Dean&Francis

ISSN 2959-6157

An Investigation of Gesture Recognition Technology Based on Artificial Intelligence and Wireless Sensing

Haile Tong

Department of Software Engineering, Dalian University of Technology, Dalian, China
Haile200412@outlook.com

Abstract:

Now human-computer interaction technology has been closely integrated with the daily life, of which gesture recognition technology has received special attention, especially those using artificial intelligence wireless sensing solutions. This wireless gesture recognition technology mainly determines hand movements by analyzing changes in radio signals, and it has many advantages, such as better protecting user privacy, not affected by ambient light, and compatible with various devices. However, this technology also has many problems, such as unstable wireless signals, insufficient algorithm adaptability, and error due to different environments, devices or users. Models include LAGER, WiGNN, Wi-AM, and WiVi-GR. Each of these models has its own characteristics. LAGER solves the problem of labelless data through adversarial mapping, WiGNN improves cross-domain adaptability with dynamic topology, Wi-AM combines meta-learning to handle sample deficiencies, and WiVi-GR combines wireless signals with visual signals to break through the limitations of a single mode. The article concludes with some problems that the technology has current aspects, such as multiple users' difficulty in distinguishing when making gestures at the same time.

Keywords: Gesture recognition; wireless sensing; deep learning.

1. Introduction

The extensive adoption of computer technology in society has resulted in the integration of Human-computer Interaction (HCI) into various aspects of daily life. The advancement of this technology is of significant importance for enhancing the efficiency of computer applications. In the context of expanding application scenarios, including smart homes, virtual reality, and interactive gaming, research and application of gesture recognition technology is receiving increasing attention.

During the period spanning from the 1960s to the 1980s, the field of gesture recognition technology was in its nascent stage, with the predominant reliance on electromagnetic induction techniques [1]. However, these methods suffered from limited operational range and susceptibility to interference from surrounding metallic objects, resulting in poor recognition accuracy and stability. During the 1990s to 2000s, traditional gesture recognition technologies utilising cameras or wearable sensors continued to evolve [2]. However, the practical implementation of these systems was constrained by factors such as lighting conditions, obstruction scenarios, and privacy concerns. Recent technological developments in the field of WiFi and artificial intelligence have paved the way for the emergence of wireless sensing technologies that are capable of 'contactless' perception. This has led to a new frontier in both research and application in this field. Wireless gesture recognition technology is predicated on the analysis of variations in radio frequency signal propagation characteristics Received Signal Strength Indication, Channel State Information, Doppler Frequency Shift(RSSI, CSI, DFS) in order to interpret human hand movements. The Label-Free Domain-Adaptive Wireless Gesture Recognition via Latent Feature Alignment and Augmentation (LAGER) [3] employs adversarial mapping and pseudo labels, utilising an Adversarial Autoencoder (AAE) to map data distributions from two domains onto a shared prior distribution space. This approach serves to reduce reliance on the original dataset, thereby achieving recognition without the need for target domain labels or additional information through feature enhancement guided by pseudo labels. The WiGNN model [4] employs a causal multi-scale time-frequency feature fusion layer (CMTF) and enhances cross-domain recognition accuracy through a dynamic topology aggregation layer, without requiring target domain data. In order to reduce deployment costs and data collection burdens, the Wi-AM system [5] incorporates a meta-learning module and an adversarial generalisation module, thereby enabling high-precision recognition with only 1-3 samples per gesture. WiVi-GR [6] integrates wireless and visual signals, employing channel-stacked convolutional networks and multimodal velocity spectrum fusion to overcome single-modality limitations, thereby enabling high-precision interaction.

Wireless sensing technology offers enhanced protection of user privacy, operates independently of ambient lighting conditions and device prevalence, and captures more nuanced vital signs. However, the system is currently facing technical challenges, including unstable wireless signal propagation and insufficient generalisation capabilities of existing algorithms. Through the utilisation of machine learning and convolutional neural networks, there has been a gradual enhancement in recognition accuracy. However, further optimisation is required to address variability in complex environments. This study provides an overview of wireless sensing and artificial intelligence-based gesture recognition technologies and their applications. It elucidates relevant theoretical frameworks within the field, summarises methodologies through a literature review of novel approaches, conducts objective comparative analysis, presents research conclusions, and offers future prospects.

