Journal of Computer and Communications, 2026, 14(4), 136-154
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A Review of Diffusion Model-Based Channel State Information Data Augmentation for Human Activity Recognition

Ruotong Zhu, Jiaxuan Li, Peng Su, Luyu Zhang, Bosheng Fu, Zhengjie Wang*

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China
Email: 2629410625@qq.com, 3506023233@qq.com, 1066785475@qq.com, 2398323202@qq.com, 2392747855@qq.com, *cieewangzj@163.com

How to cite this paper: Zhu, R.T., Li, J.X., Su, P., Zhang, L.Y., Fu, B.S. and Wang, Z.J. (2026) A Review of Diffusion Model-Based Channel State Information Data Augmentation for Human Activity Recognition. *Journal of Computer and Communications*, **14**, 136-154.
https://doi.org/10.4236/jcc.2026.144007

Received: April 5, 2026
Accepted: April 21, 2026
Published: April 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

With the rapid development of the Internet of Things and wireless communication technologies, human activity recognition based on WiFi Channel State Information (CSI) has become a core research topic in smart homes, health monitoring, and security surveillance, due to its non-invasiveness, privacy protection, and the need for no specialized equipment. CSI can accurately capture amplitude and phase perturbations of wireless channels caused by human movements, providing a reliable data basis for contactless action recognition. However, this technology faces critical challenges such as scarce labeled data, poor cross-scenario adaptability, and insufficient model generalization ability in practical applications. This paper analyzes the technical principles and research status of CSI-based action recognition, focusing on the application of diffusion models in data augmentation. By integrating generative data expansion, cross-domain feature alignment, and model robustness optimization, a complete CSI action recognition enhancement scheme is proposed. It systematically discusses how diffusion models alleviate the few-shot problem through high-fidelity sample generation and improve cross-scenario recognition performance via domain adaptive learning, providing theoretical support and technical references for the practical promotion of CSI action recognition technology.

Keywords

Channel State Information, Action Recognition, Diffusion Model, Data Augmentation, Cross-Scenario Adaptation, Generative AI

1. Introduction

1.1. Research Background

In the evolution of intelligent sensing technology, human activity recognition (HAR), as a core technology for understanding human behavior, has been widely used in health monitoring, smart home interaction, public security, and other fields [1]. Traditional action recognition schemes mainly rely on cameras, wearable sensors, or radar devices, but they have inherent defects such as privacy leakage risks, inconvenient wearing, and limitations caused by lighting conditions [2]. In contrast, CSI-based WiFi action recognition technology utilizes existing wireless communication infrastructure to achieve passive sensing by capturing the impact of human movements on wireless signal propagation paths, with significant advantages such as all-weather operation, no privacy infringement, and low deployment cost [3].

As a fine-grained channel parameter at the physical layer of WiFi systems, CSI can reflect the amplitude attenuation and phase offset characteristics of signals during multipath propagation [4]. When the human body moves in an indoor environment, it changes the propagation path of wireless channels, leading to regular changes in CSI sequences [5]. Different actions, such as walking, bending, falling, etc., correspond to significantly different CSI perturbation patterns, providing a key distinguishing feature for action recognition [6]. In recent years, the popularization of 802.11 n/ac protocols has promoted the wide application of MIMO technology, enabling CSI to carry rich channel information for multiple antennas and subcarriers. This technological advancement not only supports the improvement of wireless communication efficiency but also provides a higher-precision signal foundation for CSI-based sensing tasks. Lee *et al.* focused on the generation of high-dimensional CSI data in MIMO systems, capturing the correlation between user positions and multi-antenna channels through the conditional DDIM model to generate high-fidelity CSI samples. Their technical idea can further support the few-shot training of sensing tasks such as action recognition [7].

1.2. Research Problems

Before investigating specific applications, we introduce several basic concepts that will be mentioned later in this paper, as this helps to define the research scope and key content. Few-shot learning stands for learning scenarios with limited annotated data availability. Cross-scene adaptation presents a model ability for stable recognition under environmental and device variations. Weak motion describes micro behaviors with minor disturbance to WiFi signal responses. High-fidelity represents a strong coincidence between generated samples and real CSI measurements.

Although CSI-based action recognition technology shows great potential, it still faces three core bottlenecks in actual deployment. First, high-quality labeled data are scarce, and few-shot scenarios are prevalent. Re-collecting and labeling data consumes a lot of time and labor [3]. In addition, there are many restrictions on the actual deployment of CSI action recognition models, which ultimately make it difficult for the model to effectively adapt to new activities, new users, or new scenarios. Second, cross-scenario adaptability is poor. Changes in environmental layout, individual user differences, and device heterogeneity will lead to a data distribution shift (domain shift), seriously reducing the model's generalization ability [8]. Third, weak action perception is difficult. The CSI perturbation signals corresponding to small movements such as breathing and micro-postures are weak and are easily interfered with by environmental noise, making it difficult to guarantee recognition accuracy [9]. These challenges restrict the large-scale application of CSI action recognition technology and urgently need to be solved by advanced machine learning and generative AI technologies. Relying on the automatic feature learning ability, deep learning methods have shown significant advantages in CSI action recognition, providing a new path to break through the above bottlenecks [10].

1.3. Application of Diffusion Models

Drawing primarily on IEEE, ACM, ScienceDirect, and Springer databases, we reviewed relevant literature published over the past five years. This survey specifically focuses on studies utilizing diffusion models to generate virtual Channel State Information (CSI) data, which serves to augment available datasets and enhance the accuracy of human activity recognition (HAR). We analyze the related applications, methodologies, and implementations, investigating key aspects such as diffusion model architectures and the resulting improvements in recognition precision. Furthermore, this review strictly concentrates on the application of diffusion models and deep learning techniques in human activity recognition, explicitly excluding research that relies on conventional machine learning methods.

Early CSI action recognition research mainly relied on handcrafted features and traditional machine learning models [11]. Abdelnasser *et al.* extracted statistical features and state change features of RSSI signals and combined them with SVM to achieve gesture recognition [12]. Li *et al.* compressed CSI waveforms using the Discrete Wavelet Transform (DWT) and calculated sequence similarity with the Dynamic Time Warping (DTW) algorithm for action recognition [13].

