WiSR: Sparse Recovery for Wi-Fi Signal via Generative Adversarial Network

Mingzhu Yang, Dawei Yan, Yubo Yan,
School of Computer Science and Technology, University of Science and Technology of China, China
mzyang@mail.ustc.edu.cn, yandw@ustc.edu.cn, yuboyan@mail.ustc.edu.cn

***Abstract*— Recently, Wi-Fi based sensing technology has been widely studied to provide more convenient services for humans. Although previous arts claim to achieve diverse fine-grained sensing using Wi-Fi signals, most of them assume that the data such as *Channel State Information* (CSI) used to achieve the sensing tasks can be sufficiently collected. However, in practical, due to the competitive nature of Wi-Fi and the frequent intermittent traffic, Wi-Fi based sensing applications often encounter problems of irregular intervals and insufficient sampling. Therefore, in this paper, we propose a Wi-Fi signal sparse recovery system (WiSR) that aims to recover sufficient and uniform sensing data from unevenly spaced and under-sampled CSI. Inspired by the success of image and audio restoration, we improve the *Generative Adversarial Network* (GAN) to recover Wi-Fi CSI. However, the direct application of GAN technologies for image and audio to CSI is not effective due to the difference in data representation. First, to avoid spectral impairments after conversion from time domain to frequency domain, we directly operate on the original time series CSI waveforms, thus being able to recover continuous channel variations from intermittent sparse samples. Second, to enhance the above recovery process, we utilize two novel denoising methods to obtain clean CSI, and introduce restrictions in the time and frequency domains to optimize low-level features and high-frequency information, respectively. Real-world experiments show that WiSR can accurately recover CSI, even at a rate of 10 packets per second. Through practical applications of gait recognition and gesture recognition, WiSR significantly improves accuracy compared to traditional linear interpolation and cubic interpolation.**

***Index Terms*—Channel State Information, Signal Processing, Generative Adversarial Network, Wi-Fi Sensing**

I. Introduction

Wi-Fi infrastructure has been widely deployed indoors for internet connectivity, establishing it as one of the most pivotal wireless technologies. Through the analysis of wireless signal patterns, Wi-Fi has demonstrated considerable potential for device-free human sensing tasks, including activity recognition [1], [2], gesture recognition [3]–[6], and gait recognition [7]–[9]. Despite the unique advantages brought over by Wi-Fi sensing, such as low cost, ubiquitous infrastructure, and non-intrusiveness, a number of problems and challenges are yet to be solved to make the sensing applications commercially viable and ubiquitous. In this paper, we focus on two issues that affect the reliability and usability of Channel State Information (CSI) for high-performance Wi-Fi sensing:

- **Low Sampling.** CSI can only be estimated during data communication. Consequently, when data communication is inactive, the sampling rate of CSI significantly decreases, rendering it insufficient for higher-level applications.
- **Irregular Intervals.** The contention-based multi-access characteristic of Wi-Fi often leads to significant variations in the time intervals between adjacent frames. This results in an irregular frame arrival rate for each link, as illustrated in Fig.1. Furthermore, data caching and rate control by upper-layer protocols intensify this irregularity.

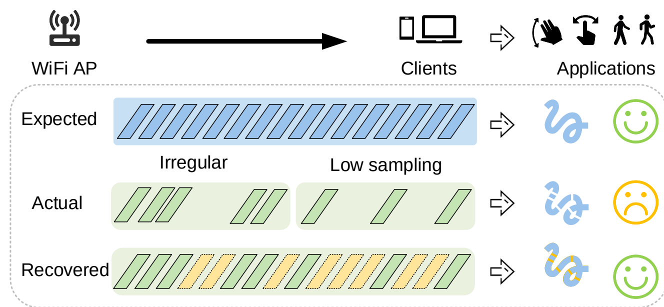


Fig. 1. **In practical, Wi-Fi based sensing suffers from irregular intervals and inadequate sampling which significantly degrades performance, our scheme can recover sufficient and uniform CSI, thus maintaining accuracy for a variety of applications.**

CSI contained in each frame is vital for Wi-Fi sensing. An inconsistent frame rate translates to a fluctuating sampling rate, potentially limiting the effectiveness of Wi-Fi sensing. It's also crucial to acknowledge that CSI data is high-dimensional. Maintaining a high frame rate for real-time applications can place considerable strain on communication, potentially disrupting the normal functioning of Wi-Fi.

Interpolation is a straightforward technique that can address low and irregular packet rates [10]. Nevertheless, this method only fills data gaps using local information to ensure a uniform distribution in the time domain. It may not be effective when parts of the signal are missing. Currently, Generative Adversarial Networks (GANs) have gained popularity in tasks such as image repair and super-resolution [11], [12]. GANs utilize a learned data distribution to fill in missing pixels. Previous studies [13]–[15] have primarily utilized the data generation capabilities of GANs to synthesize virtual samples in target domains, facilitating domain adaptation. In this paper, we introduce *WiSR*, a method that employs global information to restore CSI data. However, several challenges must be overcome to successfully implement this system and achieve accurate and reliable CSI recovery.

