MASS: Empowering Wi-Fi Human Sensing with Metasurface-Assisted Sample Synthesis

Jiaming Gu[1*][0009−0009−7325−4144], Shaonan Chen[2*][0009−0002−5991−3923], Yimiao Sun[1][0000−0001−9384−5915], Yadong Xie[1][0000−0002−2467−4240], Rui Xi[3][0000−0002−6400−9106], Qiang Cheng[2][0000−0002−2442−8357], and Yuan He[1(✉)][0000−0002−6676−4009]

[1] Tsinghua University, Beijing, China
{gujm24,sym21}@mails.tsinghua.edu.cn
{ydxie,heyuan}@tsinghua.edu.cn
[2] Southeast University, Nanjing, China
{220230750,qiangcheng}@seu.edu.cn
[3] University of Electronic Science and Technology of China, Chengdu, China
ruix.ryan@gmail.com

Abstract. Wi-Fi human sensing has attracted numerous research studies over the past decade. The rapid advancement of machine learning technology further boosts the development of Wi-Fi human sensing. However, current Wi-Fi human sensing suffers from the "data scarcity" problem: all the existing proposals require collecting a large amount of human-based datasets to train the sensing models, which is labor-intensive and may raise ethical concerns in certain scenarios. This obstacle seriously restricts the size, quality, and diversity of available datasets, thereby affecting the sensing performance in terms of accuracy and cross-domain applicability. In order to solve this problem, we in this paper propose Metasurface-Assisted Sample Synthesis (MASS), a novel approach to synthesize high-fidelity Wi-Fi sensing samples that effectively capture both the essential features of human motion and environment-specific multipath characteristics without requiring human involvement. The evaluation results show that MASS is effective in boosting the machine learning performance, improving the classification accuracy by 18%, and enhancing the cross-domain sensing accuracy by 22%. These findings underscore the potential of MASS to facilitate the creation of high-quality, diverse datasets with minimal human involvement and associated labor costs.

Keywords: Wi-Fi Sensing · Data Augmentation · Metasurface · Machine Learning · Cross-domain Sensing.

1 Introduction

Benefiting from the ubiquitous Wi-Fi infrastructures, Wi-Fi sensing technologies exhibit versatile abilities to enable diverse applications, including location

---

* Co-primary authors: Jiaming Gu and Shaonan Chen

(a) Data Collection with MASS

(b) Data Collection with Human

Original   Time Scaling   Freq. Stretching   Translation

(c) Transformation-Based Augmentation

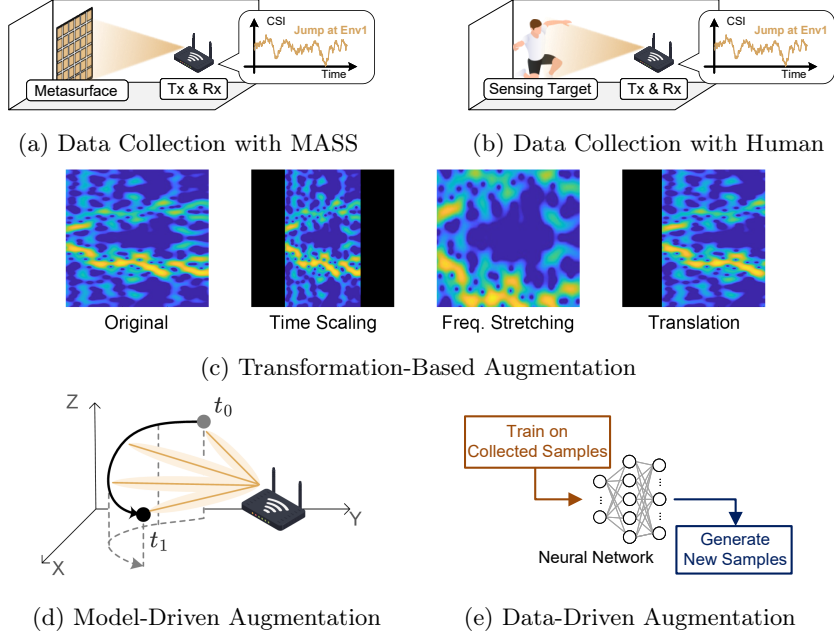(d) Model-Driven Augmentation

(e) Data-Driven Augmentation

Fig. 1: Comparison between Data Collection with MASS, Data Collection with Human, and other Data Augmentation Methods

estimation [15, 34, 37], health monitoring [17], activity recognition [20, 32], *etc.* Recent advancements further release the potential of Wi-Fi sensing by integrating deep learning (DL) models into sensing systems, with which the sensing ability and accuracy can be significantly enhanced [29, 33, 37].