In the field of WiFi CSI action recognition, in response to practical challenges such as few-shot learning, data distribution shift, and difficulty in perceiving weak movements that emerged later, researchers have successively proposed a series of generative learning methods, such as traditional data augmentation and Generative Adversarial Networks (GAN). Among them, traditional augmentation methods rely on time-series transformation and frequency-domain perturbation to achieve sample expansion, while GAN learns data distribution through adversarial training. However, the former lacks diversity, and the latter is prone to mode collapse and unstable training, making it difficult to meet the generation requirements of both high fidelity and strong generalization [2] [6]. Against this background, diffusion models, as a class of deep generative models based on probabilistic likelihood learning, have gradually become the mainstream solution due to their stable training process and excellent time-series fitting ability. Diffusion models achieve high-fidelity sample synthesis through a dual-process mechanism of forward progressive noise addition and reverse iterative denoising. In the forward diffusion stage, the model progressively adds Gaussian noise to the original CSI time-series signals, smoothly converting the real data distribution into a standard Gaussian distribution; in the reverse generation stage, the network learns noise prediction and removal rules, and gradually restores CSI samples conforming to the real distribution from pure noise [6]. Compared with traditional generative models such as GAN, diffusion models show more balanced performance in time-series signal synthesis. With the conditional control mechanism, they can also guide the generation of directional CSI samples according to action categories, scene information, etc., providing an effective technical path to solve the problems of data scarcity and insufficient generalization in WiFi sensing [7] [14]. At the level of feature optimization, feature fusion strategy has also become an important direction to improve recognition performance. The CSI-F method proposed by Niu *et al.* strengthens the discrimination of action features through CSI signal feature fusion technology, achieving more stable recognition effects in complex indoor environments [15], forming a technical complement to the generation enhancement of diffusion models and jointly helping to improve CSI action recognition performance.

Diffusion models have shown significant application value in WiFi CSI sensing tasks. The DiffAR adaptive conditional diffusion model proposed by Huang *et al.* effectively improves the fidelity and cross-scenario adaptability of synthetic CSI samples through step-specific conditional guidance [16]. Hao *et al.* introduced latent diffusion models into domain-invariant pose estimation tasks, achieving robust feature learning of WiFi signals in complex environments [14]. Xu *et al.* constructed a CSI enhancement scheme based on the conditional diffusion framework, achieving single-model multi-type action sample generation with action labels as conditions [17]. Relevant studies have confirmed that diffusion models can accurately fit the time-series perturbation laws of CSI, and their performance is significantly better than that of traditional augmentation and GAN methods in few-shot completion, cross-domain generalization, weak action enhancement, and other scenarios, making them the mainstream research direction for device-independent WiFi sensing data augmentation.

Although several studies have employed diffusion models to generate CSI data for augmenting synthetic datasets, there is a notable lack of comprehensive reviews on this topic. Therefore, this study conducts an in-depth investigation of existing works that utilize diffusion models to generate synthetic CSI data for action recognition, summarizes their implementation methodologies, analyzes typical applications, and provides a reference for future research on generating synthetic CSI data using diffusion models.

1.4. Paper Structure

The contents of this paper are arranged as follows. Section 2 systematically investigates the core technical principles, signal characteristics, sensing mechanisms, and system framework of CSI action recognition. Section 3 analyzes the core challenges of CSI action recognition, introduces the basic principles of diffusion models, the advantages of data augmentation, and their technical adaptability. Section 4 proposes a full-process enhancement scheme for CSI action recognition, integrating diffusion models, designing multi-dimensional technical strategies, and constructing the overall framework. Section 5 sorts out the key challenges faced by CSI action recognition research based on diffusion models and discusses the core future research directions. Section 6 summarizes the paper and refines the theoretical contributions and practical application values.

2. Principles and System Framework of CSI Action Recognition

2.1. CSI Signal Characteristics and Action Sensing Mechanism

WiFi systems adopt Orthogonal Frequency Division Multiplexing (OFDM) technology, dividing the channel into multiple orthogonal subcarriers, and CSI can provide an independent channel state description for each subcarrier [1]. Relevant studies show that in MIMO systems, CSI exists in the form of a matrix with dimensions of the number of transmitting antennas, the number of receiving antennas, and the number of subcarriers, fully reflecting the signal propagation characteristics between multiple antennas [18]. The amplitude and phase information of CSI jointly carry channel change characteristics, among which the amplitude information is more sensitive to environmental disturbances and less affected by hardware synchronization errors, so it is widely used in action recognition research [3].

The impact of human movements on CSI is mainly achieved through two mechanisms. First, the occlusion effect is important. The human body, as a dielectric, will occlude the direct path signal, leading to enhanced signal attenuation. Second, the multipath effect is another important factor. Human reflection will generate new propagation paths, changing the superposition relationship of multipath signals [5]. Different movements have different amplitudes, speeds, and spatial ranges, leading to significantly different CSI change patterns [6]. For example, walking movements cause periodic fluctuations in CSI sequences, falling movements correspond to sudden changes in CSI amplitude, and CSI sequences are relatively stable in a static state [1]. By extracting these feature differences, the distinction of different action categories can be realized [19]. Deep learning models can automatically capture these complex feature patterns without manual feature engineering, significantly improving the adaptability of recognition systems [10].

2.2. CSI Action Recognition System Framework

A complete CSI action recognition system (as shown in **Figure 1**) usually includes five core modules: data collection, preprocessing, feature extraction, model training, and recognition inference. The specific process is as follows [4]:

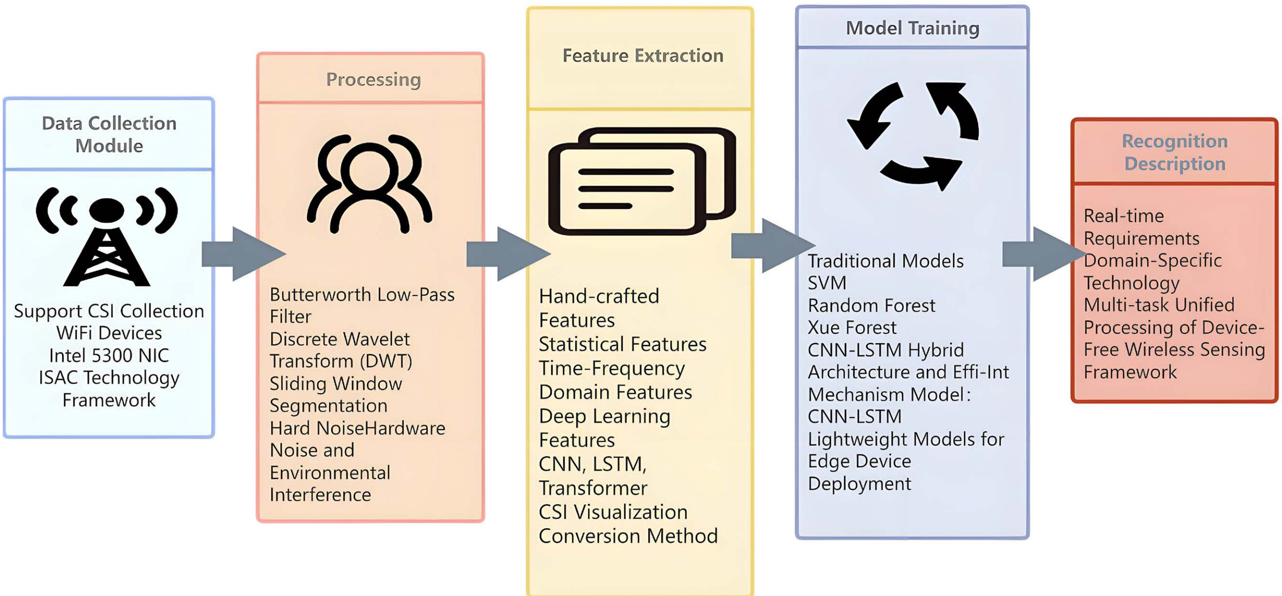

Figure 1. CSI action recognition system framework.

