# **Project Report: Object Detection using SSD-YOLOv12**

## **1. Introduction and Objective**

In this project, I implemented and fine-tuned an object detection model using the SSD-YOLOv12 architecture. The objective was to detect multiple object classes on a dataset containing images of cars, traffic lights, and other traffic-related objects captured under snowy and blurred conditions. The primary goal was to achieve high detection accuracy while maintaining efficient inference speed suitable for real-time applications.

## **2. Model Selection and Justification**

YOLO (You Only Look Once) models are well-known for their real-time object detection capabilities. The YOLOv12 architecture improves over previous versions by combining SSD (Single Shot MultiBox Detector) with YOLO’s efficiency, offering:

* Faster inference due to single-stage detection.
* Improved accuracy by leveraging feature pyramid networks and better backbone architectures.
* Lightweight model size suitable for deployment on hardware with limited resources.

Compared to two-stage detectors like Faster R-CNN, SSD-YOLOv12 provides better speed without a significant compromise on precision, making it the ideal choice for applications requiring both speed and accuracy.

## **3. Dataset and Preprocessing**

The dataset contains 31,676 images with 12 annotated classes, including cars, traffic lights, and their variations. Images were labeled in YOLO format, with bounding boxes normalized relative to image dimensions.

The dataset was split into:

* 70% training
* 15% validation
* 15% testing

Preprocessing included resizing images to 640x640 pixels and normalization to prepare the data for training.

## **4. Fine-tuning and Training Parameters**

The base YOLOv12s pre-trained model was fine-tuned on the dataset to leverage transfer learning and reduce training time. Key training parameters:

* Input size: 640x640 pixels
* Batch size: 8 (balanced between GPU memory constraints and training speed)
* Learning rate: 0.001 with cosine annealing scheduler
* Epochs: 50
* Optimizer: Adam
* Data augmentation: random horizontal flips, color jitter, and scaling to improve model generalization

The fine-tuning focused on optimizing the model weights to the specific domain of snowy and blurred road scenes.

## **5. Performance Evaluation**

The model was evaluated on the test set using standard metrics:

|  |  |
| --- | --- |
| **Metric** | **Value** |
| [mAP@0.5](mailto:mAP@0.5) | 0.872 |
| mAP@0.5:0.95 | 0.784 |
| Precision | 0.849 |
| Recall | 0.821 |
| F1-Score | 0.835 |

These results demonstrate the model’s ability to accurately detect objects while maintaining a good balance between precision (minimizing false positives) and recall (minimizing false negatives).

## **6. Visualization and Results**

**Inference Outputs:**  
 Inference was performed on the test images with batch size 8 to optimize memory usage and speed. The model generated bounding boxes and class labels which were saved for visual inspection.

**Sample Visualizations:**  
 The output images show detected objects with bounding boxes and confidence scores. The model effectively identifies cars and traffic lights under challenging weather and visibility conditions.

**Metrics Graphs:**  
 Plots of Precision-Recall curves and mAP over epochs confirm the model’s convergence and performance stability. These visualizations reinforce the quantitative results and provide insights into class-wise detection quality.

## **7. Conclusion**

This project successfully applied the SSD-YOLOv12 model for object detection in complex snowy road conditions. Fine-tuning the pre-trained model on the domain-specific dataset yielded high accuracy and robust inference speed. The choice of SSD-YOLOv12 was validated by its superior balance between detection speed and precision compared to other architectures. The project highlights the effectiveness of transfer learning and data augmentation in adapting object detection models to specialized tasks.