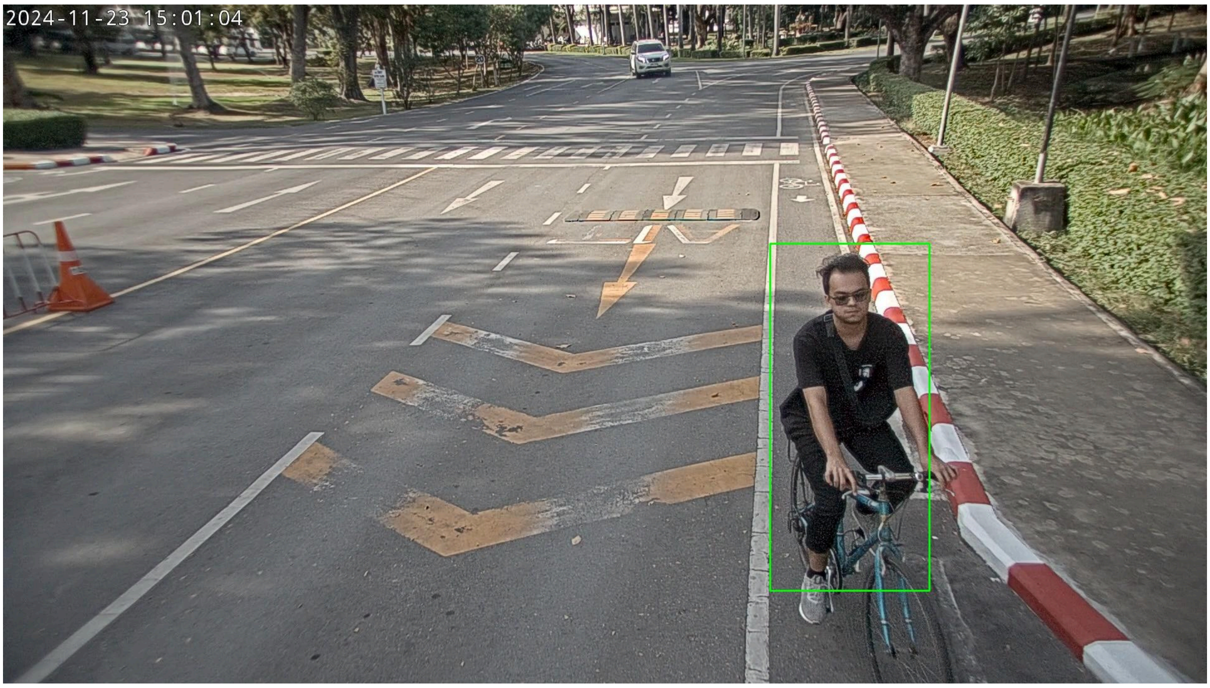
Report

Q1

In this experiment, I aimed to evaluate and compare the performance of YOLOv5, YOLOv8, and YOLOv11 models on an object detection task using a provided dataset containing 1,100 images. Initially, the dataset was utilized without addressing class imbalance, and each model was trained for 50 epochs with a batch size of 16. The primary metric for evaluation was the mean Average Precision (mAP) score, which guided the selection of the best-performing model among the three. After this phase, it became evident that class imbalance was a significant limitation in improving the models' overall performance. To partially address this issue, I augmented the dataset with additional images, increasing its size to 2,309 images. However, the augmentation was limited and did not completely resolve the class imbalance. Considering advancements in model architecture, I decided to discontinue experiments with YOLOv5, as it is generally considered outdated, and instead focused on training YOLOv8 and YOLOv11 models, which offer superior performance and efficiency.

For the expanded dataset, I trained YOLOv8 and YOLOv11 models in two variations each, with batch sizes of 16 and 32, over 1,000 epochs. These experiments were computationally intensive, fully utilizing an NVIDIA GPU with 11GB of VRAM. In parallel, recognizing the importance of a diverse and balanced dataset, I conducted additional data collection specifically for the "person" class using cameras located at the gates of the Asian Institute of Technology (AIT). This effort resulted in a dataset of 2,933 labeled person images, which will be shared during the final submission. These newly collected images were labeled carefully to ensure high-quality annotations, aiming to improve the model's accuracy for this critical class.

