Brief Write-Up: Estimate crowd size on bridge

Approach:  
I was aware of YOLO in object detection so thought of using it for detecting the objects (person) in the video. The crowd in the video is too dense challenging it to detect the person presizely. Tried my best in tuning the model provided by YOLO. I tried to to increase the confidence threshold but it failed to detect the people as it is a drone shot making it complicated to the model. I tried to balance the confidence threshold. At start it was detecting many false positives but in later half it started detecting with increased accuracy. I also tried to adjust Iou threshold but as it is a drone shot there are overlapping frames, challenging the model.

I tried using R-CNN as well but it did not work for me

Thought process:

Each frame is converted from BGR to RGB, transposed, and normalized to match the input requirements of YOLOv5.

The model performs inference to detect objects. The non\_max\_suppression function filters out overlapping bounding boxes and keeps only those with high confidence scores.

Bounding boxes around detected people are drawn on the frame, and the number of detected people is counted based on class labels.

Confidence Threshold (conf\_thres): Set to 0.05 to ensure that even low-confidence detections are considered, capturing more potential people in the crowded scene.

IoU Threshold (iou\_thres): Set to 0.6 to manage overlapping detections effectively, reducing false positives.

Alternative Solutions Considered:

Different YOLO Models: I considered larger versions of YOLOv5 (e.g., YOLOv5m or YOLOv5l) for potentially better accuracy but did not use due to higher computational requirements.

Feedback:

I would like to learn how we can fine tune the models to detect such dense objects. It was an interesting assignment got to learn a lot.