International Journal of Healthcare Information Systems and Informatics
Volume 19 • Issue 1

Application of Behavior Recognition Technology Based on Deep Learning in Elderly Care

Shihui Zhang, HeBei North University, China
Jing Mi, HeBei North University, China*
Naidi Liu, HeBei North University, China

ABSTRACT

China is currently one of the countries with the largest elderly population in the world, and the issue of population aging has become a widespread concern. The behavior recognition algorithm based on deep learning is currently the main behavior recognition algorithm and one of the basic technologies in the field of computer vision. In existing research, the method of constructing complex classification models based on manual feature representation can no longer meet the requirements of high recognition accuracy and applicability, and the introduction of deep learning has brought new development directions for behavior recognition. Therefore, this article aims to study how to apply deep learning-based behavior recognition technology more accurately and effectively in the care of elderly people in the context of "artificial intelligence."

KEYWORDS

Aged Care Services, Behavioral Recognition Technology, Deep Learning

1 INTRODUCTION

China is transitioning towards an aging society, and unlike other countries, its per capita income is lower, resulting in a lower level of social security (Chang et al., 2022). Since China is the country with the largest population in the world, the scale of population development is also the largest. Data shows that in 2010, the proportion of elderly people aged 60 or older in China accounted for 13.26% of the total population, which rose to 18.7% in the latest census. Due to uneven development across the country, some provinces have even entered a stage of deep aging. Moreover, the population of the country has exceeded 1.4 billion, and the number of elderly people even exceeds the total population of many countries (Qiu et al., 2022).

As of the end of 2018, the population aged 65 and above in China had reached 166.58 million, accounting for approximately 11.94% of the total population in China, a year-on-year increase of

DOI: 10.4018/IJHISI.336548 *Corresponding Author

0.6% (Cheng et al., 2023). In 2017, though China's population decreased by 2 million, the birth rate still exceeded 19.4%. The population statistics and predictions for Chinese population aged 65 and above from 2000 to 2035 are shown in Figure 1.

At present, the natural and social structure of the Chinese people is rapidly changing. In the next 50 years, the size of China's elderly population will sharply increase. It is expected that by 2025, the elderly population aged 60 and above in China will reach 318 million, with the proportion of the elderly population accounting for 21% of the national population. At the same time, as the capacity of family caregivers gradually decreases, there will be a large number of long-term care needs, which will have a huge impact on China's economic growth and pose challenges to the national social security system.

Nowadays, there are still many problems in elderly care in China, such as cost burden, multiple elderly problems, declining medical costs, and a lack of professional service systems for the elderly. Determining how to build a healthy aging society has become a major social problem that China urgently needs to solve. The market size of China's elderly care industry is shown in Figure 2.

According to the "*China Elderly Care Development Report (2020),*" in 2019, the number of elderly care institutions in China reached 68000, with 71000 beds. However, the number of both elderly care institutions and beds is still far from meeting the needs of the elderly. It can be foreseen that the market size of China's elderly care industry will continue to grow in the future and will drive the development of related industry chains, such as health products, medical devices, medical services, etc. Meanwhile, with the development of technology and policy support, the service quality and level of the elderly care industry will continue to improve.

It is not uncommon for an elderly person's daily life to be filled with a variety of unexpected incidents. Accidents, like falls and heart attacks, can pose a threat to safety and could even result in death if no one is there to provide emergency care and treatment in a timely manner. As a result, prompt detection of any abnormal situations in the elderly can enhance both the industry's service quality and the safety and security of their everyday life (Karthickkumar & Kumar, 2020). The behavior recognition technology uses algorithms to extract similar and different features between behaviors from the collected human target images (including RGB images, grayscale images, infrared images, RGB-D depth images, etc.), and uses features to classify (Ozcan & Basturk, 2020). Due to

Figure 1. Population statistics and projections of Chinese Population aged 65 and above from 2000 to 2035