Comprehensive and diverse labeled datasets are necessary for training a robust DL model. When it comes to Wi-Fi human sensing, a sample of sensing data is the coupling result of human motions and the complex multipath propagation of the indoor environment (commonly referred to as the sensing domain). This inherent coupling results in data scarcity through two primary avenues. First, the requirement for human participation makes the process labor-intensive and may raise ethical concerns, especially when collecting health-related samples from patients or individuals with disabilities. Second, the multipath characteristics are a crucial factor in sensing. The usability of data collected in one domain may degrade when it is used to train a DL model for sensing in other domains [4, 35, 37].

Data augmentation is deemed a feasible solution to tackle the data scarcity problem. Some existing works acquire more data by simply applying different transformations on the original signals [36] (*e.g.,* scaling, translating the axis, or adding noise), but are likely to alter the physical properties of signals. Model-driven approaches take into account the signal propagation process [11, 38], but overlook the complex multipath effects, which lead to low robustness of the

DL models. Recent works propose to employ style transfer or generative AI for sample synthesis [5, 6, 19, 31, 35], but still require a large amount of data for training and fall into the "chicken-and-egg" dilemma.

Inspired by the advancements in metasurface research [8, 14, 28, 30], we in this paper propose **M**etasurface-**A**ssisted **S**ample **S**ynthesis (**MASS**) for synthesizing Wi-Fi sensing samples. Fig. 1 compares MASS with human-based data collection and other data augmentation approaches. Leveraging its inherent capacity of waveform manipulation, a metasurface can emulate the effect of human motion on Wi-Fi signals. By replacing the human with the metasurface in the task of collecting sensing samples in real environments, MASS presents a new and much more efficient method to synthesize sensing samples without human participation, significantly reducing labor cost and avoiding ethical concerns.

Our contributions can be summarized as follows:

- We present the theoretical framework underpinning MASS and develop a complete workflow for its realization, elucidating how metasurfaces can be utilized to emulate the effect of human motion on Wi-Fi signals.
- We present the approach of MASS and elaborate on the procedure from collecting an activity template to acquiring metasurface-synthesized samples. This approach significantly reduces labor costs and alleviates ethical concerns while efficiently capturing the intrinsic domain characteristics of the sensing environment.
- We implement MASS and evaluate it under different settings. The results demonstrate that MASS significantly enhances the machine learning accuracy, yielding an 18% accuracy gain, which is 9% higher than the transformation-based augmentation by adding noise. Furthermore, MASS can also increase the accuracy up to 22% in the cross-domain sensing scenarios, where the traditional augmentation only achieves 9% improvement.

The structure of this paper is as follows: Sec. 2 provides a comparison between MASS and related studies. Sec. 3 elaborates on the design of MASS. The evaluation setup, results, and findings are presented in Sec. 4. Sec. 5 discusses the implications of MASS, and finally, Sec. 6 concludes the study.

2   Related Work

2.1   Data Augmentation for Wi-Fi Sensing

Transformation-Based Augmentation: In computer vision, data augmentation is widely used to mitigate data scarcity. Simple transformations are applied to the original images to generate new samples. As shown in Fig. 1c, Dense-LSTM [36] extends this concept to Wi-Fi sensing by adding noise, scaling the time axis, or stretching the frequency axis in spectrograms. Although these techniques improve classification performance, they alter the inherent physical properties of the Wi-Fi signal [11]. For example, stretching the Doppler frequency dimension might incorrectly transform "walking" into "running".
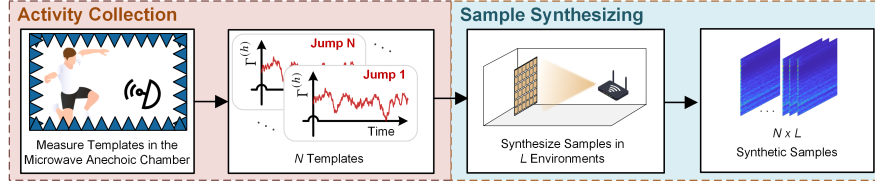
Fig. 2: MASS Overview

Model-Driven Augmentation: To preserve the intrinsic physics of Wi-Fi signals, model-driven data augmentation methods have been proposed [11, 24, 26, 27]. For example, SimHumalator [27] models human limbs as scatter points shown in Fig. 1d and computes their impact on Wi-Fi signals to synthesize samples. However, these studies ignore the modeling of complex environmental multipath effects. For example, interactions such as wall reflections and furniture absorption are often ignored, despite being critical for data diversity [4, 35].

