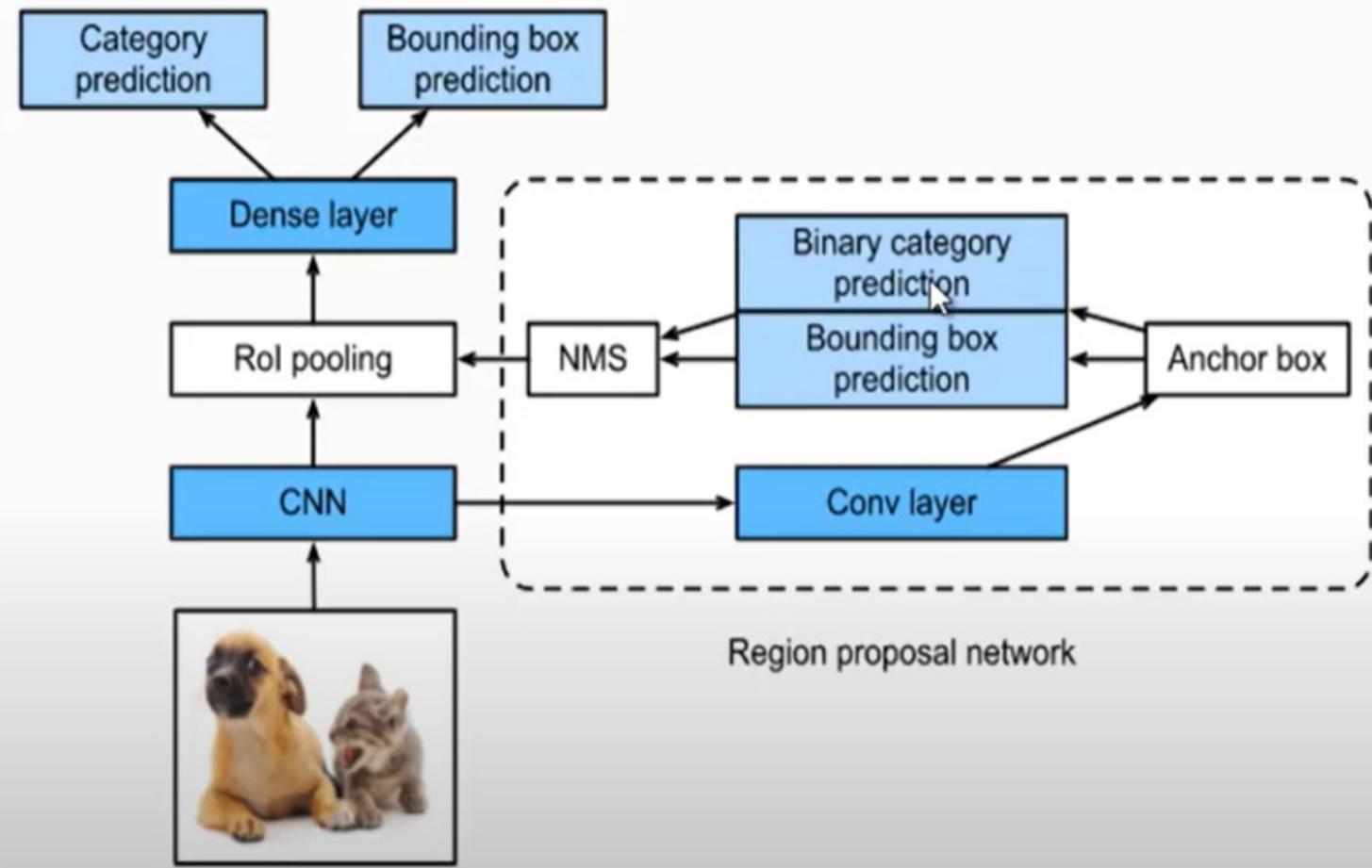
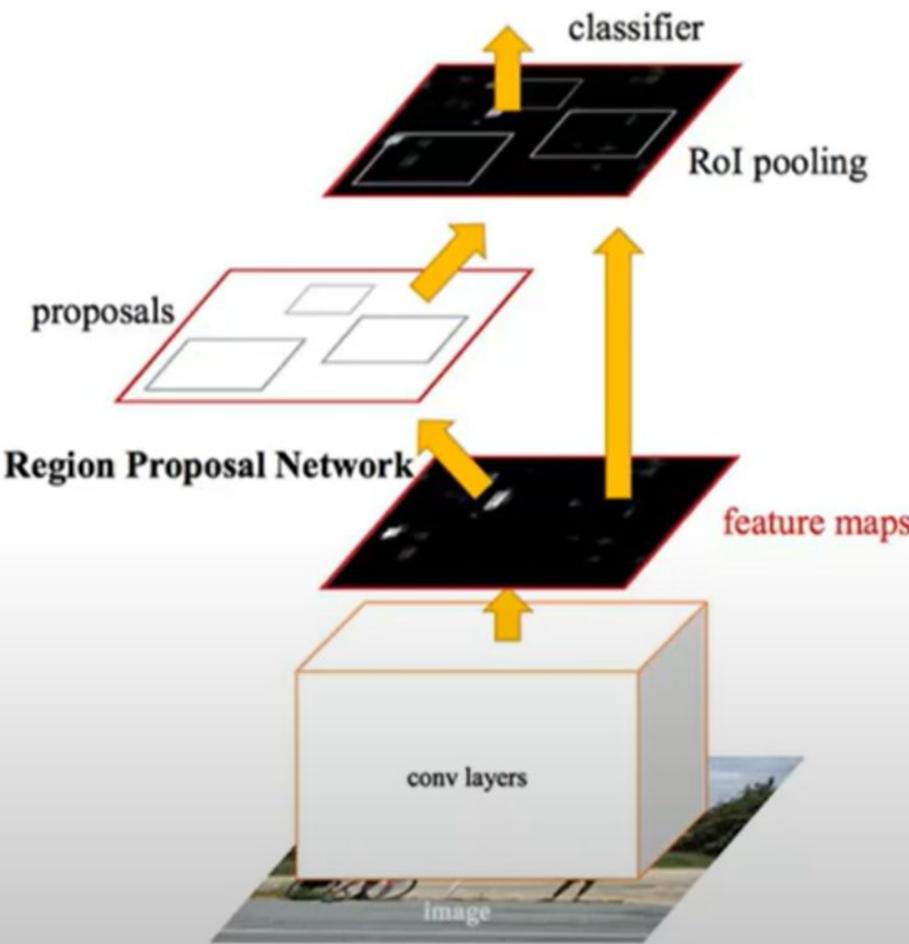
**Region Proposal Network (RPN)**

