International Journal of Modern Agriculture
ISSN: 2305-7246
Volume 9 Issue 3, 2020

EnhancedCNN-LSTM approach for Human Activity Recognition

S. Syed Husain[1], B. Natarajan[2]
[1,] Assistant Professor, Department of Electronics and Communication Engineering
K. Ramakrishnan College of Engineering, Tiruchirappalli, Tamil Nadu, India
[2,] Research Scholar, School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India
[1]apsyedhusain@gmail.com
[2]natarajan@cse.sastra.ac.in

Abstract-Research in Deep Learning (DL) based computer vision techniques attracted wider attention in the research area. Human activity recognition by digital devices still lacks to perform well in different activity environments. To address this challenge, we proposed the hybrid framework called Enhanced CNN-LSTM approach for the Human Activity Recognition model. In recent days, huge theft and violent activities are happening everywhere, unable to find the culprit even though many technologies exist. The proposed deep learning model monitors and detects human activity through the hybrid approach, which combines both Convolutional Neural Network for spatial and temporal feature extraction and finding the discriminative feature identification using the technique Long Short Term Memory algorithm. The technique LSTM aims to predict the sequences. The experiments are conducted on the UCF10 Human Activity Recognition datasets, which results in improved performance of activity recognition.

***Keywords*:**Convolutional Neural Network;Activity Recognition;Long Short-TermMemory

I. INTRODUCTION

In day-to-day life, human beings are doing their activities like brushing, bathing, eating, driving, and sleepingon a routine basis. From the infant stage itself, every human learns the environment through the eyes. Based on the activities observed he/she reacts. To implement and recognize the same activity recognition and generation process by digital devices, we need to instruct the computer to do. Advanced Artificial Intelligent Techniques were adopted to address this task. The proposed Enhanced CNN-LSTM Taxonomy addresses the various challenges faced in this area. Computer vision-based methods efficiently handle images and videos to classify them. The proposed novel Enhanced CNN-LSTM method detects discriminative features of actions in video sequences to identify the activity of humans.

The human activity recognition (HAR) model is a complex task based on classification. HAR is a way of finding the activity of humans in video sequences.It involves tracking and monitoring the activity of human daily activities.The data set UCF 101 comprises various classes of activities that are routinely performed by common people in their daily lives. The proposed model uses deep learning algorithms to handle huge amounts of data efficiently.

Activity recognition is the problem of tracking the activity of a person, often inside the home, based on images and videos captures through camera devices. CNN and LSTM methods process the

image and video data to find the patterns of images to classify the activity related to it. Disparate DL techniques used tremendously in various real-time applications like activity recognition, detecting vehicles in traffic signals, finding people using CCTV camera videos. This framework helps to monitor and track the activities of the child and elder care people.

The proceeding sections of the paper arearranged as follows. The existing methodologies for human action recognition are discussed in section 2. The proposed framework called Enhanced CNN-LSTM approach for Human Action Recognition model is discussed with system architecture and algorithms in section 3. Section 4 discusses the outcomes of the proposed model and section 5 deals conclusion.

II. RELATED WORKS

2.1 Feature Extraction using Convolutional Neural Network

CNN algorithm is a widely used successful DL architecture for analyzing images and videos, which performs a mathematical operation called convolution on neural networks. The convolution operation is performed similarly to matrix multiplication on image pixels. The activation function Relu is commonly used on CNN. CNN is a multilayer approach consists of various layers with kernels as filters. The first layer is the convolution layers called the input layer, and thenthe hidden layers arefully connected, the normalization layer,and output layers. The number of hidden layers can be increased based on the feature extraction information needed. CNN performs pooling operation on image data to reduce dimensions of Image pixels. Three types of pooling operations are the first one Max pooling, the second one average pooling, andMin pooling is the final one. Max pooling is computed by consideringthe maximum pixel value in the image matrix, min pooling considers minimum value, and average pooling is calculated by averages of all the values in the grid. In the end,the Softmax function is performed for classifying imagesusing probability estimation of values lies between 0 to 1.To remove noise in live stream data, we need to average the nearby pixel values using the weighted function w(a). The equation (2) denotes convolution operation [1].

$$s(t) \int x(a)w(ta)da \quad (1)$$

$$s(t)=(x*w)(t) \quad (2)$$

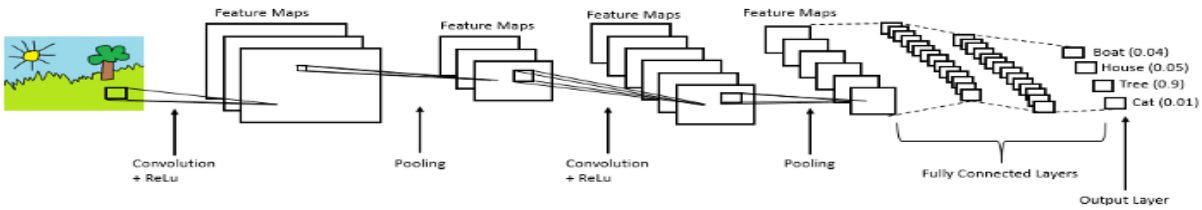

Figure 1. CNN Architecture

2.2 Predicting the sequence of activity classes through Long Short Term Memory

The proposed model uses the LSTM algorithm to inherit the concept ofa self-loop sequencemodel,a special kind of RNN. LSTM capable of remembering long-term sequences. Layers namely remember, sigmoid, and forgot helps to achieve this. The forgot layer identifies the old information which is not required and thrown it away LSTM can be used tremendously by finding missing letters in the word, sentence formation through hand actions, and prediction of upcoming words while typing [1].