Data-Driven Augmentation: With advancements in deep learning, data-driven augmentation methods have emerged. Some studies [19, 31] employ style transfer for cross-domain adaptation, while others [5, 6] utilize generative AI to create new samples. Despite the potential of these methods, they encounter issues with interpretability, as AI models may not adhere to physical laws [7, 13]. Furthermore, data-driven approaches face the "chicken-and-egg" dilemma, as they require extensive data for training before they can generate new data.

In contrast, MASS is theoretically robust and makes no simplifications to the environment. It synthesizes new Wi-Fi sensing samples that capture human motions and the environment multipath simultaneously without the need for an extensive Wi-Fi sensing dataset.

2.2   Domain-Independent Sensing

Besides data augmentation, some studies aim to extract domain-independent features, thus performing well in cross-domain sensing scenarios [12, 22]. Some other works use transfer learning to adapt models to unseen domains with minimal labeled data [3, 21]. However, DPSense [10] suggests that domain-specific features cannot be entirely removed. Although these works have improved cross-domain sensing performance, they still cannot avoid the labor-intensive sample collection process, while MASS aims to resolve data scarcity at the sample collection stage, making our work orthogonal and complementary to these methods.

3   Design

This section first presents an overview of MASS, then elaborates on its sample synthesis capability using the MASS theory, and finally outlines the procedure of sample synthesis.
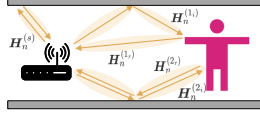
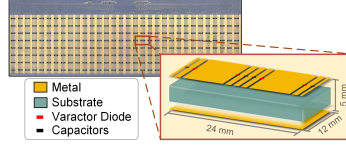Fig. 3: Wi-Fi Sensing Model in Multipath-Rich Indoor Environment

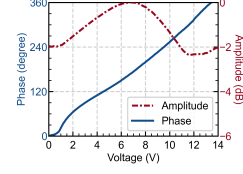Fig. 4: The Metasurface and the Structure of a Meta-Atom

Fig. 5: Reflection Coefficient of the Meta-Atom

3.1 Overview

As shown in Fig. 2, MASS operates in two stages: the activity collection stage and the sample synthesizing stage. In the first stage, activity templates are gathered to capture the impact of human activities on Wi-Fi signals. In the second stage, the metasurface is deployed alongside Wi-Fi sensing equipment across various environments. At this point, the metasurface simulates human movements based on the recorded activity templates. Taking into account the real-world setting in which the process occurs, the samples collected by the Wi-Fi sensing equipment represent a combination of human activity and environmental characteristics. By replacing humans with the metasurface, MASS avoids high labor costs and addresses potential ethical concerns, while still producing substantial sensing samples. For example, a total of $N \times L$ samples can be synthesized from N activity templates and L environments.

3.2 MASS Theory

But how does a metasurface emulate the human activity? This section presents the MASS theory, which first models the environment when either a human or metasurface is present and then methodically translates the problem of emulating human activities to finding the appropriate control voltage sequence, or coding sequence, for the metasurface.

When a human is present in a multipath-rich indoor environment, as illustrated in Fig. 3, the CSI of the n-th Wi-Fi package is modeled as [25, 34]:

$$
\begin{aligned}
\boldsymbol{H}_n(f) &= \boldsymbol{H}_n^{(s)}(f) + \boldsymbol{H}_n^{(d)}(f) \\
&= \boldsymbol{H}_n^{(s)}(f) + \sum_{p=1}^{P_n} \boldsymbol{H}_n^{(p_r)}(f)\varGamma_n^{(h)}(\boldsymbol{q})\boldsymbol{H}_n^{(p_i)}(f),
\end{aligned}
\tag{1}
$$

