



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Leveraging CNN and Transfer for Vision Based Human Activity Recognition

Kothamasu Kousik¹, Kilari Rajeev², Mrs.G.Anbu Selvi³

U.G. Student, Department of CSE, Satyabhama Institute of Science and Technology, Chennai, Tamil Nadu, India¹

U.G. Student, Department of CSE, Satyabhama Institute of Science and Technology, Chennai, Tamil Nadu, India²

Associate Professor, Department of CSE, Satyabhama Institute of Science and Technology, Chennai, Tamil Nadu, India³

ABSTRACT:—To recognition of human activity has advanced dramatically since the Internet of Things (IoT) was introduced recently. A wider range of uses for HAR exist, including monitoring systems, geriatric care, and unusual behaviour detectors. To forecast what a person would do in a specific circumstance, several machine learning techniques have been developed. However, methods for extracting features that choose the best set of features have proven to be more effective than Machine learning algorithms that are traditional. Convolutional Neural Networks and other deep learning models, on the other hand (CNN) are well known for their capacity to proactively extract features while attempting to reduce computing costs. In this study, we analyse the image dataset model and identify human activities using the CNN model. Transfer learning, in particular, is used to train classifiers that use machine learning.

KEYWORDS: Deep learnings, convolutional neural networks (CNN), and activity recognition are all terms that are related

I. INTRODUCTION

Acknowledgement of human activity is used in aged care, pc homes, and surveillance equipment (HAR) is a field of study. is currently being looked into. The recognition of human actions has already been the subject of numerous studies in the past. Some of the earlier work can be worn or cannot be worn. In HAR applications that depend on them, Portable sensors that are worn just on body are used. The nature of HAR systems based on wearables is intrusive. A non-wearable HAR system doesn't need a customer to accessorise any sensors or convey any human action recognition equipment..Nonwearable approaches also include device- and vision-based HAR systems. Sensor-based technologies use radio frequency (RF) signals from sensors to identify people, including barcode technology, PIR sensor, and Wi-Fi signals. activity. Video content, images from imaging devices, or Vision-based technology uses infrared cameras to classify human actions. Sensor-based HAR systems, while not intrusive, may not provide a superior level of accuracy. Like a result, devices which use vision to detect human activity are becoming more common. are currently receiving a lot of attention. It might be challenging to identify human behavior in streaming video. Marker-based and vision-based recognition system using Video can be classified into two types. based on motion features. A framework for electro - optic portable marker motion detection (MoCap) is used in marker-based techniques Although this technique can capture complex human movements with accuracy, it has a number of limitations. In addition to a number of camera settings, attaching sensing equipment towards the human body is also necessary. On the other hand, the vision-based method uses RGB or depth images. Users do not need to wear or carry any sort of sensing equipment. In an effort to make the HAR framework straightforward and straightforward to enforce in a wide range of situations, this scientific approach is gaining popularity. the most recent reaches for activity recognition. Among deep learning techniques, convolutional neural network (CNN) networks are most frequently used (CNN). In nine computer vision-related fields, CNNs are frequently used. It processes images using a wide range of convolutional layers. In this study, we use the Wieszmann Dataset to identify human activity using CNN. The videos' pixels for each activity were first extracted by humans. To teach classification model powered by machine learning and intensity picture features, transfer learning is specifically used. We used to categorise activities using With the identical dataset and three different CNN models, we compared our findings to earlier studies. In conclusion, the strengths-based methodology used in our work is as follows: 1. To define human recognition processes, we used 3 distinct CNN models, and the majority of the eyesight HAR processes that the research recommended used conventional algorithms for machine learning. On the other hand, deep

learning has proven to be more effective than conventional machine learning methods. With VGG-16, we had accuracy of 96.3 percent. We applied knowledge from enormous amounts of data like ImageNet to a set of data for characterising human behaviour using transfer learning.

