

# Multi-modal Fusion Sensing: A Comprehensive Review of Millimeter-Wave Radar and its Integration with other Modalities

Shuai Wang, *Senior Member, IEEE*, Luoyu Mei, *Member, IEEE*, Ruofeng Liu, *Member, IEEE*, Wenchao Jiang, *Member, IEEE*, Zhimeng Yin, *Member, IEEE*, Xianjun Deng, *Senior Member, IEEE*, Tian He, *Fellow, IEEE*

**Abstract**—Millimeter-wave (mmWave) radar, with its high resolution, sensitivity to micro-vibrations, and adaptability to various environmental conditions, holds immense potential across multi-modal fusion sensing. Although there exist review papers on mmWave radar, there is a noticeable lack of comprehensive reviews focusing on its multi-modal fusion sensing capabilities. Addressing this gap, our review offers an extensive exploration of mmWave radar multi-modal fusion sensing, emphasizing its integration with other modalities. This review discusses the complex realm of millimeter-wave radar multi-modal fusion sensing, detailing its importance, hardware and software aspects, principles, and current potential and applications. It delves into data characteristics and datasets associated with mmWave radar, focusing on Doppler, point cloud, and multi-modal data formats. The review highlights how these data types enhance multi-modal fusion sensing and discusses methodologies, including signal processing and learning algorithms. Three categories of multi-modal fusion methodologies are proposed to optimally manage and interpret fused data. Various practical applications of mmWave radar multi-modal fusion sensing are illustrated, underlining the unique capabilities it provides when integrated with other sensors. The review concludes by identifying potential future research avenues, underscoring the immense potential of this field for further exploration and advancement.

**Index Terms**—Multi-modal Fusion Sensing, Millimeter-wave Radar, Review

## I. INTRODUCTION

Millimeter-wave (mmWave) radar is a high-precision sensor for contactless targets' range, velocity, and angle measurements. With the recent advance in radar technologies, mmWave sensors have been widely utilized in many scenarios, e.g., human sensing [1], industrial monitoring [2], [3], and autonomous vehicle sensing [4], [5], [6]. A mmWave radar emits electromagnetic signals with wavelengths in the millimeter range (1 to 10mm) corresponding to a frequency range of

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62272098. (Shuai Wang and Luoyu Mei are co-first authors.) (Corresponding author: Shuai Wang.)

S. Wang, L. Mei, and T. He are with the Southeast University. E-mail: shuaiwang@seu.edu.cn, lymei@seu.edu.cn, tianhe@seu.edu.cn.

R. Liu is with the University of Minnesota Twin Cities. E-mail: liux4189@umn.edu.

W. Jiang is with the Singapore University of Technology and Design. E-mail: wenchao\_jiang@sutd.edu.sg.

Z. Yin, and L. Mei is with the City University of Hong Kong. E-mail: zhimeyin@cityu.edu.hk, luoyumei2-c@my.cityu.edu.hk.

X. Deng is with the Huazhong University of Science and Technology. E-mail: dengxj615@hust.edu.cn.

30 to 300 GHz. The short wavelength of the mmWave radar provides several advantages, such as the small size of antennas and the ability to detect millimeter-level resolution.

Millimeter-wave radar sensing exhibits advantages over other sensing modalities, including Wi-Fi, camera, and LiDAR in various aspects. It demonstrates exceptional precision and resolution, operates independently of ambient light conditions [7], features penetration capabilities through obstacles [8], [9], and ensures enhanced privacy protection [10]. Compared to Wi-Fi sensing, which is prone to signal interference and degradation, mmWave radar has high bandwidth and offers superior sensing accuracy [11]. In contrast to cameras, mmWave radar's capabilities extend to operation under diverse lighting conditions, obstruction penetration, and superior privacy due to its non-visual nature [12]. When compared with LiDAR, a technology that encounters increased errors due to atmospheric particulates, mmWave radar displays increased robustness and stability [13]. Despite these strengths, mmWave radar also encounters challenges, such as sparse point clouds and specular reflection, which need to be addressed [14].

The challenges facing individual sensor technologies illuminate a promising opportunity: multi-modal fusion. The state-of-the-art solutions architect systems that embody complementarity, by fusing mmWave radar with other sensing technologies, such as cameras and LiDAR. The fusion of mmWave radar's all-weather [15], lighting-agnostic capabilities [16], with high-resolution, color-rich visual data from cameras, proficient in object recognition and classification [17], yields a more resilient and comprehensive sensing system. For instance, the integration of mmWave radar with LiDAR [18], [19], recognized for its high-resolution, three-dimensional point cloud data, results in a heightened level of accuracy and reliability in the sensing system. Here, the mmWave radar compensates for LiDAR's vulnerabilities to adverse weather conditions, contributing to its robustness against weather variations [16], and supplementing accurate velocity information [20]. These multi-modal fusion sensing systems adeptly exploit each sensor's unique strengths whilst concurrently addressing their individual limitations. The fusion of these sensing modalities cultivates a more integrated and robust perception system. This approach not only skillfully mitigates the inherent constraints of each individual sensor, but also capitalizes on their synergistic capabilities to amplify multi-modal perception [21]. This lays the groundwork for a new

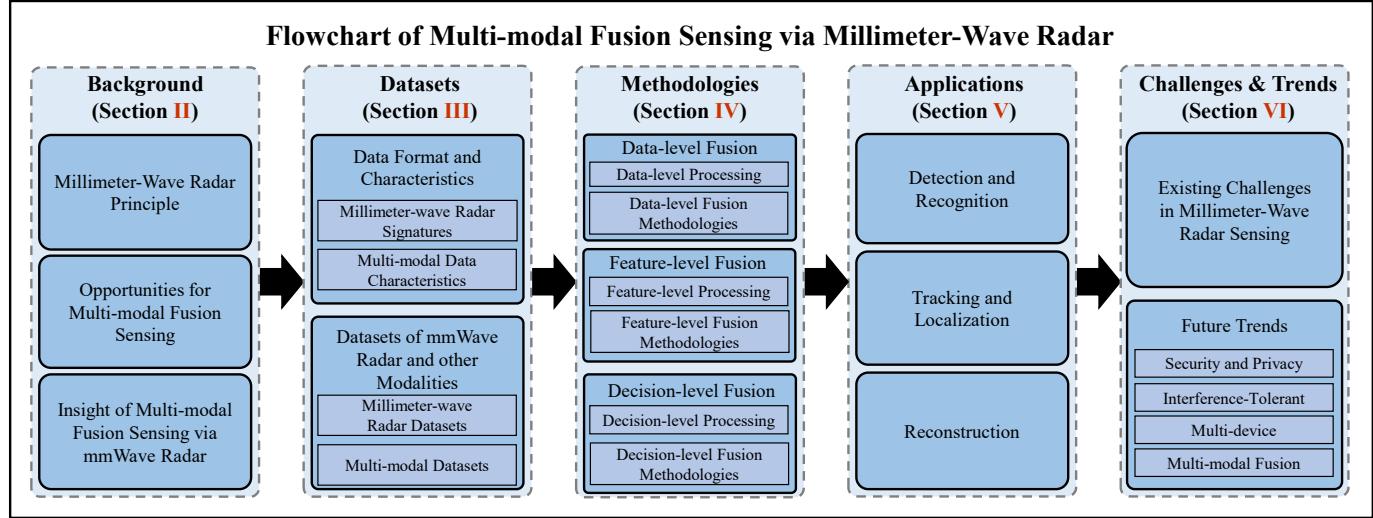


Fig. 1. Review organization

frontier of multi-modal fusion sensing [22].

While several reviews on mmWave radar fusion exist [23], [24], [25], they primarily focus on only a part of the multi-modal fusion landscape. Building upon the foundation laid by prior reviews, this paper distinguishes itself by presenting an exhaustive exploration of multi-modal fusion sensing techniques with a specific focus on the integration of mmWave radar. Specifically, we consider the fusion of mmWave radar and more modalities, including vision, lidar, near-infrared, and the inertial measurement unit. Even in the mmWave and vision fusion, we divide the fusion methods into three levels: data level, feature level, and decision level, while the existing surveys do not have such a hierarchy. Furthermore, we introduce the applications of mmWave radar in human sensing, industry sensing, and autonomous vehicle sensing.

This review presents a comprehensive examination of millimeter-wave radar multi-modal fusion sensing, spanning its underlying principles, data characteristics, methodologies, applications, challenges, and future directions, as elaborated across various sections. The organization of the review is illustrated in Fig. 1. Section II is background, delineates the hardware, and elucidates the principle governing multimodal fusion sensing of mmWave radar. Section III delves into the unique data traits of mmWave radar, including Doppler (micro-Doppler, range-Doppler), point-cloud, and multi-modal data, supplemented by overviews of existing datasets for each data type. Section IV outlines the methodologies employed in mmWave radar, spanning processing and fusion methodologies. The processing and fusion methodologies are subdivided into three tiers: data-level, feature-level, and decision-level methodologies. Section V delves into the pivotal application scenarios where mmWave radar sensing is employed, namely in human sensing, industrial sensing, and autonomous vehicle sensing. In order to facilitate a structured overview, we categorize these applications according to their functional types: detection and recognition, tracking and localization, and reconstruction. Ultimately, Section VI shows the challenges and future trends of mmWave radar sensing, both as a standalone modality and as part of multi-modal fusion sensing.

## II. BACKGROUND

This section provides an introduction to the essential concepts of multi-modal fusion sensing via millimeter-wave (mmWave) radar, with a specific emphasis on Frequency-Modulated Continuous-Wave (FMCW) radar. FMCW radar is distinguished by its high flexibility and cost-effectiveness, making it the predominant technology in mmWave radar applications [26]. It is utilized to recognize a variety of objects, extending well beyond basic range and velocity measurements. Despite the existing Orthogonal Frequency-Division Multiplexing (OFDM) radars, it is more complex and costly compared to the FMCW radar. OFDM radar systems typically require more sophisticated signal processing techniques and higher computational resources to handle the simultaneous transmission of multiple frequency signals [23]. The increased complexity leads to higher power consumption and reduced efficiency in certain applications [24].

Reflecting these considerations, the commercial mmWave radars available on the market today are mainly based on FMCW technology [27], which offers a better trade-off between performance and cost [28]. This capability, combined with its lower cost relative to other mmWave radar technologies, renders FMCW radar particularly suitable for multi-modal fusion sensing applications. Given its widespread utilization and significant advantages, this review primarily focuses on FMCW mmWave radar.

The fundamental equations for the transmitted FMCW radar signal are represented as:

$$S^T(t) = \cos(2\pi(f_c t + \frac{Bt^2}{2T})). \quad (1)$$

Here  $S^T(t)$  is the transmitted signal,  $f_c$  is the carrier frequency,  $B$  is the bandwidth of the chirp, and  $T$  is the chirp duration. The received signal, which reflects off the target and returns to the radar after a delay  $\tau$ , is expressed as:

$$S^R(t) = \alpha \cos(2\pi(f_c(t - \tau) + \frac{B(t - \tau)^2}{2T})). \quad (2)$$

Here  $\alpha$  is the attenuation factor and  $\tau$  is the delay of the signal round trip between the radar and the target. The output of the mixer multiplies the received signal with the conjugate of the transmitted signal, resulting in an intermediate frequency (IF) signal that contains the range and velocity information of the target. The IF signal is expressed as Equation 3.

Following this introduction, we introduce the fundamental principles of the mmWave radar in FMCW systems, covering its ability to measure range, velocity, and angles. We then examine the hardware components employed for multi-modal fusion sensing, providing a summary of the features and characteristics of sensors from each modality, along with the opportunities for sensor fusion. Lastly, we present an insight into multi-modalities, emphasizing the objectives and methodologies associated with multi-modal fusion sensing.

#### A. Millimeter-Wave Radar Principle

A mmWave radar functions by propagating a high-frequency mmWave signal towards objects. This signal undergoes scattering when it interacts with the objects, reflecting a portion of the scattered waves back to the radar, which is then captured by the receiving antenna. The backscattered signal encapsulates multi-dimensional data about the object across time, spatial, and frequency dimensions [29]. These data attributes are subsequently processed to enable the radar's core capabilities, namely, range, velocity, and angle measurements.

**Antenna array:** In mmWave radar, the antenna size is typically half the wavelength of the millimeter waves, which allows for compact and effective sensing. Smaller antenna arrays enable more compact and energy-efficient radar units, which is particularly important for applications that require portable or battery-powered devices [30]. Additionally, mmWave radar components, including the antenna array, are often integrated into a system-on-chip (SoC) design [31]. This integration further enhances the radar unit's compactness, reduces power consumption, and improves overall system efficiency. By carefully optimizing the bandwidth and antenna array size, along with leveraging SoC integration, mmWave radar designers create high-performance, portable, and power-efficient imaging solutions that meet the demands of various applications.

**Range measurement:** To estimate the range of the target, the radar measures the Time of Flight (ToF) of the signal by comparing it with the transmission. In particular, the received signal  $S^R(t)$  is multiplied by the transmitted signal and the phase-shifted transmitted signal  $\sin(2\pi(f_{ct} + \frac{B\tau}{2T})t)$  and then low-pass filtered. This yields a complex intermediate frequency (IF) signal denoted as:

$$S^M(t) = e^{j2\pi(\frac{B\tau}{T}t + f_c\tau)}. \quad (3)$$

Here  $\tau$  represents the Time of Flight, the IF signal has a single frequency tone and the time of flight is obtained from its tone  $f_{IF} = \frac{B\tau}{T} = S\tau$ , where  $S = \frac{B}{T}$  is the slope of the chirp. The range between the object and the radar (denoted as  $d$ ) is then calculated by multiplying half of the ToF by the speed of light (denoted as  $c$ ), that is,  $d = \frac{\tau c}{2}$ .

The range resolution  $d_{Res}$  is the minimum distance at which the radar distinguishes two separate targets. From the IF signal

in Equation 3, we infer that the frequency of the IF signal is directly proportional to the ToF  $\tau$ . For two distinct targets, the difference in their ToF must be at least one IF frequency bin to be resolved, which corresponds to the inverse of the signal bandwidth  $B$ . Thus, the ToF resolution is  $\tau_{Res} = \frac{1}{B}$ , and since the range is  $d = \frac{\tau c}{2}$ , we substitute  $\tau_{Res}$  to obtain the range resolution as  $d_{Res} = \frac{\tau_{Res}c}{2} = \frac{c}{2B}$ .

**Velocity measurement:** The velocity of the target is determined from the phase differences of the IF signal between the chirps (as depicted in Fig. 2(a)). When a target moves at a velocity of  $v_t$  relative to the radar, the target moves  $Tv_t$  in a single chirp duration, leading to an increase in the round-trip time  $\Delta\tau = \frac{2Tv_t}{c}$ . As a result, the phase of the IF signal (equation 3) changes by  $\Delta\phi = \frac{4\pi\Delta\tau}{\lambda}$ , where  $\lambda = \frac{c}{f_c}$  is the wavelength. The phase difference is utilized to obtain the small change of distance between chirps, i.e.,  $\Delta d = \frac{\lambda\Delta\phi}{4\pi}$ , which yields the velocity  $v_t = \frac{\lambda\Delta\phi}{4\pi T}$ . In practice, the radar transmits a sequence of chirps to measure velocity.

The velocity resolution  $v_{Res}$  is the minimum velocity difference the radar resolves between two targets with Doppler shifts. The phase difference  $\Delta\phi$  caused by the Doppler effect is a product of the velocity  $v_t$ . If we consider  $\Delta\phi_{Res}$  to be the minimum phase difference that leads to a resolvable frequency shift in the IF signal, we obtain  $v_{Res}$  by rearranging the velocity-phase relationship to  $v_{Res} = \frac{\lambda\Delta\phi_{Res}}{4\pi T}$ . Since the Doppler shift is discretized into bins of size  $\frac{1}{NT}$  (the inverse of the observation time for  $N$  chirps), the minimum resolvable phase shift is  $\Delta\phi_{Res} = \frac{2\pi}{N}$ , leading to  $v_{Res} = \frac{\lambda}{2NT}$ . In practice, the number of chirps  $N$  is limited by the mmWave radar hardware characteristics, such as the sampling rate, the memory, and the power consumption. Therefore, the velocity resolution cannot be arbitrarily high. Typically, the number of chirps  $N$  is in the range of 16 to 128 [32], depending on the application and the radar configuration.

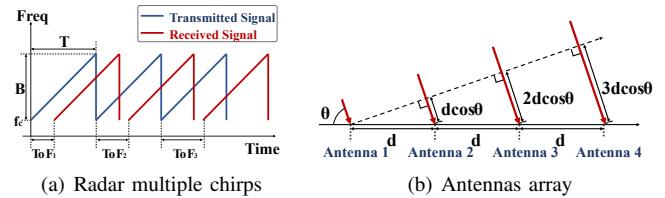


Fig. 2. (a) The range and velocity are computed by utilizing the single chirp Time of Flight (ToF) and cross-chirp ToF differences, respectively. (b) The angle of arrival (AoA) is determined by the cross-antenna signal variances.

**Angle measurement:** Finally, the angle measurement is achieved by estimating the phase differences of the IF signal across antennas. As Fig. 2(b) depicts, the angle of arrival (denoted as  $\theta$ ) leads to a ToF difference of  $l\cos(\theta)$  between adjacent antennas, where  $l$  is the distance between antennas. Similar to the velocity, the small difference of ToF causes a phase difference of the IF signal  $\Delta\phi$ . The angle  $\theta$  is calculated using trigonometry:  $\theta = \cos^{-1}(\frac{\lambda\Delta\phi}{2\pi l})$ .

The angle measurement resolution, denoted as  $\theta_{Res}$ , represents the minimum angular separation at which a radar system distinguishes between two targets originating from different directions. The phase difference between signals  $\Delta\phi$  at two

adjacent antennas, separated by a distance  $l$ , is a function of the angle of arrival  $\theta$ . The smallest resolvable phase difference  $\Delta\phi_{Res}$  corresponds to the smallest measurable time difference  $\Delta\tau_{Res}$ , which is  $\Delta\tau_{Res} = \frac{l}{c}$  for adjacent antennas. By relating the phase difference to the angle of arrival, we express the angular resolution as  $\theta_{Res} = \cos^{-1}\left(\frac{\lambda\Delta\phi_{Res}}{2\pi l}\right)$ . For an array with  $A$  antennas, the resolution improves with increasing  $A$ .

### B. Opportunities for Multi-modal Fusion Sensing

This subsection delves into the discussion of multi-modal sensor fusion opportunities, with a focus on the integration of various sensor modalities. The principal foundational sensor for multi-modal fusion sensing is the mmWave radar. Following this, the section explores the opportunities derived from the integration of sensors from diverse modalities, such as cameras, LiDAR, Near-Infrared (NIR) sensors, and Inertial Measurement Units (IMU).

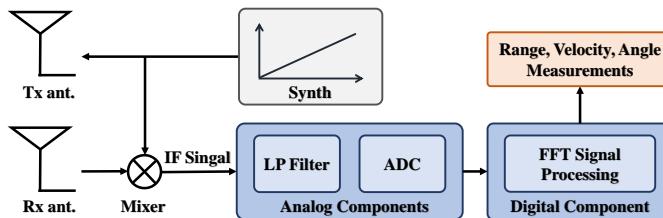


Fig. 3. Millimeter-wave radar functional structure

As depicted in Fig. 3, mmWave radar includes the transmit and receive front-end, analog components, and digital components. The transmitting antenna propagates frequency-modulated continuous wave (FMCW) chirps, which, upon hitting a target, get reflected and are received by the receive antenna. The received signal undergoes processing via the analog subsystems and is transformed to an intermediate frequency (IF) before being digitized. The digitized signal is further processed to compute critical parameters such as the range and velocity of the object.

Initially, researchers constructed their own mmWave radar systems, adhering to a standardized functional architecture. For example, multiple studies [33], [34], [35], [36], [37] discuss the development of custom-made mmWave radar systems, primarily spearheaded by academic researchers. The proliferation of commercially available mmWave radar systems has accelerated since 2019, with leading manufacturers such as Texas Instruments, Infineon, STMicroelectronics, and Adano Semiconductor driving the market [38], [39], [40], [41]. Over the past five years, these companies have introduced a range of Frequency-Modulated Continuous Wave (FMCW) mmWave radars, each designed for specific application domains.

In recent years, the perception accuracy of mmWave radar has significantly improved. Advanced imaging radars, such as those developed by Vayyar [39], have demonstrated enhanced resolution and object detection capabilities, narrowing the gap between mmWave radar and other high-accuracy sensing techniques. Moreover, compared to LiDAR, mmWave radar has the advantage of being more compact, consuming less power, and exhibiting higher sensitivity to micrometer-level vibration [2]. These attributes make mmWave radar an

attractive choice for applications that require high-precision sensing in space-constrained or power-limited environments, as well as for detecting minute movements or vibrations that are imperceptible to other sensing modalities.

As research progresses, investigators have identified significant opportunities for mmWave radar to be integrated with other sensor modalities. For instance, the fusion of mmWave radar with visual sensors enhances accurate perception under various weather conditions. The penetration capability allows mmWave radar to operate effectively in non-line-of-sight scenarios where LiDAR and cameras' efficiency decreases. Furthermore, mmWave radar is virtually unaffected by adverse weather conditions like rain and fog, which significantly degrade the performance of light-based sensing modalities. This robustness in diverse environments makes mmWave radar a valuable tool for a wide range of applications, including indoor mapping [12], through-wall sensing [42], and emergency response in challenging conditions like tunnel fires.

In a multi-modal fusion sensing framework utilizing mmWave radar, data extracted from the radar is integrated with data obtained from other sensor modalities, such as cameras and LiDAR, thereby enhancing the overall system efficacy and broadening its capabilities. The integration process of data from mmWave radar with other modalities typically involves several steps. Initially, a preprocessing phase is enacted on the data harvested from each modality, aiming to mitigate noise interference and rectify discrepancies. Subsequently, the multi-modal data are temporally and spatially synchronized to ensure congruence, signifying they correspond to the same timestamp and geographical coordinates. Ultimately, a data fusion operation is performed, yielding a holistic and precise depiction of the environment.

