

# iRadar: Synthesizing Millimeter-Waves from Wearable Inertial Inputs for Human Gesture Sensing

Huanqi Yang<sup>1</sup>, Mingda Han<sup>2</sup>, Xinyue Li<sup>3</sup>, Di Duan<sup>1</sup>, Tianxing Li<sup>4</sup>, Weitao Xu<sup>1,\*</sup>

<sup>1</sup>City University of Hong Kong, <sup>2</sup>Shandong University,

<sup>3</sup>Xidian University, <sup>4</sup>Michigan State University

**Abstract**—Millimeter-wave (mmWave) radar-based gesture recognition is gaining attention as a key technology to enable intuitive human-machine interaction. Nevertheless, the significant challenge lies in obtaining large-scale, high-quality mmWave gesture datasets. To tackle this problem, we present *iRadar*, a novel cross-modal gesture recognition framework that employs Inertial Measurement Unit (IMU) data to synthesize the radar signals generated by the corresponding gestures. The key idea is to exploit the IMU signals, which are commonly available in contemporary wearable devices, to synthesize the radar signals that would be produced if the same gesture was performed in front of a mmWave radar. However, several technical obstacles must be overcome due to the differences between mmWave and IMU signals, the noisy gesture sensing of mmWave radar, and the dynamics of human gestures. Firstly, we develop a method for processing IMU and mmWave data that can consistently extract critical gesture features. Secondly, we propose a diffusion-based IMU-to-radar translation model that accurately transforms IMU data into mmWave data. Lastly, we devise a novel transformer model to enhance gesture recognition performance. We thoroughly evaluate *iRadar*, involving 18 gestures and 30 subjects in three scenarios, using five wearable devices. Experimental results demonstrate that *iRadar* consistently achieves 99.82% Top-3 accuracy across diverse scenarios.

**Index Terms**—mmWave sensing, gesture sensing, diffusion model

## I. INTRODUCTION

### A. Background and Motivation

Radio Frequency (RF)-based gesture recognition has attracted significant attention due to its ability to enable contactless and device-free human-machine interaction. A prime example is the utilization of millimeter-wave (mmWave) signals from frequency-modulated continuous-wave (FMCW) radar for gesture recognition. This exploits that each gesture has a unique pattern, and mmWave signals can capture these differences. The applications of mmWave-based gesture recognition extend to diverse fields such as smart homes, autonomous driving, and interactive gaming [1]–[3].

Despite its potential, mmWave radar-based gesture recognition, like many other RF-based sensing tasks, confronts a fundamental challenge that requires extensive training with prior instances of individuals performing gestures in the same settings [4]–[6]. This requirement poses difficulties for the practical deployment of this technology in real-world scenarios. For instance, in interactive gaming scenarios, a radar-based gesture recognition system may struggle to accurately identify gestures that have not been previously recorded by mmWave radar. To address this issue, recent studies have explored the use of transfer learning [7] and domain adaptation [8], [9] to

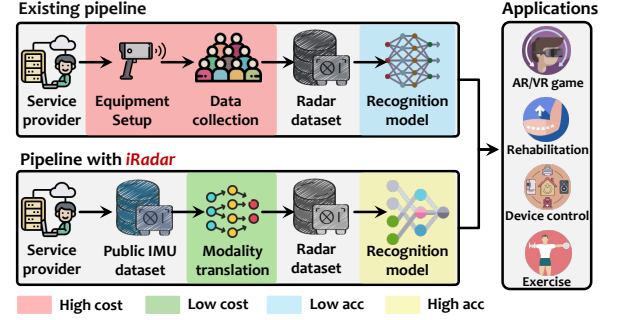


Fig. 1: Motivation of *iRadar*.

reduce the training burden, primarily focusing on minimizing the required training instances. While these approaches have shown promising results, it is crucial to acknowledge that they still necessitate prior radar data collection, which carries two key limitations: 1) the deployment of radar devices in the data collection area and 2) the pre-collection of gesture instances.

This paper introduces a novel approach that eliminates the need for prior mmWave data collection and substantially deviates from existing mmWave-based gesture recognition systems. Drawing from the recent success of diffusion models (e.g., Sora and GPT-4), we aim to investigate the viability of using alternative sensing modalities to eliminate the need for data collection. Traditional gesture recognition methods primarily rely on three sensing modalities: wireless signals [4], [6], [10], cameras [11]–[13], and wearable sensors [14]–[16]. This inspires us to leverage the advantage of one sensing modality (i.e., the ubiquity of cameras and wearable sensors) to overcome the data scarcity problem of mmWave-based gesture sensing. While previous studies have suggested the synthesis of mmWave signals from videos [17]–[19], we found that wearable sensors provide multiple advantages over video in this context due to the following reasons. Firstly, video-based methods still require the deployment of a camera in the data collection environment. Secondly, videos are prone to various practical factors, such as occlusion, lighting conditions, and viewpoint, which can introduce instability in the generated mmWave signals. Moreover, privacy concerns arise when video recording for gesture analysis is implemented. In contrast, the Inertial Measurement Unit (IMU) is widely equipped with wearable devices, such as smartwatches and rings. Therefore, utilizing the ubiquitously available IMU sensors essentially reduces device deployment costs. Additionally, there is a plethora of existing IMU-based gesture datasets, like mmGest [20] and UHH-IMU [21], which can be directly translated to mmWave-based gesture datasets without

extensive data collection. Instead of following the traditional development pipeline for mmWave gesture recognition systems, our system allows service providers to convert IMU datasets into costly-to-collect mmWave datasets with ease. These transformed mmWave datasets are essential for developing accurate mmWave-based gesture recognition models that meet end-user needs. To summarize, our research leverages the strengths of IMU-based systems to overcome data scarcity in mmWave-based human gesture sensing.

## B. Challenges and Contributions

**Challenge 1: Fundamental discrepancies between IMU and mmWave signals.** The signal properties of IMUs and mmWave radars are inherently distinct. To begin with, IMUs detect the inertial forces and joint rotations associated with a person’s gestures, whereas mmWave radar devices exploit the shadowing, diffraction, reflection, and scattering effects caused by the gestures on wireless signals [22]–[25]. Additionally, IMU signals are expressed as real numbers, contrasting with the complex number representation of mmWave radar signals as shown in Fig. 2. To address this challenge, we thoroughly analyze the IMU and mmWave signals related to human gestures. We employ a theoretical model to explore their fundamental relationship. Yet, due to the dynamic nature of human gesture patterns, translating IMU data into mmWave data through direct mathematical formulations is a difficult task. While current diffusion models [26], [27] demonstrate strong performance in tasks such as text and vision generation, they exhibit limitations in the realm of sensor data generation. To bridge this gap, we propose a novel deep diffusion framework equipped with an inertial fusion module and a translation module to efficiently transform IMU data into mmWave data.

**Challenge 2: Noisy gesture sensing in mmWave radar.** A major challenge in leveraging commercial single-chip mmWave radar units for gesture recognition lies in the accurate extraction of fine-grained features critical for interpreting subtle movements. While these devices excel at detecting target movements, they often capture a significant amount of environmental noise. As depicted in Fig. 3 (a), the Range-Doppler map exhibits substantial interference. Therefore, the Time-Frequency map derived from the Range-Doppler map still contains unavoidable noise, as demonstrated in Fig. 3 (b), which can obscure the nuanced gestures we aim to identify. To address this challenge, we propose the MC-MWIE algorithm, a sophisticated dual-stage technique designed to enhance signal clarity and resolution for gesture recognition applications. It engages a synergistic approach combining cluster analysis with morphological processing to mitigate environmental noise and refine the resolution. The result is a set of mmWave heatmaps with improved clarity, enabling accurate feature extraction.

