Hand Gesture Tracking in Videos using YOLO V8

Mian Wang, MSDS @ Portland campus

Supervisor name: Mohammad Toutiaee

Northeastern University
Khoury College of Computer Sciences

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.

Fig.1: VIDEO

Fig.2: MARK

PROBLEM Definition

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.

APPROACHES

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. Tried different versions like V10 and V11 with around 70% accuracy, V8 is best on metric.

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

OUTCOME

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	1.60982310148938E-07	FALSE	TRUE
0	0.0	0	Right	1	THUMB_CMC	0.7687110900878910	0.728232741355896	-0.03405911102890970	FALSE	TRUE
0	0.0	0	Right	2	THUMB_MCP	0.775063693523407	0.7507306933403020	-0.04516080021858220	FALSE	TRUE
0	0.0	0	Right	3	THUMB_IP	0.7694122195243840	0.7710885405540470	-0.05067367106676100	FALSE	TRUE
0	0.0	0	Right	4	THUMB_TIP	0.7590732574462890	0.7869101166725160	-0.05314527451992040	FALSE	TRUE
0	0.0	0	Right	5	INDEX_FINGER_MCP	0.7981778979301450	0.7778782844543460	-0.019442081451416000	FALSE	TRUE
0	0.0	0	Right	6	INDEX_FINGER_PIP	0.7753847241401670	0.8044007420539860	-0.03701949864625930	FALSE	TRUE
0	0.0	0	Right	7	INDEX_FINGER_DIP	0.7455855011940000	0.8041597604751590	-0.05178834870457650	FALSE	TRUE
0	0.0	0	Right	8	INDEX_FINGER_TIP	0.7217285633087160	0.7970301508903500	-0.05901299417018890	FALSE	TRUE
0	0.0	0	Right	9	MIDDLE_FINGER_MCP	0.7816058993339540	0.7773427963256840	-0.0055404906161129500	FALSE	TRUE
0	0.0	0	Right	10	MIDDLE_FINGER_PIP	0.7546167373657230	0.7994251847267150	-0.023463904857635500	FALSE	TRUE
0	0.0	0	Right	11	MIDDLE_FINGER_DIP	0.7251232862472530	0.7962070107460020	-0.039194632321596100	FALSE	TRUE
0	0.0	0	Right	12	MIDDLE_FINGER_TIP	0.7019429206848150	0.7877747416496280	-0.04654814302921300	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	-0.012244765646755700	FALSE	TRUE
0	0.0	0	Right	15	RING_FINGER_DIP	0.7158321142196660	0.7906695604324340	-0.02521284855902200	FALSE	TRUE
0	0.0	0	Right	16	RING_FINGER_TIP	0.6965320706367490	0.7829573154449460	-0.03017725795507430	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	-0.0016954537713900200	FALSE	TRUE
0	0.0	0	Right	19	PINKY_DIP	0.71258944272995	0.7818864583969120	-0.009914831258356570	FALSE	TRUE
0	0.0	0	Right	20	PINKY_TIP	0.6986371278762820	0.7757253646850590	-0.01225782185792920	FALSE	TRUE
0	0.0	1	Left	0	WRIST	0.1224750429391860	0.7169436812400820	-1.20934657843463E-07	FALSE	TRUE

video_id	Ground Truth	Prediction
1	right	1
2	right	1
3	left	1
4	left	1
5	left	1
6	right	1
7	left	0
8	right	1
9	right	1
10	right	1
11	left	1
12	left	0
13	right	1
14	right	1
15	right	1
16	left	1
17	left	1
18	left	0
19	left	1
20	right	1

Fig.3: PREDICTION

Accuracy: 85%

Conclusions

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

References

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. https://doi.org/10.1109/CVPR.2016.91