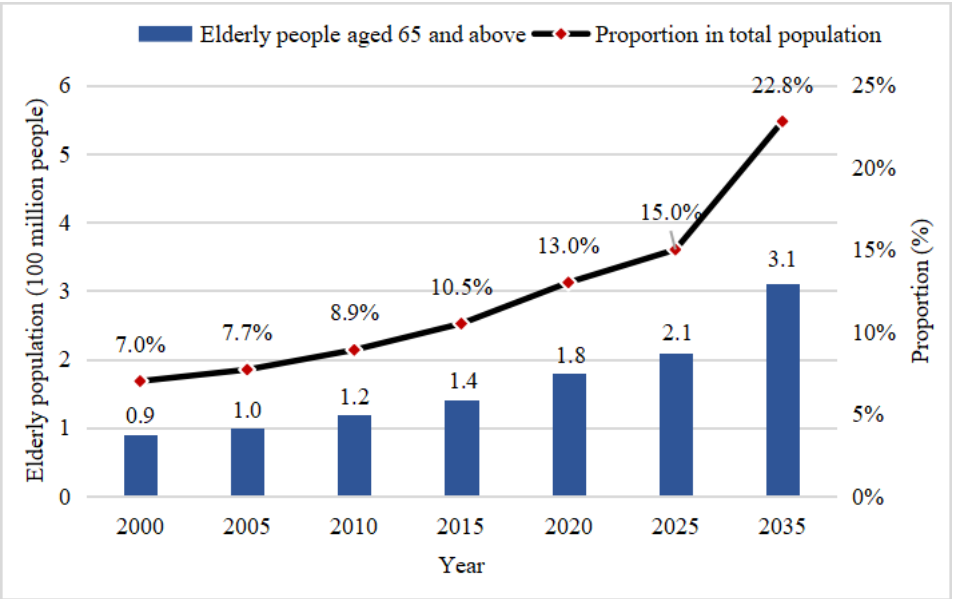
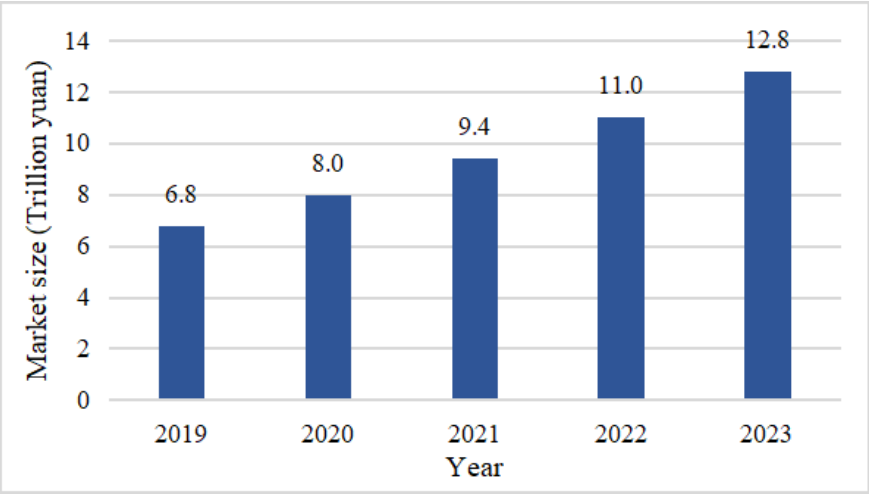


Figure 2. Market size of China's elderly care industry



the high complexity of behavior in both time and space, many identical behaviors have significant differences due to differences in executors or timing. At the same time, there may be similarities between different behaviors (Deep et al., 2019). The action's rich structural information should be adequately expressed by the action features that were built. Subsequent recognition outcomes will also be impacted if the distinctiveness of various action types and the similarity of various versions of the same action (e.g., forward and backward falls) cannot be well represented. While algorithms can automatically extract features using deep learning feature extraction approaches, manual feature extraction takes specialized skill.

This article mainly studies a human behavior recognition technology based on deep learning. At present, elderly people choose to rely mainly on home care or institutional care, and this article hopes that the recognition system can be applied to these two scenarios. This study designs an overall scheme based on a monocular RGB camera, sorts out common abnormal behaviors of the elderly, studies human pose estimation and object detection algorithms, and designs a behavior recognition scheme for human targets. The recognition system is required to accurately and timely identify the position and behavior category of human targets, and provide reminders for abnormal behavior.

