IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 14, Iss 04, 2025

UTILIZATION OF CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING FOR VISION-BASED HUMAN ACTIVITY RECOGNITION

Kuruva Rahul[1], Guguloth Harshith[2], Gajula Vivek[3], Vangari Dinesh Kumar[4], P. Nirosha[5], Prof. D Venkatesh[6]

[1,2,3&4]Ug Schoalr, Department Of Cse(Ai&Ml), Narsimha Reddy Engineering College (Ugc-Autonomous), Maisammaguda (V), Kompally, Secunderabad, Telangana-500100

[5]Assistant Professor, Department Of Cse(Ai&Ml), Narsimha Reddy Engineering College (Ugc- Autonomous), Maisammaguda (V), Kompally, Secunderabad, Telangana-500100

[6]Associate Professor & Hod, Department Of Civil Engineering, Narsimha Reddy Engineering College (Ugc- Autonomous), Maisammaguda (V), Kompally, Secunderabad, Telangana-500100

ABSTRACT

The identification of human behavior is an important problem in a variety of domains, including health monitoring, human-computer interaction, and security surveillance, among others. Through the use of a Multiscale Convolutional Neural Network (MSCNN), this study presents a unique way to human behavior identification. The objective of this research is to improve the accuracy and resilience of behavior categorization based on video data. A number of different convolutional scales are included into the MSCNN model that has been suggested in order to extract certain spatial and temporal characteristics from video sequences. In order to properly identify complex and diverse human behaviors, which are often difficult to detect using standard approaches, the model processes video frames at multiple sizes. This allows the model to efficiently recognize these behaviors. The network is able to concentrate on both fine-grained features and more general contextual information because to the multiscale methodology, which ultimately results in significant improvements in recognition performance. A number of convolutional layers, each of which operates at a different resolution, make up the MSCNN architecture. In order to provide a full picture of human behaviors, these layers are intended to extract hierarchical elements, which are subsequently fused together. The model is trained on a huge dataset consisting of annotated video sequences, which allows it to learn and generalize across a wide variety of behavioral patterns and events. Through numerous studies on benchmark datasets, the usefulness of the MSCNN has been established. These trials have shown considerable gains in recognition accuracy when compared to other approaches that are currently in implementation. The findings provide information on the capability of the model to deal with a variety of obstacles, including occlusions, changes in appearance, and a variety of climatic variables.

1.INTRODUCTION

The study on human behavior recognition may not only contribute to the development of the appropriate theoretical foundation in the area of computer vision, but it can also broaden the engineering application of machine learning. When it comes to the theoretical foundation, the subject of behavior recognition incorporates the knowledge of a wide

range of fields, including image processing, computer vision, artificial intelligence, human kinematics, and biology. The processing of video footage via the use of computer vision technology is an essential approach that involves human behavior identification. It is a significant path for the study to take. In accordance with the various types of convolution kernels, approaches for behavior identification that are based on deep learning may be classified into two distinct categories: In the field of motion recognition, a great number of researchers have used deep learning techniques, including 2D convolution networks and 3D convolution networks. Through the use of a variety of approaches, they have attempted to develop the behavior recognition technology that is based on computer vision, and they have obtained satisfactory results. These techniques of behavior recognition may be loosely split into two categories: the first category is technology for behavior recognition that is based on classical classification methods, and the second category is technology for behavior recognition that is based on deep learning. Combining the benefits of these two approaches, the current research path of behavior recognition technology is to apply the technique of manual feature extraction in conjunction with deep learning [2, 3]. This is the mainstream research direction. Furthermore, it is difficult to effectively model the relatively slow or static behavior because of the complexity of human behavior itself. Human behavior is easily disturbed by complex background, occlusion, light, and other environmental factors. The majority of the current feature extraction methods are cumbersome and prone to error transmission. This is because human behavior is easily disturbed by these factors. Additionally, the convolutional neural network that only has one scale is unable to completely characterize the features of human behavior from a variety of perspectives, which is not favorable to the final behavior recognition. As a result of the study conducted in this field, a great number of effective network architectures have come into existence, including C3R [4, eco [5, TSN [6], and many more. All of these network models have a strong modeling capacity for video data and are able to efficiently differentiate between various human activities in real scenarios, despite the fact that their structures are distinct from one another. From a theoretical standpoint, the feature description vectors that are created from various network models are sensitive to category information (using the classification job as an example), and they become linearly separable at the output layer of the network. Although they may originate from distinct modeling procedures, the feature vectors that are created have to be comparable to one another. The question of whether or not the information that is obtained by various network architectures may be taught and shared is a complex one that deserves discussion. Increasing the breadth and depth of the initial network, using the decomposition of the original parameters or the unit matrix to initialize the weight parameters, and achieving cross structure transfer learning were all things that Chen et al. [7] accomplished. The authors Ali et al. [8] used the 2D network in order to oversee the input and output of the 3D network. They also made the 3D network conform to the output characteristic