2.2.1. Data Collection Module

WiFi devices supporting CSI collection (such as Intel 5300 network cards) are used to extract original CSI data through CSI tools, and action labels such as breathing are recorded synchronously during the collection process to provide labeled data for model training [9]. It can effectively adapt to the signal sensing requirements of indoor non-contact scenarios and ensure the sensitivity of CSI data to weak breathing movements.

2.2.2. Preprocessing Module

Original CSI data contains high-frequency noise, hardware noise, and environmental interference, which need a series of processing steps to improve data quality [1]. Common methods include Butterworth low-pass filtering to remove high-frequency noise, Discrete Wavelet Transform (DWT) to decompose signal components, and sliding window segmentation to extract action segments, etc. [19]. In addition, for the Carrier Frequency Offset (CFO) and Sampling Frequency Offset (SFO) problems in CSI phase information, strategies such as phase calibration or directly discarding phase information are usually adopted. The quality of the preprocessing stage directly affects the subsequent feature extraction effect and provides high-quality input data for deep learning models [20].

2.2.3. Feature Extraction Module

This module aims to extract discriminative action features from preprocessed CSI sequences, divided into two categories: handcrafted features and deep learning features [1]. Handcrafted features include statistical features (mean, variance, peak value), time-frequency domain features (Short-Time Fourier Transform, wavelet coefficients), etc., relying on expert experience design. Deep learning features are automatically learned through models such as CNN, LSTM, and Transformer, which can capture the spatio-temporal correlation features of CSI sequences with better recognition performance [6]. The introduction of the attention mechanism [10] further strengthens the extraction of key action features and improves the model's ability to perceive weak movements. In recent years, feature extraction methods based on CSI image conversion have become a research hotspot. The ImgFi framework proposed by Zhang *et al.* converts one-dimensional CSI sequences into two-dimensional images through five CSI imaging methods (recurrence plot, Gramian Angular Summation Field, etc.), giving full play to the advantages of CNN in image recognition, and achieving 99.5% recognition accuracy with only a three-layer convolution structure [21], providing a new effective path for feature extraction.

2.2.4. Model Training Module

Classification models are used to train the extracted features and establish the mapping relationship between action categories and feature patterns [2]. Traditional models include Support Vector Machine (SVM), Random Forest, etc., while deep learning models are represented by CNN-LSTM hybrid architecture and attention mechanism models [1]. Problems such as data imbalance and overfitting need to be solved during training to improve the model's generalization ability. Lightweight deep learning models [22] reduce computational overhead while ensuring recognition accuracy through parameter optimization and structure simplification, adapting to edge device deployment requirements. The HARNN scheme proposed by Ding *et al.* adopts a Deep Recurrent Neural Network (RNN) combined with the LSTM module, and extracts Channel Power Variation (CPV) and Time-Frequency Analysis (TFA) features, achieving more than 95% accuracy in six daily action recognitions [23], verifying the effectiveness of recurrent neural networks in time-series CSI feature modeling.

2.2.5. Recognition Inference Module

The CSI sequence to be recognized is input into the trained model, and the action recognition result is output through feature matching and category decision [6]. This module needs to meet real-time requirements to ensure that the recognition delay is within the acceptable range for practical applications [24]. The integration of domain adaptive technology [20] enables the model to maintain stable performance when deployed across scenarios, expanding the application scope of the technology. Zhang *et al.* systematically sorted out the application logic of deep learning in tasks such as positioning and action recognition in the field of device-free wireless sensing, and proposed a general technical framework integrating deep learning methods, providing design ideas and technical support for integrated solutions of multi-task unified recognition inference in complex scenarios [25].

3. Core Challenges and Solutions

3.1. Key Technical Challenges

The performance of CSI action recognition systems is restricted by many factors, and the core challenges are mainly reflected in three aspects. First, at the data and deployment level, high-quality labeled data is scarce, and few-shot scenarios are prevalent. Fast sampling easily loses details, high-fidelity generation has large computational overhead, and lightweight model deployment is difficult [3][6]. Second, at the scenario generalization level, complex environmental features, individual user differences, and device heterogeneity jointly aggravate the CSI data domain shift, and it is difficult to deeply alleviate the distribution shift problem, severely limiting cross-scenario generalization ability [3]. Third, at the feature perception level, the CSI perturbation signals corresponding to weak movements such as breathing and micro-postures are weak with a low signal-to-noise ratio, and are easily submerged by environmental noise [9]. These challenges need to be solved collaboratively through strategies such as data augmentation, feature optimization, and model improvement [7], and the integration of generative AI and deep learning [14] provides a new technical path to solve these problems.

3.2. Principles and Applications of Diffusion Model Enhancement Technology

3.2.1. Basic Principles of Diffusion Models

A diffusion model is a deep learning model based on probabilistic generation, which achieves high-fidelity sample generation through two processes: forward diffusion and reverse generation (as shown in **Figure 2**). In the forward diffusion process, the model gradually adds Gaussian noise to the original data to convert the data distribution into a standard Gaussian distribution. The reverse generation process learns the inverse process of noise to restore the original data distribution from Gaussian noise [6]. Compared with GAN, diffusion models have stable training, are not prone to mode collapse, and can generate samples with more diversity and fidelity, making breakthrough progress in image generation, time-series data synthesis, and other fields [9].

The core advantage of diffusion models lies in their ability to model time-series data. For time-series signals such as CSI, diffusion models can capture their time dependence and dynamic change characteristics, generating sequence samples conforming to physical laws. Conditional diffusion models can directionally generate specific types of CSI samples by introducing conditional information such as action categories and scene parameters, making targeted data augmentation possible [7]. The DiffAR Adaptive Conditional Diffusion Model (ACDM) proposed by Huang *et al.* further improves the quality and adaptability of synthetic CSI through step-specific conditional guidance for sample generation [16]. Hao *et al.* applied latent diffusion models to domain-invariant human pose estimation from WiFi signals [14], expanding the application scenarios of diffusion models in CSI sensing.

