**Estimating the Speed of a Moving Object in Video Statement**

Lu Xuanchen 22257896

1. Project Background

With the rapid development of computer vision and deep learning technologies, analyzing video data to extract meaningful information has become increasingly feasible and valuable in real-world applications. One of these applications is estimating the speed of moving objects, which is critical in areas such as traffic monitoring, autonomous driving, sports analysis, etc.

Traditionally, speed estimation is performed using physical sensors such as radar or lidar, but these devices are expensive and have certain limitations. In contrast, video-based speed estimation provides a low-cost, scalable, and convenient solution. However, implementing an accurate video-based system remains a technical challenge due to factors such as object detection accuracy, tracking stability, and the difficulty of converting pixel distances into actual measurements.

This project aims to develop a computer vision system that estimates the speed of moving objects from video clips using deep learning and tracking techniques.

1. Problem / Improvement Areas

There are several key challenges in the field of video-based velocity estimation:

1.Object detection accuracy: objects moving at high speeds may cause tracking to be lost

2.Object tracking consistency: the identity of detected objects needs to be consistent across multiple frames.

3.Multiple object detection: if there are multiple moving objects in the video, they need to be detected together

4.Spatial calibration: convert pixel-based movement to real-world distance. Specific reference objects may be required for judgment.

5.Real-time performance: the system needs to process video frames efficiently to achieve real-time or near-real-time performance.

6.Environmental robustness: the system should be able to cope with different lighting conditions, occlusions, and object size and relative spatial changes due to viewing angles.

1. Proposed Approach

This project proposes a deep learning-based pipeline for estimating the actual speed of moving objects in video clips. The pipeline first uses a YOLOv5 or YOLOv8 model to find moving objects in the video, such as vehicles, pedestrians, or badminton balls, by finding the offset of the object's position in each video frame. These detected objects are then passed to the Deep SORT tracking algorithm, which assigns each object a consistent ID across frames, allowing accurate tracking of their position over time.

To address the perspective distortion and spatial scaling issues caused by the camera, the project integrates a pre-trained monocular depth estimation model MiDaS. MiDaS will generate a depth map for each frame, providing a relative depth value for each detected object.

By combining the 2D position changes from Deep SORT and the depth values ​​from MiDaS, the system estimates the motion of each object in 3D space. The system then calculates the actual speed of each object in meters per second (m/s) based on the frame rate of the video. The final output will include annotated video frames with bounding boxes, object IDs, and speed values, as well as structured data for analysis.