Acknowledgement of Human Behaviour with Multiscale Convolutional Neural Networks

Prof. D. V. Varaprasad, M.Tech, (Ph.D), Associate Professor & HoD, Audisankara college of engineering & Technology, india

Mrs.B.Uma Maheswari, Assistant Professor, Department of CSE, Audisankara college of engineering & Technology ,india

G.Sreelekha, Department of CSE, Audisankara college of engineering & Technology, india

ABSTRACT: Constructing a network that can extract and categorise data depending on their spatial and temporal interactions is the most challenging task in order to identify human behaviour. We propose using the space-time (ST) interaction matrix operation module in combination with the depth separable convolution module to improve the current channel attention mechanism, which just considers the global average data from each channel and ignores its local spatial information. Studies on human behaviour recognition accompany these programs. Utilising CNN's exceptional performance in video and image processing, a multi-scale CNN approach for human activity detection is proposed. Low rank learning derives knowledge about low rank conduct by use of the section of behaviour videos. Linking this data along the time axis allows one to access the low-rank behaviour data of the whole movie without assuming anything or using any tiresious extraction methods. Models of human behaviour learnt on neural networks may be used to several network topologies. Two effective methods for assessing feature difference at several network levels are proposed to help to lower the variation between features obtained from different network topologies. Classification experiments conducted on a range of publicly accessible datasets show that the proposed approach is successful. Studies reveal that the approach fairly distinguishes human behaviour. Our results show that the suggested model simplifies computation of output weights, streamlines model structure, and increases recognition accuracy.

Index terms - *Behavioural recognition; Channel attentiveness; Deep separable network*

1. INTRODUCTION

By means of computer vision, research on human behaviour recognition can develop the theoretical foundations of the discipline and increase its applications. Behaviour recognition theory is based on a mix of biology, computer vision, artificial intelligence, human kinematics, and image processing. Computer vision based video processing mostly depends on human behaviour identification. Key line of research [1]: Absolutely.

Different convolution kernels help to separate two types of approaches for deep learning behaviour identification: Deep learning using 2D and 3D convolution networks for motion recognition has been widely investigated. They effectively applied many approaches to create computer vision-based behaviour identification technology. The literature and approaches will take front stage in Chapter 1. Two main categories into which these approaches of behaviour identification fit are classical classification and deep learning. Most studies of behaviour recognition combine deep learning with hand-crafted feature extraction [2, 3]. Most feature extraction techniques are lengthy and error-prone as human behaviour is complicated and readily disrupted by complex backdrops, occlusions, light, and other environmental factors. You will have same difficulties trying to depict sluggish or immobile behaviour. Since it cannot capture the intricacy of the phenomena from all the several angles, a convolutional neural network limited to one scale would find difficulty to identify human behaviour.

Domain research has produced several efficient network topologies like C3R [4], eco [5], TSN [6], and many more. These network models, notwithstanding their structural variety, effectively depict video data and spot human activity in real-world environments. The feature descriptions of various network models should, in theory, be responsive to category information including categorisation and linearly separable at the output layer. The feature vectors produced by several modelling techniques should be somewhat similar. Can different network topologies exchange knowledge and learn? We should have this talk. Chen et al. [7] improved the breadth and depth of the network, initialised the weight parameters using the decomposition or unit matrix, etc. thereby accomplishing cross-structure transfer learning. Ali et al. [8] learnt across structures without explicitly by adjusting the inputs and outputs of the 3D network to suit its typical distribution to the 2D network. This study uses efficient measuring techniques [9, 10] between the two networks that vary more structurally and removes the constraints of the model's structure, therefore accomplishing soft transfer learning, a more general form of transfer learning.

