Network Implementation for Real-Time Exercise Behavior Recognition

Junho Kwon
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
kwonkai@keti.re.kr

Myeongseop Kim
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
myeongseopkim@keti.re.kr

Seho Park
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
sehopark@keti.re.kr

Kyung-Taek Lee
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
ktechlee@keti.re.kr

***Abstract*— With the growing societal emphasis on exercise and health care, there has been a surge in research integrating machine learning and deep learning into exercise-related studies. A critical aspect of exercise assistance is the ability to accurately recognize exercise behaviors and count repetitions. In this study, we developed and evaluated two models, the DNN (Deep Neural Network) and the LSTM (Long Short-Term Memory), to measure users' motor behaviors and their frequency. These models were trained using diverse datasets, including KETI, NTU, UCF, AI-HUB, and YouTube. Our findings revealed that the LSTM model achieved a commendable recognition accuracy of 97.35%, while the DNN model followed closely with 94.29%. Although the LSTM model excelled in recognizing exercise movements, it faced challenges in counting repetitions. Conversely, the DNN model efficiently counted exercise repetitions, showcasing its potential for real-time applications.**

Keywords—Deep Neural Network, Action Recognition, 3D Pose Estimation, Pose Classification

I. Introduction

With the advancement of machine learning and deep learning technologies, there is a continuous increase in interest in human action recognition research. Based on the rising popularity of activities like home training, personalized fitness programs, and CrossFit, there is active research on various exercise recognition AI models. Exercise behavior analysis models are primarily divided into RGB image classification methods and skeleton-based classification methods. Image-based classification focuses more on the context awareness of the image rather than recognizing specific exercise postures. Therefore, for exercise action recognition, a skeleton-based AI model is more suitable[1]. In this study, we introduce a model that conducts exercise action recognition using skeletal data and accurately quantifies the number of exercise repetitions.

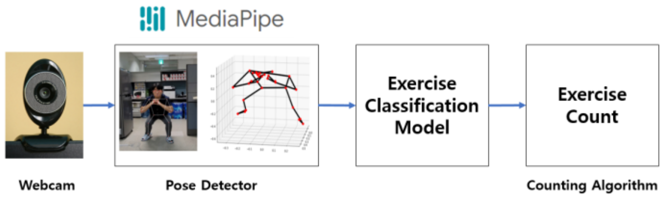

Fig. 1. Real-Time Exercise Recognition & Count Pipeline

Fig 1 illustrates the process of exercise behavior recognition and exercise repetition counting. We extract 33 3D skeletal coordinate data in real-time from webcam video using MediaPipe Pose Estimation function, based on BlazePose [1]. These extracted data points are then classified to identify the current exercise behavior. In our study of exercise action recognition, we evaluated the effectiveness of deep neural network (DNN) and long-term memory (LSTM) models in recognizing seven different exercise postures.

The research revealed that the LSTM model excels in recognizing movement postures. However, accurately counting movement repetitions presents challenges and requires additional code. In contrast, the DNN model not only adeptly identifies specific exercise movements like 'PUSHUP_DOWN' and 'PUSHUP_UP' but also provides immediate and accurate repetition counts through motion recognition.

II. REAL-TIME EXERCISE BEHAVIOR RECOGNITION METHOD

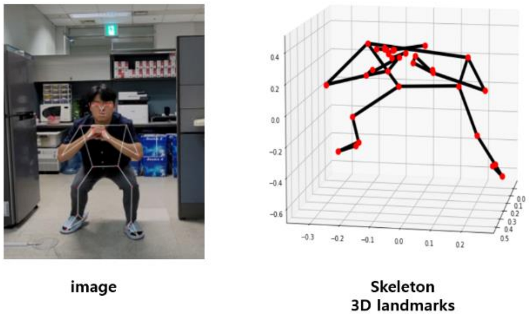

Fig. 2. 3D Pose Landmarks Data Collection

Accurate exercise behavior recognition necessitates pose estimation technology. As depicted in Fig 2, we extracted 3D coordinate values for 33 key points across the human body using Google's MediaPipe framework. These 33 3D landmarks offer a comprehensive representation of the human physique, encompassing facial and full-body movements. The coordinate data of these landmarks was employed as training data for the classification model.

Using the 3D landmarks, we trained a classification model utilizing both a Deep Neural Network (DNN) model [3] and a Long Short-Term Memory (LSTM) model [4]. The LSTM

This work was supported by the IT R&D program of MOTIE/KEIT (20009543, Development of Smart Home Manager for providing Edge AI-based Emotional Services).

model excels in recognizing exercise postures based on the sequences of landmarks during the learned sequence and accurately identifies the learned exercise postures. It can determine actions like 'PUSHUP,' 'LUNGE,' and 'JUMPING JACK.'

However, when employing the LSTM model for exercise behavior recognition, specifically classifying actions such as 'PUSHUP_DOWN' and 'PUSHUP_UP' became challenging. Additionally, accurately calculating the number of exercise repetitions using the LSTM model proved difficult, necessitating additional code for repetition tracking. However, the DNN model performs classification for each frame of the webcam-captured video. It offers a practical and user-friendly solution by efficiently recognizing specific exercise movements like 'PUSHUP_DOWN' and 'PUSHUP_UP' while providing immediate and accurate repetition counts.

