Enhancing Gesture-Based Interactions for Individuals with Disabilities through Dual-Attention Wireless Sensing Networks

Jingzhi Pan

School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China

202283270337@nuist.edu.cn

Abstract. This paper investigates the application of wireless sensing technology to develop a more natural and intuitive gesture-based interaction framework for individuals with disabilities. The study begins by identifying the primary challenges faced by this demographic when interacting with conventional devices. It then introduces a novel approach for WiFi-enabled gesture recognition and interpretation using dual attention networks. This methodology comprises two essential components: Channel State Information (CSI) preprocessing and a gesture recognition module. The paper details the implementation of these modules and elucidates their roles in enhancing gesture detection accuracy. Furthermore, the discussion extends to the future prospects of wireless sensing technologies, envisioning their integration into smart home systems, public services, and enhanced social interactions. The research underscores the transformative potential of integrating gesture-based interactions with wireless sensing to significantly elevate the quality of life for people with disabilities, suggesting a paradigm shift in how assistive technologies are developed and utilized.

Keywords: Wireless Sensing; Gesture Interaction; Dual-attention CSI networks; ResNet.

1. Introduction

Research Background: Half a million years ago, humans utilized precise hand and voice synchronization to hunt large animals like mammoths, not only fulfilling their survival needs but also establishing the fundamental role of gestures in conveying information and emotions. Despite the adequacy of traditional human-computer interaction methods such as keyboards and mice in addressing the needs of modern society, these methods exhibit considerable limitations in naturalness, intuitiveness, and convenience. With the burgeoning advancements in technologies like smart homes, virtual reality, and augmented reality, the demand for more organic and intuitive forms of human-computer interaction has intensified [1]. In this evolving landscape, wireless sensing gesture interaction technology has emerged, offering vast application potential, especially in enhancing the inclusivity and well-being of individuals with disabilities [2].

Significance of the Study: The advent of gesture interaction technology heralds a more accessible and convenient mode of engagement for people with disabilities, who often find traditional interfaces challenging or impossible to use. This difficulty can marginalize them within the digital society. Gesture interaction technology employs the capture and analysis of user gestures, converting them into commands that machines can understand, thereby facilitating effective communication with electronic devices [3]. This method significantly diminishes the reliance on physical capabilities, enhances interaction efficiency, and markedly improves the quality of life for people with disabilities. Additionally, the non-contact nature of wireless sensing technology ensures the protection of user privacy and minimizes the personal traces left during device usage.

Contribution of This Paper: The current research landscape is intensely focused on optimizing algorithms to enhance the real-time performance and robustness of gesture recognition systems, which are crucial for adapting to diverse environments and various gesture typologies. The dual attention network, a novel deep learning model, has gained prominence for its efficacy in dynamic gesture recognition [4]. This network enhances performance by concentrating on different dimensions of information during the feature extraction and recognition phases, ensuring robust operation even in complex settings. This pioneering research not only advances the field of gesture recognition technology but also offers more sophisticated interaction solutions for people with disabilities [5]. Therefore, this paper's in-depth exploration into the application of wireless sensing gesture interaction technology and dual attention networks holds significant theoretical value and promises to enhance both the quality of life for people with disabilities and the security of technology applications.

2. Methods

2.1. System Model Framework

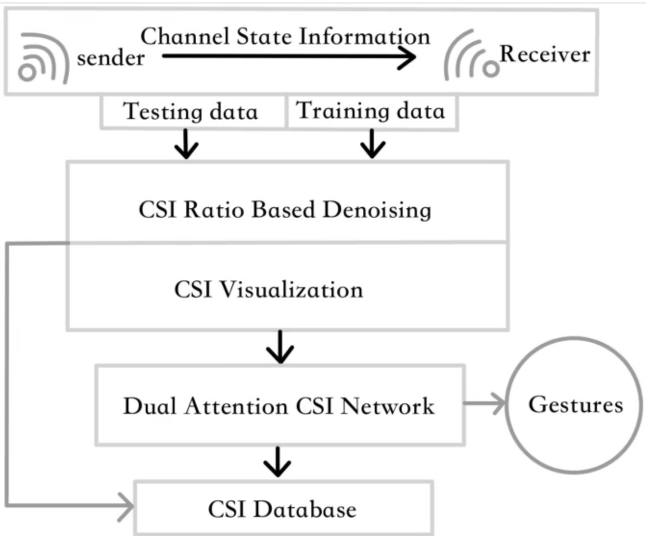

Figure 1. System model framework (Photo credit: Original).

