International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 11 | Nov 2023 www.irjet.net p-ISSN: 2395-0072

A Intensified Approach on Deep Neural Networks for Human Activity Recognition Using Computer Vision and Machine Learning Technique

Velantina.V[1], Mr.Manikandan.V[2]

[1,2]*Department of CSE, JU, Bangalore, Karnataka, India*

Abstract- *In recent years a lot of interest in human activity recognition in video analysis has been a trend in today's digital world. However, the majority of these methods assign a single activity name to a video after dissecting the entire clip or using a classifier for each instance. However, it tends to be inferred that we humans only need one instance of visual information for scene recognition when compared to the human vision system. In addition, small groups of edges or even a single video case are sufficient for precise identification. The model accepts outlines as information. It lays the groundwork for evaluating the discovery system by providing an overview of key datasets and conclusive estimation measurements in a condensed manner. Additionally, the focus examines security, intranet-class variations, conjecture, and multi-scale object identification concerns. The reviewed overview also lists the crucial steps for developing the model. In order to detect a variety of human actions, AI and vision are the subject of extensive research. It has been demonstrated that there are a number of effective strategies for action recognition, and both motion videos and still images successfully provide action information. Using the dataset that the model is being trained on, the HAR model is used to detect human actions in various situations. By identifying various activities in various scenarios, the model demonstrates accuracy and performance.*

***Key Words*: Deep learning, CNN, ANN, RNN, Machine learning, activity recognition.**

1. INTRODUCTION

The Systems for monitoring human behavior and interacting with computers have grown significantly in recent years. "Human activity recognition" is the process of analyzing sensor or video data. Sitting, walking, biking, jogging, eating, reading, washing, and other static or dynamic human actions include In recent years, deep learning has grown in popularity for pattern recognition and sensor data analysis. Since the development of deep learning, CNNs have received a lot of attention. Convolutional neural networks are used in natural language processing, medical image analysis, and recommendation systems.

Practitioners of deep learning also have a hard time manually building deep models and figuring out the right configuration through trial and error. Infusing area information into DL requires many advances, including model age, model organization, and component designing. The mechanical headways of the present knowledge designing patterns have brought about the making of novel and inventive ideas. Human action recognition is one method for identifying human-performed actions. People perform a wide range of actions in a variety of settings every day. We must be aware of its characteristics and patterns in order to accurately identify an activity. For the purpose of activity identification and analysis, a number of datasets gathered from various standard repositories are utilized.[1]

Due to their numerous applications, such as customized video following, picture commenting, etc., activity recognition in confidential recordings sent by customers has grown to be a major examination point in light of the rapid advancement of web apps and mobile phones [2]. As a result, these recordings contain a wide range of class variations inside a single semantic description. Currently, it is a challenging to recognize human movements in these recordings. The conventional framework was followed by numerous movement acknowledgement techniques. Recordings initially had a huge number of development elements removed from them. At that stage, each area is quantized into a histogram vector using back-of-words (bow) depiction. Then, to execute acknowledgment in testing and recording, vector-based classifiers, such as reinforce vector machines, are used. Nevertheless, during the quantization and extraction of nearby characteristics, a correlated information might be introduced [9]. These techniques may need an extensive preparing process with several accounts in order to get executed particularly for real recordings. In general, motionless images can also be used to reduce human action propensity.

Therefore, these techniques are typically weak and cannot be employed effectively when the video has significant camera shake, occlusion, unclear establishment, etc. Important activities, such as connected content, human appearance etc.

Human activity in order to improve recognition accuracy. Recent analyses have demonstrated the usefulness of using comparable objects or human positions. These techniques could need a lengthy preparation process with a huge number of accounts to get exceptional execution, notably for real recordings. In general, motionless images can also be used to decrease human action propensity.

The ability to adapt has been increased, and changes can now be coordinated between groups of places. The

modification procedure in a semi-coordinated learning system should be feasible in asking to investigate the surrounding tangled structures close to the preparing video information and successfully employ the unlabeled information in video space.

2. LITERATURE REVIEW

[1] Human activity recognition is a core problem in intelligent automation systems due to its far-reaching applications including ubiquitous computing, health-care services, and smart living. In this paper, we posit the feature embedding from deep neural networks may convey complementary information and propose a novel knowledge distilling strategy to improve its performance. More specifically, an efficient shallow network, i.e., single-layer feed forward neural network (SLFN), with handcrafted features is utilized to assist a deep long short-term memory (LSTM) network. On the one hand, the deep LSTM network is able to learn features from raw sensory data to encode temporal dependencies. On the other hand, the deep LSTM network can also learn from SLFN to mimic how it generalizes. Experimental results demonstrate the superiority of the proposed method in terms of recognition accuracy against several state-of-the-art methods in the literature. [2] This review article surveys extensively the current progresses made toward video-based human activity recognition. Three aspects for human activity recognition are addressed including core technology, human activity recognition systems, and applications from low-level to high-level representation.