distribution of the 2D network, which indirectly resulted in the realization of cross structure learning. Taking this as its inspiration, this research further relaxes the limitations of the model structure and uses successful measurement methodologies [9], [10] between the two networks that have bigger structural disparities in order to create a more generic notion of transfer learning, which is referred to as soft transfer. At the moment, approaches for recognizing human behavior may be primarily classified into two categories: the classic manual feature extraction methods and the deep learning methods. Methods of manually extracting features typically consist of three processes that are performed in succession: characteristic extraction and computation of the local descriptor Classification (number 11). Following the matching of the edge information with the key posture and location of the mark, Sullivan et al. [12] proceed to track between subsequent frames based on the contour information. Oikonomo et al. [13] introduced a detector that requires computing the entropy characteristics of the cylindrical neighborhood surrounding a certain space-time point. This detector was developed according to their findings. Patrona et al. [14] introduced automatic motion data and dynamic motion data weighting in order to adjust the importance of human data on the premise of action participation. This was done in order to achieve more effective action detection and recognition. This was accomplished by highlighting motion features in order to represent different positions within the video. When compared to the 2D convolution approach, the process that is based on the 3D convolution network has a network topology that is both more straightforward and more effective. It is difficult to utilize the behavior data acquired by smart phone sensors (triaxial acceleration and gyroscope sensors) directly in the investigation of har based on smart phones because of the noise that is present in the data. As a result, feature engineering is used extensively in a variety of har models in order to extract reliable human behavior traits from sensor data. There is a human behavior pattern recognition framework that is constructed and given the term human in the publication [13]. This framework provides the RF with greater recognition. A human behavior recognition framework that is based on environment perception is proposed in paper [14], which integrates data on human behavior with information about the environment. Experiments conducted using decision trees (DT), support vector machines (SVM), and k-nearest neighbors (k-NN) demonstrate that the framework for behavior identification that is based on knowledge about the environment contributes to an improvement in the recognition performance of the model. In accordance with the needs of many domains, the study [15] suggested a learning model for human behavior identification that was based on cascade integration. Extreme gradient boosting trees (egbt), random forests (RF), extreme randomized trees (ERT), and softmax regression are the components that make up each layer of the model. In the first layer, the four models are trained using sensor data, and then the probability vectors representing the distinct categories of each data are generated The initial input data and the probability vector are then concatenated together to serve as the input

for the subsequent level classifier. Finally, the prediction results are achieved in accordance with the classifier that is the final level. In comparison to the approaches that are currently in use for recognition, the results of the experiments demonstrate that this method achieves a higher level of recognition accuracy, and the process of training the model is both more straightforward and more efficient. It is necessary to manually finish the feature marking in the study on human behavior identification based on volume neural network extended model. Additionally, the calculation amount, generalization ability of the model, and feature acquisition ability of the model need to be further enhanced. In order to learn and inherit the video feature modeling capability of networks, a new network structure of MDN is built for the fundamental module of densenet [11]. Additionally, the soft migration technique is used in order to develop this structure. The structures of many network models are distinct from one another. Please keep in mind that the mdn-i3d combination is a semi-supervised "learner supervisor" combination.

2.LITERATURE REVIEW

Jia Lu, Wei Qi Yan, and Minh Nguyen showed a detection approach that is based on deep learning in order to demonstrate the ability to distinguish pedestrians. The YOLO model, which is a deep learning approach that permits real-time detection, was used in the research experiment. While the deep learning system is being trained and tested, a GPU acceleration is required in order to minimize the amount of time that is used. For the purpose of fine-tuning the model, it is essential to choose an appropriate hyperparameter with great care. This is because different hyperparameters have the potential to affect the results. The extension of the YOLO detection technique need to be the primary focus of further research in the future. The ability to detect items and place each one in the right category is a demonstration of deep learning's capabilities. For the purpose of facilitating machine learning-based dynamic behavior categorization and detection, Mayur Shitole, Jerry Zeyu Gao, Shuqin Wang, and Hanping Lin Sheng Zhou, together with Layla Reza, have proposed well-defined emoji-based human behavior patterns. Standing, moving fast, moving slowly, and sitting are the four main human movements that are the focus of this study. Additionally, a system that is provided is characterized as being able to support real-time human dynamic behavior recognition and classification based on the recommended machine learning model and behavior patterns that are based on emojis. The outcomes of several previous case studies for dynamic human behavior detection and categorization using emoji representation are also presented in this article. Streaming videos in real time as well as videos that have been recorded in advance may both be viewed on the system without any problems.