The first challenge is how to effectively transform the WiFi signal data for efficient deep learning processing. The signal data captured by WiFi equipment differs significantly from the image data. Most works [7], [8] converted the radio frequency (RF) signal data from time domain to frequency domain and fed the frequency spectrum image into the machine learning models for processing. However, crucial sensing information is lost in converting a signal from the time domain to the frequency domain. So we introduce a new generative model operating directly on the raw CSI waveform to recover continuous channel variation from the intermittently sparse samples.

In contrast to image and audio, noise comes in many forms in the raw CSI that a commercial WiFi device collects. The presence of noise can cause signal distortion, decrease the signal-to-noise ratio, and impair the reliability of the signal. In the process of signal recovery, noise is treated as part of the signal itself, resulting in errors between the recovered signal and the original signal. Therefore, we apply two noise mitigation techniques to minimize the effects of noise and enhance the quality and reliability of the recovered signal.

Moreover, signal inpainting with significant gaps presents a formidable challenge. In the context of CSI traffic, such gaps can result in substantial information loss, rendering algorithms designed for shorter intervals ineffective. The solution to this problem involves exploring the generation of missing frequencies within the sparse part. We propose a novel loss function in the frequency domain, achieved by calculating the frequency component using the Short Fast Fourier Transform (STFT). This new loss function directly references the ground truth in both the temporal and frequency domains, thereby enhancing the precision of time-series sampling and providing a balanced focus on the spectral details of the CSI data.

Our key contributions can be summarized as follows:

- To avoid the loss of Wi-Fi signals in converting a signal from the time domain to the frequency domain, we introduce a new recovering model operating directly on the raw CSI waveform to recover continuous channel variation from the intermittently sparse samples.
- Moreover, by employing two novel denoising methods and applying loss function both on temporal and frequency domains, we excels in recovering high-quality time-series signals.
- We implemented *WiSR* using the Gait dataset collected by ourselves and the public Widar3.0 [5]. We evaluated the system's performance through extensive experiments. A comparison with two interpolation methods revealed *WiSR*'s superior performance in recovering Wi-Fi signals under varying packet rates.

The remainder of the paper is organized as follows: Section II introduces the related works. Section III details the sparse system overview. The denoising strategy and sparse recovering network are introduced in Section IV and Section V, respectively. The experiments are provided in Section VI. Finally, we conclude our work in Section VII.

II. Related Work

In this section, we discuss the related work in large-scale Wi-Fi sensing systems: GAN-based WiFi sensing techniques and existing sparse recovery algorithm for WiFi-Sensing.

A. GAN-based WiFi Sensing Techniques

To the best of our knowledge, previous works have primarily used the GAN model's data generation ability to synthesize virtual samples in target domains and facilitate domain adaptation. CsiGAN [13] proposed a semi-supervised GAN network for producing complement fake data samples and improving activity recognition accuracy when performed by new users. WiGAN [14] aimed to increase both WiFi data capacity and diversity. DMNet [16] and GNG [15] also used GANs to generate more samples and improve activity recognition.

B. Sparse Recovery for WiFi Sensing

SenCom [10] uses a fitting resampling scheme to solve the problem of uneven Wi-Fi transmission and injects active detection packets to improve sampling when Wi-Fi traffic is insufficient. Muse-Fi [17] proposes a sparse recovery algorithm applying to the spectrum of the signal, which divides the signal into sparse and non-sparse parts and uses an autoencoder (AE) based on Time Convolutional Network (TCN) to recover the sparse parts. WiImg [18] converts the CSI samples into images and improves the GANs for CSI image inpainting, relaxing the requirement of a high sample rate in sensing.

III. System Overview

In this paper, we introduce a new generative model operating directly on the raw CSI waveform to recover continuous channel variation from the intermittently sparse samples due to realistic traffic. Fig.2 illustrates the overall structure of the proposed system.

Specifically, we first apply *principal component analysis (PCA)* along sub-carrier dimension and AGC removal to minimize the effects of noise and enhance the quality and reliability of the recovered signal (Section IV-A and Section IV-B). Then, To keep a balance between computational resources and contextual information, original time series CSI data is divided into 2-second fragments. And each signal sequence from the dataset is either interpolated or downsampled to a resampling frequency f_{target} (Section IV-C). After denosing and resampling, CSI data is inputted into a sparse recovery model to generate the sparse component (Section V). This output, after recovering, can be further processed (*e.g.* time-frequency analysis) to recognize the gait or gesture of the subject.

