WiOpen: A Robust Wi-Fi-based Open-set Gesture Recognition Framework

Xiang Zhang *Member, IEEE,* Jinyang Huang* *Member, IEEE,* Huan Yan, Yuanhao Feng, Peng Zhao, Guohang Zhuang, Zhi Liu *Senior Member, IEEE*, and Bin Liu*

***Abstract*—Recent years have witnessed a growing interest in Wi-Fi-based gesture recognition. However, existing works have predominantly focused on closed-set paradigms, where all testing gestures are predefined during training. This poses a significant challenge in real-world applications, as unseen gestures might be misclassified as known classes during testing. To address this issue, we propose WiOpen, a robust Wi-Fi-based Open-Set Gesture Recognition (OSGR) framework. Implementing OSGR requires addressing challenges caused by the unique uncertainty in Wi-Fi sensing. This uncertainty, resulting from noise and domains, leads to widely scattered and irregular data distributions in collected Wi-Fi sensing data. Consequently, data ambiguity between classes and challenges in defining appropriate decision boundaries to identify unknowns arise. To tackle these challenges, WiOpen adopts a two-fold approach to eliminate uncertainty and define precise decision boundaries. Initially, it addresses uncertainty induced by noise during data preprocessing by utilizing the CSI ratio. Next, it designs the OSGR network based on an uncertainty quantification method. Throughout the learning process, this network effectively mitigates uncertainty stemming from domains. Ultimately, the network leverages relationships among samples' neighbors to dynamically define open-set decision boundaries, successfully realizing OSGR. Comprehensive experiments on publicly accessible datasets confirm WiOpen's effectiveness. Code is available at https://github.com/purpleleaves007/WiOpen.**

***Index Terms*—Wi-Fi, Gesture Recognition, Open-Set Recognition, CSI, Uncertainty Reduction.**

I. Introduction

Wi-Fi based gesture recognition [1] has garnered significant attention in recent years due to its advantages in terms of ubiquitous deployment and non-intrusive sensing. However, current studies in the field all rely on a closed-set assumption [2], [3], *i.e.*, each test sample is assumed to always belong to one of the pre-defined set of gesture classes. Although this conventional presumption often proves untenable in practical applications, as gesture recognition systems can invariably encounter unseen gesture classes or even non-gestural activities, close-set techniques tend to force unknown class samples to be classified into one of the known gesture classes. This limitation not only results in an poor user experience but also undermines the practicability and reliability of Wi-Fi gesture recognition systems. Therefore, it is imperative to address this drawback and develop more robust and flexible open-set gesture recognition (OSGR) approaches that can handle open-set scenarios effectively. Such methods should properly classify unknown-class samples as "unknown" and known-class samples as one of the known classes.

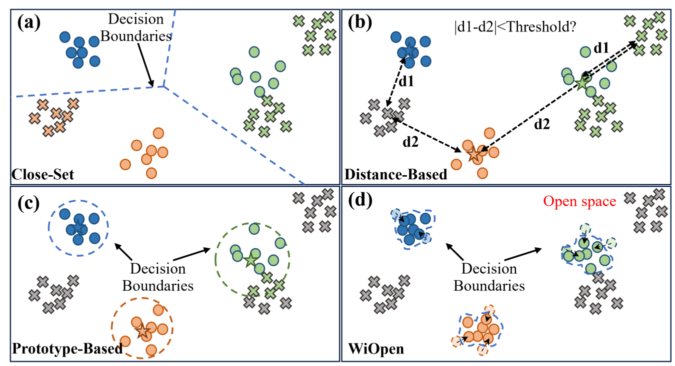

Fig. 1. The comparison between different approaches. (a) Close-set methods; (b) Distance-based methods; (c) Prototype-based methods; (d) Wiopen. Colored circles represent known classes samples, gray "x" symbols, colored "x" symbols, and stars represents unknown samples, misclassified samples and prototypes, respectively.

The objective of closed-set gesture recognition is to minimize empirical risk, which relates to the risk associated with misclassifying known classes. In contrast, OSGR not only focuses on minimizing empirical risk but also addresses open space risk. In real-world scenarios, each class's associated feature space is finite, while the area beyond these feature spaces is referred to as the open space, as depicted in Figure 1d. Labeling samples within the open space as known classes introduces open space risk. Traditional closed-set classifiers [3] usually divide the entire feature space into several known classes, and the decision boundaries of such solutions are shown in Figure 1a. These methods accept data from infinitely wide regions, implying that their open-space risk is unbounded. Recently, some studies have briefly touched upon the identification of unknowns in Wi-Fi sensing. For instance, Wione [4] constructed prototypes for all users for user identification and classified samples as unknown when their distance to the prototypes of all known classes

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62462015), in part by the Anhui Province Science Foundation for Youths (Grant No. 2308085QF230), and in part by Major scientific and technological project of Anhui Provincial Science and Technology Innovation Platform (Grant No. 202305a12020012).

Xiang Zhang, and Bin Liu, are with School of Cyber Science and Technology, University of Science and Technology of China, Hefei, 230026, China.

Jinyang Huang, Peng Zhao, and Guohang Zhuang, School of Computer and Information, Hefei University of Technology, Hefei, 230601, China.

Yuanhao Feng, Department of Computing, The Hong Kong Polytechnic University, Hong Kong, 100872, China.

Huan Yan, School of Big Data and Computer Science, Guizhou Normal University, Guiyang, 550025, China.

Zhi Liu, Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, 1828585, Japan.

* Corresponding authors.

was similar. Similarly, CAUTION [5] compared the distance between the sample and the two nearest user prototypes, classifying it as unknown when the distances were similar. Nevertheless, as illustrated in Figure 1b, these distance-based approaches are only effective for unknown samples in close proximity to the decision boundaries. In another examples, Tan et al. [6] and Multitrack [7] developed prototypes for each activity class and employed similarity comparisons for activity recognition. Samples falling below a certain similarity threshold were labeled as unknown. While prototype-based approaches construct decision boundaries for rejecting unknowns based on distance, these boundaries cannot flexibly envelop the feature space of each class, leading to misjudgments, such as green class in Figure 1c. Based on the observation that unknown samples mostly distribute around the edges of known sample clusters. Widar3 [2] utilizes a KNN-based algorithm to detect unknown samples, allowing for a more flexible decision boundary. However, these existing approaches primarily focus on improving the partitioning of decision boundaries to identify unknown samples, overlooking how to mitigate open space risk at its root and achieve a balance between open space risk and empirical risk. In contrast, this paper explores the inherent uncertainty [8] in Wi-Fi sensing and its connection to open space risk, emphasizing the investigation of the sources of open space risk in Wi-Fi OSGR and the learning of better features to achieve OSGR.

The uncertainty in Wi-Fi sensing arises from both noise and domain variability. Noise introduces random biases to the Wi-Fi sensing signals, resulting in random offsets in the data distribution and making the distribution more discrete. Domain refers to variables such as users, the location and direction of activities. According to previous studies [9], variations in the domain can cause different mappings of activities onto Wi-Fi sensing signals, leading to directed shifts in the data. The uncertainty in Wi-Fi sensing presents two significant challenges for implementing OSGR based on Wi-Fi. The first challenge emerges from the high data dispersion caused by uncertainty, leading to confusion between classes with substantial intra-class variation and limited inter-class variation. The second challenge arises from the irregularity in data distribution, making it more complex to define decision boundaries for distinguishing unknown from known samples.

In this paper, we present WiOpen, a robust Wi-Fi-based OSGR framework. To address the first challenge, WiOpen conducts an extensive analysis of the open-set challenges encountered in Wi-Fi gesture recognition. It introduces an uncertainty quantification method as a pivotal step. Leveraging this analysis, WiOpen takes measures to mitigate the uncertainty caused by noise during data preprocessing, achieved through the utilization of the CSI ratio. Then, the OSGR network inspired by uncertainty quantification comes into play. This network learns the relationship between sensing data and Doppler frequency shift (DFS), effectively eliminating the influence stemming from static path factors. Simultaneously, it unravels the intrinsic structure of Wi-Fi data characterized by a highly irregular distribution by analyzing the relationships between neighboring samples, thereby reducing the uncertainty associated with domains. With respect to the second challenge, WiOpen suggests employing the sample's neighborhood structure rather than the structure of the sample and prototype as the criterion for decision boundary construction. The decision boundaries constructed based on distances from samples to a subset of neighbors exhibit greater flexibility compared to those constructed using distances from samples to prototypes. Additionally, they can be integrated with neighbor-based feature learning stages to form an end-to-end learning and decision process. This allows for direct adjustment of the neighbor structure feature learning process based on decision results, making it more targeted in feature learning compared to methods that only use KNN for decision-making [2], as illustrated in Figure 1d.

