Models Used in Object-Based Video Summarization Project

1. YOLOv8 (You Only Look Once, Version 8)

- **Source**: Acquired from the Ultralytics GitHub repository.
- **Model Type**: Pre-trained model.
- **Training Dataset**: Trained on the COCO (Common Objects in Context) dataset.
- **Object Categories**: Contains 80 object categories.
- **Detection Capability**: Capable of detecting a wide range of everyday objects.
- **Purpose**: YOLOv8 was used to perform object detection in the input image.
- **Functionality**: It detects objects and provides bounding boxes with class labels.
- **Outcome**: This functionality allows the system to generate a list of objects that users can select for video summarization.

2. SAM2 (Segment Anything Model 2)

- **Source**: Sourced from Hugging Face's model hub as a pre-trained model developed by Meta.
- **Purpose**: Employed to identify frames in the video containing the object selected by the user from YOLOv8's detected list.
- **Functionality**: Uses a prompt-based segmentation approach to extract precise object masks across video frames.
- **Outcome**: Helps in selecting relevant frames for the final summary.
- **Characteristics**: Excels in generating high-quality masks for user-specified prompts and effectively handles complex backgrounds and overlapping objects for accurate frame selection.