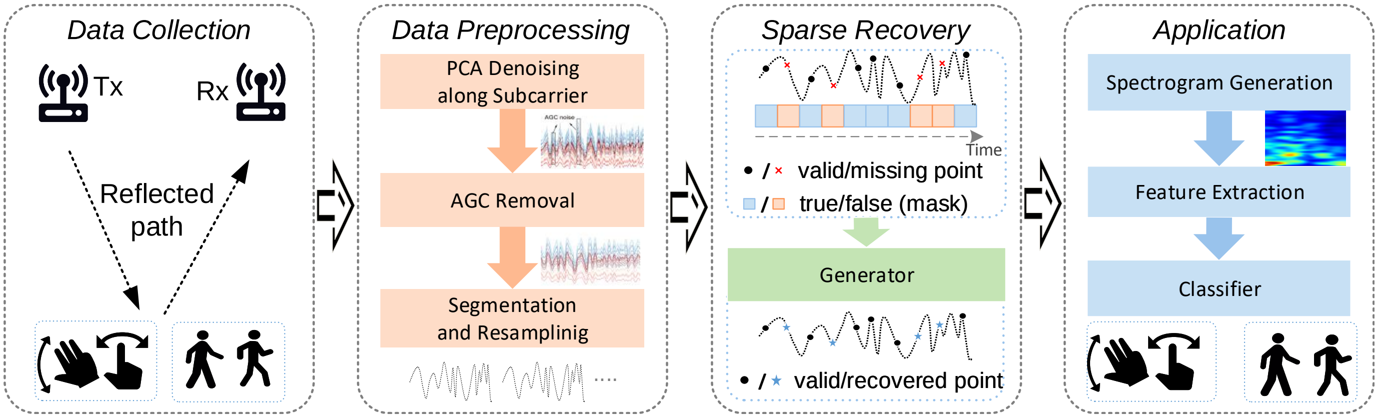
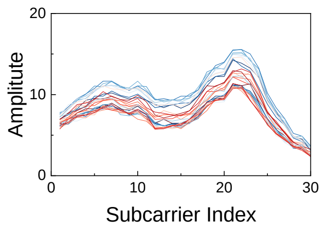
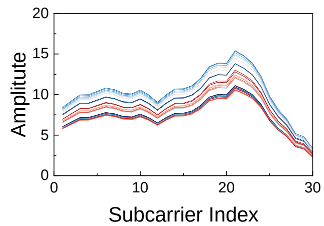


Fig. 2. **Overview of *WiSR*: consist of data collection, data preprocessing, sparse recovery and applications such as gesture and gait recognition.**

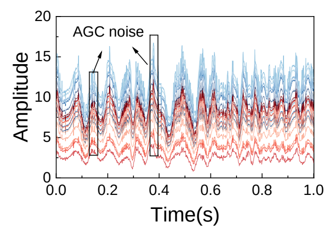


(a) CSI in sub-carrier dimension.

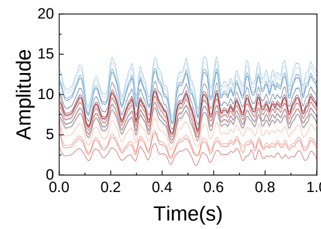


(b) CSI after PCA denoising.

Fig. 3. **Illustration of PCA denoising along sub-carrier dimension.**



(a) Illustration of AGC noise.



(b) CSI after AGC removal.

Fig. 4. **Illustration of AGC noise and removal.**

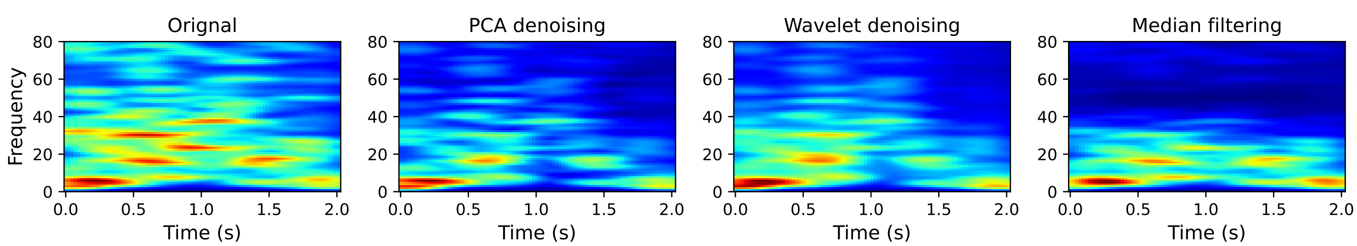


Fig. 5. **Comparison between different denoising methods in frequency domain.** From left to right: the original signal, processing signal after PCA denoising, wavelet denoising, and median filtering, respectively.

IV. Data Preprocessing

A. PCA Denoising along Sub-carrier Dimension

When the target is far away from the transceiver or the walking direction is close to parallel to the propagation path, the signal pattern may be overwhelmed by the noise. We observe that using PCA denoising at the subcarrier dimension and extracting the first principal component can better remove environmental noise while preserving the detailed features of the signal as much as possible. So we transpose the CSI matrix to represent the reflection paths of all sub-carriers in the environment over a period of time. We keep the first principal component as motion related and treat the others as unrelated components representing the ambient noise. As a result, we can remove the ambient noise by selecting the eigenvector corresponding to the maximum variance to reconstruct the signals.

To demonstrate the effectiveness of the PCA denoising algorithm along the sub-carrier dimension, we plotted the effects of different denoising methods on the signal, including time and frequency domains. The results are shown in Fig.5. In the time-frequency spectrum of PCA denoising and wavelet threshold denoising, the noise regions are significantly attenuated, and the details of the signal are well-preserved. Median filtering reduces the noise to some extent, but the preservation of signal details is not as effective as PCA and wavelet denoising, especially for high frequencies. Compared to wavelet threshold denoising, PCA denoising allows for clear separation between the signal and noise areas in spectrum, which is important to the restoration of frequency domain characteristics of signals.