2 THEORETICAL KNOWLEDGE OF DEEP LEARNING

2.1 Deep Learning

In recent years, deep learning has gradually become a very important research method in various research fields. Major domestic companies have also carried out more in-depth research, including Baidu, Google, etc. (Park et al., 2019). In the research and development process of the next generation of artificial intelligence, AlphaGo defeated Li Shi Shi to become the world Go champion, adding a strong touch to the world of human intelligence. Deep learning methods have also produced many important technological breakthroughs in other fields such as image recognition, speech recognition, and image mining (Bolhasani et al., 2021). Deep learning frameworks are able to sort through a lot of data and find the best features. Since the same convolutional kernel will process the entire image rather than utilizing several convolutional kernels to conduct operations on different portions of the image, convolution only needs to learn the parameters of a small number of convolutional kernels. It can decrease the number of convolutional kernels while at the same time allowing the same convolutional kernel to be used for the same processing in another image. Grid-structured data is

frequently utilized in the analysis and recognition of image data because it is easier to convolve (Phyo et al., 2019). With the increasing complexity of China's aging society, deep learning overcomes the limitations of previous research methods and provides new ideas for improving the pension system. One of the most effective ways to do this is through a deep confidence network.

A generation method of Deep Belief Networks (DBNs) enables a neural network to obtain training results with maximum likelihood by training the weight relationships between neurons. DBN is composed of multiple layers of neurons, which are further divided into explicit neurons and hidden neurons. The explicit neurons are used for inputting data, while the hidden neurons are used for extracting and learning data. If there is an n-layer DBN model, the model can be constructed based on the joint distribution between visible layer variables and hidden layer variables $h_i^{(k)}$, k=1, 2, 3, ..., I. Each hidden layer is composed of binary units $h_i^{(k)}$, so for the DBN model as a whole, its joint probability P(v, $h^{(1)}$, $h^{(2)}$, ..., $h^{(t)}$) is calculated as P(v, $h^{(1)}$, $h^{(2)}$, ..., $h^{(t)}$) =P(v | $h^{(1)}$) P ($h^{(1)}$,$h^{(2)}$)... P[$h^{(t-2)}$)| $h^{(t-1)}$]P[$h^{(t-1)}$)| $h^{(t)}$], where v = $h^{(0)}$. P [$h^{(k)}$)|$h^{(k+1)}$] is the distribution of factorial conditions between the kth layer and the k+1th layer, that is, $P[h^{(k)} \mid h^{(k+1)}] = \prod_i P[h_i^{(k)} \mid h^{(k+1)}]$.

Therefore, when building a model, an important step is to determine the number of nodes between the input layer and the hidden layer. In the model, the input layer nodes primarily provide data to the acquired set of information. The number of output layer nodes in the model is formed after model learning, which can help the model understand the hidden content of the information set, so as to show more complex nonlinear relationships embedded in the information set in the model. If the number of hidden layer nodes simulated is too small, the model can have errors, and conversely, if the number of hidden layer nodes is too large, overfitting can occur, resulting in more consequences than modeling bias. Therefore, it is important to determine the number of hidden layer nodes when defining the model.

2.2 Model Structure of Deep Learning

Neurons and perceptron models: The concept of neurons comes from biology. A simple neuron usually consists of several dendrites, an axon, and several axon-ends (Soufian et al., 2022). Among them, dendrites are mainly used to receive input information, axon ends transmit information to the next neuron, and axons are used to connect dendrites and axon ends (Yu et al., 2022). Thanks to the special memory and information processing mechanisms of neurons, in 1943 psychologist McCulloch and mathematician Pits abstracted the basic structure of deep neural network units: the neuron model, by referring to the structure of biological neurons (Byeon et al., 2021). The basic model of a neuron consists of three parts: input, output, and computation.

Activation function: The emergence of the activation function is mainly due to the introduction of nonlinear operations in neural networks, which improves the ability of models to express complex environments and data (Lentzas & Vrakas, 2022). At present, the activation functions commonly used in deep neural networks are Sigmoid, Tanh, and ReLU functions. Among them, the ReLU function is the most widely used activation function in CNNs, and it is also used in the activation layer of this article.

Loss function: The loss function is used to measure the gap between the predicted value of the model $\hat{y}$ and the actual value y (J represents the loss). The smaller the value of the loss function, the better the robustness of the model. Common loss functions include mean absolute error (MAE), mean square error (MSE), hinge loss function, and cross-entropy (CE). Among them, CE is the most commonly used loss function in CNNs, and it is also the loss measurement function used in this paper.

