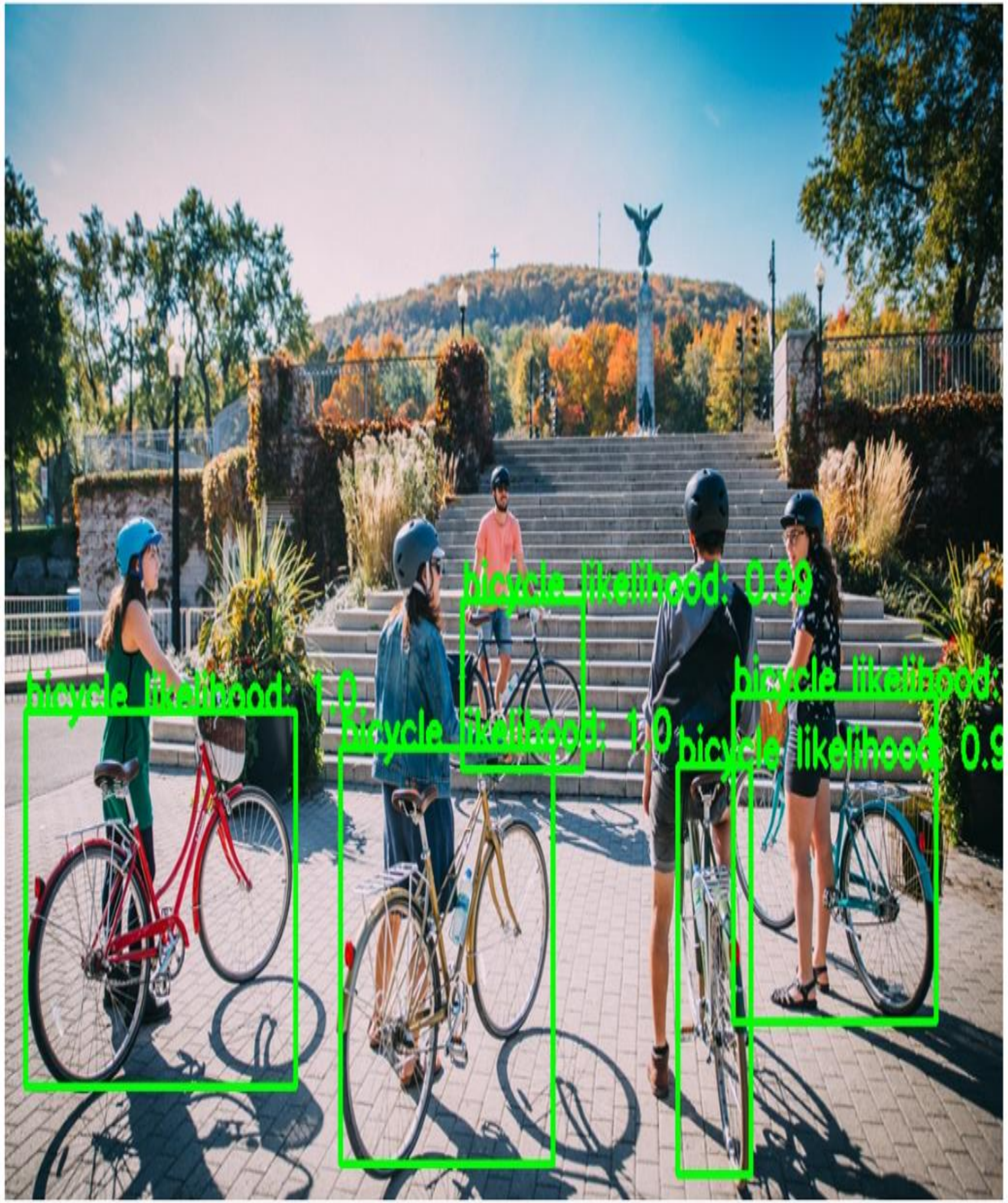
INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