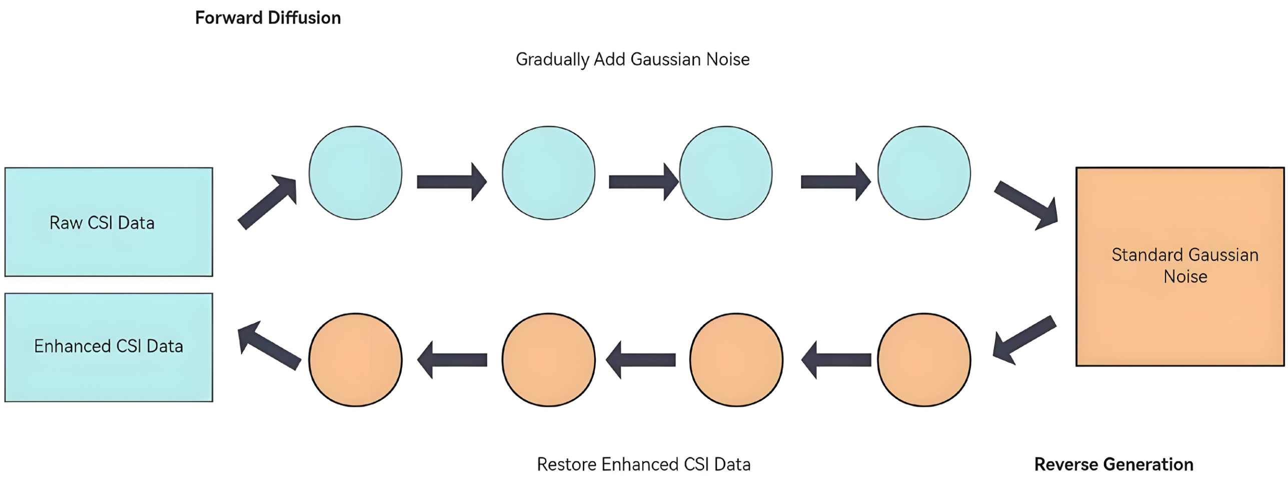

Figure 2. Forward and reverse processes of the diffusion model.

3.2.2. Advantages of Diffusion Models in CSI Data Augmentation

Diffusion models show three core advantages in CSI data augmentation. First, high-fidelity sample generation ability, which can accurately simulate the CSI perturbation patterns corresponding to different actions, with minimal numerical differences between synthetic samples and real samples. The Mean Absolute Error (MAE) can be as low as 0.718, the Mean Squared Error (MSE) as low as 1.003, and the Continuous Ranked Probability Score (CRPS) of the compatibility between the generated distribution and the real distribution can be as optimal as 1.001 in multi-scenarios, fully reflecting excellent fidelity [16]. Second, diverse generation characteristics can be achieved. By adjusting the noise scheduling strategy and conditional parameters, diffusion models can generate CSI samples covering different indoor environments, different user behaviors, and multi-user concurrent scenarios, effectively expanding the data distribution in real scenarios and alleviating the problems of data distribution imbalance and sample scarcity [19]. Third, weak action recognition can be implemented. Diffusion models achieve data generation through progressive denoising, which can effectively weaken the noise and interference components in CSI signals, making generated samples cleaner and features clearer, thus enhancing the adaptability of weak action recognition to environmental interference [9].

Compared with traditional data augmentation methods, diffusion models can generate brand-new, unseen CSI samples instead of simply modifying existing samples, which can better meet the needs of model generalization ability [6]. Compared with GAN, the samples generated by diffusion models have no mode collapse problem and can more comprehensively cover the variation space of actions and scenarios. Studies have shown that diffusion model-based CSI data augmentation can improve action recognition accuracy by 3% - 8% in few-shot scenarios, significantly better than traditional methods [16]. In addition, diffusion models also perform prominently in the combined scenario of CSI compression and generation. Kim *et al.* [26] proposed a generative diffusion model based on a compression scheme, realizing efficient compression and reconstruction of high-dimensional CSI data through conditional diffusion decoding, reducing data transmission overhead while ensuring recognition performance. Lee *et al.* [27] generated user-specific high-dimensional channel data through the conditional DDIM model, achieving compression performance comparable to full data with only 1% of real training data, providing strong support for CSI data augmentation in few-shot scenarios. Khan *et al.*'s research [28] also verified the effectiveness of deep learning models in CSI data modeling, laying a foundation for the application of diffusion models.

To more intuitively compare the core differences among diffusion models, the two methods of traditional data augmentation and GAN in CSI data augmentation, and further highlight the performance advantages of diffusion models, Table 1 summarizes the characteristics and quantitative indicators of the three methods in key dimensions such as principle, fidelity, and robustness:

Table 1. Performance comparison of three CSI data augmentation methods.

Table Head	Comparison Dimension		
	Diffusion Model [6] [9] [16]	Traditional Augmentation [20] [22]	GAN [2]
Sample Fidelity	High	Medium	Medium-Low
Sample Diversity	High	Low-Medium	Medium
Training Stability	High	High	Low
Noise Robustness	High	Low-Medium	Medium
Few-Shot Accuracy Gain	Significant	Limited	Moderate

3.2.3. Adaptability Between Diffusion Models and CSI Action Recognition

The time-series characteristics and multi-dimensional structure of CSI signals are highly compatible with the modeling ability of diffusion models [6]. As typical time-series data, CSI sequences contain action characteristics in the changing trend of the time dimension, and diffusion models can accurately capture this trend through progressive denoising and generation of time-series data [17]. At the same time, the multi-antenna and multi-subcarrier dimensional characteristics of CSI can be effectively modeled through the high-dimensional data generation ability of diffusion models, and the generated samples can completely retain the spatial structure information of the original CSI [1] [17]. The conditional latent diffusion model RF-ACCLDM further extends this adaptability to multi-modal RF data such as RFID and FMCW radar. The class-conditional synthetic data generated in the latent space have high fidelity and excellent computational efficiency, and their performance in downstream tasks such as 3D human pose tracking and HAR is comparable to or even surpasses models trained with pure real data [29]. Lee *et al.*'s research [7] further verified this adaptability. Their user position conditional diffusion model can generate high-dimensional CSI samples consistent with the real channel distribution, effectively improving the task generalization ability and performance stability in cross-user scenarios.

In practical applications, diffusion models can generate directional samples according to conditions such as action categories and scene parameters [7]. For example, for the few-shot problem of high-risk actions such as falls, many high-fidelity fall action CSI samples can be generated through conditional diffusion models [16]. For the domain shift problem in cross-scenario recognition, CSI samples covering different environmental layouts and user characteristics can be generated to help the model learn domain-invariant features [9]. Hao *et al.*'s latent diffusion model [14] is directly oriented to domain-invariant feature learning, further strengthening cross-scenario adaptability. In addition, the combination of diffusion models and signal processing technologies (such as VMD-HHT feature extraction) can further improve the quality of CSI samples corresponding to weak actions and enhance the model's ability to recognize small movements [9]. Zuo *et al.*'s unsupervised diffusion model generates diverse synthetic samples through statistical information such as mean, standard deviation, and Z-score, expanding the training set without labeled data and providing a new solution for CSI action recognition in low-resource scenarios [30]. To systematically compare the technical differences and application scopes of these diffusion-based approaches, existing studies are categorized by diffusion model type, task objective, conditioning signals, and evaluation setup, as shown in **Table 2**.

Table 2. Classification of diffusion-based CSI sensing methods.