B. AGC Removal

After removing the environmental noise, we also need to remove the influence of AGC noise. Previous schemes using

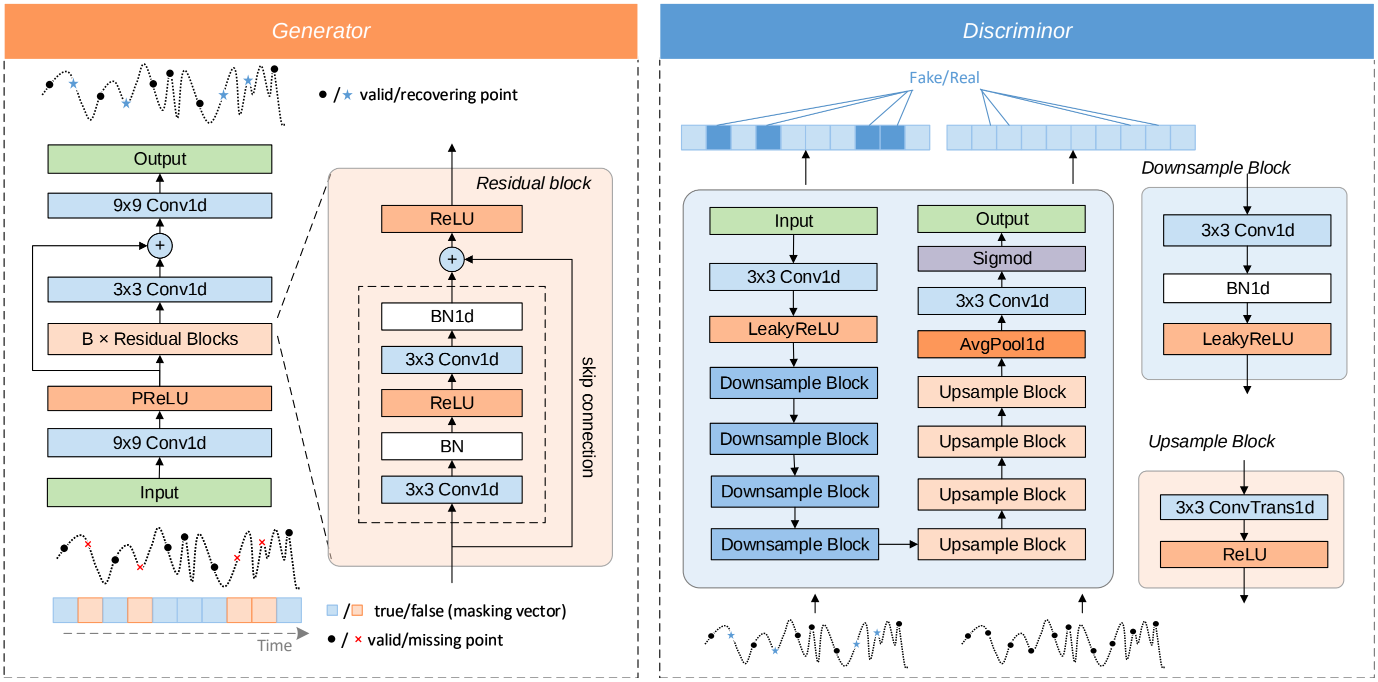


Fig. 6. **Architecture of generator and discriminator network for signal recovery.**

the ratio of two antennas can effectively remove AGC noise [19]. In this paper, to completely preserve the distribution of CSI data, we do not use the ratio method. According to our observation, the noise caused by AGC is uncertain and sparse points around the dynamic signal in the time dimension [20], so we can filter out the noise caused by AGC according to the sparse density distribution. Specifically, we utilize the DBSCAN spatial clustering method to cluster AGC-related points according to their sparse distribution density and retain only the cluster with the largest number of scatter points [21], and the results are shown in Fig.4.

C. Data Segmentation and Resampling

Our recovering models directly take masked raw waveforms as input and generate recovered waveforms as outputs. Longer sequences provide more contextual information, which helps the model capture longer temporal dependencies. However, it increases computational demands, including memory and computation time, which impact training and inference efficiency. Additionally, longer input data can present challenges in gradient propagation and optimization. Hence, we slice the original time series CSI data into 2-second fragments, striking a balance between input data length, computational resources, and model requirements.

Each signal sequence from the dataset is either interpolated or down sampled to a target frequency f_{target}, ensuring uniformity in the model's input length. The frequency f_{target} is empirically specified in different applications. For gesture and gait recognition, we set f_{target} to 1000Hz, which is commonly used in previous work [3], [4], [7].

V. Sparse Recovery

In this paper, we propose a GAN-based model operating directly on the raw CSI waveform to recover continuous channel variation Y_t from the intermittently sparse samples X_t due to realistic traffic.

We take an input time series $\{X_t \mid t = 1 \ldots n\}$ drawn from the CSI of specific antenna pair with a mask $\{M_t \mid t = i \ldots j, 1 < i < j < n\}$ as input, where t is the sampling time and X_t is the amplitude of a corresponding CSI sample. In order to utilize spatial information, we use multiple subcarrier data as channel inputs. The masked samples are set to be zeros. The model will recover the masked samples and generate an evenly and densely sampled multi-channel sequence $\{Y_t \mid t = 1 \ldots n\}$. The goal is to minimize the loss between Y_t and X_t.

In training, Y_t is a raw CSI signal with a high sampling rate, and X_t is randomly sampled from Y_t through a specific strategy. After segmentation and re-sampling, X_t has the same dimension of Y_t (*i.e.* $T \times N$, where T is the length of CSI time series and N represents the number of sub-carriers).