(1) The MAE loss function, also known as L1 loss function, takes the absolute error between the predicted value and the real value as the distance. The mathematical expression is shown in equation (1).

$$J_{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\left(y_i - \hat{y}_i\right)\right| \tag{1}$$

(2) The MSE loss function, also known as the L2 loss function or Euclidean distance, takes the sum of squares of the error as the distance and is the most commonly used loss function in regression tasks, and its expression is equation (2).

$$J_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{2}$$

(3) The Hinge loss function is a binary classification loss function, which is suitable for the classification of the maximum interval (max-marg in), so it is often used in SVM models. The expression for Hinge loss is equation (3).

$$J_{Hinge} = \sum_{i-1}^{N} max\left(0,1 - y_i\hat{y}_i\right) \tag{3}$$

where, $\widehat{y_i} \in \left(-1,1\right)$, $y_i = \pm 1$, which represents positive and negative samples in binary classification.

(4) CE loss function is one of the most used loss functions in CNN classification tasks. According to the requirements of classification tasks, it can be divided into two cases, namely, binary classification and multi classification, which are collectively referred to as CE loss function. CE is an important concept in information theory, mainly used to measure the differential information between two probability distributions. Suppose there are two probability distributions p and q, where p represents the true distribution and q represents the predicted distribution, then the difference information between them can be represented by cross entropy, and the equation is shown in equation (4).

$$\mathrm{H}\left(\mathrm{p},\mathrm{q}\right) = \sum_{\mathrm{x}}\mathrm{p}\left(\mathrm{x}\right)\log\left(\frac{1}{\mathrm{q}\left(\mathrm{x}\right)}\right) = -\sum_{\mathrm{x}}\mathrm{p}\left(\mathrm{x}\right)\log\mathrm{q}\left(\mathrm{X}\right) \tag{4}$$

2.3 Deep Neural Networks

A convolutional neural network (CNN) is a deep neural network that typically consists of one or more layers of convolution, pooling, and complete connections. The corresponding model also typically includes input and output layers, the structure of which is shown in Figure 3.

(1) Input layer. In video recognition tasks for human behavior, it is often necessary to input one or more continuous data images or intermediate layer feature matrices.
(2) Convolutional layer. Convolution is an important part of CNN feature extraction structure, with two main features: local perception and parameter sharing. Local perception means that input

Figure 3. Schematic diagram of CNN structure

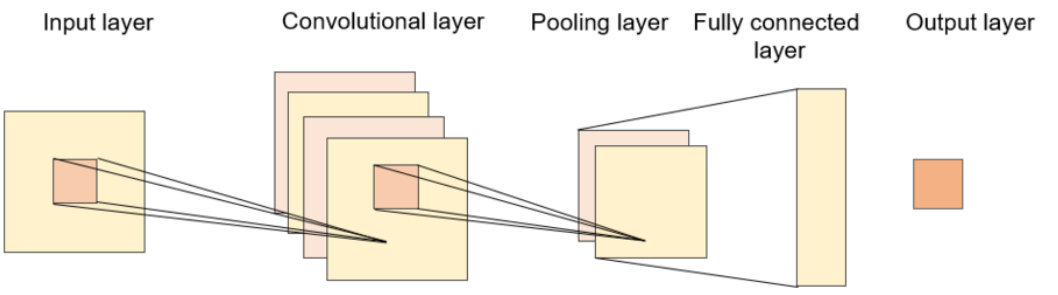


image blocks are extracted from features using convolution kernels of a specified size, which greatly reduces network parameters compared to traditional fully connected networks (Hsueh et al., 2020). Shared parameters mean that the same convolution kernel keeps its own parameters unchanged when convolving the input feature map, reducing the number of parameters that the network needs to form.

(3) Pooling layer. The pooling layer has two functions. One is to further reduce the dimensionality of the information extracted from the convolutional layer, reducing computational complexity. The second is to enhance the invariance of image features, making them more robust in terms of image offset, rotation, and other aspects. Common pool operations are maximum pool and average pool.