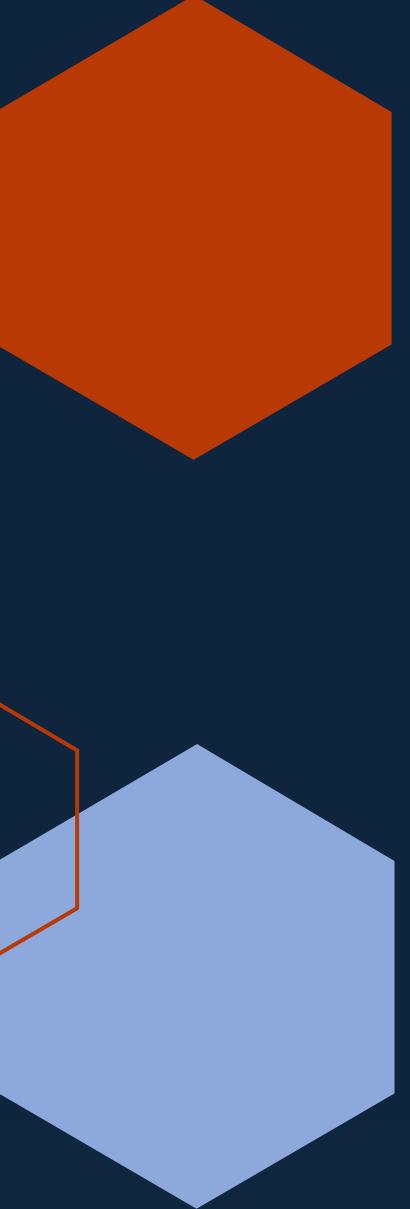
ROI Pooling

- Anchor boxes

**Classification & Bounding Box Regression**

**Non-Maximum Suppression (NMS)**





# Advantages of Faster R-CNN

**Efficient Region Proposal Generation**

**End-to-End Training**

**High-Quality