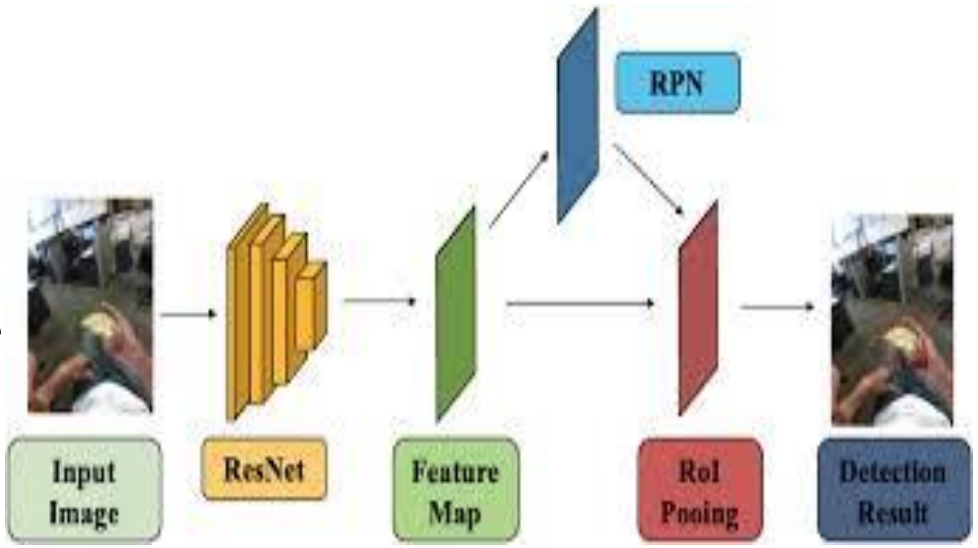
HYDERABAD

"Topic- Object detection using faster r-cnn model on COCO Dataset "



Introduction:

A computer vision task that aims to identify and localize the object of interest in an image or video is called object detection. A major advancement in this direction is the Faster R-CNN (Faster Region-based Convolutional Neural Network)-speed combined with accuracy enables nearly real-time object detection in many applications, spanning from autonomous driving to surveillance systems. Objective of this project: This project aims at including the workings of the Faster R-CNN architecture.



1

Objective:

To develop an efficient object detection model using Faster R-CNN architecture to accurately identify and localize objects within images from the COCO dataset.

2

Project Overview:

It utilizes a pre-trained Faster R-CNN having a ResNet-50 backbone with large COCO dataset. The project involves preprocessing data, training the model, evaluating its performance, and optimizing it for real-time applications.

3

Challenges:

1. Handling diverse and extensive COCO dataset variations.
2. Addressing computational resource demands for training Faster R-CNN.
3. Managing overfitting due to the complexity of the model.

.

Faster R-CNN:

Faster R-CNN consists of the following few layers:

Convolutional Layers: Using the VGG or ResNet layer, it extracts feature maps from the input image while capturing the essential spatial and semantic information.

