mmCTD: Concealed Threat Detection for Cruise Ships via mmWave radar

Danei Gong, Kezhong Liu, Xuming Zeng, Shengkai Zhang, Mozi Chen, Kai Zheng, and Dashuai Pei

***Abstract*—The safeguarding of critical zones aboard a marine vehicle, such as the engine room, wheelhouse, and pump room, assumes crucial significance while navigating through the open sea. Despite the existing pre-boarding security measures, Concealed Threat Detection (CTD) systems have emerged as a pressing need to prevent the ship from post-boarding damage with concealed dangers. Due to concerns regarding deployment cost and privacy, mmWave-based CTD systems have received significant attention. However, current solutions are not easily adapted to work in ships because of the large number of ghost targets resulting from multipath reflections in full metal cabins. To address these challenges, this paper proposes a new CTD system, called mmCTD, which utilizes two mmWave commercial radars. The proposed system addresses the multipath challenge by unifying multi-view perceptions with two distinct designs. First, we propose a ghost-point elimination algorithm that extracts the point clouds from real objects. Then, we design a multi-view domain adversarial framework to predict concealed threats in the human body using the extracted RF features. Our experimental results demonstrate that the recognition accuracy of mmCTD in three scenarios reaches 89% with a low false alarm rate.**

***Index Terms*—mmWave radar, Concealed threat detection, Ghost target, Cruise ship.**

I. Introduction

THE cruise ship possesses a substantial passenger capacity and a diversity personnel composition. It's a significant threat for navigation safety when somebody carrying hazardous objects in crowds ares. Although the pre-boarding security inspection system is well-established, the security check during the voyage receives insufficient attention. The attack of intruder on the functional compartments, such as the wheelhouse, engine room, etc. may lead to serious consequences during the voyage [1]. Thus, it is necessary to implement ubiquitous and automated non-contact detection systems for human-concealed threat items.

Many promising methods have been used for automatic CTD, such as frequency bands in the microwave, mmWave, terahertz electromagnetic wave (EMW) devices, X-ray, etc. EMW solutions are achieved by high-resolution scanning imaging or analysis of the reflected signal to distinguish material. However, these solutions require large-scale antenna arrays and special equipment that require significant cost and space for widespread deployment. Infrared solutions, which image temperature differences between targets, encounter scalability challenges in real-world scenarios due to factors such as environmental variations and detection distance [2]. Recently, emerging wireless sensing technologies have received significant attention as they are insensitive to light and easy to deploy [3]. For example, recent research based on WiFi signals detects hidden metallic objects through reflected energy or the channel state information (CSI) of signals [4], [5]. Nevertheless, accurately distinguishing multiple users using CSI is challenging in dynamic environments.

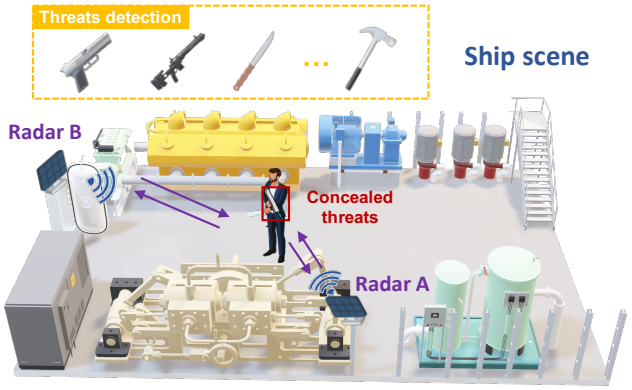


Fig. 1. Our proposed mmCTD uses mmWave radars deployed on the ship scene to implement concealed threat detection. mmCTD's multi-view deployment can achieve full coverage of the human body and overcome the interference caused by reflectors inside the ship for accurate target detection.

Single-chip mmWave radar presents new opportunities for addressing the issues mentioned above, given its stronger directionality and detection resolution compared to low-frequency signal technologies, as well as the advancements in integration. Additionally, it has the capability to penetrate thinner materials such as fabrics and paper shells. However, existing methods based on single-chip mmWave radar concealed objects detection are not well-suited for the dynamic and confined environments inside ship cabins. Methods [6] map the problem of weapon detection to gait signature recognition, identifying weapon bearers through micro-Doppler(MD) signatures. However, this approach is less effective when items are small and features are intentionally hidden. Another approach combines low-cost mmWave radar and MIMO SAR

This work was supported in part by the National Natural Science Foundation of China under Grant 51979216, in part by the Natural Science Foundation of Hubei Province, China, under Grant 2021CFA001. (Corresponding author: Xuming Zeng.)

Danei Gong, Xuming Zeng, Mozi Chen, Kai Zheng, and Dashuai Pei are with the School of Navigation, Wuhan University of Technology, Wuhan 430063, China (e-mail: dank@whut.edu.cn; zengxvming@whut.edu.cn; chenmz@whut.edu.cn; kzheng@whut.edu.cn; 291623@whut.edu.cn).

Kezhong Liu is with the Hubei Key Laboratory of Inland Shipping Technology and the School of Navigation, Wuhan University of Technology, Wuhan 430063, China (e-mail: kzliu@whut.edu.cn).

Shengkai Zhang is with the School of Information Engineering, Wuhan University of Technology, Wuhan 430063, China (e-mail: shengkai@whut.edu.cn).

technology to create a two-dimensional imaging system [7], but it is typically only suitable for near-field imaging and mechanical sliding scanning can be time-consuming [8]. The RF feature-based method utilizes the generated range-azimuth-elevation image cube and a deep learning prediction network to achieve object classification [9].

