Hand gesture tracking in videos

Student Names: Mian Wang

Northeastern University
Khoury College of Computer Sciences
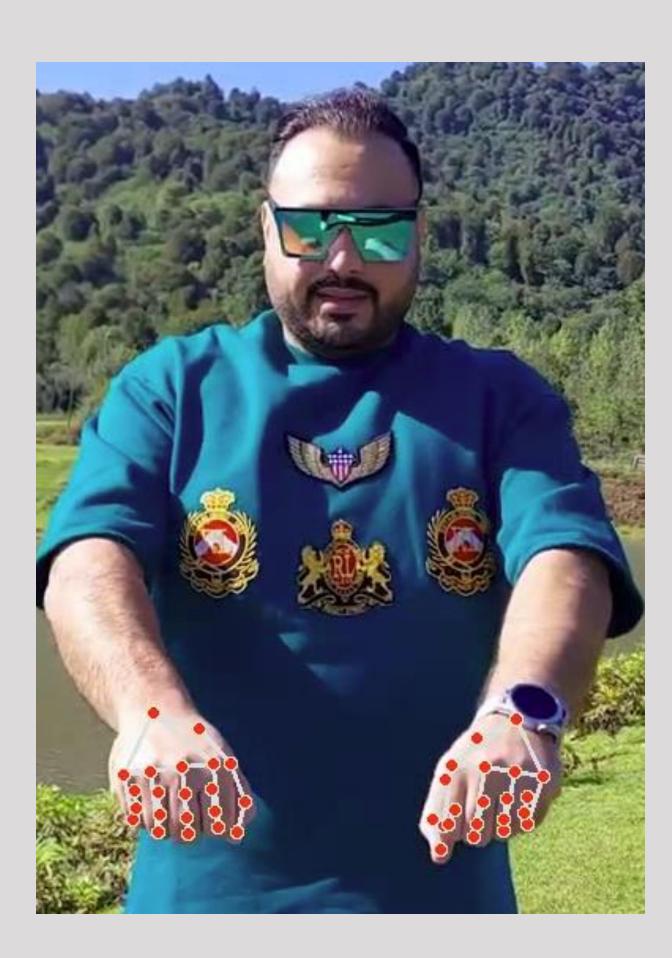
BACKGROUND (or Introduction) and PROBLEM

- The game "Guess Which Hand" is a simple game where one player hides an object in one of their hands, then presents both hands to another player to guess which one holds the object. This research aims to train an AI model to recognize body postures, hand positions, and subtle movements, using this game as a basis for developing posture recognition capabilities. The project's input would be video or image data capturing players' postures and hand movements during the game, while the output would be predictions indicating the likely hand holding the object. The apprentice researcher will receive essential resources, including a labeled dataset capturing various postures and hand positions. This dataset will consist of images or video frames annotated with key body and hand landmarks. The project will involve several training steps: first, preprocessing the dataset, including annotation and augmentation to ensure robust model performance across different scenarios. Next, a deep learning model (such as YOLO, HRNet or OpenPose) for pose estimation—will be trained on this data, using frameworks like TensorFlow or PyTorch. Iterative training and validation steps will allow for adjustments and improvements to model accuracy. The project requires skills in computer vision, machine learning, and programming, with experience in data annotation and motion tracking as advantageous.

APPROCHES (or Methods)

Dataset Preparation: Preprocessed a dataset of 20 videos, including annotation and augmentation to ensure robust model performance across different scenarios.

Model Selection: Employed the YOLO (You Only Look Once) model for pose estimation and hand gesture tracking, leveraging its real-time object detection capabilities.

Landmark Generation: Generated hand landmarks (e.g., wrist, thumb, index finger) and predicted results, which were saved as CSV files for analysis.

OUTCOME (or Results)

Successfully generated hand landmarks and predicted results from 20 videos using the YOLO model.

Produced CSV files containing detailed landmark data, including coordinates (x, y, z) for key hand points (e.g., wrist, thumb, index finger) and predictions for right-hand presence.

Example output (from Image 1): For frame 0, the model identified the right hand with landmarks such as the wrist at (0.759, 0.716, -0.169) and the index finger tip at (0.721, 0.790, -0.059), with "right_hand_present" predicted as TRUE.

Visualizations of hand landmarks and predictions were generated to validate the model's performance.

Conclusions

The YOLO model effectively tracked hand gestures and predicted the presence of the right hand in the "Guess Which Hand" game, demonstrating its potential for posture recognition tasks.

The generated CSV files provided valuable insights into hand landmark positions, enabling precise analysis of hand movements.

This project highlights the potential of AI in understanding human gestures, with applications in gaming, human-computer interaction, and motion analysis.

References