Action Recognition using Pose Estimation

Industry Graded Project - PG Program in AIML, NITW

Problem Statement:

Analysis of people's actions and activities in public and private environments are highly necessary for security. This cannot be done manually as the number of cameras for surveillance produce lengthy hours of video feed every day. Real-time detection and alerting of suspicious activities or actions are also challenging in these scenarios. This issue can be solved by applying deep learning-based algorithms for action recognition.

Project Description:

Aim of this project is to automatically recognize human actions based on analysis of the body landmarks using pose estimation. The following are the major tasks performed:

1. Implementation of Convolutional Neural Network based pose estimation for body landmark detection
2. Implementation of pose features-based action recognition and its improvement using graphical feature representation and data augmentation of body landmarks
3. Preparation and preprocessing of image datasets
4. Fine tuning and improvement of the action recognition model with better feature representation and data augmentation
5. Development, error analysis and deep learning model improvement

Dataset:

A subset of Frames Labeled in Cinema (FLIC). Training images and their annotations in action_joints.csv consists of 7 joints - 'left shoulder', 'left elbow', 'left wrist', 'right shoulder', 'right elbow', 'right wrist', 'left eye', 'right eye', 'nose'.

Project Solutions:

1. **Human Pose Estimation:**

   Pose estimation is achieved through the implementation of a CNN model. Transfer learning has been utilized for detection of body joints. VGG16 pre-trained model is used as the convolution base. The top model is customized for the problem. Training takes place in 2 groups. I was able to obtain a **R2 score of 0.92** on the test data, which seems accurate enough.

Reshma Abraham

Testing on Image outside the dataset:

```
[169.36214  185.21248  180.11754  147.68842  160.61696   81.57629
 107.488594  40.405823  68.03899   84.73467   54.323395 152.2243
  60.107597 187.63916 ]
```

Our model seems to detect the body joints fairly well.

2. Action Recognition:

I created two action recognition models.

1. **xy_actions_model.h5**: Model that trains on the (X,Y) coordinates of joints

- This Neural network was built by directly considering x and y coordinates as features and action label as target. This model gave me a **validation accuracy of 100%**. But humans can appear on any scale in a real scenario, considering raw coordinates as features is not a good idea. Therefore, I developed the below model.

2. **dist_actions_model.h5**: Model that trains on the Euclidean distance of the joint coordinates

- For this model I have extracted the distance between every joint and further normalized the distance features. These distance features were considered for training the action recognition model. I was able to achieve a **validation accuracy of 100%**. However, this model wasn't stable, ie.. this model occasionally did return a 50% accuracy.

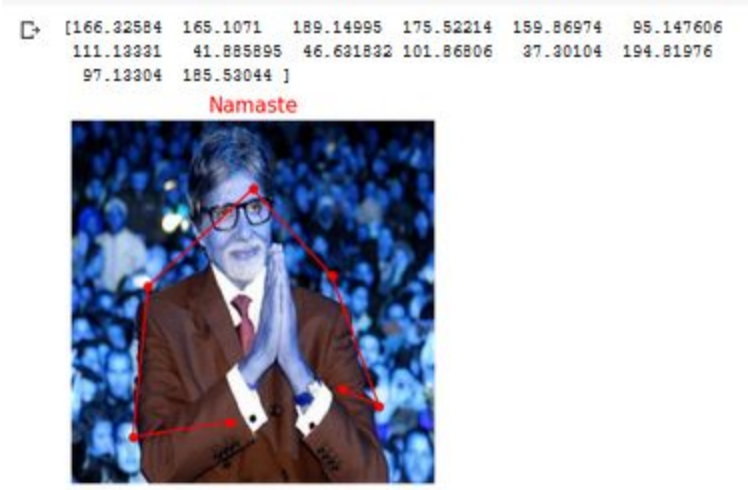
Therefore, **for action recognition in Video**, I have used the first model (**xy_actions_model.h5**)

3. Implementation of Action Recognition in videos using pose joints estimated by the CNN model

The '**pose_estimation_model.h5**' has estimated Pose fairly accurately and '**xy_actions_model.h5**' has recognized the action correctly as 'Namaste'.

Fig: Output Frames of Video Input

Testing the sequential Pose estimation and action model on a Single image:



Reshma Abraham