(4) Fully connected layer. The fully connected layer mainly consists of converting the feature output matrix of the convolutional layer or pool layer into a one-dimensional feature vector, and then transmitting it to the output layer. This is basically equivalent to a linear transformation in the space of the feature map. Assuming that the input characteristic moment matrix is $X \in R^m$, and the output result after conversion is $Y \in R^m$, then the formula for its transformation is equation (5).

$$Y = f\left(W_f X + b_f\right) \tag{5}$$

Where, $W_f \in R^{m \times n}$ is the weight, $b_f \in R^n$ is the bias, and f is the activation function.

(5) Output layer. Receive a one-dimensional feature vector with a fully connected output, and then map the probabilistic feature values using a Softmax classifier. Assuming the output is $P \in R^k$, then the expression of its mapping is shown in equation (6). If the category of the classification task is K, Softmax maps the input $Y \in R^n$ to K real numbers (0,1) and ensures that their sum is 1. The expression is shown in equation (7).

$$\mathrm{P} = \mathrm{Softmax}\left(\mathrm{W_s Y} + \mathrm{b_s}\right) \tag{6}$$

$$\sum_{i=1}^{k} P_i = 1 \tag{7}$$

Where, $Ws \in R^{n \times k}$ is the weight, and $b_s \in R^k$ is the bias.

3. HUMAN BEHAVIOR RECOGNITION TECHNOLOGY

3.1 Overview of Behavioral Recognition

There are many types of abnormal behaviors among elderly people, and there are many influencing factors. However, abnormal behavior has a low-frequency characteristic, and daily behavior is a widespread behavioral activity. This article selects falling and lying down as common abnormal behaviors among elderly people, while standing up, sitting down, walking, and standing are normal behaviors. Falls are the most common part of abnormal behavior among elderly people, and the incidence is also very high. From a medical perspective, falls are caused by loss of balance and physical instability (Pandey & Litoriya, 2019). The main reasons for falls include: physiological factors, like decreased mobility in the elderly, leading to an increased risk of falls; effects of cardiovascular and other diseases; psychotropic drugs, cardiovascular drugs, etc. can affect a person's mental state, vision, balance, and other aspects, causing falls; and external risk factors, such as poor lighting conditions and uneven and slippery road surfaces.

In experimental research, simple static motion detection forms are often identified. Action gain activation is used to improve body movement inquiry, also known as marking. The effect of representing motion features is of great significance for body motion detection (Zhu et al., 2022). An effective action representation should meet the following criteria: (1) Discriminability. Differentiating behavior provides data containing this type of information, while another behavior distinguishes different behaviors. (2) Efficiency. It should be easy to calculate and understand. (3) Low dimensionality. It means considering the quality and cost of recognition. Finally, based on the extraction of action representations, motion features can be captured in spatiotemporal sequences and grouped through data analysis. Action feature recognition can be seen as a process in which mathematical notation are trained and classified by combining prior knowledge (Hassan et al., 2018).

3.2 Mainstream Behavior Recognition Technology

(1) Method based on 3D convolutional neural network. The 3D convolutional network method refers to the use of 3D convolutional neural networks to obtain the temporal and spatial characteristics of the video together with the 3D convolution kernel (Aslan et al., 2020). The 3D convolution calculation kernel refers to the extension of the 2D convolution kernel in the time dimension. The operation of 3D convolution is somewhat similar to processing a batch of images using 2D convolution, but there are temporal connections in these image sequences. Therefore, the convolution kernel of 3D convolution is also different, adding a dimension that can simultaneously perceive adjacent frames of images and extract temporal features. (2) Method based on CNN+ time dynamic coding. Using the new technology of CNN and time dynamic coding, each product video data is imported into the convolutional neural network, and a large amount of product data is obtained on the vacuum questionnaire (Jain et al., 2022). Then, the order of the spatial convolution of each picture in the image is numbered to obtain the spatial information of the entire picture. (3) Method based on CNN+LSTM. This method takes advantage of the convolutional neural networks in image spatial feature extraction, and the advantages of long memory and long memory networks in dynamic time modeling (Tariq et al., 2019). In the field of behavioral cognition, Donahue first adopted the architecture of convolutional net + long-term memory system, which can be trained end-to-end. Skeleton information can be used for behavior recognition. The main difference from other methods is that the input information mode is different. This method has a certain threshold, and first requires obtaining joint points. Whether using depth maps or RGB maps for estimation or hardware generation, it can proceed to the next step of more accurate recognition.