We have implemented WiOpen and conducted a comprehensive evaluation using publicly available datasets and real-world experiments. The key contributions are as follows:

- We conduct an in-depth analysis of the open-set challenges within Wi-Fi-based gesture recognition, elucidating the relationship between uncertainty and Wi-Fi OSGR performance.
- We propose a uncertainty quantification method and an OSGR network inspired by uncertainty quantification, designed to mitigate uncertainty and achieve effective Wi-Fi OSGR.
- We implement WiOpen and conduct extensive experiments using publicly datasets and real-world experiments to evaluate its performance. Evaluations demonstrate the feasibility and effectiveness of our system.

II. Related Works

A. Wi-Fi based Gesture Recognition

Gesture sensing and recognition enabled by Wi-Fi [10]–[12] can be broadly categorized into two groups: handcraft-based and deep learning-based methods. Handcraft-based approaches typically involve manual characterization of signal distortions corresponding to different gestures. WiGest [13] employs manual pattern construction for each gesture in received signal strength (RSS) and uses a similarity matching method for recognition. While this work is inspiring, its reliance on coarse-grained RSS indicators limits its accuracy. WiMU [14] goes further by achieving multi-user gesture recognition using fine-grained Channel State Information (CSI). WiDraw [15] employs Angle-Of-Arrival (AOA) measurements for hand tracking, allowing users to draw in the air with minimal tracking error. QGesture [16] utilizes phase information for similar performance but requires knowledge of the initial hand position for tracking.

Learning-based methods shift the focus to automatic pattern recognition through data-driven techniques. This category can be further divided into shallow learning [17]–[19] and deep learning [20]. Shallow learning involves training a shallow learner with handcrafted features for gesture classification. Wikey [18] was one of the first to explore keystroke recognition based on Wi-Fi sensing and machine learning. WiFinger [17] employs Wi-Fi CSI for recognizing nine sign languages. While shallow learning typically requires a small training dataset, its performance is limited. Consequently, deep learning has emerged as an effective alternative. For example, WiSign [20] focuses on American Sign Language recognition, using amplitude and phase CSI profiles processed by a Deep Belief Network (DBN) for recognition. However, deep learning-based approaches face a critical challenge, namely, their dependence on domain-specific training.

WiDar3 [2] addresses this challenge by introducing a domain-independent feature, BVP, which characterizes power distribution across various velocities for cross-domain gesture recognition. Building upon the WiDar3.0 dataset, WiHF [21] derives a domain-independent motion change pattern for arm gestures, providing unique features for cross-domain recognition. WiSGP [22] employs a data augmentation-based approach to augment data and domain information, achieving a Wi-Fi gesture recognition system with domain-generalization capabilities. WiGRUNT [1], PAC-CSI [23], on the other hand, focus on attention mechanisms designed to automatically uncover critical information for gesture recognition. Wi-learner [24] utilizes autoencoders to enable small-sample cross-domain Wi-Fi gesture recognition.

B. Identify Unknows in Wi-Fi Sensing

The identification of unknown classes can be achieved by utilizing the scores provided by classifiers [25]. For instance, setting a threshold on classification scores from traditional machine learning models [26] or deep softmax learners [27] can be used to determine whether an input sample belongs to an unknown class. Nevertheless, this approach is most suitable for identifying unknown samples situated near the classifier's decision boundary [28].

In Wi-Fi sensing, some efforts have touched upon the challenge of distinguishing unknown classes. However, these efforts are mentioned as a part of a larger study and lack in-depth analysis of this specific challenge. Existing studies on Wi-Fi-based unknown class recognition primarily utilize the prototype-based method. [29]–[32]. In the works of Wione [4] and CAUTION [5], a conventional softmax classifier is used for user identity classification. After model training, they summarize a prototype for each class. When a sample is equidistant from prototypes of two or more classes, it is classified as an unknown user. Tan et al. [6] and Multitrack [7] directly learn a prototype for each activity class. Activity recognition is achieved by comparing the similarity of input samples with prototypes of all classes. When the lowest similarity with all prototypes falls below a certain threshold, the sample is classified as unknown. Compared to such methods, Widar3 [2] utilizes a KNN-based algorithm to detect unknown samples, allowing for a more flexible decision boundary.

III. Preliminaries

A. Open-Set Recognition and Open Space Risk

Open-set recognition (OSR) pertains to situations where new, previously unseen classes emerge during testing. In such cases, a classifier should not only accurately classify known classes but also effectively reject unknown ones. Considering a training set $\mathbf{D_L} = \{(x_1, y_1), ...(x_n, y_n)\}$ of n labeled gesture samples, where x_i represents each sample and $y_i \in \{1...Y\}$ denotes the label of x_i, and a testing dataset $\mathbf{D_T} = \{(t_1), ...(t_u)\}$ where the label of t_i belongs to $\{1...Y + Q\}$, with Q representing the number of unknown classes typical in real-world scenarios. The open space, denoted as $\mathcal{O}$, encompasses regions far from the known classes. And the degree of openness, which quantifies the open space in the OSR, can be described as $\mathcal{P}$ [33]:

$$\mathcal{P} = 1 - \sqrt{\frac{2 \times Y}{2 * Y + Q}} \tag{1}$$

Inevitably, designating any sample in the open space as a known class introduces risks, known as open space risk ($\mathcal{R}_o$). Qualitatively, $\mathcal{R}_o$ can be described as the relative measure of the open space $\mathcal{O}$ in comparison to the overall measurement space $\mathcal{M}$ [26]:

$$\mathcal{R}_{\mathcal{O}} = \frac{\int_{\mathcal{O}} f(x)dx}{\int_{\mathcal{M}} f(x)dx} \tag{2}$$

here, $f(x)$ represents the measurable recognition function, where $f(x) = 1$ denotes that a certain class within the known classes has been recognized, otherwise $f(x) = 0$. In other words, the more samples from the open space are classified as known classes, the higher $\mathcal{R}_{\mathcal{O}}$ becomes.

In the context of OSR and considering the concepts of open space risk and openness, the fundamental requirement for addressing the OSR problem is to determine a recognition function, denoted as $f(x)$, that minimizes the following open-set risk:

$$arg \min_f \{\mathcal{R}_{\mathcal{O}}(f, \mathbf{D_U}) + \lambda_r \mathcal{R}_{\mathcal{E}}(f, \mathbf{D_L})\} \tag{3}$$

here, λ_r serves as a regularization constant, and $\mathcal{R}_{\mathcal{O}}$ and $\mathcal{R}_{\mathcal{E}}$ represent the open space risk and the empirical risk (the risk associated with incorrectly assigning known samples), respectively. $\mathbf{D_L}$ denotes the set of known labeled training data, and $\mathbf{D_U}$ represents the potentially unknown data.

B. When Wi-Fi Based Gesture Recognition Meets Open-Set Challenge

Current Wi-Fi based sensing solutions predominantly rely on Channel State Information (CSI) [34]–[36]. CSI characterizes the signal attenuation that occurs as signals propagate through a given medium. This attenuation can be expressed through the following equation:

$$Y = HX + \mathcal{N} \tag{4}$$

where Y and X are the received and transmitted signal vectors, respectively. $\mathcal{N}$ is additive white Gaussian noise, and H stands for the channel matrix representing the CSI.

CSI can be delineated as the combination of two main components: static CSI and dynamic CSI. Static CSI is influenced by the surrounding environment and the presence of a Line-of-Sight (LoS) between the transceivers. Dynamic CSI, on the other hand, is primarily determined by the reflection path from moving objects:

$$H(r, t) = H_s(r, t) + H_d(r, t) \tag{5}$$

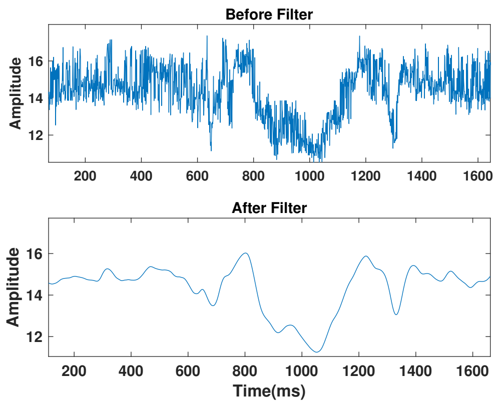

Fig. 2. Examples of samples with different noise uncertainty, where Ψ_n before filtering is 1.92, and Ψ_n after filtering is 0.96.

where r and t represent the signal frequency and the timestamp, respectively. The dynamic CSI can be further elaborated as:

$$H_d(r,t) = \sum_{k \in \mathbf{D}} h_k(r,t) e^{-j2\pi \frac{d_k(t)}{\lambda_k}} \tag{6}$$

Here, $h_k(r,t)$, $d_k(t)$, and λ_k represent the attenuation, the path length of the dynamic path, and the wavelength associated with the k^{th} path, respectively. The set $\mathbf{D}$ encompasses dynamic paths. Notably, gestures induce changes in length of the dynamic paths, subsequently altering the overall CSI sequence. Traditional Wi-Fi-based gesture recognition techniques aim to extract gesture-related information from the overall CSI and consequently interpret and recognize the original gestures.

