Solution Document

For the solution in this repository, I encapsulated everything into a single Python script capable of addressing both tasks by accepting input commands. I created two functions to perform inference. The first function, named `video_detection`, is for carrying out inference on videos, while the second function, named `image_detection`, is for inference on images.

This separation was implemented because the inference process differs between videos and images. Video processing necessitates looping through each frame until the video ends, whereas image processing requires only a single inference pass.

1. To address the inference questions regarding trucks and ambulances, I utilized an if-else selection to determine which task to perform. The first menu option is for truck inference, using a model previously trained and tested with prepared test footage. The available footage includes:
    a. Video of a Truck
    b. Video of an Ambulance
    c. Image of a Truck
    d. Image of an Ambulance

    Through testing against these four types of footage, the model successfully detects both trucks and ambulances without any misidentification of objects.

2. To address the modification question to only detect ambulances, I continued to utilize an if-else selection to determine the task to be performed. The second menu option is for ambulance inference, utilizing the pre-trained model as before, but with additional filtering to detect only objects classified as ambulances. This filtering is achieved by specifying the class label during the detection process.

    The model is then tested against the same set of prepared test footage, which includes:
    a. Video of a Truck
    b. Video of an Ambulance
    c. Image of a Truck
    d. Image of an Ambulance

    During testing, the model successfully detects ambulances in the ambulance footage, while no detection results are displayed for the truck footage due to the earlier filtering process.

1. How you would calculate accuracy of the truck detection.

   Accuracy can calculate by see the result matrix who provided by yolov8 when after train the data and can use the validate mode by yolov8. However accuracy can calculate manually by testing to othere footage like I mention before. When inference to truck footage is it the model can detect accurate and it can call as True Positive, when inference the Truck to ambulance footage is it the model detect as ambulance and it call as True Negative. Then we can calculate using the confusion matric about the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). After that we can calculate accuracy by formula :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. How you would calculate accuracy of the ambulance detection

   Accuracy can calculate by see the result matrix who provided by yolov8 when after train the data and can use the validate mode by yolov8. However accuracy can calculate manually by testing to othere footage like I mention before. When inference to Truck footage is it the model detect the Truck and if cant detect it call True Negative. When inference to Ambulance footage is it the model can detetct the ambulance and if can detect it call True Positive. Then can calculate by formula that i answer in question 1.

3. How you would calculate the speed of inference.

   Speed can be calculated by measuring the inference time per image or per frame. Then, the average can be calculated from each frame. As for YOLOv8, it displays the speed time on the terminal when performing inference.

4. How you would improve the speed of inference if asked.

   It can be achieved by optimizing the model, utilizing more compatible hardware, implementing batch processing, tolerating detection results by using lower precision, and simplifying preprocessing and postprocessing.