Table I elucidates a comparison of diverse sensor modalities, including mmWave radar along with the sensors it integrates with, such as *Vision*, *LiDAR*, *Near-Infrared (NIR)*, and the *Inertial Measurement Unit (IMU)*. These sensors exhibit complementary attributes, proving them ideal for sensor fusion applications. The mmWave radar is distinguished by its high level of light independence, weather resistance, and a favorable trade-off between cost, sensing range, and resolution. These properties make it a robust and dependable primary sensor in the sensor fusion architecture. While the mmWave radar affords vital parameters like distance, velocity, and angle, its functionality is enhanced through fusion with other sensor modalities. *Vision* and *LiDAR*, provide invaluable vision and spatial perception, thereby augmenting the overall environmental awareness. The *NIR* sensor, with its light independence sensing capabilities, enhances the granularity of the data, particularly under medium weather conditions. The *IMU*, despite not directly contributing to spatial sensing, provides indispensable motion data that assists in the accurate interpretation of spatial data from other sensors. The resolution of the IMU sensors is expressed appropriately: for gyroscopes, the resolution ranges from 0.1°/s to 0.001°/h, and for accelerometers, it spans from 100 mg to 10 µg. In conclusion, the mmWave radar acts as the central component of the sensor fusion architecture, whereas the other sensors augment its data, facilitating a more comprehensive and accurate understanding

TABLE I  
MODALITIES IN MULTI-MODAL FUSION SENSING

Modalities	Sensing Range	Cost	Resolution	Light Independence	Weather Resistance	Data Rate
mmWave Radar	0-200m	~\$200	High	High	High	0.1-0.2MB/s
Vision	0-100m	~\$100	Medium	Low	Low	20-40MB/s
LiDAR	0.7-200m	~\$5000	High	High	Medium	10-70MB/s
Near-Infrared (NIR)	10-30m	~\$1000	High	High	Medium	10-20MB/s
Inertial Measurement Unit (IMU)	-	~\$50	Acceleration-related	High	High	0.01-0.05MB/s

of the surrounding environment. The synergistic integration of these sensors results in a holistic, reliable, and high-performance system capable of functioning effectively across a broad spectrum of conditions and applications.

### C. Insight of Multi-modal Fusion Sensing via mmWave Radar

Regarding detection range and application contexts, mmWave radars are generally categorized into Long-Range Radar (LRR), Middle-Range Radar (MRR), and Short-Range Radar (SRR). Long-range radar, with detection ranges exceeding 200 meters, is primarily deployed in automotive scenarios for long-distance detection tasks, such as monitoring vehicles on the road. Middle-range radar, with detection ranges spanning 30 to 200 meters, is commonly utilized in automotive applications to scrutinize road conditions around the vehicle, facilitating features such as blind-spot detection, lane change assistance, and cross-traffic alerts. Short-range radar, capable of detecting within a 30-meter range, finds extensive application in human sensing scenarios, such as fall detection, vital sign monitoring, and pose estimation.

The mmWave radar exhibits remarkable capabilities attributed to their inherent properties, such as illumination independence, resilience to weather conditions, and adaptable detection ranges. Nevertheless, when integrated with other sensor modalities, they unveil a multitude of possibilities that considerably augment the overall system performance. In fall detection scenarios, the integration of mmWave radar and vision data leads to decreased false alarm rates [43], [44], consequently resulting in cost savings within healthcare environments. Within the domain of vital sign monitoring, the fusion of mmWave data with video data effectively addresses the challenges of visual interference frequently encountered in video-based monitoring [45], [46]. This multi-modal methodology delivers a more precise and reliable detection of live targets, enhancing the monitoring system's overall efficacy. Concerning pose estimation tasks, the fusion of mmWave radar and vision modalities resolves issues associated with point cloud sparsity and multi-path ghosting effects intrinsic to mmWave radar [14]. By integrating high-resolution spatial information from vision sensors with the robust and dependable data from mmWave radars, the system accomplishes heightened accuracy and reliability in pose detection.

Millimeter-wave radar multi-modal fusion sensing aims to improve the performance, reliability, and robustness of the sensing system by integrating mmWave radar with sensors of other modalities. The mmWave radar data are combined

with data obtained from complementary sensors, such as LiDAR, cameras, and ultrasonic transducers, yielding a more comprehensive and precise representation of the surrounding environment. Multi-modal fusion techniques are classified into data-level, feature-level, and decision-level fusion, each presenting unique advantages and trade-offs.

In these application contexts, incorporating multi-modal fusion techniques enhances the radar's capabilities by providing a more comprehensive and accurate depiction of the target objects and conditions. The implementation of mmWave radar fusion sensing varies based on the specific application requirements. For example, radar-vision fusions are often leveraged in automotive applications. The radar provides robust distance and velocity data, while the vision contributes high-resolution spatial information and object classification. The integration of these multi-modal sensors results in a more dependable and precise perception system, capable of operating under diverse environmental conditions.

## III. DATASETS

This section provides a detailed overview of the data obtained from the mmWave radar and other sensors' modalities. It mainly covers three specific data types: Doppler-based, point-cloud, and multi-modal fusion data. Before discussing these data types, we explain the generation methods of micro-Doppler, range-Doppler, and point cloud signatures, which are the intermediate steps to obtain the data from the received signals. Moreover, this section also introduces the publicly available datasets for each data type. These datasets serve as the resource for mmWave radar research and applications. Each type of dataset contains unique features. For instance, Doppler-based data (i.e., micro-Doppler, range-Doppler) reveals the velocity and motion of objects relative to the radar, while point-cloud data gives a three-dimensional representation of the scene. Multi-modal fusion data, on the other hand, merges mmWave radar data with data from other sensor modalities, resulting in more complete data on the sensing targets.

### A. Data Format and Characteristics

1) *Millimeter-wave Radar Signatures:* The raw radar signal (I/Q) contains significant noise and redundant information, so it commonly undergoes signal processing, which transforms the raw signal into various formats that highlight the features of sensing targets in specific domains (e.g., time, space, and velocity). This facilitates radar sensing applications and

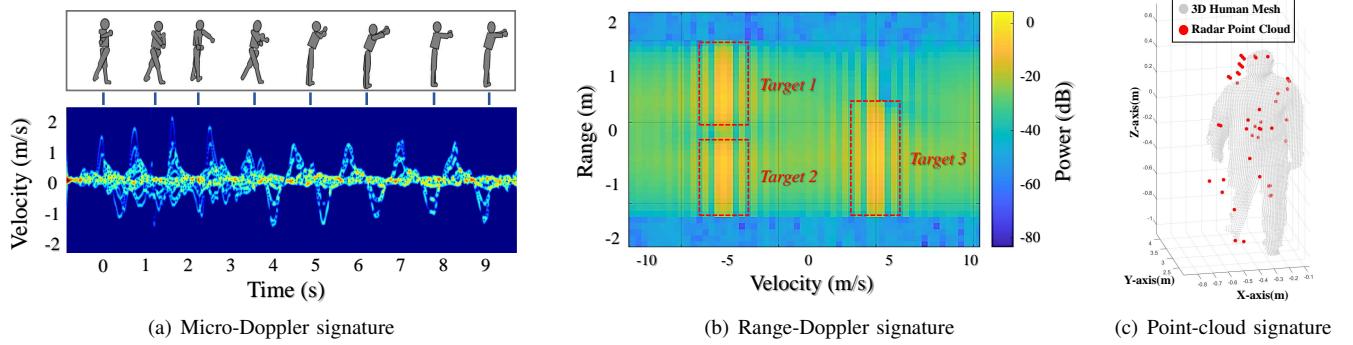


Fig. 4. Millimeter-wave radar signatures

reduces the burden of data communications. At a high level, the Doppler-based data format of mmWave radar is classified into micro-Doppler signatures and range-Doppler signatures.

**Micro-Doppler Signature:** The micro-Doppler signature captures fine-grained motions of the target, leveraging the radar signal's sensitivity towards minor movements (the characteristic discussed in Section II-A). Technically, it applies a Short-Time Fourier Transform (STFT) on the received signal to determine the velocity of the target over time. Importantly, when the target is a non-rigid body (e.g., a person), each component's micro-motion might exhibit variable velocities. This variability results in a unique temporal-velocity spectrum, revealing rich dynamic features of the motion. For example, Fig. 4(a) depicts a micro-Doppler signature of a walking human, highlighting the velocity changes in different body parts (e.g., torso, legs, and arms). Micro-Doppler signatures, due to their ability to express complex motion patterns, are utilized as the input of machine learning designs in human activity recognition, vital sign detection, and gait analysis.

**Range-Doppler Signature:** The Range-Doppler signature is another Doppler-based data format utilized in radar applications. The radar conducts range and Doppler Fast Fourier Transforms (FFTs) on the signal, creating a two-dimensional range-velocity heatmap. The radar distinguishes multiple targets in the signal based on their unique combination of range and velocity. An instance of the range-Doppler signature is demonstrated in Fig. 4(b), representing three moving individuals with distinct positions, ranges, and velocities. This format is frequently utilized in multi-target tracking and point cloud generation, as discussed in Section III-A1.

**Range-Angle Signature:** The Range-Angle signature serves as an integral intermediate stage in the creation of point cloud data and primarily presents the spatial information of targets. The radar obtains this signature by performing an Angle-FFT after the Range-FFT, measuring the phase difference caused by the angle of arrival. The horizontal antenna array provides azimuth information, while a vertical array aids in elevation estimation. In unison, an antenna matrix provides the 3D position of the target. The Range-Angle signature is typically utilized for target localization and tracking.

**Point-cloud Signature:** The point cloud signature is a data type that provides fine-grained 3D spatial information of the targets. It is generated by combining the range, velocity,

and angle estimates from the Range-Angle signature, which is obtained by performing an Angle-FFT after the Range-FFT, measuring the phase difference caused by the angle of arrival. An antenna matrix provides the target's 3D position, the utilization of a horizontal antenna array provides azimuth information, while a vertical array aids in elevation estimation. Commercial mmWave radars, leveraging Multiple-Input Multiple-Output (MIMO) capabilities, estimate angles with higher resolution. The angle estimations are combined with the range and velocity measurements to generate fine-grained 3D point clouds in real-world space (Section IV-A1 discusses the detailed steps of point cloud generations.). Fig. 4(c) presents a radar-generated 3D point cloud of a human target together with the corresponding 3D human mesh. Despite potential resolution limitations and multipath effects leading to sparse point clouds, 3D radar data offers an enhanced understanding of the targets. Leveraging recent advances in deep learning, radar point cloud enables emerging human sensing tasks (e.g., pose estimation, semantic segmentation, and mesh reconstruction) as well as autonomous driving applications [47]).

*2) Multi-modal Data Characteristics:* Each data format provided by mmWave radar – micro-Doppler, range-Doppler, and point cloud – serves as a distinct lens that allows the analysis of the environment from various perspectives, focusing on time, velocity, and spatial dimensions. These lenses offer powerful capabilities individually, and when integrated with data from other sensor modalities, they create even more compelling opportunities for a wide range of applications. Micro-Doppler data, with its ability to capture micro-vibrations of a target object, becomes invaluable in non-contact bio-signature detection when fused with vision data. This integration enables the measurement of vital signs such as respiration and heart rate without requiring physical contact [48], making it ideal for monitoring patients in healthcare settings or assessing well-being in remote or hazardous environments. Range-Doppler data excels in providing information about the velocity and distance of target objects, making it an opportunity for egomotion detecting applications [58]. In such scenarios, it is combined with GPS and IMU data to achieve accurate egomotion estimation, resulting in a comprehensive understanding of the target's movement relative to its surroundings. Point cloud data effectively represents the physical presence and

TABLE II  
AN OVERVIEW OF MICRO-DOPPLER DATASETS

Dataset	Year	Description	Frames
mBeats[48]	2020	A multi-modal datasets for heart rate sensing. Consists of micro-Doppler data of humans with eight kinds of poses collected by TI-IWR6843-ISK radar, and heart rate data collected by Polar H10.	~216k
DynamicHGR [49]	2021	DynamicHGR is a micro-Doppler hand gesture dataset collected by TI-IWR1642 radar. DynamicHGR consists of ~60 repetitions, each containing five human targets with six gestures.	~108k
ABDSVT [50]	2022	Consists of micro-Doppler gait gestures from ~100 volunteers collected by K-MC1 radar transceiver.	~125k
Single-Snapshot [51]	2022	Consists of ~1600 micro-Doppler snapshots of pedestrians walking randomly in direction collected by TI-AWR1843 radar.	~96k

\* Datasets without URL means that until this review publish, the datasets do not yet open permission for accession.

TABLE III  
AN OVERVIEW OF RANGE-DOPPLER DATASETS

Dataset	Year	Description	Frames
Soli [52]	2016	Soli is one of the first public datasets for gesture recognition. It consists range-Doppler data of 11 gestures of 10 users, each repeated 25 times. URL: <a href="https://github.com/simonsws/deep-soli">https://github.com/simonsws/deep-soli</a>	~74k
RFWash [53]	2020	RFWash is a range-Doppler dataset with ~1800 samples of human hand washing gestures collected by TI-IWR1443. Consists of nine kinds of gestures according to the hand-washing procedure recommended by WHO.	~144k
HawkEye [16]	2020	HawkEye is a multi-modal image and range-Doppler dataset. It consists of ~4000 synthesized scenes collected by a 2D antenna radar platform with 60 GHz radio.	~60k
MU-ID [54]	2020	MU-ID is a range-Doppler dataset collected by TI-IWR1642BOOST radar for human identification. It consists of 10 human targets, repeat walking 20 times.	~40k
mmVib [2]	2020	One of the most popular datasets of micrometer-level vibration measurement. Consists of ~40 traces for each setting collected by TI-IWR1642 radar and piezoelectric vibration sensor. URL: <a href="http://tns.thss.tsinghua.edu.cn/sun/researches/mmWaveSensing.html">http://tns.thss.tsinghua.edu.cn/sun/researches/mmWaveSensing.html</a>	~2100k
RadarNet [55]	2021	The largest range-Doppler dataset for gesture recognition so far, with ~558000 gesture frames and ~392000 negative frames. URL: <a href="https://github.com/nabilbenmerad/RadarNet">https://github.com/nabilbenmerad/RadarNet</a>	~4,478k
Pegoraro-RDA [56]	2021	Consists of multi-person tracking and identification data for ~10000 frames collected by INRAS RadarLog device.	~10k
JGLNet [57]	2022	Consists of posture data of 4 users in 3 environments doing five activities. Each group of activities consists of 40 samples repeated ten times. The data are collected by TI-AWR1843 radar and Intel Real-Sense D415.	~24k

\* Datasets without URL means that until this review publish, the datasets do not yet open permission for accession.

moving condition of objects in space, and due to its inherent sparse nature, it is particularly useful in privacy-sensitive situations as an alternative to traditional camera. This allows for discreet monitoring and understanding of the environment without infringing on personal privacy [59], making it a valuable tool for applications in security, surveillance, and public spaces. The following subsection delves into each data format, presenting their individual datasets. Moreover, the following subsection explores how the integration of these data formats with other sensor modalities significantly enhances their utility across a wide range of applications, from object detection and recognition to tracking and localization.

### B. Datasets of mmWave Radar and other Modalities

In this section, we present an exhaustive review of pertinent mmWave radar datasets from various research communities, including the datasets of micro-Doppler, range-Doppler, and point cloud. These datasets are classified based on their data formats, encompassing millimeter-wave radar signatures, and multi-modal data characteristics. Comprehensive information

regarding these datasets, including their publication year, intended application cases, a concise summary of data acquisition techniques, and the dataset size, is furnished in Tables II to V. Moreover, we expound on the distinctive characteristics and disparities among these datasets.

**Micro-Doppler datasets:** Table II presents a variety of micro-Doppler datasets collected for diverse applications. Zhao *et al.* [48] have assembled a multi-modal micro-Doppler dataset targeted at heart rate sensing applications. Jiang *et al.*, on the other hand, have constructed a micro-Doppler dataset consisting of six unique dynamic hand gestures [49]. Micro-Doppler gait datasets have been collected by Chen *et al.* [50] and Hor *et al.* [51], utilizing a K-MC1 radar transceiver and TI mmwave radar, respectively.

**Range-Doppler datasets:** Table III provides a comprehensive summary of various range-Doppler datasets collected by researchers. The Google Soli project [52] provided the first range-Doppler dataset for human gesture recognition. Hayashi *et al.* [55] and Khamis *et al.* [53] further supplemented the gesture datasets by increasing the gesture types and the data

TABLE IV  
AN OVERVIEW OF POINT CLOUD DATASETS

Dataset	Year	Description	Samples
mm-Pose [60]	2019	mm-Pose is a multi-modal dataset collected by TI-AWR1642 radar and Microsoft Kinect. It consists of ~32000 samples of training data, ~6000 samples of validation data, and ~1700 samples of test data with two subjects of 5 actions for pose estimation. URL: <a href="https://github.com/open-mmlab/mmPose">https://github.com/open-mmlab/mmPose</a>	~40k
RadHAR [61]	2019	RadHAR is a point cloud dataset of gestures from two human targets collected by TI-IWR1443 radar. It consists of 5 types of gestures. URL: <a href="https://github.com/nels/RadHAR">https://github.com/nels/RadHAR</a>	~12k
DLTracking [62]	2019	Consists of point cloud data for target tracking collected by TI-IWR1643BOOST radar and ground truth data collected by RGB camera.	~5k
mmGaitNet [63]	2020	mmGaitNet consists of 30 hours of 3D point cloud data from 95 volunteers collected by TI-IWR6843 and TI-IWR1443 radar. URL: <a href="https://github.com/mmGait/people-gait">https://github.com/mmGait/people-gait</a>	~1260k
RHGR [64]	2020	Consists of 50 groups of point cloud data collected by TI-AWR1642BOOST-ODS radar for gesture recognition. Each group consists of 10 human targets with six gestures.	~150k
3D-Skeleton [65]	2020	Consists of point cloud data of 14 human targets for 3D-Skeleton estimation collected by TI-IWR1443 radar and ground truth collected by Microsoft Kinect.	~50k
milimap [12]	2020	Consists of ~46000 frames from four types of objects in two buildings collected by TI-AWR1443, Velodyne VLP-16, and wheel odometry. It is one of the first published point cloud datasets for indoor mapping under harsh environments such as smoke and dark. URL: <a href="https://github.com/ChristopherLu/milliMap">https://github.com/ChristopherLu/milliMap</a>	~46k
mmFall [43]	2020	mmFall is a dataset for fall detection collected by TI AWR1843BOOST radar. URL: <a href="https://github.com/radar-lab/mmfall">https://github.com/radar-lab/mmfall</a>	~22k
mHomeGes [66]	2020	mHomeGes is a point cloud dataset of gestures from 11 males and 14 females collected by TI-IWR1443 radar. It consists of 10 types of gestures in 7 different environments. URL: <a href="https://github.com/GestureMan/mHomeGes-dataset">https://github.com/GestureMan/mHomeGes-dataset</a>	~22k
mmMesh [67]	2021	Consists of point cloud data from 20 human targets with eight actions, each target has ~3000 frames collected by TI-AWR1843BOOST radar. And ground truth 3D human pose collected by the VICON motion capture system. URL: <a href="https://github.com/HavocFiXer/mmMesh">https://github.com/HavocFiXer/mmMesh</a>	~480k
MARS [68]	2021	One of the first datasets for rehabilitation movement with labeled joints. Consists of ~2.28 million reference data points collected by Microsoft Kinect V2 and ~3.81 million data points collected by TI-IWR1443 radar. URL: <a href="https://github.com/SizheAn/MARS">https://github.com/SizheAn/MARS</a>	~40k
HPHD [69]	2021	Consists of point cloud data for target tracking collected by TI-IWR1443 radar and ground truth data collected by RGB camera.	~30k
Pantomime [70]	2021	Pantomime is a point cloud dataset of 21 types of gestures from 45 human targets in 5 environments collected by TI-IWR1443 radar. URL: <a href="http://dx.doi.org/10.5281/zenodo.4459969">http://dx.doi.org/10.5281/zenodo.4459969</a>	~7k
OMTT [71]	2021	Consists of 20 sequences of point cloud data, collected by TI-IWR1643 radar, each last for 20 seconds. Together with ground truth data collected by Logitech C920 camera and Velodyne LP16 LiDAR.	~4k
mmReID [59]	2022	One of the largest multi-modal datasets for gait recognition and human target identification, collected by TI-IWR6843-BOOST radar and Microsoft Kinect. Consists of 45 different identities, each with 17 radar records on average.	~5000k
mTransSee [72]	2022	mTransSee is a point cloud dataset for gesture recognition collected by TI-IWR1443 radar. It consists of gestures from 36 human targets, five gestures in 13 positions, and each repeated 16 times. URL: <a href="https://github.com/mmTransGes/mTransSee_Dataset">https://github.com/mmTransGes/mTransSee_Dataset</a>	~43k
Tesla-Rapture [73]	2022	Consists of 21 types of gesture with 13 anchor positions collected by TI-IWR1443 radar.	~41k
MiliPoint [74]	2023	MiliPoint is a point cloud dataset for human activity recognition collected by mmWave radar. It consists of actions from 20 human subjects, 10 actions in 4 scenarios, and each repeated 10 times. URL: <a href="https://github.com/yizzf/MiliPoint">https://github.com/yizzf/MiliPoint</a>	~80k

\* Datasets without URL means that until this survey publish, the datasets do not yet open permission for accession.

amount. Guan *et al.* [16] produced synthetic radar datasets for vehicle imaging, whereas Yang *et al.* [54] and Cao *et al.* [57] collected range-Doppler radar data for multi-person scenarios. Jiang *et al.* [2] also published a range-Doppler dataset for micrometer-level vibrations.