2.3 UCF101 - Action Recognition Data Set

The UCF101dataset [7] consists of 3320 videos, which are taken from YouTube channels to describe the daily activities of human, grouped into 101 classes. These videos are considered with all background information and illumination conditions, quality of videos, and RGB color levels. The various activities include applying eye makeup, brushing, applying lipstick, bowling, biking, typing, cutting the vegetables in the kitchen, writing on the board, and so on.

Fig. 2 UCF 101 Dataset of Human Activity Recognition

Human Action Recognition (HAR) research received high attention in reach area because of high usefulness in healthcare and eldercare [2]. Accurate detection of human action plays a very important role in designing a framework. To model the framework [3] used the probabilistic model HMM - Hidden Markov Model for action recognition and structure prediction is done by using CRF - Conditional Random Field Method. The various challenges posed by authors are recognizing the concurrent activities of human, interleaved activities, and ambiguity of action interpretation. Besides all, it has to consider the overlapping of actions, epenthesis movements [4-6].

Eunju Kim et al discuss the various challenges present in HAR which are needed to be addressed while implementing a framework for action recognition which are recognizing parallel activities and independent activities, the ambiguity of action gestures, and different living environments[8]. The author also proposed various taxonomies using HMM and CRF. This framework faces the complexity of handling automatic labeling processes and managing different

sensor environments. [9]. The author [10] discusses various types of HAR activities which includes radio-based, camera-based, and wearable device based methods.

Haoxi Zhang et al proposed the HAR framework using a multi-head CNN and attention mechanism for feature extraction and selection. This model also learns effective features used for activity classification. This model needs to be improved for recognizing long term sequences.[10]

Yan et al implemented the HAR framework using a multitask classification algorithm. This model uses wearable cameras to monitor the daily activities of humans. This model needs to be improved for outlier detection of activities. [11]

Majdi Rawashdeh et al proposed the activity recognition model using activity profiling and implemented the model using the Naive Bayes, SVM, and J48 Classifiers. This model handles the pre-processing tasks like filtering the noisy information, cleaning the irrelevant information, and overlapping of activities.[12]

Xiaoran Shi et al [13] proposed DL Methods using spectrograms for classification to find HAR. This model uses Deep Convolutional Generative Adversarial Network (DCGAN) for augmenting the training data to avoid overfitting problems and Deep CNN for feature extraction and classification.

Mohd Halim Mohd Noor et al proposed the activity recognition framework using unsupervised deep learning methods by classifying the public data set into three groups as dynamic, static, and transitional.[14]

III. PROPOSED SYSTEM

The proposed framework is called Enhanced CNN-LSTM approach for Human Activity recognition model user for monitoring and detecting human activities. The proposed system uses the dataset UCF 101 consists of a video count of 13,320 for disparateactivitiesmapped with 101 different classes. To recognize the activities of humans the proposed model is utilized CNN, and RNN- LSTM deep learning algorithms. This section discusses the algorithms for the CNN_LSTM approach along with the system architecture of the proposed model.

Algorithm 1

Labeled Videos $V_k = \{V_1, V_2, V_3 \ldots V_k\}$
1 Read input from camera devices/Videos
2 Convert the input into frames of sequences
3 Investigate spatial and temporal information
4 Extract Manual and non-manual features
5 Reshape the frame size
6 Perform gloss level annotations
7 Map the frames with training frames
8 Apply CNN Algorithms (ReLu + Softmax)
9 Apply LSTM for recognizing larger sequences
10 Evaluate the accuracy of the model