where $\boldsymbol{H}_n^{(s)}(f)$ is the static component that arises from static objects such as walls and chairs. $\boldsymbol{H}_n^{(d)}(f)$ is the dynamic component from the interaction between the human and the environment, further decomposable into P_n multipath components. For the p-th component, the signal propagates along the path p_i,
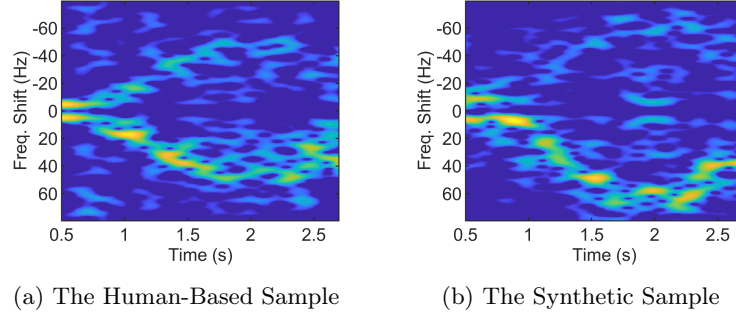
(a) The Human-Based Sample   (b) The Synthetic Sample

Fig. 6: MDS Comparison

and is reflected by a part of the human body simplified as a scatter point at $\boldsymbol{q}$ with reflection coefficient $\varGamma_n^{(h)}(\boldsymbol{q})$. The signal then propagates along the path p_r to the receiver. Each path, p_k, is modeled as $\boldsymbol{H}_n^{(p_k)}(f) = a_{n,p_k}(f) \exp\left(-j\ 2\pi f\ \tau_{n,p_k}\right)$, where $a_{n,p_k}(f)$ is the attenuation and τ_{n,p_k} is the path delay.

When the metasurface, rather than the human, is present in the same environment, the n-th CSI can be expressed as:

$$\boldsymbol{H'}_n(f) = \boldsymbol{H}_n^{(s)}(f) + \sum_{p=1}^{P_n} \boldsymbol{H}_n^{(p_r)}(f)\varGamma_n^{(m)}(\boldsymbol{q}_A)\boldsymbol{H}_n^{(p_i)}(f), \tag{2}$$

where $\varGamma_n^{(m)}(\boldsymbol{q}_A)$ denotes the reflection coefficient of the meta-atom at $\boldsymbol{q}_A$. $\varGamma_n^{(m)}(\boldsymbol{q}_A)$ is determined by the metasurface design and the voltage $V_n(\boldsymbol{q}_A)$ applied to the meta-atom. As depicted in Fig. 4 and Fig. 5, our metasurface comprises 16×8 meta-atoms, each capable of controlling phase reflection from 0 to 360 degrees, while keeping the amplitude almost unchanged.

Eq. 2 diverges from Eq. 1 solely by substituting $\varGamma_n^{(h)}(\boldsymbol{q})$ with $\varGamma_n^{(m)}(\boldsymbol{q}_A)$. Thus, the human activity is replicated by equating $\varGamma_n^{(m)}(\boldsymbol{q}_A)$ to $\varGamma_n^{(h)}(\boldsymbol{q})$. To emulate an **activity template** $\{\varGamma_n^{(h)}(\boldsymbol{q})\}$, an accurate **coding sequence** $\{V_n(\boldsymbol{q}_A)\}$ is required so that the impacts of metasurface, $\varGamma_n^{(m)}(\boldsymbol{q}_A)$, are aligned with the impacts of the human body, $\varGamma_n^{(h)}(\boldsymbol{q})$.

3.3   Sample Synthesis Procedure

Building on the MASS theory, the sample synthesis process unfolds as follows:

Activity Template Collection: activity templates are collected in a microwave anechoic chamber. In the chamber, multipath effects are minimized and the impacts of human activities on Wi-Fi signals are isolated. A vector signal analyzer [2] is used to transmit a single-frequency sinusoidal wave, a volunteer performs the activity in the chamber, and the echo signal is collected as the activity template. In practice, since the human sensing necessitates a space where
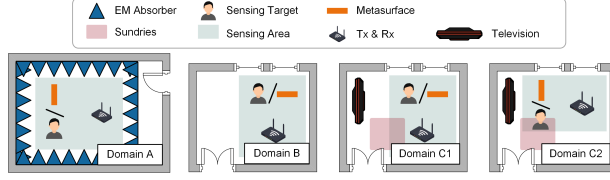
Fig. 7: Domain Layouts

Fig. 8: The Scenario of Domain C2

the volunteer can move freely, we modified an office space and surrounded the sensing area with absorbing materials to reduce reflections, as shown in Fig. 7, Domain A. During the measurement of activity templates, the antennas are placed on a 1.5-meter-high stand, directly facing the human torso.