2. Methods

2.1 Strategies for Addressing Domain Shift

2.1.1 WiGNN

Domain shift issues exist in Wi-Fi gesture recognition. This paper proposes a graph neural network model, WiGNN, featuring a dynamic topology structure [4]. It mitigates variations in data distribution caused by environmental, device, or user differences, enhancing robustness in cross-domain recognition. During WiFi wireless sensing, domain shift is predominantly triggered by environmental noise and hardware discrepancies, such as multipath effects and frequency deviations. Traditional static feature extraction methods exhibit poor generalisability. WiGNN models WiFi channel state information and Doppler shift data as a dynamic topological graph. Through node embedding and adjacency matrix updates, it captures stable patterns of gesture motion in real time. The ‚Structure-Aware - Feature Alignment' dual-layer framework filters out environmental noise while preserving domain invariance. It employs causal reasoning to suppress confusion bias, preventing spurious association learning. Requiring no target domain annotations, it achieves feature alignment through domain-adaptive loss functions, thereby enhancing accuracy across varying user scenarios. This solution can be deployed across multi-device environments, reducing retraining costs and providing robust perception support for smart homes and remote interactions.

2.1.2 Wi-AM

Wi-AM addresses domain shift in WiFi gesture recognition through a "multi-domain adversarial + meta-learning" approach [5]. Adaptation to new domains requires only 1 to 3 samples, mitigating distribution shifts caused by environmental, user, and device variations, while reducing data costs. Domain shift stems from multipath effects and hardware differences, necessitating extensive target domain data for retraining in traditional methods. Wi-AM first purifies CSI signals through filtering, PCA, denoising, and STFT-based Doppler feature extraction. It then replaces traditional single classifiers with multi-domain

ISSN 2959-6157

classifiers, combining attention-weighted features to maximise gesture label accuracy while minimising domain label recognition capability. This extracts domain-invariant features, offering advantages over conventional methods: In domain adaptation precision, multi-domain adversarial matching addresses complex distributions to reduce false associations. Regarding adaptation efficiency, meta-learning leverages task generation and prototype networks to achieve rapid fine-tuning with few samples. In terms of hardware cost, only a single WiFi transceiver pair is required. Experiments demonstrate single/three-sample accuracy rates of 82.13%/86.76% in unseen domains, providing a low-cost cross-domain solution for smart homes and similar applications.

2.2 Strategies for Addressing the Scarcity of Annotated Data

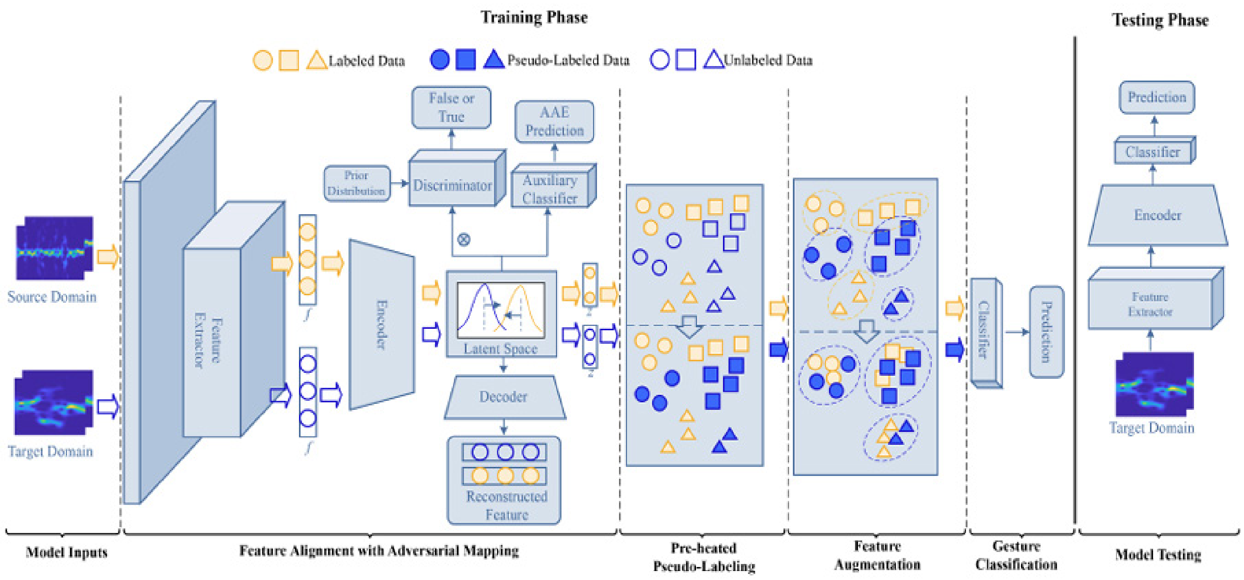

Fig. 1. Lager Via Latent Feature Alignment and Augmentation [3].

