

ABSTRACT

YOLO (You Only Look Once) is a deep learning-based object detection algorithm capable of recognizing objects in images and videos with high accuracy and efficiency. It is widely popular in the field of computer vision due to its fast performance and accurate detection results. However, YOLO has certain limitations, especially when applied to sequential data such as video. A common issue is the instability of detection results between frames. This occurs because YOLO performs detection independently on each frame without considering temporal relationships or the detection history from previous frames.

Ideally, a video detection system should produce consistent and stable outputs over time, especially for real-time applications where sudden changes in detection can reduce the system's reliability. Since YOLO does not take previous frame results into account, each detection is conducted independently per frame. As a result, the detection output tends to be unstable and frequently exhibits flickering, particularly when objects are partially occluded, blurred due to fast movement, or moving dynamically within the frame. To address this challenge, this study implements a combination of Kalman Filtering and Polling (majority voting) as an additional post-processing module on top of the YOLO inference pipeline. Several previous studies have integrated YOLO with Kalman Filtering for object tracking purposes, such as vehicle speed estimation (YOLO + Kalman Filtering), tea flower counting (YOLOv5 + Kalman Filtering + Hungarian Algorithm), and soccer player tracking (YOLOv8 + Kalman Filtering + Hungarian Algorithm). However, these approaches focus primarily on tracking rather than improving detection stability.

This research takes a different approach by focusing specifically on enhancing detection stability, not tracking. To support this objective, a custom YOLO model was developed and trained to detect three types of animals: dogs, cats, and bears. These animals were selected due to their similar visual appearance and body structure, which often pose challenges in distinguishing between objects with comparable visual characteristics—thus increasing the likelihood of flickering during detection. By integrating Kalman Filtering and Polling, the YOLO detection results become more stable and resistant to short-term uncertainty, producing smoother and more reliable outputs in video streams. Based on experimental results, it was found that Flickering was reduced after incorporating polling, FPS dropped by 18.66% due to the additional overhead from kalman filtering and polling.

Keywords: YOLO, Custom Model, Object Detection, Flickering, Polling, Kalman Filtering, Post-processing, Real-time Video, Detection Stability, Euclidean Distance