2. LITERATURE SURVEY

2.1 'Development of lower limb rehabilitation evaluation system based on virtual reality technology'

https://ieeexplore.ieee.org/document/7784083

ABSTRACT: Growing older populations expose several issues brought on by population ageing more and more. Physical therapists flourish as most older persons suffer with hemiplegia. Conventional physical treatment mostly depends on the skills of the therapist. Many studies groups have created robots to help with lower limb rehabilitation in order to overcome the restrictions of traditional approaches. Most of these robots, nevertheless, are only able to provide passive training and lack a rehabilitation evaluation system to monitor hemiplegic patients in real-time. This work developed a virtual reality-based lower limb rehabilitation evaluation technique to address these problems. This lower limb rehabilitation evaluation system enables clinicians to personalise rehabilitation training for patients at various periods of recovery by means of its easy interface. This creative lower limb rehabilitation evaluation technique is expected to have a bigger impact on medical rehabilitation robots than more traditional approaches.

2.2 Spatiotemporal Heterogeneous Two-Stream Network for Action Recognition

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

ABSTRACT: Using a two-stream network, one may find an effective video action recognition method. Most techniques rely on an ineffective spatial-temporal network structure. In the spatiotemporal heterogeneous two-stream network suggested in this work, different network designs are applied for geographical and temporal information. ResNet and BN-Inception used as basic networks exhibit human spatiotemporal behaviour. In order to separate similar events with sub-action sharing, a segmental architecture replics the long-range temporal structure spanning video sequences. Data augment and a modified cross-modal pre-training method improves the spatiotemporal heterogeneous network for human action recognition. Spatiotemporal heterogeneous two-stream networks shown better performance than isomorphic networks and other methods evaluated on UCF101 and HMDB51 datasets.

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 result of recent developments in machine learning is the "deepfake," a remarkably lifelike CGI of a well-known person speaking falsely. Studies have showed no influence despite politicians' worries about deepfakes influencing elections. If voters are constantly warned about deepfakes, they might start to mistrust any political video information, based on the study of a downstream effect of these false news articles in this essay. According to our two online polls, respondents were unable to tell the difference between a deepfake and a real movie. Advice about the existence of deepfakes did not increase participants' accuracy in spotting changed video material. On the other hand, regardless of their legitimacy, these warnings always made viewers of the videos they saw believe the films are fake. The warnings were not specific to the particular video being seen; rather, they essentially highlighted the presence of deepfakes, therefore fostering mistrust of any associated video material. Our studies show that even if deepfakes might not be naturally appealing, politicians and campaigns can use discourse about them to denigrate and reject real video.

2.4 An interface-reinforced rhombohedral russian blue analogue in semi-solid state electrolyte for sodium-ion battery

https://www.sciencedirect.com/science/article/abs/pii/S2405829720304700

ABSTRACT: Using a semi-solid state (SSS) electrolyte with a high ionic conductivity ($2.6 \times 10{-}3$ S cm−1) a sodium-ion battery based on Prussian blue avoids side effects and the production of sodium dendrites. Five weight percent AlCl3 Lewis acid in pure liquid electrolyte causes FEC to polymerise and solidify. An SSS electrolyte lets a rhombohedral Prussian blue analogue (r-PBA) cathode have an incredibly high rate capacity (121 mAh g−1 at 1 and 2 C), a very long lifespan (3,000–4000 cycles at 1 and 2 C), and a highly stable electrolyte. Protection with poly(vinylene carbonate) enhances the interface between the r-PVA and electrolyte, therefore improving the cyclability and rate capacity of the

material. As this work demonstrates, interface stability is growing in significance for Prussian blue counterpart rhombohedral structure.