- Chen Chen has presented a system for behavior recognition that is based on a deep neural network and a hidden Markov model. This research combines the advantages of conventional methods to preserve features with the benefits of deep learning techniques in order to optimize the benefits of these methods. It is possible for his proposed technique to have a favorable influence on the detection of

interactive behavior because of the advantages that deep networks, self-extraction, self-training, and temporal information processing provide. However, owing to the arduous hand extraction of characteristics that is required by traditional approaches, this method is still not timely.

- Zhengjie Wang and Yinjing Guo provide the broad techniques of behavior identification that are already in use now, as well as relevant surveys, the idea of channel state information, and an explanation of the fundamentals of CSI-based behavior recognition. In addition to this, the study provides a comprehensive analysis of the general framework for behavior recognition. This analysis covers the basic signal selection, signal preprocessing, and behavior identification methods that use pattern-based, model-based, and deep learning-based approaches. The paper takes the existing research and applications and divides them into three categories based on the recognition methodologies that were mentioned earlier. It then goes on to describe each typical application in great detail, including the test equipment, experimental situations, user count, observed behaviors, classifier, and system performance. In addition to this, it investigates a few specific applications and contains in-depth talks on the selection of recognition algorithms and the evaluation of performance. During these interactions, several helpful recommendations for the development of an identification system have been offered.
- When the AcFR system made the decision to change its viewpoint, an insufficient amount of consideration was given to the direction in which it would travel. This may be problematic since the system may opt to inspect the person from behind rather than the front, which is the normal movement that people make in order to get a better view of a subject's face. It is necessary to make an estimate of the orientation of the face in order to achieve more accurate active face recognition.

A technique that may recognize certain human behaviors or activities has been proposed by Sumin Jin, Yungcheol Byun, and Sangyong Byun. This approach makes use of electroencephalogram (EEG) brain waves. This was accomplished by identifying six different behaviors and recording the brain waves that were connected with each of those behaviors. In the trials, they employed CNN and LSTM models, and the results showed that they were able to identify 66% of behaviors by analyzing EEG brain waves. Taking into consideration the intricate nature of the relationship that exists between brain waves and behaviors, this is an encouraging finding. As a result of the presence of dynamic information, the LSTM model generated a superior result.

- Weihu Zhang and Chang Liu have suggested a technique that is based on improved deep learning and is intended to detect anomalous characteristics of human behavior. According to this technique, the rate of recognition is higher, the extraction of features is more accurate, and the model is simplified to a lesser extent than with the standard approach. Through the use of the Gauss model and the Farneback dense optical flow approach, it is possible to get accurate key regions of human motion as well as an optical flow map. The advantages of CNN and LSTM are used in

conjunction with one another to provide an accurate recognition effect.

- A technique has been proposed by Franjo Matkovi, Darijan Mareti, and Slobodan Ribari for recognizing motion patterns and anomalous crowd behavior in surveillance footage. An examination of fuzzy predicates and fuzzy logic formulae that are obtained from human interpretation of actual video sequences, (multi-agent) crowd simulators, and data from common sense are the foundations upon which it is built. Analysis of motion patterns of an individual or group of persons is performed with the help of fuzzy logic predicates. This allows for the identification and classification of motion patterns in accordance with the taxonomy of fuzzy logic predicates that has been provided. Through the use of fuzzy logic functions, the identification and categorization of peculiar crowd behavior. The fuzzy predicates are the essential building blocks of fuzzy logic functions, and the assignment functions for these predicates are formed by properly analyzing training video sequences in combination with fuzzy logic operators. This process is known as fuzzy logic. For the purpose of evaluating the suggested method, we make use of real trajectories that were accomplished by the 4pipelined multi-person tracker that was presented, as well as ground truth annotations of actual video sequences. A number of early tests have shown findings that are both positive and comforting.
- An algorithm for group feature behavior identification that is based on the attention mechanism was proposed by Feng Xiufang and Dong Xiaoyu in order to solve the problem of long-term modeling of current behavior recognition algorithms. Through the use of sparse sampling, it is possible to effectively reduce the redundant frames that occur between frames in video sequences. The original frame pictures are used in CNN for the purpose of modeling space features in order to extract motion change information in an effective manner. During the process of performing network training, progressive networks with pyramid pools are used to extract image attributes. After the features of the video frame have been encoded in a sequential manner using Bi-GRU, the final feature vector is then created by adding an attention layer. The results of the experiments that were conducted show that the data aspects of this article have the potential to effectively boost the network's capacity to express itself, and that the structure of this paper's network can successfully simulate the long-term attention that videos provide.