Our ultimate goal is to train a generating function G that estimates for a given input low sampling and irregular signals, its corresponding high sampling and regular counterpart. To achieve this, we train a generator network as a feed-forward Convolutional Neural Network (CNN) and define a discriminator network D.

A. Generator

Considering the advantages of CNN in extracting high-dimensional feature information, a CNN model is designed to achieve model generation. The architecture of the generator is shown in Fig.6. At the core of our very deep generator network

G are B residual blocks with identical layout. Inspired by SRGAN [22], we use two convolutional layers with small 9×9 kernels and 64 feature maps, followed by batch-normalization layers and ParametricReLU as the activation function.

B. Discriminator

To discriminate real signals from generated samples, we train a discriminator network. The architecture is shown in Fig.6. We follow the architectural guidelines summarized by Radford et al. [23], use LeakyReLU activation and avoid max-pooling throughout the network. It contains four downsample blocks and four upsample blocks. The resulting feature maps are followed by an average pooling layer and a final sigmoid activation function to obtain a probability for each signal point classification.

C. Loss Function

The definition of loss function $\mathcal{L}_{GAN}$ is critical for the performance of our network. We design a loss function for generator that assesses a solution with respect to Wi-Fi signal-relevant characteristics. We formulate the loss as the weighted sum of a temporal domain loss $\mathcal{L}_t$ and a frequency domain loss $\mathcal{L}_f$.

Loss on Temporal Domain. The mean-squared error (MSE) focusing on low-level features is the most widely used optimization target for image and audio inpainting models. We apply it on the temporal domain:

$$\mathcal{L}_t = \mathbb{E}_t \left[\left| \hat{Y}_t - Y_t \right|^2 \right], \tag{1}$$

Loss on Frequency Domain. Aside from loss in temporal domain, we propose a loss function for frequency domain to restore high-frequency information in sparse part. The key to sparse recovery is the exploration of generating missing frequencies from low to high frequencies. Compared to the spatial domain, specific frequencies can be clearly separated in the frequency domain. In addition, frequency components can provide global information about signals. We transform the real and recovered waveforms into frequency domains with STFT. The L1-loss of spectrum magnitude difference between recovering and ground truth is averaged to produce the total frequency loss $\mathcal{L}_f$:

$$\mathcal{L}_f = \frac{1}{UV} \sum_{u=1}^{U} \sum_{v=1}^{V} \left| |\hat{Z}|_{u,v} - |Z|_{u,v} \right|, \tag{2}$$

where $\hat{Z}$ and Z represent the spectrum magnitude of the recovered and ground truth signal, respectively, and U and V represent the number of time and frequency components, respectively.

The theoretical advantage of using supervised loss in the frequency domain is twofold. (1) Direct emphasis, particularly on high-frequency information components that are lacking, improves recovery in these locations. (2) Because of the Fourier transform's features, this loss gives global guidance during training as opposed to a loss based on local signal points in the spatial domain.

The overall loss function of our model:

$$\mathcal{L}_{GAN} = \mathcal{L}_t + \alpha \mathcal{L}_f + \beta \mathcal{L}_d, \tag{3}$$

where α and β are hyperparameters and $\mathcal{L}_d$ is the Binary Cross Entropy (BCE) loss function for the discriminator.

VI. Evaluation

To demonstrate the effectiveness of our system, we conduct extensive experiments to evaluate its performance with respect to data restoration quality and the recognition accuracy of tasks.

A. Datasets

In order to assess the performance of the method, we specifically selected two types of applications that require high sampling rates: gesture recognition and gait recognition.

Gait Dataset. We install the CSI Tool on two IPCs and collect CSI using injection/monitoring mode with the sampling rate set to 1000Hz [24]. Tx has one antenna, while Rx has three antennas and is mounted on a tripod 0.5 m above ground to detect human movement more effectively. As shown in Fig.7, we deploy our experiments with a 6 m distance between Tx and Rx. Each volunteer is asked to walk along three trajectories and repeat them three times. We recruit three volunteers of different heights and weights to record the CSI of different people.

Widar3.0 [5]. We utilize the publicly available dataset from Widar3.0. Widar3.0 is the largest WiFi sensing dataset for gesture recognition, composed of 22 categories and 43K samples. It is collected via the Intel 5300 Network Interface Card (NIC) with 3×3 pairs of antennas in many distinct environments.

B. The Target Sampling Rate

A higher sample rate requires more missing data to be recovered but does not bring much improvement in sensing performance. Besides, when too much data needs to be recovered, the recovery capability of GANs decreases. As Fig.8 shows, we observed that the sensing performance at the sample rates of 200 Hz and 1000 Hz is very similar. Based on our empirical studies, a 200 Hz sample rate is good enough for most WiFi sensing applications such as hand gesture recognition, gait identification, and activity recognition. So we set f_{target} to 200 Hz, striking a balance between computer resources and recovering quality.

C. Implementation Details

To simulate the characteristics of intermittent data transmission, we randomly mask out different proportions of the original data. For instance, a 75% means the sampling rate is about 50 Hz. Then, we follow the method in section IV to preprocesing data.