Diffusion Model Type	Test Dataset		
	Conditional Diffusion Model [16] [17]	Temporal Enhanced Diffusion Model [16] [19] [22]	Generative Diffusion Model [17] [27]
Task	Data Augmentation and Activity Recognition	Multi-User Activity Recognition	Signal Fidelity and Distribution Fitting
Conditional Signal	Action Class Labels, CSI Time-Series Signals	Time-Domain Features, Subcarrier Information	Amplitude and Phase Features
Evaluation Setup	Office, SignFi, UT-HAR Datasets	Single-User/Multi-User Scenarios	MAE

4. CSI Action Recognition Enhancement Scheme Based on Diffusion Models

4.1. Overall Framework of the CSI Action Recognition Enhancement Scheme

The overall framework of the CSI action recognition enhancement scheme based on diffusion models is shown in **Figure 3**. The data collection module is responsible for obtaining original CSI data and conducting preprocessing such as denoising and segmentation. The diffusion model module supports the generation of high-fidelity synthetic samples according to multi-dimensional conditions, which, together with real samples from the training dataset [17]. The feature extraction module adopts a CNN-self-attention hybrid architecture or CSI image conversion method to extract spatio-temporal features [19] [21], and strengthens key feature capture combined with the attention mechanism [10]. The cross-domain adaptive training module copes with environmental interference through data augmentation and robustness optimization to improve the model's generalization ability in different scenarios. Such a framework comprehensively improves the accuracy, robustness, and cross-scenario adaptability of CSI action recognition through the collaborative effect of data augmentation and model optimization [17], and integrates the lightweight design concept [22] to ensure the engineering deployment feasibility of the scheme.

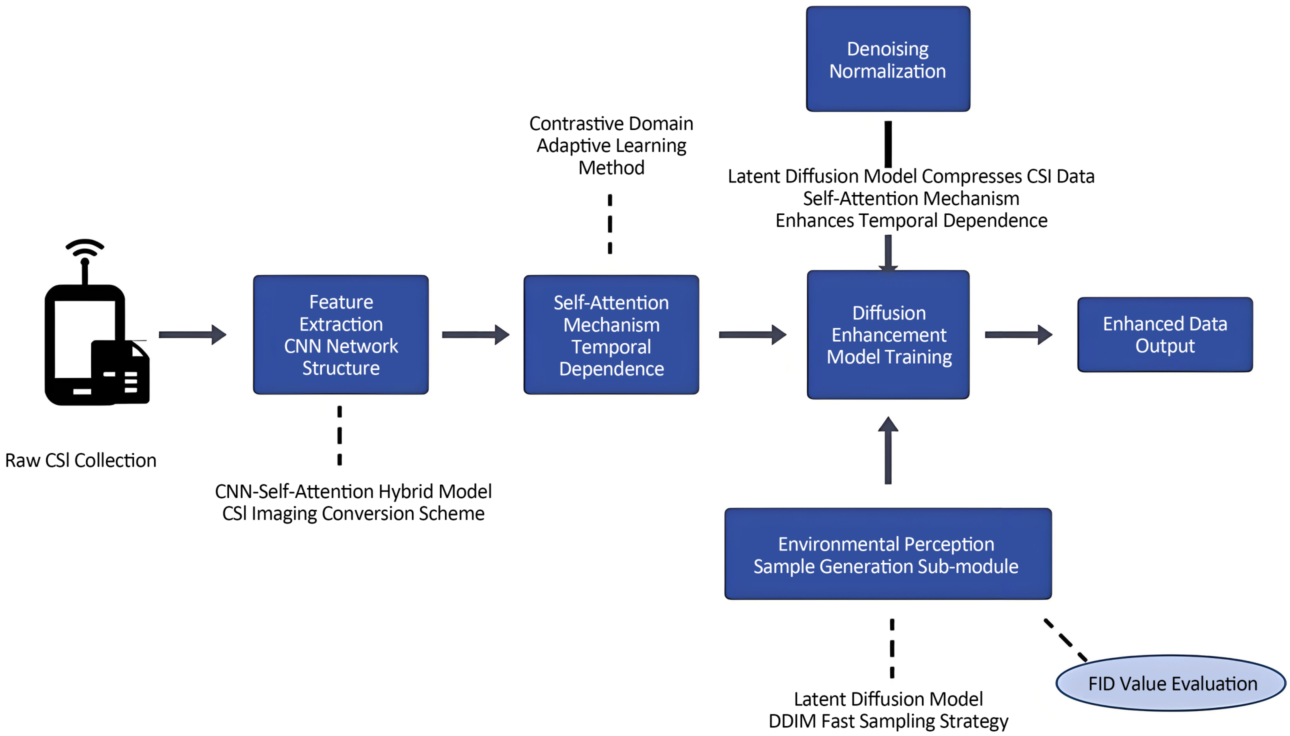

Figure 3. Framework of the CSI action recognition enhancement scheme based on diffusion models.

All performance improvements in the diffusion model-based CSI human activity recognition enhancement scheme are based on specified baseline methods, corresponding datasets, and application scenarios. The performance of virtual samples can be verified with that of real samples through three indicators, including distribution fitting degree, downstream task accuracy, and noise robustness. Approximation errors still exist between the synthetic distribution and the real distribution. High-fidelity generation brings high computational cost [7]. Test benchmarks vary in different studies and are thus inconsistent. Virtual samples also carry potential privacy leakage risks [13]. All these factors require comprehensive trade-offs in practical deployment.

4.2. High-Fidelity CSI Sample Generation Strategy

Targeted sample generation strategies are designed by combining the characteristics of diffusion models with CSI signal features [16]. Conditional diffusion models are adopted to learn CSI data distribution, where action category labels serve as conditional guidance for generation. Single models are empowered to support multi-type action sample generation, thereby reducing training and storage overhead. The reverse iteration process can be simplified by reasonably setting diffusion steps to improve inference efficiency while ensuring high-fidelity sample generation [17]. It can be applied to introduce multi-dimensional conditional information, such as action categories, user characteristics, and scene parameters, to construct conditional diffusion models and realize directional sample generation [6]. Additionally, the diffusion model compression framework proposed by Kim *et al.* can be utilized to optimize data transmission efficiency while generating high-fidelity samples, balancing recognition performance and deployment practicality [26]. Finally, the detailed features of generated samples can be optimized through adversarial training to enhance the discrimination of similar actions (such as sitting and bending) [2].

A multi-index system is adopted for the quality evaluation of generated samples [7] [16]. MAE (Mean Absolute Error) and MSE (Mean Squared Error) are used to measure the numerical similarity between synthetic samples and real samples, and CRPS (Continuous Ranked Probability Score) is used to evaluate the compatibility between the generated distribution and the real distribution. The effectiveness of generated samples is reflected through action recognition accuracy. Related studies indicate that the CSI samples generated by this strategy have a MAE as low as 0.822 and MSE as low as 1.134 on the Office dataset, and a MAE as low as 0.717 and MSE as low as 1.003 on the SignFi dataset, significantly better than existing diffusion models (such as DiffWave, WaveGrad), and can effectively improve model training effects [16]. Combined with the CSI enhancement method proposed by Shi *et al.* [20], sample quality can be further optimized, and the discrimination of action features can be strengthened. For CSI image-based feature extraction scenarios, diffusion models can be used to generate diverse CSI image samples, providing richer training data for frameworks such as ImgFi and further improving their recognition accuracy in few-shot scenarios [16] [21].

