IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Human Activity Recognition

Mr. Harishchandra Maurya[1], Mr. Raj Chorghe[2], Mr. Shubham Prasad[3], Mr. Yash Patil[4]
Assistant Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4]
Chhatrapati Shivaji Maharaj Institute of Technology, Panvel, Maharashtra, India

Abstract*: The topic of Human activity recognition (HAR) is a prominent research area topic in the field of computer vision and image processing area. It has enabled state-of-the- art applications in multiple sectors, such as surveillance, digital entertainment, and medical healthcare. Studying and predicting these movements can be interesting and intriguing, and several sensor-based approaches have also been introduced. +predict human activities such accelerometer, gyroscope, etc., This paper develops an intelligent system for recognizing human activities using convolution, which has its own advantages and disadvantages. (3D) Kernels are taught using the Kinetics data set, which has 400 classes that depict humans' daily activities and work, and consists of four.. This model uses the RESNET-34 3D CNN model, and the videos are only around a tenth of a second in length.. The model that was trained performed satisfactorily during all stages of training and testing*

Keywords: Human activity recognition

I. INTRODUCTION

The topic which has increased its importance in last few decades in the domain of Computer Vision and A.I. is “Human Activity Recognition”. As the concepts of the human activity recognition helps in understanding the concepts and issues of the human action understanding which majorly helps in medication, management, learning patterns and many situations of video retrievals.

The Human Activity Recognition Systems (HAR) is capable of recognizing physical activities like running, playing, sleeping, eating and many such activities. The detection of the physical activities by different such sensors and recognition process is a key topic of research in wireless, smartphones and mobile computing. Human Activity recognition Systems is able to perform different tasks and recognize the multi day to day actions performed by humans which can be either simple activities like sleeping or the complex activities like running and eating.

For the purpose of activity recognition of human’s different actions, multiple types of sensors and devices are required like video sensors, environmental activities sensors, body inertia sensors and many other sensors like these which record or sense the human actions.

There are many other sensors used by the HAR systems but with the limited availability of use due to the effect of outdoor environments and activities on them like GPS receiver which is limited to outdoor environments. Thus in this research paper we are trying to implement Human Activity Recognition through resnet-34 algo which is an artificial neural network (ANN) type algo which is based on the constructs of the basic things known from the pyramidal cells of the cerebral cortex. The ResNet algos specifically ResNet-34 do the process of this by the usage of skip connections and the process of jumping over some layers in the different neural networks. The general ResNet Algo and specifically the ResNet 34 algo are basically implemented with two and it’s three layer skip which generally contains the nonlinearities (ReLU) and the batch normalization for the usage in the residue neural network techniques.[1] The skip weights can be recalled by the usage of an additional weight matrix which are known as the HighWayNets term.[2] In ResNet the procedure followed by the models with multiple levels of parallel levels skips are referred as DenseNets.[3] As they are using skipping of multiple neural network layers

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, November 2023

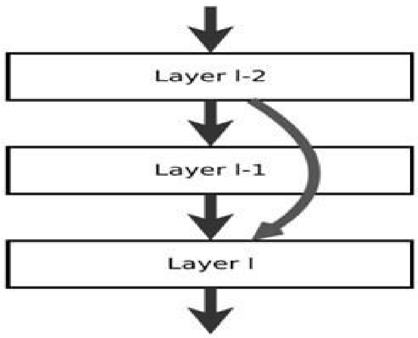
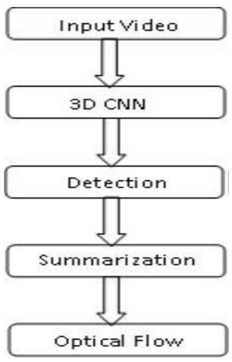


Fig.1 Skipping of Different levels in ResNet

Thus, only the weights of the adjacent layer areadapted with no link to upper layers of the neural network.[6] The best use of ResNet can be taken from it when a singlelayer is over stepped that is when all the linear layers are used as the intermediate layers and if we are not capable of doing this a particular weight matrix should bemade for the all skipped connections that is not the DenseNet should not be used rather HighwayNet should be taken in use. The procedure of skipping multiple layers in ResNet in an effective manner helps us in the simplification of the network by the use of fewer layers in the initial training stages. The learning phase gets speed up as we reduce the impact of the gradients due to the fewer layers for the propagation procedure. The skipped layers are gradually reconstructed via the

II. METHEDOLOGY

Implementation involves two major processes that are training and recognition. To proceed with the training process, we have to pick a temporal spot in a film to generate training samples using sampling.[4]A sixteen frame film is produced about the selected temporal position. We loop around the video until necessary if the video clip selected is smaller than sixteen frames. Next, we will choosea spatial position and spatial scale accordingly as per necessary. The samples are also spatially resized to 112 X 112 pixels. While training the model that is Resnet-34 from scratch the learning rate at the beginning was set to 0.1 and later reduced by a factor of 0.1 after the saturation of validation loss.[5] Then comes the recognition part where the loop begins over the frames where we first initialize the batch of frames that will be passed to the neural net. From there we will populate the batch of frames from the stream of video and resize them to a width of 400 pixels and maintain the aspect ratios.[7] The reason here is that we're building a batch of multiple images to be passed through the human activity recognition network, enabling it to take advantage of spatiotemporal information. Dataset used to train the model is the Kinetics human action video dataset. The dataset contains 400 classes of human activities, with 400 and more films for each and every action. Each film lasts around tenth of a second and is extracted from a different YouTube clip.[8] The actions are human centric and cover a wide range of classes including human-object interactions such as riding skateboard, cooking, smoking, reading book, reading newspaper ,as well as human-human interactions such as hand shaking, hugging etc.[9]