In contrast to close-set methods, Wi-Fi-based OSGR aims to accomplish the dual task of recognizing known gestures accurately while effectively identifying unknown gestures. Equation 3 highlights that the challenge in solving the Wi-Fi OSGR problem involves the simultaneous minimization of empirical classification risk for labeled known data and open-space risk for potential unknown data. Open-space risk, as quantified by Equation 2, and empirical risk can be expressed as follows:

$$\mathcal{R}_{\mathcal{E}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f(x_i), y_i) \tag{7}$$

here, N signifies the number of training samples, while x_i and y_i denote the i-th sample and its corresponding label. The loss function, $\mathcal{L}$, gauges the likelihood of the prediction $f(x_i)$ with respect to the true label y_i.

It is important to note that both $\mathcal{R}_{\mathcal{O}}$ and $\mathcal{R}_{\mathcal{E}}$ are inherently linked to the misclassification tendencies of the gesture recognition model, $f(x)$. Traditional recognition models typically transform input data from the data space to a feature space and next establish a decision boundary within this feature space to differentiate between distinct classes. Unfortunately, due to the impact of the unique uncertainty inherent in Wi-Fi sensing, both reducing data confusion and determining suitable decision boundaries face significant challenges.

We define uncertainty in Wi-Fi sensing as **the intrinsic variability, randomness, or ambiguity present in the available sensing data**. In a nutshell, uncertainty introduces a certain degree of bias to sensory data, resulting in a spatial displacement of samples in the data space, ultimately leading to irregular and discrete distributions of collected CSI samples. We posit that uncertainty in Wi-Fi sensing arises from two factors: noise and domains (such as distinct users or different gesture execution locations), and it can be expressed as:

$$\Psi = \Psi_n + \Psi_d \tag{8}$$

where Ψ, Ψ_n and Ψ_d represents the overall uncertainty, the uncertainty caused by noise and the uncertainty caused by domains.

A portion of the noise present in CSI arises from multipath effects. From equations 5 and 6, the received CSI is the summation of signals from multiple propagation paths, which makes the target path signal susceptible to interference from unrelated environmental signals [37]. Additionally, noise is also introduced by Wi-Fi devices themselves, contributing to Ψ_n:

$$\Psi_n = \Psi_n^e + \Psi_n^d \tag{9}$$

where Ψ_n^e and Ψ_n^d represent the noise uncertainty caused by environment and devices. Noise introduces random biases to the collected CSI samples, causing an expansion in the distribution range of these samples. This results in larger intra-class differences and smaller inter-class differences, ultimately making it easier for confusion to occur between classes.

Domain-induced uncertainty similarly introduces biases to the collected CSI samples. However, unlike noise uncertainty, the biases introduced by domain uncertainty are directional, closely related to the principles of Wi-Fi sensing. When there is no significant change in static paths, CSI variations are primarily driven by dynamic path changes caused by activities. According to the Fresnel zone theory [9], [38], [39], even if the user's gestures remain unchanged, variations in user positions and orientations concerning the transmitting and receiving antennas lead to corresponding changes in the dynamic path variations. Wi-Fi sensing is highly sensitive to domain changes, and as CSI data represents the superimposition of signals from multiple paths, it possesses a lower spatial resolution compared to sensory data in domains such as vision. Domain changes result in substantial data distribution shifts towards the corresponding directions. This type of uncertainty is referred to as Ψ_d:

$$\Psi_d = \Psi_d^u + \Psi_d^l + ...\Psi_d^{nd} \tag{10}$$

where Ψ_d^u, Ψ_d^d and Ψ_d^{nd} represent user, location and nd th domain. Such uncertainty is common in Wi-Fi based sensing, leading to a widely scattered and irregular distribution of CSI data [2], [21].

It can be observed that addressing the challenges of Wi-Fi-based OSGR requires eliminating the effects of uncertainty, and before eliminating uncertainty, it is advisable to quantify it. Noise adds random noise to CSI data, and the data distribution of this random noise should follow a normal distribution with an unknown variance. Therefore, quantifying noise uncertainty can be transformed into assessing the proportion of data in CSI that conforms to this normal distribution and the magnitude of the variance of this normal distribution. On the other hand, domain uncertainty results in directional biases in the data. When domain uncertainty is higher, the distance of an individual CSI sample to other samples of the same class increases, while the distance to samples of other classes decreases. Thus, it can be transformed into assessing the distances between all samples and other samples. Therefore, we propose the following formula to quantify uncertainty.

$$\begin{aligned}\Psi &= \Psi_n + \Psi_d \\ &= \frac{1}{N*G}\Big(\sum_{n=1}^{N}\Big(\sum_{k=1}^{G}(\alpha_{nk}-\bar{\alpha_n})^2+\sum_{k=1}^{G}\sigma_{nk}\Big)\Big) \\ &+ \frac{1}{N}\Big(\sum_{n=1}^{N}\sum_{i=1}^{N}d(x_n,x_i)-\sum_{n=1}^{N}\sum_{j=1}^{N}d(x_n,x_j)\Big)\end{aligned} \tag{11}$$

here, $i \neq n$ and $j \neq n$, $y_i \neq y_n$ and $y_j = y_n$.

The two terms in the formula above serve as measures of noise uncertainty and domain uncertainty. For the first term, we fit a CSI sample, denoted as x_n, into a Gaussian Mixture Model (GMM) with G components, leading to the following probability distribution:

$$p(x_{nt}) = \sum_{k=1}^{G}\alpha_{nk}\eta(x_k,\mu_{nk},\sigma_{nk}) \tag{12}$$

where α_{nk} represents the probability that the observed data belongs to the kth component, μ_{nk} and σ_{nk} represent the expectation and variance of the kth component, respectively. Here, η is the Gaussian probability density function and N is the number of samples. x_{nk} represents the reading at the kth timestamp of the nth sample. When x_n exhibits larger noise, it indicates stronger random noise. This results in one of K Gaussian components (representing random noise) having a higher α_{nk}. Moreover, greater noise leads to a more random data distribution, resulting in a larger σ_{nk}. Consequently, we design the first term to quantify Ψ_n, and the effect is shown in Figure 2.

The second term in Equation 11 is used to measure Ψ_d. In this term, y_n represents the label of x_n, and d denotes the distance metric function. The objective of this term is to gauge domain-related uncertainty by considering neighborhood relationships. In essence, when there is a lower probability that closer neighboring samples belong to the same class, uncertainty increases. Conversely, the higher the probability that a sample's neighbors originate from other classes, the greater Ψ_d becomes. We selected some data from the WiDar3.0 [2] dataset and presented their data distribution as shown in Figure 3. Figure 3a includes data from different users, while Figure 3b includes data from different users, directions, and positions. It can be observed that domain factors do increase the irregularity in data distribution. Our uncertainty quantification approach can also quantify domain uncertainty.

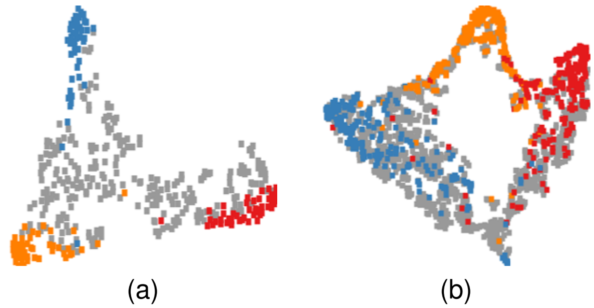

Fig. 3. Domain uncertainty.(a) 0.0901; (b) 0.2659. Colored points represent samples from known classes, while gray points represent samples from unknown classes.

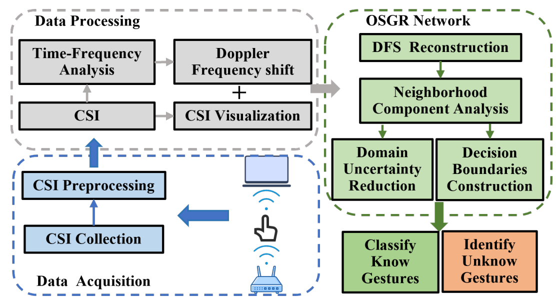

Fig. 4. Overview of the WiOpen Framework: The framework consists of sensory information being collected through the data acquisition module, processed by the data processing module, and subsequently input into the OSGR network to achieve gesture classification and unknown class identification.