**Point Cloud datasets:** The point cloud format is a popular choice for radar datasets as it offers a rich source of spatial information about target objects within a 3D coordinate system [47]. Key details of point cloud datasets are compiled in Table IV. Meng *et al.* [63] published the mmGaitNet dataset comprising the gait point clouds of 95 human subjects across various multi-person scenarios. Cao *et al.* [59] and Zhao *et*

*al.* [85] assembled point cloud datasets for gait recognition, featuring 45 and 12 subjects respectively. Various datasets are available for pose recognition and reconstruction. Singh [61], Sengupta *et al.* [62], Jin *et al.* [43], and Cui *et al.* [69] collected point cloud for posture recognition, fall detection, and tracking. For the task of 3D skeleton estimation, datasets provided by Xue *et al.* [67], An *et al.* [68], Cui *et al.* [71], Wang *et al.* [65], and Sengupta *et al.* [60] come with ground truth skeletons, obtained via the VICON system, Microsoft Kinect, or the Logitech C920 camera. On the topic of gesture recognition, point cloud datasets were created by Xia *et al.* [64], Liu *et al.* [66], Palipana *et al.* [70], Liu *et al.* [72],

TABLE V  
AN OVERVIEW OF MULTI-MODAL FUSION DATASETS

Dataset	Year	Description	Frames
nuScenes [75]	2020	nuScenes is the first large-scale dataset to provide data from the entire sensor suite of an autonomous vehicle. It contains mmWave, vision, Lidar, GPS, and IMU data from 1000 autonomous driving scenarios. URL: <a href="https://www.nuscenes.org/nuscenes#download">https://www.nuscenes.org/nuscenes#download</a>	~40k
milliEgo [58]	2020	The mmbbody is a dataset for human egomotion estimation, which consists of mmWave, vision, and IMU data of human motion. URL: <a href="https://drive.google.com/drive/folders/1cJ4w3Dj21EMIox2ZSx7TSdzflxNhhTVL">https://drive.google.com/drive/folders/1cJ4w3Dj21EMIox2ZSx7TSdzflxNhhTVL</a>	~300k
STFWSF [76]	2020	The dataset contains mmWave, vision, Lidar, and NIR data covers different weather conditions such as fog, snow, and rain and was acquired by over 10,000 km of driving in northern Europe. URL: <a href="https://github.com/princeton-computational-imaging/SeeingThroughFog">https://github.com/princeton-computational-imaging/SeeingThroughFog</a>	~100k
RCF[77]	2021	An autonomous driving dataset of 12 driving sequences collected by a test fleet, which includes 14.8 hours of mmWave and vision data in various driving scenarios, including highways, cities, and urban roads.	~107k
VDAT [78]	2021	A dataset for vehicle detection and tracking in unstructured environments. It contains real-time mmWave, Lidar, GPS, and IMU fused data.	~16k
RSOD [79]	2021	A fused dataset of mmWave and vision data for small target detection on the water surface. It consists of 12,000 frames of floating bottle image data and mmWave data collected in real-world inland waters.	~12k
RADIATE [80]	2021	RADIATE is a high-resolution dataset in adverse weather. It focuses on multi-modal sensor data fusion including mmWave, vision, Lidar, GPS, and IMU data in adverse weather conditions. URL: <a href="http://pro.hw.ac.uk/radiate/">http://pro.hw.ac.uk/radiate/</a>	~200k
mRI [81]	2022	mRI is a multi-modal 3D human pose estimation dataset with mmWave, depth vision, and internal sensor data. It consists of over 160k synchronized frames from 20 subjects performing rehabilitation exercises and supports the benchmarks of HPE and action detection. URL: <a href="https://github.com/sizhean/mri">https://github.com/sizhean/mri</a>	~160k
Vision-RF [59]	2022	Vision-RF is a dataset for pedestrians gait re-identification. It consists of mmWave and depth vision data of 45 people walking in corridors and outdoor environments.	~45k
Geryon [82]	2022	Geryon is a dataset for object detection. It contains mmWave and vision data of people, bicycles, and cars in foggy/rainy/snowy weather and bad lighting conditions.	~26k
PLD [46]	2022	It contains mmWave and vision data of pedestrians and billboards in four different road situations.	~15k
GLE-Net [83]	2022	GLE-Net is a dataset for object detection. It contains mmWave and vision data of pedestrians walking for 15 and 13 scenes on foggy days and poor lighting conditions.	~4.4k
mmbbody [84]	2022	The mmbbody provides human body data with mmWave, depth vision, motion capturer, and GT mesh data. URL: <a href="https://chen3110.github.io/mmbbody/index.html">https://chen3110.github.io/mmbbody/index.html</a>	~200k
HPJL[14]	2023	HPJL consists of 10 different activity data of 32 individuals collected by mmWave and vision in three different environments, indoor and outdoor.	~960k

and Salami *et al.* [73]. Moreover, Chris Lu *et al.* [12] have provided the first point cloud dataset aimed at indoor mapping.

Table V provides an overview of multi-modal fusion datasets, including data from different sensing modalities to improve perception and understanding of the environment for various applications. The utilization of multi-modal fusion sensing is a central theme, enhancing the robustness and reliability of systems, particularly in challenging scenarios where a single sensor modality does not provide sufficient information. To underscore the breadth of modality integration, we underline the specific modalities utilized within each dataset, highlighting the diverse approaches in multi-modal sensory data. For example, the fusion of millimeter wave radar and vision data, which combines the radar's robust sensing capabilities in adverse weather conditions with high-resolution imagery from vision, is seen in a range of datasets [75], [58], [76], [77], [80], [81], [59], [82], [46], [83], [84], [14]. In addition, some datasets fuse mmWave radar and LiDAR data, combining the object detection and velocity measurement capabilities of radar with high-resolution 3D point clouds from LiDAR [75], [76], [78], [80]. Datasets fuse mmWave radar and Near-Infrared (NIR) data, enhancing performance in low-light and night-time conditions [76]. Moreover, the fusion of mmWave radar with Inertial Measurement Unit

(IMU) data is also researched, which is vital in improving motion tracking, localization, and navigation, especially in environments that GPS do not cover [75], [76], [78], [80]. Across these combinations, the unique advantages of each sensing modality are integrated through multi-modal fusion sensing, thereby creating more robust and reliable sensing solutions for a variety of applications.

### C. Lessons Learned: Summary and Insights

In the preceding discussions, we classify data into two primary categories: mmWave radar and multi-modal data. Each data type underscores distinct characteristics of the targets, with their strengths and weaknesses deliberated below.

The mmWave radar data is further segregated into two sub-categories: Doppler-based signatures and radar point clouds. Doppler-based signatures are seamlessly converted into 2D image format, rendering them compatible with cutting-edge neural network architectures in computer vision. In contrast to point cloud-based data, Doppler-based data preserves a higher degree of original information, facilitating more granular sensing tasks such as heartbeat detection. The radar point cloud furnishes 3D information, offering an enhanced spatial understanding of the environment and the ability to model complex objects, surfaces, and structures more accurately.

The multi-modal data format represents a comprehensive approach of integrating mmWave radar data with data of other modalities, integrating multiple types of data to create a more complete and detailed representation of the target. This format merges data from different sensors or sources, such as combining millimeter-wave radar data with optical sensor data or infrared sensor data. The advantage of this approach lies in its ability to leverage the strengths of each data type, while compensating for their individual weaknesses. For instance, radar data provide reliable distance and velocity information regardless of lighting conditions, whereas optical or infrared sensors offer high-resolution imaging. By fusing these data types, a more robust and accurate representation of the scene is achieved. However, the main challenge with multi-modal fusion lies in the need for sophisticated algorithms to effectively integrate and interpret the diverse data types. These algorithms must be capable of handling potential discrepancies in resolution, scale, and temporal synchronization among different data sources. Despite these challenges, the multi-modal fusion-based approach is gaining increasing attention in the field due to its potential for enhanced target detection and recognition performance.

#### IV. METHODOLOGIES

This section presents the methodologies of multi-modal fusion sensing via millimeter-wave (mmWave) radar and other modalities. In the field of mmWave radar multi-modal fusion sensing, both mmWave radar and multi-modal fusion play crucial roles. To intuitively demonstrate the methods and advantages of integrating mmWave radar with other modalities, we initially introduce methodologies based on mmWave radar and categorize them into signal processing methodology and learning methodology for Doppler and point cloud data. Subsequently, we focus on presenting three categories of methods of integrating mmWave radar with sensors of other modalities, including data-level fusion, feature-level fusion, and decision-level fusion. In the realm of data-level fusion, we meticulously detail the entire process of integrating mmWave radar data with data from other modalities, from data storage, preprocessing, synchronization, and association. Within the context of feature-level fusion, we discuss two categories of methodologies, model-independent and model-dependent methods. Model-independent methods are based on the original multi-modal data, leveraging their inherent physical features for fusion, while model-dependent fusion is achieved by learning the correlation between multiple modal data to grasp associative features for integration. In decision-level fusion, we provide a detailed example of how to make associative decisions based on data from multiple modalities, ultimately reaching the correct conclusion. Finally, this section summarizes the insights of the methodologies for mmWave radar multi-modal fusion sensing.

##### A. Data-level Fusion

Data-level fusion is a process of combining the raw data from different modalities into a unified data set that capture the multi-modal information more effectively. Data-level fusion

improves the quality and completeness of the data, as well as reduce noise and uncertainty. In this subsection, we first introduce the various data extraction methods that are used to obtain the relevant features from the raw data, such as fast Fourier transforms (FFTs), MULTiple SIgnal Classification (MUSIC), and constant false alarm rate (CFAR). We then discuss the data-level fusion methodologies that integrate the data from different modalities, such as synchronization and association. We also present examples and applications of data-level fusion in the context of mmWave radar sensing.

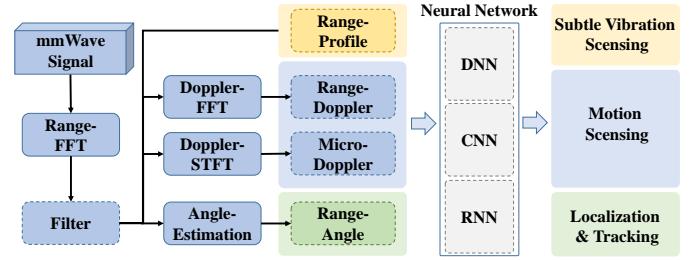


Fig. 5. Overview of Doppler-based method

**1) Data-level Processing Methodologies:** In data-level fusion, the Doppler-based approach of mmWave radar is particularly effective. In this scenario, the Doppler-based approach facilitates the fusion of raw radar data with other modalities, such as image or video data, that contemporary deep learning models efficiently process. This integration at the data level allows for a more comprehensive analysis and interpretation of the combined data, potentially leading to more accurate and robust model predictions. The core premise of the Doppler-based methodology is to convert radar data into task-relevant data forms that these neural networks are adept at processing. This transformation, as shown in Fig. 5, involves the extraction of various Doppler-based data. This subsection further introduces the data processing methodologies in data-level fusion.

**Range-FFT:** The initial phase of signal processing involves conducting a Range-FFT (Fast Fourier Transform), designed to differentiate targets based on their range. This process leverages the fact that a target situated at a specific range produces a distinct tone in the IF signal. To separate targets at varying ranges, the radar employs a Fast Fourier Transform to convert the IF signal into the frequency domain, thereby enabling an analysis of the frequency components that correspond to targets at different distances, as illustrated in Fig. 6.

The range-FFT computation mathematical equation is:

$$R[k] = \sum_{n=0}^{M-1} x[n] \cdot e^{-j \frac{2\pi}{M} nk}, k = 0, 1, 2, \dots, M. \quad (4)$$

Here the input  $x[n]$  denotes the discrete samples of the IF signal obtained from a chirp, and  $M$  represents the total number of samples. The output frequency components, termed as the range profile  $R[k]$ , include reflections from targets at the  $k^{\text{th}}$  discrete distance. Each peak value within  $R[k]$  signifies the presence of one or more targets at a specific range.

There are numerous studies focusing on the detection of subtle vibrations, for instance, heartbeat [48], [86], [87], along

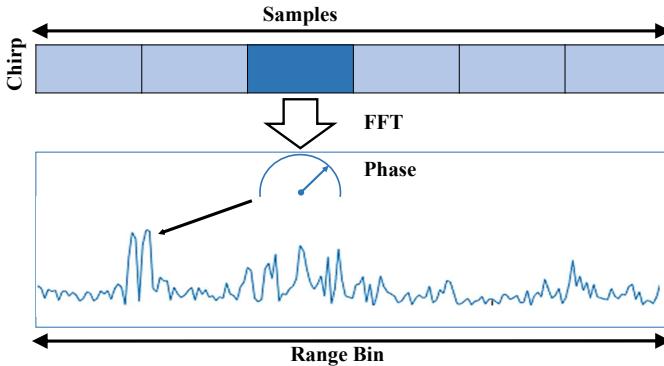


Fig. 6. Generation of range profile

with breath detection [88], and sound detection [89], [90], utilize the Range-FFT output to quantify minute displacements.

**Doppler-FFT:** The range profile undergoes further processing to differentiate targets at the same range based on their velocities, leading to the generation of a **range-Doppler** signature. As discussed in Section II-A, a moving target induces phase shifts across chirps. By observing a sequence of chirps, the radar captures the Doppler frequency resulting from phase shifts, which in turn indicates the target's velocity. To identify multiple targets with different velocities, a Doppler Fast Fourier Transform (Doppler-FFT) is performed on a radar frame composed of a sequence of  $N$  chirps, as shown in Fig. 7. The mathematical expression for this process is:

$$D[k, h] = \sum_{n=0}^{N-1} R[k, n] \cdot e^{-j \frac{2\pi}{N} nh}, h = 1, 2, 3, \dots, N. \quad (5)$$

Here the input  $R[k, n]$ , is the range profile obtained from a sequence of chirps, where  $n$  represents the index of the chirp and  $N$  is the number of chirps. Doppler-FFT is performed for each range index  $k$ , producing a 2D range-Doppler signature,  $D[k, h]$ . The index  $h$  signifies the Doppler frequency, which is translated into velocity. Each peak value on the range-Doppler signature indicates a target with a specific range and velocity.

A number of studies, such as RFWash [53] and Radar-net [55], employ Doppler-FFT for the purpose of gesture recognition and sensing. In the work presented in MU-ID [54], a temporal sequence of range-Doppler signatures is transformed into a singular image, where the velocity information is depicted through the use of varying colors. This innovative approach allows for a more intuitive and visually accessible representation of gesture-based data, facilitating the development and analysis of effective gesture recognition algorithms.

**Doppler-STFT:** The Doppler-based Short-Time Fourier Transform (Doppler-STFT) is implemented on a sequence of chirps to generate the **micro-Doppler** signature. Unlike Doppler-FFT which is executed on the entire sequence, Doppler-STFT performs the FFT within a window that covers a short time period. As this window slides along the input sequence, it reveals changes in velocity over time. Fig. 8 visualizes the process of generating the micro-Doppler signature. The Doppler-STFT equation is expressed as:

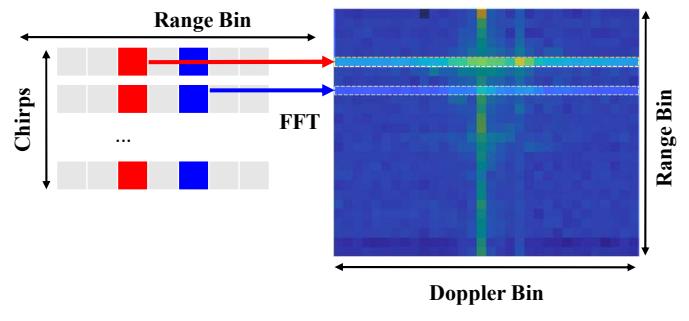


Fig. 7. Generation of range-Doppler signature

$$M(t, \omega) = \sum_{n=0}^{N-1} \sum_{\tau=1}^M \max_k R(k, \tau) w(t - \tau) \cdot e^{-j \frac{2\pi}{N} \tau \omega}. \quad (6)$$

Here  $M(t, \omega)$  denotes the STFT of the range profile  $R(k, \tau)$  at time  $t$  and Doppler  $\omega$ . The window function  $w(t - \tau)$  is applied within a short time interval  $\tau$ . The resultant STFT spectrogram provides a two-dimensional visualization, with the vertical axis representing the Doppler  $\omega$  and the horizontal axis indicating the time  $t$ .

The Doppler-STFT technique allows for the examination of localized changes in velocity over time by focusing on specific velocity components associated with distinct time segments. This localization is achieved through the utilization of appropriate window functions, such as the Hamming or Gaussian window, which effectively isolate the signal within each time segment. The STFT spectrogram reveals variations in Doppler frequency components, thereby facilitating a detailed analysis of the micro-Doppler characteristics exhibited by the target.

Incorporating the STFT equation into the explanation of the Doppler-STFT procedure underscores the rigorous mathematical and signal processing principles underlying the extraction of micro-Doppler features. This approach not only deepens our understanding of the technique, but also highlights the importance of accurate parameter selection, careful window design, and thoughtful interpretation of the resulting spectrogram in the comprehensive analysis of micro-Doppler signatures.

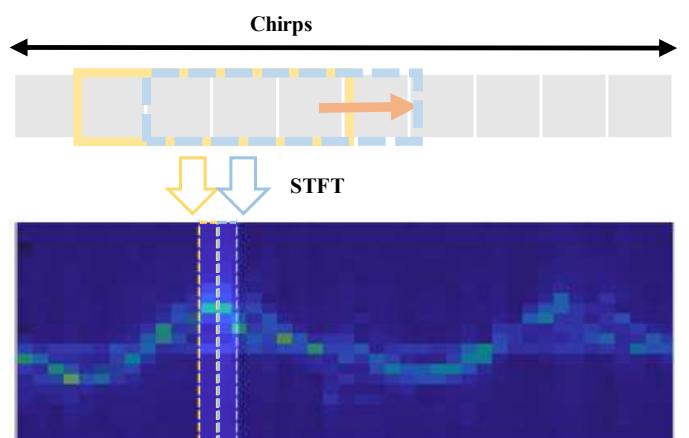


Fig. 8. Generate micro-Doppler signature

Micro-Doppler signatures have been utilized for motion sensing in a single-person context. Researchers in [91] employed micro-Doppler signatures for American Sign Language (ASL) recognition. In multi-person scenarios, individuals must be identified before generating micro-Doppler signatures. In the study by [92], clustering methods were utilized for separation, and micro-Doppler signatures were then employed to monitor the activities of patients.

- **Angle-FFT:** As outlined in Section II-A pertaining to angle measurement, the Angle of Arrival (AoA) of a signal induces phase differences across antennas. Analogous to Doppler FFT, the radar carries out an Angle Fast Fourier Transform (angle-FFT) on the signals received by the antenna array to discern its AoA. The equation for one-dimensional angle-FFT is given by:

$$A[k, h, i] = \sum_{r=0}^{R-1} D[k, h, r] \cdot e^{-j \frac{2\pi}{R} ri}, i = 1, 2, 3, \dots, R. \quad (7)$$

Here the input  $D[k, h, r]$  represents the range-Doppler profile procured from the  $r^{\text{th}}$  antenna in an antenna array. Here,  $R$  denotes the count of RX antennas. Angle-FFT is executed on each target of interest on the range-Doppler profile denoted by  $k, h$  tuples. The peak value on the output  $A[k, h, i]$  signifies a target at the  $i^{\text{th}}$  angle. For radar systems with a 2D antenna array, angle-FFT is conducted in both azimuth and elevation directions to secure the AoA in both dimensions.

- **MVDR:** The Minimum Variance Distortionless Response (MVDR), also referred to as Capon beamforming, is another algorithm utilized to estimate a signal's arrival angle, boasting a superior resolution compared to angle-FFT. The crux of MVDR beamforming lies in its utilization of a steering vector which scans various directions to identify the optimal angle that minimizes interference and noise sources while maintaining the desired signal at a specific angle. More details about MVDR are in [93].
- **MUSIC:** The MUSIC algorithm estimates the covariance matrix of the received signals and decomposes it into its eigenvalues  $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$  and corresponding eigenvectors  $\{v_1, v_2, \dots, v_M\}$ , where  $M$  represents the size of the antenna matrix. Given the number of signal sources, denoted by  $D$ , the eigenvectors are classified into signal and noise eigenvectors based on the magnitude of their respective eigenvalues. Sorting the eigenvalues in decreasing order, the eigenvectors  $\{v_1, v_2, \dots, v_D\}$  corresponding to the  $D$  largest eigenvalues constitute the signal space, while the remaining eigenvectors  $\{v_{D+1}, v_{D+2}, \dots, v_M\}$  span the noise space. It is mathematically provable that signal steering vectors and noise eigenvectors are orthogonal. Harnessing this orthogonality, MUSIC calculates the spatial spectrum using the equation:

$$P(\theta) = \frac{1}{a^H(\theta) E_n E_n^H a(\theta)}. \quad (8)$$

Here  $E_n = [v_{D+1}, v_{D+2}, \dots, v_M]$  constitutes the matrix of noise eigenvectors and  $a(\theta)$  symbolizes the steering vector for each candidate angle  $\theta$ . A steering vector  $a(\theta)$

orthogonal to the noise eigenvectors in  $E_n$  produces a peak at  $P(\theta)$ . It is important to note that the accurate knowledge of the number of signal sources is pivotal for the correct implementation of the MUSIC algorithm, as an error in this regard may misclassify eigenvectors into the signal and noise spaces.

The angle information, utilized in conjunction with range data, depicts the targets' positions. The study in [94] employs MVDR to extract angle information for localization and tracking, while RadioSES [89] utilizes the range-angle signature to determine locations. Additionally, the works in [95] and [96] generate range-azimuth-elevation signatures to capture poses.

**Filter:** Filters, in mmWave sensing, allow signals adhering to certain frequency thresholds to pass through. This feature is used to extricate desired information, such as small vibrations derived from frequency differences. In the context of mmWave-based vital monitoring [48], [86], [88], [87], band-pass filters are employed to extract subtle chest movements caused by heartbeats or respiration. On the other hand, [89] harnesses the vibration of the throat to enhance speech audio. A less benign application of similar techniques is displayed in [90], which extracts the vibration of phone speakers for eavesdropping. The work in [44] utilizes a high-pass filter to capture the high-frequency signal based on the premise that falling down occurs faster than daily activities. Similarly, due to the high-frequency nature of walking and hand-washing movements, high-pass filters are utilized in [53], [54].