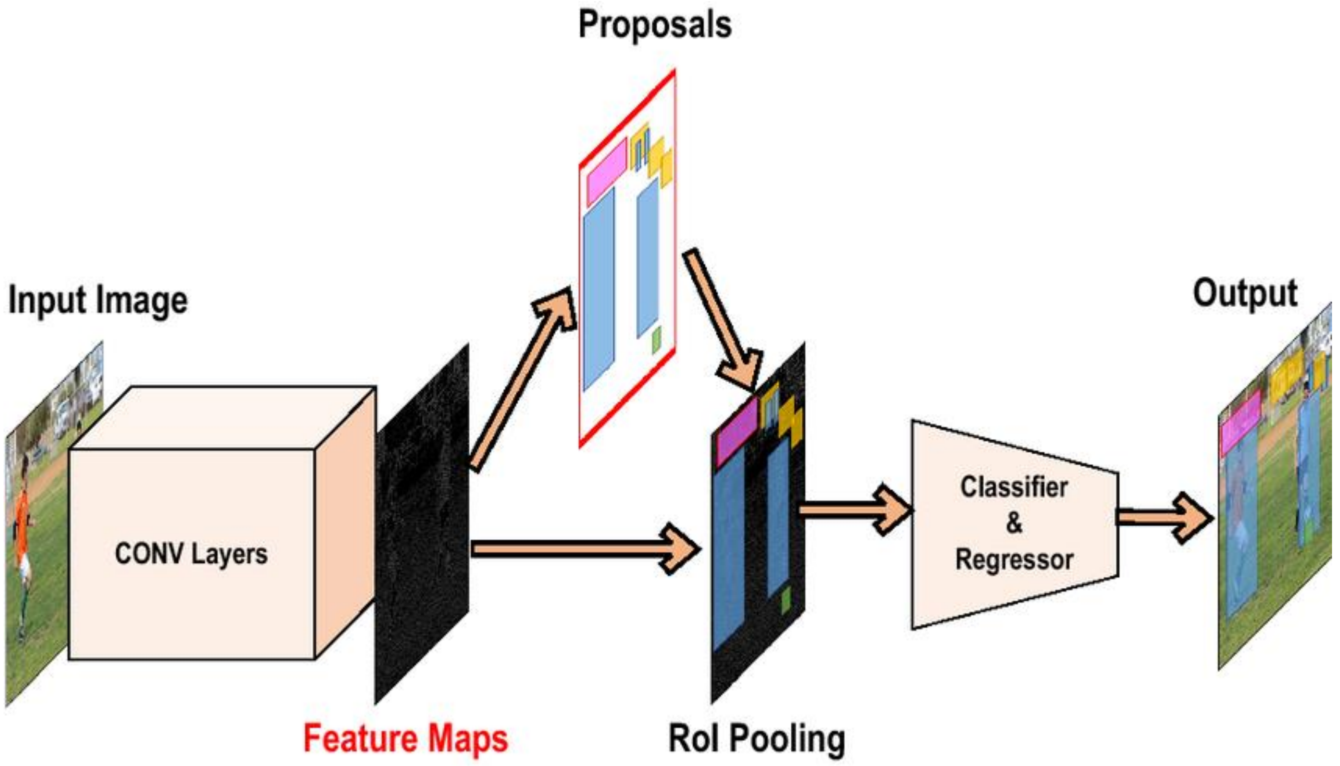
RPN: Sliding a small network over these feature maps produces a score based on object and box coordinates for each of the proposed regions.

Anchor Boxes: The predefined boxes within the RPN of different scales and aspect ratios? Anchor boxes? Getting back on the path: to cover the different sizes and shapes of objects and help later on in generating the right proposals for the region.

