Machine Learning Lab
Slot:- L47+L48
Team Name – TechnoSapiens
Explanation Project

Team Members:

VennaGiri Bhanu Sri Mohith	-23BCE20010
Shaik Arshiya Anjum	-23BCE7053
V.C.Premchand Yadav	-23BCE7167
MD Nasir Alam	-23BCE8680
P.V.Satya Laxmi Aswitha	-23BCE9302
Velaga Mouli	-23BCE9374

Faculty Name:- Prof. Sai Babu Veesam

1. Introduction

This project focuses on recognizing human actions such as walking, running, jumping, and waving from videos. It involves three main components:

 • Video Processing: Extracting frames from videos and combining them into a single video.

• Model Training: Training a Convolutional Neural Network (CNN) to classify actions based on images.

• Real-time Action Recognition: Using a trained model to predict actions from a video while drawing a bounding box around detected persons using MediaPipe.

2. Technologies and Tools Used

• **Programming Language:** Python

• **Libraries & Frameworks:**

    o OpenCV (cv2) – Video processing and frame handling

    o TensorFlow/Keras – Deep learning model creation

    o NumPy – Data handling and preprocessing

    o MediaPipe – Pose estimation and bounding box drawing

    o natsort – Natural sorting of images

3. Project Workflow

The project consists of three key scripts:

(i) Video Processing: Combining Frames into a Video

• The first script extracts image frames from different action folders (e.g., walk, run, jump, wave).

• It then combines these frames into a single video (combined_video.avi).

• Purpose: This prepares a dataset video for training and testing.

(ii) Training a CNN Model for Action Recognition

• The second script loads images from the dataset, resizes them, and normalizes pixel values.

• A Convolutional Neural Network (CNN) is designed and trained on the dataset to classify actions.

- The trained model is saved as "action_recognition_model.h5".
- Purpose: This model will later be used for recognizing actions in video frames.

(iii) Real-Time Action Recognition
- The third script loads the trained CNN model and processes frames from the video.
- Each frame is resized and passed through the model for classification.
- MediaPipe Pose Estimation is used to detect human poses and draw bounding boxes.
- The predicted action is displayed on the screen.
- Purpose: This script allows real-time action detection.

4. Output

Action: run

Action: walk

5. Conclusion

• This project successfully detects human actions from videos using deep learning and computer vision.

• The CNN model classifies actions based on frame-based features.

• MediaPipe enhances detection by ensuring the focus remains on detected humans.

• Future improvements could involve using LSTMs (Long Short Term Memory networks) to analyze motion sequences instead of only individual frames for better accuracy.