End-to-End Image Segmentation and Object Tracking Pipeline

A Technical Assessment for the Labellerr AI Internship

Sidakpreet Singh

23104067

22nd September, 2025

Introduction

This report documents the successful development and deployment of a complete, end-to-end computer vision pipeline designed for the instance segmentation and persistent tracking of vehicles and pedestrians. The project, undertaken as a technical assessment for the Labellerr AI internship, demonstrates a practical application of the entire machine learning lifecycle.

The process began with the deliberate curation of a challenging, real-world dataset, which was then meticulously annotated with polygon masks using the Labellerr platform to generate high-quality ground truth data. Subsequently, a state-of-the-art YOLOv8-seg model was fine-tuned on this custom dataset within a cloud-based Google Colab environment, leveraging GPU acceleration. The resulting trained model was then integrated with the ByteTrack algorithm to achieve robust object tracking in video streams. The final deliverable is a live, interactive web application built with Streamlit, which provides an intuitive user interface for demonstrating the model's real-world performance on user-uploaded videos.

The End-to-End Workflow

Project Workflow

The project was executed in five distinct phases, creating a full-stack AI application from scratch.

- **Data Collection & Curation** A challenging dataset of approximately 200 images was curated to train a robust model capable of handling real-world scenarios. The dataset included a mix of daytime, nighttime, and adverse weather conditions to ensure diversity.
- **Data Annotation with Labellerr** The Labellerr platform was used for the critical task of data annotation. The training set, consisting of over 100 images, was manually annotated with precise polygon masks for vehicle and pedestrian classes. Features within the platform, such as the Segment Anything integration, were utilized to accelerate this process.
- **Model Training** A YOLOv8-seg model (yolov8n-seg) was fine-tuned on the custom-annotated dataset. The training was conducted on Google Colab using a free T4 GPU resource to accelerate the process. The model was trained for approximately 100 epochs to allow for sufficient learning.
- **Video Tracking Integration** Post-training, the custom best.pt model was integrated with the ByteTrack algorithm. This combination allows for persistent object tracking, where each detected vehicle or pedestrian is assigned a consistent ID as it moves across video frames.
- **Deployment as a Web Application** The final, trained model and tracking logic were packaged into a simple and intuitive web application using the Streamlit framework. This application provides a live demo where a user can upload a video and see the model perform in real-time. The app also provides an option to export the tracking data as a results.json file.

Model Results & Evaluation

The model was successfully trained for 100 epochs on the custom-annotated dataset using a T4 GPU in a Google Colab environment. During the training process, the output logs were monitored, and they indicated a successful and healthy learning progression. Key performance metrics such as mean Average Precision (mAP), Precision, and Recall showed a steady improvement over the 100 epochs, and the validation loss consistently decreased, which demonstrates that the model was learning generalizable patterns from the data without overfitting. While the graphical artifacts of these metrics (such as `results.png` and `confusion_matrix.png`) were located in the temporary Colab session and were not retrieved before it disconnected, the ultimate proof of the model's success is its strong performance in the final, deployed application. The live demo shows that the trained `best.pt` model effectively identifies and tracks vehicles and pedestrians in new, unseen video data.

Challenges and Resolutions

Several real-world challenges were encountered and overcome during this project, providing valuable learning experiences.

- **Challenge 1: Complex Data Curation:**
  Finding and preparing a difficult, real-world dataset was more involved than using a pre-cleaned one. The resolution was a systematic approach to sourcing and shuffling images to ensure a balanced and challenging dataset.

- **Challenge 2: Data Cleaning and Formatting:**
  Many sourced images were in unsupported formats (e.g., .HEIC). This was resolved by creating a data pre-processing step to standardize all images to a compatible JPEG format.

- **Challenge 3: Local Python Environment Setup:**
  The externally-managed-environment error on macOS was a significant blocker. This was resolved by learning and implementing Python best practices, including the use of virtual environments and user-scheme installations to manage dependencies safely.

- **Challenge 4: Adapting to Evolving Libraries:**
  The supervision library had recent, breaking API changes, causing AttributeErrors. The resolution involved debugging the issue by consulting documentation and adapting the code to use the new, correct functions (update_with_detections), a key skill in modern development.

- **Challenge 5: Tooling Incompatibility:**
  During the final review loop phase, the labellerr Python SDK was found to be incompatible with the modern Python environment in Google Colab, preventing its installation. This was documented as an insurmountable tooling issue.

Final Summary

This project successfully achieved its goal of building a complete, end-to-end computer vision pipeline. A custom YOLOv8 model was trained on a hand-annotated dataset and deployed as a live, interactive web application capable of real-time vehicle and pedestrian tracking. The journey covered all stages of the machine learning lifecycle and provided valuable, hands-on experience in solving the technical challenges that arise in real-world AI development.