**Clustering Algorithm:** Clustering algorithms such as K-means and DBSCAN are employed in multi-target scenarios [86], [48], [92], [89], [97], [88], [94] for target separation. The fundamental principle of the K-means algorithm is to segregate  $N$  objects into  $K$  clusters, with the aim of grouping objects exhibiting high similarity into the same cluster, and those with significant differences into different clusters. DBSCAN, a density-based clustering algorithm, is robust to noise and continues to perform efficient clustering amidst noise.

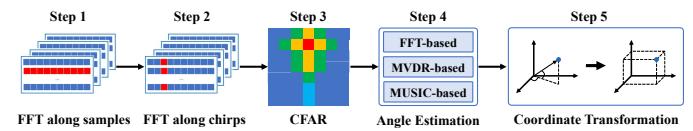


Fig. 9. Process chain for generating point clouds

**Range-Doppler profile:** Commercial mmWave radars are commonly equipped with multiple TX and RX antennas. The multiple-input-multiple-output (MIMO) capability allows radars to perform angle estimation and produce fine-grained 3D point clouds. The typical point cloud generation procedure is depicted in Fig. 9, which consists of the following steps.

As steps 1 and 2 in Fig. 9 show, range and Doppler FFT (discussed in the previous sections) are conducted to identify targets in the range-Doppler domain. A MIMO radar with  $N_{tx}$  transmitting and  $N_{rx}$  receiving antennas obtains  $N_{tx} \times N_{rx}$  streams of the received signal, a group of  $N_{tx} \times N_{rx}$  range-Doppler signatures are produced in parallel.

**CFAR algorithm:** The range-Doppler signature produced by radar suffers from environmental noise and hardware

imperfection. To accurately detect the target in noisy range-Doppler signature, radar commonly utilizes the Constant False Alarm Rate (CFAR) algorithm, which maximizes the target detection probability while maintaining a constant false alarm rate. The main objective of the CFAR is to adaptively adjust the threshold level for detecting a target in the presence of noise and clutter. The basic CFAR algorithm is demonstrated in Fig. 10(a). For each cell under test (CUT) in the range-Doppler signature, CFAR creates training cells in bands around the CUT. The training cells are utilized to estimate noise levels and dynamically adjust the detection threshold for CUT. CFAR also creates guard cells to prevent signal components from leaking into the training cell. CFAR produces a list of range and velocity tuples  $\langle k, h \rangle$  where the targets exist. By doing this, CFAR effectively reduces the amount of the data and ensures that only potential targets of interest are passed to the rest of the steps of the pipeline.

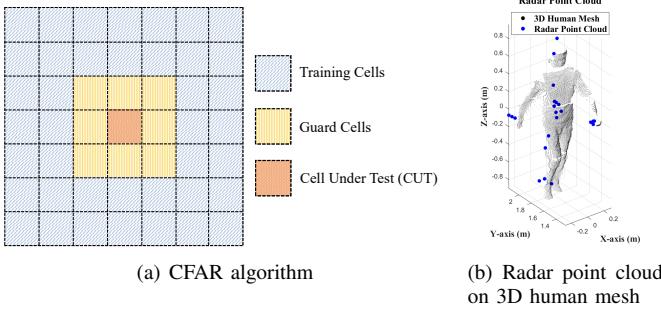


Fig. 10. CFAR algorithm and radar point cloud.

In conclusion, generating a 3D point cloud utilizing commercial mmWave radar data involves a series of processing steps, including range-Doppler profile generation, target detection with the CFAR algorithm, and angle estimation utilizing angle-FFT and MVDR. These procedures allow the radar to effectively identify and localize targets in real-world environments, even in the presence of noise and clutter. Fig. 10(b) shows an example of generated radar 3D point cloud of a human target together with the 3D human mesh. The radar point cloud is collected by TI-IWR6843-BOOST radar and the 3D human mesh is collected by Microsoft Kinect. 3D Radar data provides an opportunity for a better understanding of the targets and has numerous application in different areas, including human sensing, autonomous driving [47], etc.

2) *Data-level Fusion Methodologies:* In the realm of mmWave radar sensing methodologies, Data-level Fusion plays a crucial role. Our discussion in this section centers around its primary objectives and provides a general overview of the fusion techniques employed. Data-level fusion aims to integrate the raw data obtained from multimodal sensors at the earliest stage. This method is designed to enhance the system's performance by capitalizing on the multimodal data gathered from various sensors. The process of data-level fusion involves several steps, including data storage, preprocessing, synchronization, and association. Each of these stages is vital to successfully implementing data-level fusion, ultimately leading to a more enriched and detailed interpretation of the sensed environment in mmWave radar systems.

**Data Storage:** Three primary methodologies are utilized for storing data from mmWave radar: raw IQ signal storage, point cloud data storage, and processed image storage. Storing raw In-phase and Quadrature (IQ) signals is a comprehensive approach that guarantees no information is lost during the storage process. However, the significant volume of raw IQ data presents challenges, as it makes storage and subsequent processing resource-intensive and cumbersome. An alternative method is storing data as point clouds, as mentioned in the previous section, which are sets of data points in space. This is generally a more convenient approach due to the reduced data volume compared to raw IQ signals, but it struggles to handle erroneous point clouds that result from interference or noise in the radar signals. The third method incorporates storing data as processed images, typically created by merging point cloud data with vision data. Although this format is the most straightforward for observation and interpretation, it leads to a considerable loss of original data. The process of converting raw data into a visual image format inherently discards certain information, such as spatial information related to depth. This makes it less suitable for detailed analytical tasks that require access to the original signal data. Therefore, each storage method has its own advantages and trade-offs, and the choice depends on the specific requirements of the given application.

**Preprocessing:** The preprocessing process in mmWave radar multimodal fusion sensing includes data denoising and data clustering. Data denoising involves noise reduction techniques, such as the Kalman Filter, to reduce noise and remove outliers in mmWave radar IQ signal and point cloud data. The purpose of this process is to obtain more accurate data. On the other hand, data clustering utilizes clustering methods (e.g., k-means) to cluster the point cloud data. This process not only categorizes the data but also eliminates outlier noise points further, improving the overall quality and usability of the data. Thus, through denoising and clustering, the preprocessing stage enhances the reliability and precision of the mmWave radar data, facilitating more effective data fusion.

**Synchronization:** Data synchronization is a crucial step in the data-level fusion of the mmWave radar fusion sensing. This process involves aligning and integrating processed radar data (IQ signal or point cloud data) with data from other modalities, such as vision data, LiDAR, and gated NIR sensors. The alignment of data, both temporally and spatially, results in a comprehensive original dataset containing synchronized data from the mmWave radar and other sensing modalities.

When synchronizing with vision data, the primary challenge lies in aligning the high-resolution image data with the radar data, given their vastly different data structures and formats. This process often involves mapping the radar's range-azimuth data onto the camera's image plane, allowing for seamless integration of the two data types [59]. The resultant fused dataset provides a richer perspective of the environment, combining the high-resolution detail of the vision data with the robust range and velocity information of the radar data.

In the case of synchronizing with LiDAR, the synchronization process is similar to that of vision data, but with an added dimension of depth [78]. The point cloud data generated by LiDAR provides a detailed 3D representation

of the environment, which is combined with radar data to enhance object detection and classification capabilities. This fused dataset leverages the depth information from the LiDAR and the superior range and velocity data from the radar to provide a more complete understanding of the environment.

Synchronizing mmWave radar data with Near-Infrared (NIR) sensors' time-of-flight data presents an intricate set of challenges, primarily due to the distinct operating principles of the two sensor types. Millimeter-wave radar relies on radio frequency waves, transmitting them and interpreting the reflected signals. Conversely, NIR sensors operate based on the time-of-flight principle. It emits a light signal and measures the time it takes for the signal to return after reflecting off an object. The high-resolution depth data from NIR sensors is structured differently from the range and velocity data from the radar. The operation of these sensors occurs on different scales and they produce data in distinct formats. The alignment process, therefore, necessitates the application of advanced data processing techniques. These techniques typically involve processes such as concatenation, transformation, and fusion, [76]. This process requires careful manipulation and interpretation of the data to ensure meaningful and accurate synchronization. The synchronization of these two data types yields a powerful dataset. The high-resolution depth data from the NIR sensor complement the robust range and velocity data from the radar, providing a more comprehensive representation of the environment. This multimodal fusion enhances the system's ability to detect and classify objects, improving the overall performance and reliability of the sensing system.

In conclusion, data synchronization in the context of multimodal fusion is a complex process that involves aligning and integrating data from different sensing modalities with mmWave radar data. The resultant fused datasets leverage the strengths of each modality, providing a more complete and accurate depiction of the sensed environment, and thereby enhancing the overall performance of the sensing system.

**Association:** There are two primary ways of data association in multimodal fusion. Firstly, it is utilized to correct errors in mmWave radar data utilizing information from sensors of other modalities. Secondly, data association is employed to supplement information that is missing from the data from sensors of other modalities utilizing the mmWave radar data. The mmWave radar has specific advantages over other sensors, such as functioning effectively in poor weather conditions and its capacity to provide velocity measurements. By associating the radar data with the data from other sensors, these unique attributes are incorporated into the overall fused data, enhancing the comprehensiveness of the multimodal fusion.

The data association between mmWave radar and cameras primarily focuses on extracting spatial data from cameras with the velocity and distance data from the radar [46]. Cameras provide high-resolution images that classify objects and determine location with respect to each other while struggling under poor lighting or adverse weather conditions. Millimeter-wave Radar, on the other hand, provides accurate distance and speed information and works reliably under various environmental conditions [59]. After all, radar data supplement camera images with depth and speed information, while vision

data refine the radar's object recognition by providing color and shape details. The fusion of these two modalities provides a comprehensive understanding of the environment.

LiDAR and mmWave radar are both instrumental in producing distance and velocity measurements, yet they leverage distinct technological methodologies. LiDAR operates through the emission of light pulses to ascertain distances, creating high-resolution 3D point clouds. This makes it highly effective for environmental mapping and static, small object detection. However, LiDAR's performance is compromised under adverse weather conditions such as fog or rain. Contrarily, mmWave radar, demonstrates resilience to such challenging conditions. Millimeter-wave radar data bolster LiDAR-generated point clouds by ensuring consistent distance and speed metrics, irrespective of weather constraints. Concurrently, LiDAR refines radar's object localization efficacy by providing precise spatial and shape-specific details [98]. Thus, the fusion of radar and LiDAR data significantly enhances the sensing system's reliability and robustness.

NIR sensors generate images utilizing infrared light, which facilitates the detection of heat signatures and enables vision through certain materials. This proves particularly beneficial in liveness detection or penetrating obstructions like smoke. Conversely, mmWave radar offers velocity and distance measurements, supplementing the data from NIR sensors. Millimeter-wave radar data augment NIR images by providing depth and velocity information, unimpacted by lighting conditions. Simultaneously, NIR sensors amplify the radar's object detection capabilities by producing detailed imagery in low-light situations [76]. The fusion of these two data types enables the system to capture a more comprehensive understanding of the environment, especially under challenging conditions.

### B. Feature-level Fusion

Feature-level fusion is a process of combining the feature sets extracted from different modalities into a single feature set that represents the multi-modal data effectively. Feature-level fusion improves the recognition performance and robustness of the system by exploiting the complementary and correlated information among different modalities. In this subsection, we first introduce the various feature extraction methods that are utilized to obtain the relevant features from the raw data, such as convolutional neural networks (CNNs), voxelization, and transformers. We then discuss the feature-level fusion methodologies that integrate the feature sets from different modalities, such as model-independent and model-dependent fusion. We also present examples and applications of feature-level fusion in mmWave radar sensing.

1) *Feature-level Processing Methodologies:* Machine learning methods work like a tool to tackle classification or regression tasks using Doppler-based features as we describe above. In numerous studies, Doppler-based features are treated as conventional images or sequences of images, and Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are employed for feature extraction. For instance, Li et al. [44] apply CNN and RNN to recognize patterns in range-angle signatures to accomplish fall detection

tasks; RFWash [53] filtrate out high-frequency hand movement with a high-pass filter and generate range-Doppler heatmap as input for CNN and RNN based model. MU-ID [54] filtrate out the gait pattern and represent with range-Doppler signature, and CNN and RNN are utilized to identify different people by classifying range-Doppler signatures. Similarly, RadarNet [55] feeds range-Doppler signatures into CNN and RNN models to recognize gestures, and tinyRadar [1] utilizes CNN to classify micro-Doppler signatures for activity recognition. Zhao et al. [99] generate multiple range-angle signatures to achieve accurate drone position estimation. Pegoraro et al. [97] apply clustering on range-Doppler signatures to distinguish people and a Kalman filter is utilized for trajectory tracking. Then, person-associated micro-Doppler signatures are fed to CNN for identification. In m3track [96], in order to achieve better pose-tracking performance, range-Doppler signatures are utilized as temporal features and range-angle signatures are utilized as spatial features. Conv-LSTM models are utilized as feature extractors to analyze spatial signatures and 3D convolutional models are applied to temporal features.

In the field of speech separation, RadioSES [89] utilizes a dual-path RNN to combine mmWave and audio data, enabling the separation of speech signals from different individuals. For heart rate extraction, mBeat [48] and RF-SCG [87] employ CNNs as robust solutions. HeartPrint [86] and M-Auth [88] employ Radial Basis Function Support Vector Machines (RBF-SVMs) for user authentication based on extracted heartbeat and breath patterns. Additionally, in environment-independent American Sign Language (ASL) gesture recognition, mmASL [91] exploits the distinctive characteristics of ASL and introduces auxiliary tasks in the model design. These learning-based methodologies offer promising approaches to address various tasks and challenges in different domains.

**Voxel-based Networks:** The point cloud data collected by millimeter-wave radar is a representation of three-dimensional spatial information through a large number of discrete points. However, the distribution of point cloud data between different frames is often uneven. To facilitate the processing and analysis of point clouds, voxelization has been employed for millimeter wave point cloud data. By converting the point cloud data into regular voxel grids, the issue of data non-uniformity is addressed. Several existing deep neural networks, such as those utilizing 3D convolution operations, have proven effective in processing such voxelized point cloud data. Specifically, voxelization involves the core concept of mapping point cloud data onto a three-dimensional voxel grid. Voxels are small cubic regions that divide the space in a three-dimensional grid representation. Each voxel represents a discrete volume element within the space. By assigning the points from the point cloud to their corresponding voxels, the point cloud voxel network transform the point cloud data into a voxel representation with a regular structure. This transformation simplifies the processing and manipulation of point cloud data and provides additional geometric information.

Utilizing voxelization to process point cloud data indeed brings several benefits: firstly, voxelized point cloud data is stored in memory in an orderly manner, reducing random memory access and increasing data processing efficiency;

secondly, voxelization, with its ordered storage and downsampling, allows for the processing of point cloud data on a large scale; thirdly, voxelized data efficiently utilize spatial convolution, facilitating the extraction of multi-scale and multi-level local feature information. In the classic millimeter wave-based point cloud work, voxelization is often employed to process specific tasks such as gesture recognition and human activity recognition [85], [61].

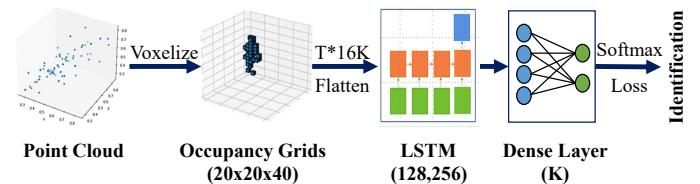


Fig. 11. Voxel-based networks architecture

Fig. 11 provides an overview of mmWave radar re-identification [85] methodologies. The process begins with the original point cloud data, which is voxelized to create a  $20 \times 20 \times 40$  occupancy grid. Voxelization converts the un-ordered point cloud into an ordered representation, where each voxel represents a specific region in the 3D space. The resulting occupancy grid is then compressed to reduce the dimensionality and remove redundant information. This compressed representation is fed into an LSTM (Long Short-Term Memory) network, and the LSTM network extracts meaningful features from the voxelized point cloud sequences. Finally, the output from the LSTM network is passed through a Dense Layer, which is a fully connected layer, to obtain the final predicted object identity. The Dense Layer combines the extracted features and performs classification based on the learned representations. By voxelizing the point cloud data and using an LSTM network for feature extraction, mID enables direct processing of point cloud sequences and performs better in identity recognition. This approach leverages the benefits of both voxelization and recurrent neural networks to effectively handle the un-ordered and sequential nature of point cloud data.

In summary, voxelization provides a practical solution for processing mmWave point cloud data. While it offers several advantages such as ordered storage, scalability, and efficient feature extraction, there are also trade-offs to consider, such as information loss, memory usage, and computational efficiency. The choice of voxelization strategy are made based on the specific requirements of the application, considering factors such as accuracy, computational resources, and storage costs.

**Transfer Learning for in Diverse Scenarios:** Transfer learning is a well-established research field that enables networks to improve their performance on smaller datasets by utilizing pre-trained models that have achieved high performance on large datasets [100]. This transfer of knowledge from a source domain to a target domain effectively reduces the cost of data collection and training. In the context of millimeter-wave sensing, the collected millimeter-wave point cloud data are often sensitive to the environment. For example, extensive experiments conducted by mTransSee [72] have shown that gesture recognition based on millimeter-wave radar requires a consistent working environment. Each time the system

enters a new environment, the data acquisition and training process needs to be repeated. These uncontrollable costs pose challenges for practical deployment. Hence, transfer learning is widely employed in these environmentally sensitive tasks.

A common approach in transfer learning is to freeze the convolutional layers of a pre-trained model and only train the fully connected layers and the remaining customized convolutional layers in the target domain. These pre-trained convolutional layers retain valuable underlying information that is beneficial for the task. In the case of mTransSee [72], a shallow neural network named mSeeNet is initially designed for the source domain to efficiently accomplish the gesture recognition task. Notably, unlike state-of-the-art works [66], mSeeNet adopts a unified feature extraction mode, enabling transferability to various environments in the target domain. For convenience, the last layer, FC3 (shown in Fig. 12), computes as follows:

$$Y_3 = \sigma(W_3 Y_2 + B_3). \quad (9)$$

Here  $Y_k$  represents the output of the  $k$ -th layer,  $W_k$  and  $B_k$  denote the trainable parameters of the  $k$ -th layer, and  $\sigma$  refers to the "SoftMax" non-linear activation function.

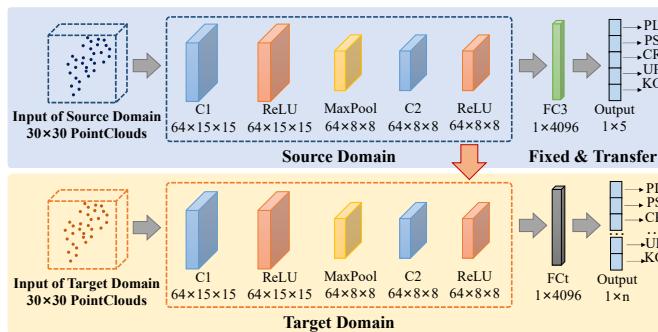


Fig. 12. Transfer architecture

To facilitate the transfer process, the mSeeNet network trained on the source domain is utilized, and its final output layer, FC3, is replaced by the fully connected layer FCt, which has a variable output size  $n$ . The FCt layer takes the output  $Y_2$  of layer C2 as its input. Since mSeeNet employs a unified representation of gesture features, the feature vector provided by C2 contains mid-level gesture features that are utilized to derive their high-level configurations. The computation of the FCt layer is,  $Y_t = \sigma(W_t Y_2 + B_t)$ , where  $W_t$  and  $B_t$  represent the trainable parameters of the  $t$ -th layer. As depicted in Fig. 12, the parameters of layers C1 and C2 are initially trained on the source domain, then transferred to the target domain and kept fixed. Only the adaptation layer FCt is trained on the target domain, where  $W_t$  and  $B_t$  represent the trainable parameters of the  $t$ -th layer. As depicted in Fig. 12, the parameters of layers C1 and C2 are trained on the source domain, then transferred to the target domain and kept fixed. Only the adaptation layer FCt is trained on the target domain.

Overall, transfer learning in the context of mTransSee employs the mSeeNet network pre-trained on the source domain and adapts it to the target domain. By leveraging the shared features and knowledge learned in the source domain, the model effectively improves its performance on the target

domain, enabling gesture recognition in diverse environments with reduced data collection and training costs.

In the following sections, we delve into the multi-modal fusion methodologies utilized in mmWave radar sensing. This involves the strategic integration of data from multiple sensors to enhance the overall system performance and precision. We specifically explore three primary fusion techniques: Data-level fusion, Feature-level fusion, and Decision-level fusion. Each method holds unique characteristics and offers distinct advantages in processing and interpreting sensor data, thereby contributing to the robustness and accuracy of mmWave radar sensing systems. Through this discussion, we aim to provide a comprehensive understanding of how these fusion-based methodologies are leveraged in mmWave radar sensing.

2) *Feature-level Fusion Methodologies:* In the domain of mmWave radar multimodal fusion sensing, feature-level fusion plays an integral role. The procedure initiates with the extraction of features from both structured data, encompassing point cloud and vision data, and unstructured data. Following this, the extracted features from the various data modalities undergo fusion, a process that is bifurcated into two distinct methodologies: model-independent and model-dependent methods. Model-independent fusion methods, also known as feature-based fusion methods, combine features directly without considering specific model, making utilization of statistical techniques or machine learning algorithms. Model-dependent fusion methods, or model-based fusion methods, incorporate the understanding of the physical world or a specific model of the environment into the fusion process, utilizing various types of learning models. The chapter aims to provide a comprehensive overview of the current methodologies in mmWave radar sensing, including the fusion of data from different sensors to enhance the performance of sensing systems.