To verify the generalization performance of the enhancement scheme on different public datasets and the credibility of the enhancement scheme, **Table 3** is provided:

Table 3. Generalization performance of the enhancement scheme on different public datasets.

Performance	Test Dataset			
	Office [16]	SignFi [16]	CSI-HAR [22]	UT-HAR [17]
Actions	7	276	7	7
Baseline Accuracy	96.97%	96.74%	85.70%	93.37%
Enhancement Method	Diffusion Augmentation	Diffusion Augmentation	Traditional Augmentation	Conditional Diffusion Augmentation
Enhanced Accuracy	98.49%	98.19%	89.30%	97.29%
Absolute Improvement	1.52%	1.45%	1.45%	3.92%
Generalization Scenario	Missing Value/Fixed Window Completion	Long Time-Series Gesture Generalization	Single-User Action Generalization	Few-Shot Generalization

4.3. Cross-Scenario Recognition Enhancement Mechanism

Aiming at the domain shift problem in cross-scenario recognition, the enhancement mechanisms combining diffusion models and domain adaptive learning were reported in existing literature [16] [31]. First, diffusion models are adopted to generate cross-scenario samples, construct a mixed dataset covering source and target scenarios, and expand the model's feature learning range [32]. Also, contrastive domain adaptive learning methods are introduced to improve the model's ability to extract domain-invariant features by narrowing the feature distance of the same type of actions in different scenarios [33]. Environment-modules are utilized to dynamically adjust model parameters, which may be adapted to different scenarios because the characteristics of CSI signals are easily affected by environmental obstacle distribution and signal propagation path changes [19]. The noise intensity and conditional weights are adjustable for sample generation to strengthen the model's adaptability to complex channel environments. Hao *et al.*'s domain-invariant pose estimation method [14] offers a valuable reference for cross-scenario feature learning and further improves the model's environmental adaptability.

The mechanism does not require labeled data from the target scenario and effectively improves cross-scenario recognition performance through the collaborative effect of generative sample expansion and domain adaptive learning [3]. Studies have reported that such methods can improve recognition accuracy by 8% - 12% compared with traditional transfer learning methods in cross-scenario settings such as Room1-Room2 and Room1-Room3, reaching above 85%. Shi *et al.*'s AFEE-MatNet scheme also achieved excellent performance in cross-scenario recognition [20], and its feature enhancement and matching network design ideas can supplement this mechanism. Gu *et al.* [34] clearly pointed out that robust feature learning is the core research direction in the field, which is highly consistent with the core demand for mining domain-invariant features for cross-scenario action recognition. The cross-scenario samples generated by diffusion models can provide sufficient data support for mining domain-invariant features.

4.4. Model Robustness and Efficiency Optimization

To improve the model's robustness to weak actions and noise interference, a strategy integrating diffusion models and signal processing technologies has been adopted in related studies [9]. CSI features of weak actions through amplitude extraction and noise suppression generate synthetic samples consistent with the real sample distribution, combined with conditional diffusion models to enhance the model's anti-interference ability in data-scarce scenarios [17]. A noise-adaptive loss function is introduced during model training to improve the model's adaptability to environmental noise [9] [20]. An unsupervised statistical feature-guided diffusion model is adopted to generate diverse synthetic samples using statistical information such as mean, standard deviation, and Z-score, further alleviating insufficient feature learning in few-shot scenarios [30]. Combined with the attention mechanism [10], the model's capture of key action features can be strengthened, and weak action recognition accuracy is improved.

In terms of model efficiency optimization, model pruning and quantization technologies have been applied to reduce the computational complexity of diffusion models and recognition models. Pruning removes redundant parameters in the model, and quantization converts model parameters from 32-bit floating-point to 8-bit integer, reducing model storage overhead and inference delay while ensuring performance loss does not exceed 3% [35]. For edge device deployment requirements, designing a lightweight diffusion model architecture (such as compressing high-dimensional RF data to low-dimensional latent through R-VAE) can optimize model parameters and computational efficiency and meet real-time requirements [3] [24]. Lee *et al.* apply a user-specific channel generation scheme [27] to reduce invalid sample generation through conditional constraints, improving data diversity while reducing training overhead. Deng *et al.*'s WiLDAR model [22] and El Zein *et al.*'s HAR-LightCNN [5] both provide mature experience for lightweight design, which can provide a reference for the efficiency optimization of this scheme. For lightweight recognition models such as ImgFi [21], diffusion models can be used to generate CSI image samples adapted to their input format, improving recognition performance without increasing model complexity.

5. Research Challenges and Future Directions

5.1. Main Research Challenges

Although diffusion models provide an effective enhancement approach for CSI action recognition, they still face the following key challenges. First, it is difficult to balance the generation efficiency, fidelity, and deployment of diffusion models. Fast sampling may lead to the loss of sample details, while high-fidelity generation requires more computing resources, has certain requirements for the computing power of hardware platforms, and is difficult to meet the requirements of real-time and lightweight deployment [6] [24]. Second, deep mitigation of cross-scenario domain shift is still difficult. CSI data distribution changes drastically in complex environments (such as multiple obstacles and dynamic interference) [3], and individual user differences and device heterogeneity further aggravate domain shift [20]. A single conditional guidance strategy is difficult to effectively alleviate cross-scenario domain shift, limiting the model generalization ability. Third, feature enhancement of weak actions still faces more challenges. The CSI perturbation signals corresponding to actions such as breathing and micro-postures are weak and easily submerged by noise, making it difficult for diffusion models to generate such samples [9].

5.2. Future Research Directions

In response to the above challenges, future research can be carried out on the following aspects. First, lightweight optimization of diffusion models should be explored. More efficient sampling strategies and network architectures, such as extracting core features through low-pass filtering, should be developed, reducing invalid calculations through adaptive neighbor manipulation, and combining structure-aware specification design to reduce model overhead [24], drawing on the lightweight design experience of WiLDAR [22] and HAR-LightCNN [5]. Second, the deepening of cross-scenario adaptive generation technology should be studied. A dynamic environment-aware diffusion model can adjust the generation strategy according to real-time scene parameters. Hao *et al.*'s domain-invariant feature learning method is useful [14]. Third, security and privacy protection should be emphasized. Wang *et al.* proposed the generative protection signal modulation technology to prevent unauthorized devices from using CSI for malicious sensing and protect user privacy [36]. Fourth, in-depth integration of self-supervised and few-shot learning is important. Xiao *et al.*'s time-series segmentation method [37] further alleviates the problem of scarce labeled data. In addition, Wang *et al.* applied AIGC technology to multi-platform radio frequency sensing data generation, providing a unified framework for cross-device action recognition [38]. The radio frequency signal synthesis method based on conditional diffusion effectively improves few-shot sensing performance [39]. In the future, the above ideas can be combined with the scheme in this paper to further enhance the system generalization ability.