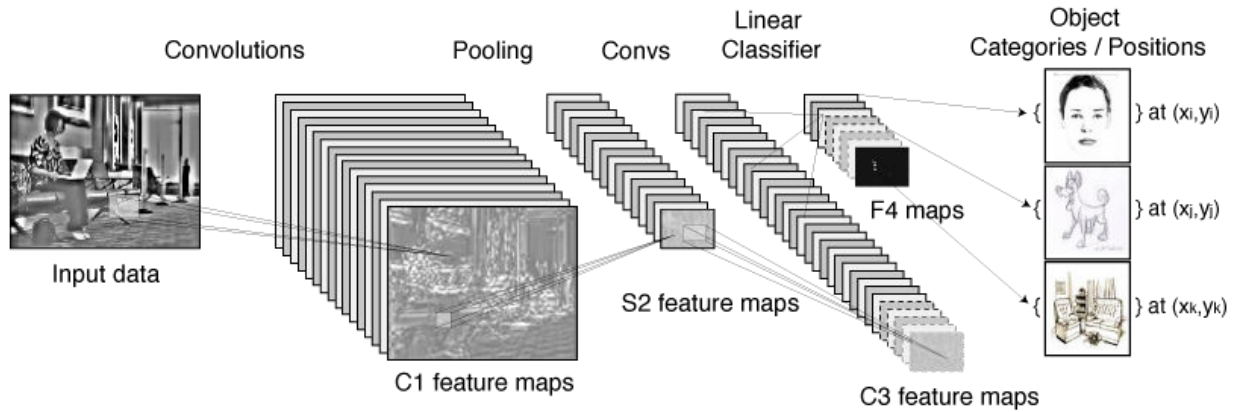
II. LITERATURE SURVEY

- [1] Summit The international accession number with this object is 19.9078016. It was created by Samundra Deep and Xi Zheng in the Department of Computing at Macquarie University in Sydney, Australia. The most recent revision of Networks and Applications in Telecom services on IEEE Xplore was made April 27, 2020. INSPEC R. Alazrai, B., M. Habbeh, and There's a conclusion at the end. Deep Learning for Human-to-Human Interaction Recognition Framework. With Wi-Fi signals, the created the end-to-end deep learning algorithm manages to achieve 86.3 reliability with all human companions recognition. Interactions in groups are not intended use of the proposed model. Additionally, this idea can only be used in interpersonal interactions. pp. 197695–197710 in IEEE Access, vol. 8, no. 2020.
- [2] V. Shitole, G. Thomas, G. Navada, T. Dobhal, Deep Learnings, and Numeric Motion Image are used to recognise human activity. The deep learning model provides a consistent detection accuracy with excellent precision in both 2D and 3D datasets when there are multiple people present in the 3D view. Procedia Computer Science, 58, p. 178–185. (2015). A powerful method for recognising KPCA outperforms A. Mohamed, M.M. Mohamad, M.Uddin, A. Almogrens by using built - in sensors and machine learning, using support vector machines (SVM), Artificial Neural Network (ANN). The support vector algorithm (SVM) and neural network models (ANNs) networks (ANN) are outperformed by KPCA (ANN). 81 (2018): 307–313 Future Generation Computer Systems.
- [3] S. Baskar, R. Janarthanan, and S. Doss The on-nodule wearable's enhanced unsupervised deep learning-aided recreated coder. Utilizing an unsupervised deep learning model enhances feature extraction and extraction for the sensor that detects human behavior. Performance suffers when dealing with big data sources and a wide range of human activities. International Journal of Measurements 164, 108050 (2020). Inception, J. Visumathi, A. Jeyanthi Suresh ResNet-based deep transfer Learning paradigm for long-term short-term memory for recognising human action, It has accuracy rate ratings of 92% and 91% for the different sets of data. It unavoidably calls for a protracted training period. Scientific Proc today (2020).
- [4] G. Cui, Y. Jia, Y. Guo, X. Zhong, Convolutional and neurocomputing-based non - linear and non and multi-domain human activity recognition. The addition of range maps to the established deep learning model improves detection performance by 1.3%. The proposed paradigm did not intend for group interactions. M. Farhan, N. Zehra, S.H. Azeem, and M. Have used ensemble learning across various convolutional neural networks to recognise human movement The proposed model seeks to promote automatic feature extractions, which reduces pretreatment time. For the purpose of recognizing concurrent activities, the model is inappropriate. CISS 2021, the 55th Annual Conference on Information Science (2021).
- [5] Long-term reliance and sparseness features learning for identifying human actions are provided by S. Ullah and D.H. Kim. Real-time data has poor precision. 2021