**Model-independent Fusion:** This subsection covers the concept of Model-independent Fusion, a method that operates by directly combining features without the need for a specific model. This approach primarily makes utilization of statistical techniques or machine learning algorithms. In the context of multimodal fusion, Model-independent Fusion plays a critical role in integrating data from various sensors or modalities. The fusion process enhances the overall system's performance, enabling a more comprehensive and precise understanding of the environment. The subsequent sections will delve into deeper details of specific techniques utilized in model-independent fusion, such as Shallow Neural Networks (SNN), and their applications in the realm of multimodal fusion sensing.

**Shallow Neural Networks:** In scenarios where mmWave radar systems have limited computing resources, shallow neural networks (SNNs) offer a viable solution for human perception tasks. SNNs are learning models that do not require the extensive computational resources of complex neural networks. However, they outperform simple neuronal structures, making them well-suited as a compromise solution when computational resources are limited.

SNNs are commonly employed for simple tasks in point cloud-based applications, such as gesture classification [66], [37]. These tasks often face limitations due to constrained computing resources or emphasize the real-time nature of the

system. For example, the system proposed in mHomeGes [66] utilizes Hidden Markov Models (HMMs) to transform the data stream and its resultant sequence into a "status" probability vector, which is then utilized to classify human gestures. In another work by Patra et al. [37], Self-Organizing Maps (SOMs) and Learning Vector Quantization (LVQ) are employed to generate a gesture classification map, where multiple nodes in the map correspond to different gesture types.

These examples demonstrate the usage of SNNs in point cloud-based tasks, where the focus is on leveraging SNNs to achieve satisfactory performance while accommodating limited computational resources or real-time requirements.

**Model-dependent Fusion:** This section discusses the concept of Model-dependent Fusion, a method that incorporates knowledge of the physical world or a specific environmental model into the fusion process. Leveraging various types of learning models, this fusion method allows for a more nuanced understanding of the environment by considering the relationships and interactions between different features. In the context of multimodal fusion, model-dependent fusion is pivotal in integrating data from various sensors or modalities, providing a higher level of environmental comprehension and enhancing the overall performance of the sensing system. The subsequent sections provide a deeper exploration of specific techniques utilized in model-dependent fusion, including Graph Neural Network (GNN), Transformer, and PointNet, and their applications in multimodal fusion sensing.

**Graph Convolutional Networks:** Graph Convolutional Networks (GCNs) are a type of convolutional neural network designed to operate directly on graphs, leveraging their structural information [101], [102]. While other deep neural networks focus on sequential and grid-like structures, GCNs generalize convolutional operations to graph representations, which are more complex structures compared to Euclidean domains. In the context of GCNs, a graph refers to a collection of vertices (also called nodes) connected by edges.

In a typical graph neural network, the features of the vertices are refined by aggregating information along the edges. In each iteration ( $t + 1$ ), the vertex features are updated using the following formulation:

$$v_i^{t+1} = g^t(\rho(e_{ij}^t | (i, j) \in E), v_i^t), \quad (10)$$

$$e_{ij}^t = f^t(v_i^t, v_j^t). \quad (11)$$

Here  $e^t$  and  $v^t$  are the edge and vertex features from the  $t^{th}$  iteration. A function  $f^t(\cdot)$  computes the edge feature between two vertices.  $\rho(\cdot)$  is a set function that aggregates the edge features for each vertex.  $g^t(\cdot)$  takes the aggregated edge features to update the vertex features. The process is repeated in the next iteration to further refine the vertex features or the final vertex features are outputted by the graph neural network.

In the context of mmWave radar, the point cloud data obtained from FMCW signals are able to be represented as a graph. However, mmWave point clouds are often sparser compared to general point cloud data. Several designs are proposed to enhance the performance of GCNs in this domain.

For instance, MMPoint-GNN [103] introduces a dynamic edge selection function  $r^t(\cdot)$  that determines the existence

of an edge based on the features of adjacent points in the point cloud. To make the edge selection function differentiable, a modified hard sigmoid activation function called MHS is utilized [103]. Another approach, STPOINTGCN [104], leverages the point state property to initialize vertex features in the first layer, allowing the GCN to adapt to the characteristics of point clouds. Furthermore, STPOINTGCN considers the temporal dependency between consecutive frame point clouds and designs a graph neural network to extract and aggregate features in the time dimension for mmWave radar point clouds. These advancements in GCNs, along with their ability to process unstructured and non-grid data, have demonstrated promising results in tasks such as processing sparse mmWave data and human activity sensing. By leveraging the inherent graph structure of the data, GCNs provide a framework for feature extraction and information aggregation, enabling enhanced analysis and understanding of complex systems.

**Transformer:** Transformer is a highly promising neural network architecture that leverages the attention mechanism. Initially proposed for natural language processing [105], its exceptional performance has been extended to various domains, including image analysis [106], video processing [107], and 3D point cloud applications [108], [109].

The original Transformer model is designed for processing sentences and capturing the relationships between words through multi-head attention. Its input consists of two components: word embeddings and position encodings. The attention mechanism operates on pairs of tokens, disregarding the sequential position of words in a sentence.

When adapted to 3D point clouds, the inherent characteristics of the data naturally align with the input requirements of the original Transformer. Point clouds, being discrete and having each point's position encoded in its coordinates, lend themselves well to the Transformer architecture. The straightforward approach is to treat the entire point cloud as a sentence and each point as a word. There exist reviews that have proposed further improvements for the characteristics of point clouds, such as the Point Cloud Transformer (PCT) [108] and Point Transformer (PT) [109].

Both PCT and PT introduce modifications to the attention mechanism to improve local feature extraction. In PCT, self-attention is replaced by offset attention, and a sampling and grouping (SG) layer is employed to aggregate local features. In PT, self-attention is performed on the local neighbors of each data point, and scalar dot-product attention is replaced by vector attention. Additionally, a trainable positional encoding based on point coordinates is utilized as an alternative to the original position encoding. Fig. 13 illustrates variant Transformers adapted for mmWave data in different formats, with modifications made to the attention implementation.

Transformer architectures are adapted for mmWave data processing in different ways. For instance, in radar gait recognition, the attention-based dual-stream vision Transformer captures gait characteristics embedded in the radar spectrum and utilizes a vision Transformer to extract features from the spectrum. In another study, the ImmFusion model [110] employs a Transformer-based fusion approach for robust mmWave-RGB fusion in 3D human body reconstruction, leveraging attention

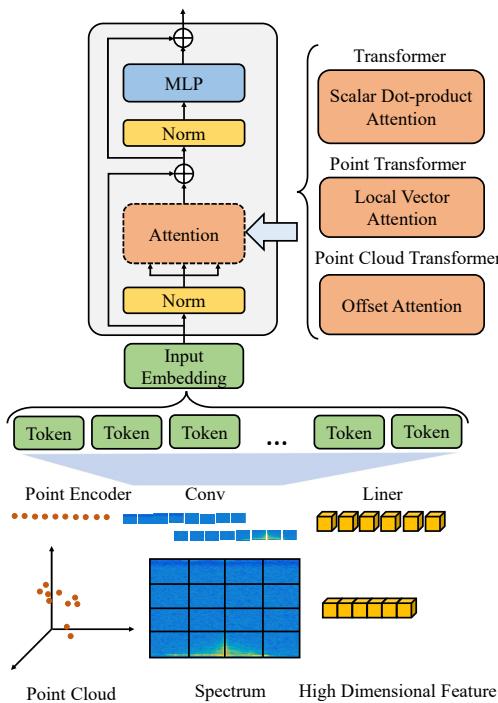


Fig. 13. Variant transformers for mmWave data in different formats, modifications are made on attention implementation

mechanisms to dynamically focus on informative features and mitigate the adverse effects of sparsity in mmWave data or noise in RGB data under extreme environmental conditions.

Compared to previous methods such as CNN, GCN, and PointNet, Transformer excels in modeling the dynamic spatial relationships between coordinate points in mmWave point clouds. It effectively captures long-term dependencies between point cloud frames, allowing for better extraction of contextual information. However, it comes with certain drawbacks. Transformer poses greater training difficulty and requires more time. Additionally, it may be less sensitive to excessively sparse mmWave point cloud inputs. Therefore, the application of Transformer in mmWave perception tasks should be carefully considered, taking into account the specific requirements and characteristics of the target scenario.

**PointNet & PointNet++:** PointNet [111] is a pioneering deep learning method specifically designed for the feature-level fusion of point cloud data and data of other modalities. It avoids to project or quantize irregular point clouds onto regular grids in 2D or 3D by directly processing the raw point clouds. This approach offers high computational efficiency. PointNet employs permutation-invariant operators, such as pointwise Multi-Layer Perceptrons (MLPs) and max pooling, to handle the unordered nature of the input points. This ensures that the results are invariant to the permutation of the points.

PointNet++ [112] builds upon PointNet and introduces three key improvements. Firstly, it incorporates sampling and grouping operations to enhance the extraction of local structural features. Secondly, it employs a hierarchical network that performs step-by-step downsampling to reduce information loss in the features. Finally, it utilizes skip connections to connect global and local features.

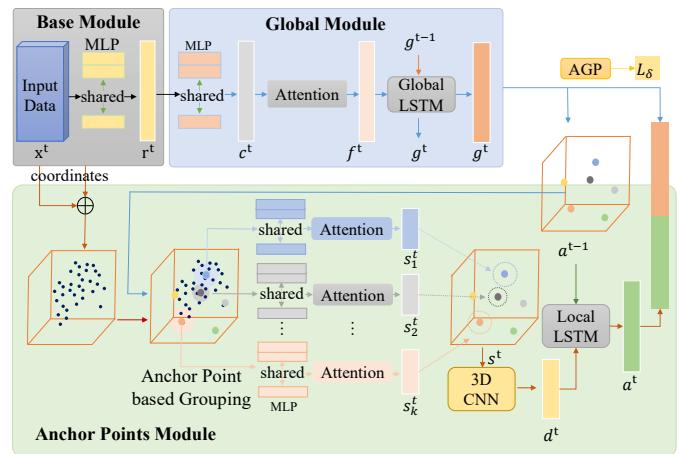


Fig. 14. PointNet++ architecture

The state-of-the-art approaches in the feature-level fusion utilize PointNet and its variants, along with additional techniques, to extract spatial and spatial-temporal features from sparse mmWave point clouds. Pantomime [70] utilizes two abstraction layers, consisting of sampling, grouping, and PointNet modules, to extract spatial features from sparse mmWave point clouds. The mmMesh [67] and m4esh [113] employ a combination of PointNet, attention mechanisms, and Long Short-Term Memory (LSTM) networks to extract spatial-temporal features from sequences of mmWave point clouds, as shown in Fig. 14. m4esh [113] further utilizes a coarse-grained human skeleton as an anchor to better extract features relevant to human reconstruction scenes.

### C. Decision-level Fusion

Decision-level fusion is a process of combining the decisions or outcomes derived independently from multiple detection systems. The aim is to increase the certainty of the final decision, reducing the probability of false positives and enhancing the overall system's reliability. In this subsection, we first introduce the methods of extracting meaningful information from the data before entering the decision classifier, such as extracting the radar cross-section (RCS) of the target. We then discuss the decision-level fusion methodologies with specific examples. We also present applications of decision-level fusion in the context of mmWave radar sensing.

1) *Decision-level Processing Methodologies:* RCS is a physical quantity that measures the intensity of the echo generated by the target illuminated by the radar wave. We obtain the RCS value of the object reflection to judge the shape characteristics of the target. But off-the-shelf radars (e.g., the TI IWR series) do not provide RCS results directly. Therefore, we need to obtain the other parameters (including the radar parameters) before calculating the RCS value.

In calculating RCS, the most critical and difficult problem is obtaining the SNR, which is the ratio of RX average signal strength and average noise strength. The noise mainly depends on the background noise of the radar circuit. Although the intensity of this noise is relatively stable, it is difficult to

measure the noise value on the integrated radar equipment directly. Therefore we do not directly obtain the noise value but consider that the RCS should always be proportional to the RX signal strength when the other parameters are determined. Therefore, we decided to use the corner reflector in the calibration phase to obtain the RX signal strength at different distances  $d$  and establish the benchmark database  $\mathcal{B}(d)$ . From  $\mathcal{B}(d)$ , we obtain the space Cartesian coordinates  $(x, y, z)$ , if the distance  $L = \sqrt{x^2 + y^2 + z^2}$ , then we calculate RX signal intensity  $P_{r_t}$  based on the following formula:

$$\sigma_p = \frac{P_{r_t}}{\mathcal{B}L} \sigma_r. \quad (12)$$

We obtain the benchmark database  $\mathcal{B}(d)$  during the calibration phase. From  $\mathcal{B}(d)$ , we obtain the signal-to-noise ratio, and then we obtain the RCS based on the following formula:

$$\sigma = \frac{(4\pi)^3 d^4 k T F S N R}{P_t G_{TX} G_{RX} \lambda^2 T_{meas}}. \quad (13)$$

Here  $k$  is the Boltzmann constant,  $T$  is the antenna temperature,  $F$  is the noise coefficient of RX,  $\lambda$  is the millimeter-wave wavelength (constant),  $P_t$  is the radar output power,  $G_{TX}$  is the TX Antenna Gain,  $G_{RX}$  is the RX Antenna Gain,  $T_{meas}$  is the measurement time, and  $d$  is the distance between the target and the millimeter-wave radar.  $d$  is computed from the radar output and the metadata. We obtain the RCS value of the target through the above method.

2) *Decision-level Fusion Methodologies*: Fig. 15 presents an illustrative example of an end-to-end liveness target detection system [46], emphasizing the implementation of decision-level fusion utilizing mmWave radar and vision data. This system's inputs consist of signal data from the mmWave radar and vision data from a camera, both of which undergo individual processing before being fused. The mmWave signal data is processed using the Radar Cross Section (RCS) calculation methodology, transforming it into a point cloud complete with RCS information. This point cloud is then projected into radar pixel images retaining the RCS information. Concurrently, vision data is processed through MediaPipe [114], producing target segmentation and posture results.

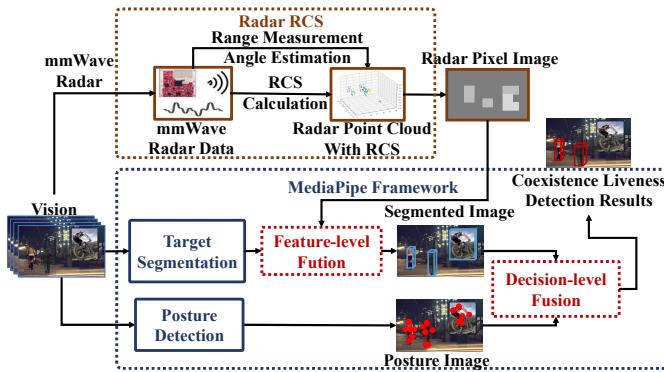


Fig. 15. Example of decision-level fusion

The processed radar and vision data are then combined using two specially designed fusion models, representing

a significant contribution to the multimodal fusion sensing approach. These fusion models function by integrating the mmWave radar and vision sequence data in a time-series format, exploiting unique features such as movements within the radar and image sequence data over time. In essence, this end-to-end liveness target detection system, bolstered by decision-level fusion, exemplifies the power and potential of multimodal fusion sensing. It enhances the robustness of the system, making it capable of efficiently detecting live targets even in challenging environments.

#### D. Lessons Learned: Summary and Insights

The methodologies of mmWave radar multi-modal fusion sensing offer invaluable insights into the integration of various sensor modalities for improved sensing. Key takeaways include the importance of proper data handling in data-level fusion, where synchronization and association play vital roles. In feature-level fusion, the distinction between model-independent and model-dependent methods highlights the versatility of fusion approaches, each leveraging different features of multi-modal data. Decision-level fusion demonstrated the potential of these methodologies in making associative decisions, bridging the gap between raw data and actionable insights. Through these methodologies, we realize the complexity of mmWave radar multi-modal fusion sensing, paving the way for future advancements in sensory technology. Furthermore, our review provides a comprehensive summary of multi-modal fusion sensing covering a wider range of modalities than existing reviews [25], [24], [23].

Specifically, in comparison to the work of Z. Wei et al. [25], our review provides a comprehensive summary of multi-modal fusion sensing covering a wider range of modalities. Our work broadens the scope by exploring how mmWave radar data are integrated with a variety of sensor modalities to enhance perception in multiple contexts and explores the challenges and solutions for mmWave radar fusion in various domains beyond autonomous driving. Besides, Yeong et al. [24] reviewed sensor and sensor fusion technology in autonomous vehicles but lacked the emphasis on the role and benefits of mmWave radar fusion. Our review highlights the integration and synergy of mmWave radar with other sensor modalities to achieve a more reliable and robust perception system in various scenarios, such as human sensing, industrial, and autonomous driving. Moreover, Y. Zhou et al. [23] summarized the current techniques and applications of mmWave radar and camera fusion, but did not provide a critical analysis of the performance and complexity trade-offs of various fusion strategies. Our review evaluates the strengths and weaknesses of various fusion methods, such as data-level, feature-level, and decision-level fusion, and identifies the key challenges and open issues for future research.

## V. APPLICATIONS

This section presents a comprehensive summary of the various applications of millimeter-wave (mmWave) radar sensing, as encapsulated in Table VI. We review various application scenarios that leverage mmWave radar sensing, including

TABLE VI  
SUMMARY OF MMWAVE RADAR MULTI-MODAL FUSION SENSING APPLICATIONS

Application Type	Application Scenario	mmWave Radar Sensing Applications
Detection and Recognition	Human Sensing	<b>Human Event Detection:</b> Heart Beat [48], [115], [86], [87], Breath [116], [88], Fall[43], [44], Posture Change [117], [103], [1], [118], Speech [89], [90], [119], [120], [121], Driving Fatigue [122], Food Intake [123] <b>Gait Recognition:</b> [63], [51], [124], [125], [126], [127], [128], [129], [130], [17], [131], [132] <b>Gesture Recognition:</b> Arm [66], [133], [73], [70], [72], Hand [49], [53], [55], [64], [134], [135], [136], [137], [37], Finger [138], [139], [140], Keystroke [141], [142]
	Industrial Sensing	<b>Vibration Detection:</b> [143], [144]
Tracking and Localization	Human, Autonomous Vehicle, and Industrial Sensing	<b>Tracking:</b> Trajectory [96], [69], [62], [145], [146], [147], ReID [59], [85], [54], [148], [149] <b>Localization:</b> [94], [85], [69], [56], [150], [151] <b>Egomotion:</b> [152], [153], [99], [82], [154]; <b>SLAM</b> [12], [155], [156]
		<b>Posture Reconstruction:</b> [95], [68], [60], [157], [158], [81], [96] <b>Fine-grain Body Reconstruction:</b> Mesh [67], [113], [96], [110] <b>3D Imaging:</b> [62], [71], [159]
Multi-modal Fusion	Human, Autonomous Vehicle, and Industrial Sensing	<b>Data-level Fusion:</b> [81], [78], [98] <b>Feature-level Fusion:</b> [160], [7], [14], [59], [82], [110], [161], [46], [154], [58], [79] <b>Decision-level Fusion:</b> [77], [76], [98]

human, autonomous vehicle, and industrial sensing. Through the implementation of data-level, feature-level, and decision-level fusion strategies, it integrates data from mmWave radar with other sensor modalities such as cameras, LiDAR, and ultrasonic sensors. Such multi-modal fusion significantly enhances the robustness and precision of sensing systems by delivering a thorough and accurate depiction of sensing data. The multi-modal fusion sensing via mmWave radar and other modalities enables various application scenarios.

Specifically, for human sensing, mmWave radar enhances the detection and recognition of human events, such as heart-beat, breath, falls, posture change, speech, driving fatigue, and food intake. It also enables the recognition of human gait and gestures, such as arm, hand, finger, and keystroke movements. These features are useful for health monitoring, biometric security, and human-computer interaction. Moreover, mmWave radar supports the reconstruction of human posture and fine-grain body mesh, which provide rich and accurate 3D data of the human body. For autonomous vehicle sensing, mmWave radar improves the tracking and localization of objects, such as trajectory, re-identification, and self-localization. It also assists the ego-motion and SLAM of the vehicle, which is essential for navigation and mapping. For industrial sensing, mmWave radar facilitates the detection of vibration and the imaging of 3D structures, which are important for quality control.

#### A. Detection and Recognition

This section presents a summary and Comparison of different Activity Recognition applications in the mmWave field. As shown in Table VII, the signal pre-processing techniques, algorithms, and performance results are summarized. For signal pre-processing, DN represents Denoise, ST represents Signal Transform, and FE stands for Signal Feature Extraction. In the algorithm column, M and L stand for modeling-based and learning-based algorithms, respectively. Learning-based

algorithms, e.g., Convolutional Neural Network(CNN) and Long Short-Term Memory Network(LSTM) are widely used. Among the 12 papers based on the deep learning model, 9 papers utilize CNN, 4 papers utilize Long Short Term Memory (LSTM), 2 papers utilize Graph Neural Network(GCN), and 2 papers utilize PointNet. Several Shallow Learning Algorithm are also utilized among the remaining 5 papers, e.g., SVM, Random Forest, SOM, and LVQ. 3DFFT is widely used in pre-processing phase to obtain point clouds. Noise reduction is usually needed after point cloud-based signal transfer. DB-SCAN are commonly utilized to tackle multi-person scenario and eliminate speckles. Kalman Filter is often utilized in trajectory tracking tasks after clustering.

#### B. Tracking and Localization

The summary of mmWave radar human localization and tracking applications is presented in Table VIII. A typical pipeline for mmWave radar human localization and tracking system consists of the following stages: detection, fusion (optional), data association, and tracking. In the detection stage, the position of subjects is measured by radar (and other sensors if it is a multi-sensor system). In the data association stage, the position of subjects in successive frames is linked into trajectories, and in the tracking stage, subjects' positions are estimated. A mmWave radar with an antenna matrix has the capacity to provide subjects' range, velocity, and angle information. Several works utilize sensor fusion techniques to achieve better localization performance. The basic idea is to make utilization of data from other sensors to enhance the location accuracy of subjects in azimuth and elevation. Most works [71], [62] choose the camera as another sensor for its high resolution in cross-range, and its shortage of depth sensing is covered by mmWave radar.

