HUMAN ACTIVITY RECOGNITION USING DEEP CONVOLUTIONAL NEURAL NETWORK

[1]Velantina.V , [2]Dr.S.N.Chandrashekara

[1]Student , [2]Head of Department Computer Science and Engineering
C. Byre Gowda Institute of Technology, Kolar, India

Abstract : ***As the technology is emerging Computers are getting better at solving some very complex problems and advances in computer vision. In the face of extremely fast-growing data with no obvious laws, the traditional feature engineering methods are becoming more and more compact. With the development of Deep Learning technology we do not need to manually extract features, by migrating deep neural network experience in image recognition a deep learning model for human activity Recognition based on video input and training a model to predict the actions performed and achieve better accuracy. Human action recognition is necessary for various computer vision applications that demand information about people's behavior, including surveillance for public safety, human computer interaction applications, and robotics. The proposed method shows consistent superior performance and has good generalization performance, when compared with state-of-the-arts.***

IndexTerms **- Deep Learning, Activity Recognition, Computer Vision, Machine Learning.**

I. INTRODUCTION

In recent years, automatic human activity recognition has drawn much attention in the field of video analysis technology due to the growing demands from many applications, such as surveillance environments, entertainment and healthcare systems. In a surveillance environment, the automatic detection of abnormal activities can be used to alert the related authority of potential criminal or dangerous behaviors. Similarly, in an entertainment environment, the activity recognition can improve the human computer interaction. Furthermore, in a healthcare system, the activity recognition can help the rehabilitation of patients, such as the automatic recognition of patient's action to facilitate the rehabilitation processes. There has been numerous research efforts reported for various applications based on human activity recognition, and healthcare applications. This proposed model uses deep learning for human acclivity Recognition based on video input - given a set of labelled videos, train a model so that it can give a label/prediction for a new video. Here, the label might represent what is being performed in the video, or what the video is about. Human action recognition is necessary for various computer vision applications that demand information about people's behavior, including surveillance for public safety, human computer interaction applications, and robotics. Due to high recognition accuracy and easy deployment, video-based human action recognition techniques have got more research attention and been widely applied into lots of industrial applications.

II. LITERATURE SURVEY

[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 behavior, 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---8x 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.

III. PROPOSED SYSTEM

A deep learning model is developed to detect a person actions and emotion like walking, jogging, running, boxing, hand waving and hand clapping using Convolutional neural network. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

The dataset contains 599 videos – 100 videos for each of the 6 categories .With these dataset we will perform following process:

- Downloading, extracting and pre-processing a video dataset.
- Dividing the dataset into training and testing data.
- Create a neural network and train it on the training data.
- Test the model on the test data.

The data is obtained from the online source. Further we will resize the image for future use. Image resizing, or image scaling, is a geometric image transformation which modifies the image size based on an image interpolation algorithm. The video dataset contains six types of human actions (boxing, handclapping, hand waving, jogging, running and walking) performed several times by 25 different subjects in 4 different scenarios - outdoors *s1*, outdoors with scale variation *s2*, outdoors with different clothes *s3* and indoors *s4*. The model will be constructed irrespective of these scenarios.

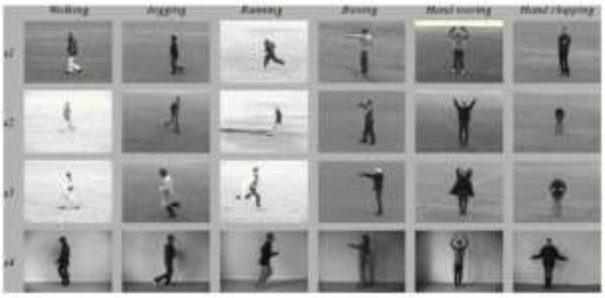


Fig 1. Frames of each actions walking, jogging, running, boxing, hand waving and handclapping.

There are a total of 6 categories - boxing, handclapping, hand waving, jogging, running and walking. One of the most important part of the project was to load the video dataset and perform the necessary pre-processing steps. A class (Videos) that had a function called (read_videos()) that can be used to for reading and processing videos. We were able to build an artificial convolutional neural network that can recognize images. Split the dataset into train and test dataset and Finally train the model using training dataset. Multiple convolutional and pooling layers are stacked together.