3.EXISTING SYSTEM

Strategies such as handcrafted feature extraction and shallow learning models are the primary strategies that are used by traditional human behavior identification systems. The extraction of certain characteristics from video data, such as motion vectors or posture keypoints, and the subsequent application of these features to the classification of behaviors are common components of these systems. For the purpose of capturing motion and appearance information, it is standard practice to make use of techniques such as optical flow, histogram of oriented gradients (hog), and several other feature descriptors. For the purpose of behavior recognition, a great number of the currently available systems make use of typical convolutional neural networks

(cnns). Conventional convolutional neural networks (cnns) are used to extract spatial data from video frames or sequences by using fixed convolutional kernels. These approaches, on the other hand, may have difficulty capturing the complex temporal dynamics and various geographical scales that are inherent in human activities. Despite the fact that cnns are excellent for the classification of static images, it is possible that they are not able to fully handle the complexities of video data. This is because activities unfold over time and entail a variety of sizes and movements at different times. In addition to cnns, some systems also make use of recurrent neural networks (rnns) or long short-term memory (lstm) networks in order to manage temporal elements. Processing sequences of frames is the goal of these hybrid techniques, which are designed to describe temporal interdependence. This integration may still be lacking in its capacity to properly handle varied and multiscale behavioral aspects, despite the fact that it enhances the ability to capture time-related patterns. Complex structures are often involved in this integration.

3.1.Drawbacks:

- Limited Integration of Temporal Context: Conventional CNN-based systems and many other approaches now in use have difficulty efficiently integrating temporal context. In spite of the fact that they are able to extract spatial characteristics from individual frames, they often fail to comprehend the time development of activities. This constraint may make it more difficult to accurately recognize behaviors that entail intricate sequences of events or interactions that last for an extended period of time
- Conventional convolutional neural networks (CNNs) are characterized by their use of fixed-scale convolutions, which may result in the loss of crucial features at varying sizes. It is possible that this fixed approach may result in performance that is less than optimum when it comes to identifying behaviors that take place at many scales or include both fine-grained and coarse-grained spatial structures. It is possible that the system's capacity to generalize across a variety of settings will be hindered by its inability to adjust over size.
- High Computational Demand: Multiscale Convolutional Neural Networks (MSCNNs) need the processing of video data at many scales, which may greatly increase the amount of computational complexity and resource use. Due to the fact that it requires many convolutional layers and operations at various resolutions, it might result in longer training durations and greater computing costs, which makes it difficult to implement in settings with limited resources.
- Difficulty in Handling Occlusions: Occlusions, which are situations in which portions of the subject are obscured from view, is something that might provide difficulties for behavior recognition systems. Obstructions may still cause problems with the proper extraction of features, even if MSCNNs are designed to enhance resilience. This is particularly true in situations when essential sections of the body or crucial activities are hidden at certain sizes.
- Overfitting Risks: The intricacy of MSCNNs, with their numerous scales and deep architectures, raises the potential of overfitting to training data. This kind of

overfitting may have negative consequences. If it is not controlled appropriately, the model may learn to perform well on certain datasets, but it may have difficulty generalizing to data that it has not before seen or to a variety of real-world settings.