Proposals**

**Practical Implementation**

# Mask R-CNN: Architecture

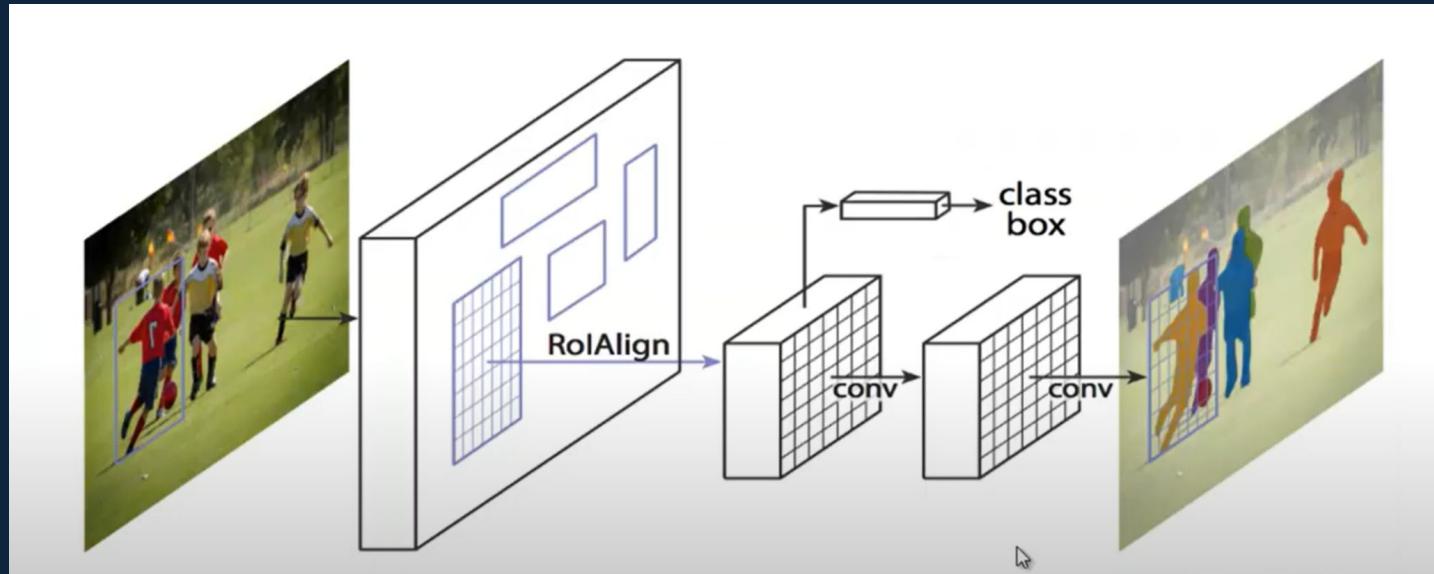
**Backbone CNN**

**Region Proposal Network  
(RPN)**

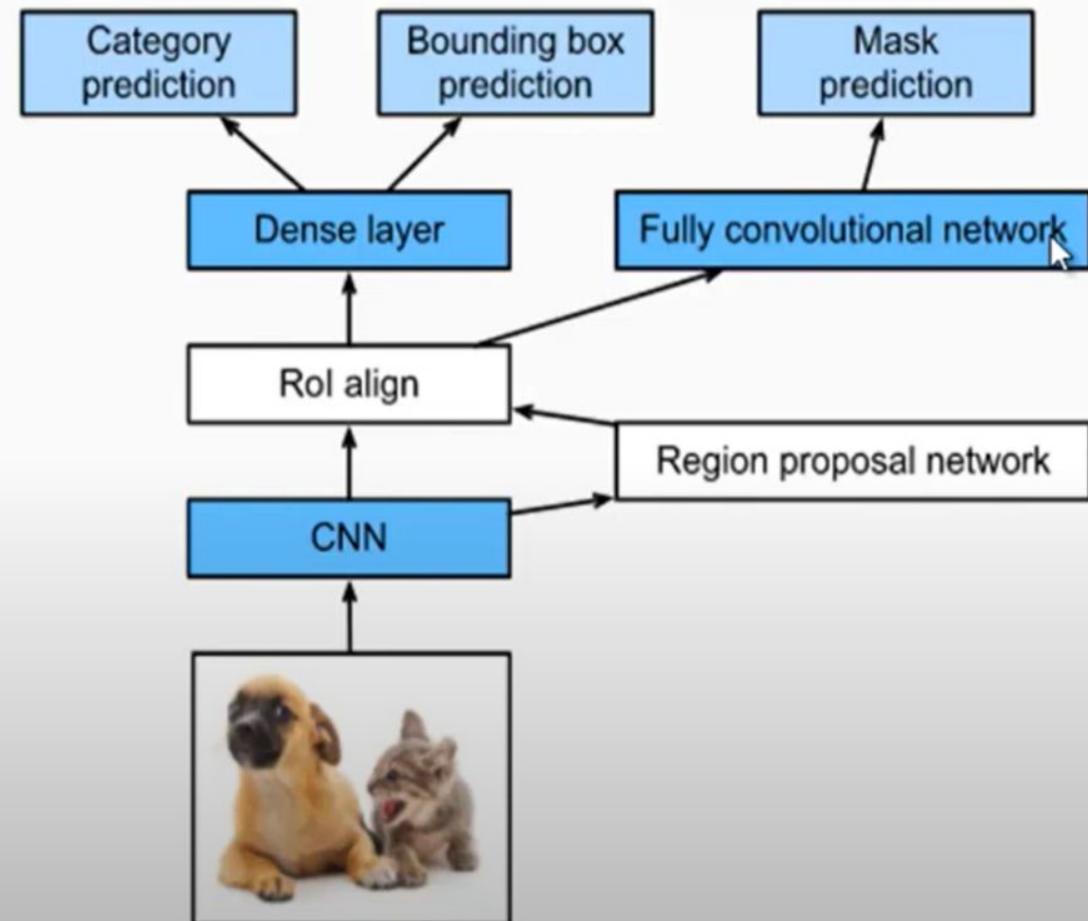
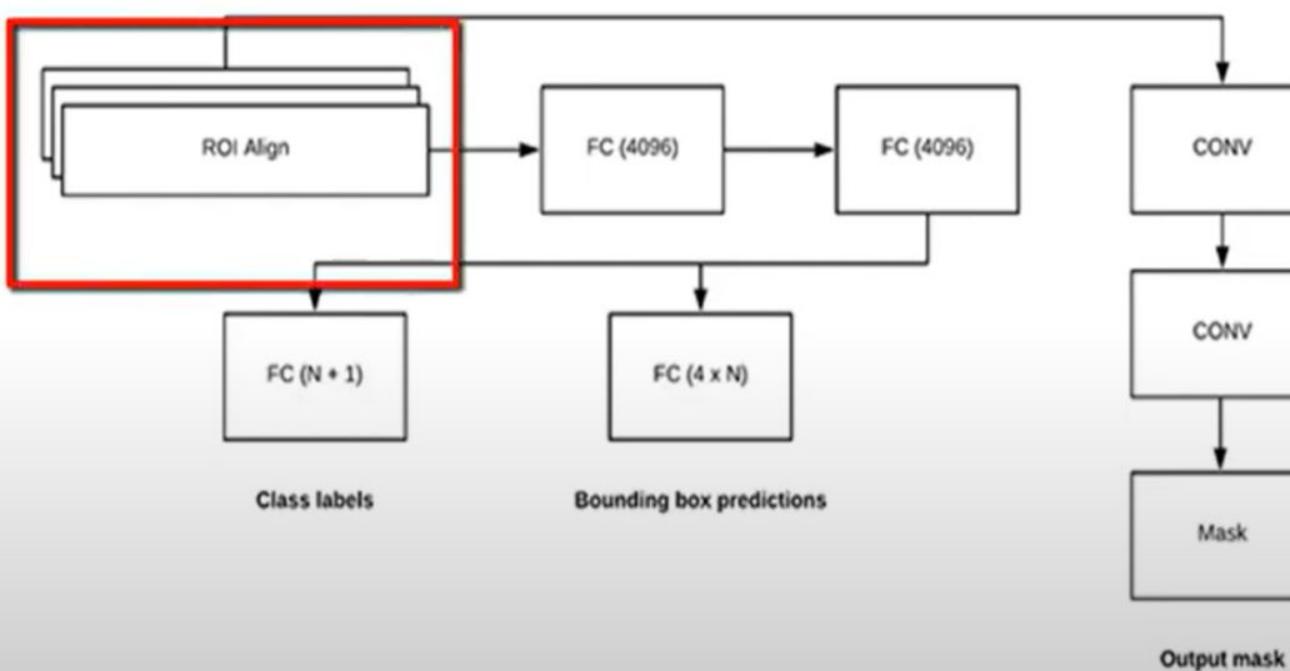
- Anchor Boxes