To achieve cost-effective deployment of Wi-Fi gesture recognition and address the scarcity of annotated data, Chen et al. proposes LAGER shown in Fig. 1—an efficient domain adaptation method for Wi-Fi-based gesture recognition. This approach maintains high recognition accuracy in novel environments without relying on target-domain annotated samples.LAGER decomposes cross-domain Wi-Fi gesture recognition into two interrelated subtasks: leveraging the technical advantages of unsupervised domain adaptation (UDA), it iteratively performs pseudo-label-guided feature alignment and feature enhancement operations within the latent feature space. To mitigate the potential negative impact of erroneous pseudo-labels during the initial training phase, the team further introduced warm-up training. This technique divides the model training process into two distinct stages, each employing tailored training strategies to reduce the interference from early-stage pseudo-label noise.Evaluations were conducted on the Widar3.0 public dataset to validate the method's efficacy,. Performance comparisons against representative baselines—including EI, CA, WiGRUNT, and Widar3.0—were performed across diverse cross-domain settings (e.g., cross-environment, cross-location, cross-orientation, cross-user). Experimental results demonstrate that the proposed LAGER method significantly outperforms all comparison benchmarks. Notably, compared to cross-domain recognition methods employed in Widar3.0, LAGER achieves an average recognition accuracy improvement of 6.89%–7.83% across different cross-domain experimental scenarios, fully validating its cross-domain transfer capability in scenarios with sparse annotations.

3. Comparison and discussion

3.1 Comparison

Table 1. Comparison of Four Models

Model	Innovation Points	Advantages	Disadvantages
LAGER	Adversarial Mapping Pseudo Labels	No target domain tag or additional information required	Multimodal fusion deficiency
WiGNN	Causal Multiscale Time-Frequency Feature Fusion Layer (CMTF)	Enhanced cross-domain recognition accuracy without requiring target domain data	Unable to resolve the issue of separating multiple users' gestures
Wi-AM	The Yuan Learning Module and the Adversarial Generalisation Module	Each gesture requires only 1-3 samples	Accuracy decreases when locations are constantly changing.
WiVi-GR	Integrating wireless signals and visual signals	Suitable for high-precision interaction	Time synchronisation error affects results

Shown in Table 1, LAGER employs adversarial mapping and pseudo-labels to enable recognition without requiring target domain labels or supplementary information. However, it suffers from a lack of multimodal fusion capabilities and exhibits unsupervised domain generalisation limitations. WiGNN introduces a causal multi-scale time-frequency feature fusion layer (CMTF), enhancing cross-domain recognition accuracy without necessitating target domain data. Nevertheless, it can only recognise single-person gestures and fails to address multi-person gesture separation challenges. Wi-AM incorporates a meta-learning module and an adversarial generalisation module, requiring only 1-3 samples per gesture for each recognition instance. However, accuracy diminishes when user and environmental positions undergo continuous variation. WiVi-GR integrates wireless and visual signals through multimodal velocity spectrum fusion, overcoming the limitations of single-modality approaches. It is suitable for high-precision interactions, though results are susceptible to time synchronisation errors between devices.

3.2 Challenges and future prospects

Wireless sensor-based gesture recognition technology has achieved significant breakthroughs, yet numerous unresolved issues persist at both technical and practical levels, making it challenging to fully integrate with complex real-world environments. The challenge of addressing stable wireless signal transmission is currently struggled with:1)Although models such as WiGNN and Wi-AM employ dynamic topology mapping or signal purification techniques to mitigate these effects, they inevitably prove unsatisfactory under certain extreme conditions such as dense multipath effects in crowded settings, severe signal attenuation caused by non-metallic obstacles, or cross-interference between multiple electronic devices within a household. 2)Existing methods struggle to distinguish non-centred gesture features formed by overlapping movements from two or more users, leading to a sharp decline in recognition accuracy under such conditions. Furthermore, the limited generalisation capability in complex environments substantially diminishes the true practical value of this novel recognition technology. 3) Although domain-adaptive algorithms reduce the requirement for target domain data samples, they cannot rapidly respond to instantaneous changes in scenarios—such as furniture being suddenly removed/rearranged overnight, or users moving at slower/faster speeds, or switching to hardware—or alterations in indoor air humidity. For such variations, the computational capacity of edge deployment in smart living environments is severely constrained, meaning the adoption of more complex adaptive models would compromise real-time performance requirements. Therefore, without compromising performance, or shifts to hardware, or alterations in indoor humidity. For such changes, the computational capacity of edge deployment in smart living is severely limited. Employing more complex adaptive models would compromise real-time requirements, making implementation challenging without sacrificing performance. 4)Finally, insufficient depth in multimodal fusion hinders overall improvement. WiVi-GR combines wireless and visual signals for training. While these signals are fused, this integration remains at the general data/feature level without semantic alignment of spatio-temporal characteristics. As the time scales between modalities are asynchronous, this introduces errors, resulting in suboptimal accuracy. Furthermore, incorporating additional modalities like IMU introduces extra data