4.PROPOSED SYSTEM

An enhanced method for recognizing human behavior is presented by the system that is being offered. This method makes use of a Multiscale Convolutional Neural Network (MSCNN), which is meant to improve the accuracy and resilience of behavior categorization in video data. This cutting-edge approach incorporates numerous convolutional scales in order to collect a broad variety of spatial and temporal data. This allows it to overcome the restrictions that are inherent in conventional techniques. Within the framework of the MSCNN architecture, the network analyzes video sequences by way of a number of convolutional layers, each of which operates and operates at a different resolution. The model is able to efficiently capture complicated behaviors that span several scales because to its multiscale approach, which enables it to simultaneously concentrate on fine-grained features and wider contextual information. The system is able to identify small subtleties in human behaviors and interactions that are often ignored by fixed-scale models. This observation is made possible by the analysis of video frames at many levels of abstraction. With the help of convolutional and recurrent processes, the MSCNN is able to manage both spatial and temporal dynamics. This is accomplished by exploiting the capabilities of both frameworks. During the process of modeling the sequential character of activities and interactions across time, temporal properties are collected by means of a sequence of convolutional layers, which are then followed by recurrent connections or temporal pooling techniques. With the help of this hybrid technique, the system is guaranteed to be able to effectively distinguish and discriminate between behaviors that occur throughout a number of frames and time steps.

4.1.ADVANTAGES:

1. **Enhanced Feature Extraction:** The Multiscale Convolutional Neural Network (MSCNN) excels in capturing features at various spatial scales. By processing video data through multiple convolutional layers operating at different resolutions, the system can detect both fine-grained details and broader contextual information, improving the accuracy of behaviour recognition.

2. **Improved Temporal Context Understanding:** MSCNNs integrate temporal dynamics by combining convolutional and recurrent mechanisms. This approach allows the network to effectively model the sequential nature of human actions and interactions over time, leading to more accurate classification of behaviours that span multiple frames.

3. **Robust to Scale Variations:** The multiscale architecture addresses the limitations of fixed-scale models by incorporating features from different resolutions. This adaptability makes the system more robust to variations in the size and scale of actions, enhancing its ability to recognize behaviours in diverse scenarios.

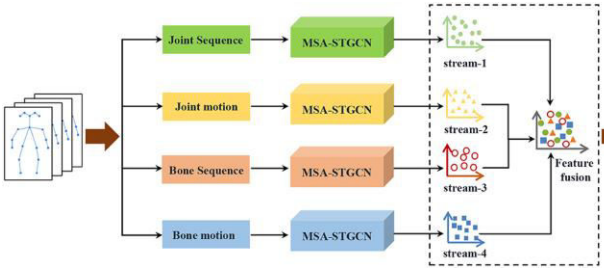
4. **Comprehensive Behaviour Representation:** By employing advanced feature fusion techniques, the MSCNN creates a comprehensive representation of human activities. Integrating information from multiple scales and temporal contexts allows the system to capture complex behaviours more accurately, leading to improved recognition performance.

5. **Reduced Computational Burden:** The proposed system incorporates optimization strategies such as model pruning and quantization. These techniques reduce the computational and storage requirements of the MSCNN, making it more efficient and suitable for deployment in resource-constrained environments.

6. **Adaptability to New Behaviours:** The MSCNN's ability to analyze and integrate multiscale and temporal features enhances its adaptability to new or unseen behaviours. This flexibility is crucial for generalizing to diverse behaviour patterns and environmental conditions without extensive retraining.

5.IMPLEMENTATION

5.1. SYSTEMARCHITECTURE



5.2. MODULES:

- **User**
- **HAR System**
- **VGG16**
- **Transfer Tearning**

MODULES DESCRIPTION:

User:

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

.

HAR System:

Video-based human activity recognition can be categorised as vision-based according. The vision based method make use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors 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.

VGG16:

VGG16 is a convolutional neural network model. Deep Convolutional Networks for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing

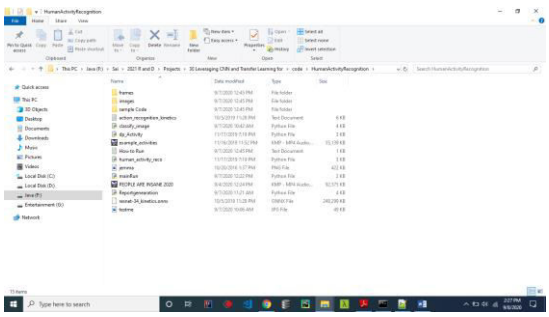
ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 14, Iss 04, 2025

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.