In the system, the gesture recognition process is divided into two distinct modules: the CSI preprocessing module and the gesture recognition module, which employs a Dual-Attention CSI Network (DACN) as illustrated in Figure 1 [6]. Initially, the system utilizes WiFi transmitters and receivers strategically placed within the environment. It applies the CSI ratio method to effectively denoise the raw measurement data, thereby eliminating environmental noise and signal interference, and enhancing the data's quality and reliability. Post-denoising, the phase tensor from each subcarrier is extracted and organized into a two-dimensional matrix [7]. This matrix, where each row or column encapsulates the phase information from various subcarriers at a particular moment, offers a rich spatio-temporal dataset for further analysis. This matrix is subsequently transformed into a time series image, which vividly depicts the spatiotemporal dynamics of gestures, providing a high-resolution foundation for accurate gesture identification [8]. The Dual Attention CSI Network (DACN) then processes these time series images to recognize gestures. At the core of this module is the integration of a dual attention mechanism [9], which significantly enhances the network's ability to discern the sequential correlations inherent in gesture data, enabling effective cross-domain recognition. The DACN employs a ResNet (Deep Residual Network) equipped with dual attention modules, capitalizing on its profound capacity for deep feature extraction and ensuring robust gesture recognition.

The WiFi CSI signals can be described in the frequency domain as.

$$\vec{Y} = \vec{H} \cdot \vec{X} + \vec{N} \tag{1}$$

The superposition of every propagation path signal is known as CSI, and this can be used to express its channel frequency response:

$$H(f,t) = \sum_{m \in \Phi} a_m(f,t) e^{-j2\pi \frac{d_m(t)}{\lambda}} \tag{2}$$

Where $a_m(f,t)$ and $d_m(t)$ stand for the complex attenuation and propagation length of the m-th multipath component, respectively, Φ for the multipath component set, λ for the signal wavelength, and f and t for the center frequency and time stamp, m for the multipath components [10].

m in CSI-based gesture recognition is made up of both static and dynamic paths:

$$\begin{aligned} H(f,t) &= H_s(f,t) + H_d(f,t) \\ &= \sum_{m_s \in \Phi_s} a_{m_s}(f,t) e^{-j2\pi \frac{d_{m_s}(t)}{\lambda}} \\ &+ \sum_{m_d \in \Phi_d} a_{m_d}(f,t) e^{-j2\pi \frac{d_{m_d}(t)}{\lambda}} \end{aligned} \tag{3}$$

Where $H_s(f,t)$ and $H_d(f,t)$ represent dynamic and static components, respectively, Φ_s representing a set of static paths, such as reflections from a wall, furniture, or stationary body, Φ_d representing a set of dynamic paths, such as reflections from a moving body [11]. When detecting gestures, the motion of the hand and arm changes the dynamic propagation distance $d_{m_d}(t)$, thereby changing the phase shift of Φ_d to $e^{-j2\pi \frac{d_{m_d}(t)}{\lambda}}$, which ultimately affects $H(f,t)$. In general, this gesture can be depicted by changes in the CSI phase shift.

