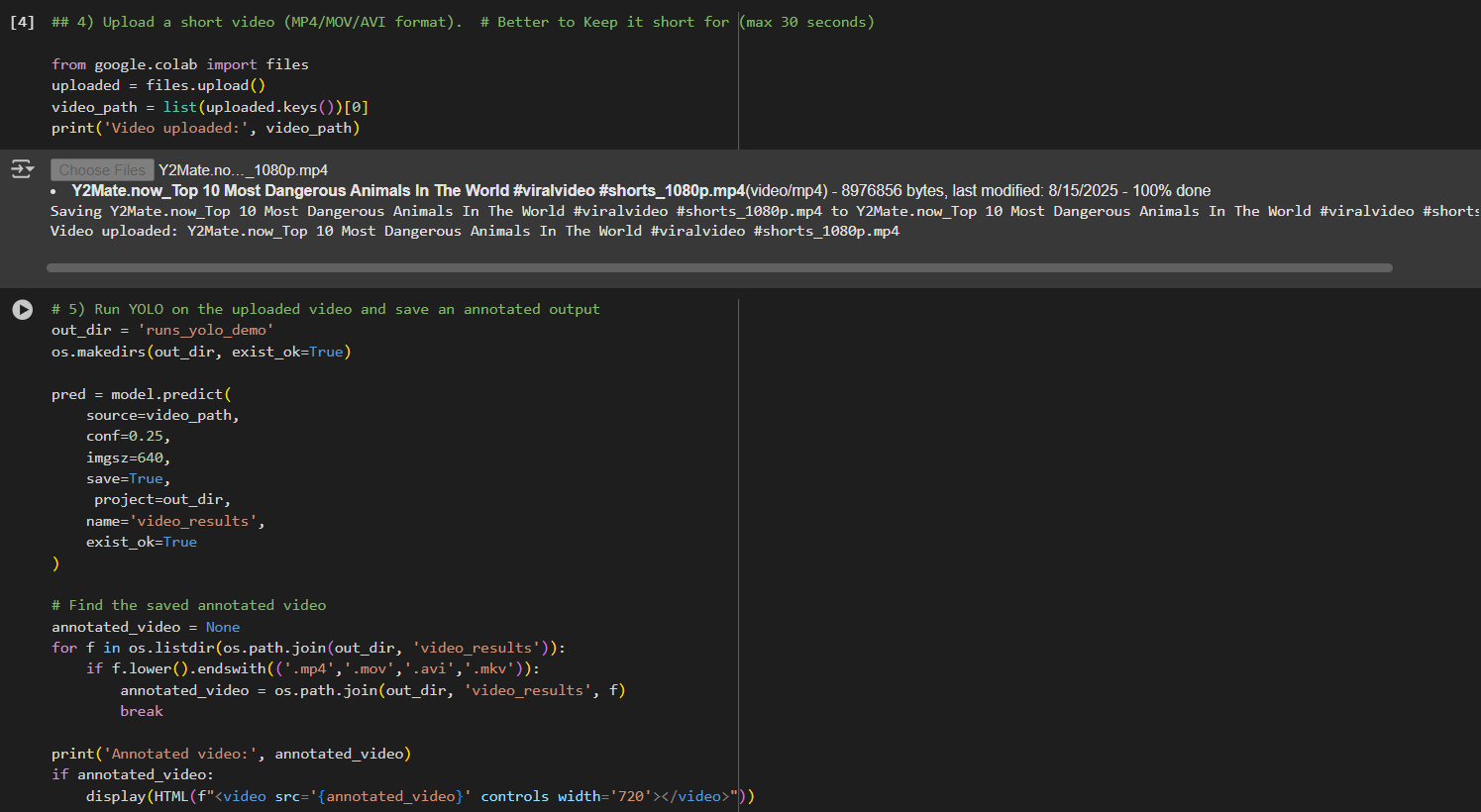
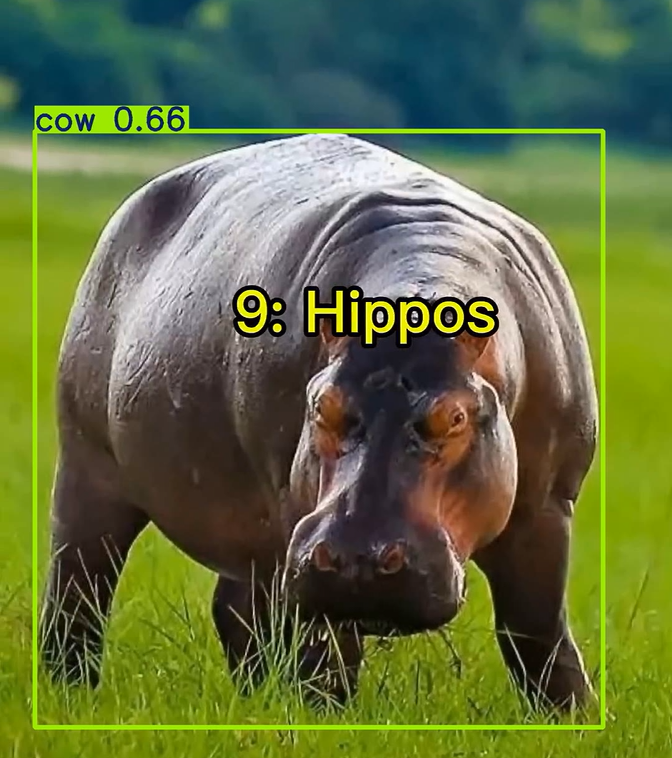
Deep Learning Video Assignment