IV. Method

Based on the previous analysis, effectively addressing the open-set challenge in Wi-Fi-based gesture recognition hinges on mitigating the impact of uncertainty in data/feature space. To mitigate uncertainty, we propose a method that involves designing strategies in both the data preprocessing and feature learning stages, each serving a distinct purpose. In the data preprocessing stage, WiOpen focuses on the reduction of noise interference. In the feature learning stage, our approach utilizes the uncertainty quantification method-inspired OSGR network to facilitate feature extraction and sample clustering in the feature space.

A. Overview of WiOpen

As shown in Figure 4, WiOpen is composed of three main components: data acquisition, data processing, and the OSGR network. The data acquisition component is responsible for obtaining data from Wi-Fi sensing devices and performing data preprocessing to eliminate some of the noise uncertainty. The data preprocessing component involves extracting DFS from CSI and visualizing it along with CSI amplitude and phase as input to the OSGR network. The OSGR network utilizes the reconstruction from CSI to DFS to eliminate influence of static paths, and learns the semantic neighborhood structure of samples to eliminate domain uncertainty and construct decision boundaries for open-set recognition.

B. Data Acquisition

The CSI is collected from wireless network cards capable of capturing CSI. As demonstrated in the previous section, the gesture can be portrayed by the change in CSI. However, with commodity Wi-Fi devices, there's a challenge due to unsynchronized transmitters and receivers, leading to a time-varying random noise $e^{-j\theta_n}$ [40]:

$$\begin{aligned} H(f,t) &= e^{-j\theta_n}(H_s(f,t)+H_d(f,t)) \\ &= e^{-j\theta_n}(H_s(f,t)+A(f,t)e^{-j2\pi\frac{d(t)}{\lambda}}) \end{aligned} \tag{13}$$

where $A(f,t)$, $e^{-j2\pi\frac{d(t)}{\lambda}}$ and $d(t)$ denote the complex attenuation, phase shift and path length of dynamic components, respectively. This random phase noise, $e^{-j\theta_n}$, hinders the direct use of CSI phase information and increase noise uncertainty.

Therefore, we need to eliminate $e^{-j\theta_n}$. Fortunately, for commodity Wi-Fi cards, this noise remains constant across different antennas on the same Wi-Fi Network Interface Card (NIC) because they share the same RF oscillator. This noise can be eliminated using the CSI-ratio model [41]:

$$\begin{aligned} H_q(f,t) &= \frac{H_1(f,t)}{H_2(f,t)} \\ &= \frac{e^{-j\theta_n}(H_{s,1}+A_1e^{-j2\pi\frac{d_1(t)}{\lambda}})}{e^{-j\theta_n}(H_{s,2}+A_1e^{-j2\pi\frac{d_2(t)}{\lambda}})} \\ &= \frac{A_1e^{-j2\pi\frac{d_1(t)}{\lambda}}+H_{s,1}}{A_2e^{-j2\pi\frac{d_1(t)+\triangle d}{\lambda}}+H_{s,2}} \end{aligned} \tag{14}$$

where $H_1(f,t)$ and $H_2(f,t)$ are the CSI of two receiving antennas. When two antennas are close to each other, $\triangle d = d_2(t)-d_1(t)$ can be regarded as a constant. According to Mobius transformation, equation 14 represents transformations such as scaling and rotation of the phase shift $e^{-j2\pi\frac{d_1(t)}{\lambda}}$ of antenna 1 in the complex plane, and these transformations will not affect the changing trend of the CSI. Following the CSI ratio-processing, we further mitigate noise uncertainty by subjecting the data to a low-pass filter for the removal of environmental noise. Next, the processed data undergoes further refinement in the data processing section.

C. Data Processing

The data processing section involves visualizing the CSI and extracting DFS, which are fundamental for providing high-quality inputs to the OSGR network. Specifically, to ensure that the OSGR network receives normalized and information-rich inputs, we adopt the CSI visualization technique detailed in [1]. This method visualizes both the amplitude and phase of CSI separately and then integrates them into a single image. The advantage of this approach is that it consolidates the various dimensions of CSI information into a two-dimensional matrix, which is suitable for processing using a Convolutional Neural Network (CNN)-based network.

In contrast to previous methods, WiOpen derives DFS from data processed with CSI ratio, as opposed to CSI conjugate multiplication [2]. The CSI ratio method is advantageous because it represents transformations such as rotation and scaling of CSI in the complex plane, thus avoiding some negative effects introduced by $H_2(f,t)$. To further mitigate cumulative error effects caused by $\triangle d$, we apply an antenna selection coefficient s_a to select $H_1(f,t)$ and $H_2(f,t)$. This coefficient is calculated as follows:

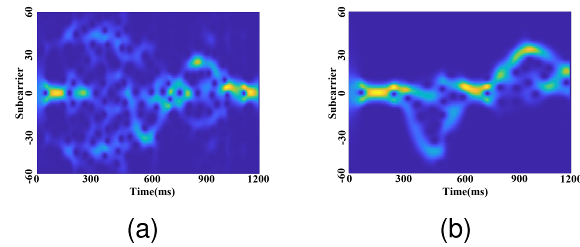

Fig. 5. The DFS visualization for a sample of "push & pull" gesture.(a) Widar3.0 [2]; (b) Wiopen.

$$s_a = \frac{1}{C}\sum_{c=1}^{C}\frac{var(|H_a(f_c,t))|}{mean(|H_a(f_c,t)|)} \tag{15}$$

where var and $mean$ denote the variance and mean value of amplitude readings for the ath antenna of the cth subcarrier. We select the antennas with the highest and lowest s_a values as $H_1(f,t)$ and $H_2(f,t)$, respectively. The rationale behind this selection is that CSI with larger variances is generally more sensitive to motion, while CSI with higher amplitude typically contains a larger static path component, making $H_{1(f,t)}$ less affected by $\triangle d$.

After obtaining denoised CSI after data acquisition, and to further reduce the impact of the static path (without introducing DFS), we apply a high-pass filter. Finally, we utilize Fast Fourier Transform (FFT) to obtain DFS. A comparison between the method used to extract DFS in previous approaches and the one adopted by WiOpen is illustrated in Figure 5. Notably, the latter produces higher-quality DFS.

D. Open-Set Gesture Recognition Network

The proposed OSGR network, depicted in Figure 6, is structured into two branches, each with specific objectives. The first branch is dedicated to constructing DFS outputs from the original inputs, while the other is focused on learning valuable class-related knowledge. Throughout the training process, these branches are guided by the neighborhood loss and construction loss, respectively, to facilitate effective knowledge acquisition. During testing, a K-Nearest Neighbors (KNN)-based decision method is employed to classify known samples and reject unknown ones.

Even after data preprocessing, the CSI retains influences from static path interference and domain uncertainty. To address the former, some existing approaches [2], [21], [24] advocate using DFS or its derivatives as input for the network, as DFS predominantly captures dynamic path characteristics and effectively mitigates interference from static paths. However, relying solely on DFS as input may result in the loss of valuable information. Therefore, our approach retains both amplitude and phase as input but incorporates a construction loss to guide the network in leveraging dynamic path-related information and mitigating interference from static paths to the fullest extent. As illustrated in Figure 6, $f_{\theta 1}(x)$ represents the backbone network responsible for feature learning. It extracts low-dimensional features, denoted as v_n, from the sample.

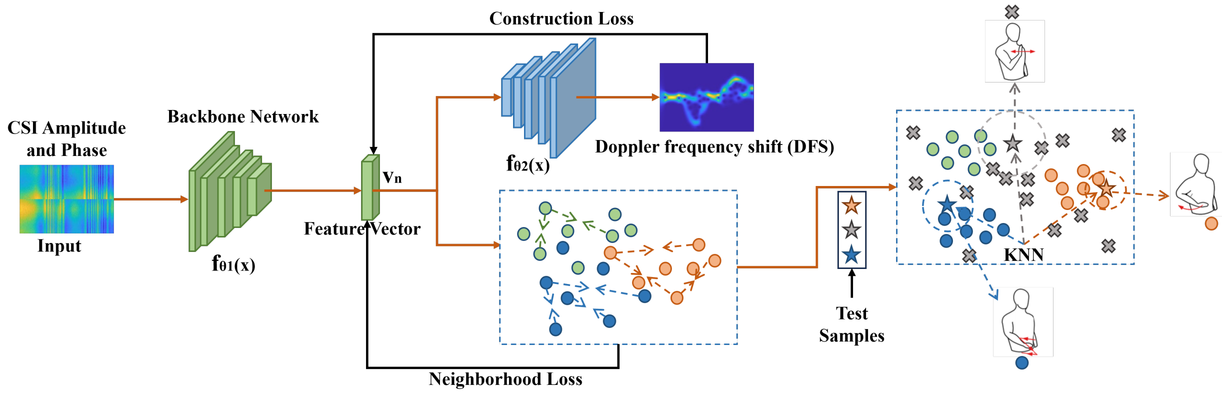

Fig. 6. Framework of the proposed open-set gesture recognition network. The feature vector v_n is the outputs of $f_{\theta 1}(x_n)$.