Another approach for improving cross-range resolution is to utilize a pair of radars set orthogonally, as the cross-range

TABLE VII  
SUMMARY OF MMWAVE RADAR MULTI-MODAL FUSION SENSING: DETECTION AND RECOGNITION APPLICATIONS

Reference	Signal Processing	Algorithm	Application	Performance
mBeats [48]	micro-Doppler	P: Band-pass Filters, DBSCAN; M: Inter-frame Detection; L: DNN	Heart Beat Detection	Accuracy: 93.03%
RF-SCG [87]	range-Doppler	P: Spatio-temporal Filters; M: Beamforming; L: Hybrid CNN	Heart Beat Detection	Median Error: 0.26% to 1.29%
HeartPrint [86]	range-Doppler	P: Band-pass Filters, DBSCAN; M: Wavelet Packet Transform, Recursive Feature Elimination; L: RBF-SVM	Heart Beat Detection & Authentication	Authentication Accuracy: >95%
M-Auth [88]	range-Doppler	P: Band-pass Filters, Adaptive Filter, K-means; L: RBF-SVM	Breath Detection & Authentication	Authentication Accuracy: 96%, Attack Detection Rate: >95%
Jin-mmFall [43]	Point Cloud	P: Kalman Filter, DBSCAN; M: Motion Pattern Estimation; L: Hybrid RNN	Fall Detection	Accuracy: 98%
Li-mmFall [44]	range-Doppler	P: High-pass Filters; L: CNN	Fall Detection	Accuracy: 98.8%
Jin-2019 [117]	micro-Doppler & Point Cloud	P: Kalman Filter, DBSCAN; L: CNN	Posture Change Detection	Accuracy: 98.94%
tinyRadar [1]	range-Doppler	L: CNN	Posture Change Detection	Accuracy: 96.43%
m-Activity [118]	Point Cloud	P: DBSCAN; M: Voxelization; L: Hybrid CRNN	Posture Change Detection	Real-time/Offline Accuracy: 93.25%/91.52%
RadioSES [89]	range-Doppler	P: High-pass Filters, DBSCAN; M: CIR; L: Hybrid RNN	Speech Detection	Accuracy: 98.97%
mmPhone [90]	micro-Doppler	P: How-pass Filters; M: Transmission Line model; L: DNN	Speech Detection	Accuracy: >93%
RFWash [53]	ST: FFT;DN:high-pass filter	L:CNN+BiLSTM	Gesture Recognition	gesture error rate:<8%
Tesla-Rapture [73]	ST:Range-FFT,Doppler-FFT,CFAR,Angle-FFT	L:TFNet+Temporal Graph kNN+mlp	Gesture Recognition	Accuracy:98.1%
Pantomime [70]	ST:Range-FFT,Doppler-FFT,CFAR,Angle-FFT,AHC,K-means; DN:DBSCAN	L:PointNet++ +LSTM+FC	Gesture Recognition	Accuracy:95%, AUC:99%
Soli [138]	ST:FFT; FE:RDMD,RDMD,RDTD, SPMI	L:Random Forest+Bayesian filter;M:Scattering center model	Gesture Recognition	Accuracy:92.10%
RadarNet [55]	ST:ARD,CRD	L:CNN+LSTM	Gesture Recognition	inference time:0.147ms; Accuracy:90%
mmASL [91]	ST:Doppler shifts,STFT	L:CNN+Multitask Learning	Gesture Recognition	Accuracy:87%
mTransSee [72]	ST:Range-FFT,Doppler-FFT,CFAR,Angle-FFT;DN:UDAN	L:CNN+FC	Gesture Recognition	Accuracy:98%
D. Cao et al. [59]	ST:Range-FFT,Doppler-FFT,CFAR,Angle-FFT;DN:DBSCAN	L:PointNet++ +LSTM+FC	Gait Recognition	Top1 Accuracy:92.5%
MU-ID [54]	ST:Range-FFT,Doppler-FFT;DN:high-pass filter	L:CNN+FC	Gait Recognition	Identification accuracy:97%(single-person)/92%(multi-person)
mmGaitNet [63]	ST:Range-FFT,Doppler-FFT; DN:DBSCAN	L:COV+FC	Gait Recognition	Accuracy:90% for single-person scenarios,88% for five co-existing persons
Stpointgcn [151]	ST:Range-FFT,Doppler-FFT	L:GCN	Gait Recognition	Accuracy:60%

TABLE VIII  
SUMMARY OF MMWAVE RADAR MULTI-MODAL FUSION SENSING: TRACKING AND LOCALIZATION APPLICATIONS

Ref.	Signal Processing	Algorithm	Application	Performance
Zhao-19 [85]	Point cloud	P:DBSCAN, M:Hungarian Algorithm & KF, L:LSTM (identification)	Tracking and Identification	position error: 0.16m, identification accuracy: 86%
Sengupta-19 [62]	Point cloud and image (fusion with camera)	M:Decision Box (fusion); L:Camera Bounding-Box DNN & LSTM	Tracking	Mean error: 0.01m at cross-range, 0.002m at depth-range
Pegoraro-21 [97]	Range-Doppler or Range-Doppler-Angle	P:DBSCAN; M:Hungarian Algorithm & KF; L:DCNN with IBs	Tracking and Identification	identification accuracy: 95% for RD data, 98% for RDA data
Cui-21 [71]	Point clouds and image (fusion with camera)	P:DBSCAN, M:Multiple Hypothesis Tracking & EKF	Tracking	Multiple Object Tracking Accuracy (MOTA): 89.5%
Kong-22 [96]	Range-Doppler	P:MVDR, M:K-means+coordinate-corrected EKF	Tracking	Overall joint tracking error: 32.4mm, 34.9mm, 38.6mm, and 42.4mm for 1-user, 2-user, 3-user, and 4-user scenarios
Cui-21 [69]	Point clouds(multi-radars)	P:DBSCAN, M:KF.	Tracking	Sensitivity on human detection:98.6%
Zhang-22 [145]	Range-Doppler	P:Binary-phase multiplex, M:Sequential Monte Carlo & Recursive Bayesian.	Tracking	Detection probability:98%
Wu-21 [146]	Point clouds	P:DBSCAN.	Tracking	Detection Precision:98.6% , Detection Sensitivity:97.1%
Shastri-22 [147]	Point clouds(multi-radars)	M:SVD and LS & Hungarian Algorithm & KF.	Tracking	Position error:11cm, Orientation error:2.78°
Cao-22 [59]	Point clouds	P:DBSCAN; L:DNN.	Tracking and Re-identification	TOP-1 Accuarcy:92.5%
Yang-20 [54]	Range-Doppler	P:High-pass filter; L:CNN.	Tracking and Identification	Identification Accuarcy:97%(single-person),92%(multi-people)
Zhao-21[99]	Point clouds	P:DBSCAN; M:Hungarian Algorithm & KF; L:LSTM.	Tracking and Identification	Identification Accuarcy:99%, Position errors:0.16 m
Wu-20[94]	Range-Doppler	P:k-means; M:KM (Kuhn and Munkres) algorithm; L:LSTM.	Localization	Identification Accuarcy:99%, Localization errors:9.9 cm(dynamic targets), 19.74 cm(static targets)
Pegoraro-20[150]	Range-Doppler-Angle	P:DBSCAN, L:DAE and S2S.	Localization and Tracking	Tracking errors : 12 cm
Wang-22[151]	Point clouds	L:GNN.	Tracking and Identification	Identification Accuarcy:68.88%(single-person),75.51%(multi-people)
Zhou-22[152]	Point cloud and image (fusion with camera)	M:3D KF.	Egomotion	Multiple Object Tracking Accuracy (MOTA): 60.6%
Wang-21[153]	Range-Azimuth Heatmaps and image (fusion with camera)	M:location-based nonmaximum suppression,L:CNN.	Egomotion	Detection Precision: 85.62%
Zhao-21[99]	Heatmaps	L:CNN & LSTM.	Egomotion	Localization error: 8.92cm
Deng-22[82]	Point cloud and image (fusion with camera)	M:Fusion module; L:DNN.	Egomotion	mAP: 96.51%
Deng-22[154]	Point cloud and image (fusion with camera)	L:CNN & DNN.	Detection and Tracking	AP: 91.5%
Lu-20[12]	Point cloud	L:GAN & CNN.	SLAM	Material Classification Accuracy: 92%, Reconstruction Error:0.83m
Blanco-22[155]	CSI	P:FTM, M: mDtrack.	SLAM	location error:18cm
Gao-22[156]	Point cloud	M:RSD, L:KNN.	SLAM	translation error:0.29m, rotation error:0.0017rad

TABLE IX  
SUMMARY OF MMWAVE RADAR MULTI-MODAL FUSION SENSING: RECONSTRUCTION APPLICATIONS

Reference	Signal Processing	Algorithm	Application	Performance
mm-Pose [60]	Radar-to-image	L:Forked CNN	Human Pose Estimation	Average localization errors: 3.2 cm (X), 7.5 cm (Y), 2.7 cm (Z) of 17 joints
Li-2020 [95]	range-angle	L:Forked CNN	Human Pose Estimation	Average OKS: 0.705 AP 50: 0.877 of 14 joints
mmPose-NLP [157]	Point cloud	P:DBSCAN L:GRU & Attention	Human Pose Estimation	<3cm (XYZ) of 25 joints
MARS [68]	Point cloud	P:Sorting L:CNN	Rehabilitation & Human Pose Estimation	5.87cm of 19 joints
mmMesh [67]	Point cloud	P:Highest point Picking L:PointNet & Attention & Dynamic Anchor	Mesh Reconstruction	2.47cm of mesh & 2.18cm of 24 joints
m3Track [96]	range-azimuth & range-elevation	P: MVDR; M:EKF L:Forked-ConvLSTM	Human Posture Tracking	Single user: 3.24cm Four-users: 4.24cm of 17 joints
An-2022 [158]	Muliti-Frame Point Cloud	L:Meta-Learning	Human Pose Estimation	4X faster than MARS
MI-Mesh [162]	Point cloud	P:phase calibration L:Attention & DNN	Mesh Reconstruction	MPJPE=2.18cm
3DRIMR [159]	range-azimuth-elevation	M:SAR L:Conditional GAN	Car Reconstruction	2D depth image: range error = 8cm, orientation error = 4.8 deg
ImmFusion [110]	Point cloud & image	L:Transformer based fusion	Mesh Reconstruction	MPJPE=5.9cm

TABLE X  
SUMMARY OF MMWAVE RADAR MULTI-MODAL FUSION SENSING: FUSION APPLICATIONS

Reference	Fusion Modalities	Algorithm	Application	Performance
CRAFT [160]	mmWave radar and vision	M: Feature Encoder;	3D Object Detection	mAP:41.1% NDS:52.3%
DepthEstimation [7]	mmWave radar and vision	M: MLP; L: Encoder-Decoder	Depth Estimation	MAE:2179.3, RMSE:4898.7
Vision-RF [59]	mmWave radar and vision	L:PointNet++ & LSTM & FC	Gait Recognition	Top1 Accuracy:92.5%
Geryon [82]	mmWave radar and vision	M:Fusion module; L:DNN	Egomotion	mAP: 96.51%
STFWF [76]	mmWave radar, vision, Lidar, and NIR	P: BeV projection; L: VGG	Object Detection	AP: 90.60%
ImmFusion [110]	mmWave radar and vision	L:Transformer based fusion	Mesh Reconstruction	MPJPE=5.9cm
CramNet [161]	mmWave radar and vision	L: Cross-Attention;	3D Object Detection	AP:62.07%
PLD [46]	mmWave radar and vision	L: Attention	Liveness Detection	mAP:97.7%
GLE-Net [154]	mmWave radar and vision	L:CNN & DNN	Detection and Tracking	AP: 91.5%
VDAT [78]	mmWave radar, LiDAR, and IMU	P:DBSCAN; M: Geometric Model	Vehicle Detection and Tracking	precision rate:99.88%
milliEgo [58]	mmWave radar, LiDAR, and IMU	L:CNN & RNN & LSTM	Egomotion Estimation	3D error drift 1.3%
RadarNet [98]	mmWave radar and LiDAR	P: BeV projection; L: Feature Pyramid Network	3D Object Detection	AP:87.9%

dimension of one radar is covered by the range dimension of another [69]. Kalman Filter (KF) or Extended Kalman Filter (EKF) is utilized for trajectory tracking, while learning-based methods have emerged in recent years. Most works employ KF or EKF for tracking with modifications based on different data formats or specific task requirements. Another work [62] utilizes learning-based methods for tracking. The location of the subject at each timestamp is fed into an LSTM module to generate trajectory. There are numerous researches [56], [85] combine tracking tasks with identification tasks. This is due to the complementarity of the two tasks. Especially in the multi-person scenario, identification helps determine the correct association of two subjects in successive frames when

intersections or occlusions exist in their trajectory.

### C. Reconstruction

The summary of mmWave-based reconstruction applications is presented in Table IX. For reconstruction applications, most of the current papers focus on skeletal pose reconstruction. There is also a paper [67] on reconstructing mesh and a paper [158] focusing on meta-learning to improve the generalization ability of reconstruction tasks. Papers utilize deep learning models due to the challenges that mmWave pose to the reconstruction task. In terms of model input, papers utilize point clouds generated by mmWave signals, while others utilize heat maps. The models utilized to process point

clouds include CNN, GRU, and PointNet, and the heatmaps are mainly processed by CNN and ConvLSTM due to their excellent feature extraction ability on image structure. Due to the sparsity of the point cloud, it is naturally not suitable to utilize CNN to process point cloud data, so the work [68] preprocess point cloud to map to low dimensional space. The existing works continuously improve the accuracy of human reconstruction tasks and complete human pose information. The mmMesh [67] combined with the human body model shows that mmWave perception information contains abundant human body information, including body shape, gender, and so on, which shows the great potential of mmWave perception.

#### D. Multi-modal Fusion

Table X elaborates on the diverse applications of millimeter-wave radar sensing with multi-modal fusion. Multi-modal fusion is a central theme in these applications, demonstrating the crucial role of mmWave radar in various fields. The fusion applications span numerous areas, including object detection, gait recognition, ego-motion estimation, mesh reconstruction, liveness detection, and vehicle detection and tracking. The fusion methodologies that combine mmWave radar with LiDAR [76], [78], [98], vision [160], [7], [59], [82], [76], [110], [161], [46], [154], [23], [24], [25], IMU [76], [78], [58], and NIR [76] data. These fusion techniques are implemented utilizing a variety of algorithms, such as MLP, Encoder-Decoder, PointNet++, LSTM, and Transformers, to process and integrate sensor data effectively.

#### E. Insights of mmWave Radar Multi-modal Fusion Sensing

Millimeter-wave radar excels at detecting micro-vibrations, making it an ideal technology for various applications. Its ability to sense subtle movements sets it apart from traditional sensors. The high-resolution imaging capabilities of mmWave radar, resulting from its high frequency and short wavelength, provide superior resolution compared to lower-frequency alternatives like microwave radar. In multi-modal fusion sensing, mmWave radar works synergistically with other sensing modalities, such as cameras, infrared sensors, and ultrasonic sensors, to achieve more accurate and reliable results. By integrating data from multiple sources, the system overcomes individual limitations and enhance overall performance. When combined with multi-modal fusion sensing, mmWave radar is effectively applied in areas such as real-time vital sign monitoring, gesture recognition, and structural health monitoring. These applications benefit from the improved resolution and detection capabilities of mmWave radar, enabling new possibilities and innovations across diverse domains.

## VI. CHALLENGES AND FUTURE TRENDS

This section discusses the challenges and future trends of mmWave radar sensing, both as a standalone modality and as part of multi-modal fusion sensing. We first introduce the existing challenges in mmWave radar sensing, such as hardware limitations, signal processing complexity, and data scarcity. We then explore the future trends of mmWave radar sensing, such as security and privacy, interference-tolerant, and multi-device

sensing. We also examine the future trends of multi-modal fusion sensing with mmWave radar and other sensors, such as cross-modal data generation, self-supervised learning, and end-to-end fusion. We aim to provide a comprehensive and forward-looking perspective on the development and potential of mmWave radar sensing.

#### A. Existing Challenges in Millimeter-Wave Radar Sensing

In human sensing, the precision of the mmWave radar is critical, as it is often utilized for vital sign monitoring [163], [164] and gesture recognition. The sensor must accurately distinguish between minor movements and the inherent noise present in the signal. In the context of autonomous vehicle sensing, robustness becomes a primary concern. These sensors must perform reliably under various environmental conditions, such as rain, fog, and dust, which attenuate mmWave signals. For industrial sensing, challenges often revolve around integrating mmWave radar sensors into complex industrial systems. These environments are highly reflective or absorptive, leading to signal attenuation and multipath interference issues, which significantly affects sensor performance.

Specifically, mmWave radar sensing exhibits significant advantages over other radio frequency sensing technologies in resolution and anti-interference capabilities. However, it also faces challenges in practical applications, mainly classified into hardware, data, and algorithmic aspects. In terms of hardware, a major challenge of mmWave radar lies in its low angular resolution [165], primarily due to limitations of the antenna hardware, such as size and layout. These constraints limit the angular resolution of the mmWave radar, adversely affecting the imaging quality. This is particularly problematic in high-precision sensing and positioning scenarios, where insufficient accuracy might fail to meet user needs [166].

Data-wise, the issues with Doppler-based data primarily revolve around the high level of speckle noise, which implies that the actual data often includes a lot of irrelevant or erroneous information. Thus, this impacts the accuracy and reliability of radar sensing. Doppler data do not intuitively represent three-dimensional spatial information, potentially leading to difficulties processing complex scenarios or performing advanced analysis. Furthermore, point-cloud-based data also face challenges. On one hand, the sparsity of data points might degrade the data quality and sensing precision. On the other hand, due to the specular reflection property of mmWave radar, the perceived information is often incomplete, which might hinder its application effectiveness.

Algorithmically, the majority of current methods for extracting useful information from mmWave radar raw signals are still based on traditional processing techniques, such as the Fourier Transform and Constant False Alarm Rate (CFAR). However, these methods might not handle interference signals perfectly. They mistakenly eliminate signals containing target information or retain signals with speckle noise. At present, there is no perfectly effective method for processing raw signals, eliminating interference information while preserving target information. This represents a problem that requires further research and resolution.

### B. Security and Privacy in mmWave radar Sensing

Due to its sparse and unintuitive output, the mmWave radar is considered a potential substitute for the camera in sensing and detection. This characteristic brings mmWave radar advantages in private scenarios. As the increasing number of works provide accurate detection of location, human pose, and subtle movements using mmWave radar, it has also become a threat to privacy. Moreover, mmWave radars' advantage, for example, its capacity of through-wall and working under poor illumination, make it a more concerning and troublesome threat. Thus, attack and defense techniques [167] are the topic of interest in mmWave sensing [168], [169], [170]. Spying attacks are conducted using mmWave radar on phone calls, screens, and even on physically isolated servers. Several works [171], [119], [120], [90] focus on the feasibility of spying on phone call contents at a long distance using commercial mmWave radar. The basic idea of these works is to catch the phone earpiece's vibration and recover the speaker's contents on the other side of the phone. WaveSpy [8] explores the feasibility of spying on the contents displayed on liquid-crystal display (LCD) screens, utilizing the fact that mmWave signal reflected by LCD screens varies on different liquid crystal nematic patterns. A more creative attack is carried out in SpiralSpy[172], where the researchers utilize mmWave radar to accumulate, convert channel, and filter data from physically isolated devices with the help of a hacked cooling fan. Data are transmitted utilizing modulated fan speed, which is caught by the mmWave radar. On the other hand, RF-protect[173] provides a defense technique against human tracking. This is realized by creating fake reflections, generating "ghost" trajectories, and causing disruption in the eavesdropper's information.

### C. Interference-Tolerant in mmWave radar Sensing

Multiple mmWave radar sensing provides more sensing information, but interference among multiple radars is also becoming a significant issue. For example, in autonomous vehicles, the mutual interference of radar signals will lead to disastrous consequences [174], such as false targets and reduced detection performance. Improving the anti-jamming capability of radar signals is helpful in improving the robustness and versatility of mmWave perception. To Date, detection and repair is the mainstream anti-jamming scheme among multiple radars. This scheme first detects the affected received signal and then recovers the waveform to make the signal approach the original waveform without interference [175], [176], [177], [178], [179], [180], [181], [182], [183], [184], [185]. Because the target signal and the interfering signal have different characteristics in different variation domains, the interfering signal is filtered better in the transformation domain. The transformation-based scheme is the most advanced anti-interference scheme [182]. Short-time Fourier transform (STFT) and wavelet transform are commonly utilized. For example, STFT is utilized to convert radar signals from the time domain to the frequency domain and filter the signal in the frequency domain to eliminate interference.

Monitoring and avoidance is another anti-interference scheme. The traditional avoidance scheme reduces the overlap

of radar signals in time, frequency, and space. For example, using the idea of time division multiple access (TDMA), the chirp slope parameters of FMCW are randomly generated to avoid the time overlap of radar signals. Frequent division multiple access (FDMA) is adopted to randomly generate radar transmission frequency range to avoid overlapping radar signal frequencies. However, traditional avoidance schemes are unsuitable for dense RF environments such as autonomous vehicles, and more and more mmWave sensors need more robust monitoring and avoidance solutions. To avoid conflicts, mmWave signals monitor and change operating parameters before transmissions, such as transmission direction, waveform, or center frequency [186], [187], [188], [189]. Monitoring and avoidance schemes are being worked on by the industry.

With the integrated development of mmWave communication and perception, the industry has begun to focus on the interference-tolerant scheme of multiple radar sensor collaboration at this stage, which is a supplement to the monitoring and avoidance schemes. This scheme is based on the communication between multiple mmWave radars to coordinate the allocation of frequency and time resources [190], [191], [192], [193], [194], [195]. Based on the feature that mmWave spectrum is multiplexed in the spatial domain, frequency domain, time domain or combined with coding method, media access control (MAC) technology in time, frequency, and the coding domain is widely utilized to deal with interference. For example, in VANET-Assisted [190], coordination between radar sensors is established through inter-vehicle communication based on time division multiple access MAC protocol specification. SS-RTS [192] proposes the concept of the coded waveform to spread the inter-radar interference to a broader spectrum to reduce interference. With the development of Vehicular ad-hoc network, the sensing and communication cooperation between radars further improves the interference-tolerant ability of mmWave radars.