**Region of Interest Align  
(RoIAlign)**

- Pixel-level precision
- Preserves spatial information
- Precise alignment



# Mask R-CNN



# YOLOv8: Architecture

## Center to Front (C2F) Module

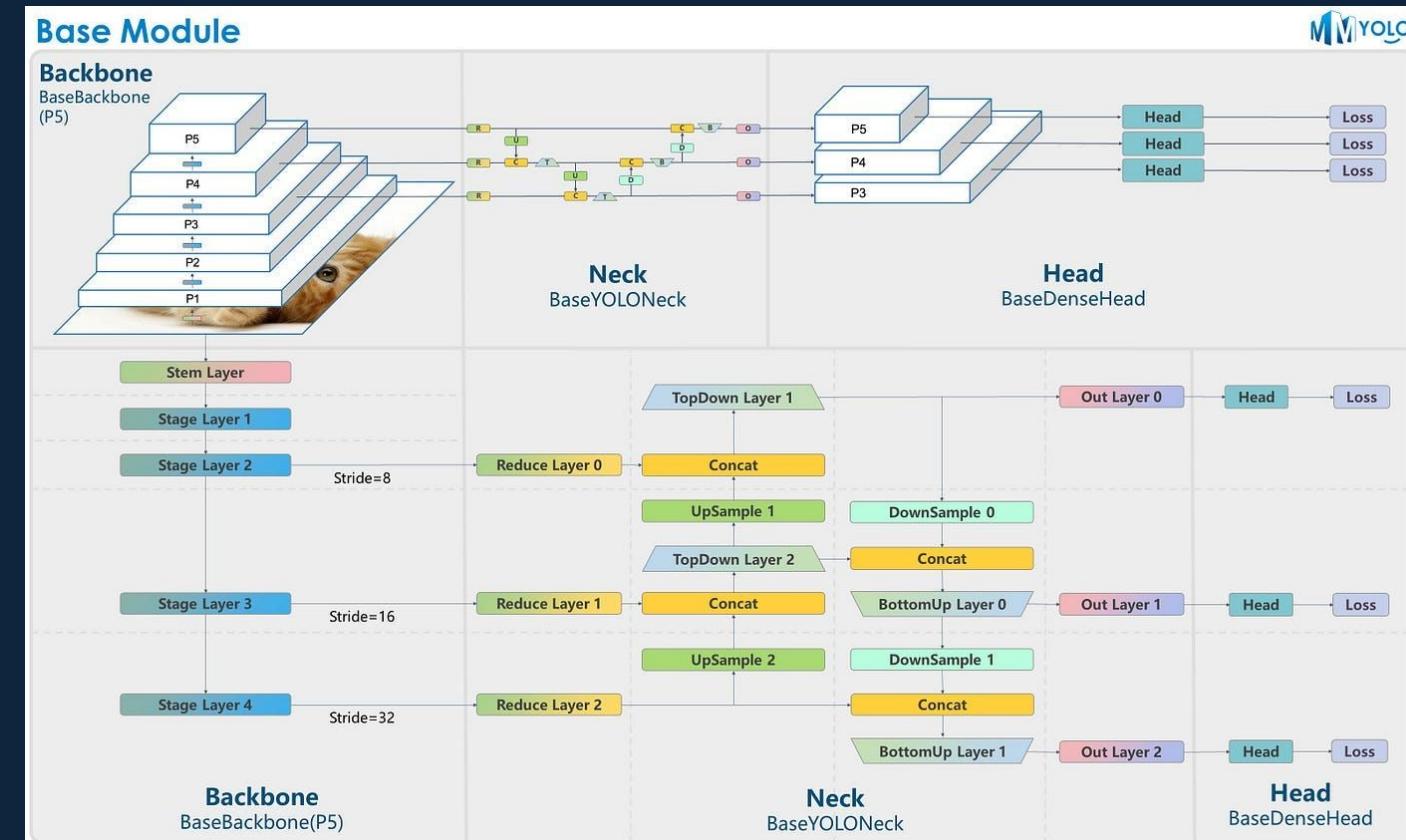
- Keeps the model lightweight
- Capturing gradient information
- Center sampling

## Anchor-Free Model

## Non-Maximum Suppression (NMS)

## Loss Function

- Complete IoU (CIoU)
- Distribution Focal Loss (DFL)



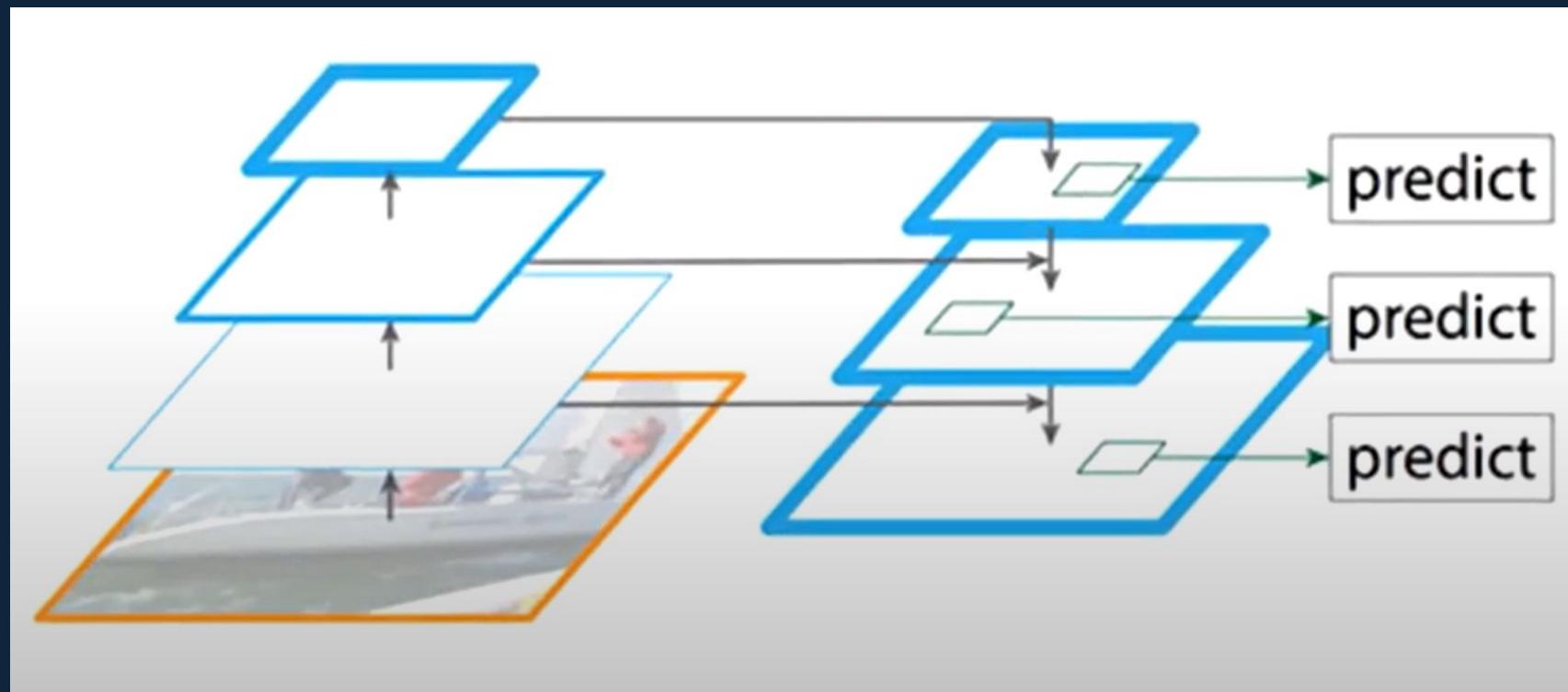
# Feature Pyramid Network (FPN)

