**Human Activity Recognition through Ensemble Learning of Multiple Convolutional Neural Networks**

**Alternate title:** Deep learning based video classification using Convolution Neural Network

**Aim:**

This paper aim to classify the different types of videos using deep learning framework with convolution neural network (CNN)

**Abstract:**

Video classification has been extensively researched in computer vision due to its wide spread use in many important applications such as human action recognition and dynamic scene classiﬁcation. It is highly desired to have an end-to-end learning framework that can establish effective video representations while simultaneously conducting efficient video classiﬁcation. Deep learning plays a vital role in image processing. We use Convolutional neural network algorithms for classification. The convolution 3-D (C3-D) are ﬁrst deployed to extract temporal and spatial features from the input videos cooperatively, which establishes comprehensive and informative representations of videos.

**Synopsis:**

In this paper, we developed three different CNN based models, using which, a series of ensemble learning models were also created. Ensemble learning is a paradigm by which multiple models or learners are trained to solve the same problem. The advantage of this paradigm lies in its ability to generalize. It has the ability to boost the learning effect of learners that are weaker, leading to an improved performance output. The following sub-sections describe the overview and characteristics of the models.

**Existing System:**

In existing system, mostly we use CCTV camera for the surveillance or monitoring. It was a traditional method to identify what types of works are done. It’s hard to watch our CCTV footage for many hours. In Existing system, there is lot of man work to monitor the works. In our proposed system, we decrease the man work.

**Problem Definition:**

With the advancements in deep learning, this process has been accelerated with unprecedented improvements, especially with the availability of more data and higher computation power. The feature selection process no longer needs to be task-dependent as deep learning models can extract features automatically and they can be applied to a range of classification tasks. With the provision of deep learning, it makes a logical flow to maximise the use of this state-of-the-art research for concentrated implementation of HAR. Therefore, the aim of this work is to develop a systematic process for a more streamlined feature representation for improved activity recognition through ensemble learning of CNN.

**Proposed System:**

In this proposed system, we propose the convolution neural network method for action recognition in video. The input video will be captured by using the webcam. The input video is converted into number of frames. Then the CNN (Convolution Neural Network) algorithm is used in order to detect the particular part of the frame. Then the maximum weight values are taken from the feature extraction frames by using the Convolution neural network. Finally the action will be detected in the videos and then the label (action name) is identified. Then that output taken to the firebase and the firebase value given to the user via android notification.

**Advantage:**

We trained and tested our models and their ensembles using dataset. Training accuracies and losses of the three CNN models. After about Several epochs, the accuracies crossed and approached above 80% accuracy over the course of the 50 epochs. The validation accuracies, however, lay between 70 - 80% with a higher degree of variation observed as compared to the training procedure.

**Module Description:**

* Video Streaming
* Deep learning algorithm
* Classification

**Video streaming**

The input video will be captured by using the webcam. The input video is converted into number of frames. Video sequences present great variability due to huge scale changes, viewpoint variation and camera motion which pose great challenges for both video representations and classiﬁcation.

**Deep learning algorithm**

The convolution neural networks (CNN) have been studied in the video domain for a large variety of classiﬁcation tasks. The captured video is feed into the convolution neural network and it is highly desired to have an end-to-end learning framework that can establish effective video representations while simultaneously conducting efficient video classiﬁcation.

**Classification**

Then the maximum weight values are taken from the feature extraction frames by using the Convolution neural network. According to video labels the entire network is trained for action recognition. Finally the action will be detected in the videos and then label (action name) is identified. After that the detected value passed to the user through firebase via android notification.

**Software Requirements:**

* Operating System : Windows 10 (64 bit)
* Software : Python and Anaconda
* Tools : Jupyter Note Book and Spyder IDE

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Architecture Diagram:**

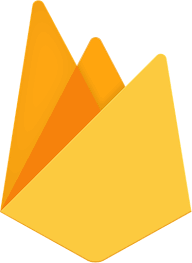
Classification

Convolutional

Neural Network

Activity Recognition

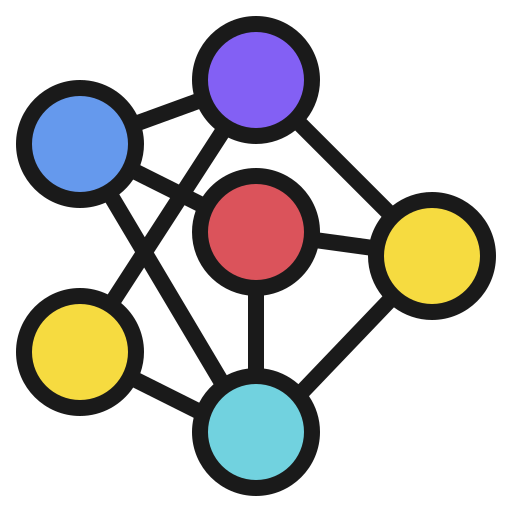
Firebase



Android

Notifications

Video Stream



**Conclusion:**

We presented three CNN based models as well as their ensembles for dataset. It was found that the performance of the ensemble model is better than that of individual models. One of the ensemble model performed better than the methods in the literature. In the dataset we used, we see a class imbalance such that we have 38% samples for walking class but hardly 5% for sitting and standing. In future, the results might be improved even more, if we can remove the class imbalance from dataset.

**Future Work:**

Moreover, currently an ensemble of average of the three models is created but for further exploration, a future direction can be performing weighted ensemble learning such that the best performing model has the most effect in the ensemble. Furthermore, we can explore the area of ensemble learning for a hybrid model, that is, ensemble learning of CNN and RNN models.