### D. Multi-device in mmWave radar Sensing

Various millimeter-wave radar sensing tasks leverage data sourced from multiple devices to enhance task performance or cater to specific tasks, such as continuous target tracking in indoor environments. The affordability and ease of deployment of commercial millimeter-wave radars forecast an increase in the prevalence of these devices across diverse scenarios, including beadhouses, workstations, offices, crossroads, sports fields, malls, and beyond. Fundamental applications of mmWave radar involve localization [155], [153], [156] and various classification tasks, such as gesture recognition [133], [66], [138]. Early research utilized mmWave radars for indoor localization by installing multiple devices at anchor points. These radars, positioned at known locations, facilitated more accurate target localization through their respective signal transmission and reception. GaitCube [125] leverages the raw data collected by multiple radars to achieve continuous target recognition in indoor environments.

Sensors of other modalities, such as LiDAR, also combine point clouds from multi-devices for higher performance and efficiency. VI-Eye [196] simultaneously utilizes the point clouds

generated by the vehicle's radar and the point clouds provided by the roadside infrastructure. It aligns them to achieve real-time registration of point clouds, providing greater sensing range and object detection capability.

### E. Multi-modal in mmWave radar Fusion Sensing

The fusion of millimeter-wave (mmWave) radar with diverse sensor modalities is set to redefine the sensing landscape in various scenarios. The synergy between mmWave radar and cameras [197], [20], [198], [199], [46], LiDAR [13], [19], [18], [19], [154], or a fusion of both [76], is particularly transformative in autonomous driving, where it enhances object detection and localization. The distinct sensing modalities capture varied target information, leading to a comprehensive situational awareness. mmWave radar excels in adverse weather, complementing the high cost and deployment challenges of LiDAR, and the rich detail from optical sensors. The fusion approach has been revolutionized by deep learning, yielding impressive outcomes. Nonetheless, the deep learning paradigm has highlighted a critical challenge—the scarcity of robust datasets for training complex models, prompting a need for multi-sensor data collection across diverse scenarios.

Future trends indicate a surge in sensor technology advancements, algorithmic innovation, and expansion in application areas. The integration of artificial intelligence within multi-modal sensor frameworks is poised to further elevate system autonomy. Sensor miniaturization and cost reduction will democratize the use of advanced sensor fusion, making it accessible across industries. Novel fusion techniques are also anticipated, which could elevate performance even in the most challenging environmental settings. The demand for large-scale, diverse, and realistic datasets is becoming ever more pressing, as they are essential for training and validating the next generation of fusion systems. This necessitates a concerted effort to design and compile data resources that capture the complexity of real-world dynamics. Another aspect shaping the future of mmWave radar fusion sensing is the growing need for real-time processing capabilities. This is crucial for applications that require rapid decision-making, such as autonomous vehicles and dynamic environmental monitoring. The mmWave radar's potential extends beyond vehicular applications, with fusion possibilities involving audio signals [200] for vibration detection, IMUs [58] for motion estimation, and RGBD cameras [59] for intricate re-identification tasks. The exploration of cross-modal dataset transformation [201], [84] is an innovative response to the rarity of mmWave data, enabling the effective training of models in the absence of extensive real-world radar data.

In conclusion, mmWave radar is capable of forming the core of a robust sensory network when fused through various levels and stages with other sensors. The technological trajectory suggests an evolution toward systems with unparalleled sensing capabilities under any weather condition, driven by advancements in sensor technology, fusion algorithms, and the imperative for large-scale, realistic datasets.

## VII. CONCLUSION

In conclusion, this review provides a comprehensive analysis of mmWave radar multi-modal fusion sensing, with a specific focus on human sensing applications. Through an evaluation of its significance, underlying principles, and opportunities of multi-modal fusion sensing via mmWave and other modalities, we highlight the transformative potential of mmWave radar in sensing applications. We examine the characteristics of mmWave radar data and available datasets, emphasizing the advantages and challenges associated with Doppler-based and point cloud-based data. Furthermore, we explore various methodologies proposed to tackle these challenges, including signal processing techniques and learning algorithms. The review showcases a wide range of applications for mmWave radar multi-modal fusion sensing, emphasizing its unique strengths in detection and recognition, tracking and localization, and reconstruction. By identifying potential future research directions, we aim to inspire readers to delve deeper into the study of mmWave radar multi-modal fusion sensing and contribute to overcoming current challenges. Ultimately, through ongoing advancements and expanding the application scope, we can propel the field of mmWave radar multi-modal fusion sensing, enabling innovative solutions and enhancing human perception capabilities.

## REFERENCES

- [1] S. S. Yadav, R. Agarwal, K. Bharath, S. Rao, and C. S. Thakur, "tinyradar: mmwave radar based human activity classification for edge computing," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2022, pp. 2414–2417.
- [2] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "mmvib: micrometer-level vibration measurement with mmwave radar," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.
- [3] J. Guo, Y. He, C. Jiang, M. Jin, S. Li, J. Zhang, R. Xi, and Y. Liu, "Measuring micrometer-level vibrations with mmwave radar," *IEEE Transactions on Mobile Computing*, vol. 22, no. 4, pp. 2248–2261, 2023.
- [4] Z. Sun, S. Balakrishnan, L. Su, A. Bhuyan, P. Wang, and C. Qiao, "Who is in control? practical physical layer attack and defense for mmwave-based sensing in autonomous vehicles," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3199–3214, 2021.
- [5] R. Komissarov and A. Wool, "Spoofing attacks against vehicular fmew radar," in *Proceedings of the 5th Workshop on Attacks and Solutions in Hardware Security*, 2021, pp. 91–97.
- [6] X. Chen, Z. Li, B. Chen, Y. Zhu, C. Lu, Z. Peng, F. Lin, W. Xu, K. Ren, and C. Qiao, "Metawave: Attacking mmwave sensing with meta-material-enhanced tags," in *30th Annual Network and Distributed System Security Symposium*. The Internet Society, Mar. 2023, pp. 1–17, the 30th Network and Distributed System Security (NDSS) Symposium 2023, NDSS 2023 ; Conference date: 27-02-2023 Through 03-03-2023.
- [7] A. D. Singh, Y. Ba, A. Sarker, H. Zhang, A. Kadambi, S. Soatto, M. Srivastava, and A. Wong, "Depth estimation from camera image and mmwave radar point cloud," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9275–9285.
- [8] Z. Li, F. Ma, A. S. Rathore, Z. Yang, B. Chen, L. Su, and W. Xu, "Wavespy: Remote and through-wall screen attack via mmwave sensing," in *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020, pp. 217–232.
- [9] C. Wang, F. Lin, Z. Ba, F. Zhang, W. Xu, and K. Ren, "Wavesdropper: Through-wall word detection of human speech via commercial mmwave devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 2, jul 2022.
- [10] J. Liu, C. Xiao, K. Cui, J. Han, X. Xu, and K. Ren, "Behavior privacy preserving in rf sensing," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 784–796, 2022.

- [11] T. Gu, Z. Fang, Z. Yang, P. Hu, and P. Mohapatra, "Mmsense: Multi-person detection and identification via mmwave sensing," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 45–50.
- [12] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See through smoke: robust indoor mapping with low-cost mmwave radar," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020, pp. 14–27.
- [13] K. Qian, S. Zhu, X. Zhang, and E. L. Li, "Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 444–453, 2021.
- [14] S. Wang, D. Cao, R. Liu, W. Jiang, T. Yao, and C. X. Lu, "Human parsing with joint learning for dynamic mmwave radar point cloud," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–22, 2023.
- [15] Z. Hong, Y. Petillot, and S. Wang, "Radarsslam: Radar based large-scale slam in all weathers," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5164–5170.
- [16] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh, "Through fog high-resolution imaging using millimeter wave radar," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11464–11473.
- [17] U. Niazi, S. Hazra, A. Santra, and R. Weigel, "Radar-based efficient gait classification using gaussian prototypical networks," in *2021 IEEE Radar Conference (RadarConf21)*. IEEE, 2021, pp. 1–5.
- [18] H. Yin, R. Chen, Y. Wang, and R. Xiong, "Rall: End-to-end radar localization on lidar map using differentiable measurement model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 6737–6750, 2022.
- [19] Y.-J. Li, J. D. Park, M. O'Toole, and K. Kitani, "Modality-agnostic learning for radar-lidar fusion in vehicle detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 908–917, 2022.
- [20] Y. Long, D. D. Morris, X. Liu, M. Castro, P. Chakravarty, and P. Narayanan, "Full-velocity radar returns by radar-camera fusion," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16 178–16 187, 2021.
- [21] L. Fan, L. Xie, X. Lu, Y. Li, C. Wang, and S. Lu, "mmmic: Multi-modal speech recognition based on mmwave radar," in *The 42nd International IEEE Conference on Computer Communications (INFOCOM)*, 2023.
- [22] T. Liu, F. Lin, C. Wang, C. Xu, X. Zhang, Z. Li, W. Xu, M.-C. Huang, and K. Ren, "Wavoid: Robust and secure multi-modal user identification via mmwave-voice mechanism," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '23. New York, NY, USA: Association for Computing Machinery, 2023.
- [23] Y. Zhou, Y. Dong, F. Hou, and J. Wu, "Review on millimeter-wave radar and camera fusion technology," *Sustainability*, vol. 14, no. 9, 2022.
- [24] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, 2021.
- [25] Z. Wei, F. Zhang, S. Chang, Y. Liu, H. Wu, and Z. Feng, "Mmwave radar and vision fusion for object detection in autonomous driving: A review," *Sensors*, vol. 22, no. 7, 2022.
- [26] A. Shastri, N. Valecha, E. Bashirov, H. Tataria, M. Lentmaier, F. Tufvesson, M. Rossi, and P. Casari, "A review of millimeter wave device-based localization and device-free sensing technologies and applications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1708–1749, 2022.
- [27] J. Yang, H. Huang, Y. Zhou, X. Chen, Y. Xu, S. Yuan, H. Zou, C. X. Lu, and L. Xie, "Mm-fi: Multi-modal non-intrusive 4d human dataset for versatile wireless sensing," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [28] H. Lu, M. Mazaheri, R. Rezvani, and O. Abari, "A millimeter wave backscatter network for two-way communication and localization," in *Proceedings of the ACM SIGCOMM 2023 Conference*, ser. ACM SIGCOMM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 49–61.
- [29] G. R. Curry, *Radar system performance modeling*. Artech, 2004.
- [30] J. M. Schellberg and S. Sur, "Visar: A mobile platform for vision-integrated millimeter-wave synthetic aperture radar," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '21 Adjunct. New York, NY, USA: Association for Computing Machinery, 2021, p. 69–71.
- [31] R. Nikandish, A. Yousefi, and E. Mohammadi, "Spurs in millimeter-wave fmcw radar system-on-chip," *IEEE Transactions on Radar Systems*, vol. 1, pp. 21–33, 2023.
- [32] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," *Texas Instruments*, pp. 1–8, 2017.
- [33] S. Scherr, S. Ayhan, H. Gulen, M. Pauli, and T. Zwick, "61 ghz ism band fmcw radar for applications requiring high accuracy," in *2014 Asia-Pacific Microwave Conference*. IEEE, 2014, pp. 1118–1120.
- [34] T. Jaeschke, C. Bredendiek, S. Küppers, and N. Pohl, "High-precision d-band fmcw-radar sensor based on a wideband sige-transceiver mmic," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 12, pp. 3582–3597, 2014.
- [35] S. Scherr, B. Göttel, S. Ayhan, A. Bhutani, M. Pauli, W. Winkler, J. C. Scheytt, and T. Zwick, "Miniaturized 122 ghz ism band fmcw radar with micrometer accuracy," in *2015 European Radar Conference (EuRAD)*. IEEE, 2015, pp. 277–280.
- [36] M. Pauli, B. Göttel, S. Scherr, A. Bhutani, S. Ayhan, W. Winkler, and T. Zwick, "Miniaturized millimeter-wave radar sensor for high-accuracy applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1707–1715, 2017.
- [37] A. Patra, P. Geuer, A. Munari, and P. Mähönen, "mm-wave radar based gesture recognition: Development and evaluation of a low-power, low-complexity system," in *Proceedings of the 2nd ACM Workshop on Millimeter Wave Networks and Sensing Systems*, 2018, pp. 51–56.
- [38] "Ti-awr2944," [Website], <https://www.ti.com/product/cn/AWR2944>.
- [39] "Vayyar 4d imaging radar," [Website], <https://www.vayyar.com/>.
- [40] "Infineon 60ghz radar," [Website], <https://www.infineon.com/cms/en/product/sensor/radar-sensors/radar-sensors-for-iot/60ghz-radar>.
- [41] "Novelic-eval-tinyrad," [Website], <https://www.novelic.com/#>.
- [42] B. Xie, M. Cui, D. Ganeshan, and J. Xiong, "Wall matters: Rethinking the effect of wall for wireless sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 4, jan 2024.
- [43] F. Jin, A. Sengupta, and S. Cao, "mmfall: Fall detection using 4-d mmwave radar and a hybrid variational rnn autoencoder," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 2, pp. 1245–1257, 2022.
- [44] W. Li, D. Zhang, Y. Li, Z. Wu, J. Chen, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, "Real-time fall detection using mmwave radar," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 16–20.
- [45] S. Wang, L. Mei, Z. Yin, H. Li, R. Liu, W. Jiang, and C. X. Lu, "End-to-end target liveness detection via mmwave radar and vision fusion for autonomous vehicles," *ACM Trans. Sen. Netw.*, 2023.
- [46] H. Li, R. Liu, S. Wang, W. Jiang, and C. X. Lu, "Pedestrian liveness detection based on mmwave radar and camera fusion," *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 262–270, 2022.
- [47] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4338–4364, 2021.
- [48] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, "Heart rate sensing with a robot mounted mmwave radar," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2812–2818.
- [49] W. Jiang, Y. Ren, Y. Liu, Z. Wang, and X. Wang, "Recognition of dynamic hand gesture based on mm-wave fmcw radar micro-doppler signatures," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4905–4909.
- [50] S. Chen, W. He, J. Ren, and X. Jiang, "Attention-based dual-stream vision transformer for radar gait recognition," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3668–3672.
- [51] S. Hor, N. Poole, and A. Arbabian, "Single-snapshot pedestrian gait recognition at the edge: A deep learning approach to high-resolution mmwave sensing," in *2022 IEEE Radar Conference (RadarConf22)*. IEEE, 2022, pp. 1–6.
- [52] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 2016, pp. 851–860.

- [53] A. Khamis, B. Kusy, C. T. Chou, M.-L. McLaws, and W. Hu, "Rfwash: a weakly supervised tracking of hand hygiene technique," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 572–584.
- [54] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "Mu-id: Multi-user identification through gaits using millimeter wave radios," *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pp. 2589–2598, 2020.
- [55] E. Hayashi, J. Lien, N. Gillian, L. Giusti, D. Weber, J. Yamanaka, L. Bedal, and I. Poupyrev, "Radarnet: Efficient gesture recognition technique utilizing a miniature radar sensor," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–14.
- [56] J. Pegeraro, F. Meneghelli, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2020.
- [57] Z. Cao, W. Ding, R. Chen, J. Zhang, X. Guo, and G. Wang, "A joint global-local network for human pose estimation with millimeter wave radar," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 434–446, 2022.
- [58] C. X. Lu, M. R. U. Saputra, P. Zhao, Y. Almalioglu, P. P. B. de Gusmão, C. Chen, K. Sun, N. Trigoni, and A. Markham, "milliego: single-chip mmwave radar aided egomotion estimation via deep sensor fusion," *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020.
- [59] D. Cao, R. Liu, H. Li, S. Wang, W. Jiang, and C. X. Lu, "Cross vision-rf gait re-identification with low-cost rgb-d cameras and mmwave radars," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–25, 2022.
- [60] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, 2020.
- [61] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 51–56.
- [62] A. Sengupta, F. Jin, and S. Cao, "A dnn-lstm based target tracking approach using mmwave radar and camera sensor fusion," in *2019 IEEE national aerospace and electronics conference (NAECON)*. IEEE, 2019, pp. 688–693.
- [63] Z. Meng, S. Fu, J. Yan, H. Liang, A. Zhou, S. Zhu, H. Ma, J. Liu, and N. Yang, "Gait recognition for co-existing multiple people using millimeter wave sensing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 849–856.
- [64] Z. Xia, Y. Luomei, C. Zhou, and F. Xu, "Multidimensional feature representation and learning for robust hand-gesture recognition on commercial millimeter-wave radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 4749–4764, 2020.
- [65] K. Wang, Q. Wang, F. Xue, and W. Chen, "3d-skeleton estimation based on commodity millimeter wave radar," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. IEEE, 2020, pp. 1339–1343.
- [66] 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," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 4, no. 4, pp. 1–28, 2020.
- [67] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "Mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 269–282.
- [68] S. An and U. Y. Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare," *ACM Transactions in Embedded Computing Systems*, 2021.
- [69] H. Cui and N. Dahnoun, "High precision human detection and tracking using millimeter-wave radars," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 1, pp. 22–32, 2021.
- [70] S. Palipana, D. Salami, L. A. Leiva, and S. Sigg, "Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–27, 2021.
- [71] F. Cui, Y. Song, J. Wu, Z. Xie, C. Song, Z. Xu, and K. Ding, "Online multi-target tracking for pedestrian by fusion of millimeter wave radar and vision," in *2021 IEEE Radar Conference (RadarConf21)*. IEEE, 2021, pp. 1–6.
- [72] 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," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 1, pp. 1–28, 2022.
- [73] 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 Transactions on Mobile Computing*, 2022.
- [74] H. Cui, S. Zhong, J. Wu, Z. Shen, N. Dahnoun, and Y. Zhao, "Milipoint: A point cloud dataset for mmwave radar," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [75] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11621–11631.
- [76] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. C. J. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11679–11689, 2020.
- [77] X. Dong, B. Zhuang, Y. Mao, and L. Liu, "Radar camera fusion via representation learning in autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1672–1681.
- [78] N. Li, C. Lu, X. Yu, X. Liu, and B. Su, "Real-time 3d-lidar, mmw radar and gps/imu fusion based vehicle detection and tracking in unstructured environment," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13339–13345.
- [79] Y. Cheng, H. Xu, and Y. Liu, "Robust small object detection on the water surface through fusion of camera and millimeter wave radar," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15263–15272.
- [80] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "Radiate: A radar dataset for automotive perception in bad weather," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 1–7.
- [81] S. An, Y. Li, and U. Ogras, "Multi-modal 3d human pose estimation using mmwave, RGB-d, and inertial sensors," in *NeurIPS 2022 Workshop on Learning from Time Series for Health*, 2022.
- [82] K. Deng, D. Zhao, Q. Han, S. Wang, Z. Zhang, A. Zhou, and H. Ma, "Geryon: Edge assisted real-time and robust object detection on drones via mmwave radar and camera fusion," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, sep 2022.
- [83] K. Deng, D. Zhao, Q. Han, Z. Zhang, S. Wang, and H. Ma, "Global-local feature enhancement network for robust object detection using mmwave radar and camera," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4708–4712.
- [84] A. Chen, X. Wang, S. Zhu, Y. Li, J. Chen, and Q. Ye, "mmbbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 3501–3510.
- [85] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, A. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 33–40, 2019.
- [86] Y. Wang, T. Gu, T. H. Luan, M. Lyu, and Y. Li, "Heartprint: Exploring a heartbeat-based multiuser authentication with single mmwave radar," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25324–25336, 2022.
- [87] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proceedings of the 26th annual international conference on mobile computing and networking*, 2020, pp. 1–14.
- [88] Y. Wang, T. Gu, T. H. Luan, and Y. Li, "Your breath doesn't lie: Multi-user authentication by sensing respiration using mmwave radar," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2022, pp. 64–72.
- [89] M. Z. Ozturk, C. Wu, B. Wang, M. Wu, and K. R. Liu, "Radio ses: mmwave-based audioradio speech enhancement and separation system," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1333–1347, 2023.
- [90] C. Wang, F. Lin, T. Liu, Z. Liu, Y. Shen, Z. Ba, L. Lu, W. Xu, and K. Ren, "mmphone: Acoustic eavesdropping on loudspeakers via