All CSI data is preprocessed using Matlab R2023b on a computer with an AMD-6800HS 3.2GHz CPU. The deep learning model training and prediction are carried out on a server with an NVIDIA 3090 graphics card running Python

TABLE I
RECOVERING PERFORMANCE ON THE GAIT AND WIDAR DATASETS.

Dataset		**Gait dataset**			**Widar3.0** [5]		
Proportion	Method	MAE($\times 10^{-2}$)	MSE($\times 10^{-2}$)	PSNR	MAE($\times 10^{-2}$)	MSE($\times 10^{-2}$)	PSNR
50%	Linear interpolation	1.450±0.970	0.247±0.117	22.436±72.942	2.287±4.392	6.885±505.152	5.733±30.210
	Cubic interpolation	2.834±0.445	0.304±0.029	13.213±52.547	7.853±1.355	2.805±0.783	9.804±19.612
	Ours	**1.373±0.635**	**0.150±0.028**	**24.930±22.815**	**0.475±0.075**	**0.591±0.345**	**11.246±12.338**
75%	Linear interpolation	3.847±4.722	0.951±1.200	6.942±88.071	4.998±23.338	9.892±4.598	4.779±28.059
	Cubic interpolation	5.404±1.770	0.818±0.139	3.003±61.327	12.123±5.000	8.099±4.405	-6.209±27.249
	Ours	**3.409±2.599**	**0.485±0.153**	**10.896±20.024**	**2.151±4.712**	**0.880±0.560**	**6.139±12.976**
90%	Linear interpolation	7.067±11.607	2.220±5.640	-3.159±65.241	14.047±18.707	18.591±38.136	-12.148±78.552
	Cubic interpolation	8.447±2.299	1.560±0.234	-3.975±111.248	23.415±19.770	9.544±10.255	-11.884±80.954
	Ours	**5.784±5.275**	**1.043±0.503**	**2.191±9.349**	**5.845±7.355**	**1.314±1.296**	**1.930±9.490**
95%	Linear interpolation	26.910±94.598	13.085±5.154	-32.200±37.148	17.492±119.623	33.993±10.915	-23.080±76.761
	Cubic interpolation	18.224±11.301	6.174±3.603	-17.556±315.106	40.266±42.697	24.196±18.376	-31.146±59.120
	Ours	**9.342±8.577**	**2.135±1.362**	**-5.918±13.782**	**8.039±14.970**	**3.334±12.300**	**-5.011±7.800**

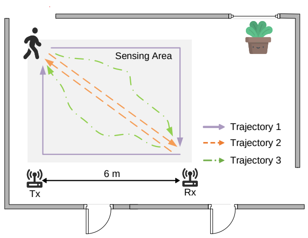


Fig. 7. **Experimental scenario of Gait dataset.**

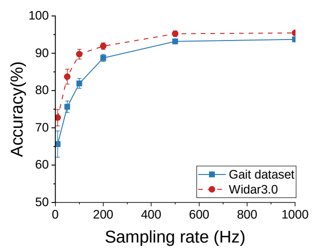
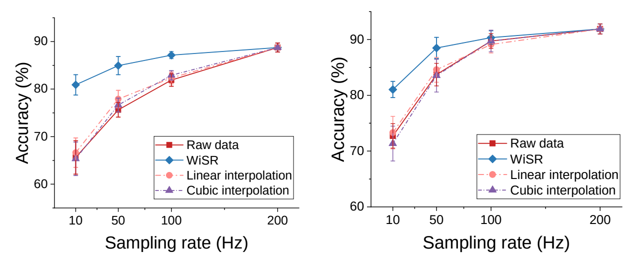


Fig. 8. **Performance under different sampling rate.**



(a) Gait recognition (b) Gesture recognition

Fig. 9. **Comparison between our method and baselines in Gait dataset and Widar3.0.**

3.10 with CUDA 11.8 and Pytorch 2.0. We use the Adam optimizer with a learning rate of 0.001, and the batch size is fixed at 64. The ratio of training and testing splits is 8:2 for all datasets using stratified sampling. The hyperparameter α for the loss function is set to 10, and β is set to 1.

D. Baselines and Criteria

We compare our system to two common interpolation methods utilized in Wi-Fi sensing systems [10], [17]. We employ the mean absolute error (MAE), MSE, and Peak Signal-to-Noise Ratio (PSNR) computed over the segments to assess performance.

E. Overall Performance

Overall recovering quality. Table I shows the evaluation results for signal recovery on Gait dataset and Widar3.0 datasets. As the proportion of missing data increases, the MSE and MAE values of all methods increase, while the PSNR values decrease. On the Gait and Widar3.0 datasets, the performance of the linear interpolation and cubic interpolation methods are relatively low, with higher MSE and MAE values and lower PSNR values regardless of the data ratio. In contrast, our method significantly improves recovery performance compared to interpolation methods on both datasets.

Accuracy of different tasks. We also provide a quantitative assessment of the proposed method in Wi-Fi sensing applications. We use CNN-LSTM as a classifier for evaluation. The comparison results are shown in Fig.9. We can see that the accuracy improvements of two interpolation methods are very limited, while our system can achieve high accuracy comparable to the accuracy of high-rate data. When the packet rate is 50 Hz, our system can achieve an accuracy of 85.34%, which is only 5.7% lower than the performance of 200Hz. These results demonstrate that our system can significantly improve the sensing performance, outperforming the two interpolation methods.