1) Upload a video of animals?

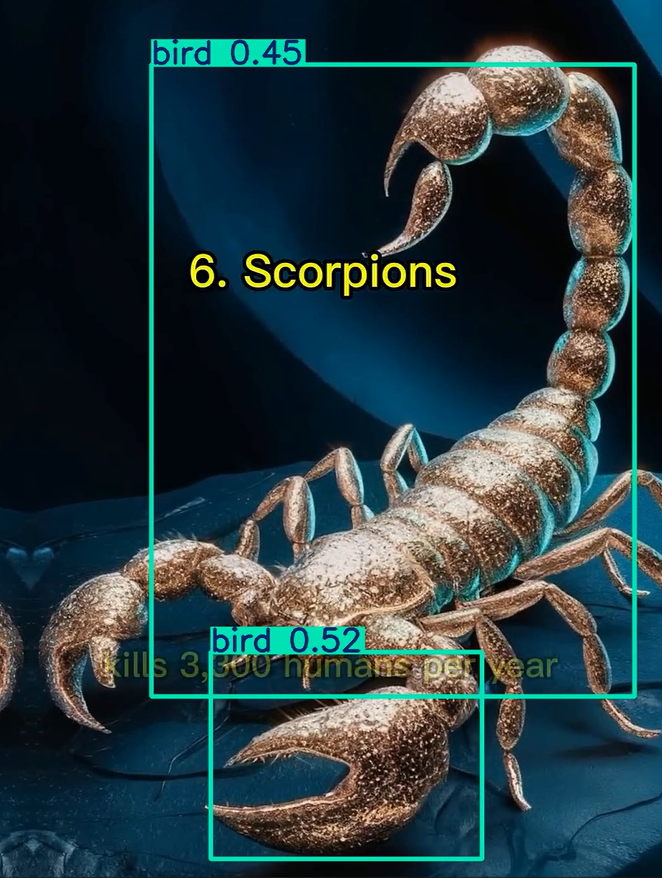
Code:-



Output:-

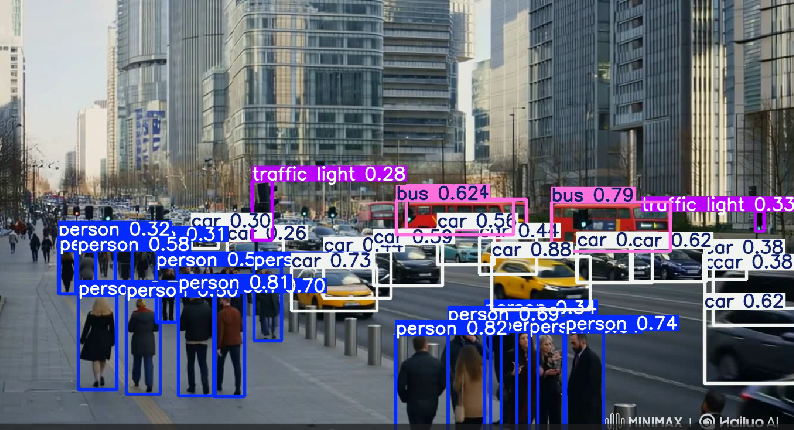


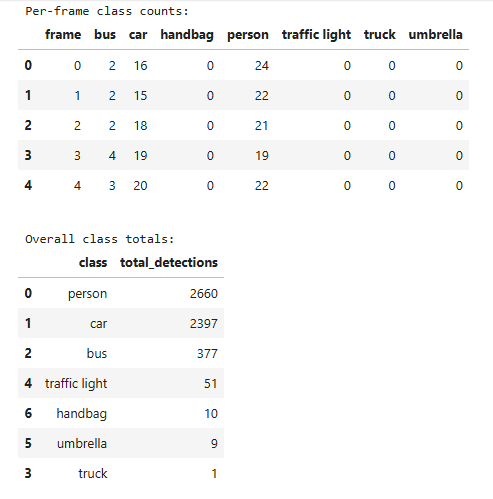




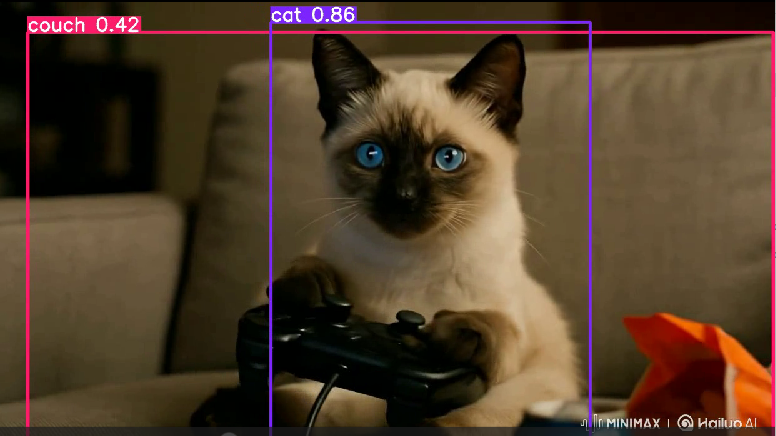
Conclusion:  
The YOLO model was successfully executed for deep learning–based video classification of animals. The model detected and classified multiple objects with confidence scores, such as elephants (0.86), humans (0.89), hippos (misclassified as cow, 0.66), scorpions (0.45–0.52), and dogs (0.91). While the detections demonstrate the model’s capability to identify diverse objects within video frames, there are also signs of misclassification (e.g., hippo detected as cow, scorpion labelled as bird in some bounding boxes), indicating potential areas for model fine-tuning or dataset enhancement. Overall, the experiment confirms YOLO’s effectiveness in real-time object detection but highlights the need for improved accuracy in distinguishing similar or less-represented classes.

1. Will it detect cartoons?