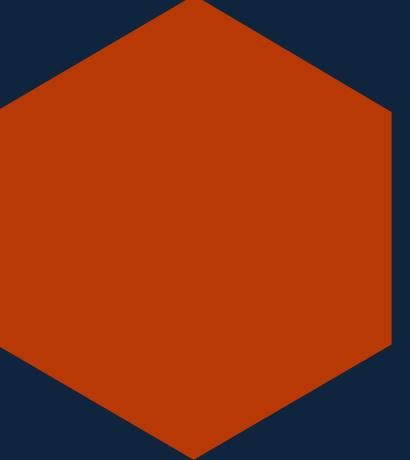
## Bottom-Up Pathway

- Traditional CNN & Pooling
- Feature Pyramid

## Top-Down Pathway

- Fuses Multi-scale Features
- High & Low resolution



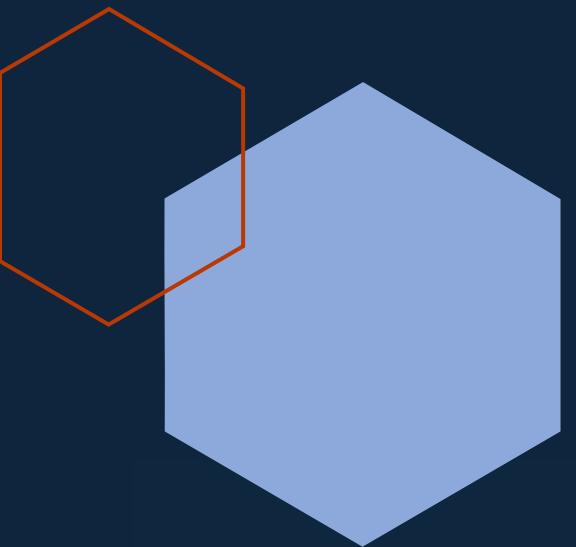


# Advantages of YOLOv8

**Real-Time Object Detection**

**Quick and Accurate**

**Small object detection**



# Overview of Models

R-CNN	Fast R-CNN	Faster R-CNN	Mask R-CNN	YOLOv8
<ul style="list-style-type: none"><li>Introduced selective search for generating region proposals.</li><li>Slow inference due to processing each region independently.</li><li>Not end-to-end trainable.</li></ul>	<ul style="list-style-type: none"><li>Introduced Region of Interest (RoI) pooling for efficient feature extraction.</li><li>Shared convolutional features for the entire image.</li><li>Faster inference compared to R-CNN.</li></ul>	<ul style="list-style-type: none"><li>Introduced Region Proposal Network (RPN) for generating region proposals.</li><li>End-to-end trainable.</li><li>Achieves faster inference than Fast R-CNN.</li></ul>	<ul style="list-style-type: none"><li>Extends Faster R-CNN by adding a parallel branch for mask prediction.</li><li>Provides pixel-level segmentation alongside object detection.</li><li>Enhances understanding of object shapes and boundaries.</li></ul>	<ul style="list-style-type: none"><li>Single-stage object detection model.</li><li>Divides the input image into a grid and predicts bounding boxes and class probabilities directly.</li><li>Known for real-time inference speed..</li></ul>

# Overview of Project

Datasets utilized	Challenges	Performance comparison	Conclusion
<ul style="list-style-type: none"><li>COCO dataset: Large-scale dataset for object detection, segmentation, and captioning tasks.</li><li>Custom dataset for military tank detection.</li><li>Custom dataset for SAM.</li></ul>	<ul style="list-style-type: none"><li>Robustness to environmental variations.</li><li>Mislabeling of objects.</li><li>Model stability during training.</li></ul>	<ul style="list-style-type: none"><li>Evaluation metrics: Accuracy, speed, robustness.</li><li>Benchmarking on standardized datasets like COCO.</li></ul>	<ul style="list-style-type: none"><li>Each model has its strengths and weaknesses.</li><li>Choice of model depends on specific requirements such as speed, accuracy, and deployment..</li></ul>