III. ALGORITHM

A. CNN

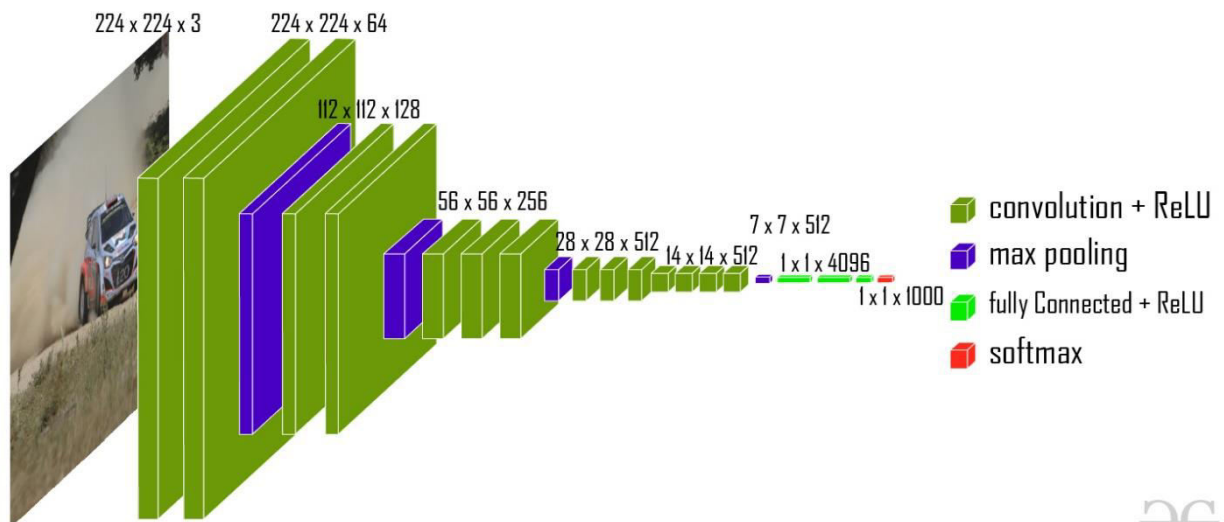
Convolutional neural networks, which are also known as convnets or CNNs, are a type of machine learning. It's a sub - set of the multiple models of artificial neural networks used for different applications and information sets. A CNN is a type of neural network structure for deep learning algorithms that is used for projects like picture recognising but also pixel data processing. Although other types of neural network models exist in deep learnings, CNNs are the favoured network architecture for object identification and recognition. As a result, they are ideal for device vision (CV) tasks and specific applications in which exact object recognition is critical, such as face as well as self-driving car systems



B. VGG-16

VGG16 proved to be a watershed moment in humanity's quest to allow computer systems to "see" this same world. The domain of vision systems (CV) has made substantial progress in this area since many decades ago. VGG16 is one of the major technologies that paved the way for further advancements in this field.

Almost any network interface is assumed to have a corrected 224 x 224 image as its input, with three colour channels, G, and B. The only which was before done is to standardise each pixel's RGB values. To achieve this, the average values of each pixel was calculated.



IV.METHODOLOGY

A.User

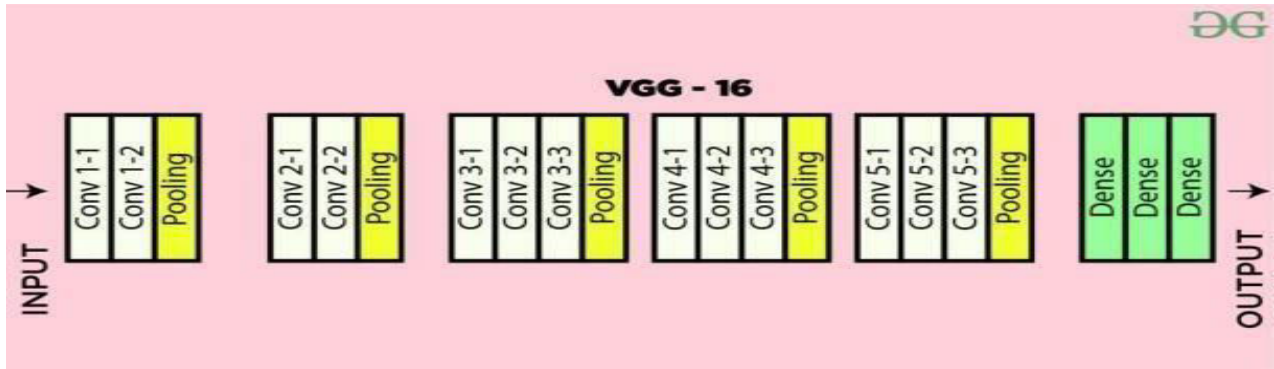
Run the mainrun.py file to launch the project. The direction towards the video file must be entered by the user. The system's primary camera is represented by the computer vision open-source class Video Capture(0), while the secondary camera is represented by Video Capture(1). Congratulations to video capture, we were still able to pack the video file from the disc without a camera (Video file path). In the past, Vgg16 and Vgg19 had been programmed. The user has a variety of options for how to run the code and can also choose a different model.

