Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

Mobile Phone Behavior Recognition Based on Convolutional Neural Network

Xiangbo Zhang[1], Yuxuan Wang[2]

[1]China Agricultural University, Yantai 100083, Shandong Province
[2]Korea University South Korea 02841

Abstract: ***With the continuous enhancement of computational capabilities in smartphones and the increasing integration of various sensors into them, smartphones have transcended their conventional role in communication and gained significant application value in areas like human-computer interaction, fall detection, and information security. Utilizing smartphone sensor data for human activity recognition has also become a prominent research focus. Based on mobile phone behavior recognition data from Baidu Feizi, this paper constructs convolutional neural network and uses batch normalization method to accelerate the convergence speed, and the final accuracy rate reaches 96.7. The algorithms proposed in this paper are effective in recognizing six distinct behaviors. The study concludes with an analysis of various models, discussing their strengths and weaknesses in the context of behavior recognition applications.***

Keywords: Behavior Recognition; MLP; CNN; Batch Normalization.

1. INTRODUCTION

1.1 Background and significance

In recent years, the popularity of smartphones has made it more convenient for sensors to obtain data. Compared to computer vision methods, sensor based behavior recognition can better track human behavior, respond faster, have a wider range of applications, and have higher practical value. At the same time, the portability of smartphones also avoids unnecessary trouble for subjects, as they can obtain data from the daily lives of the general public when obtaining access to mobile data permissions. Therefore, researchers are increasingly concerned about the various sensor data collected by smartphones and their application value.

Based on mobile phone sensors for human behavior recognition, the purpose is to analyze and recognize human action types, behavior patterns, etc. through sensor observation data, and use natural language to describe or enable smart devices to complete corresponding behaviors, in order to better serve people's daily lives. At present, sensor based behavior recognition has important applications in healthcare, intelligent environments, and other fields. This article is based on the mobile behavior recognition dataset provided by Baidu Feijiang, and uses deep learning methods for behavior recognition to achieve good accuracy.

1.2 Literature Review

Xi et al. [1] enhanced problem-solving abilities by integrating reinforcement learning with large language models (LLMs), while Lyu et al. [2] optimized convolutional neural networks (CNNs) for efficient 3D point cloud object recognition. In healthcare, Pang et al. [3] leveraged electronic health records (EHRs) for diabetes risk prognosis, and Wu et al. [4] explored how supply chain digitalization contributes to carbon neutrality in energy sectors. Computer vision research has seen significant progress, exemplified by Peng et al. [5]'s dual-augmentor framework for domain generalization in 3D human pose estimation. Meanwhile, Liu et al. [6] improved mathematical reasoning in LLMs via hallucination detection, and Liu et al. [7] introduced Tool-Planner, a multi-tool task planning system. In biomedical research, Wang et al. [8] mapped the immune microenvironment in gastrointestinal cancers, while Wang et al. [9] and Li et al. [10] applied predictive modeling to e-commerce logistics and drug trial monitoring, respectively. Multimodal AI applications are expanding rapidly. Yuan et al. [11] utilized GPT-4 for EHR processing, and Song et al. [12] enhanced e-commerce content generation via AI-human collaboration. Data quality and scalability were addressed by Chen et al. [13, 14], who proposed frameworks for gig economy platforms. Legal and risk management innovations include Wang et al. [15]'s study on enterprise naming rights and Gong et al. [16]'s ensemble ML-based risk decision support system. In industrial applications, Bohang et al. [17] improved image steganalysis via active learning, while Zhao et al. [18] and Yao et al. [19] optimized steel production scheduling and drone-3D printing for post-disaster shelters, respectively. Financial analytics were advanced by Yang et al. [20]'s CNN-based stock sentiment analysis and Ji et al. [21]'s AI-driven retail strategies. Cross-domain risk management was tackled by Yang et al. [22] using LLMs for real-time market monitoring, and Li et al. [23] refined clinical trial strategies via Bayesian optimization. Healthcare innovations continued with Peng et al. [24]'s IoT-enhanced cognitive training study and Ming et al. [25]'s telemedicine feasibility analysis for pediatric care. Legal automation was explored by Wang et al. [26] using explainable LLMs for compliance audits. Sustainable AI applications were highlighted by He et al. [27] in architecture and Ge et al. [28] in urban planning, demonstrating AI's role in green technology.