2.2. Preprocessing

$$e^{-j\theta_{offset}} \left(H_s(f,t) + A(f,t) e^{-j2\pi \frac{d(t)}{\lambda}} \right) \tag{4}$$

$$\begin{aligned} H_q(f,t) &= \frac{H_1(f,t)}{H_2(f,t)} \\ &= \frac{e^{-j\theta_{offset}} \left(H_{s,1} + A_1 e^{-j2\pi \frac{d_1(t)}{\lambda}} \right)}{e^{-j\theta_{offset}} \left(H_{s,2} + A_2 e^{-j2\pi \frac{d_2(t)}{\lambda}} \right)} \\ &= \frac{A_1 e^{-j2\pi \frac{d_1(t)}{\lambda}} + H_{s,1}}{A_2 e^{-j2\pi \frac{d_2(t)+\Delta d}{\lambda}} + H_{s,2}} \end{aligned} \tag{5}$$

Where the two receiving antennas' CSIs are represented by $H_1(f,t)$ and $H_2(f,t)$, Δd can be regarded as a constant, formula (5) table.Shows the rotation and scaling transformations of the phase shift $-j2\pi \frac{d_1(t)}{\lambda}$ antenna 1 in the complex plane, which do not affect the change of phase shift a lot.

2.3. Classification and Recognition

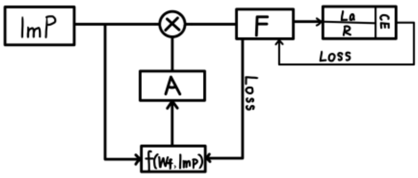

Figure 2. Basic recognition neural networks based on attention (Photo credit: Original).

The significance of each $\operatorname{Im}P$ image's pixels for gesture identification should be evaluated in order to maintain the current study method's search for gestures' fundamental cues while stifling the other data. As show in the figure 2.

$$P_A = \operatorname{Im}P \cdot A \tag{6}$$

$$A = f\left(W_f, \operatorname{Im}P\right) \tag{7}$$

Where the attention map W_f is symbolized by A. The values of the parameters of the function f are represented by W_f, which is initialized at random. P and A have the same dimensions, and each pixel in A represents its weight in $\operatorname{Im}P$; the greater the weight, the more significant the pixel in $\operatorname{Im}P$.

$$R = F\left(P_A\right) \tag{8}$$

Cross-entropy is used to process the real label La and the classification result R given by the training set.

$$L = CE\left(R, La\right) \tag{9}$$

$$\frac{\partial L}{\partial W_f} = \frac{\partial L}{\partial F}\frac{\partial F}{\partial W_f} \tag{10}$$

As a consequence, the network uses gradient descent to update the parameter W_f:

$$W_f = W_f - \alpha \frac{\partial L}{\partial W_f} \tag{11}$$

The process of figuring out the loss L and modifying the parameter W_f keeps repeating until the model converges, where α stands for the learning rate.

3. Experiment and Results Analysis

3.1. Experimental Setup and Parameter Configuration

This study uses the Widar 3.0 dataset, which covers a variety of gestures and their corresponding labels, including the following six gestures: All gestures are recorded and analyzed in a Hall environment.

3.2. Gesture Recognition Process

Based on the premise of using ResNet18, the DACN process can be expressed as:

$$\begin{aligned}
&\operatorname{Im} P' = A_{tsa}\left(\operatorname{Im} P\right) \otimes \operatorname{Im} P \oplus \operatorname{Im} P \\
&\operatorname{Im} P'' = \operatorname{Re} sNet18\left(\operatorname{Im} P'\right)\backslash \\
&\operatorname{Im} P''' = A_{tsb}\left(\operatorname{Im} P''\right) \otimes \operatorname{Im} P'' \\
&RecognitionResults = Softmax\left(Avgpool\left(\operatorname{Im} P'''\right)\right)
\end{aligned} \tag{12}$$

In this case, $\operatorname{Im} P$ stands for the phase matrix P heat map, $\operatorname{Im} P'$ for the first attention processing output and the ResNet18 input, $\operatorname{Im} P''$ for the ResNet18 output and the second attention processing input, and $\operatorname{Im} P'''$ for the second attention processing output and the recognition processing input.

