Dept Scientific Computing, Modelling and Simulation, SPPU

Title: Comparative Analysis of Object Detection using CNN Variants and YOLO using PyTorch

- Name: Bhushan Zade
- Roll No: MS2421
- Guide: Dr. Kaveri Kale
- Academic year: 2025–26
- Date: 27 DEC 2025

INTRODUCTION:

Object detection is a fundamental problem in computer vision that involves not only identifying objects present in an image but also localizing them using bounding boxes. With the rapid growth of applications such as autonomous driving, video surveillance, medical imaging, and robotics, efficient and accurate object detection techniques have become critically important. Traditional Artificial Neural Networks (ANNs) are not suitable for object detection tasks due to their inability to preserve spatial information. Convolutional Neural Networks (CNNs) address this limitation by learning hierarchical spatial features directly from images. Over time, several CNN-based object detection architectures such as R-CNN, Fast R-CNN, and Faster R-CNN have been developed to improve detection accuracy by incorporating region proposal mechanisms.In contrast, YOLO (You Only Look Once) introduced a regression-based approach that treats object detection as a single end-to-end prediction problem. YOLO performs detection in a single forward pass of the network, making it significantly faster than region-based CNN methods while maintaining competitive accuracy.This project aims to conduct a systematic comparative analysis of CNN-based object detection variants and YOLO using the PyTorch deep learning framework, focusing on accuracy, speed, and computational efficiency.

PROBLEM STATEMENT:

Objective : To compare the performance of CNN-based object detection models and YOLO in terms of detection accuracy, speed, and computational efficiency.

Input

- 1. RGB images containing one or more objects
- 2. Images from standard object detection datasets (PASCAL VOC)

Output

- Predicted object class labels
- Bounding box coordinates for detected objects

Comparison Criteria

- Detection accuracy (mAP)
- Inference time per image
- Training time
- Computational efficiency

MOTIVATION:

Choosing the appropriate object detection algorithm is crucial for real-world applications where both accuracy and speed are important. Region-based CNN models provide high detection accuracy but suffer from increased computational cost and slower inference. YOLO, on the other hand, enables real-time object detection but may face localization challenges in certain scenarios.

This project is motivated by the need to:

- Understand architectural differences between region-based CNNs and YOLO
- Analyze trade-offs between accuracy and real-time performance
- Gain practical experience in implementing modern object detection models using PyTorch
- Strengthen foundational understanding of deep learning for computer vision

LITERATURE SURVEY:

LeCun et al. (1998) laid the foundation of CNNs through LeNet, demonstrating the effectiveness of convolutional layers for image-based tasks. Girshick et al. introduced R-CNN, which combined region proposals with CNN-based feature extraction, significantly improving object detection accuracy.

To address R-CNN's high computational cost, Fast R-CNN and Faster R-CNN were developed, with Faster R-CNN introducing the Region Proposal Network (RPN) to speed up detection. These models improved detection precision but still faced challenges in real-time applications.

Redmon et al. proposed YOLO, a single-stage detector that performs object detection as a regression problem. YOLO predicts bounding boxes and class probabilities in one pass, achieving real-time performance. Subsequent versions of YOLO further improved accuracy and robustness, making it a strong alternative to region-based detectors.

The comparative study presented in the referenced IEEE paper highlights that while CNN variants achieve strong localization accuracy, YOLO significantly outperforms them in terms of detection speed, especially for video and real-time systems.

proposed AlexNet, which revolutionized image recognition by achieving breakthrough performance on the ImageNet dataset. Subsequent models such as VGGNet, ResNet, and Inception further improved CNN capabilities, emphasizing depth and modular design. These works collectively highlight the evolution of neural network architectures from fully connected ANNs to highly efficient CNNs

PROPOSED METHODOLOGY:

The proposed project will be carried out in the following stages:

1. Dataset Selection

- Use a standard object detection dataset such as:
    - PASCAL VOC
    - MS COCO

2. Data Preprocessing

- Image resizing and normalization
- Annotation parsing (bounding boxes and labels)
- Dataset splitting into training and testing sets

3. Model Development

Implement and train the following models using PyTorch:

- CNN-based object detectors:
    - R-CNN / Fast R-CNN / Faster R-CNN
- YOLO-based detector

4. Training Configuration

- Optimizer: Adam / SGD
- Loss functions:

        - Classification loss
        - Bounding box regression loss
    - Batch size and epochs selected based on hardware constraints

5. Evaluation Metrics

    - Mean Average Precision (mAP)
    - Precision, Recall, and F1-score
    - Inference time per image
    - Training time comparison

6. Comparative Analysis

    - Quantitative comparison using metrics
    - Qualitative analysis using detected images
    - Visualization of bounding boxes and confidence scores

Dataset and Evaluation:

The selected dataset consists of labeled images with object class annotations and bounding box coordinates. The dataset will be preprocessed and resized according to model requirements. Evaluation will be conducted using:

- Mean Average Precision (mAP)
- Precision, Recall, and F1-score
- Confusion matrix (where applicable)
- Detection speed (frames per second)

References:

- Girshick, R. et al., “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” IEEE TPAMI.
- Redmon, J., & Farhadi, A., “You Only Look Once: Unified, Real-Time Object Detection,” CVPR.
- Krizhevsky, A. et al., “ImageNet Classification with Deep Convolutional Neural Networks,” NIPS.
- Comparative Study on Image Detection using Variants of CNN and YOLO, IEEE COM-IT-CON 2022.
- Girshick, R. et al., “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” IEEE TPAMI.