Coding Sequence Derivation: the coding sequence is then derived from the activity template. Each activity template is filtered by a bandpass filter to remove the direct current component and focuses on the human motion. The control voltage series is then determined by referring to the phase extracted from the filtered template according to the relationship depicted in Fig. 5.

Sample Synthesis: Upon obtaining the coding sequences, the metasurface is prepared to emulate human activities. The metasurface and Wi-Fi devices are deployed in a real-world environment, with coding sequences applied to the metasurface. Consequently, Wi-Fi devices capture the CSI as synthesized samples.

As illustrated in Fig. 6, a comparison of the micro-Doppler spectrum (MDS) for the "running approach" activity reveals strong similarities between the human-based sensing sample and the synthesized sample. Both spectra demonstrate a clear sequence: initiation of running around 1 second, speed maintenance until approximately 2 seconds, and deceleration around 2.6 seconds. This demonstrates MASS's ability to capture the dynamic characteristics of human activities. We acknowledge slight discrepancies between the two spectra, and attribute it to the inherent, non-deterministic nature of human movement.

It is important to note that our method focuses on simulating the entire activity as a whole, rather than simulating specific, isolated time slots. Fig. 6 serves as an intuitive demonstration of the overall similarity for all activities. For quantified similarity results and a more rigorous analysis, readers are referred to Sec. 4.

4   Evaluation

4.1   Experiment Setup

Metric: The evaluation of MASS is conducted through end-to-end experiments, with activity recognition accuracy serving as the metric. LeNet [16] is employed as the classifier. Before classification, CSI preprocessing [18, 22] is performed to correct the phase error and filter out noise.
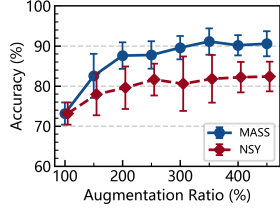
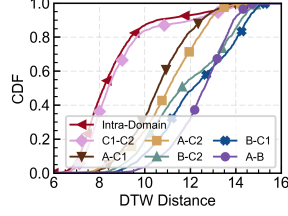Fig. 9: Classification Accuracy across Augmentation Ratios
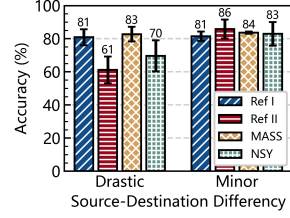

Fig. 10: DTW Distance between Domains

Fig. 11: Classification Accuracy for Different Cross-Domain Sensing Groups and Methods

Experiment Environments: Sensing samples are collected in four distinct environments, as depicted in Fig. 7. Domain A, the microwave anechoic chamber, serves both as the activity template collection site and a sensing sample collection site. Domains B and C emulate typical indoor environments using the same room, with Domain B being more spacious and Domain C containing more miscellaneous items. The key difference between Domain C1 and C2 is the positioning of the sensing target and the Wi-Fi equipment, and Fig. 8 shows the scenario of Domain C2.

Sensing Equipment: PicoScenes is selected as the Wi-Fi sensing platform, which is compliant with the standard Wi-Fi protocol [1]. For data collection, PicoScenes is configured to transmit packets at 800 Hz over a duration of 3 seconds, resulting in each sample comprising 2400 packets.

Collected Dataset: Four volunteers of varying heights and weights participate in performing seven common daily actions within the specified domains. These actions include: Walk Approach (WA), Walk Away (WW), Run Approach (RA), Run Away (RW), Sit Down (SD), Stand Up (SU), and Jump (JMP). Each action is repeated multiple times to ensure the diversity of the dataset. After filtering out corrupted samples (*e.g.*, instances where a volunteer executed the wrong action, or when accidental interference occurred), the dataset comprises 140 activity templates, 659 human-based samples, and 1520 synthesized samples.