- mmwave-characterized piezoelectric effect," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 820–829.
- [91] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "mmasl: Environment-independent asl gesture recognition using 60 ghz millimeter-wave signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–30, 2020.
- [92] F. Jin, R. Zhang, A. Sengupta, S. Cao, S. Hariri, N. K. Agarwal, and S. K. Agarwal, "Multiple patients behavior detection in real-time using mmwave radar and deep cnns," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–6.
- [93] T. S. Kiong, S. B. Salem, J. K. S. Paw, K. P. Sankar, and S. Darzi, "Minimum variance distortionless response beamformer with enhanced nulling level control via dynamic mutated artificial immune system," *The Scientific World Journal*, vol. 2014, 2014.
- [94] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 2400–2409.
- [95] G. Li, Z. Zhang, H. Yang, J. Pan, D. Chen, and J. Zhang, "Capturing human pose using mmwave radar," *pervasive computing and communications*, 2020.
- [96] H. Kong, X. Xu, J. Yu, Q. Chen, C. Ma, Y. Chen, Y.-C. Chen, and L. Kong, "m3track: mmwave-based multi-user 3d posture tracking," *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022.
- [97] J. Pegoraro, F. Meneghelli, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2020.
- [98] B. Yang, R. Guo, M. Liang, S. Casas, and R. Urtasun, "Radarnet: Exploiting radar for robust perception of dynamic objects," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 496–512.
- [99] P. Zhao, C. X. Lu, B. Wang, N. Trigoni, and A. Markham, "3d motion capture of an unmodified drone with single-chip millimeter wave radar," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5186–5192.
- [100] C. Zhang, S. Yu, Z. Tian, and J. J. Yu, "Generative adversarial networks: A survey on attack and defense perspective," *ACM Comput. Surv.*, 2023.
- [101] C. Zhang, S. Zhang, J. J. Yu, and S. Yu, "Sam: Query-efficient adversarial attacks against graph neural networks," *ACM Trans. Priv. Secur.*, 2023, just Accepted.
- [102] Y. Duan, N. Chen, S. Shen, P. Zhang, Y. Qu, and S. Yu, "Fdsa-stg: Fully dynamic self-attention spatio-temporal graph networks for intelligent traffic flow prediction," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9250–9260, 2022.
- [103] P. Gong, C. Wang, and L. Zhang, "Mmpoint-gnn: Graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–7.
- [104] C. Wang, P. Gong, and L. Zhang, "Sptointcnn: spatial temporal graph convolutional network for multiple people recognition using millimeter-wave radar," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3433–3437.
- [105] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [106] 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.
- [107] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *ICML*, vol. 2, no. 3, 2021, p. 4.
- [108] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. Martin, and S.-M. Hu, "Pct: Point cloud transformer," *Computational Visual Media*, vol. 7, pp. 187–199, 06 2021.
- [109] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 16 259–16 268.
- [110] A. Chen, X. Wang, K. Shi, S. Zhu, Y. Chen, B. Fang, J. Chen, Y. Huo, and Q. Ye, "Immfusion: Robust mmwave-rgb fusion for 3d human body reconstruction in all weather conditions," 2022.
- [111] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [112] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.
- [113] H. Xue, Q. Cao, Y. Ju, H. Hu, H. Wang, A. Zhang, and L. Su, "M4esh: mmwave-based 3d human mesh construction for multiple subjects," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 391–406.
- [114] C. Lugaesi, J. Tang, H. Nash, C. McClanahan, E. Ubweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *Third Workshop on Computer Vision for AR/VR*, 2019.
- [115] E. Al-Masri and M. Momin, "Detecting heart rate variability using millimeter-wave radar technology," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 5282–5284.
- [116] D. T. Petkie, C. Benton, and E. Bryan, "Millimeter wave radar for remote measurement of vital signs," in *2009 IEEE Radar Conference*. IEEE, 2009, pp. 1–3.
- [117] F. Jin, R. Zhang, A. Sengupta, S. Cao, S. Hariri, N. K. Agarwal, and S. K. Agarwal, "Multiple patients behavior detection in real-time using mmwave radar and deep cnns," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–6.
- [118] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-activity: Accurate and real-time human activity recognition via millimeter wave radar," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8298–8302.
- [119] P. Hu, Y. Ma, P. S. Santhalingam, P. H. Pathak, and X. Cheng, "Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 11–20.
- [120] C. Wang, F. Lin, T. Liu, K. Zheng, Z. Wang, Z. Li, M.-C. Huang, W. Xu, and K. Ren, "mmeve: eavesdropping on smartphone's earpiece via cots mmwave device," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 338–351.
- [121] P. Hu, W. Li, Y. Ma, P. S. Santhalingam, P. Pathak, H. Li, H. Zhang, G. Zhang, X. Cheng, and P. Mohapatra, "Towards unconstrained vocabulary eavesdropping with mmwave radar using gan," *IEEE Transactions on Mobile Computing*, 2022.
- [122] X. Xu, J. Yu, C. Ma, Y. Ren, H. Liu, Y. Zhu, Y.-C. Chen, and F. Tang, "mmeeg: Monitoring human cardiac cycle in driving environments leveraging millimeter wave," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 90–99.
- [123] Y.-H. Wu, Y. Chen, S. Shirmohammadi, and C.-H. Hsu, "Ai-assisted food intake activity recognition using 3d mmwave radars," in *Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management*, 2022, pp. 81–89.
- [124] H. Abedi, A. Ansariyan, P. P. Morita, A. Wong, J. Boger, and G. Shaker, "Ai-powered non-contact in-home gait monitoring and activity recognition system based on mm-wave fmcw radar and cloud computing," *IEEE Internet of Things Journal*, 2023.
- [125] M. Z. Ozturk, C. Wu, B. Wang, and K. R. Liu, "Gaitcube: Deep data cube learning for human recognition with millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 546–557, 2021.
- [126] M. Canil, J. Pegoraro, and M. Rossi, "Millitrace-ir: contact tracing and temperature screening via mmwave and infrared sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 2, pp. 208–223, 2021.
- [127] T. Li, Z. Zhao, Y. Luo, B. Ruan, D. Peng, L. Cheng, and C. Shi, "Gait recognition using spatio-temporal information of 3d point cloud via millimeter wave radar," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [128] Y. Huang, Y. Wang, K. Shi, C. Gu, Y. Fu, C. Zhuo, and Z. Shi, "Hdnet: Hierarchical dynamic network for gait recognition using millimeter-wave radar," *arXiv preprint arXiv:2211.00312*, 2022.
- [129] J. Wu, J. Wang, Q. Gao, M. Pan, and H. Zhang, "Path-independent device-free gait recognition using mmwave signals," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11 582–11 592, 2021.
- [130] T. Li, X. Cao, H. Liu, C. Shi, and P. Chen, "Mtgcgt: Multi-person gait recognition with spatio-temporal information via millimeter wave radar," in *2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2021, pp. 660–666.

- [131] Y. Shi, L. Du, X. Chen, X. Liao, Z. Yu, Z. Li, C. Wang, and S. Xue, "Robust gait recognition based on deep cnns with camera and radar sensor fusion," *IEEE Internet of Things Journal*, 2023.
- [132] L. Senigagliesi, G. Ciattaglia, and E. Gambi, "Contactless walking recognition based on mmwave radar," in *2020 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2020, pp. 1–4.
- [133] Z. Xia, Y. Luomei, C. Zhou, and F. Xu, "Multidimensional feature representation and learning for robust hand-gesture recognition on commercial millimeter-wave radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 4749–4764, 2020.
- [134] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. R. Liu, "mmwrite: passive handwriting tracking using a single millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 291–13 305, 2021.
- [135] M. Arsalan, A. Santra, and V. Issakov, "Radarsnn: A resource efficient gesture sensing system based on mm-wave radar," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 4, pp. 2451–2461, 2022.
- [136] 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*, vol. 213, p. 119042, 2023.
- [137] Y. Li, D. Zhang, J. Chen, J. Wan, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, "Di-gesture: Domain-independent and real-time gesture recognition with millimeter-wave signals," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 5007–5012.
- [138] J. Lien, N. Gillian, M. E. Karagozler, A. P. M. C. Schwesig, E. M. Olson, H. K. B. Raja, and I. Poupyrev, "Soli: ubiquitous gesture sensing with millimeter wave radar," *international conference on computer graphics and interactive techniques*, 2016.
- [139] Y. Liu, S. Zhang, M. Gowda, and S. Nelakuditi, "Leveraging the properties of mmwave signals for 3d finger motion tracking for interactive iot applications," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 6, no. 3, pp. 1–28, 2022.
- [140] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, "M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3397–3415, 2021.
- [141] Y. Hu, B. Wang, C. Wu, and K. R. Liu, "mmkey: Universal virtual keyboard using a single millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 510–524, 2021.
- [142] H. Wei, Z. Li, A. D. Galvan, Z. Su, X. Zhang, K. Pahlavan, and E. T. Solovey, "Indexpen: Two-finger text input with millimeter-wave radar," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–39, 2022.
- [143] J. Guo, M. Jin, Y. He, W. Wang, and Y. Liu, "Dancing waltz with ghosts: Measuring sub-mm-level 2d rotor orbit with a single mmwave radar," in *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*, 2021, pp. 77–92.
- [144] J. W. Smith, O. Furxhi, and M. Torlak, "An fcnn-based super-resolution mmwave radar framework for contactless musical instrument interface," *IEEE Transactions on Multimedia*, vol. 24, pp. 2315–2328, 2021.
- [145] Z. Zhang, X. Wang, D. Huang, X. Fang, M. Zhou, and Y. Zhang, "Mrpt: Millimeter-wave radar-based pedestrian trajectory tracking for autonomous urban driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–17, 2021.
- [146] J. Wu, H. Cui, and N. Dahnoun, "A novel high performance human detection, tracking and alarm system based on millimeter-wave radar," in *2021 10th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 2021, pp. 1–4.
- [147] A. Shastri, M. Canil, J. Pegoraro, P. Casari, and M. Rossi, "Mmscale: self-calibration of mmwave radar networks from human movement trajectories," in *2022 IEEE Radar Conference (RadarConf22)*. IEEE, 2022, pp. 1–6.
- [148] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "Human tracking and identification through a millimeter wave radar," *Ad Hoc Networks*, vol. 116, p. 102475, 2021.
- [149] J. Pegoraro, F. Meneghelli, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, pp. 2994–3009, 2021.
- [150] J. Pegoraro, D. Solimini, F. Matteo, E. Bashirov, F. Meneghelli, and M. Rossi, "Deep learning for accurate indoor human tracking with a mm-wave radar," in *2020 IEEE Radar Conference (RadarConf20)*. IEEE, 2020, pp. 1–6.
- [151] C. Wang, P. Gong, and L. Zhang, "Stpointcnn: spatial temporal graph convolutional network for multiple people recognition using millimeter-wave radar," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3433–3437.
- [152] T. Zhou, K. Jiang, S. Wang, Y. Shi, M. Yang, W. Ren, and D. Yang, "3d multiple object tracking with multi-modal fusion of low-cost sensors for autonomous driving," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 1750–1757.
- [153] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: Radar object detection using cross-modal supervision," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 504–513.
- [154] K. Deng, D. Zhao, Q. Han, Z. Zhang, S. Wang, and H. Ma, "Global-local feature enhancement network for robust object detection using mmwave radar and camera," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4708–4712, 2022.
- [155] A. Blanco, P. J. Mateo, F. Gringoli, and J. Widmer, "Augmenting mmwave localization accuracy through sub-6 ghz on off-the-shelf devices," in *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022, pp. 477–490.
- [156] P. Gao, S. Zhang, W. Wang, and C. X. Lu, "Dc-loc: Accurate automotive radar based metric localization with explicit doppler compensation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4128–4134.
- [157] A. Sengupta and S. Cao, "mmpose-nlp: A natural language processing approach to precise skeletal pose estimation using mmwave radars," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [158] S. An and U. Y. Ogras, "Fast and scalable human pose estimation using mmwave point cloud," *design automation conference*, 2022.
- [159] Y. Sun, Z. Huang, H. Zhang, Z. Cao, and D. Xu, "3drimr: 3d reconstruction and imaging via mmwave radar based on deep learning," in *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*. IEEE, 2021, pp. 1–8.
- [160] Y. Kim, S. Kim, J. W. Choi, and D. Kum, "Craft: Camera-radar 3d object detection with spatio-contextual fusion transformer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp. 1160–1168.
- [161] J.-J. Hwang, H. Kretschmar, J. Manela, S. Rafferty, N. Armstrong-Crews, T. Chen, and D. Anguelov, "Cramnet: Camera-radar fusion with ray-constrained cross-attention for robust 3d object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 388–405.
- [162] H. Ding, Z. Chen, C. Zhao, F. Wang, G. Wang, W. Xi, and J. Zhao, "Minmesh: 3d human mesh construction by fusing image and millimeter wave," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–24, 2023.
- [163] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 211–220.
- [164] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 289–304.
- [165] P. Kumari, A. Mezghani, and R. W. Heath, "A low-resolution adc proof-of-concept development for a fully-digital millimeter-wave joint communication-radar," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 8619–8623.
- [166] K. V. Mishra, M. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 100–114, 2019.
- [167] G. Wu, X. Chen, Z. Gao, H. Zhang, S. Yu, and S. Shen, "Privacy-preserving offloading scheme in multi-access mobile edge computing based on madrl," *Journal of Parallel and Distributed Computing*, vol. 183, p. 104775, 2024.
- [168] R. Ding, H. Jin, and D. Shen, "Rotation speed sensing with mmwave radar," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, 2023, pp. 1–10.

- [169] Y. Xie, R. Jiang, X. Guo, Y. Wang, J. Cheng, and Y. Chen, "Universal targeted adversarial attacks against mmwave-based human activity recognition," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, 2023, pp. 1–10.
- [170] R. Reddy Vennam, I. K. Jain, K. Bansal, J. Orozco, P. Shukla, A. Ranganathan, and D. Bharadvia, "mmspoof: Resilient spoofing of automotive millimeter-wave radars using reflect array," in *2023 IEEE Symposium on Security and Privacy (SP)*, 2023, pp. 1807–1821.
- [171] S. Basak and M. Gowda, "mmspy: Spying phone calls using mmwave radars," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 1211–1228.
- [172] Z. Li, B. Chen, X. Chen, H. Li, C. Xu, F. Lin, C. Lu, K. Ren, and W. Xu, "Spiralspy: Exploring a stealthy and practical covert channel to attack air-gapped computing devices via mmwave sensing," in *Network and Distributed Systems Security (NDSS) Symposium 2022*. The Internet Society, Apr. 2022.
- [173] J. Shenoy, Z. Liu, B. Tao, Z. Kabelac, and D. Vasisht, "Rf-protect: privacy against device-free human tracking," in *Proceedings of the ACM SIGCOMM 2022 Conference*, 2022, pp. 588–600.
- [174] K. Sood, M. R. Nosouhi, N. Kumar, A. Gaddam, B. Feng, and S. Yu, "Accurate detection of iot sensor behaviors in legitimate, faulty and compromised scenarios," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 288–300, 2023.
- [175] M. Toth, P. Meissner, A. Melzer, and K. Witrisal, "Performance comparison of mutual automotive radar interference mitigation algorithms," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–6.
- [176] J. Bechter, F. Roos, M. Rahman, and C. Waldschmidt, "Automotive radar interference mitigation using a sparse sampling approach," in *2017 European Radar Conference (EURAD)*. IEEE, 2017, pp. 90–93.
- [177] F. Jin and S. Cao, "Automotive radar interference mitigation using adaptive noise canceller," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3747–3754, 2019.
- [178] Z. Xu and M. Yuan, "An interference mitigation technique for automotive millimeter wave radars in the tunable q-factor wavelet transform domain," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 12, pp. 5270–5283, 2021.
- [179] F. Uysal and S. Sanka, "Mitigation of automotive radar interference," in *2018 IEEE Radar Conference (RadarConf18)*. IEEE, 2018, pp. 0405–0410.
- [180] S. Neemat, O. Krasnov, and A. Yarovoy, "An interference mitigation technique for fmcw radar using beat-frequencies interpolation in the stft domain," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 3, pp. 1207–1220, 2018.
- [181] Z. Xu, S. Xue, and Y. Wang, "Incoherent interference detection and mitigation for millimeter-wave fmcw radars," *Remote Sensing*, vol. 14, no. 19, p. 4817, 2022.
- [182] Z. Xu, "Bi-level  $l_1$  optimization-based interference reduction for millimeter wave radars," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [183] J. Jung, S. Lim, J. Kim, S.-C. Kim, and S. Lee, "Interference suppression and signal restoration using kalman filter in automotive radar systems," in *2020 IEEE International Radar Conference (RADAR)*. IEEE, 2020, pp. 726–731.
- [184] J. Wu, S. Yang, W. Lu, and Z. Liu, "Iterative modified threshold method based on emd for interference suppression in fmcw radars," *IET Radar, Sonar & Navigation*, vol. 14, no. 8, pp. 1219–1228, 2020.
- [185] S. Lee, J.-Y. Lee, and S.-C. Kim, "Mutual interference suppression using wavelet denoising in automotive fmcw radar systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 887–897, 2019.
- [186] J. Bechter, C. Sippel, and C. Waldschmidt, "Bats-inspired frequency hopping for mitigation of interference between automotive radars," in *2016 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*. IEEE, 2016, pp. 1–4.
- [187] G. Hakobyan, K. Armanious, and B. Yang, "Interference-aware cognitive radar: A remedy to the automotive interference problem," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 2326–2339, 2019.
- [188] S. Alland, W. Stark, M. Ali, and M. Hegde, "Interference in automotive radar systems: Characteristics, mitigation techniques, and current and future research," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 45–59, 2019.
- [189] F. Uysal, "Phase-coded fmcw automotive radar: System design and interference mitigation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 270–281, 2019.
- [190] M. Zhang, S. He, C. Yang, J. Chen, and J. Zhang, "Vanet-assisted interference mitigation for millimeter-wave automotive radar sensors," *IEEE Network*, vol. 34, no. 2, pp. 238–245, 2020.
- [191] C. Aydogdu, M. F. Keskin, N. Garcia, H. Wymeersch, and D. W. Bliss, "Radchat: Spectrum sharing for automotive radar interference mitigation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 416–429, 2019.
- [192] R. Singh, D. Saluja, and S. Kumar, "Spread spectrum coded radar for r2r interference mitigation in autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [193] C. Aydogdu, M. F. Keskin, G. K. Carvajal, O. Eriksson, H. Hellsten, H. Herbertsson, E. Nilsson, M. Rydstrom, K. Vanas, and H. Wymeersch, "Radar interference mitigation for automated driving: Exploring proactive strategies," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 72–84, 2020.
- [194] S. Jin and S. Roy, "Fmcw radar network: Multiple access and interference mitigation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 968–979, 2021.
- [195] C. Aydogdu, N. Garcia, L. Hammarstrand, and H. Wymeersch, "Radar communications for combating mutual interference of fmcw radars," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–6.
- [196] Y. He, L. Ma, Z. Jiang, Y. Tang, and G. Xing, "Vi-eye: semantic-based 3d point cloud registration for infrastructure-assisted autonomous driving," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 573–586.
- [197] Y. Long, D. D. Morris, X. Liu, M. Castro, P. Chakravarty, and P. Narayanan, "Radar-camera pixel depth association for depth completion," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12 502–12 511, 2021.
- [198] Y. Wang, G. Wang, H.-M. Hsu, H. Liu, and J.-N. Hwang, "Rethinking of radar's role: A camera-radar dataset and systematic annotator via coordinate alignment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2815–2824.
- [199] M. Meyer and G. Kuschk, "Deep learning based 3d object detection for automotive radar and camera," *2019 16th European Radar Conference (EuRAD)*, pp. 133–136, 2019.
- [200] T. Liu, M. Gao, F. Lin, C. Wang, Z. Ba, J. Han, W. Xu, and K. Ren, "Wavoice: A noise-resistant multi-modal speech recognition system fusing mmwave and audio signals," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 97–110.
- [201] X. Zhang, Z. Li, and J. Zhang, "Synthesized millimeter-waves for human motion sensing," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 377–390.



**Shuai Wang** is a young chief professor in the School of Computer Science and Engineering at Southeast University. He has been serving as the head of the computer engineering department since Nov. 2022. He received his Ph.D. degree in the Department of Computer Science and Engineering at the University of Minnesota, Twin Cities in March 2017, under the guidance of Prof. Tian He. His research interests include artificial intelligence, data analytics, Internet of Things (IoT), cyber-physical systems, wireless networks and sensors

and related areas.



**Zhimeng Yin** is an Assistant Professor at the Department of Computer Science at the City University of Hong Kong. He obtained his Ph.D. degree from the University of Minnesota (supervised by Prof. Tian He) in 2020. Before that, I received my bachelor's degree and master's degree (advised by Prof. Hongbo Jiang) from Huazhong University of Science and Technology in 2011 and 2014.



**Luoyu Mei** received the B.S. degrees from Southeast University, China. He is currently pursuing the Ph.D. degree of the joint program with the School of Computer Science and Engineering, Southeast University, and the Department of Computer Science, City University of Hong Kong. His research interests include the Wireless Sensing, Internet of Things, Wireless Network and mobile systems.



**Xianjun Deng** (Senior Member, IEEE) received the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2014. He is currently a Professor with the School of Cyber Science and Engineering, Huazhong University of Science and Technology. He has published more than 50 technical papers in international top journals and conferences. His research includes security and reliability, reinforcement learning-enabled networking, coverage optimization, and Internet of Things. Prof. Deng obtained the IEEE TCSC Award.



**Ruofeng Liu** received the B.S. and M.S. degrees from Northwestern Polytechnical University, China. He graduated with a Ph.D. degree in Computer Science at the University of Minnesota, advised by Dr. Tian He. He is currently a research scientist at Bosch Sunnyvale, CA. His research interests include the Internet of Things, Wireless Network, backscatter systems, and mobile systems.



**Tian He** (Fellow, IEEE) received the Ph.D. degree in engineering from the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, in 2008. He is currently a Full Professor with the Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA. He is the author and coauthor of over 280 papers in premier network journals and conferences with over 22 000 citations (H-Index 65). His research interests include wireless networks, networked sensing systems, cyber-physical systems, Internet of Things, and distributed systems in general. Dr. He is the recipient of the NSF CAREER Award in 2009, the McKnight Land-Grant Chaired Professorship in 2011, the George W. Taylor Distinguished Research Award in 2015, the China NSF Outstanding Overseas Young Researcher I and II in 2012 and 2016, and eight best paper awards in international conferences, including MobiCom, SenSys, and ICDCS. He has served as a few general/program chair positions in international conferences and on many program committees and also has been an Editorial Board Member for six international journals, including ACM Transactions on Sensor Networks, IEEE Transactions on Computers, and IEEE/ACM Transactions on Networking. He is a Fellow of ACM.



**Wenchao Jiang** has been an Assistant Professor in the pillar of ISTD, SUTD (Singapore) since September 2019. He received his PhD degree from the Department of Computer Science, at the University of Minnesota, Twin Cities in August 2019 under the supervision of Prof. Tian He, ACM/IEEE fellow. His thesis focuses on building direct communication between COTS heterogeneous wireless technologies in the PHY layer and also explores the extension and applications of such technology in the whole network stack.

His research interest also expands broadly on the Internet of Things (IoT), wireless communication, embedded system, low-power sensor networks, mobile computing, and localisation. He has published in top conferences and journals in wireless networking (MobiCom, SenSys, INFOCOM, SigComm, NSDI, TON, TMC, TPDS).