**Challenge 3: Dynamics of human gestures.** Gestures involve the coordinated movements of multiple body parts, which encompass various actions and involve numerous joints and muscles [28]. Capturing the full dynamics of these complex 3D motions with a single sensing modality proves difficult. Additionally, the variability and subtlety among different ges-

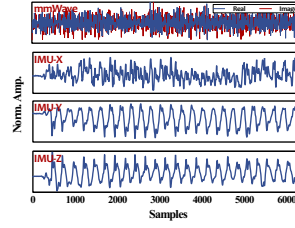


Fig. 2: Signals difference.

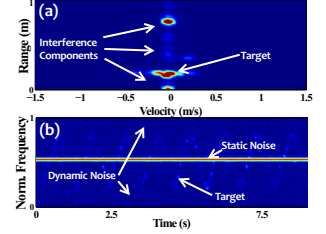


Fig. 3: Noisy gesture sensing.

tures often impede gesture recognition accuracy. To enhance the precision of gesture recognition, it is imperative to move beyond the traditional convolution-based method [7], [29], [30], which are insufficient for the complexity of gesture dynamics. Transformers have demonstrated effectiveness in handling vision tasks [31]; however, their direct application to mmWave heatmaps of gestures presents challenges, as the unique time-frequency properties of gesture signals differ significantly from typical image data. In response, we introduce a novel transformer for the unique mmWave gesture heatmaps. This model boosts recognition by integrating specialized components to better capture the intricacies of gestures.

In this paper, we propose *iRadar*, a cross-modal **iMU-to-Radar** gesture recognition framework. This system eliminates the necessity for the initial deployment of mmWave radar devices and the collection of explicit data. This progress pushes mmWave-enabled gesture recognition technologies into a realm of real-world usability. Through a comprehensive evaluation that included eighteen gestures, thirty participants, tested across three distinct settings, and utilizing five different mobile devices, *iRadar* has proven its efficacy by attaining an average accuracy of 92.3% in settings ranging from indoors and outdoors to through-obstacle scenarios. Our key contributions are outlined as follows:

- We present *iRadar*, which, to our best knowledge, is the first cross-modal IMU-to-mmWave gesture recognition framework that avoids the installation of mmWave device and explicit data collection, significantly reducing the burden of service providers.
- *iRadar* offers threefold specialized approaches, which include a diffusion-driven translation method, a mmWave heatmap enhancement method, and a Doppler transformer recognition method. Collectively, these methods tackle the above challenges and ensure accurate gesture recognition.
- We develop a system prototype and conduct extensive experiments to evaluate the performance of *iRadar* across various scenarios and mobile devices. Experimental results indicate *iRadar* consistently achieves an average Top-3 accuracy of 99.82%.

## II. PRELIMINARIES

### A. mmWave Sensing

The mmWave radar utilizes the FMCW signal, often referred to as the chirp signal. The chirp signal’s frequency increases linearly with time  $t$  according to the equation  $f(t) = f_c + St$ , where  $f_c$  denotes the starting frequency

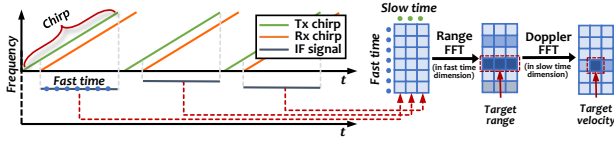


Fig. 4: FMCW signals and processing methods.

and  $S$  represents the frequency modulation slope [32], [33]. Assuming the amplitude of the transmitted signal at time  $t$  is  $A_1$ , the transmitted FMCW signal  $S_{Tx}(t)$  is expressed as

$$S_{Tx}(t) = A_1 \cos \left[ 2\pi \left( f_c t + \frac{St^2}{2} \right) \right]. \quad (1)$$

When the transmitted signal encounters an obstacle, such as the user's hand, at a distance  $d$ , the radar receives a delayed version of the transmitted signal  $S_{Tx}(t)$ , denoted as  $S_{Rx}(t)$ . This received signal can be expressed as

$$S_{Rx}(t) = \alpha A_1 \cos \left[ 2\pi \left( f_c (t - \tau) + \frac{S(t - \tau)^2}{2} \right) \right], \quad (2)$$

where  $\alpha$  represents the path loss,  $\tau = 2d/c$  denotes the time delay, and  $c$  is the speed of light. Finally, the transmitted signal  $S_{Tx}(t)$  is mixed with the received signal  $S_{Rx}(t)$ , and a low-pass filter is employed to extract the sum frequency components, resulting in the intermediate frequency (IF) signal:

$$S_{IF}(t) = LPF\{S_{Tx}(t) \cdot S_{Rx}(t)\} = A_2 \cos(2\pi f_{IF} t + \phi_{IF}), \quad (3)$$

where  $A_2$  is the amplitude of the IF signal,  $f_{IF} = 2dS/c$  is the beat frequency, and  $\phi_{IF}$  represents the phase of the IF signal. The FMCW mmWave radar enables the extraction of crucial target information such as range and velocity. Specifically, as depicted in Fig. 4, the range information is determined by applying the Fast Fourier Transform (Range FFT). The velocity information is obtained by performing the Fast Fourier Transform (Doppler FFT) on multiple IF signals spanning the slow time dimension.

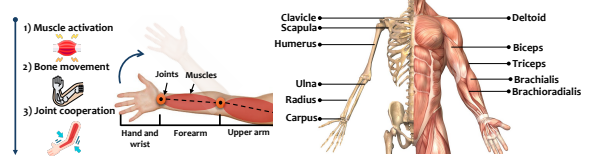
### B. Human Gesture

Human gestures are characterized by the intricate movements and postures adopted by various body segments [28]. These gestures originate from the unique interactions among specific body components. As shown in Fig. 5, the human gesture is driven by three steps. 1) Muscle activation: the biceps brachii muscle flexes the elbow, while the deltoid muscle aids in moving the arm at the shoulder joint. 2) Bone movement: the rotation of the radius over the ulna allows pronation and supination of the forearm. 3) Joint cooperation: the combined action of the shoulder's ball-and-socket joint, the elbow's hinge joint, and the wrist's complex array of plane and hinge joints. These synchronized elements facilitate a diverse range of nonverbal expressions.

### C. Cross-Modal Relationship Analysis

We now explore the relationship between mmWave and IMU gesture sensing, as shown in Fig. 6.

**IMU signal.** As detailed in Sec. II-B, human gestures involve distinct movements of specific body parts. The IMU data from an on-arm wearable device (denoted as  $I_s(t)$ , collected



(a) Movement process. (b) Skeletal muscles of arm.

Fig. 5: Understanding human gesture.

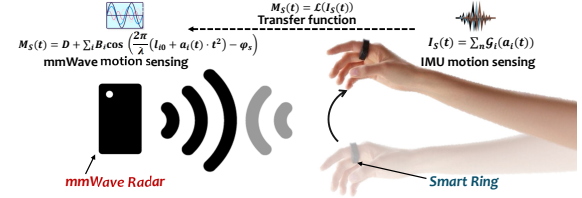


Fig. 6: Cross-modal relationship analysis.

during these gestures) is a composite representation of the accelerations experienced by the engaged body segments:  $I_s(t) = \sum_n g_i(a_i(t))$ , where  $a_i(t)$  is the acceleration of the  $i^{\text{th}}$  body part involved in the execution of the gesture at time  $t$ , and  $g_i(\cdot)$  represents the transfer function mapping the acceleration due to the movement of each body part to the on-arm device. **mmWave signal.** As discussed in Sec. II-A, the movement of human body parts during the execution of the gesture can lead to phase changes in the mmWave signal. Therefore, we can express the mmWave signal variations induced by different gestures mathematically as follows:

$$M(f, t) = H_0(f, t) + \sum_i A_i(f, t) e^{-j \left( \frac{2\pi}{\lambda} (l_{i0} + a_i(t) \cdot t^2) - \varphi_i \right)}, \quad (4)$$