The impressive performance of our system can be attributed to several key factors: 1) We directly recover the signal in the time domain, reducing information loss in frequency domain conversion and improving the quality of signal recovery; 2) The PCA denoising method effectively removes noise, reduces the impact of noise on feature extraction, and provides higher signal-to-noise ratio for subsequent tasks. 3) Compared with simple interpolation algorithms, our system extracts signal features more comprehensively, ensuring a more stable quality of the recovered signal.

F. Impact of Denoising Methods

To validate the efficacy of the denoising method, we compare the recovery performance with different denoising methods on the Gait dataset. As depicted in Table. II, the denoising method we introduce does contribute to significantly reducing

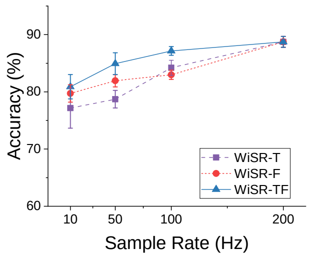


Fig. 10. **Impact of loss function.**

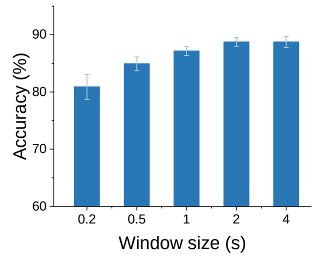


Fig. 11. **Impact of recovering window size.**

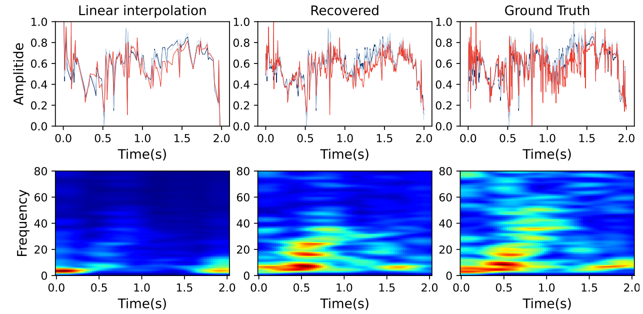


Fig. 12. **Example of CSI recovering in time and frequency domain.**

TABLE II
IMPACT OF DENOISING METHODS.

	MAE($\times 10^{-2}$)	MSE($\times 10^{-2}$)	PSNR
W/o denoising	10.934±3.409	3.712±0.510	1.672±86.335
Wavelet denoising	4.445±2.328	2.513±0.298	3.141±31.682
Median filtering	7.934±4.831	2.118±0.597	4.425±28.922
PCA denoising	**3.445±2.292**	**0.563±0.145**	**9.891±23.223**

TABLE III
IMPACT OF LOSS FUNCTION.

	MAE($\times 10^{-2}$)	MSE($\times 10^{-2}$)	PSNR
WiSR-T	**2.934±0.831**	**0.237±0.129**	6.672±4.335
WiSR-F	7.455±3.136	0.725±0.238	10.141±7.682
WiSR-TF	3.409±2.599	0.485±0.153	**10.896±20.024**

recovery error. It is clear that denoising is important to obtain accurate signal recovery. In addition, PCA denoising surpasses the other two methods in terms of all metrics. The results show that PCA denoising along sub-carriers successfully retains as much signal information as feasible while reducing ambient noise, hence improving recovery performance.

G. Impact of Loss Function

We also assess the efficacy of the generator network when it operates without the two loss functions, $\mathcal{L}_f$ (*WiSR*-F) and $\mathcal{L}_t$ (*WiSR*-T). We label *WiSR* with both loss functions as *WiSR*-TF. The quantitative results are summarized in Table III. *WiSR*-TF offers solutions with the highest MAE and MSE values. However, it appears less convincing perceptually compared to the results obtained with the loss function for the frequency domain. This is because the frequency domain loss function is more sensitive to the specific sensing task, as demonstrated in Fig.10.

H. Impact of Recovering Window Size

We evaluate our system using various sizes of the recovery window for the Gait dataset. We expand on the default window size used in prior experiments, using both larger and smaller sizes. The recognition accuracy for these various sizes is depicted in Fig.11. Our findings reveal that accuracy generally rises with an increase in window size. This is because longer sequences offer additional contextual information, facilitating the model's ability to capture extended temporal dependencies. However, this also escalates computational requirements, including memory usage and computation time, thereby potentially affecting training and inference efficiency.

I. Visualization of Sparse Recovery

In the low-frequency range of 0-40 Hz, our method successfully recovers a significant portion of the input signal. The waveform shows a clear resemblance to the original signal, with most of the major features preserved. The spectrogram further verifies a robust reconstruction of the frequency components, mirroring the same peaks and patterns as the original signal. However, it is important to note that while our system excels at recovering low-frequency information, it does encounter challenges in the high-frequency range of 40-80 Hz. In this region, the missing minor details and the relatively poorer recovery effect are evident in both the waveform and the spectrogram. These results highlight the strengths and limitations of our approach. While we achieve remarkable performance in reconstructing the low-frequency components, further improvements are needed to enhance the recovery of high-frequency details.