6. Conclusion

This paper presents a systematic review of research advances in diffusion model-based data augmentation for WiFi channel state information (CSI) human activity recognition. A clear distinction is made between existing literature findings and the integrated conceptual framework proposed in this study. Device-free activity recognition based on CSI features presents several advantages, such as non-intrusive deployment, low cost, and strong privacy protection. Critical limitations still exist in practical deployment, including limited labeled training data, significant cross-scene domain shift, and difficulties in weak motion perception. Diffusion models support high-fidelity time-series generation, stable training, and favorable generalization. Such models provide effective technical solutions to the above bottlenecks. This paper reviews the sensing mechanism, core challenges, and technical adaptability of diffusion models in CSI activity recognition. Mainstream technical paradigms are summarized, covering sample generation, cross-domain adaptation, and model robustness and efficiency optimization. Existing research challenges and future directions are outlined, including the trade-off between generation efficiency and fidelity, deep domain shift mitigation, and weak motion feature enhancement. This paper offers references and insight for further data augmentation and engineering application of CSI activity recognition in few-shot and cross-domain scenarios.

Funding

The work is funded by the Foundation of the Innovation and Entrepreneurship

Training Program for College Students (S202510424022).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Yadav, S.K., Sai, S., Gundewar, A., Rathore, H., Tiwari, K., Pandey, H.M., *et al.* (2022) CSITime: Privacy-Preserving Human Activity Recognition Using WiFi Channel State Information. *Neural Networks*, **146**, 11-21. https://doi.org/10.1016/j.neunet.2021.11.011

[2] Wang, D., Yang, J., Cui, W., Xie, L. and Sun, S. (2021) Multimodal CSI-Based Human Activity Recognition Using GANs. *IEEE Internet of Things Journal*, **8**, 17345-17355. https://doi.org/10.1109/jiot.2021.3080401

[3] Wang, Y., Yao, L., Wang, Y. and Zhang, Y. (2021) Robust CSI-Based Human Activity Recognition with Augment Few Shot Learning. *IEEE Sensors Journal*, **21**, 24297-24308. https://doi.org/10.1109/jsen.2021.3111030

[4] Sheng, B., Xiao, F., Sha, L. and Sun, L. (2020) Deep Spatial-Temporal Model Based Cross-Scene Action Recognition Using Commodity WiFi. *IEEE Internet of Things Journal*, **7**, 3592-3601. https://doi.org/10.1109/jiot.2020.2973272

[5] El Zein, H., Mourad-Chehade, F. and Amoud, H. (2024) CSI-Based Human Activity Recognition via Lightweight CNN Model and Data Augmentation. *IEEE Sensors Journal*, **24**, 25060-25069. https://doi.org/10.1109/jsen.2024.3414168

[6] Xiao, C., Han, Y., Yang, W., Hou, Y., Shi, F. and Chetty, K. (2024) Diffusion-Model-Based Contrastive Learning for Human Activity Recognition. *IEEE Internet of Things Journal*, **11**, 33525-33536. https://doi.org/10.1109/jiot.2024.3429245

[7] Lee, T., Park, J., Kim, H. and Andrews, J.G. (2025) Generating High Dimensional User-Specific Wireless Channels Using Diffusion Models. *IEEE Transactions on Wireless Communications*, **25**, 2907-2921. https://doi.org/10.1109/TWC.2025.3600286

[8] Zhang, Y., He, F., Wang, Y., Wu, D. and Yu, G. (2023) CSI-Based Cross-Scene Human Activity Recognition with Incremental Learning. *Neural Computing and Applications*, **35**, 12415-12432. https://doi.org/10.1007/s00521-023-08389-0

[9] Yang, X., Zhang, T., Zhou, M., Chen, B., Wang, Y. and Tong, R. (2026) CSI-Based Respiratory Pattern Recognition Using Diffusion Model and Deep Learning. *IEEE Transactions on Network Science and Engineering*, **13**, 3894-3910. https://doi.org/10.1109/TNSE.2025.3638894

[10] Chen, Z., Zhang, L., Jiang, C., Cao, Z. and Cui, W. (2019) WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM. *IEEE Transactions on Mobile Computing*, **18**, 2714-2724. https://doi.org/10.1109/tmc.2018.2878233

[11] Shalaby, E., ElShennawy, N. and Sarhan, A. (2022) Utilizing Deep Learning Models in CSI-Based Human Activity Recognition. *Neural Computing and Applications*, **34**, 5993-6010. https://doi.org/10.1007/s00521-021-06787-w

[12] Abdelnasser, H., Youssef, M. and Harras, K.A. (2015) WiGest: A ubiquitous WiFi-Based Gesture Recognition System. 2015 *IEEE Conference on Computer Communications* (*INFOCOM*), Hong Kong, 26 April-1 May 2015, 1472-1480. https://doi.org/10.1109/infocom.2015.7218525

[13] Li, M., Meng, Y., Liu, J., Zhu, H., Liang, X., Liu, Y., *et al.* (2016) When CSI Meets Public WiFi: Inferring Your Mobile Phone Password via WiFi Signals. *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security*,

Vienna, 24-28 October 2016, 1068-1079. https://doi.org/10.1145/2976749.2978397

[14] Hao, Y., Nallanathan, A. and Liu, Y. (2025) Latent Diffusion-Inspired Domain-Invariant Human Pose Estimation from WiFi Signals. 2025 *IEEE*/*CIC International Conference on Communications in China* (*ICCC*), Shanghai, 10-13 August 2025, 1-6. https://doi.org/10.1109/iccc65529.2025.11149317

[15] Niu, J., He, X., Fang, B., Han, G., Wang, X. and He, J. (2024) CSI-F: A Human Motion Recognition Method Based on Channel-State-Information Signal Feature Fusion. *Sensors*, **24**, Article No. 862. https://doi.org/10.3390/s24030862

[16] Huang, S., Chen, P. and McCann, J. (2023) DiffAR: Adaptive Conditional Diffusion Model for Temporal-Augmented Human Activity Recognition. *Proceedings of the* 32*nd International Joint Conference on Artificial Intelligence*, Macao, 19-25 August 2023, 3812-3820. https://doi.org/10.24963/ijcai.2023/424

[17] Xu, M., Li, Y., Han, K., Zhao, Y., Jiang, J. and Du, Y. (2024) A Conditional Diffusion Model Based WiFi Sensing Enhancement Method. 2024 *IEEE* 35*th International Symposium on Personal*, *Indoor and Mobile Radio Communications* (*PIMRC*), Valencia, 2-5 September 2024, 1-6. https://doi.org/10.1109/pimrc59610.2024.10817390