B.Har System

Vision-based technology can recognise human behaviour in videos. The RGB or depth image is used by the vision-based method. Users do not need to wear or carry any sort of sensing equipment. As a result, this methodology, which aims to make the very same HAR blueprint simple and quick to enforce in a number of different applications, is gaining popularity. The pictures for each activity were first taken from the videos. To train classifiers using machine learning and depth photo features, transfer learning is specifically used. Based on their parameters, such as RGB, RGB-D (Depth), and multi view, HAR data sets have a vibrant colours variety of features that were successfully captured in a controlled environment. Other requirements include "in the wild" recording, thorough sentence analysis, and annotation that only notes actions. To develop in this field, label data sets with details like the data gathering input, the amount of activities, movie clips, data-set surroundings, as well as when it was released. We observe that the intense ease, small size, and subpar performance of the vast bulk of the HAR data sets prevented them from being adopted by computer-vision researchers. Nevertheless, there are no most precise standard datasets that can be used as a benchmark for HAR tech evaluation. As we can see, the most widely used datasets among researchers are UCF101 and. Furthermore, the actions in the video recorded clips are carried out by a variety of different actors, unlike earlier datasets where occupations and actions are typically offered by a single actor.

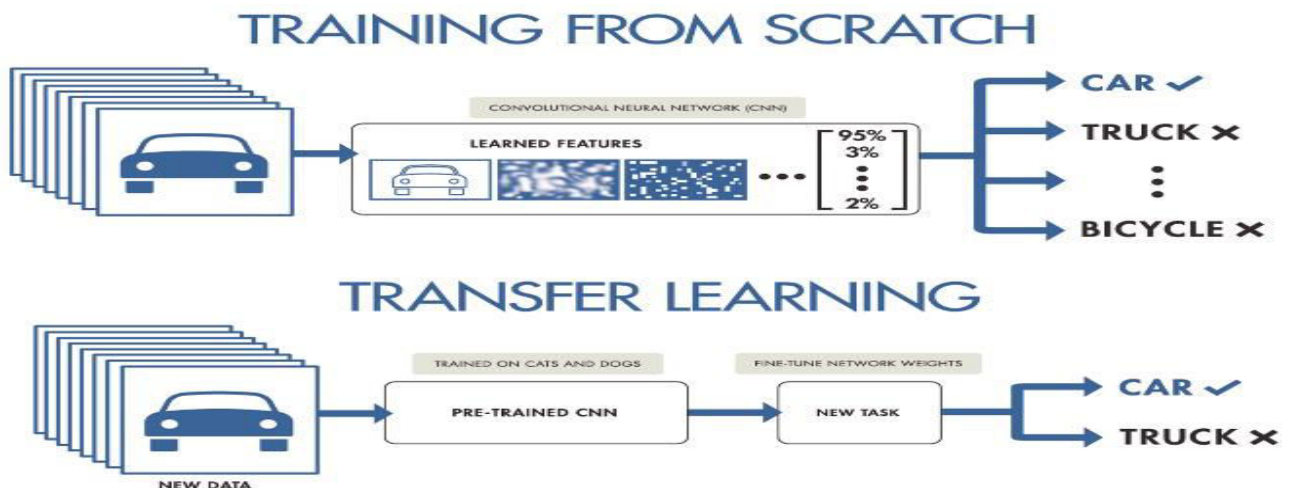
C.VGG-16

Model of the convolutional neural network VGG16. Large-Scale Image Recognition Using Deep Convolutional Networks. In the top 5 tests on Data sources, a set of data with over 14 million images divided in and out of 1000 classes, the technique performs 92.7% accurately. The model managed to enter into the ILSVRC-2014 was a well-liked one. It outperforms AlexNet by sequentially changing 33 kernel-sized filters for the large kernel-sized filter in AlexNet (11 and 5, respectively, for the first and 2nd convolutional layers). Utilizing Geforce Titan Black GPUs, VGG16 was trained over a period of weeks. Architecture VGG-16 The gadget receives a dimensionally precise image as input (224, 224, 3). With a 3*3 filter size, the first layer has 64 streams and the same additional material. The highest pool layer, which has two pace levels (2, 2) and 128 convolution layers with a step size after that (3, 3). The next layer is essentially an exact replica of the previous layer, being at most a step-(2, 2) layer. Then there are 256 filters spread across two convolutions with three and three convolution filters. Two sets of three convolutional layers are then added, followed by a maximal pool surface. There are 512 filters in total, each with exactly equal extra material (3, 3). The surface of two convolution