These features are then transformed into DFS corresponding to the sample through the first branch $f_{\theta 2}(x)$. During the process of constructing DFS, a bottleneck structure exists between $f_{\theta 1}(x)$ and $f_{\theta 2}(x)$. This structure ensures that the learned features v_n contain as much information related to dynamic paths as possible to achieve high-quality DFS construction. Therefore, training this branch using the construction loss serves the purpose of eliminating influence associated with static paths. The construction loss is defined as:

$$\mathcal{L}_c = MSE(x_n^{dfs}, f_{\theta 2}(f_{\theta 2}(x_n))) \tag{16}$$

where MSE represents Mean Squared Error loss, x_n is the input sample, and x_n^{dfs} is the DFS corresponding to x_n.

For addressing domain uncertainty, drawing inspiration from the uncertainty quantification method outlined in Equation 11, we have devised a scheme based on neighbor component analysis (NCA) [42]. In this approach, NCA computes the distance of each training sample to all other samples within the embedding space. Subsequently, guided by the distance metric and class labels, it minimizes the distances between samples of the same class while pushing samples from other classes further apart, thus effectively reducing domain uncertainty. The second branch is specifically dedicated to domain uncertainty reduction. Considering a training sample $x_n \in [x_1 ... x_N]$ along with its corresponding label y_n and its feature vector $v_n = f_{\theta 1}(x_n)$, we employ similarity as the distance metric. The similarity between x_n and another sample $x_j \in [x_1 ... x_N]$ is defined as the cosine similarity:

$$s_{nj} = \frac{v_n^T}{||v_n||||v_j||} = v_n^T v_j \tag{17}$$

The probability that sample x_n selects x_j as its neighbor is:

$$p_{nj} = \frac{exp(s_{nj}/\gamma)}{\sum_{k \neq n} exp(s_{nk}/\gamma)}, p_{nn} = 0 \tag{18}$$

γ is the parameter to control the scale of the neighborhood.

During the training process, for each sample x_n, the OSGR network computes its probability of being classified correctly as follows:

$$p_n = \sum_{j \in S_n} p_{nj} \tag{19}$$

S_n represents the indices of all samples with labels identical to that of x_n. The loss for the second branch, referred to as neighbor loss, is defined as:

$$\mathcal{L}_e = \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_e^n = -\frac{1}{N} \sum_{n=1}^{N} log(p_n) \tag{20}$$

The gradient calculation of Loss L_e with respect to the feature vectors v_n is as follows:

$$\frac{\partial \mathcal{L}_e^n}{\partial v_n} = \frac{1}{\gamma} (\sum_{k \in S_n} p_{nk} v_k - \sum_{k \in S_n} \tilde{p}_{nk} v_k) \tag{21}$$

where $\tilde{p}_{nk} = p_{nk} / \sum_{j \in S_n} p_{nj}$ is the normalized distribution within the class y_n. The overall loss of the OSGR network is:

$$\mathcal{L} = \mathcal{L}_e + \lambda \mathcal{L}_c \tag{22}$$

where λ is the hyperparameter controlling the impact of the construction branch. Throughout the training process, $\mathcal{L}$ induces samples of the same class to converge in the feature space while simultaneously distancing them from samples of different classes. This process not only eradicates domain uncertainty but also enables each sample to harness knowledge associated with all other samples, facilitating the learning of class-related information. Consequently, this leads to the formation of a feature space characterized by semantic separability. Utilizing this semantic space allows for the classification of known classes and the discrimination of unknown classes.

For recognizing known gestures and rejecting unknowns, all feature vectors of training samples are stored in a feature database, referred to as T. During the testing phase, a query sample x_t has its feature vector v_t extracted using the pre-trained network $f_{\theta 1}(x_t)$. Following that, v_t is used to query a set of K nearest samples, referred to as S_k, from T based on a similarity measure. The majority class within this set is then defined as the candidate label:

$$y_t^c = max(p_c | p_c = \frac{sum(y_j = c)}{K}) \tag{23}$$

The final label for x_t is determined based on y_t^c and a threshold t:

$$y_t = \begin{cases} y_t^c, \sum_{j \in S_t^c} d(x_t, x_j) < t \\ y_u, \sum_{j \in S_t^c} d(x_t, x_j) \geq t \end{cases} \tag{24}$$

here, y_u signifies that x_t belongs to an unknown class. The threshold t is defined as:

$$t = \xi * \frac{1}{B} \sum_{b=1}^{B} max(\sum_{j \in S_r} d(x_r, x_j)) \tag{25}$$

Where B represents the total number of training batches, x_r refers to all the samples in batch b, and ξ is a hyper-parameter used to control the threshold. In contrast to previous approaches, the proposed OSGR network formulates uncertainty measurement, feature learning objective, and classifier decision criteria based on inter-sample neighborhood relationships. This approach integrates the optimization objectives for uncertainty reduction, meaningful feature learning, and decision boundary definition in an end-to-end manner. Setting decision boundaries through neighborhood relationships overcomes the limitations of prototype-based schemes, allowing for dynamic and flexible determination of decision boundaries based on the distribution of surrounding samples. Furthermore, the proposed threshold determination approach facilitates adaptive threshold setting during training, thereby enhancing the flexibility of decision boundary determination.

V. Experimental Analysis

A. Datasets

In this section, we introduce two datasets, Widar3.0 and ARIL, which are used for experimental evaluation.

Widar3.0: The public dataset WiDar3.0 [2] contains 16125 samples collected from 3 environments, and gestures include Push Pull, Sweep, Clap, Slide, Draw-O(Horizontal), Draw-Zigzag(Horizontal), Draw-N(Horizontal), Draw-Triangle(Horizontal) and Draw-Rectangle(Horizontal). To verify the performance of the WiOpen system, in section V-C, we evaluate the know samples recognition and unknow samples rejection performance under different openness $\mathcal{P}$ with all the data from 1st environment (10125 samples, 9 users $\times$ 5 positions $\times$ 5 orientations $\times$ 9 gestures $\times$ 5 instances). In section V-D, we also use all samples to evaluate the in-domain and cross-domain performance of WiOpen same as other state-of-the-art researches [22], [23].

ARIL: The ARIL dataset [43] comprises six distinct gestures (namely, hand up, hand down, hand left, hand right, hand circle, and hand cross), executed by a single user at 16 distinct locations within a confined space. Notably, the ARIL dataset, while featuring only one environmental variable (location), incorporates the use of universal software radio peripheral (USRP) devices for CSI data collection. We employ this dataset to assess the versatility of WiOpen under different openness $\mathcal{P}$ across different wireless devices. Otherwise, we also use ARIL to evaluate the cross-domain performance of WiOpen. The ARIL dataset encompasses 1392 samples totally.

B. Implementation Details

In our implementation, the data acquisition and preprocessing were realized using Matlab, whereas the OSGR network was constructed using the PyTorch framework. Following extensive experimentation, we determined the optimal hyper-parameters, which were set as follows: $\xi = 2$, $\lambda = 1$, and $\gamma = 0.05$. The number of neighbors, K, used for selecting test sample labels was set to 50. During training, we utilized an initial learning rate of 0.001, which was reduced by a factor of 10 every 15 epochs for a total of 50 epochs. We employed the Adam optimizer for our model, and the feature extraction backbone network, $f_{\theta 1}(x)$, is a ResNet18 network, while the $f_{\theta 2}(x)$ is a 7- layer CNN. We evaluate Wiopen using Widar3.0 [2] and ARIL [43] datasets.

TABLE I
AUROC comparison between different methods on Widar3.0. The best performance values are highlighted in bold.