# Areas of focus

Performance evaluation

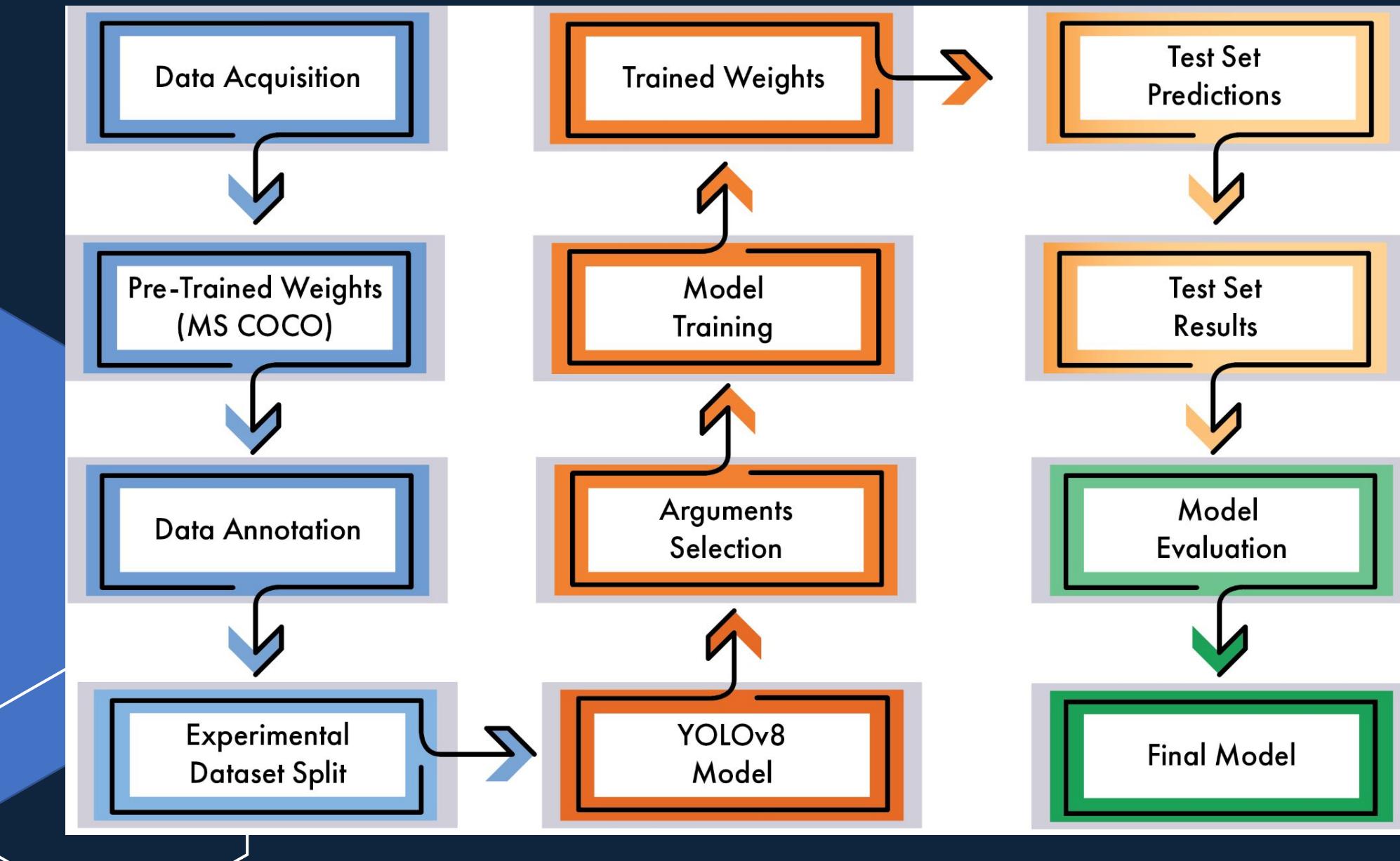
Resource Efficiency

Algorithmic Enhancements

Implementation

Real-World Applications

# Analysis



# Training

46433072405\_3b553d3749\_k.jpg



47491235592\_9d730bfe29\_k.jpg