is then applied to this image. We use 3*3 filtration in every single one of these. Instead of AlexNet's 11*11 filtration or ZF-7*7 Net's filters, convolution and maximum pooling layers were used. Furthermore, it employs a 1*11 display. with a resolution across a few layers to alter the quantity of data the quantity of input channels. One pixel of padding is added after every convolution layer to remove the geographic aspect of the image. After attempting to add a pooling layer and a layer of pooling to the stack, we successfully extracted features (7, 7, 512). This result is flattened to create an index-based feature vector (1, 25088). There are then three completely inter - connected layers: the first layer generates a variable of size (1, 4096) using the most recent several vectors as input; the second level generates a variable of dimension (1, 4096); and the stage three generates a variable of dimensions (1, 4096). a vector of dimensions (1, 1000), which is used to apply the activation function of softmax in order to classify 1000 classes. Convolution layer's activation function must be ReLU in its entirety. ReLU promotes accelerated learning, therefore It improves computational efficiency and lessens the possibility of disappearing gradient problems.



D. Transfer Learning

A machine learning technique that utilises a model from one task as a category of transfer learning. the basis for another model. It's a common approach in deep learning where pre-trained features serve as its starting point on machine vision as well as natural language transaction processing. This is due to the enormous compute and time resources needed to create models that use neural networks to focus on all these issues, and also to the massive leaps in talent that they provide on problems related. This article will demonstrate how transfer learning can be used to expedite training and Boost the accuracy of an algorithm that uses deep learning. Transfer learning is defined as the process by which a model applies knowledge acquired during previous rigorous training to that of an existing design. Learning Transfer enables the teaching of deep neural network models with a great deal less data. It has been applied to decrease training time and increase model precision. In this study, we employ transfer learning to gather information from sizable datasets like ImageNet. The frames for each activity were first taken from the videos. To attain deep photo features and functionality and learned machine learning classifiers, we use transfer learning. Transfer learning begins with weights that have already been trained on ImageNet for all CNN models. ImageNet [6] is a 20000 activity category dataset. So because variety of tasks recognised in this career falls within the scope of ImageNet, data is transferred from ImageNet pre-trained weights to the Weizmann set of data. The characteristics are obtained from the penultimate layer of CNNs.