ISSN 2959-6157

and computational burdens.[7]

In future, it is also available to explore solutions to some of the aforementioned issues. Development directions: 1) Develop physically informed deep learning techniques to enhance signal robustness: Embed and integrate physical knowledge of electromagnetic wave propagation into end-to-end networks such as convolutional neural networks, Transformers, and LSTMs. This improves noise reduction and the accuracy of effective information extraction, potentially avoiding cumbersome and bulky preprocessing methods like STFT/PCA. 2) Employ federated continuous learning to construct generalisable models: Federated training enables interaction and communication between edge devices while safeguarding user privacy during collaborative model training. Continuous learning prevents "catastrophic forgetting" [8, 9], allowing machines to adapt to local environmental changes without retraining, utilising limited initial training data for new requirements. 3) Semantic-level fusion for multimodal data enables superior alignment of wireless temporal features with visual spatial features at higher levels, achieving more precise synchronisation between them. This approach further reduces costs through its lightweight design; Furthermore, extending these technologies to other high-risk scenarios—such as medical hand-held devices, elderly assistive equipment, and home care appliances—enhances performance while improving acceptability through privacy-friendly signal processing [10].

4. Conclusion

This study reviews gesture recognition schemes based on artificial intelligence and wireless sensing. The LARGER, Wi-AM, WiVi-GR, and WiGNN models were subjected to systematic analysis. This paper analyses and reviews the principles, innovations, and advantages and disadvantages of these four models. While the accuracy and privacy of gesture recognition have been significantly enhanced, challenges remain, including insufficient cross-domain generalisation capabilities, inadequate multimodal fusion capabilities, and limitations imposed by deployment conditions. In the future, federated continuous learning will be employed to construct generalisable models, while semantic-level fusion of multimodal data will achieve precise feature alignment and cross-domain generalisation. Wireless gesture recognition will find increasing application in everyday life.

References

[1] Sluyters A, Lambot S, Vanderdonckt J. Hand Gesture Recognition for an Off-the-Shelf Radar by Electromagnetic Modeling and Inversion. IUI '22: Proceedings of the 27th International Conference on Intelligent User Interfaces. 2022; 506–522. doi:10.1145/3490099.3511107.

[2] Tchantchane R, Zhou H, Zhang S, Alici G. A Review of Hand Gesture Recognition Systems Based on Noninvasive Wearable Sensors. Adv Intell Syst. 2023; doi:10.1002/aisy.202300207.

[3] Chen J, Bi S, Lin XH, Quan Z. LAGER: Label-Free Domain-Adaptive Wireless Gesture Recognition via Latent Feature Alignment and Augmentation. IEEE Internet Things J. 2024; 11(23): 37928–37941.

[4] Chen Y, Huang X. WiGNN: WiFi-Based Cross-Domain Gesture Recognition Inspired by Dynamic Topology Structure. IEEE Wireless Commun Mag. 2023; doi:10.1109/MWC.023.2200610.

[5] Xie J, Li Z, Feng C, Lin J, Meng X. Wi-AM: Enabling Cross-Domain Gesture Recognition with Commodity Wi-Fi. Sensors. 2024; 24(5): 1354.

[6] Liu X, Tang S, Zhang B, Wu J, Ma X, Wang J. WiVi-GR: Wireless-Visual Joint Representation-Based Accurate Gesture Recognition. IEEE Internet Things J. 2024; 11(2): 2701–2711.

[7] Li W, Shi P, Yu H. Gesture Recognition Using Surface Electromyography and Deep Learning for Prostheses Hand: State-of-the-Art, Challenges, and Future. Front Neurosci. 2021; 15: 621885. doi:10.3389/fnins.2021.621885.

[8] Li L, Fan Y, Tse M, Lin KY. A review of applications in federated learning. Comput Ind Eng. 2020; 149: 106854.

[9] Mammen PM. Federated learning: Opportunities and challenges. arXiv preprint arXiv:2101.05428. 2021.

[10] Mohamed N, Mustafa MB, Jomhari N. A Review of the Hand Gesture Recognition System: Current Progress and Future Directions. IEEE Access. 2021; 9: 157422–157436.