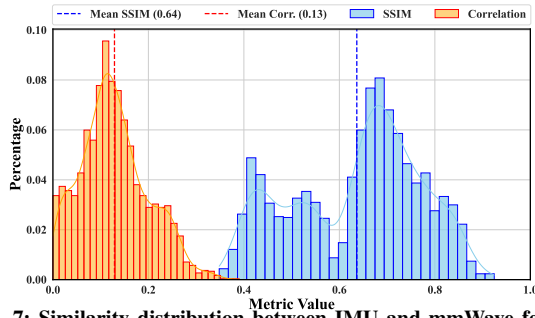
where  $H_0(f, t)$  represents the complex static path signal generated by the human body and surrounding environmental objects [33]–[40], while  $A_i(f, t)$  denotes the amplitude of the dynamic path signal reflected from the  $i^{\text{th}}$  arm/hand segment during gesture execution. The term  $l_{i0}$  signifies the initial signal propagation distance, and  $a_i(t)$  denotes the acceleration of the arm/hand segment. Let  $M_s(t)$  represent the magnitude square of the baseband signal  $M(f, t)$ . Assuming  $|H_0(f, t)| \gg |A_i(f, t)|$ , we can represent  $M_s(t)$  as

$$M_s(t) = D + \sum_i B_i \cos \left( \frac{2\pi}{\lambda} (l_{i0} + a_i(t) \cdot t^2) - \varphi_s \right), \quad (5)$$

where  $D = |H_0(f, t)|^2 + \sum_i |A_i(f, t)|^2$  represents the DC component of  $M_s(t)$ ,  $B_i = 2 |H_0(f, t) \cdot A_i(f, t)|$ , and  $\varphi_s$  denotes the phase of the complex static path signal.

**Relationship analysis.** As mentioned above, since both  $I_s(t)$  and  $M_s(t)$  are functions of acceleration  $a_i(t)$ ,  $I_s(t)$  incorporates frequency components that either coincide with or closely resemble the frequencies present in  $R(t)$ . Both data types are defined by functions that use acceleration as the independent variable, suggesting the potential to transform IMU data into mmWave data through a non-linear function  $\mathcal{L}(\cdot)$ :  $M_s(t) = \mathcal{L}(I_s(t))$ . We first examined the linear relationship between the features extracted from IMU spectrograms and mmWave heatmaps by calculating the coefficient of linear correlation. This analysis revealed a mean correlation value





**Fig. 7: Similarity distribution between IMU and mmWave features.** of 0.13, as illustrated by the orange line in Fig. 7. This low correlation coefficient suggests a weak linear relationship between these modalities. To further investigate the structural similarities between them, we utilized the Structural Similarity Index Measure (SSIM) [41], which measures the similarity between two features based on an understanding of visual perception. In our analysis, the mean SSIM between IMU and mmWave features is 0.64, as shown in the blue line in Fig. 7, indicating a moderate structural similarity. The above results indicate the complexity of defining  $\mathcal{L}(\cdot)$  that accurately characterizes the relationship between the IMU spectrogram and mmWave heatmap features. Due to this complexity and the inherent non-linearity between the two modalities, along with other complicating factors such as noise, signal attenuation, and multi-path effects [42], [43], conventional mathematical approaches to defining  $\mathcal{L}(\cdot)$  are insufficient. Therefore, we propose employing diffusion techniques, which are well-suited for capturing complex, non-linear relationships to establish an accurate mapping between them.

### III. SYSTEM DESIGN

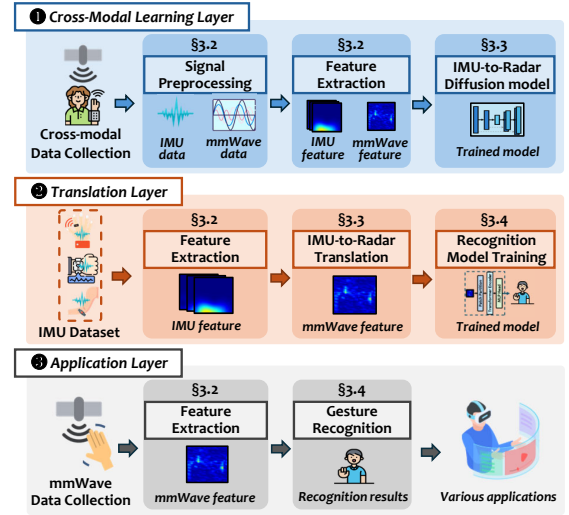
Fig. 8 shows iRadar’s overview, comprising three layers. **Cross-Modal Learning Layer.** In the cross-modal learning layer, we devise a deep diffusion model aimed at converting the IMU feature into the mmWave feature. Initially, the raw IMU data and mmWave data undergo preprocessing to mitigate noise. Subsequently, IMU and mmWave features are generated from each dataset using our proposed signal processing algorithms. Finally, the proposed diffusion model is trained for IMU-to-mmWave translation.

**Translation Layer.** In the translation layer, service providers have the option to utilize either publicly available or proprietary IMU data. This data is then translated into mmWave heatmaps using the trained IMU-to-Radar diffusion model. Once the translation is complete, the resulting mmWave heatmaps are used to train the gesture recognition network.

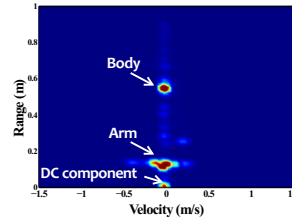
**Application Layer.** In the application layer, users can conveniently utilize mmWave radar-embedded smart devices to directly capture mmWave gesture data and conduct real-time gesture recognition for diverse applications.

#### A. Feature Extraction

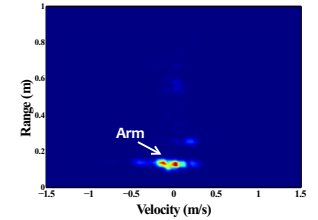
1) *mmWave Heatmap Generation:* As illustrated in Sec. II-A, the IF signal obtained by mixing is used for sensing gesture. In addition to the user’s gesture information, the



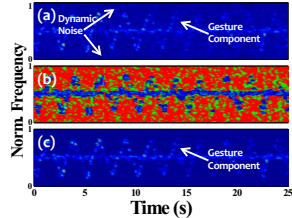
**Fig. 8: iRadar overview.**



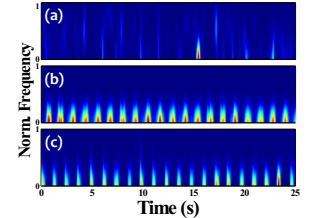
**Fig. 9: Raw RDM.**



**Fig. 10: Denoised RDM.**



**Fig. 11: MC-MHWE process.**



**Fig. 12: Extracted IMU feature.**

IF signal contains a lot of static noise generated by static objects such as walls, tables, and chairs as shown in Fig. 9. For each frame, we use the average of all IF signals as the static noise vector and subtract this static noise vector from each IF signal to obtain the denoised data as shown in Fig. 10. Then, we use the Range FFT and Doppler FFT to obtain a Range-Doppler Map (RDM) for each frame, which reflects the range and velocity information of the user while performing gesture in the current frame. Finally, to obtain the velocity change information during the execution of the gesture, we transform the RDMs of all frames into a 2D time-velocity feature map. As discussed in Sec. II-A, the Doppler FFT responds to changes in phase difference, which are proportional to frequency. Therefore, we can derive the normalized frequencies by normalizing the extracted velocities as shown in Fig. 11(a).