The majority of the human body starts off standing when falling, which is a dynamic process that resembles a normal state of life. During the fall stage, there will be an acceleration followed quickly by a displacement, regardless of whether the fall is from a chair, bed, or standing posture. The body will begin to move in the direction of the floor. This stage differs from regular sitting, lying down,

etc. in that it moves at a speed that is comparable to a free fall. The head will finally be on the ground, usually motionless or supine. Eventually, there can be a phase of recuperation during which the person can stand up on their own or with assistance from others. The two positions of sitting and lying are static, and when elderly persons are feeling under the weather, they may continue to sit or lie down for extended periods of time. Since it's unclear at this point if the elderly person has a condition, the detection model classifies this as second level alert data.

3.3 Difficulties of Human Action Recognition Methods Based on Deep Learning

At present, human action recognition methods based on deep learning still face the following difficulties (Wang et al., 2020). Firstly, based on CNN temporal dynamic codes, although current technologies can provide the temporal dynamic characteristics required for connecting layer CNN encoding, the CNN at each level does not include spatial information or changes in the surrounding environment, and cannot effectively represent the visual characteristics of moving objects (Zin et al., 2021). Therefore, some technologists have proposed end-to-end motion attention maps to assist in determining the range of object motion in the image (Liciotti et al., 2020). Although these technologies can overcome the influence of various complex conditions, the attention graph for end-to-end training is usually not fixed and requires rich training information. In addition, videos often contain a large number of redundant images, greatly reducing the accuracy of action recognition and increasing the workload of network design.

Secondly, the use of CNN+ dynamic time encoding has achieved great success in the past. The sequence pooling method and the dynamic graph method for input video achieved good results. However, since these methods do not consider the direction of motion of the lens, they can obtain relatively high-quality dynamic images for videos performed with fixed lenses. However, once the camera moves in the video, the quality of the dynamic image will be greatly affected. Researchers also pointed out that for 3D convolutional networks, the longer the input video, the higher the accuracy. However, longer input videos also increase the computational complexity of a network composed of thirty convolutions.

4. APPLICATION OF BEHAVIOR RECOGNITION TECHNOLOGY IN ELDERLY CARE

4.1 The Main Work of Behavioral Recognition for the Elderly

Recently, the main work in human behavior recognition is to recognize human behavior through video. Videos contain an added time dimension; therefore, in videos, it is important to realize that the difficulty of human behavior lies in how to handle the clock information in the video. Human dynamic cognition based on reinforcement learning can be divided into three main concepts: through 3D methods, through 2D networks, and through multidimensional methods adopted by neural networks. At the same time, a 3D version based on the combination of traditional printing, 3D, and longer printing processes has also been included in this process. This method is based on a dual flow network, including temporal and spatial characteristics. The method of using recurrent neural networks is to process information about time. Before processing temporal information, a pre trained convolutional neural network is usually used to process spatial information, and the time series function of the feature sequence is set first. Then, this feature sequence is fed into a recursive neural network to complete behavioral analysis. The analysis method for convolutional 3D videos is the most direct way to input video stream images. These image videos contain video information with a short duration, and the information processing and output matrix are the spatiotemporal characteristics of the videos. Finally, the classification layer recognizes behavior.

4.2 Application of Behavior Recognition Technology Based on Deep Learning in Elderly Care

In nursing homes, behavior recognition systems should function when the elderly are left unattended, identifying monitoring areas where the elderly have abnormal behavior or unexpected events, and triggering rapid and accurate alerts after analysis. (1) In actual monitoring scenarios, elderly people can receive treatment in a timely manner, and the system response is relatively fast. Once abnormal behavior appears in the monitoring image, an alarm message can be issued in real time. (2) If the object to be recognized is an uncut video, the recognition error using these algorithms will be high. This is because in uncut videos, there will be video content unrelated to actions. These contents often result in a large number of training sets, which undoubtedly affects the formation of the network, making the judgment of abnormal network behavior different. Of course, this cannot be applied in real scenarios. (3) The monitoring scene may have more complex backgrounds, and the camera itself may have various hardware issues. In complex scenarios that are not affected by the background, it is more necessary to improve the robustness of the algorithm. (4) An important feature of videos is that consecutive images are stacked in chronological order. Behaviors can be divided into static behaviors, including sitting, lying, standing, etc. Dynamic behavior, including falling, standing up, walking, etc. The recognition of dynamic behavior requires context correlation in order to make judgments in the future. (5) Compared to images, the number of parameters in videos increases exponentially, resulting in a significant increase in training time and hardware requirements for the dataset. The entire system recognition system includes input video, preprocessing module, object and behavior control module, and other identifiable behavior points. The input video will always be captured by a fixed camera and then run to the host through a series of consecutive RGB sequences, such as through a local area network. Mainly through the improvement of on-site samples and photos, it was detected to improve the accuracy and effectiveness of behavior detection and reduce noise interference, in order to identify benchmarks and marginal frameworks, allowing people to enter the target area of interest. The module will calculate the 2D coordinates of the critical point of the body. The authors jointly conducted disruptive self-learning on behavior detection software and then conducted classical learning through classical modes.