III. Experimental Resuslts

The motor behavior recognition experiment was conducted in a controlled environment, such as an office or conference room. The experiment was conducted using a computer with the following specifications: The computer featured an Intel(R) Core(TM) i7-10700K 16 Core CPU, an Nvidia GeForce RTX 2080 Ti GPU, and 64GB of RAM.

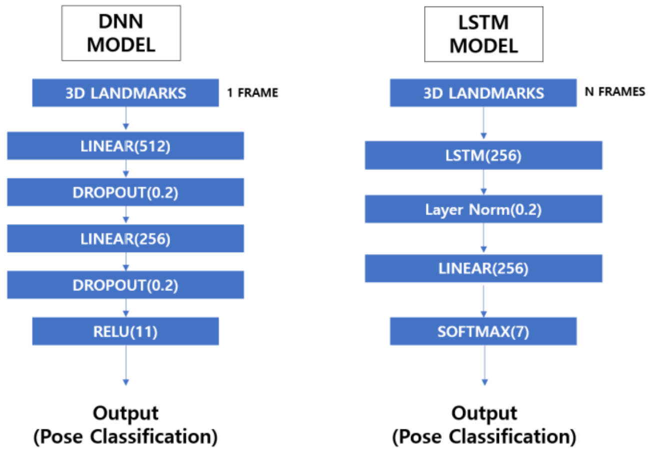

Fig. 3. Model Architecture

Fig 3 displays the DNN and LSTM model structures used for motion recognition. DNN models employ deep neural networks with linear and dropout layers to classify actions based on landmarks extracted from a single frame. In contrast, the LSTM model utilizes LANDMARKS data extracted from multiple frames for motion classification and employs a deep neural network with an LSTM layer, a normalization layer, and a linear layer.

The action recognition dataset used for learning was a complex dataset including KETI, NTU[6], UCF[7], AI-hub[8], Youtube, Waseda-Univ[9], and Kaggle/HMDB51[10]. The dataset categories even distinctive postures: 'stand' 'jumping jack,' 'push up,' 'lunge,' 'squat,' 'wall pushup,' and 'fall down.'

For the DNN model, our dataset encompasses a total of 23,593 training sheets and 2,757 test sheets, exclusively dedicated to the classification of these seven distinct poses. This meticulous partitioning ensures the provision of an extensive and representative dataset, enabling rigorous experimentation and robust model training. For the LSTM model, our dataset encompasses a substantial 29,511 training samples and 2,947 test samples. This judicious selection ensures the provision of a realistic and representative dataset, facilitating rigorous experimentation.

The accuracy of the DNN model achieved 94.29%, while the LSTM model achieved an accuracy of 97.35%. LSTM Model used an additional code that measures the number of movements based on the amount of movement change in the facial bounding box (B-BOX). However, if there is no movement in the B-BOX for a certain period of time, the counting function is not triggered, or a sequence delay of 1 to 2 frames occurs until the movement is detected and counted. In contrast, the DNN model excels in promptly counting the number of exercises as soon as it recognizes a relevant exercise change, providing the convenience of not requiring additional functions.

IV. Conclusion

Accurate recognition of exercise actions and precise calculation of repetitions are crucial in the field of home training and fitness. In this study, real-time motor action recognition technology highlighted the advantages of using deep neural network (DNN) models. While LSTM models exhibit excellent recognition performance, DNN simplifies the process by offering a solution for recognizing specific motor actions and concurrently tallying the number of repetitions. However, accurately tracking action repetitions within a sequence remains a significant challenge for LSTM models.

Future research should continue to explore exercise assistance services through DNN-based motion recognition and repetition counting. Additionally, investigating real-time applications and enhancing model efficiency are promising directions for ongoing research.

Acknowledgment

This work was supported by the IT R&D program of MOTIE/KEIT (20009543, Development of Smart Home Manager for providing Edge AI-based Emotional Services).

References

[1] Wang, L., Su, B., Liu, Q., Gao, R., Zhang, J., & Wang, G. (2023). Human Action Recognition Based on Skeleton Information and Multi-Feature Fusion. Electronics, 12(17), 3702.

[2] Zhou, X., Wang, C., & Koltun, V. (2019). "BlazePose: On-device Real-time Body Pose Tracking." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12587-12596.

[3] Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory in Recurrent Neural Networks." In Proceedings of the International Conference on Neural Information Processing (ICONIP), Vol. 2, pp. 173-178

[4] I. Chaudhary, N. Thoiba Singh, M. Chaudhary, and K. Yadav, "Real-Time Yoga Pose Detection Using OpenCV and MediaPipe," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-5.

[5] Lo, Y. H., Yang, C. C., Ho, H., & Sun, S. W. (2021, November). "richyoga: An interactive yoga recognition system based on rich skeletal

joints." In 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), pp. 256-257

[6] Shahroudy, A.; Liu, J.; Ng, T.-T.; and Wang, G. 2016. NTU RGB+D: A large scale dataset for 3D human activity analysis. In IEEE Conference on Computer Vision and Pattern Recognition

[7] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah, 2012 "UCF101: A dataset of 101 human actions classes from videos in the wild," arXiv preprint arXiv:1212.0402,

[8] Daehyung Lee, 2020, Slick Corporation, https://aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=realm&dataSetSn=231

[9] Ogata, R., Simo-Serra, E., Iizuka, S., & Ishikawa, H. (2019). Temporal distance matrices for squat classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

[10] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A large video database for human motion recognition. In ICCV, 2011. 1, 2, 6, 8