The resulting time-velocity feature map may exhibit some noisy components stemming from the superposition effects of RDMs, which can obscure significant velocity changes in the gestures. To address this issue, we propose Morphological Clustering for mmWave Heatmap Enhancement

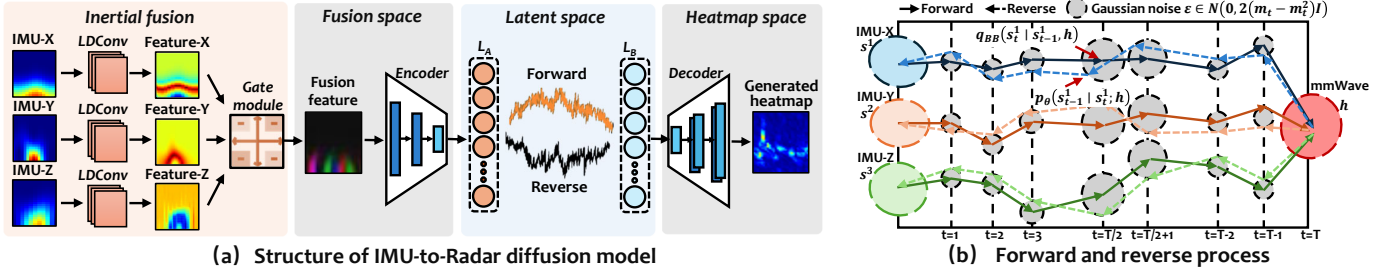


Fig. 13: I2R diffusion model.

(MC-MWHE), a mmWave feature enhancement algorithm based on image morphological operations. Specifically, we first apply Gaussian blurring to the original feature map to mitigate noise. Subsequently, we utilize K-means clustering to segregate the pixels into two discrete categories, as depicted in Fig. 11(b), where blue represents the gesture component and red indicates the dynamic noise component. Subsequently, pixels are reassigned based on the predominant clusters to accentuate the primary features. Following this, the feature map undergoes conversion to grayscale and binarization using a mean threshold. Finally, morphological closure operations are employed to bridge the discontinuities in the gesture component, which are marked in green. The final enhanced mmWave time-velocity heatmap is illustrated in Fig. 11(c).

2) *IMU Spectrogram Generation*: In iRadar, to isolate gesture-related signals, the acceleration data along the X, Y, and Z axes are decomposed into four levels utilizing maximal overlap discrete wavelet transform (MODWT). Finally, the three processed IMU signals are employed to derive three spectrograms using short-time Fourier transform (STFT), which are shown in Fig. 12(a)-(c), respectively. The obtained spectrograms from the three directions are utilized as the inputs of the inertial fusion module in Sec. III-B1.

### B. IMU-to-Radar Diffusion Model

This section introduces IMU-to-Radar (I2R), a diffusion model designed to convert IMU spectrograms to mmWave heatmaps. Fig. 13 shows the I2R structure, which consists of an inertial fusion module and a translation module.

1) *IMU Inertial Fusion*: IMU data contains a wealth of information related to signal frequency and motion intensity. Furthermore, the frequency spectrum shapes of IMU data can differ significantly between gestures. As a result, effectively modeling and extracting features from the IMU spectrogram poses a significant challenge. Consequently, we propose an inertial fusion module comprising a Learnable Dilated Convolutional Neural Network (LDCNN)—a novel convolutional approach to extract the features from IMU spectrograms. Additionally, a gating mechanism assigns varying weights to the features extracted by the LDCNN.

**LDCNN-based feature extraction.** Recall that the standard convolution operation which is characterized as  $O(\sigma) = \sum_{\sigma' \in S} I(\sigma + \sigma') * K(\sigma')$ , where  $O(\sigma)$  is the output feature map,  $I(\sigma)$  is the input feature map,  $K(\sigma')$  is the convolution kernel, and  $S$  is the neighborhood around the pixel  $\sigma$ . DCNN (Dilated Convolutional Neural Network) is designed to expand

the convolutional kernel by periodically inserting spaces (i.e., zeros) between the kernel elements [44]. As a result, the spacing between the elements' dilation rate can be described as  $O(\sigma) = \sum_{\sigma' \in S} I(\sigma + d \cdot \sigma') * K(\sigma')$ , where  $I$  represents the input feature map,  $K$  is the dilated convolution kernel, and  $d$  is the dilation rate. In LDCNN, the positions of non-zero elements within the convolutional kernel are learned using a gradient-based method. However, since the positions in the kernel are integer values, it poses a challenge in terms of differentiability. To overcome this issue, we utilize interpolation. The main motivation behind LDCNN is to explore the potential of enhancing the fixed grid imposed by the standard DConv by learning the spacing in an input-independent manner. Unlike the grid-like arrangement of convolutional kernel elements in standard and dilated convolutions, LDCNN allows for a flexible number of kernel elements [45]:  $O(\sigma) = \sum_{\sigma' \in S} I(\sigma + L(\sigma) \cdot \sigma') * K(\sigma')$ , where  $L(\sigma)$  is the learnable dilation rate function, which is updated through backpropagation.

**Gate module.** Following the LDCNN stage, the extracted features are fed into a gate module. The purpose of this module is to selectively fuse information, leveraging a gating mechanism to filter and combine pertinent features from the three orthogonal axes of IMU data. For IMU spectrogram features refined by the LDCNN, the gate module operates as  $G(s^1, s^2, s^3) = F(LDCNN(s^1, s^2, s^3); \eta)$ , where  $s^1, s^2, s^3$  is the input IMU spectrograms,  $G(\cdot)$  denotes the gated feature output,  $F(\cdot)$  is a fully connected fusion layer that integrates the derived features, and  $\eta$  is the learned parameters.

2) *Bridge Diffusion-based Translation*: Based on Brownian Bridge Diffusion (BBDM) [46], our approach incorporates a bilateral framework specifically designed to bridge the gap between IMU and mmWave data. By considering both the correlation and the unique characteristics of IMU spectrograms and mmWave heatmaps, BBDM effectively captures the nuanced mapping relationship between the two. Fig. 13 (b) illustrates the mathematical process proposed by the translation method, which includes the forward process and the reverse process.

**Forward process.** The forward process describes the diffusion of IMU spectrograms to mmWave heatmaps, which initiates from the IMU spectrograms and progressively incorporates noise and drift, gradually transitioning towards the mmWave heatmaps. The IMU spectrograms are represented by a set of inputs  $s = \{s^1, s^2, s^3\}$ . We let  $(s, h)$  denote the paired training data from IMU spectrograms and mmWave heatmaps. We take the ground truth mmWave heatmap conditional input  $h$  as its destination. It is assumed that  $s$  and  $h$  are approximately

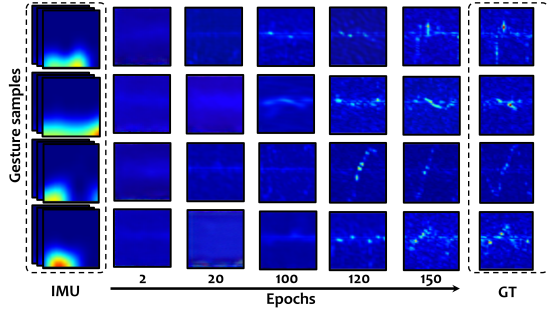


Fig. 14: Training progress for various gestures. Sequentially from top to bottom: front raise, lateral-to-front raise, push, and forearm supination.

independent and normally distributed as  $s, h \sim \mathcal{N}(0, I)$ . Given initial state  $s_0$  (as the blue, orange, green circles illustrated in Fig. 13 (b)), the intermediate state  $s_t$  (as the grey circle illustrated in Fig. 13 (b)) and destination state  $h$  (as the red circle illustrated in Fig. 13 (b)), the forward diffusion process of Brownian Bridge can be defined as:

$$q_{BB}(s_t | s_0, h) = \mathcal{N}(s_t; (1 - m_t)s_0 + m_th, \delta_t I), \quad m_t = \frac{t}{T}, \quad (6)$$

where  $T$  is the total steps of the diffusion process,  $\delta_t$  is the variance. For training and inference purposes, we need to deduce the forward transition probability  $q_{BB}(x_t | x_{t-1}, h)$  (as the grey line illustrated in Fig. 13 (b)):

$$q_{BB}(s_t | s_{t-1}, h) = \mathcal{N}\left(s_t; \frac{1-m_t}{1-m_{t-1}}s_{t-1} + \left(m_t - \frac{1-m_t}{1-m_{t-1}}m_{t-1}\right)h, \delta_{t|t-1}I\right). \quad (7)$$

According to Eq. 6, when the diffusion process reaches the destination, i.e.,  $t = T$ , we can get that  $m_T = 1$ . The forward diffusion process defines a fixed mapping from IMU spectrograms to mmWave heatmaps.

