Capstone Project

Machine Learning Engineer Masters Program

TABLE OF CONTENTS

Background

Objective of this project is to automatically recognize human actions based on analysis of the body landmarks from pose estimation.

Learning Outcomes

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

Process Flow

Expected Challenges:

The project will aim at tackling the following research challenges:

1. Body landmarks detection of human from different viewpoints
2. Detection of human actions which involves pose detection of humans appearing at any scale in the video frame with action performed using either left or right or both body parts
3. System must meet the real-time processing requirements where the Deep Learning Model must detect pose and actions from video (with 30 fps) at the rate of 33 ms per frame

DATASET

The recommended datasets are shared on your LMS.

TARGET ENVIRONMENT

You can use Edureka's Cloud Lab, a cloud based Jupyter Notebook, which is pre-installed with Python and other required packages to work on this Project.

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.

METHODOLOGY

Pose Estimation

Input to the human pose estimation model is closely cropped pedestrian or human image. Each training set image has one human inside where pixels are considered as features and target as pair of body joint coordinates. Pre-processed Pose Estimation FLIC dataset is used for this modelling. Set of landmarks from human image can be detected by training convolutional neural networks model with convolution, pooling and fully connected layers with finally landmark point regression as output. This Model can be trained based on two strategies:

STRATEGY 1:

1. Training a model from scratch where model layers can be designed manually, and weights of the model will be trained
2. Loading a pretrained base CNN architecture such as VGG16, MobileNet, removing the final softmax layer adding custom layers and training the newly added layers

STRATEGY 2:

This strategy makes use of "transfer learning" where we can choose whether to retrain the whole network or train only newly added layers. Model can be inferred using a single test sample which is not part of the training/test set. The detected pose points can be further plotted.

Action Recognition

Set of human actions can be further recognized using analysis of pose landmarks as features and action label as target. Deep neural networks can be directly trained using dense layers by considering pose point x and y coordinates. A custom dataset with two actions - Namaste, Hello will be used for this task. Action recognition neural network can be built by directly considering x and y coordinates as feature and action label as target. Since human can appear in any scale in a real scenario, considering raw coordinates as features is not a good idea. This can be solved by considering the skeleton of human pose points as graph, extracting the distance between every joint and further normalizing the distance features. These distance features can be considered for training the action recognition model. Dataset contains Hello actions performed using right-hand only. Hence, the model cannot recognize a Hello human action performed with left-hand. This issue can be solved by augmenting the action dataset by duplicating the existing actions further flipping the coordinates horizontally. Now with the augmented action dataset, the trained model can detect Hello action performed in both right as well as left-hand. Finally, the pose estimation and action recognition models can be integrated. The model can be tested with offline videos and action recognition results will be displayed for each video frame.