VII. CONCLUSION

This paper introduces *WiSR*, an innovative generative recovery model designed specifically for Wi-Fi signals, operating on raw CSI waveforms. *WiSR* stands out for its ability to recover high-fidelity time-series signals. It achieves this by utilizing two innovative denoising methods and applying a loss function in both the temporal and frequency domains. This process allows it to capture the complex characteristics of Wi-Fi signals across the spatial, temporal, and frequency domains. Comprehensive experimental results demonstrate that *WiSR* can enhance the accuracy of wireless sensing systems, particularly under low and irregular frame rates, thereby offering significant potential for key wireless tasks.

REFERENCES

[1] Han Zou, Yuxun Zhou, Jianfei Yang, Hao Jiang, Lihua Xie, and Costas J Spanos. Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.

[2] Jianfei Yang, Han Zou, Hao Jiang, and Lihua Xie. Carefi: Sedentary behavior monitoring system via commodity wifi infrastructures. *IEEE Transactions on Vehicular Technology*, 67(8):7620–7629, 2018.

[3] Ruiyang Gao, Mi Zhang, Jie Zhang, Yang Li, Enze Yi, Dan Wu, Leye Wang, and Daqing Zhang. Towards position-independent sensing for gesture recognition with wi-fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–28, 2021.

[4] Ruiyang Gao, Wenwei Li, Yaxiong Xie, Enze Yi, Leye Wang, Dan Wu, and Daqing Zhang. Towards robust gesture recognition by characterizing the sensing quality of wifi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–26, 2022.

[5] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-effort cross-domain gesture recognition with wi-fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, pages 313–325, 2019.

[6] Wei Meng, Xingcan Chen, Wei Cui, and Jing Guo. Wihgr: A robust wifi-based human gesture recognition system via sparse recovery and modified attention-based bgru. *IEEE Internet of Things Journal*, 9(12):10272–10282, 2021.

[7] Jin Zhang, Bo Wei, Fuxiang Wu, Limeng Dong, Wen Hu, Salil S Kanhere, Chengwen Luo, Shui Yu, and Jun Cheng. Gate-id: Wifi-based human identification irrespective of walking directions in smart home. *IEEE Internet of Things Journal*, 8(9):7610–7624, 2020.

[8] Dazhuo Wang, Jianfei Yang, Wei Cui, Lihua Xie, and Sumei Sun. Caution: A robust wifi-based human authentication system via few-shot open-set recognition. *IEEE Internet of Things Journal*, 9(18):17323–17333, 2022.

[9] Jin Zhang, Zhuangzhuang Chen, Chengwen Luo, Bo Wei, Salil S Kanhere, and Jianqiang Li. Metaganfi: Cross-domain unseen individual identification using wifi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(3):1–21, 2022.

[10] Yinghui He, Jianwei Liu, Mo Li, Guanding Yu, Jinsong Han, and Kui Ren. Sencom: Integrated sensing and communication with practical wifi. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2023.

[11] Yu Zeng, Zhe Lin, Huchuan Lu, and Vishal M. Patel. Cr-fill: Generative image inpainting with auxiliary contextual reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14164–14173, October 2021.

[12] Jitesh Jain, Yuqian Zhou, Ning Yu, and Humphrey Shi. Keys to better image inpainting: Structure and texture go hand in hand. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 208–217, January 2023.

[13] Chunjing Xiao, Daojun Han, Yongsen Ma, and Zhiguang Qin. Csigan: Robust channel state information-based activity recognition with gans. *IEEE Internet of Things Journal*, 6(6):10191–10204, 2019.

[14] Dehao Jiang, Mingqi Li, and Chunling Xu. Wigan: A wifi based gesture recognition system with gans. *Sensors*, 20(17):4757, 2020.

[15] Shelly Vishwakarma, Chong Tang, Wenda Li, Karl Woodbridge, Raviraj Adve, and Kevin Chetty. Gan based noise generation to aid activity recognition when augmenting measured wifi radar data with simulations. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6. IEEE, 2021.

[16] Lijian Wei, Jun Feng, Yufei Liu, Tuo Zhang, Qirong Bu, and Baoying Liu. A dense margin network for human activity recognition based on augmented channel state information. In *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, pages 67–74. IEEE, 2021.

[17] Jingzhi Hu, Tianyue Zheng, Zhe Chen, Hongbo Wang, and Jun Luo. Muse-fi: Contactless muti-person sensing exploiting near-field wi-fi channel variation. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[18] Kun Yang, Xiaolong Zheng, Jie Xiong, Liang Liu, and Huadong Ma. Wiimg: Pushing the limit of wifi sensing with low transmission rates. In *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9. IEEE, 2022.

[19] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–26, 2019.

[20] Shengjie Li, Zhaopeng Liu, Yue Zhang, Qin Lv, Xiaopeng Niu, Leye Wang, and Daqing Zhang. Wiborder: Precise wi-fi based boundary sensing via through-wall discrimination. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–30, 2020.

[21] Youwei Zhang, Feiyu Han, Panlong Yang, Yuanhao Feng, Yubo Yan, and Ran Guan. Wi-cyclops: Room-scale wifi sensing system for respiration detection based on single-antenna. *ACM Transactions on Sensor Networks*, 2023.

[22] Baozhong Liu and Ji Chen. A super resolution algorithm based on attention mechanism and srgan network. *IEEE Access*, 9:139138–139145, 2021.

[23] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[24] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review*, 41(1):53–53, 2011.