1.2.1 Overview of Research on Mobile Behavior Recognition

With the development of network communication technology and microelectronics technology, smartphones integrate various sensors such as acceleration sensors and gyroscopes, which have strong perception capabilities for the surrounding environment. At present, the commonly used method for behavior recognition using sensors is to collect data from acceleration sensors and gyroscopes, and establish a behavior recognition model from them. Yang Chenchen and others combined acceleration and gyroscope sensors to real-time read the acceleration and angular velocity information of human motion, in order to identify falling behavior.

1.2.2 Research Status of Deep Learning in Mobile Behavior Recognition

The difference between deep learning methods and traditional machine learning methods is that they change the extraction and feature selection steps of traditional methods. During classification model training, features can be automatically generated instead of manually designed features, which can more accurately represent real human behavior features for complex scenes and data types.

Shi Daiwei et al. used CNN to automatically extract features and combined it with Softmax classifier for classification and recognition, achieving high recognition rates on six human behavior patterns. Jiang Xiangyu designed three schemes that achieved higher accuracy by changing the convolution process of the convolutional layers. Yu Tao proposed an adaptive sampling rate algorithm based on reinforcement learning, which utilizes RNN (Recurrent Neural Network) and LSTM (Long Short Time Memory) networks to extract time-dependent features from samples. This algorithm ensures model recognition accuracy while minimizing energy consumption.

Jinbo et al. introduced two types of attention mechanisms based on the Time Convolutional Neural Network (TCN), achieving an accuracy of up to 98.25% on the public human behavior dataset PAMAP2, which is a significant improvement compared to the original TCN. Li Dongdong added attention mechanism on the basis of RNN and LSTM models, and the accuracy was significantly improved. Huang Junqiang effectively improved the accuracy by using a transformer based on multi-sensor information fusion method.

2. OVERVIEW OF RELATED TECHNOLOGIES

2.1 Sensor based human behavior recognition framework

The sensor based human behavior recognition technology uses slightly different sensors, algorithms, and target behaviors in various studies, but its overall research framework is basically similar.

Firstly, data collection. Then, the collected sequence data is corrected, filtered, resampled, and other steps are taken to remove noise and errors.

Based on machine learning technology, it is necessary to first extract features from the original sequence data, such as time domain, frequency domain, time-frequency domain, etc. If a large feature dimension is extracted, feature dimensionality reduction is also required, such as using LDA and PCA to extract and select variables that have a significant impact on the target. Finally, machine learning methods are used to model and optimize parameters to complete classification tasks.

Behavior recognition based on deep learning technology can usually automatically extract features through shallow networks and directly construct models.

Finally, during the model training phase, the preprocessed training samples are fed into the model for learning, and tested on test samples for model comparison and optimization, selecting the model that meets the requirements and has excellent performance.

2.2 Overview of Related Technologies

2.2.1 Convolutional Neural Networks (2D-CCN)

Convolutional neural networks (CNNs) mimic human vision and can learn from multidimensional data such as images. The three characteristics of local perception, weight sharing, and pooling enable CNN to focus on spatial locality information. Local perception is implemented using convolutional kernels, which move at a certain step size to extract local information to the next layer; Weight sharing uses a convolutional kernel to extract features from a two-dimensional tensor, reducing the number of convolutional kernels and the parameters that need to be learned; Pooling improves the feature extraction capability of the network while reducing the computational scale, while preserving local features and expanding the receptive field of the underlying network. CNN is mainly composed of input layer, convolutional layer, pooling layer, fully connected layer, and output layer. Figure 2-8 is a CNN structure diagram. For the classification task in this article, simply add the soft Max function at the last layer.

2.2.2 Evaluation criteria for the model in this article

(1) Accuracy:

The proportion of correctly classified samples to the total sample size, with the specific evaluation metric being the percentage of correctly predicted samples on the test set to the total sample size, ranging from 0 to 1.

