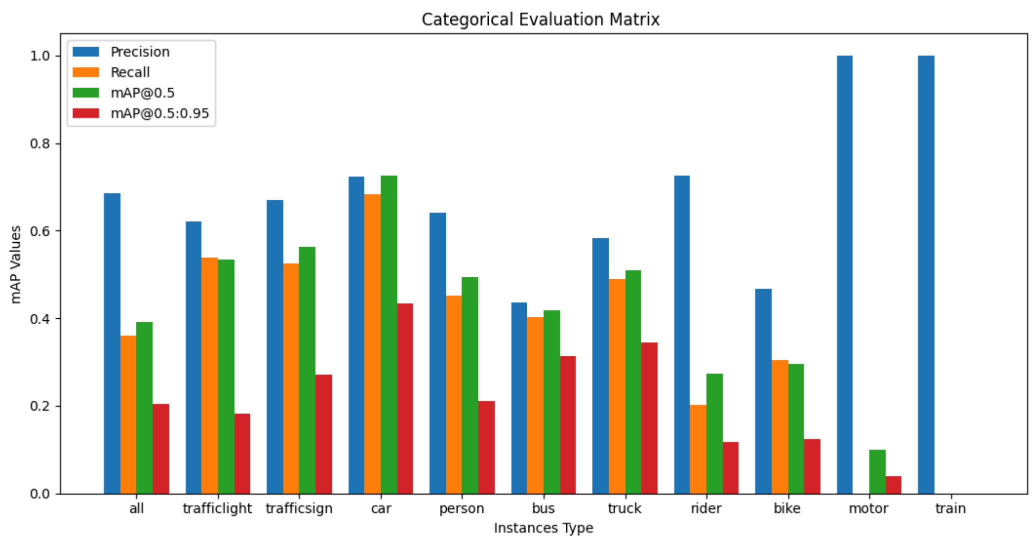
Modal Evaluation on Validation Dataset

The model training is evaluated based on below mentioned parameters:

a) Precision
b) Recall
c) mAP@0.5
d) mAP@0.5:0.95

Categorical Evaluation Matrix



Overall Evaluation:

- High precision indicating robust model classification
- Average recall value indicating problem with object detection in all image types.
- The mean average precision at an IoU threshold of 0.5 is around **0.6**, showing that the model can reliably localize objects but struggles at higher precision-recall tradeoffs.
- The steep drop (~0.4) in mAP reveals **localization challenge**—bounding boxes for detected objects are not tightly aligned with ground truth boxes indicates a potential issue in regression quality for bounding boxes.

Class-Specific Analysis

Traffic Light and Traffic Sign: Precision is moderate, but both recall and mAP values are relatively low. Small object detection is likely challenging for the model. Small instances per image also contribute lower performance values.

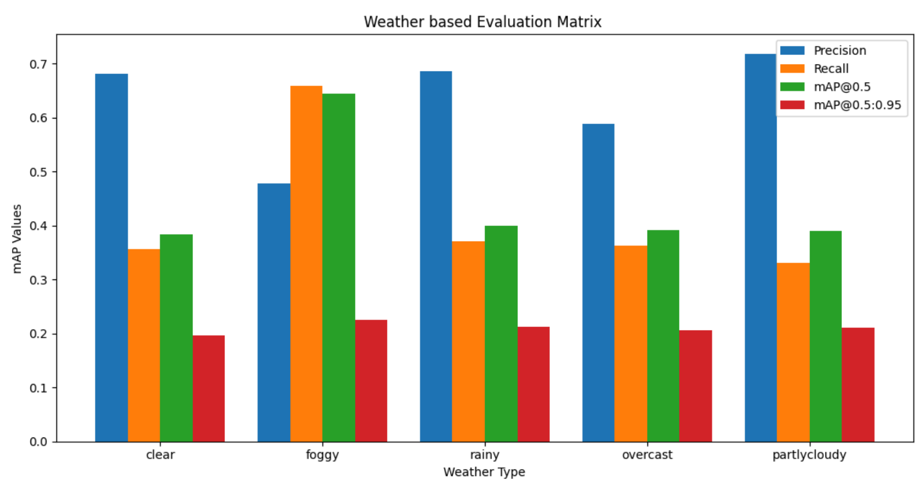
Car and Person: High performance across all metrics suggests the model is well-trained for these classes due to their higher number of instances in the training dataset.

Bus and Truck: Moderate recall and mAP values suggest the model struggles to detect these larger objects consistently. Larger vehicles may overlap with other objects (e.g., cars) or appear in varying perspectives, leading to detection challenges.

