Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume7-Issue1, January 2024
Paper: 12

HUMAN ACTION RECOGNITION SYSTEM USING TRANSFER LEARNING SYSTEM

[1] S. SARANYA, Abarna. S[2], Dhivya Bharathi. R[3], Gayathri. B[4],Kavya. A[5.]

[1]Assistant Professor, Department of Computer Science and Engineering,
K.Ramakrishnan College of Engineering,
Trichy.

[2,3,4,5]UG Scholar, Department of Computer Science and Engineering,
K.Ramakrishnan College ofEngineering,

Trichy.

saranyasambathraj@gmail.com[1],abarnaselvaganapathy@gmail.com
[2],rdhivyabharathi2001@gmail.com [3],gayugayugayugayu36@gmail.com[4],
kavyaaruna2001@gmail.com[5]

ABSTRACT:

The dififcult issues in computer vision is the recognition of human activities. It is more challenging to identify human activity from still photographs than from sensor-based or video-based methods since there is no timing information available. Recently, many deep learning-based solutions have been put forth one after another, and their effectiveness is continually growing. By assembling multi-channel attention networks that are based on transfer learning, we suggested a convolutional neural architecture in this research. In this case, feature fusion-based ensembling was created using four CNN branches, and in all sector, an attention module was employed to retrieve the data from the feature map created by the existing pre-trained models. In order to obtain the final recognition output, the derived feature maps from the sectors were combined and given into a fully connected network.

INTRODUCTION:

An action is a thing that may be seen by using either the eye or a sensors, according to Human Activity Recognition (HAR) [1]. For example, walking is an action that need constant monitoring of the person in your range of vision.Body parts are devided by activities : gesture, action, interaction, and group activity [2].In order to achieve nonverbal communication, gestures use the movement of the hands, face, or other body parts. Human motions like running, walking, jumping, crawling, etc. are referred to as actions. A human's actions with an object or another human are referred to as interaction. When several people interact with various items through gestures, actions, and interactions, this is referred to as group activity. In the past few years, HAR has the result of intensive research in the fields of computer view and pattern recognition., which has since gained traction as a hot scientific topic.

Depending one type of data input, HAR may be broadly classified into two types: sensor-

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 12

type HAR and vision-type HAR [3], [4].

Additionally, they were distinct types of vision-type HAR: video-based HAR and image-type HAR. While vision-based HAR examines images or videos taken by optical components, sensor-type HAR examines natural data from bio-sensors and remote monitor . Wearable gadgets are prime examples of sensor-type HAR since they are wornned by users because it itself detect and measure a variety of movements like sitting, walking, jumping, relaxing . A sensor won't work if anyone is either outside of its coverage area or engaging in unusual behaviour .

On the other hand, HAR systems that rely on vision have long utilised CCTV systems [5].Vision based har is less expensive and easy to build in the way its is very much felexible compared to the other sensor type har .The sensor type har are much complex compared to the other har.

This system was introduced in the way that it can help to identify many activities in the different fields there as many possible apllication that it can cover by its specific benefits which is very much easy to implement in many fields.now a days any one can build this system and can use it over wide range and it has large scope in many fileds.

Over the past ten years, a lot of research has been done on HARs that use video, and the results are consistently excellent .There will be number of features contained in a single frame of a video which is captured, Each will provide different informations in based on the video.. Or, to put it another way, in image-based HARs,it is very diffiult to classify the activity of human beings in a oneimage captured by the system. Many researchers not seen at this particular behavioural domain because the data itself is not enough to identify the information which was required to identify the process. vision based har is less expensive and easy to build in the way its is very much flexible compared to the other sensor type har are much complex compared to the other har.

With little input data, traditional machine learning approaches have worked admirably in tightly regulated environments. However, they do necessitate laborious hand engineering and lengthy pre-processing operations, which are difficult and time-consuming. [6].Usage of systems with low performance is not applicable to the users with great expectancy of usage. However, following the release of AlexNet in 2012, attention has keen on the development of convolutional neural networks and the production [7]-[8]. Strong characteristics from raw photographs can now be extracted thanks to convolutional neural networks like VGG [9] and Inception [10], which produce excellent results and continue to boost recognition accuracy in huge. CNN's impressive performance, deep CNN training has overfitting problems because action recognition lacks considerable amounts of labelled data.

RELATED WORKS:

This system was introduced in the way that it can help to identify many activities in the different fields there as many possible application that it can cover by its specific benefits which is very much easy to implement in many fields. Nowadays any one can build this system and can use it over wide range and it has large scope in many fields. The most widely used framework for action recognition in still pictures is known as the BoW framework [11]. A typical

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 12

architecture gathers traits from the entire image and represents them as histograms. However, features will always be present in the image if it contains backdrops or items that is not related to the action, decreasing the classification accuracy. In the past, [12] classified human activities using a support vector machine classifier and a bag of words.

They concentrated on a key indicator in finite mixture models, the estimate parameters given a full covariance matrix[13].they developed smart study method that blends an expectation maximisation strategy with a fixed-point covariance matrix estimator. [14] made the argument that insignificant objects and backgrounds are easily misconstrued based on their appearance. Additionally,unique mask degradation was used to easily identify the feature map activation to the intended person performing the activity, doing away with deceptive context activation. According to experiments performed by [15], the problem with still image-based HAR is the absence of time info.To address this issue, they put out a special visual representation that both accurately depicts the subject's anticipates. They used a transfer learning-based technique.