4.2   Accuracy Improvement with MASS

This section verifies the effectiveness of MASS in improving activity recognition accuracy by augmenting human-based sensing samples with synthetic samples at different augmentation ratios. The augmentation ratio is defined as the ratio of the training set size to the number of human-based samples in that set. The baseline method, which involves duplicating real samples and adding Gaussian noise, is referred to as NSY. The results are shown in Fig. 9.

Given the inherent instability of AI training, we utilize a **repeated k-fold cross-evaluation** procedure [9]. All human-based samples are partitioned into k folds. For each fold and augmentation ratio, M different subsets of synthetic

samples are merged with $k - 1$ folds to train the model, and the remaining fold is used for testing. It ensures that each fold is tested M times with distinct sets of synthetic samples. The mean accuracy and standard deviation provide a more reliable metric.

At 100% augmentation, representing no augmentation, both methods achieve the identical accuracy of 73% with only human-based samples. As the augmentation ratio increases, noticeable accuracy improvements are observed for both methods. At 350% augmentation, MASS achieves an accuracy of 91%, which is 18% higher than no augmentation and 9% higher than NSY at its maximum accuracy of 82% at 450% augmentation. Importantly, MASS consistently outperforms NSY across all augmentation ratios, demonstrating higher accuracy and more stability as indicated by tighter standard deviations.

These outcomes suggest that MASS effectively synthesizes samples that capture essential features of human motion, offering more substantial benefits for training compared to simple noise perturbation.

4.3   Cross-Domain Sensing Performance

This section evaluates the capability of MASS to characterize varying sensing environments and facilitate cross-domain generalization. In cross-domain trials, the model is trained in one domain (source) and tested in another (destination), resulting in 12 distinct combinations derived from our four-domain dataset. For convenience, we denote training on domain X and testing on domain Y as $X \rightarrow Y$.

To illustrate the differences between domains, Fig. 10 depicts the cumulative distribution function (CDF) of Dynamic Time Warping (DTW) [23] distances for the same activity across domains, along with intra-domain DTW distances labeled "Intra-Domain" for comparison. The figure reveals that inter-domain distances substantially exceed intra-domain distances. Specially, $C1 - C2$ has a mean DTW distance of 8.92, close to the intra-domain mean distance of 8.61, indicating minimal variance due to minor differences in sensing positions and directions. In contrast, other inter-domain distances are considerably larger, signaling significant changes in multipath characteristics. Notably, room size and layout do not solely determine domain differences. $A - C1$ and $A - C2$ exhibit smaller DTW distances than $B - C1$ and $B - C2$.

Based on these observations, we categorized the 12 cross-sensing trials into two groups: (1) **Minor Change**: domains with similar multipath characteristics (*i.e.,* $C1 \rightarrow C2$ and $C2 \rightarrow C1$). (2) **Drastic Change**: domains with significantly different multipath characteristics (*i.e.,* $A \rightarrow C1$, $A \rightarrow C2$, $B \rightarrow C1$, $B \rightarrow C2$, $A \rightarrow B$, and their reverse directions). These groups reflect different levels of difficulty in cross-domain tasks. The drastic change group is more challenging, whereas the minor change group poses less difficulty.

During each trial, synthetic samples from the destination domain are added to enhance the training set, allowing the model to learn domain-specific features. We compare MASS with multiple reference and baseline approaches. Ref I represents an ideal, albeit unrealistic scenario where the training set includes samples

from both the source and destination domains. Ref II trains solely on the source domain, reflecting real-world cross-domain sensing scenarios. NSY serves as the baseline by enhancing the source domain training set with Gaussian noise, a common practice in traditional data augmentation.

The results are summarized in Fig. 11. In the drastic change group, MASS gives an average accuracy of 83%, outperforming Ref II by 22% and NSY by 13%. In the minor change group, MASS achieves an accuracy of 84%, which is comparable to Ref I and Ref II. These results suggest that synthetic samples effectively capture domain-specific multipath characteristics, thereby validating the feasibility of MASS as a solution for cross-domain sensing.

5   Discussion

While the theory behind MASS is promising, its practical implementation presents several challenges. As noted in Sec. 3, the method for template collection relies on human-based measurements, which limits the diversity of activity templates and, consequently, causes accuracy improvements to reach a plateau. We anticipate that in the future, activity templates will be collected from a broader range of individuals and activities. Furthermore, since these templates are exclusively related to human motion, MASS could potentially be expanded to include simulations for template collection. Additionally, the manufacturing of metasurfaces remains costly, and their operation requires specialized equipment. Despite these challenges, the evaluation in Sec. 4 demonstrates that MASS remains effective in synthesizing sensing samples. We plan to extend future verifications of MASS to encompass more activities and environments.