The aim of the project is to create a model that can identify the basic human actions like running, jogging, walking, clapping, hand-waving and boxing. The model will be given a set of videos where in each video, a person will be performing an action. The label of a video will be the action that is being performed in that particular video. The model will have to learn this relationship, and then it should be able to predict the label of an input (video) that it has never seen. Technically, the model would have to learn to differentiate between various human actions. These are followed by some fully-connected layers, where the last layer is the output layer. The output layer contains 6 neurons (one for each category). The network gives a probability of an input to belong to each category/class. Finally, the model that performed the best on the validation data is loaded. Once the model has been trained it is possible to carry out testing. During this phase a test set of data is loaded. This data set has never been seen by the model and therefore its true accuracy will be verified. Finally, the saved model can be used in the real world. The weights of the model which gave the best performance on the validation data were loaded. The model was then tested on the test data and the results were obtained.

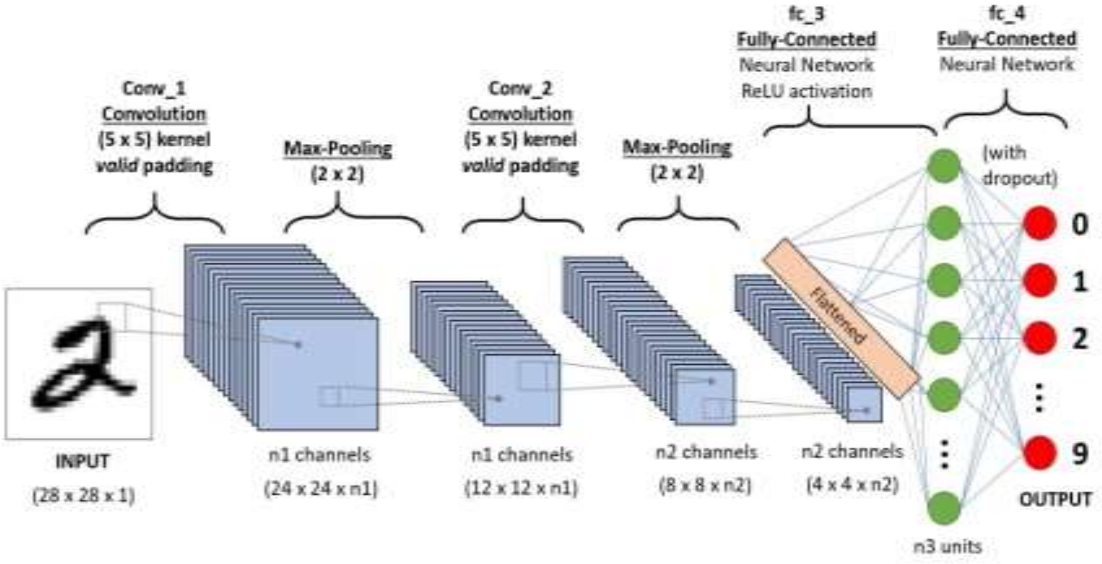


Fig 2. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) is networks specially designed to deal with images or more generally with translation invariant data. Current applications include a wide variety of image classifiers, CNNs are beginning to be reconsidered as a good alternative to recurrent networks when using sequential data. The feature extraction technique is applied on the dataset the train and test data are obtained. The model is trained with train data and test data and the test data is used to validate the model .The algorithm used is CNN as the model is trained and tested the results are obtained .The action performed is predicted and the performance, accuracy is achieved.

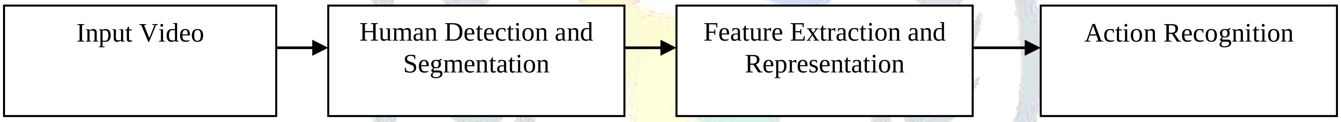
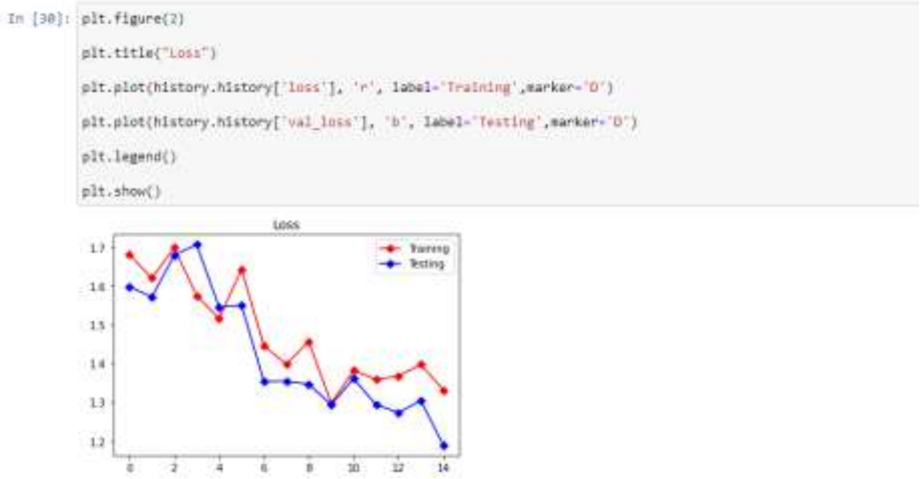