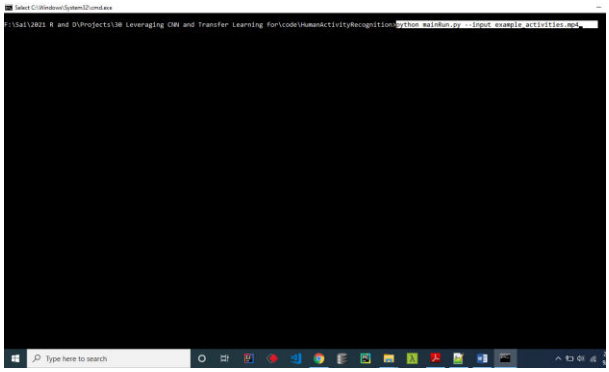
Transfer Tearning:

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 skill that 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 deep learning model.

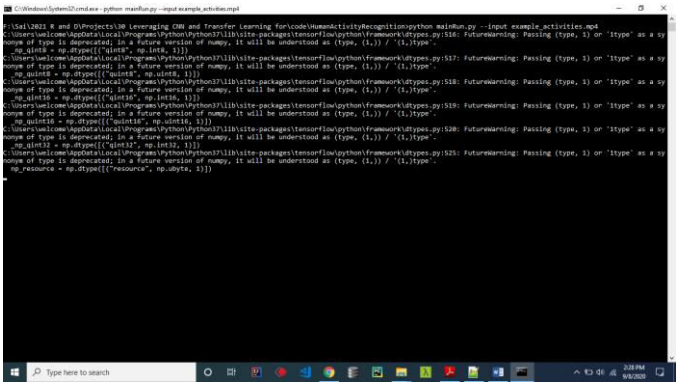
6.RESULTS



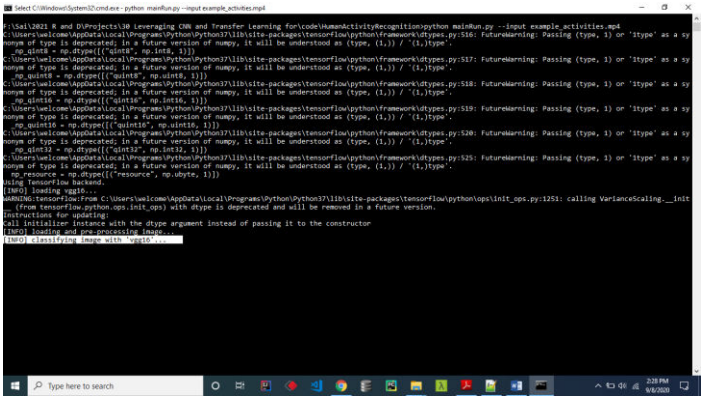
Run the main Program



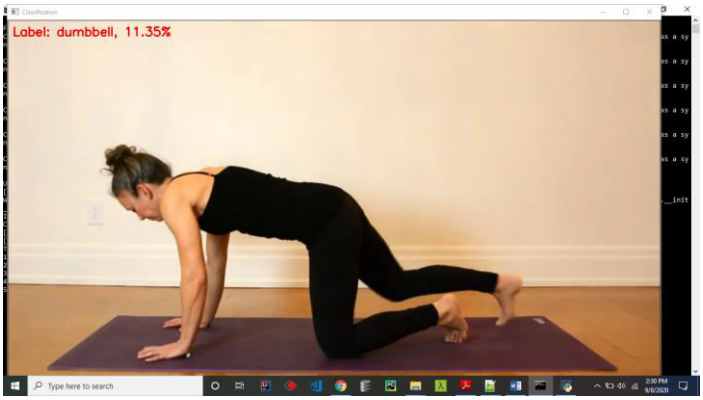
Loading Tensor flow Libraries



Classification with vgg16

