The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
19(2): S.I (1), 234-237, 2024 www.thebioscan.com

RECOGNITION OF HUMAN BEHAVIOUR UTILISING MULTISCALE CONVOLUTIONAL NEURAL NETWORKS

K BANGARU LAKSHMI [1], P GOWTHAMI DEVI [2], G MADHURI [3]

1 Assistant Professor, Kakaraparti Bhavanarayana(KBN) College, lakshmi089@kbncollege.ac.in

2 Assistant professor, Kakaraparti Bhavanarayana(KBN) College, gowthamidevi@kbncollege.ac.in

3 Assistant professor, Kakaraparti Bhavanarayana(KBN) College, madhurigongati@kbncollege.ac.in

DOI: https://doi.org/10.63001/tbs.2024.v19.i02.S.I(1).pp234-237

KEYWORDS
Behavioural recognition;
Channel attentiveness;
Deep separable network
Received on:
04-08-2024
Accepted on:
22-11-2024

ABSTRACT

In order to recognise human conduct, the most difficult thing is to construct a network that can extract and classify features based on their spatial and temporal relationships. To enhance the existing channel attention mechanism, which only takes into account the global average data from each channel and disregards its local spatial information, we suggest using the space-time (ST) interaction matrix operation module in conjunction with the depth separable convolution module. These modules are accompanied by studies on human behaviour recognition. A multi-scale CNN method for human behaviour recognition is suggested, taking advantage of CNN's superior performance in video and image processing. Low rank learning takes the behaviour video segments and uses them to derive knowledge about low rank behaviour. Without making any assumptions or enduring any tedious extraction techniques, the complete video's low-rank behaviour data can be obtained by linking this data along the time axis. Human behaviour models trained on neural networks can be reused across many network topologies. In order to reduce the disparity between features derived from different network topologies, two efficient approaches for measuring feature difference at various network levels are presented. The suggested method is effective, according to classification tests conducted on a wide variety of publicly available datasets. Experiments show that the method accurately identifies human conduct. According to our findings, the proposed model improves recognition accuracy, streamlines model structure, and makes computing output weights easier.

INTRODUCTION

Studying human behaviour recognition through computer vision can advance the field's theoretical underpinnings and broaden its practical uses. A combination of biology, computer vision, AI, human kinematics, and picture processing forms the basis of behaviour recognition theory. Video processing using computer vision relies heavily on human behaviour recognition. Crucial course of investigation [1]: Yes.
Two groups of methods for identifying behaviour via deep learning are distinguished by the use of different convolution kernels: The use of deep learning with 2D and 3D convolution networks for motion recognition has been extensively studied. They successfully implemented computer vision-based behaviour recognition technology using many methodologies. Chapter 1 will focus on the literature and methodologies. Classical classification and deep learning are two broad groups into which these methods of behaviour recognition fall. The majority of behaviour recognition studies integrate deep learning with manual feature extraction [2, 3]. Human behaviour is complex and easily disrupted by complex backgrounds, occlusions, light, and other environmental influences, making most feature extraction approaches laborious and error-prone. When trying to represent behaviour that is slow or stationary, you will face similar obstacles. A convolutional neural network that only works on one scale will struggle to recognise human conduct since it can't capture the complexity of the phenomenon from all the different perspectives.
Several effective network topologies have emerged in domain research, including C3R [4], eco [5], TSN [6], and many more. Despite their structural diversity, these network models do a good job of representing video data and identifying human behaviour in real-world settings. Different network models' feature description vectors should be linearly separable at the output layer and sensitive to category information, such as categorisation, in theory. The feature vectors that come out of different modelling processes ought to be comparable. Can various network topologies learn and share information? We need to have this conversation. In order to accomplish cross-structure transfer learning, Chen et al. [7] enhanced the network's breadth and depth, initialised the weight parameters using the decomposition or unit matrix, etc. By controlling the inputs and outputs of the 3D network and fitting its characteristic distribution to the 2D network, Ali et al. [8] learnt across structures without consciously doing so. Soft transfer learning, a broader type of transfer learning, is accomplished by this article by utilising effective measurement approaches [9, 10] between the two networks that differ more structurally and by removing the restrictions of the model's structure.

LITERATURE SURVEY

2.1 'Development of lower limb rehabilitation evaluation system based on virtual reality technology'
https://ieeexplore.ieee.org/document/7784083
ABSTRACT: With the growing older population, various challenges caused by population ageing are becoming more visible. Physical therapists thrive because hemiplegia plagues most elderly people. Traditional physical treatment relies heavily on the ability of the therapist. To address the limitations

of conventional methods, numerous research groups have developed robots to assist with lower limb rehabilitation. However, the majority of these robots are only capable of passive training and do not have a rehabilitation evaluation system to track the progress of hemiplegic patients in real-time. To tackle these issues, this study created a method for evaluating lower limb rehabilitation using virtual reality. Thanks to its user-friendly interface, this lower limb rehabilitation evaluation system allows doctors to customise rehabilitation training for patients at different phases of recovery. Medical rehabilitation robots are anticipated to be more affected by this innovative lower limb rehabilitation evaluation method compared to conventional methods.

2.2 Spatiotemporal Heterogeneous Two-Stream Network for Action Recognition

https://ieeexplore.ieee.org/abstract/document/8688399

ABSTRACT: There is a successful algorithm for video action recognition using a two-stream network. The majority of methods use an inefficient spatial-temporal network structure. Different network architectures are used for geographical and temporal information in the spatiotemporal heterogeneous two-stream network that is proposed in this study. The use of ResNet and BN-Inception as foundational networks demonstrates human spatiotemporal behaviour. A segmental architecture simulates the long-range temporal structure spanning video sequences in order to differentiate comparable occurrences with sub-action sharing. The spatiotemporal heterogeneous network is enhanced for human action recognition by using data augment and a modified cross-modal pre-training technique. When tested on UCF101 and HMDB51 datasets, spatiotemporal heterogeneous two-stream networks demonstrated superior performance compared to isomorphic networks and other approaches.

2.3 Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments

Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discernment: Evidence from Two Experiments

ABSTRACT: One product of recent machine learning advancements is the "deepfake," an impressively realistic CGI of a famous person making fraudulent statements. Despite policymakers' concerns that deepfakes could sway elections, studies have shown little impact. Voters may become suspicious of all political video content if they are continually warned about deepfakes, according to this essay's analysis of a downstream effect of these fake news stories. Our two online poll studies revealed that respondents could not distinguish between an authentic film and a deepfake. Warnings regarding the presence of deepfakes did not improve participants' proficiency in identifying altered video footage. Conversely, these cautions consistently led participants to perceive the videos they viewed as fabricated, regardless of their authenticity. The warnings were not tailored to the individual video being seen; merely indicating the existence of deepfakes heightened scepticism towards any related video content. Our research indicates that although deepfakes may lack inherent persuasiveness, discourse surrounding them can be exploited by politicians and campaigns to discredit and reject authentic footage.

METHODOLOGY

i) Proposed Work:

Two modules for improved channel attentiveness are proposed: interactivity module for space-time (ST) that captures detailed spatiotemporal features through matrix operations, and a depth-wise separable convolution module that processes spatial and channel information separately for better feature extraction. This model uses a multi-scale CNN to handle sequential input, applying low-rank learning to each segment and connecting them along the time axis to generate an activity representation. Introducing feature similarity algorithms that reduce discrepancies in extracted features at different network levels improves cross-architecture flexibility and model recognition transferability across network structures. The method improves classification accuracy while reducing computing load, making it ideal for practical applications that require efficiency and performance.

ii) System Architecture:

Classical categorisation methods and deep learning-based approaches are the two types of behaviour recognition systems that are now available. In order to maximise their respective strengths, researchers in the field of behaviour recognition are currently concentrating on merging deep learning with manual feature extraction. Background, occlusion, lighting, and other environmental factors muddy human conduct, making feature extraction approaches difficult and prone to errors. It is challenging to model slow or inactive behaviour. Furthermore, behaviour recognition is hindered by the fact that a single-scale convolutional neural network is unable to adequately depict human behaviour from several perspectives.

Here author uses 3DCNN algorithm for human behaviour prediction because all existing algorithms directly use global average information of each channel (taking all channels of images as single data) which ignores spatial and depth information from image features, resulting in inaccurate recognition. Models may accurately anticipate image shapes if they have accurate information. So the author used ST interaction module of matrix operation, depth separable convolution module, and human behaviour recognition research in this work. With CNN's image and video processing capabilities, a multi-scale CNN strategy for human behaviour recognition is suggested. Multiscale Convolution Neural Networks combine spatial and depth separable modules. The proposed model is tested on UCI HAR dataset, which records smartphone activities. Provide the most accurate CNN2D or LSTM model.

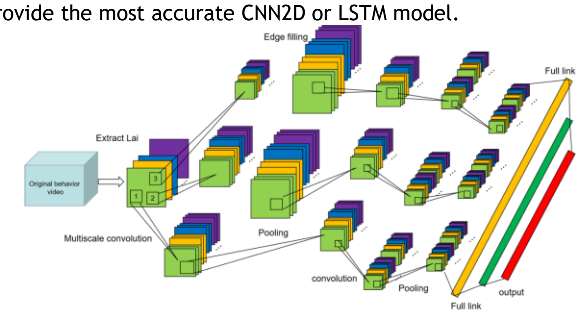

Fig 1 Proposed architecture

iii) Dataset collection:

Standing, lying, sitting, going upstairs, downstairs, and walking are all part of the human activity dataset that the proposed study employs. Our phones capture all of this. Click on the link below to download the dataset.

https://www.kaggle.com/datasets/drsaeedmohsen/ucihar-dataset/data

So, these are the dataset values

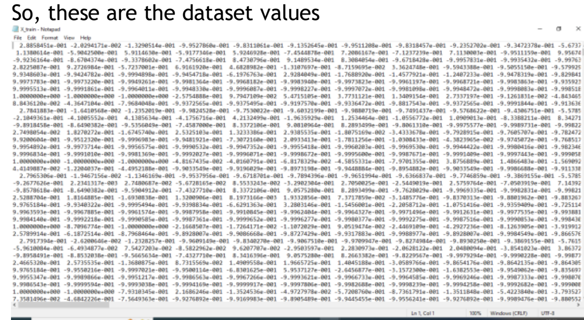

Fig 2 dataset values

iv) Data Processing:

Unstructured or semi-structured datasets include extraneous information. Training the model takes longer with this extraneous data, which could lead to worse results. It is necessary to pre-process data in order to optimise computational resources and the efficacy of machine learning models. In order for the model to make good predictions, text preparation is essential. Kenization, case normalisation, stopword removal, and numeral elimination are all part of the pre-processing. Because of case sensitivity, ML models will recognise "MACHINE" and "machine" as separate words. Lowercase data must be preprocessed.

v) Feature selection:

In order to build a trustworthy model, it is necessary to select features that are important, non-redundant, and of high reliability. With the proliferation of both large and diverse datasets, it is crucial to systematically reduce their dimensions. Enhancing a predictive model's efficacy while decreasing computing costs associated with modelling is the primary objective of feature selection.

One of the most important parts of feature engineering is feature selection, which involves finding the best features to feed into ML algorithms. In order to train a machine learning model with a smaller set of input variables, feature selection algorithms are used to filter out irrelevant features and duplicates. When compared to letting the ML model select the most important attributes, there are several benefits to selecting them beforehand.

EXPERIMENTAL RESULTS

Accuracy: One way to measure how well a model performs in a classification task is by looking at its accuracy, which is the percentage of right predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Fig 4 Accuracy graph

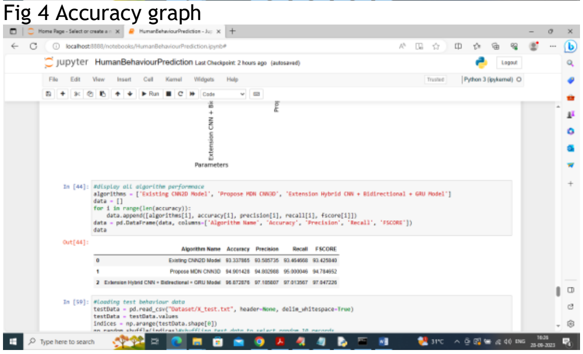

Fig 5 Performance Evaluation

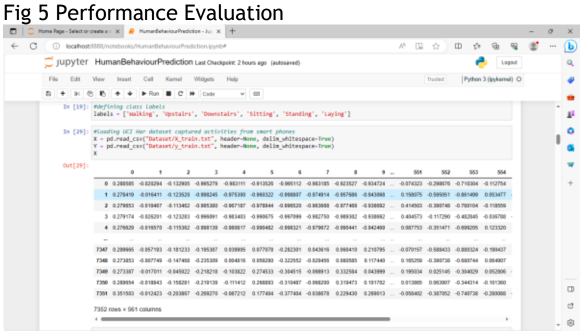

Fig 6 Dataset values page

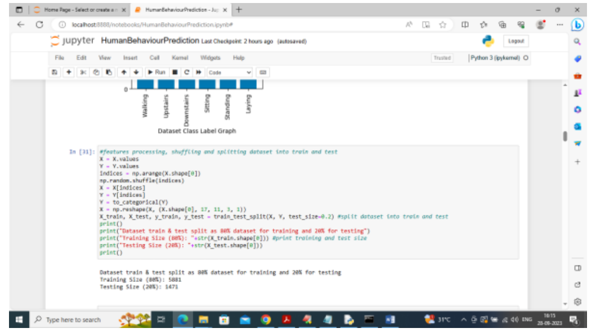

Fig 7 spitting dataset page

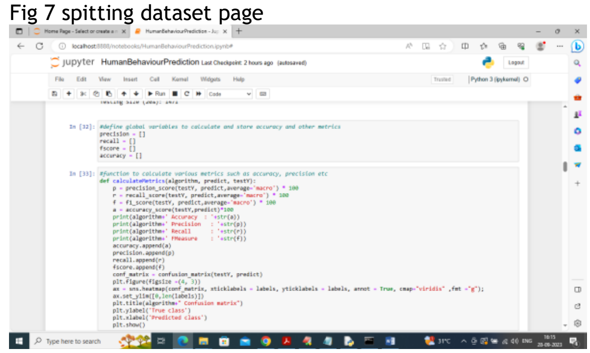

Fig 8 accuracy calculation page

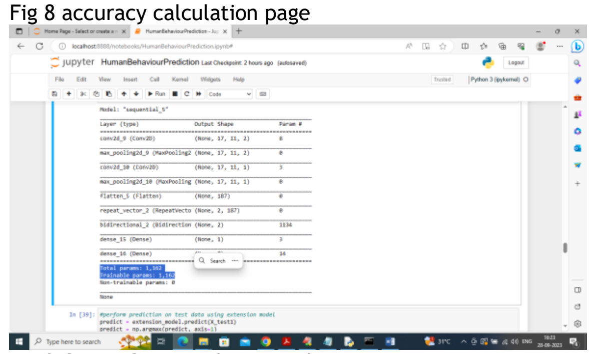

Fig 9 CNN + GRU + Bidirectional

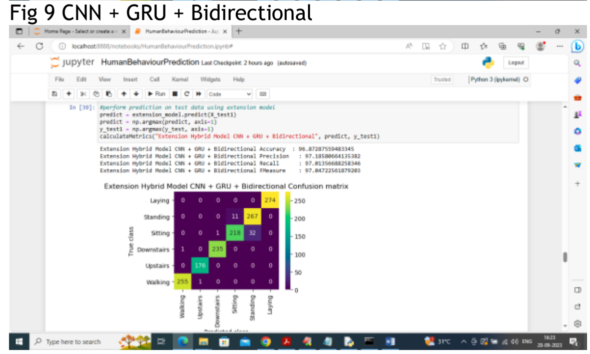

Fig 10 Run all algorithms

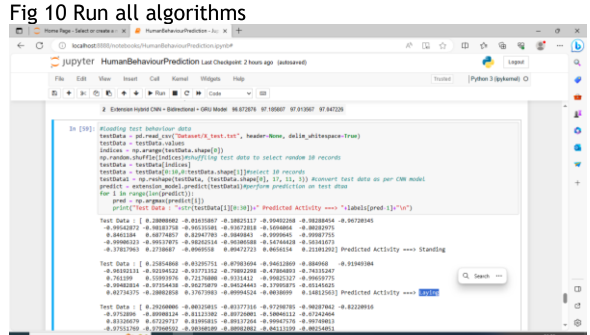

Fig 11 Accuracy Results

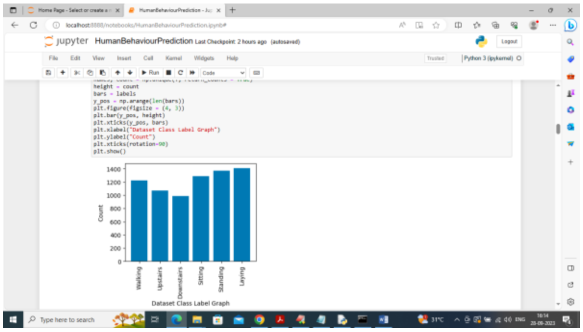

Fig 12. comparison graph

CONCLUSION

In this study, we provide a system for human behaviour recognition that makes use of an improved attention mechanism. By investigating the shortcomings of the channel attention mechanism, we suggest a better attention module. To prove that the enhanced attention module is functional, we examine visualisation results, increased network accuracy, additional network parameters, and so on. The practicality of cross-structure learning is demonstrated by the use of a multi-scale convolution kernel to extract behaviour traits across various receptive fields, which are then refined by a decently constructed convolution, pool, and complete connection layer. We require multi-stage progressive supervision since comparing supervision in different phases is so obvious. Additionally, the effect of model structure on soft migration is investigated. Convergence is effortless when the topology of the monitoring network is similar to that of the learning network. Greater sensor density can enhance data dimensionality and recognition accuracy in future studies. Model lightweight will be the focus of future study, given our method's module includes numerous parameters.

FUTURE SCOPE

Future studies can use additional sensors to improve the data dimension and accuracy of recognition. We will be focussing on making our method's model module less in weight in future efforts, as it now has a large number of parameters.

REFERENCES

- C. Ying and S. Gong, ''Human behavior recognition network based on improved channel attention mechanism,'' J. Electron. Inf., vol. 43, no. 12, pp. 3538–3545, 2021.
- 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. 132–135, 2022.
- 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. 233–239, 2022.
- S. Huang, ''Progress and application prospect of video behavior recognition,'' High Tech Ind., vol. 27, no. 12, pp. 38–41, 2021.
- [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. 375–381, 2021.
- 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. 154–158, 2022.
- 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. 1166–1172, 2022, doi: 10.16208/j.issn1000-7024.2022.04.036.
- Y. Lu, L. Fan, L. Guo, L. Qiu, and Y. Lu, ''Identification method and experiment of unsafe behaviors of subway passengers based on Kinect,'' China Work Saf. Sci. Technol., vol. 17, no. 12, pp. 162–168, 2021.
- Z. Zhai and Y. Zhao, ''DS convLSTM: A lightweight video behavior recognition model for edge environment,'' J. Commun. Univ. China, Natural Science Ed., vol. 28, no. 6, pp. 17–22, 2021.
- 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. 134–136, 2022.