# Road Object Detection using Deep Learning

## Introduction:

Object detection is a crucial task in computer vision that involves identifying and localizing objects within an image. This project aims to implement a road object detection system using the YOLOv3 pretrained model. The objective is to accurately detect and classify objects in images.

Approach:

2.1. Model Selection:

Several pretrained object detection models were evaluated for their performance and suitability for the project. After careful consideration, the YOLOv3 model was chosen due to its strong balance between accuracy and inference speed. The model had demonstrated impressive object detection capabilities in previous research and real-world applications.

YOLOv3(You Only Look Once v3) is a popular object detection algorithm, it is designed to detect objects in images and videos by dividing the input image into a grid and predicting bounding boxes and class probabilities for each grid cell.

2.2. Model Integration:

The YOLOv3 pretrained model was integrated into the project framework. This involved installing the necessary dependencies and libraries required for model inference.

To integrate YOLOv3 in our project which works on the Darknet 53 architecture, and to implement this Pytorch along with GPU was used.

The integration ensured seamless utilization of the model for object detection tasks.

2.3. Pre-processing:

Prior to feeding the images into the YOLOv3 model, pre-processing steps were implemented to prepare the input data. These steps included resizing the images to the model's input size (i.e., 416x416), normalizing pixel values, and adjusting any necessary transformations. Pre-processing was crucial to ensure compatibility between the input data and the model's requirements.

2.4. Object Detection:

The YOLOv3 model was employed to perform object detection on the pre-processed images. The model utilized a single neural network to simultaneously predict bounding boxes and class probabilities for multiple objects within an image. This approach facilitated efficient and accurate object detection.

2.5. Post-processing:

After the YOLOv3 model generated the object detection results, post-processing techniques were applied to refine the outputs. Non-maximum suppression (NMS) was utilized to eliminate redundant bounding boxes and retain the most confident detections. The NMS & IOU threshold values was set to 0.5. Additionally, bounding box filtering and thresholding were employed to remove false positives and improve the overall accuracy of the detected objects.

2.6. Evaluation:

The performance of the object detection system was evaluated using a separate test dataset. Various evaluation metrics, including confidence level and intersection over union (IoU), were calculated. These metrics provided insights into the system's effectiveness in correctly identifying and localizing objects.

## Results and Conclusion:

After implementing the above approach using the YOLOv3 pretrained model, the following outcomes and conclusions were observed:

The YOLOv3 model achieved an impressive accuracy range lies between 0.7 to 0.99 out of [0,1] on the test dataset. This demonstrated its capability to accurately detect objects in images.

Throughout the project, certain limitations and challenges were encountered. These included handling occluded objects, dealing with complex backgrounds, and optimizing the system for real-time performance.

Despite the challenges, the implemented solution utilizing the YOLOv3 pretrained model met the project's objectives of accurate and efficient object detection.

# Future Work:

Building upon the current project, several avenues for future work were identified:

* Explore other pretrained models or variations of the YOLO architecture to further enhance the accuracy and performance of the object detection system.
* Investigate the integration of real-time object detection capabilities, such as deploying the system on edge devices or optimizing the model for faster inference.
* Extend the project to handle object detection in video sequences or other types of data, enabling real-time tracking and analysis of objects.