In the core technology, three critical processing stages are thoroughly discussed mainly: human object segmentation, feature extraction and representation, activity detection and classification algorithms. In the human activity recognition systems, three main types are mentioned, including single person activity recognition, multiple people interaction and crowd behaviour, and abnormal activity recognition.

Finally the domains of applications are discussed in detail, specifically, on surveillance environments, entertainment environments and healthcare systems. [4] The model presents a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers deeper than VGG nets but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. [5] A fused convolution layer without loss of performance, but with a substantial saving in parameters that it is better to fuse such networks spatially at the last convolutional layer than earlier, and that additionally fusing at the class prediction layer can boost accuracy. Three aspects for human activity recognition are addressed including core technology, human activity recognition systems, and applications from low-level to high-level representation. In the core technology, three critical processing stages are thoroughly discussed mainly: human object segmentation, feature extraction and representation, activity detection and classification. In any event, it makes sense that existing ML models would search for clear names for more operations. As they will be put through additional testing when experiencing different but equivalent workouts, their accuracy may decrease. More data grouping may be pointless if the relationships between the data are not understood [10].

Additionally, it is difficult to update the model structure without addressing the model's foundational elements, which are typically described as the machine. In light of these problems, the practical PC□based knowledge approach provides a way to understand the viewpoint of the computer while conceivably incorporating such information to further promote estimated precision or comfort in actual situations with diverse areas. The majority of recent effort has concentrated on organising instructional assortments and fine-tuning models, with little attention paid to the post-taking care related to the model gauges findings.

Second, security has evolved into a problem that dramatically disrupts a variety of human-PC structures. Security has been a difficult union mark to check, especially given the need for explicit regulations like the General Data Protection Rule (GDPR). Huge video picture variations and the ability to cope with limits have recently brought forth a variety of thoughts for everyone. For instance, having the option to anticipate people's daily activities (ADL) and carry out auto tracking will significantly aid in the development of a system for evidence-based medicine in the clinical benefits area.

Human activity recognition model describe Test results demonstrate the assessment's effectiveness and better change execution, especially when only a small number of specified preparation tests are administered. Due to its ability to open up information in a variety of fields including surveillance, gaming, free automobiles, clinical imaging, human activity affirmation, and the video dealing attracts enormous amounts of interest from both the academic community and industry. Thus, video monitoring and evaluation have developed into a popular research area in the fields of artificial intelligence and machine learning. With the use of significant learning, the presentation of video dealing with initiatives has been elevated to a new level. Irrelevant for a given use scenario. In any event, it makes sense that existing ML models would search for clear names for more operations.

As they will be put through additional testing when experiencing different but equivalent workouts, their accuracy may decrease. More data grouping may be pointless if the relationships between the data are not understood. In any event, without information channel management collection through video can be incredibly apparent. It is challenging to balance the estimated precision

and security using the current ML models.[6] Therefore, a method for balancing is revealed by the results of the video examination with insurance while maintaining the viability of the system and people's security concerns is required. Due to limited processing resources, CNN convolutional layers with four layers, including a dropout layer and a fully linked layer, can improve overall sensitivity, according to preliminary tests. During this phase, the human action recognition process implemented genetic algorithms with an effective search function. after the CNN classifier's topology is initialized and the chromosomes are decoded for fitness evaluation. Gap in the research: In order to collect data, an individual must wear multiple sensors, and the placement of those sensors has an impact on the results. It is hard to get attributes and the most important characteristics from sensor data.

It may be difficult to use embedded sensors to map the activities of all of a person's residents if they have multiple residences. Classification techniques have different effects on time complexity and accuracy in relation to time precision and complexity as occurs frequently.

3. DISCUSSION

A CNN for feature extraction and a bidirectional LSTM layer for time series prediction across three fully connected channels are combined in the proposed multichannel deep learning model. On each head or channel, a Max pooling layer is followed by three stacked Conv1D layers. For the purpose of feature extraction, a Conv1D layer with 64 filters directly maps and abstracts the inputs.

