iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IV Month of publication: April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50726

Vision Based Anomalous Human Behaviour Detection using CNN and Transfer Learning

Dr. K. Upendra Babu[1], P. Rani[2], P. Harshitha[3], P. Geethika[4], E. Nithya[5]
Bharath University

Abstract: With the advent of the Internet of Things (IoT), there have been significant advancements in the area of human activity recognition (HAR) in recent years. HAR is applicable to wider application such as elderly care, anomalous behavior detection and surveillance system. Several machine learningalgorithms have been employed to predict the activities performed by the human in an environment. However, traditional machine learning approaches have been outperformed by feature engineering methods which can select an optimal set of features. Onthe contrary, it is known that deep learning models such as Convolutional Neural Networks (CNN) can extract features and reduce the computational cost automatically. In this paper, we use CNN model to detect human activities from Image Dataset model. Specifically, we employ transfer learning to get deep image features and trained machine learning classifiers. Our experimental results showed the accuracy of 96.95% using VGG-16. Our experimental results also confirmed the high performance of VGG-16 as compared to rest of the applied CNN models.
Keywords: Activity recognition, deep learning, convolutional Neural network.

I. INTRODUCTION

Human activity recognition (HAR) is an active research area because of its applications in elderly care, automated homes and surveillance system. Several studies has been done on human activity recognition in the past. Some of the existing work are either wearable based or nonwearable based . Wearable based HAR system make use of wearable sensors thatare attached on the human body. Wearable basedHAR system are intrusive in nature. Non- wearable based HAR system do not require anysensors to attach on the human or to carry any device for activity recognition. Nonwearable based approach can be further categorised into sensor based and visionbased HAR systems.Sensor based technology use RF signals from sensors, such as RFID, PIR sensors and Wi-Fi signals to detect human activities.

Vision based technology use videos, image frames from depth cameras or IR cameras to classify human activities. Sensor based HAR system are non-intrusive in nature but may not provide high accuracy. Therefore, vision-based human activity recognition system has gained significant interest in the present time.Recognising human activities from the streaming video is challenging. Video-based human activity recognition can be categorised as marker-based and vision-based according to motion features. Markerbased method make use of optic wearable markerbased motion capture (MoCap) framework. It can accurately capture complex human motions but this approach has some disadvantages.

It require the opticalsensors to be attached on the human and also demand the need of multiple camera settings. Whereas, the vision based method make use of RGB or depth image. It does not require the userto carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR frameworksimple and easy to be deployed in many applications. Most of the visionbased HAR systems proposed in the literature used traditional machine learning algorithms for activity recognition. However, traditional machine learning methods have been outperformed by deep learning methods in recent time.

The most common type of deep learning method is Convolutional Neural Network (CNN). CNN are largely applied in 9 areas related to computer vision. It consists series of convolution layers through which images are passed for processing. In this paper, we use CNN to recognise human activities fromWiezmann Dataset.

We first extracted theframes for each activities from the videos. Specifically, we use transfer learning to get deepimage features and trained machine learning classifiers. We applied 3 different CNN modelsto classify activities and compared our results with the existing works on the same dataset.

In summary, the main contributions of our work are as follows: 1. We applied three different CNN models to classify human recognition activities and we showed the accuracy of 96.95% using VGG-16. 2. We used transferlearning to leverage the knowledge gained fromlarge-scale dataset such as ImageNet to the human activity recognition dataset.

II. METHODOLOGY

A. User

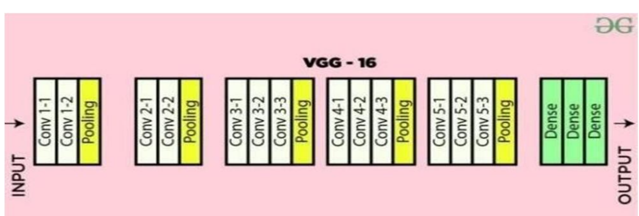
The User can start the project by running mainrun.py file. User has to give –input (Video file path).The open cv class Video Capture(0)means primary camera of the system, Video Capture(1) means secondary camera of thesystem. Video Capture(Vide file path) means with out camera we can load the video file from the disk. Vgg16, Vgg19 hasprogrammatically configured. User can change the model selection in the code and can run inmultiple ways.