Method \ $\mathcal{P}$	0.11	0.16	0.22	0.29	0.40
Softmax [44]	0.67	0.76	0.73	0.74	0.71
WiGRUNT [1]	0.69	0.76	0.73	0.75	0.72
Wione [4]	0.68	0.75	0.75	0.76	0.69
WiOpen	**0.74**	**0.80**	**0.81**	**0.79**	**0.76**

TABLE II
Close-set accuracy comparison between different methods on Widar3.0. The best performance values are highlighted in bold.

Method \ $\mathcal{P}$	0.11	0.16	0.22	0.29	0.40
Softmax [44]	96.27%	95.11%	95.11%	95.56%	95.33%
WiGRUNT [1]	**96.81%**	95.22%	96.11%	95.85%	96.89%
Wione [4]	96.30%	95.29%	96.22%	96.15%	96.22%
WiOpen	96.59%	**95.82%**	**96.56%**	**96.59%**	**97.11%**

C. Overall Performance

In this study, we have chosen to assess OSGR performance using close-set accuracy and the area under the Receiver Operating Characteristic curve (AUROC). Close-set accuracy focuses on evaluating the classification performance with respect to known classes, while AUROC [32] serves as a robust measure for assessing the ability to distinguish unknown classes. An AUROC value of "1" indicates complete separability between known and unknown classes. We also conducted a comparative analysis between WiOpen and three reference models: a baseline model based on Softmax, a SOTA Wi-Fi gesture recognition system WiGRUNT [1] and a prototype based method Wione [4].

Results on Widar3.0. To evaluate the performance of the WiOpen in open-set scenarios, we conducted experiments using 10,125 samples from the first environment and performed testing at different openness levels ($\mathcal{P} = [0.11, 0.16, 0.22, 0.29, 0.40]$, corresponding to [3, 4, 5, 6, 7] unknown classes.) and $\mathcal{P}$ is calculated by Equation 1. When $\mathcal{P} = 0.29$, it indicates that our model was trained using 2,700 samples from three known classes. Within the test set, we allocated 675 samples from these three known classes as "known" samples, while the remaining 6,750 samples from six other classes were considered "unknown." This setup mirrors a real-world scenario where a substantial number of unknown samples are encountered. For AUROC calculations, we assigned label "0" to known classes and label "1" to unknown classes.

The experimental results, as outlined in Tables I and II, underscore the superior performance of WiOpen in open-set

TABLE III
AUROC COMPARISON BETWEEN DIFFERENT METHODS ON ARIL. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD.

Method \ $\mathcal{P}$	0.05	0.11	0.18	0.29
Softmax [44]	0.64	0.65	0.59	0.71
WiGRUNT [1]	0.65	0.66	0.58	0.72
Wione [4]	0.66	0.66	0.57	0.70
WiOpen	**0.73**	**0.67**	**0.72**	**0.76**

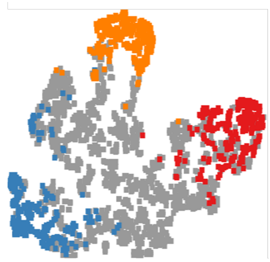

Fig. 7. The distribution of samples in the feature space on Widar3.0. Colored points represent samples from three known classes, while gray points represent samples from unknown classes.

scenarios compared to traditional methods. WiOpen not only maintains competitive recognition accuracy but also excels in identifying unknown samples, demonstrating a consistently higher AUROC ranging from 0.03 to 0.07 compared to other methods. Notably, the transition from conventional gesture recognition systems, such as WiGRUNT, to OSGR systems often involves a complex search process to determine the optimal score threshold. In contrast, WiOpen dynamically adapts the threshold during the training process. WiOpen's training and unknown rejection strategies, grounded in neighborhood structures, offer an improvement over prototype-based methods like Wione. As evident in Tables I and II, WiOpen achieves a balanced performance in recognizing known samples and rejecting unknown samples. It's crucial to note that when $\mathcal{P} = 0.11$, the AUROC for all approaches is notably low. This phenomenon is attributed to the high similarity between the sixth and seventh gestures (Zigzag and N), which can be considered essentially the same gesture rotated in space. This emphasizes the importance of addressing the distinction between intra-set and extra-set gestures, especially when they exhibit varying levels of similarity in open-set scenarios. This presents an intriguing avenue for future research. To visually represent the impact of the OSGR network, we plotted the distribution of samples in the feature space after processing, as depicted in Figure 7. The figure illustrates that the OSGR network effectively mitigates a portion of domain uncertainty compared to the state depicted in Figure 3b. For $\mathcal{P} = 0.11$, the closed-set accuracy of WiOpen is slightly lower than that of WiGRUNT. This difference arises because WiOpen is specifically designed for open-set scenarios, and this design trade-off may marginally impact its performance in closed-set settings. However, WiOpen achieves a higher AUROC compared to WiGRUNT.

Results on ARIL. For ARIL, we use all 1,392 samples to evaluate WiOpen. Testing was conducted across a range of openness levels ($\mathcal{P} = [0.05, 0.11, 0.17, 0.29]$, corresponding to [1, 2, 3, 4] unknown classes.). The parameters and experimental procedures remained consistent with those employed for the WiDar3.0 dataset. It's worth noting that when evaluating the softmax and WiGRUNT systems, a new search for the optimal threshold was required, introducing notable inconvenience and limitations.

TABLE IV
CLOSE-SET ACCURACY COMPARISON BETWEEN DIFFERENT METHODS ON ARIL. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD.

Method \ $\mathcal{P}$	0.05	0.11	0.18	0.29
Softmax [44]	86.25%	88.02%	93.06%	91.67%
WiGRUNT [1]	87.50%	88.54%	93.75%	92.71%
Wione [4]	87.95%	88.79%	92.96%	90.52%
WiOpen	**89.58%**	**93.23%**	**93.75%**	**94.79%**

TABLE V
AUROC COMPARISON BETWEEN DIFFERENT METHODS ON REAL-WORLD. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD.

Method \ $\mathcal{P}$	0.11	0.16	0.22	0.29	0.40
Softmax [44]	0.50	0.51	0.53	0.53	0.50
WiGRUNT [1]	0.51	0.52	0.53	0.51	0.52
Wione [4]	0.50	0.52	0.51	0.53	0.52
WiOpen	**0.53**	**0.54**	**0.54**	**0.55**	**0.56**

The experimental results, presented in Tables III and IV, firmly establish WiOpen's superiority in both recognizing known samples and effectively rejecting unknown samples. These results underscore WiOpen's robust performance.

Results on real-world experiments. We also conducted experiments in a real-world environment to evaluate WiOpen. In a larger space, we deployed one transmitter with a single antenna and three receivers, each equipped with three antennas. The distance between the receiving antennas is 6 cm, while the distances between the transmitter and the receiver are 0.5 m, 1.5 m, and 2 m, respectively. The experimental environment include various sources of interference, such as other WiFi devices and human activities, to simulate realistic and challenging scenarios. Two users performed nine gestures at five different positions and in three orientations, with each gesture repeated three times. The user's five positions are located at the four vertices and the center of a square with a side length of 0.8 m. The gestures were identical to the nine gestures used in the Widar3.0 dataset. Data from two gesture trials were used for training, with the remaining trial used for testing, and three-fold cross-validation was performed. As shown in Table V and VI, the results demonstrate that WiOpen achieves an advantage over others, highlighting its effectiveness in real-world conditions. Compared to the performance on publicly available datasets, gesture recognition accuracy in real-world environments significantly decreased. This is primarily because we did not control the cleanliness of the experimental environment to better reflect real-world conditions, and the limited amount of collected data resulted in insufficient training. This highlights the urgent need for large-scale WiFi recognition datasets in real-world scenarios to advance related research. Addressing this challenge will be a key focus of our future work.

TABLE VI
CLOSE-SET ACCURACY COMPARISON ON REAL-WORLD. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD.

Method \ $\mathcal{P}$	0.11	0.16	0.22	0.29	0.40
Softmax [44]	36.11%	43.33%	48.61%	64.82%	66.67%
WiGRUNT [1]	38.87%	45.56%	52.78%	70.37%	69.44%
Wione [4]	37.04%	45.56%	51.39%	68.52%	72.22%
WiOpen	**40.74%**	**47.78%**	**55.56%**	**72.22%**	**77.78%**

TABLE VII
OPEN-SET + CROSS DOMAIN COMPARISON ON WIDAR3.0. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD.

Method	Cross-Ori		Cross-Loc	
	AUROC	Acc	AUROC	Acc
Softmax [44]	0.70	90.80%	0.76	97.19%
WiGRUNT [1]	0.72	91.70%	0.76	**97.33%**
Wione [4]	0.69	90.80%	0.74	96.87%
WiOpen	**0.78**	**92.59%**	**0.81**	97.19%

Results on Open-Set + Cross Domain. To ascertain WiOpen's performance in more demanding environments, we combined open-set scenarios with cross-domain scenarios. Specifically, we conducted experiments utilizing data from the 1st environment of the WiDar3.0 dataset, applying open-set setting across different locations and orientations, all under an openness level of 0.29.