3.3. Experimental Results and Evaluation

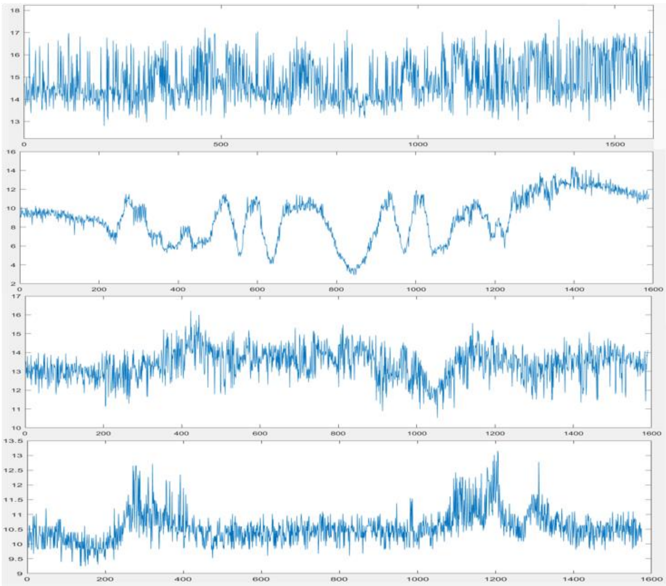

Figure 3. Phase waveform of different pairs of antennas (from top to first to fourth pair of antennas) (Photo credit: Original).

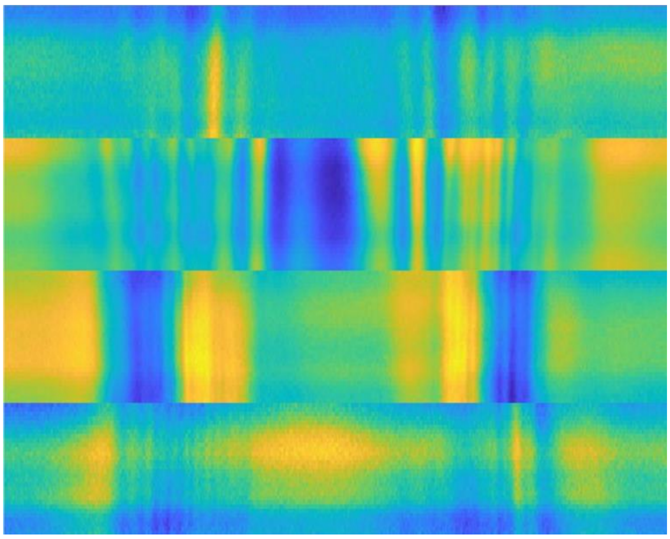

Figure 4. Original input diagram (first to fourth antenna pairs from top to bottom (Photo credit: Original).

The CSI phase waveform following noise reduction is displayed in figure 3 Gestures can cause changes in the CSI phase. The temporal dimension of figure 3 makes it simple to identify the important times for gesture recognition. In the input figure 4, the phase waveform images corresponding to the first to fourth pairs of antennas are shown.

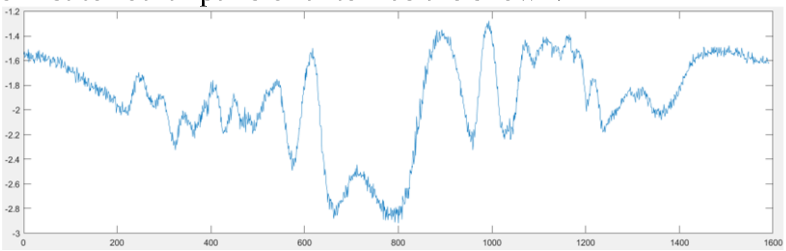

Figure 5. 5Phase waveform of subcarrier 2 (Photo credit: Original).

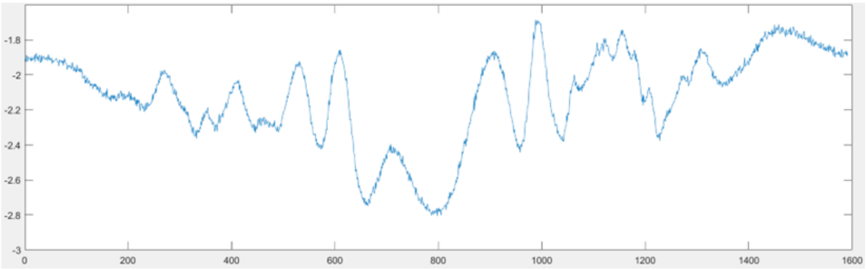

Figure 6. Phase waveform of subcarrier 25 (Photo credit: Original).