The result of each Conv1D layer is shipped off a maximum pooling layer to separate the learned elements into more modest, more reasonable pieces without forfeiting precision. When it comes to altering the internal state, bidirectional LSTM layers are an excellent choice because they employ both forward and backward dynamics. A bidirectional LSTM's inputs are typically the time series and derived features from the convolution process. A typical network consists of an input layer, one or more hidden layers, and an output layer. The three-dimensional structure of neurons is comparable to the human neural network., and, depending on how they are arranged, the normalization layers might be the hidden layers.

The pre-processing of video- based or sensor-acquired data is the first step in this framework. The user must choose the search space parameters, such as training parameters and termination criteria. Convolutional and completely associated layer plans for CNNs are created in the subsequent segment. This is because a trained classifier's performance is greatly influenced by its initial parameter weights. Due to limited processing resources, CNN convolutional layers with four layers, including a dropout layer and a fully linked layer, can improve overall sensitivity, according to preliminary tests. During this phase, the human action recognition process implemented genetic algorithms with an effective search function. after the CNN classifier's topology is initialized.

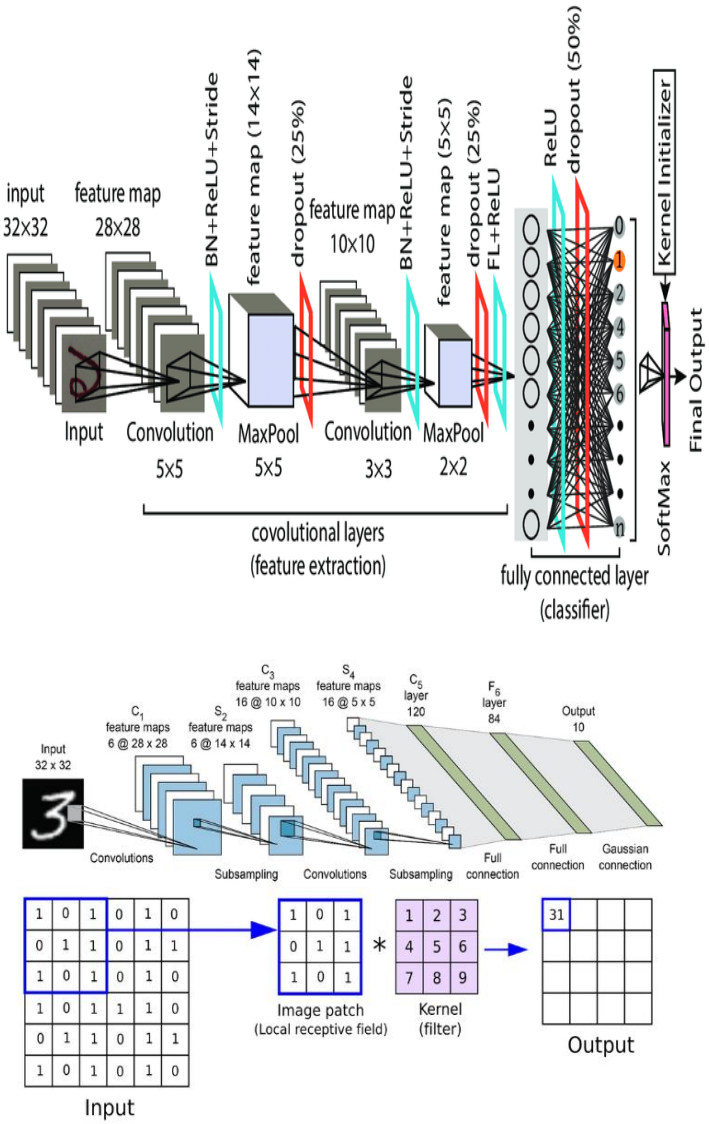

Fig-1: CNN network of HAR activity recognition model

The CNN's architecture will be arbitrary and may have a negative impact on classification performance. The set of constraints can include time limits, the maximum number of neurons that can be connected, model depth, and the number of filters that can be used. The CNN structure's identification, coding, training, and optimization demonstrated that model optimization is constrained and does not permit in-depth investigation of specific structures in the model search space. A vision sensor that looks like a camera is installed in the environment where the person uses vision-based approaches to do their daily tasks.

As a result, they are limited by issues like privacy, lighting, position, obstruction, and occlusion. Wi-Fi-based techniques, on the other hand, use advancements in public wireless infrastructure and Wi-Fi signals to identify a user's activity

and detect changes in the patterns of Wi-Fi signals reflected by the user's body.

