2024「中技社科技獎學金」

2024 CTCI Foundation Science and Technology Scholarship

境外生研究獎學金

Research Scholarship for International Graduate Students

Design of a Lightweight Framework for Human Action Detection on Edge IoT Devices

建置於邊緣物聯網設備之輕量級人體動作識別框架設計

1st year Master student: *Le Hoang Cong* Advisor: *Prof. Chen-Chien Hsu*

Dept. of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan.

1. Abstract

Human action recognition (HAR) is an evolving technology with the potential to revolutionize how we understand human behavior, which finds applications across various domains such as elderly care, surveillance systems, and human-robot interaction. As HAR continues to advance, there's a growing interest in integrating it into Internet of Things (IoT) systems. To minimize response time between clients and servers, researchers have explored embedding models into edge devices, though achieving optimal results remains a challenge. Balancing model size and performance is particularly problematic; while reducing parameters can limit model complexity, larger models often yield superior performance, posing challenges for implementation on memory-constrained edge devices. In this paper, we introduce a novel lightweight framework specifically designed to address these challenges. Through experimentation on a renowned benchmark dataset, our proposed approach demonstrates both superior performance and minimal model size.

2. Methodology

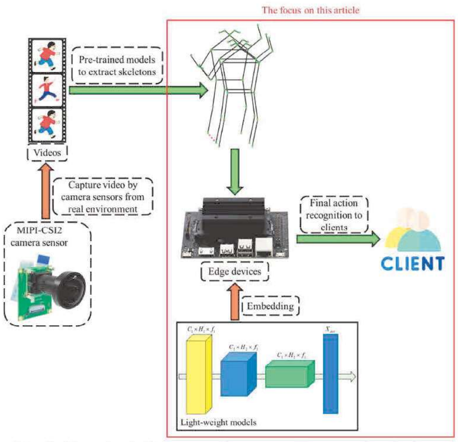

Fig. 1. Flowchart of human action recognition on edge devices.

Figure 1 illustrates the comprehensive operational process of skeleton-based online action recognition for edge devices within an IoT system. Video data captured by sensor cameras is processed through a pre-trained model (e.g., the MoveNet model) to extract human poses (skeleton coordinates). These coordinates are then sent to edge devices, where a lightweight model is stored and processes the data to generate final results. Finally, the results are sent to clients via messaging or email notifications.

2.1. Data augmentation techniques

To ensure the highest accuracy across all datasets, we implemented a data augmentation technique for 2D skeletons. This technique comprises two distinct methods, as depicted in **Figures 2b** and **2c**. First, we apply a random global jitter to shift the skeleton's overall position (**Figure 2c**), affecting the entire input representation. This method enables the network to learn that action recognition is not overly dependent on the global position of joint trajectories, preserving the intrinsic relationships among different joints. In the second method, the skeleton is divided into four groups, as shown in **Figure 2b**. Each group undergoes the same amount of random jitter, focusing on adjusting each group's location independently. While the first method shifts the entire skeleton, the second method modifies the location of each group independently.

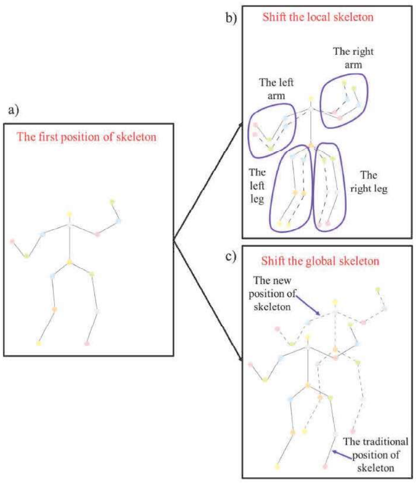

Fig. 2. a) Illustration of the skeleton shape from the JHMDB dataset. b) Depiction of the first data augmentation method for global position augmentation. c) Representation of the second data augmentation method for local position augmentation.

3. Results

TABLE I

Comparison of accuracy and parameters between our proposed method and SOTA methods utilizing only 2D skeleton Where (M) Denotes Million

Method	Years	Parameters	Accuracy
EHPI	2019	1.22M	65.5%
Transformer	2022	1.45M	74.7%
DD-Net	2019	1.82M	77.2%
Two-branch Transformer	2023	1.34M	77.9%
TD-Net	2022	1.88M	79.3%
Mattoo et. al.	2023	1.97M	81.3%
Ours*	**2024**	**0.25M**	**83.2%**

4. Conclusion

This paper introduces a novel lightweight framework specifically designed for deployment on tiny devices (i.e., Raspberry Pi 5, STM32MP157D-DK1.) within IoT systems, addressing memory limitations. We aim to extract comprehensive spatial-temporal information through skeleton while minimizing the impact of noise, light, and dynamic backgrounds typically present in RGB frames. Remarkably, our approach achieves maximum accuracy and minimal model size, particularly noteworthy when applied to challenging datasets.

5. Publications

(A) Journal publications

1. **Hoangcong Le**, Van Su Luong, and Minhhuy Le, "Development of magnetic image super-resolution model for nondestructive testing," *Measurement Science and Technology*, Vol. 34, no. 12, pp. 125907, 2023. (SCI)
2. **Hoangcong Le**, Le Quang Trung, Thuy Vu Phuong, Minhhuy Le, Jinyi Lee, and Naoya Kasa, "Optimising signal quality for corrosion detection using the innovative Mag_FSRCNN model to minimise hall sensor array in eddy current testing," *Nondestructive Testing and Evaluation*, pp. 1-24, 2024. (SCI)
3. **Hoangcong Le**, Saeed Saeedvand, and Chen-Chien Hsu, "A Comprehensive Review of Mobile Robot Navigation Using Deep Reinforcement Learning Algorithms in Crowded Environments," *Journal of Intelligent & Robotic Systems*, Vol. 110, no. 158, Nov. 9, 2024. (SCI)

(B) Conference publications

1. **Hoangcong Le**, Chen-Chien Hsu, Cheng-Kai Lu, Wei-Yen Wang, and Pin-Yen Monica Kuo, "Human Action Recognition on Edge Devices: A Novel Light-Weight Model," *2024 IEEE 13th Global Conference on Consumer Electronics* (GCCE), Kokura, Japan, Oct. 29-Nov.1, 2024.