This suggests that the second subcarrier is not gesture-sensitive and thus warrants less attention. As show in the figure 5, figure 6, figure 7 and table 1.

Table1. The result of gesture recognition is compared with the most advanced scheme.

Different Gesture Tags	1~3	4~6
Training Data Accuracy	98.4%	98.0%%
Data Accuracy of This Study	99.3%	99.5%

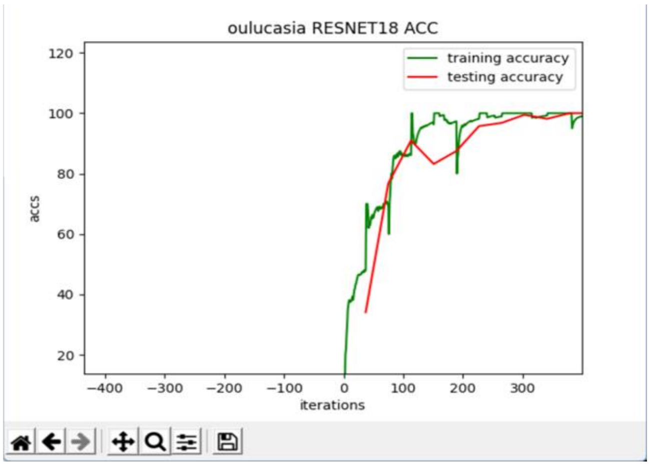

Figure 7. Part of the label data set training data accuracy and test data accuracy (Photo credit: Original).

4. Discussion And Future Research Directions

However, its practical deployment faces a myriad of technical challenges and environmental adaptability issues. One primary obstacle is the reliance on WiFi or other wireless signals for data transmission and perception. In real-world settings, these signals are often subject to interference from various electronic devices, which can degrade signal quality and subsequently impact the accuracy of gesture recognition. In complex environments, signals can reach the receiver through multiple paths, a situation referred to as the multipath effect. This phenomenon leads to the superposition of multiple signals, resulting in distortion of key signal features such as phase and amplitude. Consequently, the overall performance of gesture recognition may suffer, necessitating advanced techniques to mitigate these effects. Environmental adaptability poses additional hurdles. For instance, when multiple users perform gestures simultaneously within a shared space, interference with the wireless signal can occur, adversely affecting the recognition accuracy of the system. As such, the technology must be capable of distinguishing between gestures from different users. Moreover, environmental noise—ranging from background conversations to mechanical sounds—can further complicate the gesture recognition process. To address this, a robust system is essential, one capable of resisting the impacts of such ambient noise.

Furthermore, the discourse surrounding data privacy and security is increasingly pertinent in the context of wireless sensing gesture recognition technology. Upon data collection, the subsequent transmission, storage, and processing of this information raise significant concerns regarding privacy. Factors such as the chosen data storage locations (whether on-premises devices or cloud servers), the duration of data retention, and the methods of data processing all play critical roles in maintaining data integrity and confidentiality. In the realm of cybersecurity, the technology is vulnerable to various threats during data transmission. Potential risks include man-in-the-middle attacks, packet sniffing, and forgery of data packets. Successful execution of such attacks could enable malicious actors to access sensitive information or even tamper with it, underscoring the necessity for robust security measures to safeguard data. Healthcare professionals can leverage gesture controls to interact with medical devices, significantly diminishing the need for physical contact in sterile environments. Additionally, monitoring patients' body movements and gesture changes through wireless sensing technology can greatly assist in rehabilitation and physical therapy, providing valuable insights for practitioners evaluating patient recovery.