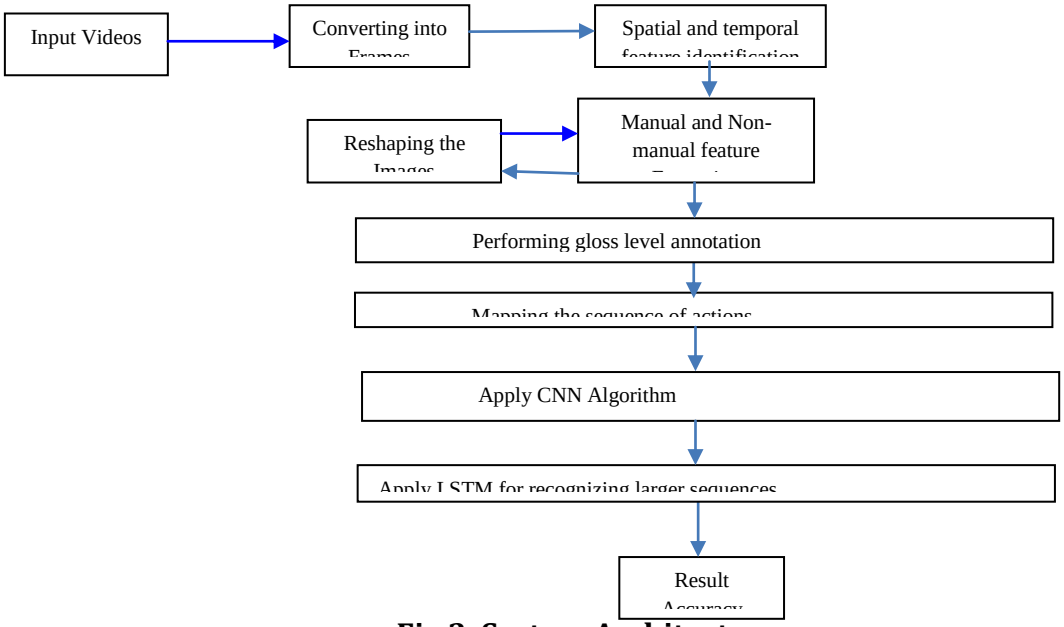

Fig 3. System Architecture

The data set UCF101 – Action Recognition Data Set is partitioned into two groups for training as well astesting. The dataset spitted into 70% for testing and 30% for training. Based on the training performance of the model, the evaluation of the testing dataset is done and this validation test results in good recognition accuracy.

IV. EXPERIMENTAL RESULTS

The experiments on UCF 101 data were conducted in Workstation CPU with operating system version Windows 10 Pro for a 64-bit machine, Intel Xeon Processor, and RAM 128 GB Environment. The dataset was randomly divided into 70% for training purposes and 30% for testing purposes.

The videos are converted into images, totally of 73,844 are generated with the size of (224, 224, 3). Next step, we have used the VGG-16 pre-trained model for the validation set.The dataset comprises of 1,000 classes are used for training the proposed model. Then the reshaped images are passed into the CNN-LSTM Architecture for feature extraction and identifying the activity. The categorical cross-entropy loss function was employed and the optimizer is Adam. The proposed

The model uses the ReLu activation function and the Softmax layer for finding the probability of prediction. The LSTM portion of the proposed model finds the activity embedded with the sequences of the Images. The proposed model achieves significant recognition accuracy compared with existing

approaches.

V. CONCLUSION

The existing techniques for human activity recognition depend on machine learning models which incur high computation time results in less recognition accuracy. Hence, Enhanced LSTM is proposed in this work. The Enhanced CNN_LSTM technique is used for extracting features and

classifying the type of activity performed by the human. Here, the LSTM techniques are powerfully used for finding the sequences of activities. Hence, this framework provides better recognition accuracy than conventional Machine Learning techniques. When compared to existing techniques, this proposed system provides better recognition accuracy and consumes less computation time.

References

[1] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, p. 2). Cambridge: MIT Press.
[2] Davide Anguita et al. 2013. A Public Domain Dataset for Human Activity Recognition using Smartphones. In ESANN
[3] Allan Stisen et al. 2015. Smart Devices Are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. ACM, 127–140.
[4] Ma, Yuchao, et al. "Transfer Learning for Activity Recognition in Mobile Health." arXiv preprint arXiv:2007.06062 (2020).
[5] Chen, Kaixuan, et al. "Deep learning for sensor-based human activity recognition: overview, challenges and opportunities." arXiv preprint arXiv:2001.07416 (2020).
[6] Khurram Soomro, Amir Roshan Zamir and Mubarak Shah, UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild., CRCV-TR-12-01, November, 2012.
[7] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012.
[8] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A PublicDomain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposiumon Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.
[9] Kim, E., Helal, S., & Cook, D. (2009). Human activity recognition and pattern discovery. IEEE pervasive computing, 9(1), 48-53.
[10] Zhang, H., Xiao, Z., Wang, J., Li, F., & Szczerbicki, E. (2019). A Novel IoT-Perceptive HumanActivity Recognition (HAR) Approach Using Multihead Convolutional Attention. IEEE Internet of Things Journal, 7(2), 1072-1080.
[11] Yan, Y., Ricci, E., Liu, G., & Sebe, N. (2015). Egocentric daily activity recognition via multitask clustering. IEEE Transactions on Image Processing, 24(10), 2984-2995.
[12] M. Rawashdeh, M.G. Al Zamil, S. Samarah, M.S. Hossain, G.Muhammad, A knowledge-driven approach for activity recognition in smart homes based on activity profiling, Future Generation Computer Systems (2017), https://doi.org/10.1016/j.future.2017.10.031
[13] Shi, Xiaoran, Yaxin Li, Feng Zhou, and Lei Liu. "Human activity recognition based on deep learning method." In 2018 International Conference on Radar (RADAR), pp. 1-5. IEEE, 2018.
[14] Noor, M. H. M., Ahmadon, M. A., & Osman, M. K. (2019). Activity Recognition using Deep Denoising Autoencoder. In 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE) (pp. 188-192). IEEE.