5 CONCLUSION

This article focuses on the research of human behavior recognition technology, integrating intelligent technology into surveillance cameras to provide safety assurance for the elderly by identifying abnormal behaviors. This article analyzes the causes of common abnormal behaviors in elderly people, and then establishes an abnormal behavior model to comprehensively consider the risk. This article mainly identifies falls, sitting still, and lying still. A method based on human appearance features extracts human joint points from two-dimensional images, and then classifies them using a classifier. The relative position and velocity features of the main joint points are used as thresholds for judgment. Extracting spatiotemporal features through deep learning and using a mixture of 2D and 3D convolutions, combined with contextual information for behavior recognition, has significantly improved compared to manual feature extraction methods. Finally, the method based on deep learning for feature extraction has achieved satisfactory verification results in terms of accuracy and speed, which can meet the requirements of practical work.

DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was supported by the Hebei Provincial Department of Education Scientific Research Project of colleges and universities in Hebei Province, Research on recognition methods of elderly falls in indoor scenes, ZC2023160; and the Hebei North University Natural Science Research Program Project, Research and Application of Generalized Additive Model in the Relationship between Air Pollution and Respiratory Diseases, QN2020013; and the Hebei North University Natural Science Research Program Project, Research on the relationship between air pollution and Respiratory disease based on generalized additive model, JYT2023007.

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

al Zamil, M. G., Samarah, S., Rawashdeh, M., Karime, A., & Hossain, M. S. (2019). Multimedia-oriented action recognition in Smart City-based IoT using multilayer perceptron. *Multimedia Tools and Applications*, *78*(21), 30315–30329. doi:10.1007/s11042-018-6919-z

Aslan, M. F., Durdu, A., & Sabanci, K. (2020). Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization. *Neural Computing & Applications*, *32*(12), 8585–8597. doi:10.1007/s00521-019-04365-9

Bolhasani, H., Mohseni, M., & Rahmani, A. M. (2021). Deep learning applications for IoT in health care: A systematic review. *Informatics in Medicine Unlocked*, *23*, 100550. doi:10.1016/j.imu.2021.100550

Byeon, Y. H., Kim, D., Lee, J., & Kwak, K. C. (2021). Body and hand–object ROI-based behavior recognition using deep learning. *Sensors (Basel)*, *21*(5), 1838. doi:10.3390/s21051838 PMID:33800776

Chang, C. W., Chang, C. Y., & Lin, Y. Y. (2022). A hybrid CNN and LSTM-based deep learning model for abnormal behavior detection. *Multimedia Tools and Applications*, *81*(9), 11825–11843. doi:10.1007/s11042-021-11887-9

Cheng, L., Gou, X., Tang, C., Li, Y., Jiang, X., & Zuo, H. (2023, April). Indoor person behavior recognition based on deep learning and knowledge graph. In *Second International Conference on Digital Society and Intelligent Systems (DSInS 2022)* (Vol. 12599, pp. 78-87). SPIE. doi:10.1117/12.2673563

Deep, S., Zheng, X., Karmakar, C., Yu, D., Hamey, L. G., & Jin, J. (2019). A survey on anomalous behavior detection for elderly care using dense-sensing networks. *IEEE Communications Surveys and Tutorials*, *22*(1), 352–370. doi:10.1109/COMST.2019.2948204

Hassan, M. M., Uddin, M. Z., Mohamed, A., & Almogren, A. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, *81*, 307–313. doi:10.1016/j.future.2017.11.029

Hsueh, Y. L., Lie, W. N., & Guo, G. Y. (2020). Human behavior recognition from multiview videos. *Information Sciences*, *517*, 275–296. doi:10.1016/j.ins.2020.01.002

Jain, D. K., Liu, X., Neelakandan, S., & Prakash, M. (2022). Modeling of human action recognition using hyperparameter tuned deep learning model. *Journal of Electronic Imaging*, *32*(1), 011211.