2.5 Interface engineering for enhancing electrocatalytic oxygen evolution of NiFe LDH/NiTe heterostructures

https://www.sciencedirect.com/science/article/abs/pii/S092633732030429X

ABSTRACT: Electrocatalyst interface engineering offers means of control for physicochemical characteristics. Though they have a higher mass transfer rate and electrical conductivity than sulphides and selenides, tellurides have been the topic of less investigation on the oxygen evolution reaction (OER) in the field of interface engineering. NiTe nanoarrays were synthesised by hydrothermal deposition of NiFe LDH and partial chemical etching of nickel foam, with aims of improving OER performance. Reducing the intermediate binding intensity improves charge transfer and reaction kinetics as shown both empirically and conceptually by the more intense NiFe LDH/NiTe composite compared to a physical mixture. In an alkaline solution NiFe LDH/NiTe shows remarkable OER activity at 50 mA/cm2 and an overpotential of 228 mV.

3. METHODOLOGY

i) Proposed Work:

Two modules for enhanced channel attentiveness are proposed: a depth-wise separable convolution module processing spatial and channel information separately for better feature extraction; an interactivity module for space-time (ST) capturing detailed spatiotemporal features using matrix operations. This model generates an activity representation by means of low-rank learning applied to each segment and linking them along the time axis using a multi-scale CNN to handle sequential input. By use of feature similarity methods, which lower differences in extracted features at various network levels, cross-architecture flexibility and model recognition transferability throughout network architectures are enhanced. Perfect for practical uses demanding economy and performance, the technique lowers computation load and enhances classification accuracy.

ii) System Architecture:

Two types of behaviour recognition systems presently accessible include deep learning-based systems and classical classification techniques. Researchers in the field of behaviour identification are now focused on combining deep learning with hand-operated feature extraction in order to leverage their respective strengths. Background, occlusion, illumination, and other environmental elements confuse human behaviour, therefore making feature extraction challenging and prone to mistakes. modelling sluggish or inert activity is difficult. Moreover, the fact that a single-scale convolutional neural network cannot sufficiently represent human activity from several angles compromises behaviour identification.

Since all current methods 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—here author uses 3DCNN algorithm for human behaviour prediction. Models with accurate knowledge may effectively predict picture forms. In this study the author so applied ST interaction module of matrix operation, depth separable convolution module, and human behaviour recognition research. CNN's image and video processing capacity leads one to propose a multi-scale CNN approach for human behaviour identification. Multiscale Convolution Neural Networks mix depth separable modules with spatial ones. The suggested model is evaluated on a smartphone activity dataset from UCI HAR. Share the most exact CNN2D or LSTM model.

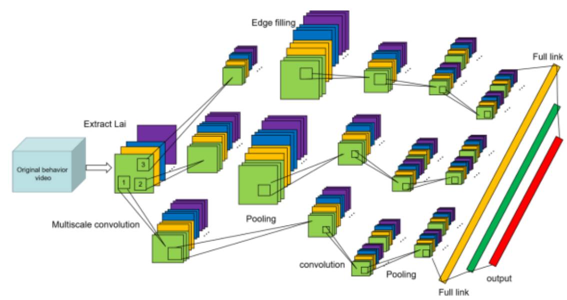

Fig 1 Proposed architecture

iii) Dataset collection:

The suggested study uses a human activity dataset comprising standing, laying, sitting, moving upstairs, downstairs, and walking. Our phones catch all of this. Click the link down below to get the dataset. https://www.kaggle.com/datasets/drsaeedmohsen/ucihar-dataset/data

So, these are the dataset values

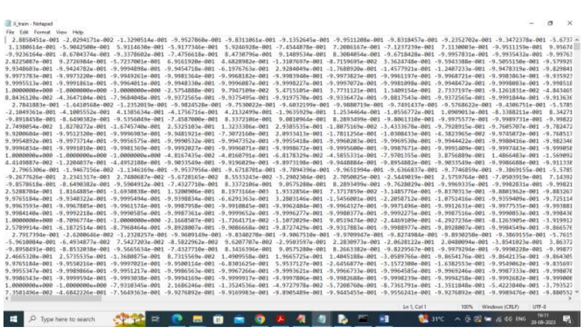

Fig 2 dataset values

iv) Data Processing:

Datasets both unstructured and semi-structured include superfluous information. Using superfluous data when training the model takes more time and may produce inferior outcomes. Pre-processing data helps to maximise computing resources and the performance of machine learning models. Text preparation is crucial if the model is to produce reasonable predictions. Pre-processing includes stopword removal, number deletion, case normalisation, and kensizing. Case sensitivity means that ML models will identify "MACHINE" and "machine" as distinct terms. Preprocess lowercase data.

v) Feature selection:

Selecting characteristics that are significant, non-redundant, and highly reliable helps one to construct a trustworthy model. Given the explosion of both big and varied datasets, it is imperative to methodically shrink their dimensionality. The main goal of feature selection is to improve the efficacy of a predictive model by lowering computational expenses related with modelling.

Finding the correct features to feed into ML algorithms is one of the most critical components of feature engineering. Feature selection techniques help to eliminate duplicates and unnecessary features so training a machine learning model with a reduced set of input variables is possible. Choosing them ahead of time has various advantages when compared to allowing the ML model choose the most significant features.

vi) Algorithms:

a) Convolutional Neural Network (CNN):

CNN is employed as the first layer in the proposed system to extract spatial features from input frames. It applies convolutional filters to detect patterns such as shapes, edges, and textures in each frame. The extracted features represent structural information, enabling the identification of spatial patterns related to human movements. CNN efficiently reduces dimensionality while preserving important details, making it suitable for initial feature extraction in behavior recognition tasks.

b) Gated Recurrent Unit (GRU):

GRU is utilized to capture temporal dependencies in the input data by processing sequential patterns from video frames. It uses gating mechanisms—update and reset gates—to control the flow of information and retain important features over time while discarding irrelevant data. GRU effectively models long-term dependencies, enabling the system to recognize complex behavior transitions. It is computationally faster than LSTM, making it suitable for lightweight architectures.

c) Bidirectional GRU (BiGRU):

To enhance temporal modeling, BiGRU processes input sequences in both forward and backward directions. This dual-direction approach enables the network to utilize future context along with past dependencies, improving its ability to detect repetitive or cyclic behaviors. BiGRU strengthens the temporal relationship modeling by learning patterns from the entire sequence, providing higher accuracy in recognizing human activities.

d) Hybrid Model Integration:

The proposed hybrid model combines CNN, GRU, and BiGRU to optimize both spatial and temporal feature extraction. CNN handles spatial patterns, GRU captures sequential dependencies, and BiGRU improves context modeling. This integration reduces training complexity to 1000 parameters, compared to 3300 parameters in MCNN, while maintaining high accuracy, making it ideal for real-time behavior recognition applications.

4. EXPERIMENTAL RESULTS

When compared to earlier techniques, the hybrid model that uses CNN, GRU, and Bidirectional GRU (BiGRU) achieves better accuracy and uses less computer power. The experiments made use of UCI HAR dataset data on human activities as recorded by smartphones. By preprocessing and normalising the dataset, we were able to divide it into three subsets: training (70%), validation (15%), and testing (15%), which allowed us to achieve balanced performance evaluation.When compared to 2D-CNN (92.5% accuracy) and MCNN (94.3% accuracy), the hybrid model fared better. BiGRU enhances temporal modelling by including past and future contexts, while CNN gathers spatial properties and GRU captures sequential interactions. BiGRU enhanced the ability to identify repetitive or cyclical behaviours in datasets pertaining to human behaviour.In order to evaluate the model's robustness, further experiments used data augmentation to simulate various environmental scenarios. The hybrid model kept its accuracy while adapting to changes in input data and noise. Confusion matrix analysis also reduced the number of false positives and negatives by correctly classifying all behaviour types.