The experimental results, as detailed in Table VII, affirm that WiOpen maintains its proficiency in recognizing known samples and effectively rejecting unknown samples, even in cross-domain open-set scenarios. Notably, the introduction of domain discrepancies poses additional challenges. The successful performance of WiOpen in cross-domain open-set recognition is especially promising, as it closely aligns with real-world applications and merits further exploration.

Results on Sparse Distribution. We conducted experiments with sparse data on the Widar3.0 dataset with $\mathcal{P} = 0.29$, utilizing two different forms of sparse data distribution. For the first scenario, we removed particular types of sample points from the data space to create sparse distribution. This experiment corresponds to the open-set + cross-domain experiments described in Section V-C of the paper. As shown in Table VII, the results indicate that WiOpen consistently outperforms other approaches under these conditions. For the second from, we randomly removed sample points from the data space. The results, presented in Table VIII, demonstrate that WiOpen outperforms alternative methods even when 50% or 75% of the sample points are removed.

TABLE VIII
OPEN-SET + RANDOM SPARSE DISTRIBUTION COMPARISON ON WIDAR3.0. THE BEST PERFORMANCE VALUES ARE HIGHLIGHTED IN BOLD, AND S REPRESENTS THE PROPORTION OF RETAINED SAMPLES.

Method	$S = 0.5$		$S = 0.25$	
	AUROC	Acc	AUROC	Acc
Softmax [44]	0.70	92.15%	0.61	90.22%
WiGRUNT [1]	0.70	94.37%	0.65	91.56%
Wione [4]	0.69	94.07%	0.63	92.15%
WiOpen	**0.75**	**94.67%**	**0.66**	**93.04%**

D. Cross Domain Performance

To provide a more comprehensive comparison of WiOpen with additional SOTA Wi-Fi-based gesture recognition systems (which are not open-sourced) and to emphasize the effectiveness of our proposed uncertainty reduction method, we extended our experiments to include cross-domain tasks reported in their respective papers. To ensure compatibility with these established solutions, we evaluated performance on the Widar3.0 dataset under two distinct settings: one utilizing all six pairs of transmitter-receivers and another employing only a single pair of transmitter-receiver. For the six-pair configuration, our dataset settings closely align with those of the SOTA PAC-CSI method [23]. It is worth noting that PAC-CSI selected 80% of the testing samples for their test set, whereas we did not perform any such sample selection for our test set. For the single-pair configuration, our dataset settings mirror those of the WiSR method [3]. Additionally, we conducted experiments on the ARIL dataset to further evaluate the performance of WiOpen in a broader context.

The experimental results are summarized in Table IX. In the WiDar3.0-6D setting, WiOpen exhibits slightly lower performance than the state-of-the-art PCA-CSI [23] in the cross-orientation task. However, it outperforms both WiGRUNT and PCA-CSI in other settings, showcasing superior overall performance. In the WiDar3.0-1D and ARIL scenarios, WiOpen lags behind WiSGP and WiSR in the cross-orientation task. Nevertheless, it significantly outperforms WiSGP [22] and WiSR [3] in other tasks. Particularly noteworthy is WiOpen's substantial lead of over 41% in cross-environment scenarios.

The key to WiOpen's success in cross-domain tasks lies in its innovative approach to uncertainty elimination. While initially designed to address open-set challenges in Wi-Fi gesture recognition, the domain uncertainty reduction facilitated by the OSGR network proves to be advantageous for cross-domain applications. The elimination of domain uncertainty induces the convergence of feature spaces for individual classes. The features acquired through OSGR learning remove domain-specific irregularities, fundamentally alleviating domain disparities and augmenting domain generalization capabilities. This further underscores the effectiveness of WiOpen, highlighting the central role played by its foundational principle of uncertainty reduction.

E. Sensitivity and Misclassified Analysis

In this section, we conduct sensitivity analyses on Widar3.0 to assess the impact of various factors on its performance.

Impact of ξ. To evaluate the sensitivity of WiOpen to the threshold ξ, we conducted experiments with ξ values ranging from 1 to 3 while keeping $\mathcal{P} = 0.29$. The results, displayed in Figure 8, demonstrate a clear relationship between the threshold and system performance. As the threshold increases, the close-set accuracy decreases, but the system's effectiveness in rejecting unknown samples improves. To strike a balance between these performance aspects, we selected $\xi = 2$.

Impact of λ. To analyze the influence of the construction loss, we performed experiments with λ values ranging from 0 to 2 while maintaining $\mathcal{P} = 0.29$. The results, as illustrated

TABLE IX
CROSS DOMAIN GESTURE RECOGNITION RESULTS COMPARED WITH STATE-OF-THE-ART SOLUTIONS.(I-D, C-L, C-O, C-E AND C-U MEANS IN-DOMAIN, CROSS LOCATION, ORIENTATION, ENVIRONMENT AND USER, RESPECTIVELY. 6D AND 1D MEANS USE 6 AND 1 PAIRS OF TRANSMITTER-RECEIVERS, RESPECTIVELY).

Method	Pub	WiDar3.0-6D					WiDar3.0-1D				ARIL
		I-D	C-L	C-O	C-E	C-U	C-E	C-L	C-O	C-U	C-L
WiDar3.0 [2]	IEEE TPAMI	92.7%	89.7%	82.6%	92.4%	-	-	-	-	-	-
WiHF [21]	IEEE TMC	97.65%	92.07%	82.38%	89.67%	-	-	-	-	-	-
WiGRUNT [1]	IEEE THMS	99.71%	96.62%	93.85%	93.73%	-	-	-	-	-	-
PAC-CSI [23]	IEEE JSAC	99.46%	98.77%	**98.90%**	96.47%	97.54%	-	-	-	-	-
SelfReg [45]	IEEE ICCV	-	-	-	-	-	39.11%	76.71%	86.67%	53.10%	44.45%
WiSGP [22]	IEEE TMC	-	-	-	-	-	43.17%	78.49%	88.46%	56.77%	48.74%
WiSR [3]	IEEE TMC	-	-	-	-	-	42.52%	77.51%	**88.80%**	55.18%	48.64%
WiOpen	-	**99.78%**	**98.81%**	98.05%	**97.99%**	**98.47%**	**84.44%**	**86.40%**	77.67%	**82.71%**	**73.61%**

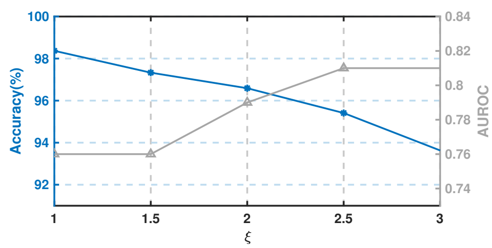

Fig. 8. Impact of ξ, and ξ affects the threshold value used for identifying unknown classes on Widar3.0.

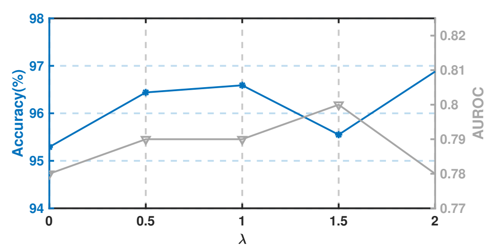

Fig. 9. Impact of λ, and λ determines the influence of the construction loss on the training process on Widar3.0.

in Figure 9, show that the construction loss significantly contributes to WiOpen's ability to acquire valuable knowledge. In our experiments, we opted for $\lambda = 1$.

Misclassified Analysis. We conducted experiments on the Widar dataset with $\mathcal{P} = 0.29$, comprising six known classes and three unknown classes. The confusion matrix, displayed in Figure 10, assigns all identified unknown classes uniformly to the label 6. For the known classes, simpler gestures are more prone to being misclassified as unknown classes, such as the gestures "sweep" and "slide." Regarding the unknown classes, they are often misclassified into visually or structurally similar known classes. For example, "draw-N" is frequently misclassified as "draw-Z," likely because "N" can be perceived as a rotated version of "Z." Similarly, "draw-rectangle" tends to be misclassified as "draw-O."