Karthickkumar, S., & Kumar, K. (2020, January). A survey on Deep learning techniques for human action recognition. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE. doi:10.1109/ICCCI48352.2020.9104135

Lentzas, A., & Vrakas, D. (2022). Machine learning approaches for non-intrusive home absence detection based on appliance electrical use. *Expert Systems with Applications*, *210*, 118454. doi:10.1016/j.eswa.2022.118454

Liciotti, D., Bernardini, M., Romeo, L., & Frontoni, E. (2020). A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing*, *396*, 501–513. doi:10.1016/j.neucom.2018.10.104

Ozcan, T., & Basturk, A. (2020). Human action recognition with deep learning and structural optimization using a hybrid heuristic algorithm. *Cluster Computing*, *23*(4), 2847–2860. doi:10.1007/s10586-020-03050-0

Pandey, P., & Litoriya, R. (2019). Elderly care through unusual behavior detection: A disaster management approach using IoT and intelligence. *IBM Journal of Research and Development*, *64*(1/2), 15–1.

Park, Y. J., Ro, H., Lee, N. K., & Han, T. D. (2019). Deep-care: Projection-based home care augmented reality system with deep learning for elderly. *Applied Sciences (Basel, Switzerland)*, *9*(18), 3897. doi:10.3390/app9183897

Phyo, C. N., Zin, T. T., & Tin, P. (2019). Deep learning for recognizing human activities using motions of skeletal joints. *IEEE Transactions on Consumer Electronics*, *65*(2), 243–252. doi:10.1109/TCE.2019.2908986

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, *80*, 241–265. doi:10.1016/j.inffus.2021.11.006

Soufian, M., Nefti-Mezian, S., & Drake, J. (2022). Toward Kinecting cognition by behaviour recognition-based deep learning and big data. *Universal Access in the Information Society*, *21*(1), 33–51. doi:10.1007/s10209-020-00744-5

Tariq, Z., Shah, S. K., & Lee, Y. (2019, December). Speech emotion detection using iot based deep learning for health care. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 4191-4196). IEEE. doi:10.1109/BigData47090.2019.9005638

Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., An, Y., Wang, K., & Li, S. (2020). Wearable sensor-based human activity recognition using hybrid deep learning techniques. *Security and Communication Networks, 2020*, 1-12.

Yu, J., de Antonio, A., & Villalba-Mora, E. (2022). Deep learning (CNN, RNN) applications for smart homes: A systematic review. *Computers*, *11*(2), 26. doi:10.3390/computers11020026

Zhu, J., Goyal, S. B., Verma, C., Raboaca, M. S., & Mihaltan, T. C. (2022). Machine learning human behavior detection mechanism based on python Python architecture. *Mathematics*, *10*(17), 3159. doi:10.3390/math10173159

Zin, T. T., Htet, Y., Akagi, Y., Tamura, H., Kondo, K., Araki, S., & Chosa, E. (2021). Real-time action recognition system for elderly people using stereo depth camera. *Sensors (Basel)*, *21*(17), 5895. doi:10.3390/s21175895 PMID:34502783

Shihui Zhang was born in Hebei, China, in 1990. She works at HeBei North University, Hebei. Her research interests include Image Processing and Artificial Intelligence.

Jing Mi was born in Hebei, China, in 1990. She works at Hebei North University. She has published six papers, two of which has been indexed by EI. Her research interests includes mobile communication and big data.

Naidi Liu is working as a lecturer at the School of Information Science and Engineering at Hebei North University, Zhangjiakou, China. She received the B.S. degree in Electronic and Communication Engineering from Hebei University of Technology. Her current research interests include electronic information science and technology.

Xiqing Zhao was born in 1970 in Hebei province. He is now the dean of the College of Information Science and Engineering at Hebei North University, majoring in computer application technology.