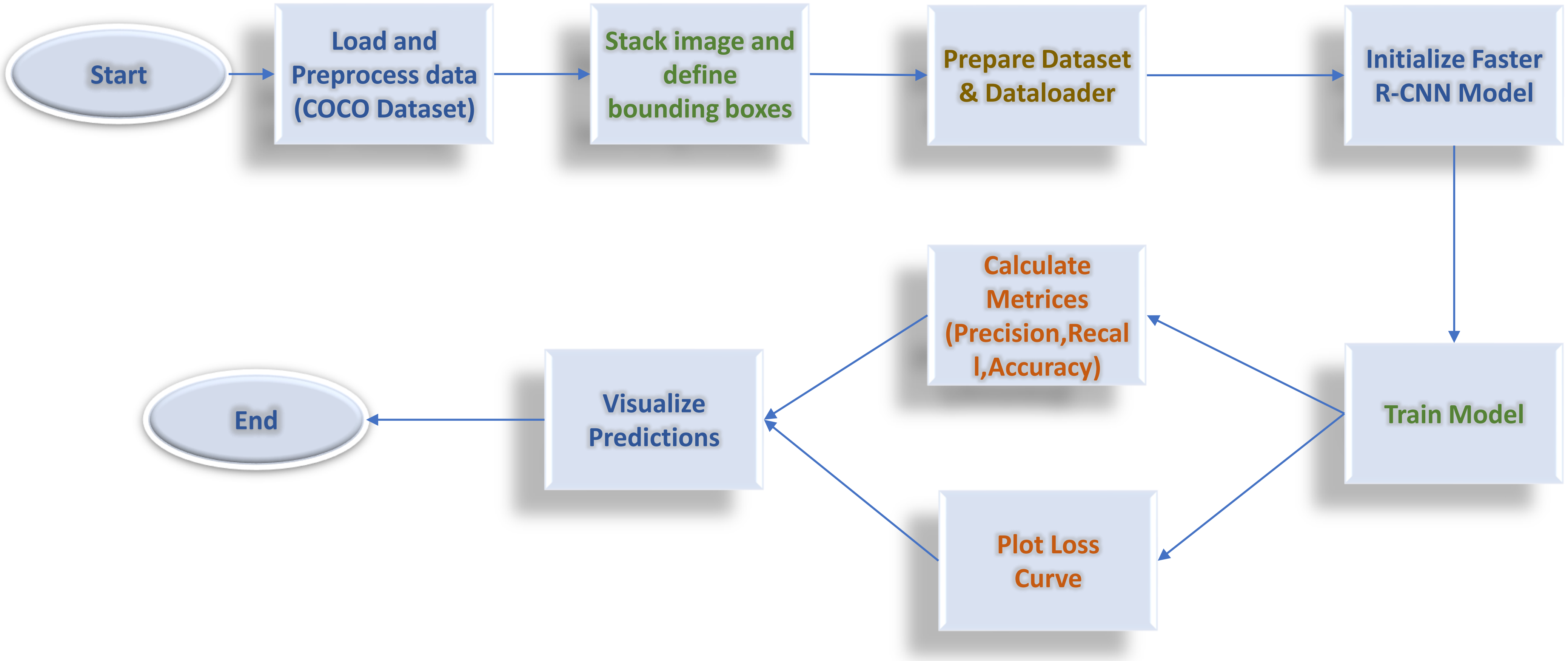
RoI pooling layer: Pooling in feature maps from each proposed region is in a fixed size. This guarantees that any feature extraction will process these very regions identically for purposes of classification and refinement of bounding box position.

Fully Connected Layers: Further refine the RoI features with some last classifier and bounding box regression layers to yield the final detection.

Classification and Bounding Box Regression: The final layer assigns labels to each region based on class and refines box positions for object localization.



Methodology:



Literature Survey:

Author(s) &Year	Title	Objective	Methodology	Key Contributions	Remarks
Ren, S., He, K., Girshick, R., & Sun, J. (2017)	Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks	To develop a faster, more accurate object detection system for real-time applications.	Introduces Region Proposal Networks (RPN) for efficient object proposal generation, integrated into the Faster R-CNN framework.	Proposed RPNs, significantly improving detection speed and accuracy over previous methods.	Faster R-CNN became foundational for real-time object detection tasks, influencing various subsequent models.
Zhu, Y., Liu, Y., Shi, W., & Chen, X. (2020)	Research on Vehicle Detection and Classification Algorithm Based on Improved Faster R-CNN	To improve vehicle detection and classification in real-world conditions using Faster R-CNN.	Enhances Faster R-CNN by tuning parameters and structures specific to vehicle detection challenges.	Demonstrates significant improvements in detecting and classifying vehicles in complex environments.	Provides insight into adapting existing models for specific domains, such as transportation surveillance.
Girshick, R. (2015)	Fast R-CNN	To speed up R-CNN-based object detection by eliminating the need for per-region convolution.	Fast R-CNN uses RoI pooling to apply a single CNN pass on the entire image, followed by region classification and bounding box regression.	Significantly reduces computational cost, making R-CNNs more practical for real-time applications	Fast R-CNN laid the groundwork for Faster R-CNN, reducing computational demands in object detection.
Dosovitskiy et al., 2021	An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale	Assesses transformers as a viable alternative to convolutional networks in vision task	Applies transformer architecture to image patches instead of traditional CNNs	Demonstrated compe titive performance of transformers in large-scale vision tasks	Set a precedent for transformer-based architectures in vision