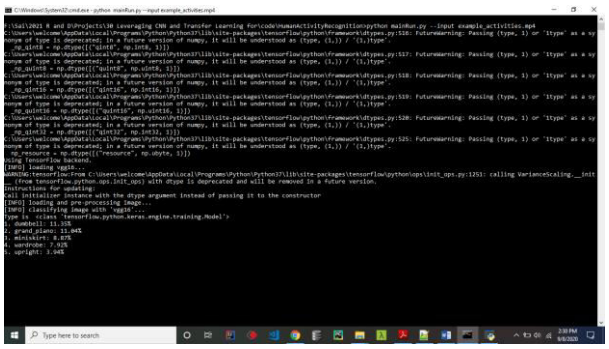


Get Image label

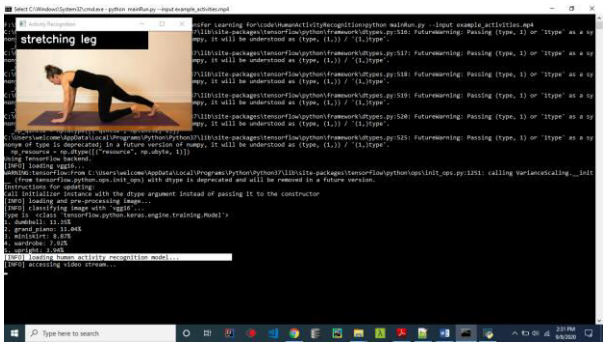


Result from image

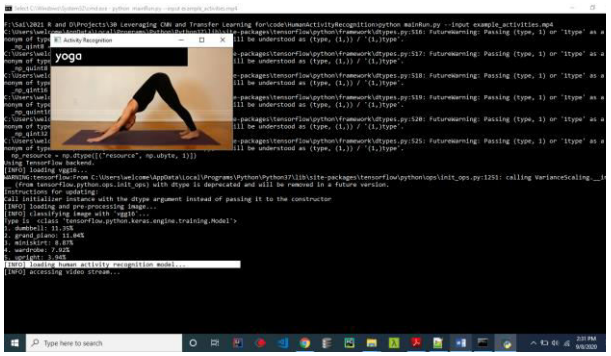
ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 14, Iss 04, 2025



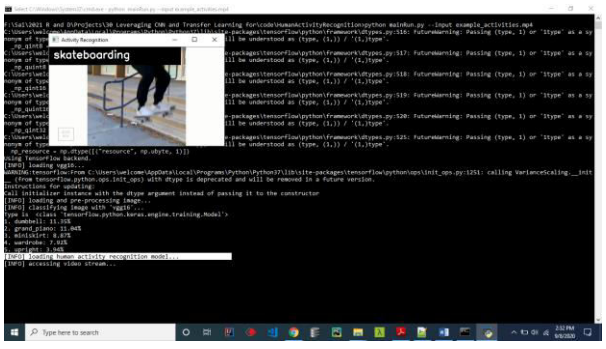
Loading model HAR



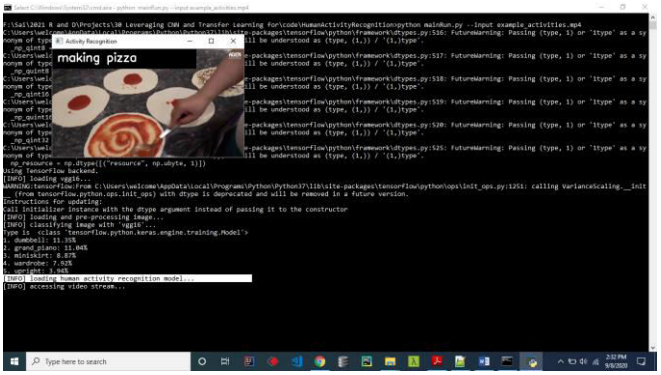
Result 1



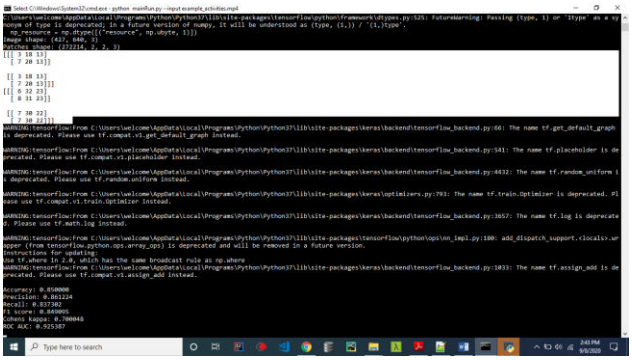
Result 2



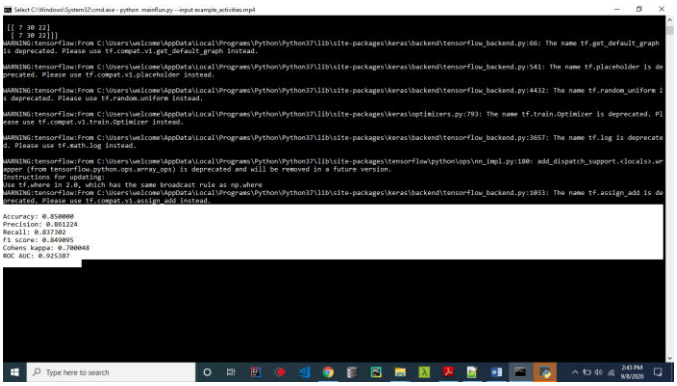
Result 3:



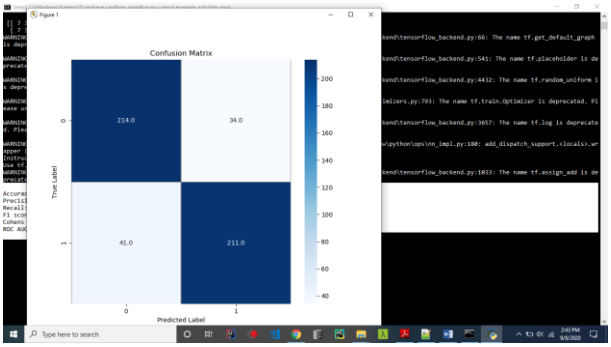
Patches from image:



Accuracy:



Confusion matrix:



CONCLUSION

The purpose of this study is to provide a technique for the identification of human behavior that is based on developed attention mechanisms. After conducting an analysis of the deficiencies of the current channel attention mechanism, a proposal for an enhanced attention module has been made. It is necessary to conduct tests in order to validate the efficacy of the enhanced attention module. These experiments are conducted from the perspectives of visualization results, network accuracy improvement, extra network parameters, and so on. In order to obtain the behavior characteristics under various receptive fields, the multi-scale convolution kernel is utilized. Additionally, the convolution layer, pool layer, and full connection layer are designed in a reasonable manner in order to further refine the characteristics, which demonstrates that the cross structure learning is feasible. The requirement of a multi-stage progressive supervision method is shown by comparing the supervision that occurs in the various phases. Additionally, the role of model structure on the effect of soft migration is examined. When the structure of the monitoring network is comparable to that of the learning network, it has been discovered that the organization of the network is more straightforward to converge. In the work that will be done in the future, more sensors may be employed to expand the data dimension, which will further improve the accuracy of the recognition. Considering that the model module of our technique has a great deal of parameters, the work that will be done in the future will concentrate on finding ways to make the model more lightweight.

REFERENCES

[1] X.-J. Gu, P. Shen, H.-W. Liu, J. Guo, and Z.-F. Wei, ``Human behavior recognition based on bone spatio-temporal map," Comput. Eng. Des., vol. 43, no. 4, pp. 11661172, 2022, doi: 10.16208/j.issn1000- 7024.2022.04.036.

[2] M. Z. Sun, P. Zhang, and B. Su, ``Overview of human behavior recognition methods based on bone data features," Softw. Guide, vol. 21, no. 4, pp. 233239, 2022.

[3] Z. He, ``Design and implementation of rehabilitation evaluation system for the disabled based on behavior recognition," J. Changsha Civil Affairs Vocational Tech. College, vol. 29, no. 1, pp. 134136, 2022.

[4] C. Y. Zhang, H. Zhang, W. He, F. Zhao, W. Q. Li, T. Y. Xu, and Q. Ye, ``Video based pedestrian detection and behavior recognition," China Sci. Technol. Inf., vol. 11, no. 6, pp. 132135, 2022.

[5] X. Ding, Y. Zhu, H. Zhu, and G. Liu, ``Behavior recognition based on spatiotemporal heterogeneous two stream convolution network," Comput.Appl. Softw., vol. 39, no. 3, pp. 154158, 2022.

[6] S. Huang, ``Progress and application prospect of video behavior recogni- tion," High Tech Ind., vol. 27, no. 12, pp. 3841, 2021.

[7] Y. Lu, L. Fan, L. Guo, L. Qiu, and Y. Lu, ``Identication method and experiment of unsafe behaviors of subway passengers based on Kinect," China Work Saf. Sci. Technol., vol. 17, no. 12, pp. 162168, 2021.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 14, Iss 04, 2025

[8] X. Ma and J. Li, ``Interactive behavior recognition based on low rank sparse optimization," J. Inner Mongolia Univ. Sci. Technol., vol. 40, no. 4, pp. 375381, 2021.

[9] Z. Zhai andY. Zhao, ``DS convLSTM:Alightweight video behavior recognition model for edge environment," J. Commun. Univ. China, Natural Science Ed., vol. 28, no. 6, pp. 1722, 2021.

[10] C. Ying and S. Gong, ``Human behavior recognition network based on improved channel attention mechanism," J. Electron. Inf., vol. 43, no. 12, pp. 35383545, 2021.