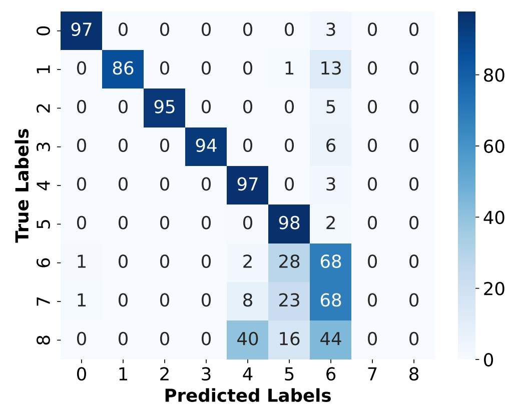

Fig. 10. Classification Confusion Matrix on Widar3.0, Values Represented as Percentages (%). Label 0-8 represents Push Pull, Sweep, Clap, Slide, Draw-O(Horizontal), Draw-Zigzag(Horizontal), Draw-N(Horizontal), Draw-Triangle(Horizontal), Draw-Rectangle(Horizontal) respectively.

VI. CONCLUSIONS

In this paper, we have introduced WiOpen, a pioneering Wi-Fi-based OSGR system. We commenced our study with a comprehensive analysis of open-set challenges within the realm of Wi-Fi sensing, shedding light on the intrinsic correlation between Wi-Fi-based OSGR and uncertainty. Building upon this correlation analysis, we have put forth solutions designed to eradicate uncertainty in Wi-Fi sensing and establish decision boundaries. Our experimental results underscore the remarkable effectiveness of WiOpen in open-set gesture recognition task. This substantiates WiOpen's unique ability to glean meaningful knowledge, making it a robust system for real-world applications. For future works, we will continue to address the intricate challenges posed by the high similarity between unknown and known gestures in open-set scenarios. Furthermore, we aim to tackle the complexities introduced when the target domain contains unknown samples during domain adaptation, further enhancing their suitability for real-world applications.

REFERENCES

[1] Y. Gu, X. Zhang, Y. Wang, M. Wang, H. Yan, Y. Ji, Z. Liu, J. Li, and M. Dong, "Wigrunt: Wifi-enabled gesture recognition using dual-attention network," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 736–746, 2022.

[2] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8671–8688, 2021.

[3] S. Liu, Z. Chen, M. Wu, C. Liu, and L. Chen, "Wisr: Wireless domain generalization based on style randomization," *IEEE Transactions on Mobile Computing*, 2023.

[4] Y. Gu, H. Yan, M. Dong, M. Wang, X. Zhang, Z. Liu, and F. Ren, "Wione: One-shot learning for environment-robust device-free user authentication via commodity wi-fi in man–machine system," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 630–642, 2021.

[5] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Caution: A robust wifi-based human authentication system via few-shot open-set recognition," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 323–17 333, 2022.

[6] S. Tan, J. Yang, and Y. Chen, "Enabling fine-grained finger gesture recognition on commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 21, no. 8, pp. 2789–2802, 2020.

[7] S. Tan, L. Zhang, Z. Wang, and J. Yang, "Multitrack: Multi-user tracking and activity recognition using commodity wifi," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[8] X. Zhang, Y. Lu, H. Yan, J. Huang, Y. Ji, and Y. Gu, "Resup: Reliable label noise suppression for facial expression recognition," *arXiv preprint arXiv:2305.17895*, 2023.

[9] K. Niu, F. Zhang, Z. Chang, and D. Zhang, "A fresnel diffraction model based human respiration detection system using cots wi-fi devices," in *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers*, 2018, pp. 416–419.

[10] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 27–38.

[11] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 617–628.

[12] Z. Wei, W. Chen, S. Ning, W. Lin, N. Li, B. Lian, X. Sun, and J. Zhao, "A survey on wifi-based human identification: Scenarios, challenges, and current solutions," *ACM Transactions on Sensor Networks*, 2024.

[13] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[14] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using wifi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 401–413.

[15] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 2015, pp. 77–89.

[16] N. Yu, W. Wang, A. X. Liu, and L. Kong, "Qgesture: Quantifying gesture distance and direction with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–23, 2018.

[17] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger: talk to your smart devices with finger-grained gesture," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 250–261.

[18] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 90–102.

[19] J. Huang, J.-X. Bai, X. Zhang, Z. Liu, Y. Feng, J. Liu, X. Sun, M. Dong, and M. Li, "Keystrokesniffer: An off-the-shelf smartphone can eavesdrop on your privacy from anywhere," *IEEE Transactions on Information Forensics and Security*, 2024.

[20] L. Zhang, Y. Zhang, and X. Zheng, "Wisign: Ubiquitous american sign language recognition using commercial wi-fi devices," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 3, pp. 1–24, 2020.

[21] C. Li, M. Liu, and Z. Cao, "Wihf: Gesture and user recognition with wifi," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 757–768, 2020.

[22] S. Liu, Z. Chen, M. Wu, H. Wang, B. Xing, and L. Chen, "Generalizing wireless cross-multiple-factor gesture recognition to unseen domains," *IEEE Transactions on Mobile Computing*, 2023.

[23] J. Su, Q. Mao, Z. Liao, Z. Sheng, C. Huang, and X. Zhang, "A real-time cross-domain wi-fi-based gesture recognition system for digital twins," *IEEE Journal on Selected Areas in Communications*, 2023.

[24] C. Feng, N. Wang, Y. Jiang, X. Zheng, K. Li, Z. Wang, and X. Chen, "Wi-learner: Towards one-shot learning for cross-domain wi-fi based gesture recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–27, 2022.

[25] P. R. Mendes Júnior, R. M. De Souza, R. d. O. Werneck, B. V. Stein, D. V. Pazinato, W. R. de Almeida, O. A. Penatti, R. d. S. Torres, and A. Rocha, "Nearest neighbors distance ratio open-set classifier," *Machine Learning*, vol. 106, no. 3, pp. 359–386, 2017.

[26] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 7, pp. 1757–1772, 2012.

[27] D.-W. Zhou, H.-J. Ye, and D.-C. Zhan, "Learning placeholders for open-set recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 4401–4410.

[28] X. Sun, Z. Yang, C. Zhang, K.-V. Ling, and G. Peng, "Conditional gaussian distribution learning for open set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 480–13 489.

[29] H.-M. Yang, X.-Y. Zhang, F. Yin, Q. Yang, and C.-L. Liu, "Convolutional prototype network for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2358–2370, 2020.

[30] G. Chen, L. Qiao, Y. Shi, P. Peng, J. Li, T. Huang, S. Pu, and Y. Tian, "Learning open set network with discriminative reciprocal points," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 507–522.

[31] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8065–8081, 2021.

[32] H. Huang, Y. Wang, Q. Hu, and M.-M. Cheng, "Class-specific semantic reconstruction for open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 4, pp. 4214–4228, 2022.

[33] C. Geng, S.-j. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3614–3631, 2020.

[34] J. Huang, B. Liu, C. Miao, X. Zhang, J. Liu, L. Su, Z. Liu, and Y. Gu, "Phyfinatt: An undetectable attack framework against phy layer fingerprint-based wifi authentication," *IEEE Transactions on Mobile Computing*, vol. 23, no. 7, pp. 7753–7770, 2023.

[35] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "Wiact: A passive wifi-based human activity recognition system," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2019.

[36] X. Zhang, J. Zhang, Z. Ma, J. Huang, M. Li, H. Yan, P. Zhao, Z. Zhang, Q. Guo, T. Zhang *et al.*, "Camlopa: A hidden wireless camera localization framework via signal propagation path analysis," *arXiv preprint arXiv:2409.15169*, 2024.

[37] X. Zhang, Y. Gu, H. Yan, Y. Wang, M. Dong, K. Ota, F. Ren, and Y. Ji, "Wital: A cots wifi devices based vital signs monitoring system using nlos sensing model," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 629–641, 2023.

[38] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: Do user location and body orientation matter?" in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 2016, pp. 25–36.

[39] F. Zhang, K. Niu, J. Xiong, B. Jin, T. Gu, Y. Jiang, and D. Zhang, "Towards a diffraction-based sensing approach on human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–25, 2019.

[40] Y. Gu, X. Zhang, H. Yan, Z. Liu, and Y. Ji, "Real-time vital signs monitoring based on cots wifi devices," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 1320–1324.

[41] D. Wu, R. Gao, Y. Zeng, J. Liu, L. Wang, T. Gu, and D. Zhang, "Fingerdraw: Sub-wavelength level finger motion tracking with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–27, 2020.

[42] Z. Wu, A. A. Efros, and S. X. Yu, "Improving generalization via scalable neighborhood component analysis," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 685–701.

[43] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with wifi fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.

[44] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1563–1572.

[45] D. Kim, Y. Yoo, S. Park, J. Kim, and J. Lee, "Selfreg: Self-supervised contrastive regularization for domain generalization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9619–9628.