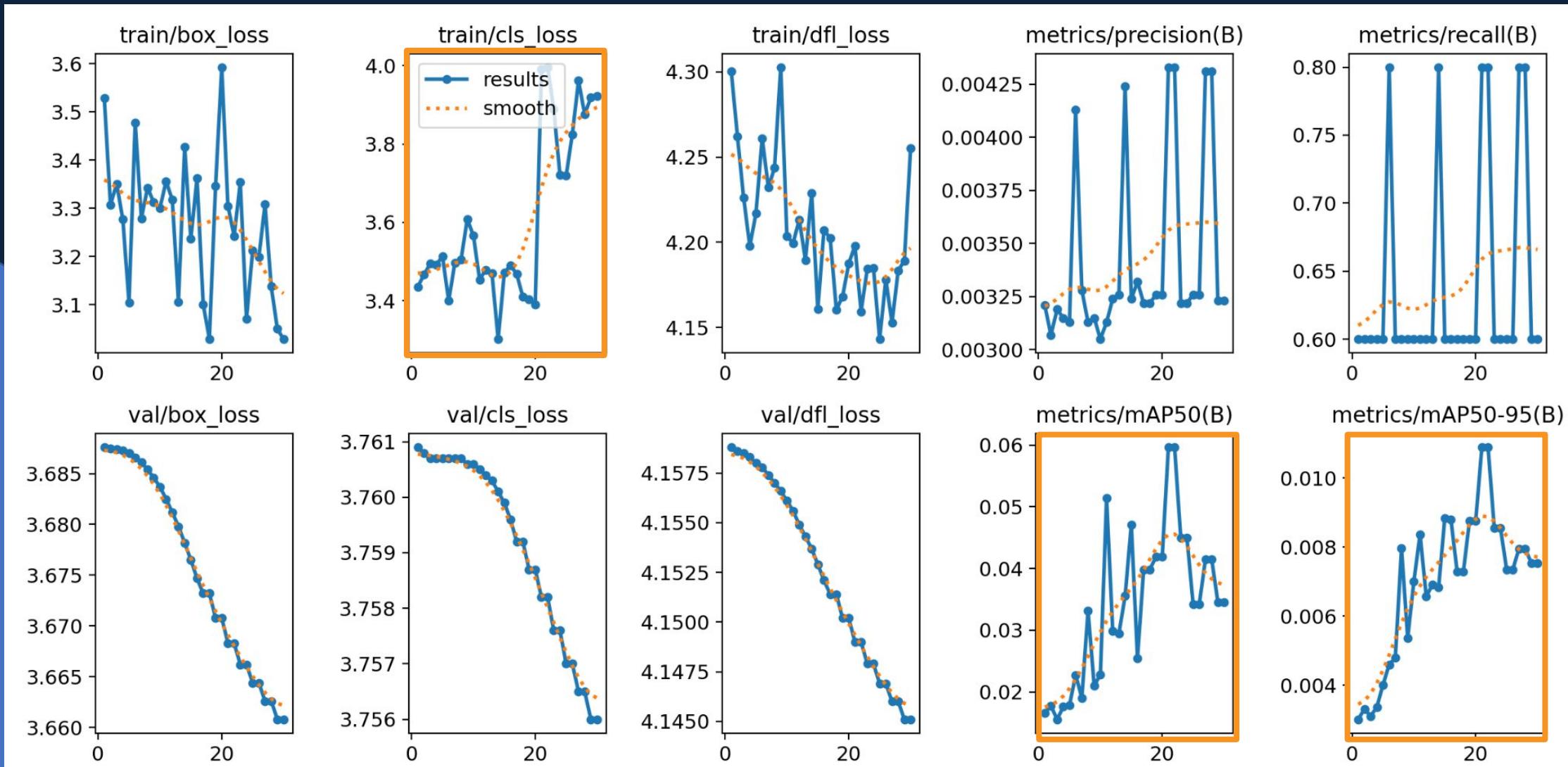
51707316241\_34ff790208\_c.jpg



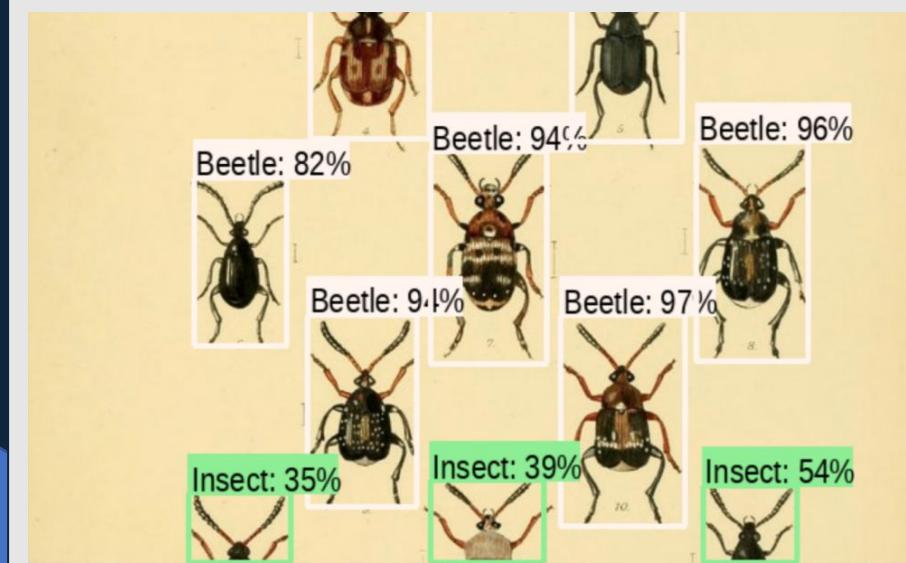
40022757021\_b526b862\_n\_w.jpg



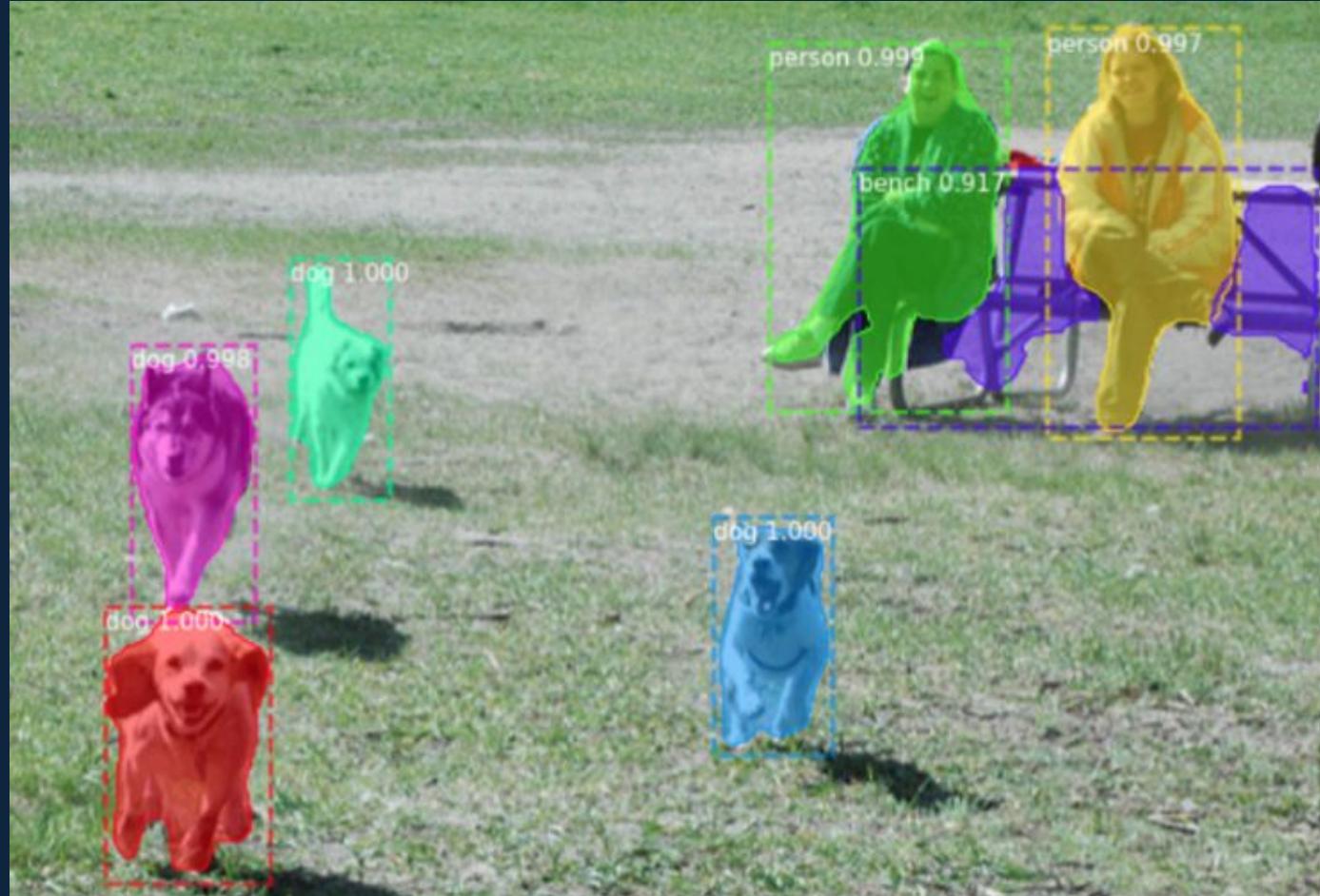
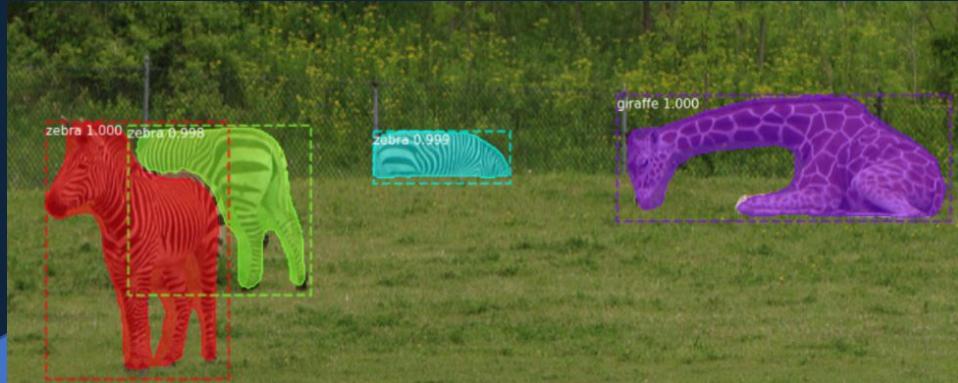
# Performance