Due to the three-layered data input, 2D convolutional layers are frequently employed for the common collecting of images. The convolutional collaboration iteratively passes a channel (segment) through the image, checking individual pixels until everything is in order. The channel describes feature map establishments as the touch result of the pixel values in the continuous channel window with the heaps. As opposed to the 2D CNN, the 1D CNN uses convolutional executives to isolate features from fixed‐length chunks of the entire dataset. It is appropriate for the time-sensitive progressions of multimodal sensor data and acoustic signals to confirm human actions. Memory blocks serve as stand-out components for both the inner and outer rehash of significant learning LSTM model data features and their transient situations.

LSTM layers frequently include memory impediments that are drearily connected to a memory unit or cell. These phones are designed with techniques to decide when to stop remembering previous mysterious states of the memory cell and to update them further as necessary to allow the association to use ephemeral information. As illustrated in Figure 4, the design of the proposed multi-channel significant learning model combines CNN for feature extraction and Bidirectional LSTM layers for course of action assumption in three channels that are then coupled together. Three stacked Conv1D layers are present in each channel or head, and they are all followed by a largest pooling layer. Direct pre-processing and acceptable portrayal of sensor inputs for feature extraction are carried out by the 64-channel Conv1D layers. Fig.2 Demonstration of the model. The accuracy of the convolution heads allows for the part extraction. Conv1D layers and max-pooling layers result in a bidirectional LSTM layer with configurations of 128 units and a dropout capacity of 0.25 percent.

The extraction of features presents the greatest obstacle for sensor-based HAR systems. A lot of research has been done on the capability of inferring a particular class from a sensor stream. Despite the fact that spectral and frequency-based statistical machine learning (SML) methods have been proposed in the past, convolutional neural networks and deep learning (DL) in general have seen tremendous growth in the context of HAR. Parameter tuning is needed to get better results from deep learning models.

Each procedure takes a long time, regardless of the method. We used a grid and search method with a learning rate of 1e-3 for the work in this paper. We started with a value and then used it to find the best parameter value. Additionally, we varied the window segments to find the ideal segment sizes for the datasets.

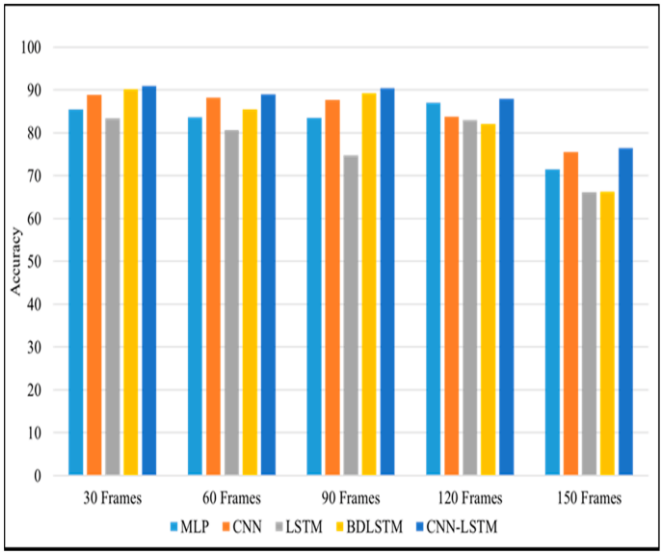

Fig-2: Various activity recognition by the models using CNN algorithm

4. CHALLENGES

Obtaining attributes and finding the most important characteristics is challenging.

- Time precision and complexity uses classification methods that have a different impact on time complexity and accuracy. as is frequently observed.
- Real-time information when a real-time dataset is used.
- Numerous activities When a person is engaged in multiple activities at once, it is difficult to identify them.
- Vision-based activity recognition: Due to crowds and live data streaming, activity recognition may be challenging. Identifying activities based on where they are although the Positioning System utilized to locate locations is challenging to locate locations installing
- Both excessive and inadequate can either over fit or under-fit when there is insufficient training data. As a result, the data must be incorporated into the implementation strategy.

5. CONCLUSION

The research conducted by HAR has significant strategic implications for healthcare surveillance, support for the elderly and people with cognitive disabilities, and surveillance. Her HAR can be enhanced and extended using deep learning, according to related studies. Classifying text, audio, and images for speech recognition with good results. However, they have not yet been incorporated into an

architecture for the multi-channel detection of human activity. This article presents a multichannel method for extracting features from multimodal sensor devices for activity detection. CNN layers of model map sensors are directly input and abstractly rendered. The forward and backward sequences of extracted features generated by the convolution process are utilized to their full potential in the internal state of a bidirectional LSTM layer. The proposed model was evaluated using two datasets that are accessible to the public. The model delivered was able to distinguish a variety of body parts and movements from a variety of complex activities.