SUGGESTIVE ARCHITECTURE:

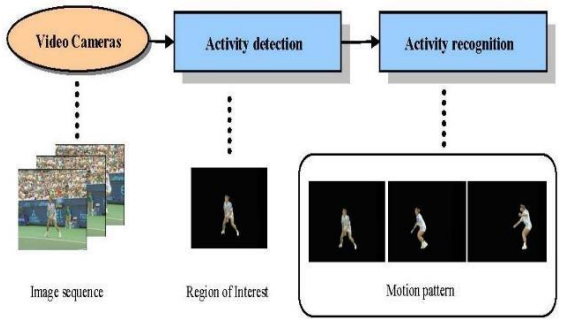

Fig:1, Diagrammatical representation

An ensembled multi-channel convolutional neural architecture to recognise human actions has been proposed in this research. Many pre-trained convolutional neural networks were involved throughout the architectural design process.

Pre trained models was given input photos in order to find significant features. Second, after employing every pre-trained architecture, a channel attention module first suggested by squeeze and excitation networks [16] was used.

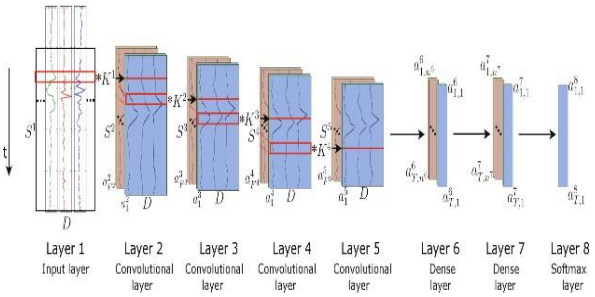

Fig:2 Process

The pre-trained feature maps' channels' weights can be adaptively adjusted by the attention module to select more potent features for type differentiation. Due to the great clear of view, it can detect information by capturing the image's global context.

MEASURESOF PERFORMANCE:

Performance indicators are present in every machine learning workflow.

The effectiveness of classification can be evaluated in many different ways. Accuracy is measured as the proportion of all the data points that were correctly predicted.

Consequently, using a formula, precision can be calculated (2).

Precision equals TP + TP + FP (2)

A confusion matrix offers an overview of the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume7-Issue1, January 2024
Paper: 12

results of the classification problem predictions..

DISCUSSION:

Performance improved as a result of four key factors. First off, because the models werepre-trained, there was a decrease in the amount of training are done to identify the important patterns. Second, employing a strong attention mechanism guarantees that the salient elements will endure and the less salient features will be eliminated. At last, the performance was improved even more thanks to the ensembled feature fusion technique that was suggested.

CONCLUSION:

In the past, numerous researchers have tested sensor-type and video-type HARs. The dearth of photographs for each class and the complexity of human being activities are two additional significant obstacles to accurately identifying human activity from still images. In this article, we discussed this area of research and suggested multi-channel attention networks based on ensembled transfer learning for recognising human being actions in still photos. First, the "lack of photos per class" issue was addressed using four pre-trained models. Pre-trained models, wholely linked layers were added after an attention module.

REFERENCES

[1] P. Pareek and A. Thakkar, ‘‘A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications,’’ Artif. Intell. Rev., vol. 54, no. 3, pp. 2259–2322, Mar. 2021.

[2] J.K. Aggarwal and M.S. Ryoo, ‘‘Human activity analysis: A review,’’ ACM Comput. Surv., vol. 43, no. 3, pp. 1–43, 2011.

[3] Z.S.Abdallah, M.M.Gaber,B. Srinivasan, and S. Krishnaswamy, ‘‘Activity recognition with evolving data streams: A review,’’ ACM Comput. Surv., vol. 51, no. 4, pp. 1–36, 2018.

[4] S. Herath, M. Harandi, and F. Porikli, ‘‘Going deeper into action recognition: A survey,’’ Image Vis. Comput., vol. 60, pp. 4–21, Apr. 2017.

[5] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, ‘‘Robust human activity recognition from depth video using spatiotemporal multi-fused features,’’ Pattern Recognit., vol. 61, pp. 295– 308, Jan. 2017.

[6] I. Portugal, P. Alencar, and D. Cowan, ‘‘The use of machine learning algorithms in recommender systems: A systematic review,’’ Expert Syst. Appl., vol. 97, pp. 205–227, May 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 1097–1105.

[8] C. Patel, D. Bhatt, U. Sharma, R. Patel, S. Pandya, K. Modi, N. Cholli, A. Patel, U. Bhatt, M.A. Khan, S. Majumdar, M. Zuhair, K. Patel, S. A. Shah, and H. Ghayvat, ‘‘DBGC: Dimension-based generic convolution block for object recognition,’’ Sensors, vol. 22, no. 5, p. 1780, Feb. 2022.

[9] K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ 2014, arXiv:1409.1556.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ‘‘Going deeper with convolutions,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1–9.