Hand Gesture Tracking in Videos using YOLO

Professor Names: Jeongkyu Lee, Mohammad Toutiaee, Student Names: Mian Wang

Northeastern University
Khoury College of Computer Sciences

BACKGROUND (or Introduction)

In the "Guess Which Hand" game, a player hides an object in one hand, and another guesses its location. This project trains an AI to recognize hand positions and postures using 20 videos with annotated landmarks. The goal is to advance posture recognition for applications in gaming and human-computer interaction.

PROBLEM

The task is to predict which hand holds the object in the "Guess Which Hand" game using video data. The AI must track hand gestures, detect landmarks (e.g., wrist, fingers), and handle challenges like occlusions and lighting variations. The output is a prediction of the object’s likely hand location.

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)

00001_data_20250322_205703

frame	timestamp	hand_index	hand_label	landmark_index	landmark_name	x	y	z	left_hand_present	right_hand_present
0	0.0	0	Right	0	WRIST	0.7592348456382750	0.7164531350135800	1.6098231014893 8E-07	FALSE	TRUE
0	0.0	0	Right	1	THUMB_CMC	0.7687110900878910	0.728232741355896	-0.034059111028909 70	FALSE	TRUE
0	0.0	0	Right	2	THUMB_MCP	0.775063693523407	0.7507306933403020	-0.045160800218582 20	FALSE	TRUE
0	0.0	0	Right	3	THUMB_IP	0.7694122195243840	0.7710885405540470	-0.050673671066761 00	FALSE	TRUE
0	0.0	0	Right	4	THUMB_TIP	0.7590732574462890	0.786910116672516 0	-0.053145274519920 40	FALSE	TRUE
0	0.0	0	Right	5	INDEX_FINGER_MCP	0.7981778979301450	0.7778782844543460	-0.019442081451416 000	FALSE	TRUE
0	0.0	0	Right	6	INDEX_FINGER_PIP	0.7753847241401670	0.8044007420539860	-0.037019498646259 30	FALSE	TRUE
0	0.0	0	Right	7	INDEX_FINGER_DIP	0.7455855011940000	0.8041597604751590	-0.051788348704576 50	FALSE	TRUE
0	0.0	0	Right	8	INDEX_FINGER_TIP	0.7217285633087160	0.7970301508903500	-0.059012994170188 900	FALSE	TRUE
0	0.0	0	Right	9	MIDDLE_FINGER_MCP	0.7816058993339540	0.7773427963256840	-0.005540490616112 9500	FALSE	TRUE
0	0.0	0	Right	10	MIDDLE_FINGER_PIP	0.7546167373657230	0.7994251847267150	-0.023463904857635 500	FALSE	TRUE
0	0.0	0	Right	11	MIDDLE_FINGER_DIP	0.7251232862472530	0.7962070107460020	-0.039194533231596 100	FALSE	TRUE
0	0.0	0	Right	12	MIDDLE_FINGER_TIP	0.7019429206848150	0.7877747416496280	-0.046548143029213 00	FALSE	TRUE
0	0.0	0	Right	13	RING_FINGER_MCP	0.7628980875015260	0.7741098403930660	0.004749728366732600	FALSE	TRUE
0	0.0	0	Right	14	RING_FINGER_PIP	0.7399324178695680	0.7925633788108830	-0.012244765645755700	FALSE	TRUE
0	0.0	0	Right	15	RING_FINGER_DIP	0.7158321142196660	0.7906695604324340	-0.025212848559022 00	FALSE	TRUE
0	0.0	0	Right	16	RING_FINGER_TIP	0.6965320706367490	0.7829573154449460	-0.030177257595074 30	FALSE	TRUE
0	0.0	0	Right	17	PINKY_MCP	0.7457074522972110	0.7689496278762820	0.012580928392708300	FALSE	TRUE
0	0.0	0	Right	18	PINKY_PIP	0.7292870283126830	0.7829254269599920	-0.001695453771390020 0	FALSE	TRUE
0	0.0	0	Right	19	PINKY_DIP	0.712589442729 95	0.7818864583969120	-0.009914831258356570	FALSE	TRUE
0	0.0	0	Right	20	PINKY_TIP	0.6986371278762820	0.7757253646850590	-0.012257821857929 20	FALSE	TRUE
0	0.0	1	Left	0	WRIST	0.1224750429391860	0.7169436812400820	-1.209346578434 63E-07	FALSE	TRUE

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

YOLO: Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *(CVPR)*.