REFERENCES

[1] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, ‘‘Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges,’’ Inf. Fusion, vol. 35, pp. 68–80, May 2017.

[2] N. Y. Hammerla, S. Halloran, and T. Ploetz. (2016). ‘‘Deep, convolutional, and recurrent models for human activity recognition using wearables.’’ [Online].

[3] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, and R. Jafari, ‘‘Enabling effective programming and flexible management of efficient body sensor network applications,’’ IEEE Trans. Human-Mach. Syst., vol. 43, no. 1, pp. 115–133, Jan. 2013.

[4] C. Xu, J. He, X. Zhang, C. Yao, and P.-H. Tseng, ‘‘Geometrical kinematic modeling on human motion using method of multi-sensor fusion,’’ Inf. Fusion, vol. 41, pp. 243–254, May 2017.

[5] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, ‘‘Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey,’’ in Proc. 23rd Int. Conf. Archit. Comput. Syst. (ARCS), Feb. 2010, pp. 1–10.

[6] Yuan, J., Mcdonough, S., You, Q., & Luo, J. (2013, August). Sentribute: image sentiment analysis from a mid-level perspective. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining* (pp. 1-8).

[7] Ana-Cosmina Popescu, Irina Mocanu, Bogdan Cramariuc, “Fusion mechanism using automated machine learning for movement detection”, IEEE Access, Volume 8, 2020,pp.143996-144014.

[8] Jaskirat Kaur & Williamjeet Singh ,“Tools, techniques, datasets and application areas for object detection in an image: a review”, Springer Publication, 2022,pp.38297–38351.

[9] Yuya Yoshikawa , Yutaro Shigeto, Akikazu Takeuchi ,“MetaVD: A Meta Video Dataset for enhancing human action recognition datasets”, Elsevier Computer Vision and Image Understanding, 2021,pp.1077-3142.

[10] Enoch Arulprakash , Martin Aruldoss ,“A study on generic object detection with emphasis on future research directions”, Journal of King Saud University Computer and Information Sciences Elsevier,2021,pp.7347–7365.

[11] Shubham Shindea, Ashwin Kotharia , Vikram Gupta b ,“YOLO based Human Action Recognition and Localization”,International Conference on Robotics and Smart Manufacturing (RoSMa) Procedia Computer Science,2018,pp.831–838.

[12] Sarita Chaudharya, Mohd Aamir Khana, Charul Bhatnagara,“Multiple Anomalous Activity Detection in Videos”,International Conference on Smart Computing and Communications, ICSCC Elsevier, 2017,pp.336–345.

[13] U. Anithaa, R.Narmadhab , D. Raja Sumanthc, D. Naveen Kumar c “Robust Human Action Recognition System via Image Processing”,International Conference on Computational Intelligence and Data Science (ICCIDS) Elsevier Procedia Computer Science, 2019,pp.870-877.

[14]Xiaohong Han, Jun Chang, Kaiyuan Wang, “Real□time object detection based on YOLO-v2 for tiny vehicle object”, International Conference of Information and Communication Technology (ICICT)Elsevier,2020,pp.61-72.

[15] Oumaima Moutika , Smail Tigani , Rachid Saadanec , Abdellah Chehri ,“Hybrid Deep Learning Vision-based Models for Human Object Interaction Detection by Knowledge Distillation”, International Conference on Knowledge-Based and Intelligent Information&EngineeringSystems,Elsevier,2021,pp. 5093–5103.

[16]Shruthi a, Prakash Pattan b, Shravankumar Arjunagi c,“A human behavior analysis model to track object behavior in surveillance videos”, Elsevier,2022,pp.100454-100459.

[17] Y. Abramson and Y. Freund, ‘‘SEmi-automatic VIsuaL LEarning (SEVILLE): A tutorial on active learning for visual object recognition,’’in Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2005.

[18] H. H. Aghdam, A. Gonzalez-Garcia, A. Lopez, and J. Weijer, ''Active learning for deep detection neural networks,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 3672–3680.

[19] D. Arpit, S. Jastrzebsk, N. Ballas, D. Krueger, E. Bengio, M. S. Kanwal,T. Maharaj, A. Fischer, A. Courville, Y. Bengio, and S. Lacoste-Julien,''A closer look at memorization in deep networks,'' in Proc. Int. Conf. Mach. Learn. (ICML), 2017, pp. 233–242.

[20] M.-F. Balcan, A. Blum, and K. Yang, ''Co-training and expansion: Towards bridging theory and practice,'' in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2004, pp. 89–96.