Bike and Motor: A significant drop in recall and mAP indicates difficulty in detecting these categories. Both are small and partially occluded in traffic scenes, making them harder to detect.

Train: The precision is perfect (~1.0) but recall and mAP are very low. This likely indicates very few predictions for this class. The very small instances in the training set have contributed to this worst performance.

Weather based Evaluation



This graph shows the performance of an object detection model under different weather conditions, analyzing precision, recall, and mAP metrics.

Class Specific Analysis

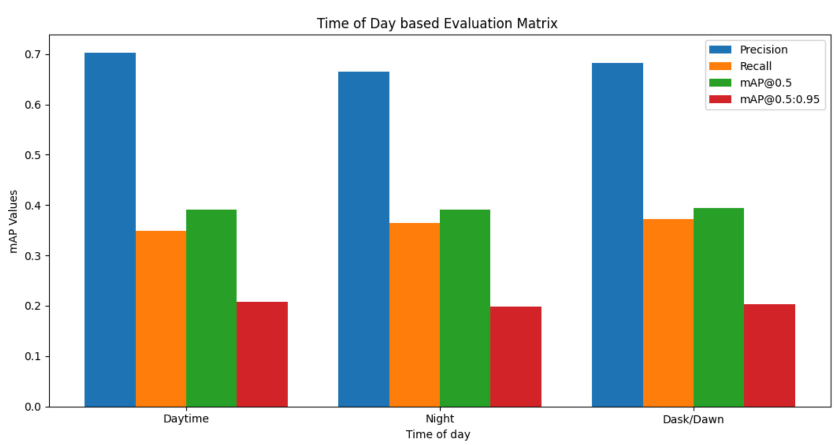
Clear Weather: The model performs well but could miss some objects, potentially due to dataset bias favoring clear weather. Localization errors are apparent and thus require tighter bounding box alignment.
Foggy Weather: Visibility issues in fog challenge the model's ability to detect smaller or occluded objects. Surprisingly the model shows balanced performance for fog as compared to clear weather, possibly due to robust features or training data diversity.

Rainy Weather: Reflections, motion blur, and water droplets on cameras likely degrade performance. The model is overconfident in a limited subset of detections.

Overcast Weather: Overcast weather creates diffused lighting, reducing contrast and detail in objects. Detection performance is relatively stable but not ideal.

Partly Cloudy Weather: Mixed lighting in partly cloudy weather creates inconsistent shadows, possibly confusing the model. Performance is better than fog or rain but still limited in recall.

Time of Day based Evaluation



This graph provides insights into how the performance of a system or model varies across different times of day: daytime, nighttime, and dusk/dawn. The metrics used for evaluation are Precision, Recall, Mean Average Precision (MAP), and MAP at different Intersection over Union (IoU) thresholds.

Class Specific Analysis

Daytime: Highest Precision, indicating that most detected objects are indeed correct. Lower Recall compared to night, suggesting some missed detections. Highest MAP, indicating overall better performance in terms of both precision and recall.

Night: Lower Precision compared to daytime, suggesting some false positives. Highest Recall, meaning the model detects a larger proportion of actual objects, even if some are false positives

Dusk Dawn: Lowest Precision, indicating a higher rate of incorrect detections. Lowest Recall, indicating a high rate of missed detections. Lowest MAP, indicating the most significant performance degradation.

Root Cause:

- **Lighting Conditions:** Nighttime and dusk/dawn present challenges due to lower light levels, which can affect object detection algorithms.
- **Environmental Factors:** Other factors like weather conditions, atmospheric haze, or background clutter can also impact performance, especially during dusk/dawn.
- **Data Imbalance:** If the training data is not well-balanced across different time of day conditions, the model may perform better.
- **Lower Training Iteration:** Lower number of epoch number has contributed to average performance of the model.

Recommendations:

- **Data Augmentation:** Increase the amount of training data for nighttime and dusk/dawn conditions, especially with images that simulate low-light scenarios.
- **Model Architecture:** Explore model architectures that are more robust to variations in lighting conditions, such as those using attention mechanisms or feature fusion techniques.

- **Higher Epoch Training:** Currently, the model is trained on 10 epoch for testing the model performance, but given higher epoch training model can outperform its current state.
- **Post-Processing:** Implement post-processing techniques like non-maximum suppression and confidence thresholding to improve the quality of detections, especially during challenging conditions.
- **Time-Aware Training:** Consider training separate models for different time periods or using time-aware loss functions to improve performance.
- Use anchor tuning or adaptive anchor boxes to account for larger objects.
- Conduct targeted error analysis to understand if occlusion or dataset imbalance is affecting performance.