For future development, a crucial direction involves enhancing the interconnectivity of wireless sensing technology with other smart devices to create an intelligent ecosystem. This integration would enable gesture operations that transcend single-device constraints, facilitating seamless interactions across multiple devices. As the technology continues to evolve, its adaptability and application scope promise to expand, transforming how individuals engage with their environments and enhancing diverse sectors, including healthcare, retail, and entertainment.

5. Conclusion

This study has demonstrated that the Dual Attention Network offers substantial accuracy improvements over existing advanced gesture recognition systems. By evaluating 750 data samples across a range of gestures, torso positions, facial orientations, and case scenarios, this research has established a robust foundation for the accurate and reliable recognition of gestures. The innovative approach adopted here discards traditional handcrafted features, which simplifies the preprocessing steps and mitigates potential information loss during feature extraction. The adaptive feature extraction method effectively captures essential gesture characteristics, enhancing the overall recognition performance and streamlining the operation of the system. This advancement is particularly beneficial for people with disabilities, offering them a more natural and efficient way to communicate and interact with their environment, thereby enhancing their independence and quality of life.

Looking forward, there is significant potential to expand the application of gesture recognition technology, especially in the realm of assistive technologies and barrier-free communication. This could have profound implications for enhancing social inclusivity and equity. Future research should focus on further refining the accuracy and responsiveness of gesture recognition systems, exploring their integration into more complex and varied environments. Additionally, extending these technologies to support a broader spectrum of disabilities could profoundly impact accessibility and usability. By continuing to advance gesture recognition technology, future developments can open new avenues for communication and interaction, promoting greater participation of disabled individuals in all aspects of society.

References

[1] Y. Gu, X. Zhang, Y. Wang, M. Wang, H. Yan, Y. Ji, M. Dong, WiGRUNT: WiFi-enabled gesture recognition using dual-attention network, IEEE Trans. Hum. -Mach. Syst. 52 (2022) 736-746.

[2] Y. Gu, H. Yan, X. Zhang, Attention-based gesture recognition using commodity WiFi devices, IEEE Sens. J. 23 (2023) 9685-9696.

[3] X. Zhu, Y. Zhang, Z. Zhao, et al., Radio frequency sensing based environmental monitoring technology, in: Fourth International Workshop on Pattern Recognition, Vol. 11198, SPIE, 2019, pp. 187-191.

[4] C. Li, M. Liu, Z. Cao, WiHF: Gesture and user recognition with WiFi, IEEE Trans. Mobile Comput. 21 (2020) 757-768.

[5] R. Gao, M. Zhang, J. Zhang, Y. Li, E. Yi, D. Wu, D. Zhang, Towards position-independent sensing for gesture recognition with Wi-Fi, Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. 5 (2021) 1-28.

[6] X. Zhu, H. Xu, Z. Zhao, X. Wang, X. Wei, Y. Zhang, J. Zuo, An environmental intrusion detection technology based on WiFi, Wireless Personal Communications, 119 (2) (2021) 1425-1436.

[7] R. Zhang, C. Jiang, S. Wu, Q. Zhou, X. Jing, J. Mu, Wi-Fi sensing for joint gesture recognition and human identification from few samples in human-computer interaction, IEEE J. Sel. Areas Commun. 40 (2022) 2193-2205.

[8] M. Yang, H. Zhu, R. Zhu, F. Wu, L. Yin, Y. Yang, Witransformer: A novel robust gesture recognition sensing model with WiFi, Sensors 23 (2023) 2612.

[9] R. Xiao, J. Liu, J. Han, K. Ren, Onefi: One-shot recognition for unseen gesture via cots wifi, in: Proc. 19th ACM Conf. Embedded Networked Sens. Syst., ACM, New York, 2021, pp. 206-219.

[10] X. Zhu, Z. Zhao, X. Wei, et al., Action recognition method based on wavelet transform and neural network in wireless network, in: 2021 5th International Conference on Digital Signal Processing, 2021, pp. 60-65..

[11] S. Tan, J. Yang, Y. Chen, Enabling fine-grained finger gesture recognition on commodity WiFi devices, IEEE Trans. Mobile Comput. 21 (2020) 2789-2802.