III. RESULT AND DISCUSSION

The trained model gave an accuracy of 79% on the kinetic dataset. We observed that the accuracy was very high for activities like running, standing, etc. but it was reduced considerably for activities like cooking, doing yoga, etc., since there are several ways of performing these activities. For further improvement of results, we can amore detailed dataset which separates the different yoga asanas into different labels. We observe that datasets with more detailed class labels give better results. So, instead of using the broad term cooking, splitting the class into different labels like cooking rice, boiling water, etc.

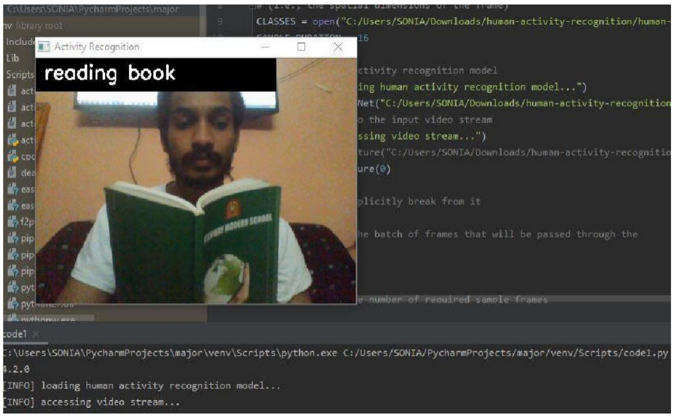


Fig.2 Demonstration Of Activity(Reading Book)

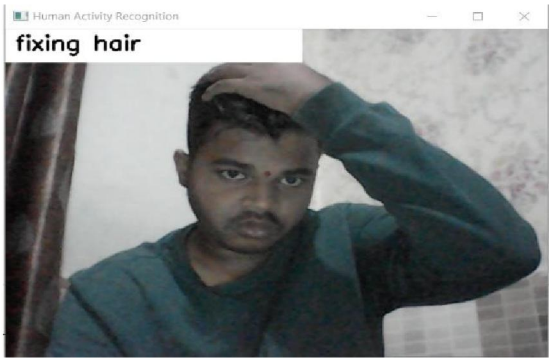


Fig.3 Fixing hair

Fig.4 Reading newspaper

The dataset contains a single activity in each entry, examples like 2 people in the same frame performing different activities were not considered. For such entries, first performing some video processing to determine personof interest in the frame and then using the model to determine different activities will be sufficient.

IV. CONCLUSION

In this paper Human Activity Recognition System, weproposed a model trained using Convolutional neural network (CNN) with spatiotemporal three-dimensional kernels on Kinetic data set to recognize almost 400 human activities with satisfactory accuracy level. The designed system can be used to automatically categorizing a dataset of videos on disk, training and monitoring a new employee to correctly perform a task, verify food worker services, monitoring bar/restaurants patrons and ensuring they are well served. For future work, we can use a dataset coveringmore than 400 activities to make the system more versatile.It is also observed that increasing the number of samples foran activity in the dataset improves the performance of the system significantly

V. ACKNOWLEDGMENT

We the team members of the research project sincerely want to thank our guide Assistant Professor Ms. Shikha our supervisor and the reputed Electronics andCommunication Department of Delhi TechnologicalUniversity, Delhi, India for their helpful and motivating encouragement and the must needed support for the completion of this project work by providing the golden opportunity in the form of Major Project.

REFERENCES

[1]. He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2015-12- 10). "Deep Residual Learning forImage Recognition".
[2]. Srivastava, Rupesh Kumar; Greff, Klaus; Schmidhuber, Jürgen (2015-05-02). "HighwayNetworks".
[3]. Huang, Gao; Liu, Zhuang; Weinberger, Kilian Q.; van der Maaten, Laurens (2016-08-24). "Densely Connected Convolutional Networks".
[4]. S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici,B. Varadarajan, and S. Vijayanarasimhan. YouTube-8M: A large-scale video classification benchmark. arXiv preprint, arXiv:1609.08675, 2016
[5]. Z. Qiu, T. Yao, and T. Mei. Learning spatio-temporal representation with pseudo-3d residualnetworks. In Proceedings of the International Conference on Computer Vision (ICCV), 2017.
[6]. L. Wang, Y. Qiao, and X. Tang. Action recognitionwith trajectorypooled deep-convolutional descriptors.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4305–4314, 2015.
[7]. BishoySefen, Sebastian Baumbach et. al. / HumanActivity Recognition Using Sensor Data of Smart phones and Smart watches/ ICAART 2016[1]
[8]. un X, Chen C, Manjunath BS , ―Probabilisticmotion para-meter models for human activity recognition,‖ In: Proceedings of 16th internationalconference on pattern recognition, pp 443–450
[9]. Kwon W, Lee TW, ― Phoneme recognition usingICA-based feature extraction and transformation,‖ Signal Process 84(6):1005– 1019, 2004
[10]. Lee SI, Batzoglou S, ―Application of independent compo-nent analysis to microarrays,‖ Genome Biol 4(11):R76.1–21, 2003