[18] Chen, Z., Shin, H. and Nallanathan, A. (2025) Generative Diffusion Model-Based Variational Inference for MIMO Channel Estimation. *IEEE Transactions on Communications*, **73**, 9254-9269. https://doi.org/10.1109/tcomm.2025.3556753

[19] Kobir, M.I., Machado, P., Lotfi, A., Haider, D. and Ihianle, I.K. (2025) Enhancing Multi-User Activity Recognition in an Indoor Environment with Augmented Wi-Fi Channel State Information and Transformer Architectures. *Sensors*, **25**, Article No. 3955. https://doi.org/10.3390/s25133955

[20] Shi, Z., Cheng, Q., Zhang, J.A. and Yi Da Xu, R. (2022) Environment-Robust WiFi-Based Human Activity Recognition Using Enhanced CSI and Deep Learning. *IEEE Internet of Things Journal*, **9**, 24643-24654. https://doi.org/10.1109/jiot.2022.3192973

[21] Zhang, C. and Jiao, W. (2023) Imgfi: A High Accuracy and Lightweight Human Activity Recognition Framework Using CSI Image. *IEEE Sensors Journal*, **23**, 21966-21977. https://doi.org/10.1109/jsen.2023.3296445

[22] Deng, F., Jovanov, E., Song, H., Shi, W., Zhang, Y. and Xu, W. (2024) WiLDAR: WiFi Signal-Based Lightweight Deep Learning Model for Human Activity Recognition. *IEEE Internet of Things Journal*, **11**, 2899-2908. https://doi.org/10.1109/jiot.2023.3294004

[23] Ding, J. and Wang, Y. (2019) WiFi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network. *IEEE Access*, **7**, 174257-174269. https://doi.org/10.1109/access.2019.2956952

[24] Xiao, C., Tong, Y., Lu, J., Xue, Y., Tai, W., Cheng, Z., *et al.* (2025) Graph-Enhanced Multi-Scale Contrastive Learning for Graph Anomaly Detection with Adaptive Diffusion Models. *IEEE Transactions on Big Data*, **11**, 2946-2957. https://doi.org/10.1109/tbdata.2025.3566619

[25] Zhang, R., Jing, X., Wu, S., Jiang, C., Mu, J. and Yu, F.R. (2021) Device-Free Wireless Sensing for Human Detection: The Deep Learning Perspective. *IEEE Internet of Things Journal*, **8**, 2517-2539. https://doi.org/10.1109/jiot.2020.3024234

[26] Kim, H., Lee, T., Kim, H., De Veciana, G., Arfaoui, M.A., Koc, A., *et al.* (2025) Generative Diffusion Model-Based Compression of MIMO CSI. *ICC* 2025-*IEEE International Conference on Communications*, Montreal, 8-12 June 2025, 6323-6328. https://doi.org/10.1109/icc52391.2025.11161629

[27] Lee, T., Park, J., Kim, H. and Andrews, J.G. (2024) High Dimensional User-Specific Channel Generation for CSI Compression Using Diffusion Models. 2024 58*th Asilomar Conference on Signals*, *Systems*, *and Computers*, Pacific Grove, 27-30 October 2024, 759-763. https://doi.org/10.1109/ieeeconf60004.2024.10942761

[28] Khan, D. and Ho, I.W. (2021) Deep Learning of CSI for Efficient Device-Free Human Activity Recognition. 2021 *IEEE* 7*th World Forum on Internet of Things* (*WF-IoT*), New Orleans, 14 June-31 July 2021, 19-24. https://doi.org/10.1109/wf-iot51360.2021.9595661

[29] Wang, Z. and Mao, S. (2024) AIGC for RF Sensing: The Case of RFID-Based Human Activity Recognition. 2024 *International Conference on Computing*, *Networking and Communications* (*ICNC*), Big Island, 19-22 February 2024, 1092-1097. https://doi.org/10.1109/icnc59896.2024.10556013

[30] Zuo, S., Fortes, V., Suh, S., Sigg, S. and Lukowicz, P. (2023) Unsupervised Diffusion Model for Sensor-Based Human Activity Recognition. *Adjunct Proceedings of the* 2023 *ACM International Joint Conference on Pervasive and Ubiquitous Computing & the* 2023 *ACM International Symposium on Wearable Computing*, Cancun, 8-12 October 2023, 205. https://doi.org/10.1145/3594739.3610797

[31] Wang, Z. and Mao, S. (2025) AIGC for Wireless Sensing: Diffusion-Empowered Human Activity Sensing. *IEEE Transactions on Cognitive Communications and Networking*, **11**, 657-671. https://doi.org/10.1109/tccn.2025.3525588

[32] Li, Z., Shi, Y., Gao, Y. and Xu, D. (2026) Diffusion-Based Data Augmentation for Image Recognition: A Systematic Analysis and Evaluation. *International Journal of Computer Vision*, **134**, Article No. 126. https://doi.org/10.1007/s11263-026-02754-x

[33] Custance, O., Khan, S., Parkinson, S. and Sheng, Q.Z. (2026) Parameter-Efficient Domain Adaption for CSI Crowd-Counting via Self-Supervised Learning with Adapter Modules. arXiv:2601.02203.

[34] Gu, F., Chung, M., Chignell, M., Valaee, S., Zhou, B. and Liu, X. (2021) A Survey on Deep Learning for Human Activity Recognition. *ACM Computing Surveys*, **54**, 1-34. https://doi.org/10.1145/3472290

[35] Erak, O. and Abou-Zeid, H. (2023) Accelerating and Compressing Deep Neural Networks for Massive MIMO CSI Feedback. *ICC* 2023-*IEEE International Conference on Communications*, Rome, 28 May-1 June 2023, 1029-1035. https://doi.org/10.1109/icc45041.2023.10279462

[36] Wang, J., Feng, C., Sun, G., Du, H., Niyato, D., Quek, T.Q.S., *et al.* (2025) Security Enhancement of CSI-Based Wireless Sensing via Generative AI. 2025 *IEEE International Workshop on Radio Frequency and Antenna Technologies* (*iWRF* & *AT*), Shenzhen, 23-26 May 2025, 83-88. https://doi.org/10.1109/iwrfat65352.2025.11103416

[37] Xiao, C., Chen, S., Zhou, F. and Wu, J. (2022) Self-Supervised Few-Shot Time-Series Segmentation for Activity Recognition. *IEEE Transactions on Mobile Computing*, **22**, 6770-6783. https://doi.org/10.1109/tmc.2022.3199015

[38] Wang, Z., Yang, C. and Mao, S. (2025) AIGC for Rf-Based Human Activity Sensing. *IEEE Internet of Things Journal*, **12**, 3991-4005. https://doi.org/10.1109/jiot.2024.3482256

[39] Wang, Z. and Mao, S. (2024) AIGC for Wireless Data: The Case of RFID-Based Human Activity Recognition. *ICC* 2024-*IEEE International Conference on Communications*, Denver, 9-13 June 2024, 4060-4065. https://doi.org/10.1109/icc51166.2024.10622401