(2) Loss function:
The loss function is an indicator for evaluating the difference between the predicted results of a model and the actual results. As it is a multi classification problem, the loss function used in this paper is cross entropy. For each sample, calculate the cross entropy between the probability distribution of the true label and the probability distribution of the model predicted label, and then average the cross entropy of all samples. The smaller the cross entropy loss function, the smaller the difference between the predicted and actual results of the model, and the better the classification performance of the model.

(3) F1 Score:

F1 score is an indicator used to evaluate the performance of classification models, which is a weighted harmonic mean of accuracy and recall. It combines the advantages of precision and recall, and can more comprehensively evaluate the classification accuracy of the model. The calculation formula is:

$$F1 = 2 \times (Accuracy\ rate)/Accuracy\ rate + recall$$

Among them, accuracy refers to the proportion of samples predicted as positive cases and actually positive cases to the proportion of samples predicted as positive cases, and recall refers to the proportion of samples predicted as positive cases to the proportion of samples actually positive cases. The range of F1 values is 0-1, with higher values indicating better model performance.

3. FEATURE EXTRACTION AND PREPROCESSING OF DATASETS

The original dataset for this experiment is the UCI-HAR dataset, which was collected by 30 volunteers aged 19-48 wearing smartphones (Samsung Galaxy S II) on their waist for six activities (walking, walking upstairs, walking downstairs, sitting, standing, lying down). The experiment captures data on 3-axis linear acceleration and 3-axis gyroscope angular velocity at a constant rate of 50Hz. A sequence data with six dimensions in total.

Based on the original dataset, the six dimensional variables of three-axis acceleration and angular velocity are treated as a sequence, and feature extraction is performed in both time and frequency domains for each dimensional variable based on the collected data over a period of time. After feature extraction, 561 features and 1 classification label were generated and retained for each sample. Retain 8000 training set samples and 2000 testing set samples with equal probability. Formed the dataset used in this experiment.

Remove outliers from each label in the training set through a box plot, save the results as tensor, float32 format data, and standardize the feature data. Draw a box plot for observation:

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

From the above figure, it can be clearly seen that there are outliers in various movements, but the two behaviors where outliers frequently appear are lying and sitting, which may be caused by the difference in angular velocity between the accelerometer and the stationary state. For example, when a person is sitting, they may turn sideways from side to side. So the existence of these outliers may be a knowledge that the model needs to learn. To verify the correctness of the above idea, the following will learn on two datasets with and without outlier removal, and ultimately compare their accuracy on the test set. Finally, divide the 8000 training set samples into a training set and a validation set in a 4:1 ratio.

4. MODEL CONSTRUCTION

4.1 Model Design

Further analysis of the features reveals that there are 66 accelerometer features, 40 angular velocity features, and 1 comprehensive feature for the x, y, and z axes, respectively. So, the acceleration and angular velocity features based on the x, y, and z axes can be divided into three channels, with data format for each channel (14, 14). Among them, zero padding is used for each channel data to fill 187 features into 196 features, making each channel a "square image" data format. Based on a two-dimensional convolutional neural network, three models were constructed with three convolutional layers, two fully connected layers, and batch normalization layers for differentiation. In the construction of this model, based on the experience of one-dimensional convolutional neural network models, all models used average pooling layers. The specific reasons will be analyzed in the conclusion. As shown in Table 1:

Table 1: Structure diagram of two-dimensional convolutional neural network model

Model Number	Model Structure	Composition	Model Training
Model 6	Convolutional layer 1 Convolutional layer 2 Convolutional layer 3 Convolutional layer 4 Fully connected layer 1 Fully connected layer 2	Conv2D1+ReLU+AvgPooling Conv2D2+ReLU+AvgPooling Conv2D3+ReLU+AvgPooling Conv2D4+ReLU+AvgPooling Linear1+Relu Linear2	Step size: 0.001 Training epochs: 100 Loss function: Cross entropy Optimization algorithm: Adam
Model 7	Convolutional layer 1 Convolutional layer 2 Convolutional layer 3 Convolutional layer 4 Fully connected layer 1 Fully connected layer 2	Conv2D1+ReLU+BN+AvgPooling Conv2D2+ReLU+BN+AvgPooling Conv2D3+ReLU+BN+AvgPooling Conv2D4+ReLU+BN+AvgPooling Linear1+Relu Linear2	Step size: 0.01 Training epochs: 100 Loss function: Cross entropy Optimization algorithm: Adam (Add batch normalization layer, you can choose a larger step size)
Model 8	Convolutional layer 1 Convolutional layer 2 Fully connected layer 1 Fully connected layer 2	Conv2D1+ReLU+MaxPooling Conv2D2+ReLU+MaxPooling Linear1+Relu Linear2	Step size: 0.001 Training epochs: 100 Loss function: Cross entropy Optimization algorithm: Adam