**Reverse process.** The reverse diffusion process can be utilized to infer the possible initial IMU spectrograms that could have resulted in the observed mmWave heatmaps by exploiting UNet to learn the mapping functions between IMU spectrograms and their corresponding mmWave heatmaps by minimizing the difference. It serves as the inverse of the forward diffusion process. Starting from the mmWave heatmaps, the backward diffusion process gradually eliminates the noise and drift through reverse operations, restoring the distribution towards the IMU spectrograms. Different from the existing diffusion models, the Brownian Bridge process directly starts from the conditional input by setting  $s_T = h$ . The reverse process aims to predict  $s_{t-1}$  based on  $s_t$ :

$$p_\theta(s_{t-1} | s_t, h) = \mathcal{N}(s_{t-1}; \mu_\theta(s_t, t), \delta_t I), \quad (8)$$

where  $\mu_\theta(s_t, t)$  is the predicted mean value of the noise, and  $\delta_t$  is the variance of noise at each step. The mean value  $\mu_\theta(s_t, t)$  is required to be learned by a neural network with parameters  $\theta$  based on the maximum likelihood criterion.

**Implementation.** We utilize input IMU spectrograms in the inertial fusion module as the source domain and mmWave heatmaps as the target domain to train the diffusion model. Additionally, we implement a U-Net neural network and employ it in the backward process. The U-Net architecture

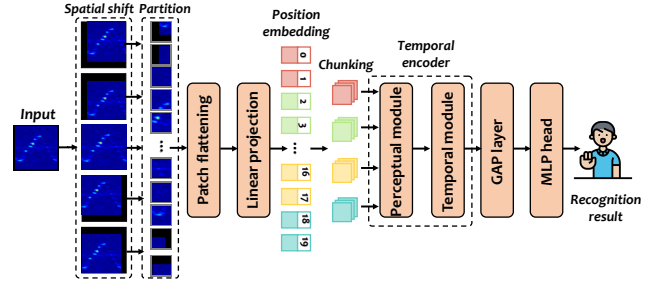


Fig. 15: Doppler transformer.

consists of four encoders and four decoders, utilizing ReLU as the activation function and max pooling for pooling operations. The inertial fusion module, forward process, and reverse process are closely connected as they are jointly trained to optimize the overall performance. As shown in Fig. 14, training progress from different gesture samples shows the convergence and stability of the model.

### C. Gesture Recognition

As outlined in Sec. I, accurate gesture recognition is a hard task due to the inherent complexity of arm movements and the diverse array of gesture patterns. Traditional approaches to feature extraction in gesture recognition often rely on convolutional neural networks [7], [29], [30], which may be inherently constrained by their local receptive fields and weight-sharing properties, potentially limiting their capacity to capture the global dependencies within complex gesture data. Drawing inspiration from the success of transformer architectures [31], [47], we introduce a novel Doppler transformer tailored for interpreting Doppler heatmaps of gestures in the following.

**Spatial heatmap shift and patch embedding.** As shown in Fig. 15, our approach enhances traditional vision transformers' limited receptive fields [31] by employing spatial heatmap patches shifted along various diagonal axes [47], resulting in an enriched representation of the time-doppler landscape through overlapping and merging with original heatmaps.

**Temporal attention layer.** To effectively learn heatmap details, we use the temporal attention mechanism to concentrate on the most significant temporal information. Specifically, we first divide the patch embedding into chunks by temporal sequence. Then we use a perceptual module to extract spatial features and a temporal module [48] to integrate these features over time, ensuring a comprehensive understanding of both space and time within the data.

**Implementation.** The model comprises two temporal transformer layers, each projecting patches into a 64-dimensional embedding space using a single attention head. It processes input data in chunks of eight. Training proceeds for 1000 epochs with a learning rate and weight decay both set at 0.001.

## IV. EVALUATION

### A. Experimental Methodology

**System implementation.** The setup for evaluating iRadar comprises the experimental devices depicted in Fig. 16. As illustrated in Fig. 16(a), data collection for mmWave sensing



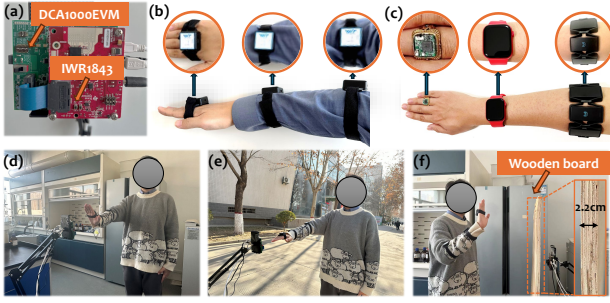


Fig. 16: Experimental setup.

TABLE I: Wearable device and mmWave radar specifications

Name	CPU Freq	RAM	OS	IMU Model*
WT901WIFI	168 MB	N/A	N/A	IS MPU9250
Smart Ring	64 MB	64 kB	N/A	IS MPU9250
Myo Armband	N/A	N/A	N/A	IS MPU9150
Apple Watch S7	1.8 GHz	1 GB	Watch OS 9	Unknown
Huawei Watch GT2	200 MHz	32 MB	Lite OS 11	STM LSM6DSO
Name	Start Freq	ADC Samples	Chirp Loops	Idle Time
TI IWR1843	77 GHz	256	255	100 $\mu$ s

\* IS: TDK InvenSense, STM: STMicroelectronics.

is conducted using a 77 GHz IWR1843 FMCW radar coupled with a DCA1000EVM real-time data capture adapter. Fig. 16(b) and (c) display the WT901WIFI motion sensor and four wearable sensors employed for acquiring IMU data, essential for the training and testing of the base model. In addition, our evaluation considers a variety of wearable technologies including a smart armband, smartwatch, and smart ring. Details on these devices can be found in Tab. I. The IMUs operate at a default sampling rate of 100 Hz.

**Data collection.** To validate iRadar, we enlisted 30 volunteers consisting of 17 females and 13 males, ranging in age from 15 to 64 years. All participants were in good health and took part in a series of controlled experiments<sup>1</sup>. The data collection spanned over a three-month period. Our study involved eighteen distinct gestures, as depicted in Fig. 17, encompassing a variety of wrist, elbow, and shoulder movements. To test the I2R diffusion model, we randomly selected half of the participants (15 individuals). Each participant executed all eighteen gestures 15 times while positioned in front of the mmWave radar equipped with an IMU sensor. These sessions were conducted under diverse conditions: indoors, outdoors, and through-obstacle scenarios, as illustrated in Fig. 16(d)(e)(f). For the evaluation of gesture recognition, the remaining 15 participants were instructed to complete two distinct sets of gesture trials. The first set aimed at gathering data for the translation recognition model, required participants to perform each of the eighteen gestures 15 times while carrying mobile devices across different settings, including indoor, outdoor, and through-obstacle environments. The second set focused on collecting gesture recognition data, with participants repeating each gesture 15 times in the presence of the radar.