Literature Survey:

Zhou et al., 2022	Improved Multi-Class Object Detection via Faster R-CNN with Enhanced RPN	Enhance Faster R-CNN's multi-class detection capability	Modified Faster R-CNN's RPN for better proposal generation in complex scenarios	Increased detection accuracy across multiple classes	Improved Faster R-CNN for diverse environments
Zhang, Yang & Zhang, 2022	A Comprehen sive Review on Object Detec tion Based on Deep Learning	Review advances in object detection using deep learning	Extensive survey covering methods like Faster R-CNN, YOLO, SSD, and transformer-based models	Provided insights into current challenges and future directions in object detection	A valuable resource for understanding the landscape of deep-learning-based object detection
Hong et al., 2023	Efficient Object Detection in the Wild: Faster R-CNN with Edge Features	Improve Faster R-CNN's performance in complex, real-world scenarios	Added edge features to enhance RPN's proposal accuracy in the wild	Achieved more robust detection in variable environments by leveraging edge features	Improved Faster R-CNN for practical applications in diverse contexts
Tang et al., 2023	Faster R-CNN and YOLO Approaches Combined for Improved Real-Time Detection	Combine Faster R-CNN and YOLO to leverage strengths of both	Hybrid approach using YOLO's speed with Faster R-CNN's accuracy	Increased real-time detection accuracy by combining strengths of both models	Shows potential for hybrid model development
Ma et al., 2023	Object Detection with Dynamic Pruning and Transformer-Based Improvements in Faster R-CNN Models	Enhance Faster R-CNN with dynamic pruning and transformers for efficiency	Uses transformer layers with dynamic pruning to reduce computation	Achieved faster and more efficient detection, with reduced computational load	Latest advancements combining transformers with Faster R-CNN

Image Stacking And Bounding Boxes:

1 **Stacked Image:**

Sampling 4-digit rows per image..

2 **Bounding Box Generation:**

Defined bounding boxes for each digit based on the column position in the row

3 **Key Code Snippet:**

for col_idx in range(num_per_row): x_min = col_idx * image_size x_max = x_min + image_size row_bboxes.append([x_min, 0, x_max image_size])

Visualizing Stacked Image:

Training Process and Metrics Calculation:

1. **Training Loop:**

   Forward pass, loss calculation, back-propagation.

2. **Metrics:**

   Precision, Recall, Accuracy calculated per epoch.

3. **Key Snippet:**

   ```
   precision = precision_score(true_labels, pred_labels, average='micro')
   ```

Future Scope:

- **Real-Time Optimization**

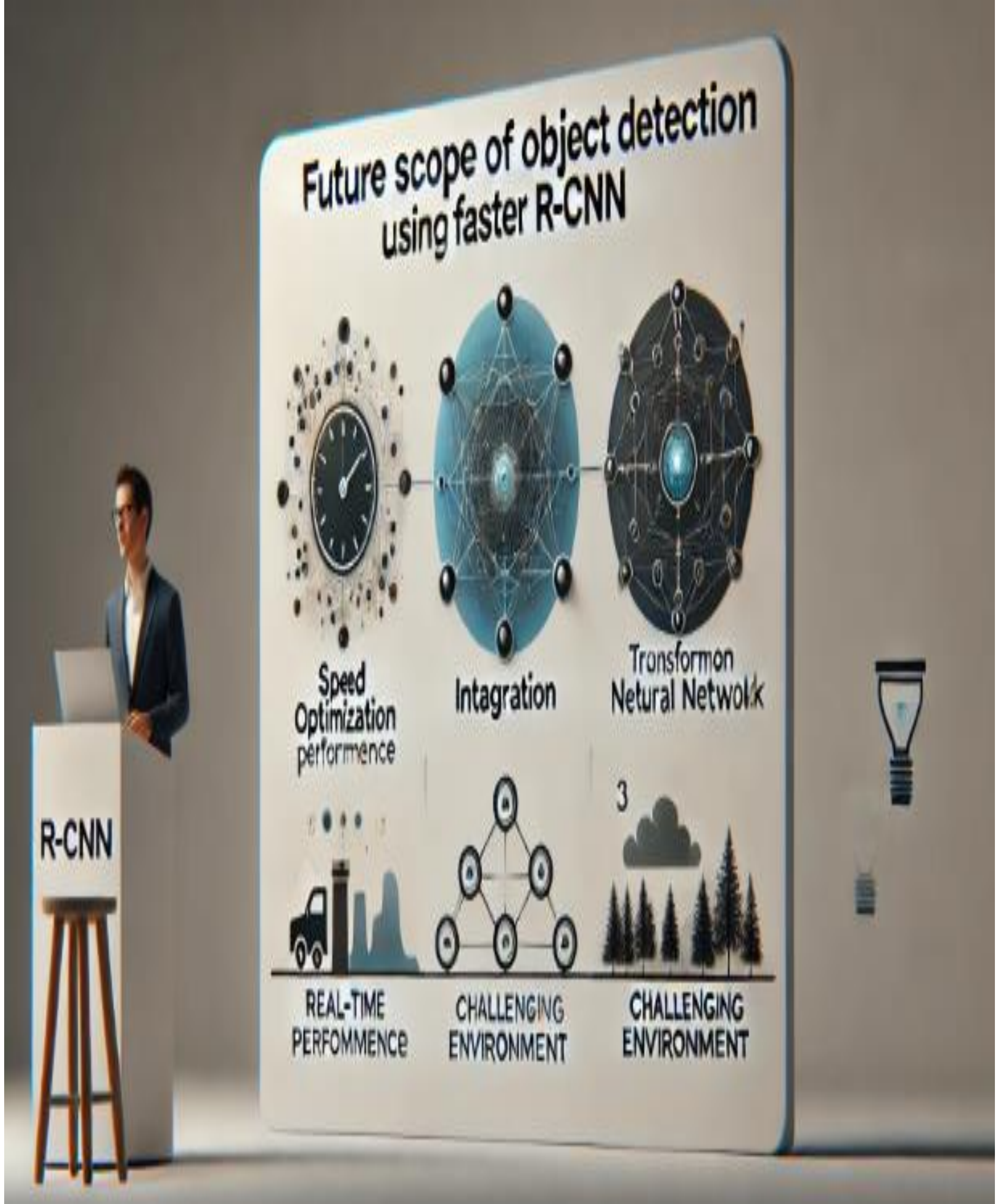
Edge computing and model compression will probably soon enable the acceleration of application and optimization of efficiency to extend into the world of real-time usage.

- **Integration with Transformers**

Application with transformers can boost the power of detection, and thus can extend its usage towards complex scenarios such as autonomous driving or even medical imaging.

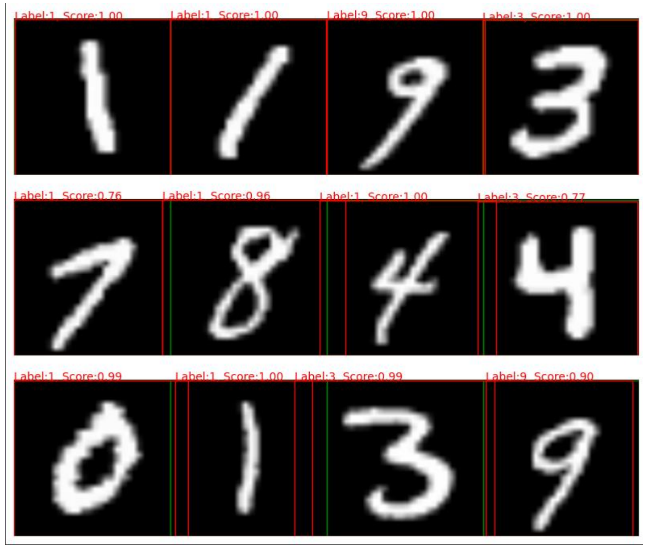
- **Enhanced Performance in Demanding Conditions**