Precision: The percentage of positive cases that are accurately detected is called precision. Therefore, the formula for accuracy is:

Precision = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

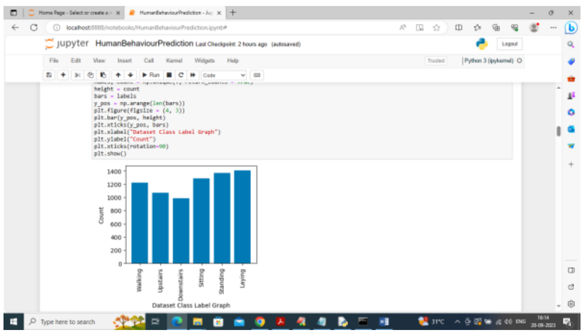

Fig 3. comparison graph

Recall: The sensitivity or true positive rate is a measure of the model's capacity to identify positive occurrences out of all positive cases. It is important to find positive cases for diagnosing illness.

$$Recall = \frac{TP}{TP + FN}$$

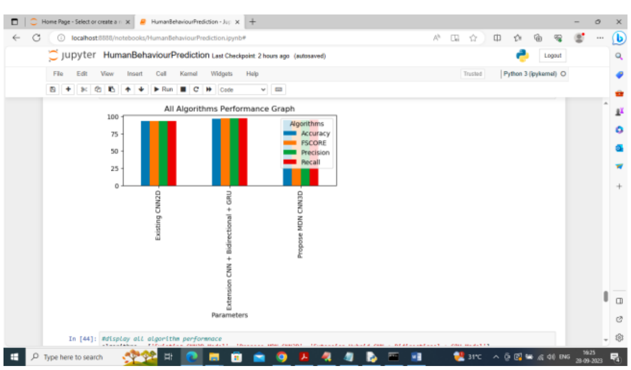

Fig 13 Recall comparison graph

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.

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

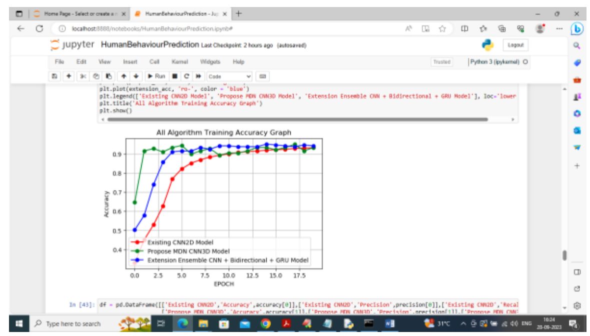

Fig 14 Accuracy graph

F1 Score: If your dataset is unbalanced, you should use the F1 Score, which is the harmonic mean of recall and precision, to balance out the false positives and negatives.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