4.2 Overview of Experimental Results

After feeding the corresponding data into 10 models for training, the performance comparison criteria are the accuracy and F1 value on the validation set and the accuracy on the test set. The model number and its characteristics, namely the scores of the three indicators, are shown in Table 2, with units of%:

Table 2: Summary of Experimental Results

Model category	Model Number	Model features	Verification set accuracy (Unit:%)	F1 Value (Unit:%)	Test set accuracy (Unit:%)
2D Convolutional Neural Network	Model 6	Deeper network+average pooling layer	95.15	95.1	95
	Model 7	Deeper network+average pooling layer+batch normalization layer	96.9	96.9	96.7
	Model 8	Shallow network+max pooling layer	93.2	93.2	92.55

For convolutional neural networks, adding batch normalization layers to deeper network structures can preserve features near the input end, avoid gradient vanishing caused by network deepening, and enhance model computational performance. Its accuracy is also the best among all models. Similarly, average pooling is superior to maximum pooling and increasing model depth can achieve higher accuracy.

In addition, regarding the question of whether outliers in sitting and lying movements mentioned in Chapter 3 need to be addressed, the conclusion is that they do contain unique information of stationary movements, and without deletion, they score slightly higher on the test set. The experimental results for Model 4 and Model 7 are shown in Table 3 below:

Table 3: Comparison of Handling Outliers or Not

Model category	Model Number	Model features	Accuracy of removing outliers (Unit:%)	Accuracy without removing outliers (Unit:%)
2D Convolutional Neural Network	Model 7	Deeper network+average pooling layer+batch normalization layer	95.85	96.7

5. CONCLUSION

As described in Chapters 3, 4, and 5, after performing well in feature extraction, convolutional neural networks perform the best due to their ability to extract dependency information between multiple sensors. The multi-layer perceptron model is difficult to pay attention to the interdependence between different sensor data, and its computing performance is poor based on the whole. In the data after feature extraction, the cyclic neural network may perform poorly because the model can only learn sequence data in multiple sensor variable inputs, ignoring the differences between different sensors. The cyclic neural network and LSTM model may be more suitable for learning the original sequence.

REFERENCES

[1] Xi, Kai, et al. "Enhancing Problem-Solving Abilities with Reinforcement Learning-Augmented Large Language Models." 2024 4th International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). IEEE, 2024.
[2] Lyu, T., Gu, D., Chen, P., Jiang, Y., Zhang, Z., & Pang, H. & Dong, Y.(2024). Optimized CNNs for Rapid 3D Point Cloud Object Recognition. arXiv preprint arXiv:2412.02855.
[3] Pang, H., Zhou, L., Dong, Y., Chen, P., Gu, D., Lyu, T., & Zhang, H. (2024). Electronic Health Records-Based Data-Driven Diabetes Knowledge Unveiling and Risk Prognosis. arXiv preprint arXiv:2412.03961.
[4] Wu, W., Bi, S., Zhan, Y., & Gu, X. (2025). Supply chain digitalization and energy efficiency (gas and oil): How do they contribute to achieving carbon neutrality targets?. Energy Economics, 142, 108140.
[5] Peng, Qucheng, Ce Zheng, and Chen Chen. "A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2240-2249. 2024.
[6] Liu, M., Bo, S., & Fang, J. (2025). Enhancing Mathematical Reasoning in Large Language Models with Self-Consistency-Based Hallucination Detection. arXiv preprint arXiv:2504.09440.
[7] Liu, Yanming, et al. "Tool-Planner: Task Planning with Clusters across Multiple Tools." arXiv preprint arXiv:2406.03807 (2024).
[8] Wang, Y., Yang, T., Liang, H., & Deng, M. (2022). Cell atlas of the immune microenvironment in gastrointestinal cancers: Dendritic cells and beyond. Frontiers in Immunology, 13, 1007823.
[9] Wang, J. (2025). Predictive Modeling for Sortation and Delivery Optimization in E-Commerce Logistics.
[10] Li, T. (2025). Enhancing Adverse Event Monitoring and Management in Phase IV Chronic Disease Drug Trials: Applications of Machine Learning.
[11] Yuan, J. (2024). Exploiting gpt-4 for multimodal medical data processing in electronic health record systems. Preprints, December.
[12] Song, X. (2024). Leveraging aigc and human-computer interaction design to enhance efficiency and quality in e-commerce content generation.
[13] Chen, J. (2025). Data Quality Quantized Framework: Ensuring Large-Scale Data Integration in Gig Economy Platforms.
[14] Chen, J. (2025). Efficient and Scalable Data Pipelines: The Core of Data Processing in Gig Economy Platforms.
[15] Wang, H. (2024). The Restriction and Balance of Prior Rights on the Right of Enterprise Name.

