1. **Data Annotation and Dataset Preparation**

**Introduction:** The goal of this task was to create a high-quality dataset for training an object detection model. This involved collecting images, annotating objects of interest, and organizing the dataset for training and validation.

**Data Collection:** To ensure a diverse dataset, a video of a road with passing vehicles and pedestrians was used as the primary source. The video provided dynamic scenes with varying object positions and orientations.. A total of 160 screenshots were manually extracted from the video to serve as individual image samples. The video link: [Object detection using deep learning dataset cctv road video](https://www.youtube.com/watch?v=cDuVtH0CZek&t=1136s&ab_channel=BrainyNeurals)

**Annotation Process:** The annotation of the dataset was performed using **CVAT** (Computer Vision Annotation Tool). The following steps were taken:

* **Objects Annotated:** Three object classes were labeled – Bike, Car, and Pedestrian. Each image contained an average of 3 to 4 objects.
* **Annotation Type**: Rectangular bounding boxes were used to enclose the objects.
* **Annotation Format:** The annotated dataset was exported in YOLO format. The YOLO format consists of text files where each annotation contains the class label and normalized bounding box coordinates relative to the image dimensions.

**Dataset Splitting:** To prepare the dataset for training, it was divided into training (80%) and validation (20%) sets.

1. **Model Training**

**Introduction:** The pretrained YOLOv8n has been chosen for this task. YOLOv8n is a lightweight and efficient object detection model optimized for real-time applications. It builds upon previous YOLO versions. This model is well-suited for tasks requiring fast inference, such as real-time video processing.

**Choice of YOLOv8n:**

* YOLO models are widely recognized for real-time object detection, and YOLOv8n balances speed and accuracy well.
* **Benefitting from Dynamic Augmentation**: No manual preprocessing is required as it's handled by the model, and more importantly, we can also benefit from the automatic augmentation of the model as it can perform augmentation on the fly during training.
* **Lightweight for Efficient Inference**: YOLOv8n’s small size allows for faster inference, making it ideal for our application.

### **Evaluation Metrics:** The model was evaluated using the built-in YOLO metrics:

* **Precision**: Measures the percentage of correctly identified objects among detections.
* **Recall**: Measures how well the model detects all relevant objects (fewer missed detections).
* **mAP@50**: Evaluates detection accuracy at an IoU threshold of 0.5.
* **mAP@50-95**: Measures accuracy across multiple IoU thresholds, providing a more comprehensive evaluation.

**Results on Validation Set:**

| **Class** | **Images** | **Instances** | **Precision** | **Recall** | **mAP@50** | **mAP@50-95** |
| --- | --- | --- | --- | --- | --- | --- |
| All | 32 | 156 | 0.95 | 0.994 | 0.981 | 0.696 |
| Bike | 30 | 108 | 0.973 | 0.983 | 0.988 | 0.641 |
| Car | 29 | 36 | 0.968 | 1 | 0.992 | 0.782 |
| Pedestrian | 7 | 12 | 0.909 | 1 | 0.963 | 0.667 |

1. **Literature Review**

**Video Seal: Open and Efficient Video Watermarking**

<https://ai.meta.com/research/publications/video-seal-open-and-efficient-video-watermarking/>

With the rapid growth of AI-generated content and advanced video editing tools, ensuring the authenticity and traceability of digital videos has become a significant challenge. Traditional watermarking techniques, often designed for static images or low-resolution videos, struggle to keep up with the demands of modern digital platforms, especially when faced with common video transformations like compression, cropping, and re-encoding. To address these challenges, this paper introduces VideoSeal, an open-source, neural network-based video watermarking framework designed for efficiency, robustness, and scalability. By leveraging innovative techniques such as temporal watermark propagation and multistage training, VideoSeal offers a powerful solution to embed imperceptible watermarks that withstand various video manipulations. This framework not only sets a new benchmark for watermarking performance but also fosters transparency and reproducibility in the field through its open-source code and public demo (<https://aidemos.meta.com/videoseal>).

**Key Contributions:**

* **VideoSeal Framework**: Introduces an open-source, efficient, and robust neural network-based video watermarking system.
* **Temporal Watermark Propagation:** A novel technique that applies watermarks efficiently without embedding them in every video frame, saving time and computing power.
* **Open Source Tools:** Provides code, models, and a demo to encourage further research.

**Methodologies:**

* **Embedder-Extractor Model:** Trains both components together, ensuring the watermark remains hidden but can still be detected after video modifications.
* **Resilience through Augmentations:** The model is tested with video distortions like compression, cropping, and color changes to ensure robustness.
* **Efficiency Techniques:** Uses downscaling and watermark propagation to reduce computational costs while maintaining high detection accuracy.

**Potential Applications:**

* **AI-Generated Content Verification:** Helps identify AI-generated videos, combating misinformation.
* **Piracy Control:** Assists in tracking unauthorized distribution of digital content.
* **Regulatory Compliance:** Aligns with emerging regulations requiring watermarks in AI-generated media.