6   Conclusion

This study introduces MASS, a novel approach to synthesizing Wi-Fi sensing samples with the metasurfaces. MASS harnesses the unique properties of metasurfaces to synthesize samples that effectively capture both the key features of human motion and the characteristics of the sensing environment. We demonstrate that MASS can significantly enhance classification accuracy and improve the generalizability, which underscores the potential of metasurfaces to empower Wi-Fi sensing by synthesizing substantial amounts of high-fidelity data at minimal cost and human involvement.

7   Acknowledgment

References

1. Picoscenes: Enabling modern wi-fi isac research!, https://ps.zpj.io/
2. pxie-5841, https://www.ni.com/en-us/shop/model/pxie-5841.html
3. Bu, Q., Yang, G., Ming, X., Zhang, T., Feng, J., Zhang, J.: Deep transfer learning for gesture recognition with wifi signals. Personal and Ubiquitous Computing pp. 1–12 (2022)
4. Chen, C., Zhou, G., Lin, Y.: Cross-domain wifi sensing with channel state information: A survey. ACM Computing Surveys **55**(11), 231:1–231:37 (2023)
5. Chen, X., Zhang, X.: Rf genesis: Zero-shot generalization of mmwave sensing through simulation-based data synthesis and generative diffusion models. In: Proceedings of the ACM SenSys. pp. 28–42 (2023)
6. Chi, G., Yang, Z., Wu, C., Xu, J., Gao, Y., Liu, Y., Han, T.X.: Rf-diffusion: Radio signal generation via time-frequency diffusion. In: Proceedings of the ACM MobiCom. pp. 77–92 (2024)
7. Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F.: Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing **92**(3), 88 (2022)
8. Feng, C., Li, X., Zhang, Y., Wang, X., Chang, L., Wang, F., Zhang, X., Chen, X.: Rflens: Metasurface-enabled beamforming for iot communication and sensing. In: Proceedings of the ACM MobiCom. pp. 587–600 (2021)
9. Fushiki, T.: Estimation of prediction error by using k-fold cross-validation. Statistics and Computing **21**, 137–146 (2011)
10. Gao, R., Li, W., Xie, Y., Yi, E., Wang, L., Wu, D., Zhang, D.: Towards robust gesture recognition by characterizing the sensing quality of wifi signals **6**, 11:1–11:26 (2022)
11. Hou, W., Wu, C.: Rfboost: Understanding and boosting deep wifi sensing via physical data augmentation. ACM IMWUT **8**(2), 1–26 (2024)
12. Jiang, W., Miao, C., Ma, F., Yao, S., Wang, Y., Yuan, Y., Xue, H., Song, C., Ma, X., Koutsonikolas, D., Xu, W., Su, L.: Towards environment independent device free human activity recognition. ACM IMWUT pp. 289–304 (2018)
13. Kang, B., Yue, Y., Lu, R., Lin, Z., Zhao, Y., Wang, K., Huang, G., Feng, J.: How far is video generation from world model: A physical law perspective. arXiv:2411.02385 (2024)
14. Khalily, M., Yurduseven, O., Cui, T.J., Hao, Y., Eleftheriades, G.V.: Engineered electromagnetic metasurfaces in wireless communications: Applications, research frontiers and future directions. IEEE Communications Magazine pp. 88–94 (2022)
15. Kotaru, M., Joshi, K., Bharadia, D., Katti, S.: Spotfi: Decimeter level localization using wifi. In: Proceedings of the ACM SIGCOMM. pp. 269–282 (2015)
16. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Handwritten digit recognition with a back-propagation network. In: Proceedings of the NIPS. vol. 2 (1989)
17. Lee, S., Park, Y.D., Suh, Y.J., Jeon, S.: Design and implementation of monitoring system for breathing and heart rate pattern using wifi signals. In: Proceedings of the IEEE CCNC. pp. 1–7 (2018)
18. Li, X., Zhang, D., Lv, Q., Xiong, J., Li, S., Zhang, Y., Mei, H.: Indotrack: Device-free indoor human tracking with commodity wi-fi. ACM IMWUT **1**(3), 1–22 (Sep 2017)
19. Li, X., Chang, L., Song, F., Wang, J., Chen, X., Tang, Z., Wang, Z.: Crossgr: Accurate and low-cost cross-target gesture recognition using wi-fi. ACM IMWUT **5**(1), 21:1–21:23 (Mar 2021)