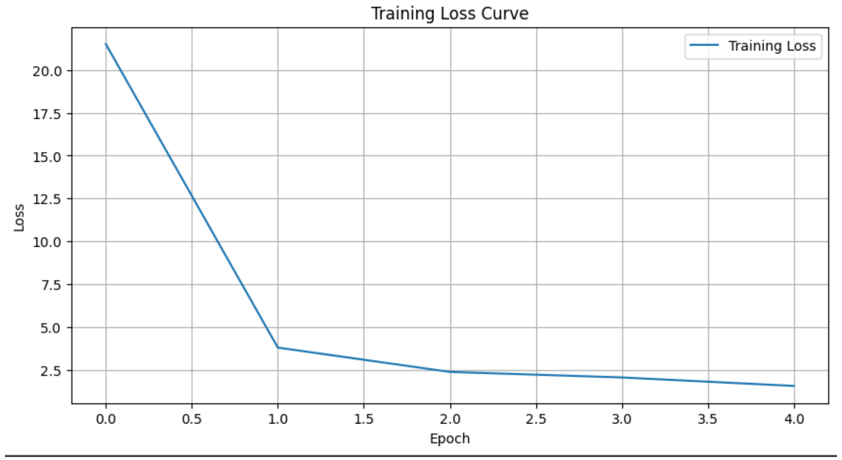
Low light, occlusion, and crowding adaptations are enhanced for applications that include surveillance, wildlife monitoring, and emergency response.



Training Results:

```
Epoch [1/5], Total Loss: 21.5007, Precision: 0.8300, Recall: 0.8300, Accuracy: 0.8300
Epoch [2/5], Total Loss: 3.8146, Precision: 1.0000, Recall: 1.0000, Accuracy: 1.0000
Epoch [3/5], Total Loss: 2.3982, Precision: 0.7100, Recall: 0.7100, Accuracy: 0.7100
Epoch [4/5], Total Loss: 2.0740, Precision: 1.0000, Recall: 1.0000, Accuracy: 1.0000
Epoch [5/5], Total Loss: 1.5810, Precision: 1.0000, Recall: 1.0000, Accuracy: 1.0000
```





1

2

3

Epoch-wise Performance:

Loss decreased and metrics impr oved at 5 epochs.

Sample Output:

Total loss and accuracy metrics at each epoch.

Graph:

Loss curve which illustrates the training.

REFERENCE:

1. Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.
2. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. "You Only Look Once: Unified, Real-Time Object Detection," CVPR, 2016.
3. Carion, N., et al. "End-to-End Object Detection with Transformers," European Conference on Computer Vision (ECCV), 2020.
4. Dosovitskiy, A., et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations (ICLR), 2021.
5. Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. "YOLOX: Exceeding YOLO Series in 2021." arXiv:2107.08430, 2021.
6. Zhou, C., et al. "Improved Multi-Class Object Detection via Faster R-CNN with Enhanced Region Proposal Network," IEEE Access, 2022.
7. Zhang, Y., Yang, F., & Zhang, H. "A Comprehensive Review on Object Detection Based on Deep Learning," IEEE Access, 2022.
8. Hong, Y., et al. "Efficient Object Detection in the Wild: Faster R-CNN with Edge Features," Pattern Recognition, 2023.
9. Tang, Z., et al. "Faster R-CNN and YOLO Approaches Combined for Improved Real-Time Detection," Computer Vision and Image Understanding, 2023.
10. Ma, L., et al. "Object Detection with Dynamic Pruning and Transformer-Based Improvements in Faster R-CNN Models," arXiv:2310.04829, 2023.

Team Name: Innovators

Group Members :

S.No.	Name
1	Devansh Patel
2	Vishesh Shahu
3	Akshay Gupta
4	Pradeep Shahu
5	Sumit Patil

TA: Anirvinya Gururajan