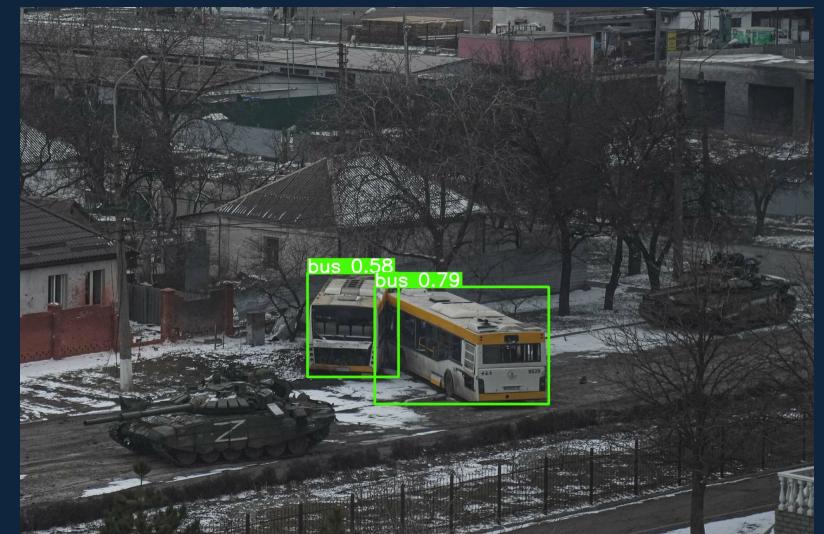
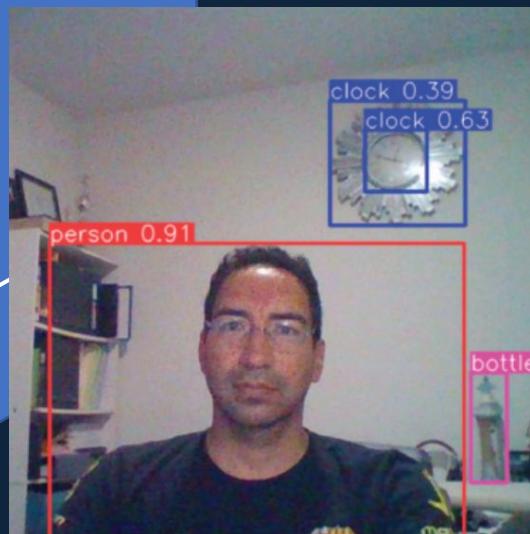
# Results of Faster R-CNN



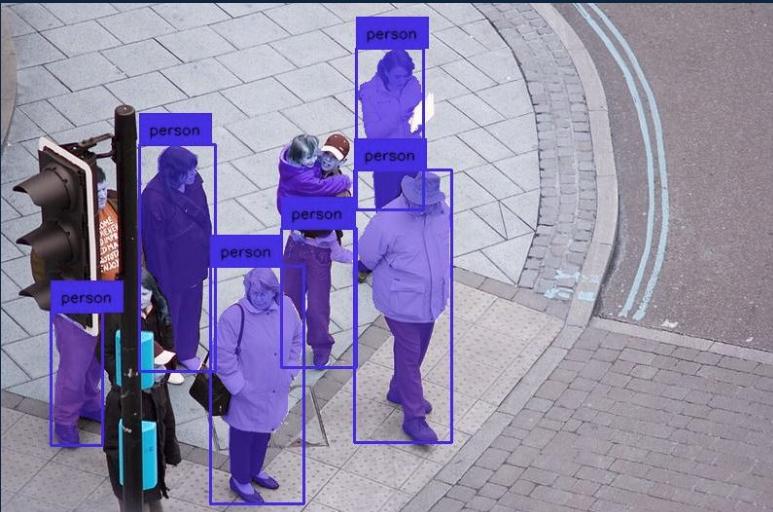
# Results of Mask R-CNN



# Results of YOLOv8



# Segment Anything SAM



helicopter 0.95



YOLO-World



# YOLOv9



## Information Bottleneck

- Addresses Data Loss
- Reversible Function

## Programmable Gradient Information (PGI)

- Reliable Gradient Information
- Update Network Weights
- Auxiliary Reversible Branch

## Generalized ELAN (GELAN)

- CSPNet + ELAN
- Gradient Path Planning

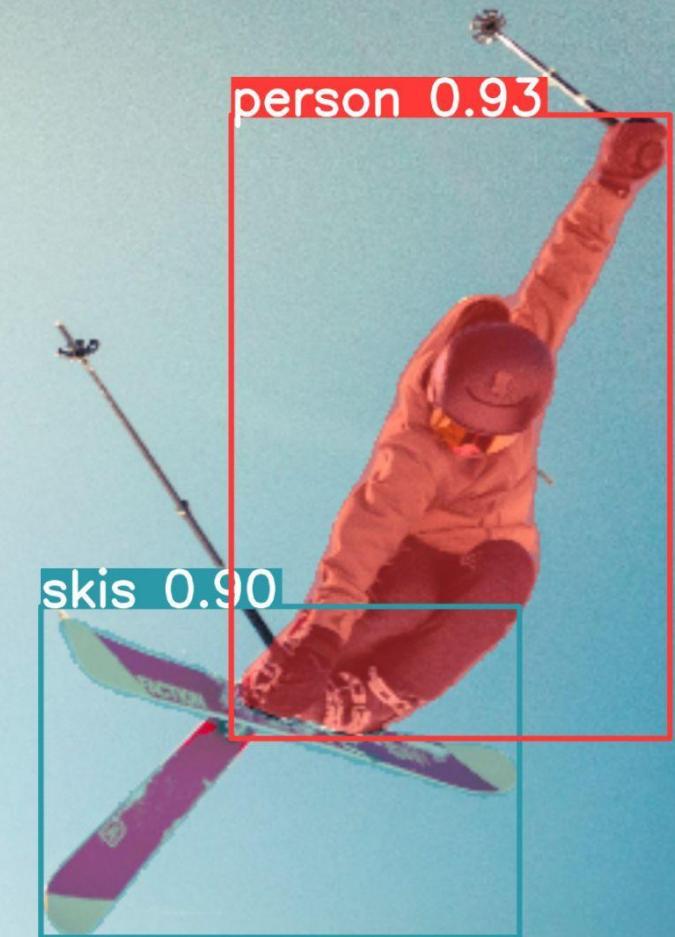
C. Wang, I. Yeh, and H. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv preprint arXiv:2402.13616v1*, 2024.

# Conclusion

**YOLOv8 is best for real-time  
Object Detection and Segmentation  
Balance in Accuracy & Speed**

Faster R-CNN does not offer real-time capabilities

Mask R-CNN excels in high accuracy



# Web Apps

RCNN-Webcam

RCNN-YOLOv8

# Questions