[16] Gong, C., Lin, Y., Cao, J., & Wang, J. (2024, October). Research on Enterprise Risk Decision Support System Optimization based on Ensemble Machine Learning. In Proceeding of the 2024 5th International Conference on Computer Science and Management Technology (pp. 1003-1007).

[17] Bohang, L., Li, N., Yang, J. et al. Image steganalysis using active learning and hyperparameter optimization. Sci Rep 15, 7340 (2025). https://doi.org/10.1038/s41598-025-92082-w

[18] Zhao, H., Chen, Y., Dang, B., & Jian, X. (2024). Research on Steel Production Scheduling Optimization Based on Deep Learning.

[19] Yao, T., Jian, X., He, J., & Meng, Q. (2025). Drone-3D Printing Linkage for Rapid Construction of Sustainable Post-Disaster Temporary Shelters.

[20] Yang, W., Lin, Y., Xue, H., & Wang, J. (2025). Research on Stock Market Sentiment Analysis and Prediction Method Based on Convolutional Neural Network.

[21] Ji, F., Zheng, X., Xue, H., & Wang, J. (2025). A Study on the Application of Artificial Intelligence in Personalized Go-to-Market Strategy in Retail Industry.

[22] Yang, J., Tang, Y., Li, Y., Zhang, L., & Zhang, H. (2025). Cross-Asset Risk Management: Integrating LLMs for Real-Time Monitoring of Equity, Fixed Income, and Currency Markets. arXiv preprint arXiv:2504.04292.

[23] Li, T. (2025). Optimization of Clinical Trial Strategies for Anti-HER2 Drugs Based on Bayesian Optimization and Deep Learning.

[24] Peng, Y., Zhang, G., & Pang, H. (2025). Exploring the effects of IoT-enhanced exercise and cognitive training on executive function in middle-aged adults. Alexandria Engineering Journal, 120, 106-115.

[25] Yang, Jie, et al. "Dynamic Hedging Strategies in Derivatives Markets with LLM-Driven Sentiment and News Analytics." arXiv preprint arXiv:2504.04295 (2025).

[26] Ming DY, Li T, Ross MH, et al. Feasibility of post-hospitalization telemedicine video visits for children with medical complexity. J Pediatr Health Care. 2022;36(2):e22–e35

[27] Wang, J., Yuan, J., Liu, J., & Evans, L. (2025). Simple Legal Compliance: Automating Regulatory Audits with Explainable LLMs.

[28] He, J., Xu, H., Li, X., & Meng, Q. (2024). Research on Innovative Applications of AI in Sustainable Architecture: Blueprint for Future Building Technology.

[29] Ge, Minyue, Zhang Feng, and Qian Meng. "Urban planning and green building technologies based on artificial intelligence: Principles, applications, and global case study analysis." Applied Science and Engineering Journal for Advanced Research 3.5 (2024): 18-27.