B. HAR System

Video-based human activity recognition can be categorized as vision-basedaccording. The vision based method make useof RGB or depth image. It does not require theuser to carry any devices or to attach anysensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR framework simple and easy to be deployed in many applications. We first extracted the frames for each activities from the videos. Specifically, we use transfer learning to get deep image features and trained machine learning classifiers. HAR datasets are a vivid variety of qualities based upon their parameters, such as RGB, RGB-D(Depth), Multiview, recorded in a controlled environment. Other parameters are – recorded "In the wild," annotated with a complete sentence, annotatedwith only action label datasets, etc, such as thesource of data collection, number of actions, video clips, nature of datasets, and released year to show the progress in this area. We observe that 20 most of the HAR datasets could not become a popular choice among computer-vision researchers due to their over simplicity, small size, and unsatisfactory performance. However, there is no such thing as the most accurate standard datasets, i.e., on which researchers measure the HAR method to set as a benchmark, but of course, as we observe UCF101 and are the dominating datasets for researchers interest. Also, the actions played in the recorded clips are, by various individuals, while in other datasets, theactivities and actions are usually performed byone actor only.

C. VGG 16

VGG16 is a convolutional neural network model. Deep Convolutional Networksfor Large- Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 millionimages belonging to 1000 classes. It was one of the famous model submitted to ILSVRC- 2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU's. VGG-16 Architecture The input to the network is an image of dimensions(224, 224, 3). The first two layers have 64 channels of a 3*3 filter size and the samepadding. Then after a max pool layer of stride(2, 2), two layers have convolution layers of 128 filter size and filter size (3, 3). This is followed by a max-pooling layer of stride (2, 2) which is the same as the previous layer. Then there are 2 convolution layers of filter size (3, 3) and 256 filters. After that, there are2 sets of 3 convolution layers and a max pool layer. Each has 512 filters of (3, 3) size with the same padding. This image is then passed tothe stack of two convolution layers. In these convolution and maxpooling layers, the filterswe use are of the size 3*3 instead of 11*11 in AlexNet and 7*7 in ZFNet. In some of the layers, it also uses 1*1 pixel which is used to manipulate the number of input channels. 21 There is a padding of 1-pixel (same padding) done after each convolution layer to prevent the spatial feature of the image.



After the stack of convolution and maxpoolinglayer, we got a *(7, 7, 512)* feature map. We flatten this output to make it a *(1, 25088)* feature vector. After this there is *3 fully* connected layer, the first layer takes input fromthe last feature vector and outputs a *(1, 4096)* vector, the second layer also outputs a vector of size *(1, 4096)* but the third layer output a *1000* channels for *1000* classes of ILSVRC challenge i.e. 3rd fully connected layer is used to implement softmax function to classify 1000 classes. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problems.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

D. *Transfer Learning*

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre- trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skillthat they provide on related problems. In this post, you will discover how you can use transfer learning to speed up training and improve the performance of your deeplearning model.

Transfer learning is a method of transferring *knowledge* that a model has learned from earlier extensive training to the current model.The deep network models can be trained with significantly less data with transfer learning. Ithas been used to reduce training time and improve accuracy of the model. In this work, we use *transfer learning* to leverage the knowledge gained from large-scale dataset such as ImageNet. We first extract the frames for each activities from the videos. We use *transfer learning* to get deep image features and trained machine learning classifiers. For all CNN models, pre-trained weights onImageNet are used as starting point for transferlearning. ImageNet [6] is a dataset containing 20000 categories of activities. The knowledgeis transferred from pretrained weights on ImageNet to Weizmann dataset, since set of activities recognised in this work fall within the domain of ImageNet. The features areextracted from the penultimate layer of CNNs.