Example of collected data:



The results of these experiments led to the identification of two best-performing models: q1_balanced_yolov11_b32_epch1000_02, trained on the partially balanced dataset, and q1_yolov8_best_res_b16_epch1000. The first model benefits from the inclusion of the additional person class data and the larger batch size, which allows for more stable gradient updates. The second model, trained with a smaller batch size, shows robust generalization. Both models demonstrate competitive mAP scores and precision across key classes. During the final evaluation, I will test these models extensively on unseen data and select the one that performs best for submission. This comprehensive approach highlights the iterative process of model improvement and the critical role of dataset balancing and augmentation in achieving optimal results.

Q2

In this experiment, the primary goal was to develop and evaluate classification models capable of accurately predicting ten distinct vehicle types using YOLOv8n-cls and YOLOv11n-cls architectures. The experiment began by training these models on a baseline dataset without performing any additional data augmentation or addressing potential class imbalances. This baseline dataset consisted of images with limited class diversity and served as the foundation for the initial phase of model evaluation. Each model was trained using two different configurations: batch sizes of 16 and 32, over 200 epochs. These configurations were chosen to test the model's learning capacity under varying conditions while ensuring optimal utilization of available computational resources. After training, the models were evaluated based on their Macro F1 scores, a metric specifically chosen for its ability to fairly assess performance across all classes, especially in imbalanced datasets. This step was critical in determining the strengths and weaknesses of each model architecture on the baseline dataset.

Following the initial evaluation, it became clear that additional data was necessary to improve class diversity and address the potential underrepresentation of certain vehicle types. To enhance the dataset, public datasets from Roboflow were integrated into the training pipeline. These included Samlor-Tuktuk-Vehicles, Detect_Dum_Em, and Vehicle_Car. These datasets introduced additional samples of various vehicle types, contributing to a more balanced and representative training dataset. The inclusion of these datasets was carefully considered to ensure that the model could generalize well to unseen data while maintaining robust performance across all classes.

With the augmented dataset, the models were retrained using the same experimental configurations of batch sizes 16 and 32, and an epoch count of 200. This ensured a fair comparison between the baseline results and the results achieved with the enriched dataset. Throughout the training process, the Macro F1 score remained the primary evaluation metric, aligning with the requirements for final submission. The retrained models demonstrated improved performance, particularly in handling previously underrepresented vehicle types, validating the importance of data augmentation and diversification in achieving robust classification.

At the end of the experiment, I identified three models as the most promising candidates for final evaluation. These models were selected based on their Macro F1 scores during training and validation. On submission day, I will evaluate these three models on the provided unseen test dataset and select the best-performing model based on its results. This step ensures that the final submission reflects the most robust and well-performing model for live inference. Overall, this systematic approach, starting with a baseline evaluation and progressing to dataset enhancement, retraining, and final selection, underscores the importance of iterative experimentation, careful dataset curation, and metric-driven evaluation in achieving optimal results.

Q3

In this experiment, the primary objective was to enhance an existing dataset by introducing a new class, "dog" (in my case), and subsequently train a YOLOv11n detection model capable of detecting three distinct classes: "person," "box," and "dog." The process began by leveraging a pre-trained YOLOv8n model to detect dog instances across the provided dataset. This model was chosen for its lightweight architecture, which offers a balance between speed and accuracy, making it well-suited for this preliminary detection task. After detecting the locations of dogs within the dataset, the results were integrated into the dataset's existing annotation files. Each detected dog instance was appended to the corresponding annotation file in YOLO format, ensuring compatibility with subsequent training pipelines.

Once the dataset was enriched to include dog annotations, the next phase involved training a YOLOv11n detection model. This model was trained to detect three classes: "person," "box," and "dog." The training was conducted on the updated dataset to ensure that the model could accurately learn and generalize across all three classes. Specific attention was paid to preserving the original dataset's annotations while seamlessly integrating the new dog annotations. The training process utilized a batch size and epoch count carefully selected to balance computational efficiency with model performance, considering the available hardware resources.

The YOLOv11n model, being a newer architecture, was chosen for its capability to handle larger datasets and provide improved accuracy over its predecessors while maintaining computational efficiency. By completing this process, the experiment successfully created a robust detection model capable of identifying all three target classes. The resulting model represents a significant enhancement over the initial dataset and model configuration, aligning well with the requirements for practical deployment and evaluation. This iterative process of dataset augmentation and model training underscores the importance of combining automated detection with manual validation to achieve high-quality results in object detection tasks.