### B. Overall Performance

**Overall accuracy.** Fig. 18(a) provides the accuracy percentages for gesture recognition across indoor, outdoor, and through-obstacle scenarios, utilizing the recognition model

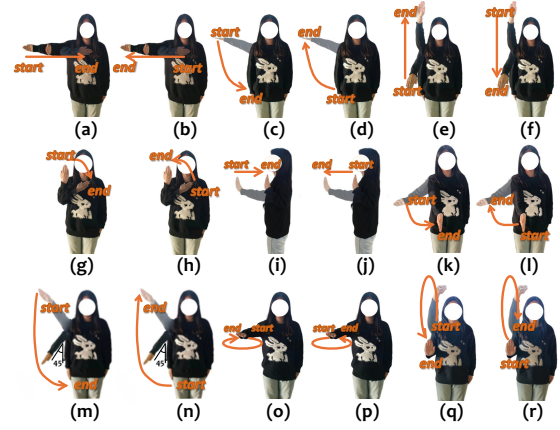
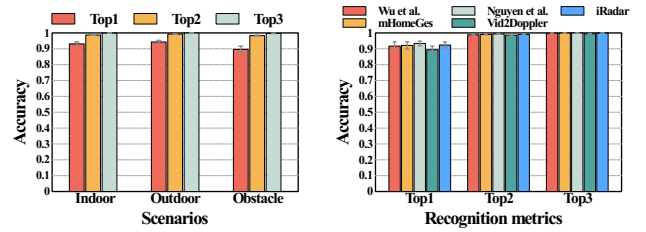


Fig. 17: Gestures in the dataset. 18 distinct gestures, including lateral-to-front raises, lateral raises, front raises, forearm supination/pronation, push, pull, swipes, 45° lateral raises, horizontal rotations, and vertical rotations.



(a) Overall accuracy.

(b) Comparison with baselines.

Fig. 18: Overall Performance.

trained with iRadar. The top-N accuracy criterion indicates the rate at which the correct gesture is identified within the top-N selections. Specifically, the Top-1 accuracy achieved in indoor settings was 93.1%, while outdoor settings saw a slightly higher of 94.3%. For through-obstacle conditions, the Top-1 accuracy is lower at 89.6%, attributable to the attenuating effects of obstacles on mmWave signal propagation. For Top-2 accuracy, the values remained notably high at 98.7% for indoor, 99.3% for outdoor, and 98.3% for through-obstacle scenarios. For Top-3 accuracies, with 99.8% for indoor, 99.9% for outdoor, and 99.7% for through-obstacle conditions. The results exhibit a high effectiveness of iRadar across a diverse array of environmental conditions.

**Comparison with baselines.** We compare iRadar with three types of state-of-the-art gesture recognition systems. (i) mmWave-based: Wu et al.'s system [49], utilizing mmWave Doppler heatmaps and mHomeGes [1], based on mmWave point clouds; and (ii) IMU-based: Nguyen et al.'s IMU-based system [50]; (iii) video to mmWave translation: Vid2Doppler [17] with video translated mmWave heatmaps. To ensure fairness, we adjusted each system to work best with our dataset. As presented in Fig. 18(b), iRadar is only 1.2% less accurate than the system by Nguyen et al., yet it is 0.7% more accurate than the system by Wu et al., 0.2% more accurate than mHomeGes, and 4% more accurate than Vid2Doppler. The IMU-based system shows higher accuracy because wearables are attached to the body, resulting in lower environmental noise. The results show that iRadar achieves comparable accuracy to the state-of-the-art systems.

<sup>1</sup>Ethical approval has been obtained from the corresponding organization.

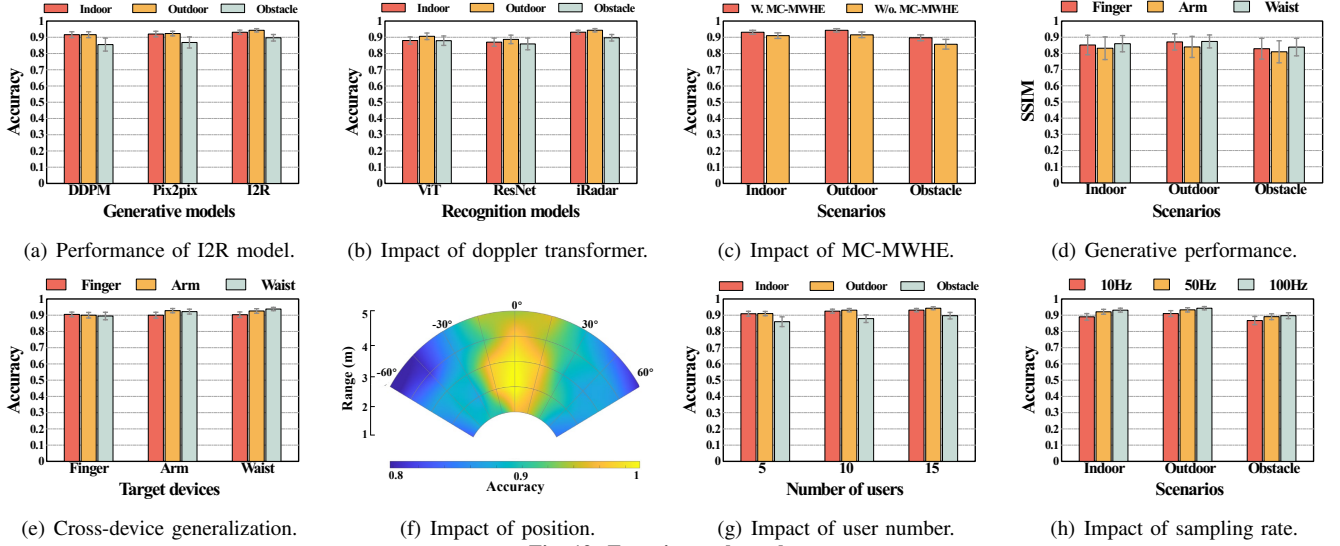


Fig. 19: Experimental results.

### C. Micro-Benchmark Evaluation

**Performance of I2R diffusion model.** The effectiveness of the I2R diffusion model is shown in Fig. 19(a), where we compare the accuracy utilizing the I2R model against those employing DDPM [26] and pix2pix [51]. For integrating the IMU inputs, both DDPM and pix2pix are accompanied by an inertial fusion module. The results indicate that I2R surpasses the comparative models in every environment. Notably, I2R achieves an increase in accuracy over DDPM by 1.64%, 2.94%, and 4.91% in indoor, outdoor, and through-wall settings, respectively. When measured against pix2pix, we observe accuracy improvements of 1.20%, 2.17%, and 3.41%.

**Evaluation on Doppler transformer.** We then assess the proposed gesture recognition model, the Doppler transformer, by benchmarking its accuracy against that of the Vision Transformer (ViT) [31] and the Residual Network (ResNet) [52] across various settings, including indoor, outdoor, and through-wall scenarios. Fig. 19(b) presents a comparison of the performance of different models in these environments. The results show that iRadar exceeds the mean accuracy of ViT and ResNet by margins of 4.13% and 5.73%, respectively. This substantial improvement is indicative of iRadar's advanced capability to capture and interpret gesture-related features.

**Evaluation on MC-MWHE method.** The impact of our proposed mmWave heatmap enhancement technique, MC-MWHE, was assessed by comparing performance metrics with and without the application of this method. As depicted in Fig. 19(c), the deployment of MC-MWHE resulted in accuracy improvements of 2.3%, 3.3%, and 4.7% for indoor, outdoor, and through-obstacle scenarios, respectively. These enhancements underscore the effectiveness of MC-MWHE in noise reduction and the overall refinement of the recognition.

**Generative information loss.** The generative results of the I2R diffusion model, which transforms IMU spectrograms into mmWave radar heatmaps, are presented in Fig. 19(d). We use the Structural Similarity Index Measure (SSIM) [41] to assess the similarity between the generated heatmaps and their

TABLE II: Cross-scenario results.