20. Liu, J., Liu, H., Chen, Y., Wang, Y., Wang, C.: Wireless sensing for human activity: A survey. IEEE Communications Surveys & Tutorials **22**(3), 1629–1645 (2019)
21. Pan, S.J., Zheng, V.W., Yang, Q., Hu, D.H.: Transfer learning for wifi-based indoor localization. In: Proceedings of the AAAI Workshop. vol. 6 (2008)
22. Qian, K., Wu, C., Zhang, Y., Zhang, G., Yang, Z., Liu, Y.: Widar2.0: Passive human tracking with a single wi-fi link. In: Proceedings of the ACM MobiSys. pp. 350–361 (2018)
23. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing **26**(1), 43–49 (1978)
24. Su, J., Mao, Q., Liao, Z., Sheng, Z., Huang, C., Zhang, X.: A real-time cross-domain wi-fi-based gesture recognition system for digital twins. IEEE Journal on Selected Areas in Communications **41**(11) (2023)
25. Tse, D., Viswanath, P.: Fundamentals of wireless communication. Cambridge University Press (2005)
26. Virmani, A., Shahzad, M.: Position and orientation agnostic gesture recognition using wifi. In: Proceedings of the ACM MobiSys. pp. 252–264 (2017)
27. Vishwakarma, S., Li, W., Tang, C., Woodbridge, K., Adve, R., Chetty, K.: Simhumalator: An open-source end-to-end radar simulator for human activity recognition. IEEE Aerospace and Electronic Systems Magazine **37**(3), 6–22 (Mar 2022)
28. Wang, S.R., Dai, J.Y., Ke, J.C., Chen, Z.Y., Zhou, Q.Y., Qi, Z.J., Lu, Y.J., Huang, Y., Sun, M.K., Cheng, Q., Cui, T.J.: Radar micro-doppler signature generation based on time-domain digital coding metasurface. Advanced Science p. 2306850 (2024)
29. Wu, D., Zhang, D., Xu, C., Wang, H., Li, X.: Device-free wifi human sensing: From pattern-based to model-based approaches. IEEE Communications Magazine **55**(10), 91–97 (Oct 2017)
30. Wu, G.B., Dai, J.Y., Shum, K.M., Chan, K.F., Cheng, Q., Cui, T.J., Chan, C.H.: A universal metasurface antenna to manipulate all fundamental characteristics of electromagnetic waves. Nature Communications **14**(1) (2023)
31. Xiao, C., Han, D., Ma, Y., Qin, Z.: Csigan: Robust channel state information-based activity recognition with gans. IEEE Internet of Things Journal **6**(6), 10191–10204 (Dec 2019)
32. Yan, H., Zhang, Y., Wang, Y., Xu, K.: Wiact: A passive wifi-based human activity recognition system. IEEE Sensors Journal **20**(1), 296–305 (2019)
33. Yang, J., Chen, X., Zou, H., Lu, C.X., Wang, D., Sun, S., Xie, L.: Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing. Patterns **4**(3), 100703 (Mar 2023)
34. Yang, Z., Zhou, Z., Liu, Y.: From rssi to csi: Indoor localization via channel response. ACM Computing Surveys **46**(2), 25:1–25:32 (Dec 2013)
35. Zhang, J., Tang, Z., Li, M., Fang, D., Nurmi, P., Wang, Z.: Crosssense: Towards cross-site and large-scale wifi sensing. In: Proceedings of the ACM MobiCom. pp. 305–320 (2018)
36. Zhang, J., Wu, F., Wei, B., Zhang, Q., Huang, H., Shah, S.W., Cheng, J.: Data augmentation and dense-lstm for human activity recognition using wifi signal. IEEE Internet of Things Journal **8**(6), 4628–4641 (Mar 2021)
37. Zheng, Y., Zhang, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., Yang, Z.: Zero-effort cross-domain gesture recognition with wi-fi. In: Proceedings of the ACM MobiSys. pp. 313–325 (2019)
38. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI CAI. vol. 34, pp. 13001–13008 (2020)