System Specification

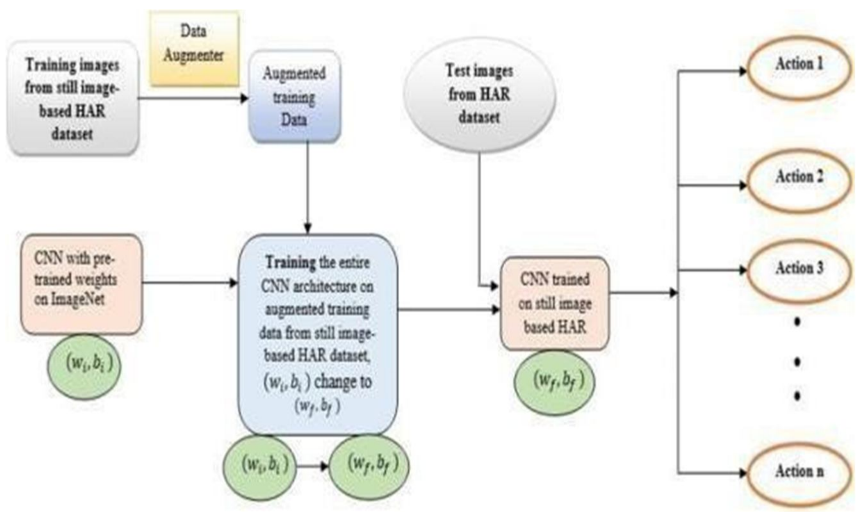


Figure 1: System Architecture

REFERENCES

[1] Hernandez, N.; Lundström, J.; Favela, J.; McChesney, I.; Arnrich, B. Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology. SN Comput. Sci. 2020, 1, 1–16. [CrossRef]

[2] Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.;Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning.Proc. IEEE 2020, 109, 43–76. [CrossRef]

[3] Deep, S.; Zheng, X. Leveraging CNN andTransfer Learning for Vision-based Human Activity Recognition. In Proceedings of the 2019 29th International Telecommunication Networks and Applications Appl. Sci. 2021, 11, 7660 26 of 28

[4] Casserfelt, K.; Mihailescu, R. Aninvestigation of transfer learning for deep architectures in group activity recognition. In Proceedings of the 2019 IEEE InternationalConference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan, 11–15 March2019; pp. 58–64.

[5] Alshalali, T.; Josyula, D. Fine-Tuning of Pre-Trained Deep Learning Models with Extreme Learning Machine. In Proceedings of the 2018 International Conference on Computational Science andComputational

[6] Cook, D.; Feuz, K.D.; Krishnan, N.C.Transfer learning for activity recognition: A survey Knowl. Inf. Syst. 2013, 36, 537–556. [CrossRef]

[7] Hachiya, H.; Sugiyama, M.; Ueda, N.Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. Neurocomputing 2012, 80, 93–101. [CrossRef]

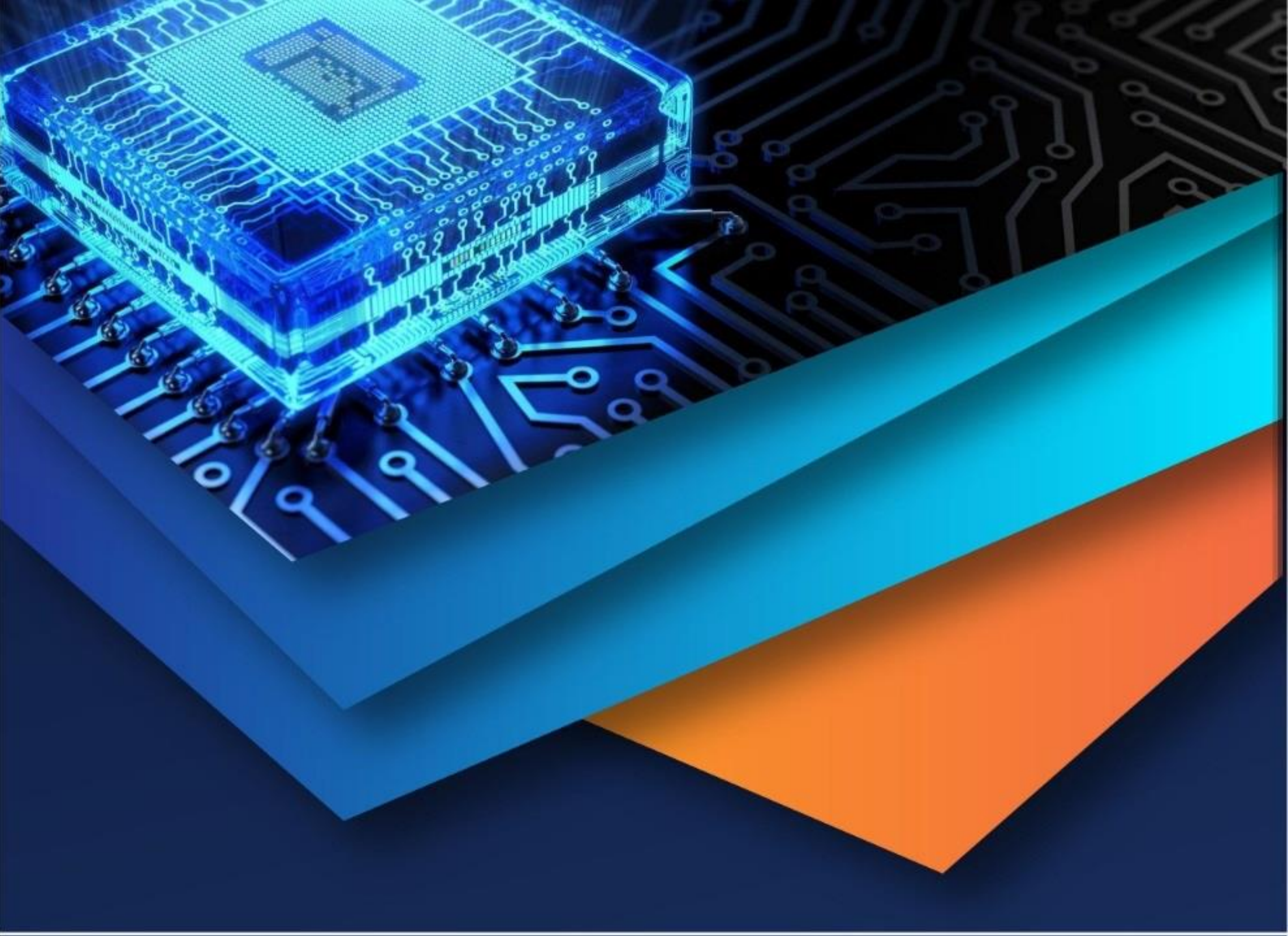
[8] van Kasteren, T.; Englebienne, G.; Kröse, B. Recognizing Activities in MultipleContexts using Transfer Learning. In Proceedings of theAAAI AI in Eldercare Symposium, Arlington, VA, USA, 7–9 November 2008. [