Input Video
Human Detection and Segmentation
Feature Extraction and Representation
Action Recognition

Fig 3. Human Activity Recognition Model

The Fig 3 describes the steps involved in human action recognition model. The input video is given as input to the model once the video is processed the detection of human and segmentation of frames is performed. Then the features are extracted from the model finally the activity is recognized and this model works efficiently in detecting actions. The activation function to be used for processing each layer. ReLU is proven to work best with deep neural networks because of its non-linearity, and its property of avoiding the vanishing gradient problem. The weights of the model which gave the best performance on the validation data were loaded. The model was then tested on the test data.

IV. Result



```
In [30]: plt.figure(2)

         plt.title("Loss")

         plt.plot(history.history['loss'], 'r', label='Training',marker='D')

         plt.plot(history.history['val_loss'], 'b', label='Testing',marker='D')

         plt.legend()

         plt.show()
```

Fig 4. A graph showing the accuracy of training and testing data of Human Activity Recognition Model.

The activity recognition model achieved a maximum accuracy by detecting the actions when the input is loaded from the dataset. Performance analysis is a type of operational analysis that consists of a collection of fundamental quantitative relationships between performance variables. Each video was given as an input to the model, the model processed each frames in the video and the action was detected and resulted the output. The above graph shows that the train and test data were evaluated by which the model achieved a good performance.

V. Conclusion

In this paper, we conceptually proposed a model for human activity recognition in videos by using convolution neural network algorithm, the model is trained and tested on different inputs. The dataset consist of activities performed by human, as each video is processed in individual frames the activity is recognized. The proposed model shows superior performance and has good generalization performance on the used dataset. For our future work direction, we may explore further the problem of data imbalance in real-life human activity recognition application.

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]. Available: https://arxiv.org/abs/1604.08880

[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] J. Margarito, R. Helaoui, A. M. Bianchi, F. Sartor, and A. G. Bonomi, ‘‘User-independent recognition of sports activities from a single wristworn accelerometer: A template-matching-based approach,’’ IEEE Trans. Biomed. Eng., vol. 63, no. 4, pp. 788–796, Apr. 2016.

[7] P. C. Roy, S. Giroux, and B. Bouchard, ‘‘A possibilistic approach for activity recognition in smart homes for cognitive assistance to Alzheimer’s patients,’’ in Activity Recognition in Pervasive Intelligent Environments. Paris, France: Atlantis Press, 2011, pp. 33–58.

[8] A. Bux, P. Angelov, and Z. Habib, ‘‘Vision based human activity recognition: A review,’’ Advances in Computational Intelligence Systems. Cham, Switzerland: Springer, 2017, pp. 341–371.

[9] Z. Chen, L. Zhang, Z. Cao, and J. Guo, ‘‘Distilling the knowledge from handcrafted features for human activity recognition,’’ IEEE Trans. Ind. Informat., vol. 14, no. 10, pp. 4334–4342, Oct. 2018. [10] S.-R. Ke, H. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, ‘‘A review on video-based human activity recognition,’’ Computers, vol. 2, no. 2, pp. 88–131, 2013.

[10] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770–778.

[11] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso, ‘‘Preprocessing techniques for context recognition from accelerometer data,’’ Pers. UbiquitousComput., vol. 14, no. 7, pp. 645–662, 2010. [13] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, ‘‘Learning precise timing with LSTM recurrent networks,’’ J. Mach. Learn. Res., vol. 3, pp. 115–143, Aug. 2003.

[12] F. J. Ordóñez and D. Roggen, ‘‘Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,’’ Sensors, vol. 16, no. 1, p. 115, 2016.