IMU Sce.*			mmWave Sce.			Accuracy		
I.D.	O.D.	T.O.	I.D.	O.D.	T.O.	Top-1	Top-2	Top-3
●	○	○	●	○	○	93.12%	98.54%	99.71%
○	●	○	●	○	○	93.21%	98.67%	99.81%
○	○	●	○	○	○	89.32%	98.22%	99.63%
●	○	○	○	●	○	92.43%	98.99%	99.91%
○	●	○	○	○	○	94.32%	98.93%	100%
○	○	●	○	○	○	88.23%	98.31%	99.55%
●	○	○	○	○	●	89.47%	98.35%	99.67%
○	●	○	○	○	○	88.26%	98.23%	99.54%
○	○	●	○	○	○	89.71%	98.41%	99.81%

\* ● for chosen, ○ for unchosen, I.D.: indoor, O.D.: outdoor, T.O.: through-obstacle.

authentic counterparts. The I2R model demonstrates superior performance, with SSIM scores indicating high levels of similarity in different scenarios: in indoor settings, the model achieves 85.21%, 83.27%, and 86.01% for three different wearable devices; in outdoor environments, SSIM scores are 87.12%, 84.07%, and 87.39%; and through-obstacle conditions resulted in 82.92%, 81.18%, and 83.98%. These results are indicative of the model's ability to produce outputs that are closely aligned with the original mmWave signal.

**Generalization ability.** The ability of iRadar to adapt to diverse environments and wearable device placements is rigorously evaluated. Tab. II displays the system's recognition capabilities across a range of IMU and mmWave data collection scenarios. The table illustrates that iRadar has a consistently high level of accuracy in various settings, with all Top-3 accuracy surpassing 99.5%. It is observed, however, that performance slightly dips in through-obstacle scenarios due to the impact of obstructions on the mmWave signal. Conversely, the system excels in outdoor scenarios, benefiting from the absence of obstructions and reduced interference. To further assess iRadar's adaptability to different wearable device positions, we establish a baseline using a model from one wearable position and then enhance models for two additional positions with ten epochs of further training. The results, as depicted in Fig. 19(e), show that each position attains an accuracy of over 89.5%. This underscores the model's robust generalization across wearable positions.

**Evaluation on user position.** Fig. 19(f) illustrates the in-



fluence of the user's relative position to mmWave radar on the accuracy of iRadar, with distances ranging from 1 m to 5 m and angular displacement spanning  $-60^\circ$  to  $60^\circ$ . It is observed that the system's accuracy initially increases with distance but subsequently diminishes. To be precise, average accuracy increases by 12.43% when the distance extends from 1 m to 3 m and then declines by 7.56% as the distance further grows to 5 m. This suggests that at closer proximities, the mmWave radar's sensitivity to user orientation can negatively impact recognition accuracy. Additionally, the system exhibits better performance within  $-30^\circ$  to  $30^\circ$ , because users are more prominently within the radar's optimal sensing radius.

**Evaluation on user number.** In our investigation of the effect of user group size on recognition accuracy within iRadar, we observe notable trends as illustrated in Fig. 19(g). Accuracy exhibits a positive correlation with group size, where an increase from 5 to 10 members resulted in a 2.45% improvement in accuracy, and a further increase to 15 members led to an additional 1.13% improvement. This enhancement can be attributed to the greater information of distinguishable gestures present within larger groups, which contributes to the system's ability to correctly recognize them. iRadar achieves over 88% accuracy even in groups as small as five.

**Evaluation on sampling rate.** We then delve into the relationship between the sampling rate and the accuracy of recognition results in iRadar. We present our findings for sampling rates of 10 Hz, 50 Hz, and 100 Hz. Fig. 19(h) illustrates that when the sampling rate is reduced from 100 Hz to 50 Hz, there is a marginal decrease in accuracy by approximately 1.13%. A further reduction of the sampling rate to 10 Hz leads to a more pronounced decline of about 2.60% in accuracy. This significant drop is largely due to the omission of vital gesture data that occurs at lower sampling rates, thereby negatively affecting the system's ability to recognize patterns accurately.

**Evaluation on time-varying performance.** Human gestures are inherently variable, which requires a thorough evaluation of the temporal reliability of iRadar. We engaged five individuals using a smartwatch continuously over two months, with evaluations every ten days. As shown in Fig. 20, a gradual decrease in recognition accuracy was observed. More precisely, there was an average reduction in accuracy of 2.17% at the one-month mark and an additional 2.55% reduction by the end of the second month. The observed decline can be ascribed to natural variations in the users' physiological characteristics, such as muscle tone and joint flexibility, which subtly alter gesture patterns. Despite these changes, iRadar maintained an average accuracy rate of 88.07% after 60 days, showcasing its resilience. To maintain efficacy over time, methods like continuous learning are recommended [53].

## V. RELATED WORK

**Cross-modality translation for mmWave.** Current research in cross-modality translation for mmWave predominantly focuses on generating mmWave signals from videos [17]–[19]. Notably, Vid2Doppler [17] converts human activity captured in videos into highly realistic synthesized mmWave radar data via

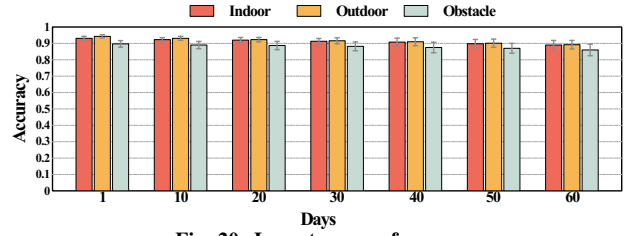


Fig. 20: Long-term performance.

a transformer-based model. Similarly, the Midas system [19] employs an enhanced transformer model in conjunction with VS-Net to generate believable radar data and pinpoint salient video segments. SynMotion [18] uses existing video datasets to translate video information into synthetic mmWave data for human motion sensing. Despite their advancements, these video-based methods are vulnerable to environmental factors such as occlusions and varying lighting conditions. To overcome these challenges, iRadar leverages less susceptible IMU data for modality translation.

**mmWave-based gesture recognition.** Current mmWave-based gesture recognition methods are broadly divided into heatmap-based approaches [29], [54] and point cloud-based approaches [1], [6], [55]. Yuan et al. [29] utilized CNNs to recognize digital gestures from hand motion trajectories depicted in heatmaps. On the other hand, point cloud-based methods, such as Pantomime [6], process sparse 3D point clouds derived from radar signals through deep learning frameworks. Similarly, mHomeGes [1] recognizes arm gestures in real-time by reconstructing point clouds and employing shallow neural networks for classification. Despite the progress, point cloud-based methods often struggle with sparsely filled point clouds that fail to capture gestures with high fidelity, as evaluated by recent study [56]. Additionally, traditional heatmap-based approaches that rely on CNNs may be constrained by their local receptive fields and shared weights, which limits their capacity to understand gestural data.

## VI. CONCLUSION

In this work, we introduce iRadar, which, to the best of our knowledge, is the first framework for cross-modal IMU-to-mmWave gesture recognition that circumvents the necessity of additional mmWave hardware installation and obviates the need for explicit data collection, substantially alleviating the service providers' burden. iRadar encompasses a diffusion-driven translation method, a novel mmWave heatmap enhancement technique, and a Doppler transformer recognition algorithm. Our comprehensive evaluation reveals that iRadar achieves an average Top-3 accuracy rate of 99.82%.

## ACKNOWLEDGMENT

This project was supported by National Key R&D Program of China (Grant No. 2023YFE0208800), the Research Grants Council of the Hong Kong SAR, China (Project No. CityU 11202124 and CityU 11201422), NSF of Guangdong Province (Project No. 2024A1515010192), the Innovation and Technology Commission of Hong Kong (Project No. MHP/072/23). \*Weitao Xu is the corresponding author.