2. Will it detect AI generated videos?

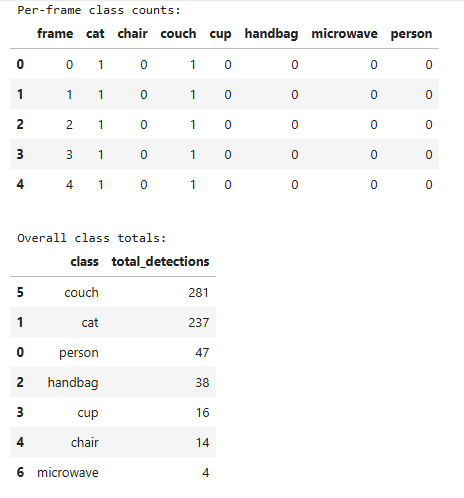




Conclusion:

The YOLO model was successfully applied for AI-based video object detection and tracking. Across the analyzed frames, the model consistently detected high-frequency objects such as persons (2,660 detections) and cars (2,397), indicating strong performance for prominent and recurring classes. Buses (377 detections) and traffic lights (51) were also reliably identified, while lower-frequency classes such as handbags (10), umbrellas (9), and trucks (1) were detected infrequently likely due to their limited presence in the video or reduced detection confidence. The per-frame counts demonstrate stable recognition patterns for frequently occurring traffic-related objects, validating YOLO’s capability for urban scene understanding in real time, with potential for further fine-tuning to improve detection of rare or partially occluded objects.

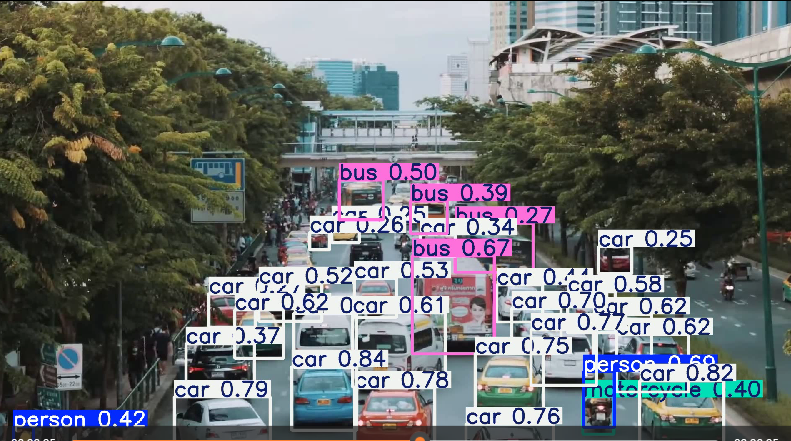


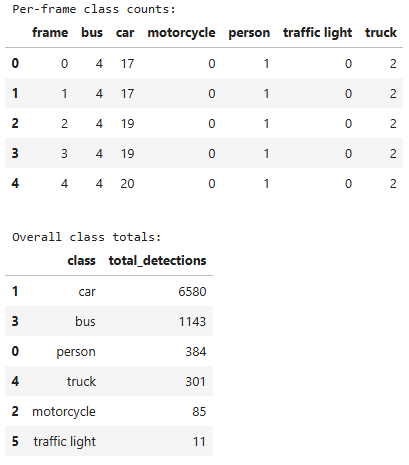


Conclusion:

The YOLO model was successfully applied for AI-based video object detection. Across video frames, the model consistently detected and classified multiple objects such as couches (281 detections), cats (237), persons (47), handbags (38), cups (16), chairs (14), and microwaves (4). The per-frame counts indicate stable detection patterns for recurring objects like cats and couches, demonstrating the model’s strong ability to track stationary and frequently appearing items. However, the lower counts for certain classes (e.g., microwaves) suggest either their rarity in the video or lower detection confidence. Overall, the results confirm YOLO’s effectiveness in real-time object detection and tracking, with reliable performance for prominent classes and scope for further tuning to enhance detection of less frequent objects.

1. What if we have objects of similar classes? (Not one of the 80 classes)

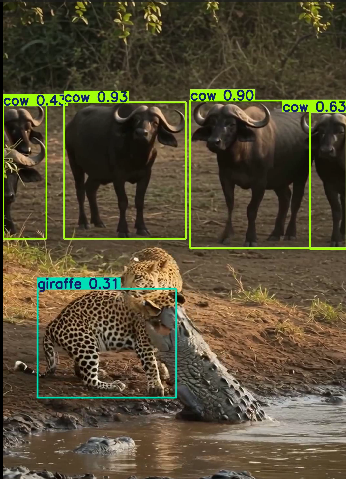


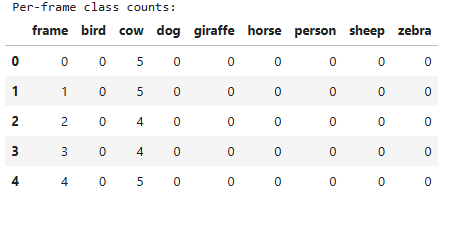


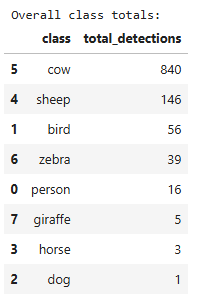
Conclusion:

The YOLO model was applied for same class video object detection in a traffic scene. While the model achieved high detection counts for cars (6,580) and buses (1,143), along with moderate counts for persons (384), motorcycles (85), and traffic lights (11), it also recorded 301 detections for trucks. On reviewing the per-frame counts, it appears that in some frames, cars and buses were misclassified as trucks. This suggests potential confusion between visually similar large vehicles, likely due to overlapping object features, partial occlusion, or limitations in the training dataset’s representation of distinct vehicle classes. Although YOLO performed well in identifying high-frequency traffic objects, fine-tuning the model with more diverse and class-specific vehicle images could reduce such misclassifications and improve precision for closely related categories.

1. What if we have objects other than 80 classes?









The provided frames reveal significant misclassifications by the detection model when identifying wildlife. In both examples, African buffalo are consistently labeled as “cow” with high confidence scores (~0.90), likely due to visual similarity in body shape and color to domestic cattle. The leopard is misclassified differently across frames in the first as a “giraffe” (0.31 confidence) and in the second as a “zebra” (0.59) and “bird” (0.29). These errors suggest that the model struggles with species that are underrepresented or visually complex in the training set, especially in scenes with multiple animals, occlusion, or uncommon interactions (such as the leopard biting a crocodile).

This indicates a need for:

* **Dataset enrichment** with more diverse wildlife images under varied lighting, angles, and interactions.
* **Fine-tuning** on domain-specific animal datasets to reduce confusion between similar or unfamiliar species.
* **Post-processing verification** using contextual cues (e.g., habitat, co-occurring species) to filter unlikely classifications.

The YOLO model was applied to a real wildlife video for object detection. Across frames, the model consistently labeled African buffalo as “cow,” resulting in 840 total detections for the class. Several other classes such as sheep (146), bird (56), zebra (39), giraffe (5), horse (3), and dog (1) appeared in the detection totals despite those animals not being present, indicating misclassifications. These errors are likely due to the model’s reliance on visually similar domestic animal categories when encountering less-represented wild species, as well as the effects of background features and partial occlusion. While the model demonstrated reliable detection of large quadrupeds, its species-specific accuracy was limited, suggesting the need for fine-tuning with domain-specific wildlife datasets to improve performance in natural environments.