V. SYSTEM ARCHITECTURE

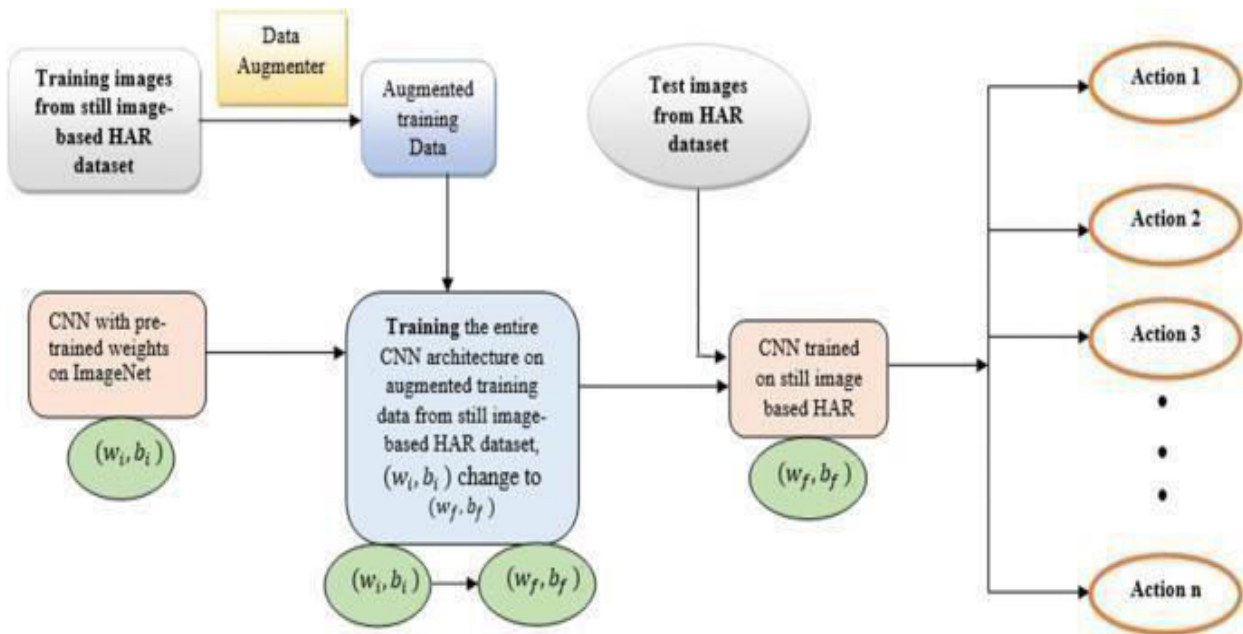


Fig.

VI. CONCLUSION

The massive Keras VGG16 network will contain millions of parameters. It is implemented using the Python dataset. There is now a depth training algorithm with pre-trained weights available. The model is used to create prediction models, extract features, and fine-tune them. And although latest HAR strategies have already had huge success up to this point, applying them to genuine systems remains difficult. VGG-16 is another image recognition algorithm that is used on a large scale. In Datasets, a collection of information including over fourteen hundred million images partitioned into 1000 classes, the model works 92.7% accurately in the top five tests.

REFERENCES

- [1] N. Hernandez, J. Lundström, J. Favela, I. McChesney, and B. Arnrich. Review of the Literature on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices and Environmental Technology. SN Computing Science 2020, 1, 1-16. [CrossRef] F. Zhuang; Z. Qi; K. Duan; D. Xi; Y. Zhu; H. Zhu; H. Xiong; He, Q. A Comprehensive Transfer Learning Survey. IEEE Proc. 2020.
- [2] S. Deep and X. Zheng. Using CNN and Transfer Learning for Vision-based Human Activity Recognition. In the Proceedings of the 2019 29th International Telecommunication Networks and Applications Conference, Appl. Sci. 2021, 11, 7660.
- [3] Mihailescu, R.; Casserfelt, K. An exploration of transfer learning in group activity recognition for deep architectures. Proceedings of the 2019 IEEE International Conference on Pervasive Computing and (PerComWorkshops), 11-15 March 2019, Kyoto, Japan.
- [4] D. Cook, K.D. Feuz, and N.C. Krishnan. Transfer learning for activity recognition: A survey. 2013; Knowl. Inf. Syst.
- [5] A probabilistic classifier based on importance-weighted least-squares for covariate shift adaptation with applications to human activity recognition. H. Hachiya, M. Sugiyama, and N. Ueda. Neurocomputing, 2012.



Organized by

Department of CSE, Velammal Engineering College, Chennai, India

- [6] B. Kröse, G. Englebienne, and T. van Kasteren. Using Transfer Learning to Recognize Activities in Different Contexts. Proceedings of the AAAI AI in Eldercare Symposium, Arlington, VA, USA, November 7-9, 2008.
- [7] L. Cao, Z. Liu, and T.S. Huang. Detecting cross-dataset actions. Proceedings of the IEEE Computer Society Conference on Computer Visualization and Pattern Recognition, held June 13-18, 2010, in San Francisco, California, USA.
- [8] H.M.S. Hossain, M.A.A.H. Khan, and N. Roy. DeActive: Scaling Activity Recognition with Active Deep Learning. ACM Interact.Mob.Wearable Ubiquitous Technol. 2018, Proc.
- [9] S.J. Pan and Q. Yang. A Transfer Learning Survey. IEEE Transactions on Expertise and Data Science, volume 22, number 2, pages 1345-1359, 2010.
- [10] Activity Recognitions: A Hierarchical Active Transfer Learning Approach, M.A.U. Alam and N. Roy. Proceedings of the 37th IEEE International Conference on Distributed Computing Systems, 2017.



SJIF: 8.423



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INTERNATIONAL CENTRE



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details