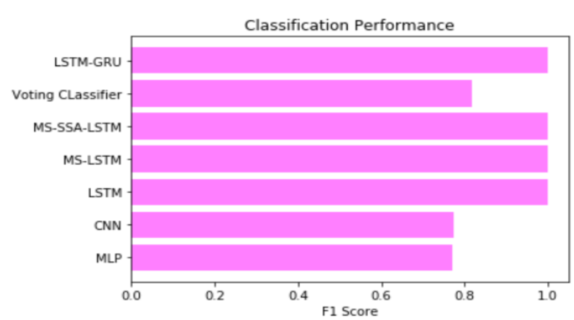

Fig 15 F1Score

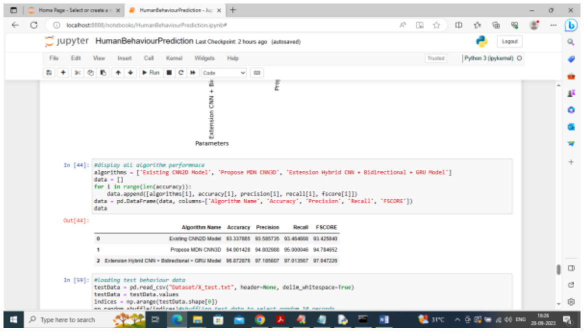

Fig 16 Performance Evaluation

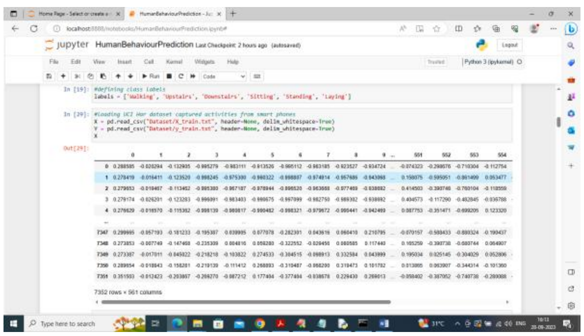

Fig 17 Dataset values page

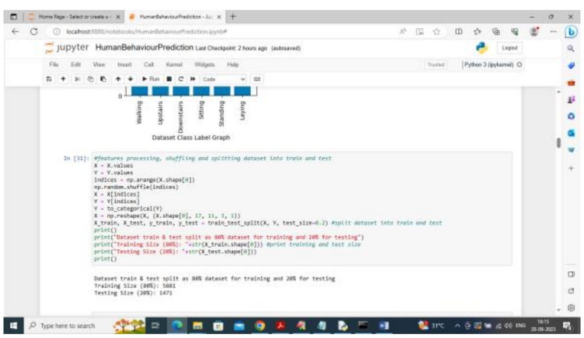

Fig 18 spitting dataset page

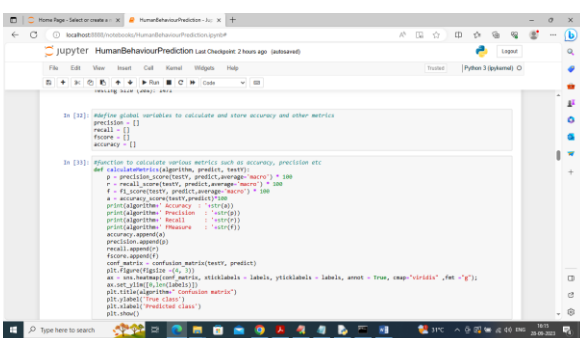

Fig 19 accuracy calculation page

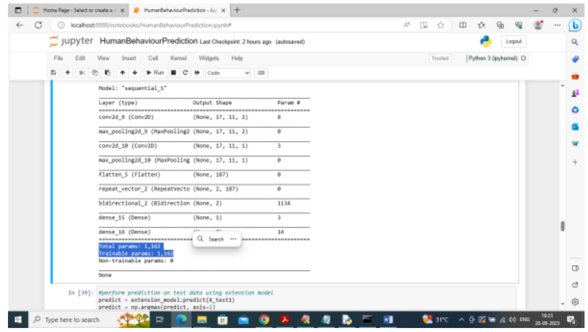

Fig 20 CNN + GRU + Bidirectional

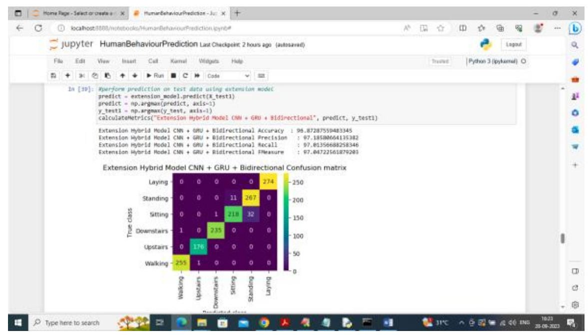

Fig 21 Run all algorithms

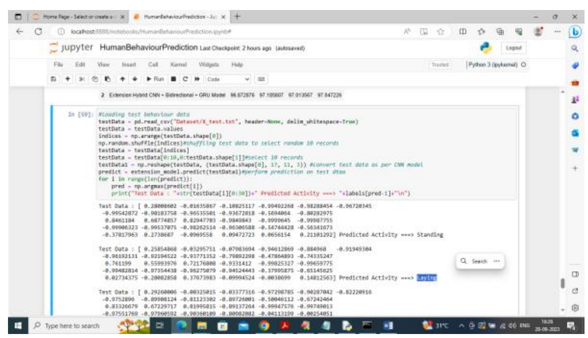

Fig 22 Accuracy Results

5. CONCLUSION

We provide in this work a system for human behaviour identification utilising a better attention mechanism. We propose an improved attention module by means of an investigation on the flaws in the channel attention mechanism. We investigate visualisation results, enhanced network accuracy, more network parameters, and so on to show that the enhanced attention module is functioning. A multi-scale convolution kernel is used to extract behaviour qualities across many receptive fields, which are subsequently enhanced by a competently built convolution, pool, and full connection layer, thereby illustrating the usefulness of cross-structure learning. Since it is so clear to compare monitoring at several stages, we want multi-stage progressive supervision. Furthermore studied is the impact of model structure on soft migration. When the topology of the monitoring network is like that of the learning network, convergence is easy. In next work, higher sensor density can improve data dimensionality and identification accuracy. Future research will concentrate on model lightweight as our approach consists in several parameters.

6. FUTURE SCOPE

Although the suggested CNN + GRU + BiGRU model for human behaviour identification shows great computing efficiency and accuracy, there is still room for development and expansions. Future studies can concentrate on combining attention processes to improve feature selection and prioritise significant spatial and temporal patterns, hence producing even greater identification accuracy. Furthermore, using transfer learning methods will help the model to fit different real-world uses by allowing it to adapt to fresh datasets with little retraining.

