

Automated Face & Hand Segmentation Using SAM2 and MediaPipe

Introduction

This project focuses on building an automated pipeline for segmenting the face and hands in static images using SAM2.1 and MediaPipe. MediaPipe is used to detect the bounding boxes of the face and hands, which are then directly used as box prompts for SAM2.1 Large to perform segmentation. A minimal Gradio interface is included for deployment, allowing users to upload images and view both the detected bounding boxes and the corresponding segmentation masks in real time. The entire system runs automatically without any manual annotations or clicks.

Approach

The pipeline begins by reading the input image into an RGB NumPy array and converting it to BGR solely for OpenCV drawing routines. No other preprocessing is applied, so MediaPipe and SAM2 both see the original resolution.

Bounding Box Detection

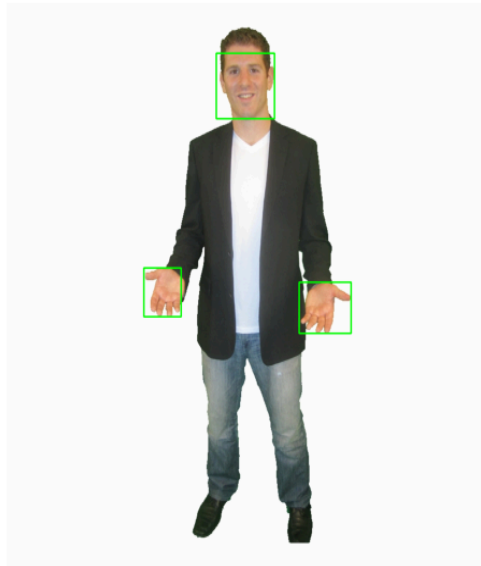
In the first stage, MediaPipe's FaceDetection and Hands modules work in tandem to locate regions of interest. The face detector runs with a confidence threshold of 0.5, while the hand detector is configured for static images and up to two hands. Each detection yields relative coordinates, which we translate into pixel-level boxes and expand by a six-pixel padding carefully clamped to the image boundaries. These boxes are drawn onto a BGR copy of the frame so we can immediately visualize what the model has found, then converted back to RGB for consistency with the segmentation stage.

Results:

Detected Face/Hands



Detected Face/Hands



Segmentation with SAM2

Next, each outlined region is used to prompt the segmentation model. The model produces a fine-grained mask for every face and hand, isolating these areas from the background. We apply each mask back onto the original frame as a semi-transparent overlay green for faces, blue for hands while retaining the original outlines for context. The result is a clear, immediate comparison of initial detections versus final, precise segmentations.

Results:

Segmented Output



Segmented Output



Gradio Deployment

I built a minimal Gradio app so anyone can try the pipeline instantly: you just upload a photo, click go, and within seconds you get two images side by side one showing the MediaPipe bounding boxes and the other showing the overlaid segmentation masks.

Deployment video :[Video](#)