[9] Cao, L.; Liu, Z.; Huang, T.S. Cross-dataset action detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.

[10] Yang, Q.; Pan, S.J. A Survey on TransferLearning. IEEE Trans. Knowl. Data Eng. 2010, 22, 1345– 1359.

[11] Hossain, H.M.S.; Khan, M.A.A.H.; Roy, N.DeActive: Scaling Activity Recognition with Active Deep Learning. Proc.ACM Interact. Mob. Wearable Ubiquitous Technol. 2018, 2, 1–23.[CrossRef]

[12] Alam, M.A.U.; Roy, N. Unseen Activity Recognitions: A Hierarchical Active Transfer Learning Approach. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems .
[13] Civitarese, G.; Bettini, C.; Sztyler, T.; Riboni, D.; Stuckenschmidt, H. NECTAR: Knowledge-based Collaborative Active Learning for ActivityRecognition. In Proceedings of the 2018IEEE International Conference onPervasiveComputing andCommunications, Athens, Greece, 19–23March 2018.
[14] Civitarese, G.; Bettini, C. newNECTAR:Collaborative active learning for knowledge-based probabilistic activity recognition. Pervasive Mob. Comput. 2019,56, 88–105. [CrossRef]
[15] Wang, S.; Chang, X.; Li, X.; Sheng, Q.Z.; Chen, W. Multi-Task Support Vector Machines for Feature Selection with Shared Knowledge Discovery. Signal Process. 2016, 120, 746–753. [CrossRef]
[16] Feuz, K.D.; Cook, D.J. Collegial activitylearning betweenheterogeneous sensors. Knowl. Inf. Syst. 2017, 53, 337– 364. [CrossRef] [PubMed]
[17] Rokni, S.A.; Ghasemzadeh, H. Autonomous Training of Activity Recognition Algorithms in Mobile Sensors: A Transfer Learning Approach in Context-Invariant Views. IEEETrans. Mob. Comput. 2018, 17, 1764– 1777. [CrossRef]
[18] Kurz, M.; Hölzl, G.; Ferscha, A.; Calatroni, A.; Roggen, D.; Tröster, G. Real- Time Transfer and Evaluation of Activity Recognition Capabilities in an Opportunistic System. In Proceedings of the Third International Conference on Adaptive and Self-Adaptive Systems and Applications, Rome, Italy, 25– 30 September 2011.
[19] Roggen, D.; Förster, K.; Calatroni, A.; Tröster, G. The adARC pattern analysisarchitecture for adaptive human activityrecognition systems. J. Ambient. Intell. Humaniz. Comput. 2013, 4, 169–186.[CrossRef] [20] Calatroni, A.; Roggen, D.;Tröster, G. Automatic transfer of activity recognition capabilities



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY