LipHS : A Lightweight WiFi-enabled Human Sensing For Multi-Class Scenarios

Chunhao Xue, Lixing Wang(✉)

School of Computer Science and Engineering, Northeastern University, Shenyang 110000, China
wanglixing@mail.neu.edu.cn

Abstract. WiFi-based human sensing technology utilizing Channel State Information (CSI) has garnered significant attention due to its reduced privacy concerns and the widespread availability of existing infrastructure, demonstrating broad development prospects in the field of intelligent computing. Deployment on edge devices represents its most prevalent application scenario. However, the high-complexity algorithms commonly employed to enhance sensing accuracy face substantial challenges when deployed on devices with limited computational resources. Furthermore, most existing studies conduct experiments only on datasets with a small number of categories which fail to meet practical sensing requirements. Consequently, developing high-accuracy, and practical WiFi-based human sensing systems remains considerably challenging. To address the core computational bottleneck of WiFi sensing systems, we presents LipHS, a feature extraction framework capable of simultaneously capturing multi-level information from CSI signals. To further reduce the number of model parameters, we employ a channel pruning method based on Layer-Adaptive Magnitude-based Pruning (LAMP) scores. Experimental results demonstrate that the proposed LipHS method achieves superior recognition accuracy compared with mainstream general-purpose backbone networks on complex multi-class gesture datasets.

Keywords: WIFI Sensing • Channel State Information • Human Activity Recognition • Lightweight

1 Introduction

With the development and proliferation of wireless networks, research and applications in the field of wireless sensing have been increasingly growing. Compared to other sensing solutions, WiFi-based human sensing offers advantages such as low cost, no need for wearable sensors, and reduced privacy intrusion, which has garnered widespread attention from both academia and industry [1,2]. Notably, recent studies demonstrate that WiFi Channel State Information (CSI) signals hold significant potential for diverse device-free human sensing applications, such as occupancy monitoring, fall detection, gesture recognition, and human identification, as depicted in Fig.1. Unlike

coarse-grained Received Signal Strength Indicator (RSSI), CSI captures more fine-grained information.

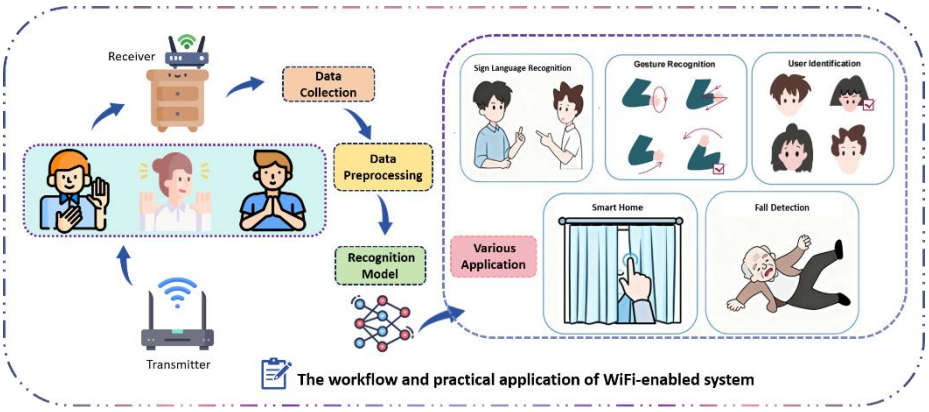

Fig. 1. Workflow and application of WiFi-enabled system

Motivated by the need for secure, device-free and pervasive gesture recognition solutions, WiFi-enabled sensing approaches have attracted extensive research attention. Nevertheless, the research area of WiFi-based human activity recognition still encounters various challenges.

First, some studies adopt sophisticated models such as deep residual networks and Transformers. These models usually have large parameter sizes and high computational costs, making them difficult to deploy on resource-constrained edge devices (e.g., development boards, smart home gateways), which represent the actual application scenarios of WiFi-based human sensing.

Second, most existing studies primarily conduct recognition evaluations on carefully selected subsets of gestures or behaviors with a small number of categories, typically 5 to 6 classes. Such settings often yield exceptionally high accuracy on test datasets; however, model performance degrades significantly when tackling complex multi-class classification tasks. Recognition on gesture datasets with a larger number of categories remains under-explored, which represents an essential step toward the real-world deployment of WiFi-based human sensing technology from academic research.

To address these issues, LipHS, a lightweight WiFi CSI human sensing system empowered by model pruning algorithms has been proposed. All experiments are conducted on a fused 22-class gesture dataset from Widar 3.0, which adopts a unified data collection protocol and strict random split to avoid data leakage and shortcut learning. The main contributions of this paper are summarized as follows:

1) LipHS, an accurate and lightweight WiFi based human sensing framework has been proposed, which investigates how to achieve high sensing performance with low computational complexity in challenging environments.
2) Based on the ConvNeXt V2 block, the LipHS module serves as the fundamental feature extraction unit of the LipHS backbone network, while incorporating an attention mechanism to simplify the model structure while maintaining high

recognition accuracy. A channel pruning method based on LAMP scores is used to remove redundant weights, thereby achieving the lightweighting objective.

3) Experimental results demonstrate that, compared with mainstream backbones adopted in WiFi sensing, LipHS achieves the highest recognition accuracy on the fused dataset, while fully meeting the constraints of edge devices.

2 Related Work

In the field of WiFi-based CSI human sensing, deep learning methods have gradually replaced traditional physical model-based approaches and become the mainstream technical paradigm for various sensing tasks owing to their powerful automatic feature extraction capabilities. To pursue higher recognition accuracy, existing studies have extensively migrated mainstream deep architectures from computer vision and natural language processing. Starting from early Multi-Layer Perceptrons (MLPs) and basic Convolutional Neural Networks (CNNs), the architectures have progressively evolved into Recurrent Neural Networks (RNNs) and their variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), as well as sophisticated models including CNN-RNN hybrid structures and Transformers, achieving continuous performance improvements in core tasks such as human activity recognition, gesture recognition, and user identity authentication [3,4,5,6]. To further enhance the feature fitting capacity of models, many studies have continuously deepened network depths by adopting deep residual networks like ResNet-18/50/101 or stacking multi-head attention modules to construct complex Transformer architectures. Although such non-lightweight models achieve recognition accuracy exceeding 98% on specific experimental datasets and settings [7], they inevitably lead to a drastic expansion of parameter volume and computational complexity. Benchmark results indicate that the FLOPs of the Transformer-based Vision Transformer (ViT) model can reach 501.64M, and the computational overhead of deep ResNet-101 is also far higher than that of shallow CNN architectures, creating a prominent contradiction with the practical deployment requirements of WiFi sensing.

WiFi sensing is primarily applied in common scenarios such as smart homes and intelligent buildings, where sensing functions are generally deployed on edge devices like routers and IoT terminals. These devices typically face strict limitations on computing capability, memory, and power consumption. Heavyweight models impose excessive computational overhead, leading to high inference latency that fails to meet real-time requirements for tasks such as fall detection and vital sign monitoring. Additionally, they significantly increase device power usage, hindering large-scale and long-term commercial deployment. Meanwhile, experimental results demonstrate that deep non-lightweight models suffer from severe overfitting in cross-domain and multi-environment scenarios, with generalization performance far inferior to shallow network architectures [7]. The complex model structure further raises the costs of data annotation and model training, which deviates from the development goals of low cost and high generalization in WiFi sensing. Therefore, designing lightweight, low-complexity WiFi sensing models that maintain high recognition accuracy has become a critical issue to be urgently addressed in this field, bearing important research value and practical application significance.

To lower model complexity and facilitate practical deployment, research efforts have increasingly focused on developing lightweight network architectures. Xing et al. [8] designed a lightweight neural network that enables fast and accurate recognition on low-cost edge devices. This architecture leverages pointwise convolution to fuse multi-channel features, which significantly reduces computational costs. Sheng et al. [9] introduced LiteWiSys, a lightweight end-to-end deep learning framework for human activity recognition. It integrates depthwise separable convolution with channel attention mechanisms into a multi-scale convolutional backbone, and constructs two network branches via feature channel splitting, alongside a joint loss function for model training. Feng et al. [10] proposed a lightweight autoencoder to learn meaningful representations from WiFi signals, and further developed a novel meta-learner that supports rapid domain adaptation with only one sample per gesture. Shafiqul et al. [11] presented HHIAttentionNet, a model built to enhance recognition performance for human-human interaction tasks. It captures key discriminative features from WiFi CSI data using customized convolutional modules and antenna-frame-subcarrier attention mechanisms. In prior work [12], Liu et al. developed LWiHS, a lightweight WiFi CSI-based human sensing system. It leverages a feature fusion-driven time series-to-image conversion strategy to achieve efficient human activity and gesture recognition with greatly simplified model structure. However, this framework is only validated on datasets with a small number of action categories, and its generalization ability on large-scale fine-grained gesture datasets remains unexplored.

To sum up, existing lightweight WiFi sensing works mainly focus on reducing model complexity for low-class recognition tasks, while the lightweight backbone design for fine-grained multi-class WiFi sensing scenarios remains significantly under-explored. Most lightweight models suffer from severe performance degradation in complex multi-class tasks, which cannot meet the requirements of practical multi-scene applications. This core research gap is the key motivation of our work.

3 Signal Preprocessing

CSI captured by commodity Wi-Fi devices characterizes indoor multipath effects at packet arrival time t and subcarrier frequency f :

$$In\widehat{H}(f,t) = (\sum_{l=1}^{L} \alpha_l(f,t) e^{-j2\pi f \tau_l(f,t)}) e^{j\in(f,t)} \quad (1)$$

where L is the number of paths, α_l and τ_l are the complex attenuation and propagation delay of the l-th path, and $\in (f,t)$ is the phase error caused by timing alignment offset, sampling frequency offset and carrier frequency offset. By representing phases of multipath signals with the corresponding DFS, CSI can be transformed as [13]:

$$\widehat{H}(f,t) = (H_s(f) + \sum_{l\in P_d} \alpha_l(t) e^{j2\pi \int_{-\infty}^{t} f_{D_l}(u)du}) e^{j\in(f,t)} \quad (2)$$

where the constant $H_s$is the sum of all static signals with zero DFS, and $P_d$is the set of dynamic signals with non-zero DFS.

With conjugate multiplication of CSI of two antennas on the same Wi-Fi NIC calculated, and out-band noises and quasi-static offsets filtered out, random offsets can be removed and only prominent multipath components with non-zero DFS are retained

[14]. As subcarriers of CSI are correlated with each other and each of them embodies different center frequency and DFS that is related to the wavelength of the signal, then adopt PCA-based algorithm proposed in CARM [15] to extract principal components of CSI streams. Further applying short-term Fourier transform yields power distribution over the time and Doppler frequency domains. We denote each time snapshot in spectrograms as a DFS profile. Based on DFS profile from multiple links, we can derive domain-independent BVP.

Then define a user-centered two-dimensional body coordinate system: the origin is at the center of the user's torso, the positive direction of the *x*-axis coincides with the user's facing direction, and the coordinate system transforms synchronously with the user's position/orientation, achieving decoupling between gesture representation and global domain information. In the body coordinate system, the Doppler frequency shift generated by an arbitrary velocity vector $\vec{v}^{(b)}$on the *i*-th Wi-Fi link satisfies:

$$f^{(i)}(\vec{v}^{(b)}) = a_x^{(i)} v_x^{(g)} + a_y^{(i)} v_y^{(g)} \tag{3}$$

where $\vec{v}^{(g)}$represents the mapping result of $\vec{v}^{(b)}$in the global coordinate system, and $a_x^{(i)}$ and $a_y^{(i)}$ are the Doppler projection coefficients of the link, which are determined solely by the transceiver positions and the Wi-Fi signal wavelength λ.

The estimation of BVP represents an underdetermined inverse problem. Utilizing the sparse characteristic of gesture motion (a single gesture includes few dominant velocity components, and BVP is highly sparse), it is formulated as an l_0-norm optimization problem with sparsity constraints:

$$min_V \sum_{i=1}^{M} |EMD(A^{(i)}V, D^i)| + \eta||V||_0 \tag{4}$$

where *M* represents the total number of WiFi links, EMD denotes the Earth Mover's Distance used to measure differences in spectral distribution, η is the sparsity coefficient, and $||V||_0$ represents the l_0-norm enforcing sparsity. By solving this optimization problem, the BVP matrix can be estimated from the multi-link DFS spectrum, and after window-based processing, a complete BVP temporal sequence is obtained.

It should be noted that the CSI preprocessing pipeline adopted in this work is a general and widely used framework in mainstream WiFi sensing research. All operations in this pipeline are linear transformations and sparse optimization with fixed computational steps, which can be executed in real time on low-power edge devices with extremely low computational overhead. The core computational bottleneck of the entire WiFi sensing system lies in the deep learning-based feature extraction backbone, which is the core focus of our lightweight design in this paper.

4 Lightweight Model Design

In this section, we provide a detailed description of the design of the feature extraction network in LipHS.

4.1 Framework Design

The backbone of the LipHS architecture is built upon ConvNeXt V2 [16]. Built on the original ConvNeXt architecture, ConvNeXt V2 incorporates the Masked Autoencoder (MAE) strategy, which suppresses redundant activations during training and enhances feature diversity. Furthermore, Global Response Normalization (GRN) layers are embedded into each module to handle high-dimensional features, effectively boosting model performance. ConvNeXt V2 has delivered outstanding performance across a wide range of downstream tasks.

Nevertheless, ConvNeXt V2 was initially developed for the ImageNet 1K image classification task, which covers 1000 distinct classes and thus requires the model to learn intricate feature representations. By comparison, human sensing tasks involve a far smaller number of classification categories than the ImageNet benchmark. With these factors in mind, this work presents a lightweight LipHS module built upon the ConvNeXt V2 block, which is specially optimized for Channel State Information (CSI) classification in human sensing applications.

As illustrated in Figure 2, the left half displays the original ConvNeXt V2 architecture, into which we integrate PartialConv and SimAM modules. The detailed design of the dedicated downsampling component is also elaborated.

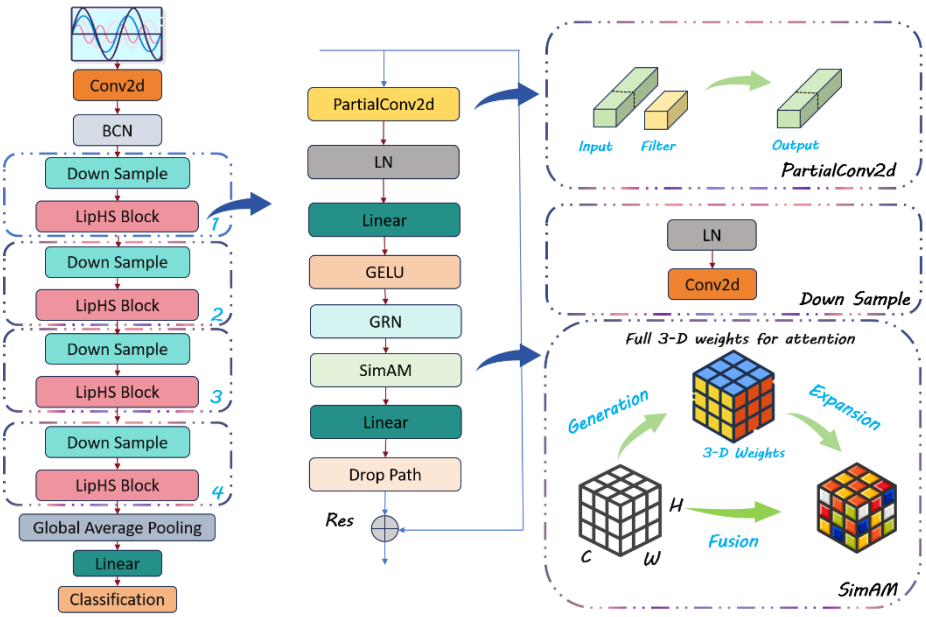

Fig. 2. LipHS framework

PartialConv. While the depthwise convolution (DWConv) in the core module layers of ConvNeXt V2 reduces parameter count and computational overhead, it fails to effectively integrate cross-channel information. To compensate for this, ConvNeXt V2 adopts pointwise convolution (PWConv) after DWConv and expands the channel

dimension to 4× the original size, forming an inverted residual structure. However, this PWConv design introduces significant memory access overhead and degrades overall computational efficiency.To address this issue, we introduce PartialConv to replace the original DWConv [17]. In this implementation, PartialConv conducts convolution solely on the initial quarter of channels in the input feature matrix, where each convolutional kernel is assigned to a single input channel, while the remaining channels are kept unaltered. Combined with the subsequent PWConv layer, this module constructs a T-shaped effective receptive field on the input feature map, which places greater emphasis on the central region of feature representations. This design is well-suited for CSI signal feature extraction, as gesture-discriminative features in CSI maps are primarily localized in the central area. In comparison to the original DWConv-PWConv architecture, PartialConv cuts down total FLOPs while preserving efficient cross-channel information fusion. Notably, the standalone PartialConv module introduces a modest rise in model parameters, a drawback fully compensated by the subsequent LAMP-based channel pruning. After pruning, the proposed LipHS achieves a substantial reduction in both parameter count and FLOPs relative to the original ConvNeXt V2 backbone, completely fulfilling our lightweight design objectives.

SimAM. It is a parameter-free attention module for convolutional neural networks (SimAM), by giving different weights to the functions of different positions so that the network can pay more attention to the non-trivial positions[18]. To this end, we integrate the SimAM attention mechanism into every building block, which enables the model to effectively extract input features and selectively amplify task-relevant information throughout the training process. A distinctive advantage of SimAM lies in its energy function, which can automatically quantify the importance of individual neurons and obtain their corresponding closed-form analytical solutions. The operator of its energy function is defined by the energy function solution, so it does not change the structure of the original network. The formula is

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{5}$$

Where $\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i$; $\hat{\sigma}^2 = \sum_{i=1}^{M}(x_i - \hat{\mu})^2$; the regularization term is represented by λ; t_k is the k neuron of the input feature map on a single channel; $\hat{u}$ and $\hat{\sigma}^2$are the mean and variance of all neurons on a single channel, respectively. By introducing SimAM, the model can adaptively adjust the receptive field at each position to enhance the feature representation and classification capabilities of the model.

4.2 Pruning Based LAMP Scores

We employ model pruning techniques to remove redundant connections to reduce the computational cost and memory footprint of the model while maintaining performance, thereby decreasing the number of model parameters and computational complexity. It does not require an additional large teacher model during training, which makes it particularly suitable for human sensing scenarios where high-performance teacher models are unavailable. We adopt the Layer-Adaptive Magnitude-based Pruning (LAMP) pruning scheme, which effectively reduces redundant parameters while preserving the

fundamental structure of the model [19]. LAMP serves as a global pruning strategy that mitigates the layer collapse issue encountered in prior global pruning techniques. The corresponding score is defined as follows:

$$score(u;W) = \frac{(W[u])^2}{\sum_{v \geq u}(W[v])^2} \tag{6}$$

where u and v represents the index of weights; $W[u]$ represents the weight term mapped by index u. $\sum_{v \geq u}(W[v])^2$ represents the sum of the squared weights of all remaining connections in the same layer. Larger channel weights yield higher LAMP scores, while low-scoring weights are regarded as less important. All channel weights are sorted by their LAMP values, and channels with lower scores are pruned until the preset pruning ratio is met. This results in a more concise network, which is subsequently fine-tuned to recover recognition accuracy.

Fig. 3 shows the channel pruning and fine-tuning process of the LipHS.

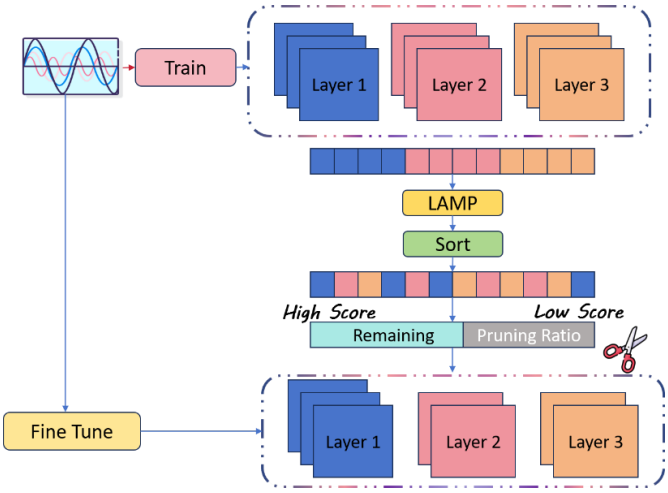

Fig. 3. LipHS pruning process

5 Experiments

5.1 Datasets

This study utilizes the wireless radio frequency signal gesture recognition dataset constructed by Tsinghua University in their Widar 3.0 work[20]. The data collection scenarios cover multiple different environments, making it currently the largest WiFi sensing dataset for gesture recognition. The dataset comprises 12 distinct gesture actions and 10 handwritten Arabic numeral actions, with a total of 43,000 samples. Figure 4 lists the specific gesture categories. While most current studies only select 5-6 actions from one subset, we integrate these two significantly different datasets for combined use, encompassing a total of 22 distinct gesture actions collected from multiple users in different rooms, which sufficiently demonstrates the diversity of categories. It should

be emphasized that the 12-gesture subset and 10-digit subset from Widar 3.0 adopt exactly the same hardware configuration, unified CSI acquisition framework, and consistent preprocessing pipeline.

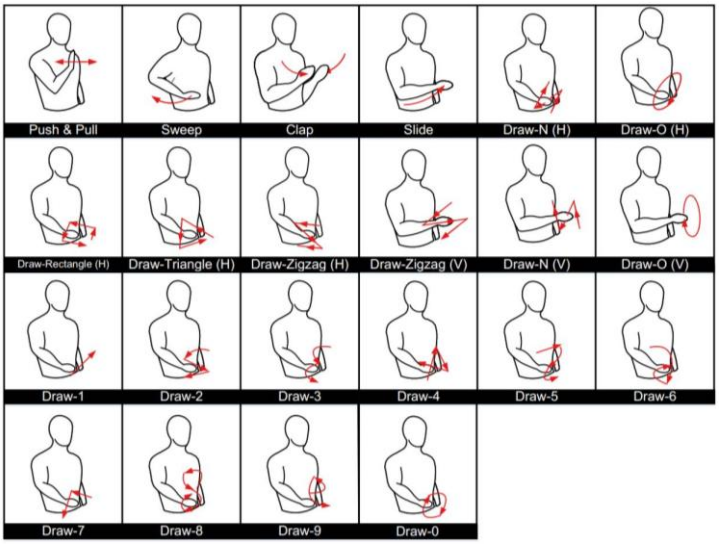

Fig. 4. Gesture categories included in the dataset

5.2 Training parameters

The dataset was divided into training, validation, and test sets using a 6:2:2 ratio. The validation set is used for comparisons among the methods proposed in this paper, while the test set is employed for comparisons with other baseline models. The model was trained using PyTorch on an RTX 4070 GPU with the Adam optimizer. The learning rate was set to 0.001. Each training session was conducted over 20 epochs.

5.3 Performance Comparison

Model pruning validation test The sparsity ratio in model pruning quantifies the sparsification level of parameters in the pruned network, which is generally defined as the percentage of zero or near-zero parameters within the compressed model. A higher sparsity ratio means fewer non-zero weights remain in the model, effectively cutting down memory usage and computational complexity. In resource-limited scenarios like mobile terminals and embedded devices, pruned models with a high sparsity ratio deliver more prominent practical advantages.Table 1 presents the comparative results of the model after pruning and fine-tuning in terms of sparsity rate, FLOPs, Params, and classification accuracy.

This paper designs a comprehensive scoring metric to balance accuracy and efficiency for holistically evaluating the impact of pruning rate. The scoring formula is as follows:

$$Composite - Score_p = \lambda_1 \times Acc_p + \lambda_2 \times FLOPs_p + \lambda_3 \times Params_p \quad (7)$$

$Composite - Score_p$ represents the score under pruning rate p. Acc_p, $FLOPs_p$ and $Params_p$ are expressed as normalized data. The calculation formula corresponding to each pruning ratio is denoted as:

$$Acc_p = \frac{Accuracy_p - Accuracy_{min}}{Accuracy_{max} - Accuracy_{min}} \quad (8)$$

$$FLOPs_p = \frac{FLOPs_{max} - FLOPs_p}{FLOPs_{max} - FLOPs_{min}} \quad (9)$$

$$Params_p = \frac{Params_{max} - Params_p}{Params_{max} - Params_{min}} \quad (10)$$

λ_1, λ_2 and λ_3 are the weights of the corresponding indicators, which are used to balance the significance of accuracy and efficiency indicators. Considering that the model accuracy has a greater impact on the model performance, we set the ratio to 5:1:1.

After computation, when the model pruning ratio is 20%, the model achieves the best. Therefore, we chose to set the pruning ratio to 20%.

Table 1. The performance of different pruning rates

Pruning Ratio(%)	Sparsity rate(%)	FLOPs(MB)	Params(MB)	Acc(%)
0	\	81.77	1.41	74.01
10	20.27	60.82	1.27	73.90
20	36.42	41.44	0.96	73.88
30	50.31	36.73	0.84	72.67
40	62.75	27.45	0.77	66.45
50	73.44	20.18	0.59	49.32
60	81.90	16.32	0.41	22.06
70	88.78	12.81	0.29	11.14

Compared with other advanced models As shown in Table 2, the LipHS model maintains high accuracy both before and after pruning, demonstrating robust classification capability. Compared to the pre-pruned LipHS, the post-pruning model achieves significant compression in FLOPs and Params. The reduction in network layers and channels during the pruning process leads to some loss of feature information, resulting in a slight decrease in model classification performance. Compared to the original ConvNeXt V2-atto, LipHS reduces FLOPs and parameters by 52.75% and 36.42%, while achieving higher detection performance metrics. This indicates that the model successfully eliminates redundant parameters and computational complexity while preserving essential parameters and model structure to maintain representational power and performance. When compared with other models[3,4,5,6,7], LipHS demonstrates substantial advantages in model performance, FLOPs, and Params. The LipHS model significantly reduces computational cost while maintaining high classification performance. Lightweight development boards such as ESP32 can typically accept FLOPs in the range of 0.5G to 2G and can accommodate Params between 0.5MB and 2MB. For

practical edge deployment, our proposed LipHS can meet both the FLOPs and Params requirements while achieving the highest recognition accuracy.

Table 2. Performance comparison with mainstream general-purpose backbones widely used in WiFi sensing research

Model	Acc(%)	FLOPs(MB)	Params(MB)
LipHS	**74.36**	81.77	1.41
LipHS(+Pruning)	73.88	41.44	0.96
ConvNeXt v2-Atto	70.27	87.72	1.51
ViT	67.72	9.28	0.106
ResNet18	71.70	38.39	11.250
ResNet50	68.56	69.70	23.640
ResNet101	68.71	145.87	42.660
CNN	70.19	3.38	0.299
GRU	62.50	1.98	0.091
CNN + GRU	63.19	3.34	0.092
RNN	47.05	0.66	0.031
LSTM	63.35	2.64	0.121
BiLSTM	63.43	5.28	0.240

Ablation Experiments To validate the impact of individual components within LipHS on overall performance, we conducted ablation studies.

As shown in Table 3, after introducing PartialConv into ConvNeXt V2, model accuracy slightly improves and FLOPs decrease, but the Params increases. This is because while the standalone introduction of PartialConv enhances computational efficiency, it also increases Params, and the model still requires further parameter compression. After incorporating the SimAM module into the ConvNeXt V2 block, FLOPs and Params remain largely unchanged, while detection accuracy increases substantially. This indicates that SimAM provides additional effective information and can significantly enhance performance, suggesting that SimAM optimizes feature selection. Through optimization, the proposed LipHS model achieves reduced FLOPs while maintaining detection accuracy higher than the original ConvNeXt V2-Atto model, though with a slight increase in Params.

Table 3. Model Ablation Experiments

Model Block	Acc(%)	FLOPs(MB)	Params(MB)
LipHS	0.7436	81.77	1.41
LipHS(-PartialConv,- SimAM)	0.7027	84.71	1.36
LipHS(-SimAM)	0.7318	81.88	1.40
LipHS(-PartialConv)	0.6982	81.76	1.33

6 Conclusion

In this paper, we propose LipHS, a lightweight feature extraction backbone for WiFi-based human sensing, which is tailored to achieve a balance between fine-grained multi-category gesture classification accuracy and edge deployment feasibility with a compact model size. We construct a lightweight CSI feature extraction network, LipHS, and employ channel pruning based on LAMP scores to trim the network. The pruned LipHS achieves extremely high accuracy with a highly lightweight architecture. Experimental results demonstrate that LipHS outperforms existing state-of-the-art systems and models in terms of accuracy, within the acceptable range of FLOPs and Params for lightweight development boards. LipHS strikes an effective balance between classification performance and model complexity, making it deployable on edge devices with limited computational resources.

In future work, we will further optimize LipHS from the following aspects:

(1) Conduct comprehensive on-device deployment and measurement of latency, memory, and power consumption on typical edge devices such as ESP32, to fully verify the actual deployment performance of LipHS;

(2) Conduct cross-domain evaluation on Widar 3.0's standard cross-domain protocol and more public WiFi sensing datasets, to further verify the generalization performance of LipHS;

(3) Conduct fair experimental comparisons with state-of-the-art domain-specific lightweight WiFi sensing models under a unified benchmark;

(4) Explore user-specific fine-tuning strategies to further improve the recognition accuracy for practical HCI deployment scenarios.

References

1. Zhang, Y., Wu, D., Wang, Y., Zhang, Y., Ji, G., & Ai, J. CSI-GLSTN: A Location-Independent CSI Human Activity Recognition Method Based on Spatio-Temporal and Channel Feature Fusion. IEEE Transactions on Instrumentation and Measurement, 73, 1–12 (2024).
2. Gong, D., Liu, K., Pei, D., Zhang, H., Zhang, S., & Chen, M. Wi-Watch: Wi-Fi-Based Vigilant-Activity Recognition for Ship Bridge Watchkeeping Officers. IEEE Transactions on Instrumentation and Measurement, 73, 1–17 (2024)
3. Zou, H., Zhou, Y., Yang, J., Jiang, H., Xie, L., & Spanos, C. J. DeepSense: Device-Free Human Activity Recognition via Autoencoder Long-Term Recurrent Convolutional Network. 2018 IEEE International Conference on Communications (ICC), 1–6 (2018)
4. Yang, J., Zou, H., Zhou, Y., & Xie, L. Learning Gestures From WiFi: A Siamese Recurrent Convolutional Architecture. IEEE Internet of Things Journal, 6(6), 10763–10772 (2019)
5. Yousefi, S., Narui, H., Dayal, S., Ermon, S., & Valaee, S. A Survey on Behavior Recognition Using WiFi Channel State Information. IEEE Communications Magazine, 55(10), 98–104 (2017)
6. Li, B., Cui, W., Wang, W., Zhang, L., Chen, Z., & Wu, M. Two-Stream Convolution Augmented Transformer for Human Activity Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 35(1), 286–293 (2021)
7. Halperin, D., Hu, W., Sheth, A., & Wetherall, D. Tool release. ACM SIGCOMM Computer Communication Review, 41(1), 53–53 (2011)

8. Xing, T., Yang, Q., Jiang, Z., Fu, X., Wang, J., Wu, C. Q., & Chen, X. WiFine: Real-Time Gesture Recognition Using Wi-Fi with Edge Intelligence. ACM Transactions on Sensor Networks, 19(1), 1–24 (2022)
9. Sheng, B., Li, J., Gui, L., Guo, Z., & Xiao, F. LiteWiSys: A Lightweight System for WiFi-based Dual-task Action Perception. ACM Transactions on Sensor Networks, 20(4), 1–19 (2024)
10. Feng, C., Wang, N., Jiang, Y., Zheng, X., Li, K., Wang, Z., & Chen, X. Wi-Learner. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(3), 1–27 (2022)
11. Shafiqul, I. M., Jannat, M. K. A., Kim, J.-W., Lee, S.-W., & Yang, S.-H. HHI-AttentionNet: An Enhanced Human-Human Interaction Recognition Method Based on a Lightweight Deep Learning Model with Attention Network from CSI. Sensors, 22(16), 6018 (2022)
12. Liu, C., Hao, Y., Liu, Y., Zhang, X., Wang, X., & Liu, Y. LWiHS: A Lightweight Wi-Fi-Enabled Human Sensing Using Feature Fusion Strategy. IEEE Transactions on Instrumentation and Measurement, 74, 1–15 (2025)
13. Qian, K., Wu, C., Yang, Z., Liu, Y., & Jamieson, K. Widar. Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 1–10. ACM (2017)
14. Li, X., Zhang, D., Lv, Q., Xiong, J., Li, S., Zhang, Y., & Mei, H. IndoTrack. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(3), 1–22 (2017)
15. Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. Device-Free Human Activity Recognition Using Commercial WiFi Devices. IEEE Journal on Selected Areas in Communications, 35(5), 1118–1131 (2017)
16. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16133–16142. IEEE (2023)
17. Chen, J., Kao, S., He, H., Zhuo, W., Wen, S., Lee, C.-H., & Chan, S.-H. G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 12021–12031. IEEE (2023)
18. L. Yang, L., Zhang, R.-Y., Li, L., & Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. International Conference on Machine Learning (2021)
19. Lee, J., Park, S., Mo, S., Ahn, S., & Shin, J. Layer-adaptive sparsity for the Magnitude-based Pruning (Version 2). Version 2. arXiv. (2020)
20. Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. Widar3.0: Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1 (2021)