## REFERENCES

- [1] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *ACM IMWUT*, 2020.
- [2] K. A. Smith, C. Csech, D. Murdoch, and G. Shaker, "Gesture recognition using mmwave sensor for human-car interface," *IEEE Sens. Lett.*, 2018.
- [3] A. J. Akbar, Z. Sheng, Q. Zhang, and D. Wang, "Cross-domain gesture sequence recognition for two-player exergames using cots mmwave radar," *ACM HCI*, 2023.
- [4] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *IEEE INFOCOM*, 2015.
- [5] W. He, K. Wu, Y. Zou, and Z. Ming, "Wig: Wifi-based gesture recognition system," in *IEEE ICCCN*, 2015.
- [6] S. Palipana, D. Salami, L. A. Leiva, and S. Sigg, "Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds," *ACM IMWUT*, 2021.
- [7] H. Liu, K. Cui, K. Hu, Y. Wang, A. Zhou, L. Liu, and H. Ma, "Mtranssee: Enabling environment-independent mmwave sensing based gesture recognition via transfer learning," *ACM IMWUT*, 2022.
- [8] S. Bhalla, M. Goel, and R. Khurana, "Imu2doppler: Cross-modal domain adaptation for doppler-based activity recognition using imu data," *ACM IMWUT*, 2021.
- [9] Y. Li, D. Zhang, J. Chen, J. Wan, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, "Towards domain-independent and real-time gesture recognition using mmwave signal," *IEEE TMC*, 2022.
- [10] Y. Wang, J. Shen, and Y. Zheng, "Push the limit of acoustic gesture recognition," *IEEE TMC*, 2020.
- [11] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *IJCI*, 2009.
- [12] Y. Zhu, Z. Yang, and B. Yuan, "Vision based hand gesture recognition," in *IEEE ICSS*, 2013.
- [13] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, 2015.
- [14] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," *IEEE THMS*, 2014.
- [15] P.-G. Jung, G. Lim, S. Kim, and K. Kong, "A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors," *IEEE THI*, 2015.
- [16] R. Wu, H. Yang, and W. Xu, "Xsolar: A generative framework for solar-based human gesture sensing via wearable signals," in *BodySys*, 2024.
- [17] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2doppler: Synthesizing doppler radar data from videos for training privacy-preserving activity recognition," in *ACM CHI*, 2021.
- [18] X. Zhang, Z. Li, and J. Zhang, "Synthesized millimeter-waves for human motion sensing," in *ACM SenSys*, 2022.
- [19] K. Deng, D. Zhao, Q. Han, Z. Zhang, S. Wang, A. Zhou, and H. Ma, "Midas: Generating mmwave radar data from videos for training pervasive and privacy-preserving human sensing tasks," *ACM IMWUT*, 2023.
- [20] M. Georgi, C. Amma, and T. Schultz, "Recognizing hand and finger gestures with imu based motion and emg based muscle activity sensing," in *Scitepress BIOSIGNAL*, 2015.
- [21] D. Jirak, S. Tietz, H. Ali, and S. Wermter, "Echo state networks and long short-term memory for continuous gesture recognition: A comparative study," *Springer Cognitive Computation*, 2023.
- [22] H. Xue, Q. Cao, C. Miao, Y. Ju, H. Hu, A. Zhang, and L. Su, "Towards generalized mmwave-based human pose estimation through signal augmentation," in *ACM MobiCom*, 2023.
- [23] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi," in *ACM MobiHoc*, 2017.
- [24] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single wi-fi link," in *ACM MobiSys*, 2018.
- [25] S. Ding, Z. Chen, T. Zheng, and J. Luo, "Rf-net: A unified meta-learning framework for rf-enabled one-shot human activity recognition," in *ACM SenSys*, 2020.
- [26] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *NIPS*, 2020.
- [27] B. Kavar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," *NIPS*, 2022.
- [28] A. Freivalds, *Biomechanics of the upper limbs: mechanics, modeling and musculoskeletal injuries*. CRC press, 2011.
- [29] C. Yuan, Y. Zhong, J. Tian, and Y. Zou, "A real-time digit gesture recognition system based on mmwave radar," in *IEEE ICMLA*, 2022.
- [30] Q. Chen, Y. Li, Z. Cui, and Z. Cao, "A hand gesture recognition method for mmwave radar based on angle-range joint temporal feature," in *IEEE IGARSS*, 2022.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [32] S. Rao, "Introduction to mmwave sensing: Fmcw radars," *Texas Instruments (TI) mmWave Training Series*, 2017.
- [33] M. Han, H. Yang, T. Ni, D. Duan, M. Ruan, Y. Chen, J. Zhang, and W. Xu, "mmsign: mmwave-based few-shot online handwritten signature verification," *ACM TOSN*, 2023.
- [34] S. Zhang, T. Zheng, H. Wang, Z. Chen, and J. Luo, "Quantifying the physical separability of rf-based multi-person respiration monitoring via sinr," in *ACM SenSys*, 2022.
- [35] S. Zhang, T. Zheng, Z. Chen, and J. Luo, "Can we obtain fine-grained heartbeat waveform via contact-free rf-sensing?" in *IEEE INFOCOM*, 2022.
- [36] S. Ji, X. Zhang, Y. Zheng, and M. Li, "Construct 3d hand skeleton with commercial wifi," in *ACM SenSys*, 2023.
- [37] H. Yang, M. Han, M. Jia, Z. Sun, P. Hu, Y. Zhang, T. Gu, and W. Xu, "Xgait: Cross-modal translation via deep generative sensing for rf-based gait recognition," in *ACM SenSys*, 2023.
- [38] M. Han, H. Yang, M. Jia, W. Xu, Y. Yang, Z. Huang, J. Luo, X. Cheng, and P. Hu, "Seeing the invisible: Recovering surveillance video with cots mmwave radar," *IEEE TMC*, 2024.
- [39] H. Yang, M. Han, S. Shi, Z. Yan, G. Xing, J. Wang, and W. Xu, "Wave-for-safe: Multisensor-based mutual authentication for unmanned delivery vehicle services," in *ACM MobiHoc*, 2023.
- [40] K. Cui, Q. Yang, Y. Zheng, and J. Han, "mmripple: Communicating with mmwave radars through smartphone vibration," in *ACM/IEEE IPSN*, 2023.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TIP*, 2004.
- [42] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, and Z. Wu, "Accelerometer-based gait recognition by sparse representation of signature points with clusters," *IEEE Trans. Cybern.*, 2014.
- [43] F. Adib and D. Katabi, "See through walls with wifi!" in *ACM SIGCOMM*, 2013.
- [44] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [45] I. Khalfaoui-Hassani, T. Pellegrini, and T. Masquelier, "Dilated convolution with learnable spacings," in *ICLR*, 2023.
- [46] B. Li, K. Xue, B. Liu, and Y.-K. Lai, "Bbmd: Image-to-image translation with brownian bridge diffusion models," in *IEEE/CVF CVPR*, 2023.
- [47] S. H. Lee, S. Lee, and B. C. Song, "Vision transformer for small-size datasets," *arXiv preprint arXiv:2112.13492*, 2021.
- [48] A. Didolkar, K. Gupta, A. Goyal, N. B. Gundavarapu, A. M. Lamb, N. R. Ke, and Y. Bengio, "Temporal latent bottleneck: Synthesis of fast and slow processing mechanisms in sequence learning," *NIPS*, 2022.
- [49] J. Wu, J. Wang, Q. Gao, M. Cheng, M. Pan, and H. Zhang, "Toward robust device-free gesture recognition based on intrinsic spectrogram of mmwave signals," *IEEE IoT-J*, 2022.
- [50] K. Nguyen-Trong, H. N. Vu, N. N. Trung, and C. Pham, "Gesture recognition using wearable sensors with bi-long short-term memory convolutional neural networks," *IEEE Sens. Lett.*, 2021.
- [51] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE CVPR*, 2017.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016.
- [53] T. Anderson, *The theory and practice of online learning*. Athabasca University Press, 2008.
- [54] B. Yan, P. Wang, L. Du, X. Chen, Z. Fang, and Y. Wu, "mmgesture: Semi-supervised gesture recognition system using mmwave radar," *Expert Systems with Applications*, 2023.
- [55] D. Salami, R. Hasibi, S. Palipana, P. Popovski, T. Michoel, and S. Sigg, "Tesla-rapture: A lightweight gesture recognition system from mmwave radar sparse point clouds," *IEEE TMC*, 2022.
- [56] P. Jiang, E. Fassman, and T. Li, "Evaluating the impact of noisy data on time-sensitive point clouds from millimeter wave gesture recognition systems," 2023.