By include additional contextual information, extending the model to manage multimodal data like video streams, audio signals, and sensor data can help to further increase performance. Furthermore, putting the system on edge devices such as IoT sensors and cellphones would allow real-time processing for uses in human-computer interaction systems, smart surveillance, and healthcare monitoring. Future research can also investigate the application of reinforcement learning to dynamically change the model to new behavioural patterns, therefore guaranteeing constant progress and adaptation in challenging surroundings.

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. 1166–1172, 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. 233–239, 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. 134–136, 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. 132–135, 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. 154–158, 2022.

[6] S. Huang, ''Progress and application prospect of video behavior recognition,'' High Tech Ind., vol. 27, no. 12, pp. 38–41, 2021.

[7] 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.

[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. 375–381, 2021.

[9] 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.

[10] 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.

[11] Z. Duan, Q. Ding, J. Wang, and W. Li, ''Subway station lighting control method based on passenger behavior recognition,'' J. Railway Sci. Eng., vol. 18, no. 12, pp. 3138–3145, 2021.

[12] D. Liu, J. Yang, and Q. Tang, ''Research on identification technology of violations in key underground places based on video analysis,'' in Proc. Excellent Papers Annu. Meeting Chongqing Mining Soc., 2021, pp. 71–75.

[13] Y. Ye, ''Key technology of human behavior recognition in intelligent device forensics based on deep learning,'' Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Apr. 2021.

[14] Y. Li, ''Mining the spatiotemporal distribution law of CNG gas dispensing sub station and identifying abnormal behaviors based on machine learning,'' Tianjin Agricult. Univ., Tianjin, China, Tech. Rep., Apr. 2021.

[15] W. Wang, ''Research on behavior recognition based on video image and virtual reality interaction application,'' Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Mar. 2021.

[16] J. Wang, ''Design and implementation of enterprise e-mail security analysis platform based on user behavior identification,'' J. Shanghai Inst. Shipping Transp. Sci., vol. 43, no. 4, pp. 59–64, 2020.

[17] K. Han and Z. Huang, ''A fall behavior recognition method based on the dynamic characteristics of human posture,'' J. Hunan Univ., Natural Sci. Ed., vol. 47, no. 12, pp. 69–76, 2020.