README: Object Based Video Summarization with SAM2 and YOLO-v8

Overview

This readme provides instructions on setting up and running a object based video summarization pipeline using the SAM2 video predictor and YOLO-v8 object detection for video summarization. The script extracts frames from a video, performs object detection, annotates the frames, and extracts the annotated frames and then generates a final summary video based on the user selected objects.

Prerequisites

Make sure you have the following installed:

- Python 3.6+
- Jupyter Notebook or any environment supporting IPython (preferably Google Colab)

Setup Instructions

Install the required Python libraries:
Run the following commands to install the necessary libraries:

```
!pip install huggingface_hub
!pip install ultralytics
!pip install opencv-python Pillow ipywidgets
!pip install sam2
!pip install -q supervision[assets] jupyter_bbox_widget
```

Set Up Paths:
Update the `SOURCE_VIDEO` and `image_path` variables with the paths of your video and image files:

```
SOURCE_VIDEO = "/path/to/your/demo.mp4"
image_path = '/path/to/your/sample_pics.png'
```

Extract Frames from the Video:
Configure `START_IDX`, and `END_IDX` to specify the range of frames and the scaling factor for frame extraction.

Model Evaluation:

1. **Select the objects to be summarized**
   Select atleast two objects for which video need to be summarised.

2. **Submit the choice**
   After selection submit the objects and then run the final model to get the summarised output.

3. **Output**
   Final output are saved in the root directory.

To run Streamlit UI

Prerequisites:

Make sure you have the following installed:

- Python 3.9+
- Vs Code
- GPU (Recommended 100 GB)

Setup Instructions:

1. **Install the required Python libraries:**

   Run the following commands to install the necessary libraries:

   !pip install huggingface_hub
   !pip install ultralytics
   !pip install opencv-python Pillow ipywidgets
   !pip install sam2

2. **Run the script:**

   Run : streamlit run app.py

3. **Upload the video and the image**
4. **Select the objects to be summarised**

5. **Select the video time stamp**
6. **Final Summarized video will be displayed as result**

   When the user selects objects, the SAM2 model segments the frames, and then combines the segmented frames to generate the final result as a segmented video.

For MAC:

This model cannot be run in Mac as it doesn't support NVIDIA GPUs. Currently SAM2 model can be used only on devices that are cuda compatible.