In this paper, we propose a 77 GHz COTS mmWave radar-based concealed threat detection system, called mmCTD, which is suitable for use in the restricted areas of cruise ships, as depicted in Fig. 1. mmCTD has three key features to achieve our objectives: 1) Automatic detection of concealed threat carried by ship personnel, covering various sides of the human body; 2) Accurate target detection achieved by overcoming the interference introduced in RF maps by strong reflectors such as metal bulkheads and machinery in ships; and 3) A robust feature representation from the RF map is constructed, taking into account the motion speed of the target and the impact of different scenes on the feature.

To realize mmCTD system, several practical challenges must be addressed. Firstly, the presence of metal walls, machinery, and other complex structures within ship cabins can result in severe multipath propagation and multiple-order reflections for the radar signal, leading to random false alarms ("ghost targets") in the radar RF features [10] that can interfere with the accurate extraction of target features [11]. Secondly, the field of view (FOV) of a single radar is often not sufficient to cover the entire human body, which can result in missed detections. Thus, multiple view radars must be combined for detection, which will be another mode. Lastly, variations in distance that pedestrian to radar and different ship scenes can cause changes in the RF features, which can also impact detection robustness. Thus, these factors need to be considered to improve the system's overall performance.

To tackle the above challenges, the mmCTD framework comprises three main components: data collection and pre-processing, target detection and ghost target removal, and a multi-view detection model. To be more specific, two radar views are used to scan the forward and lateral directions of the human body and generate the RF maps from the collected raw data. To determine the location of the pedestrian and eliminate ghost targets from the RF maps, we propose a coordinate transformation and correlation calculation method. Secondly, we propose a metric that correlates the velocity with the radar frames and crops the feature rectangle block (TRB) from the RF maps. Finally, we introduce a multi-view domain adversarial deep learning framework that takes three input from two views radars to predict concealed threats carried by people in various scenes such as cabin, wheelhouse, aisle, etc.

We have conducted extensive experiments in a real-world ship and established mmWave radar dataset containing multiple pedestrians with different threats in ship scenarios. We further evaluated the performance of our proposed mmCTD under diverse factors. Our findings indicate that mmCTD can accurately distinguish moving targets from high levels of noise and numerous strong reflections, demonstrating its efficacy in the ship.

In summary, three main novel contributions are presented in this study:

- We propose a target detection method associating multiple view radars to overcome the interference introduced in RF maps by strong reflectors in ship scenes and to solve the CTD problem covering all sides of the human body.
- We design an RF feature processing method that considers the distance and velocity variations of the target, and propose a novel multi-view deep learning solution. Multi-view feature representations of the target are extracted to achieve environment-independent detection.
- We perform an experimental evaluation and analysis of some influencing factors. Extensive experiments formalized the effectiveness of the system on ships.

II. Related Work

The detection of concealed threats requires the capability to identify dangerous metal weapons or other offensive items that may be hidden underneath a person's clothing. In this section, we provide a succinct overview of several relevant studies that are primarily focused on CTD.

A. CTD Using High-resolution EMW Device

During the past few decades, traditional concealed weapons detection with sufficient penetration usually requires various types of sensors operating in the microwave (MW), mmWave, and terahertz bands. They are based on the capability to penetrate fabric materials and are mainly classified into two categories- imaging [12], [13] and non-imaging [14]. The non-imaging system only needs to detect the echo characteristics reflected by the concealed weapons and does not need the image of itself [15]. Imaging systems are further classified as active [16]and passive [17]depending on their illumination ways [18]. Active imaging can be achieved by mechanically moving raster scans and synthetic apertures radar techniques. Passive imaging is achieved by analyzing the thermal radiation emitted by the scanning target, which takes a long time [19]. However, these imaging systems achieve high-resolution imaging through large-scale antenna apertures and high transmit power [20]. They are not suitable for ubiquitous placement in narrow scenes due to their large size. Some scanning-based detection methods require active cooperation and high time costs, which are unnecessary for real-time threat detection.

B. CTD Using IR camera

Cameras are commonly used for vision-based automatic weapon detection systems. Hand-held weapon detection is achieved by human pose estimation or steel object detection [21]. Since the vision-surveillance is not enough to provide reliable performance in CTD applications, image fusion becomes a viable technology to achieve CTD [22]. For instance, the fusion image methods combine visual and infrared (IR) images [23] from multi-camera systems, which come with a complete range of optics and achieve the detection and classification of concealed threats. The IR and THz/mmWave imager solutions have recently attracted much attention. Their most advantageous properties for CTD applications are the

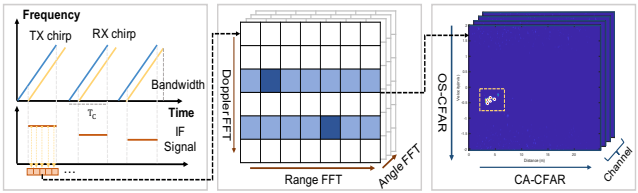
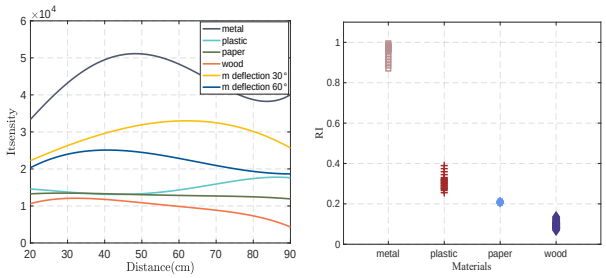
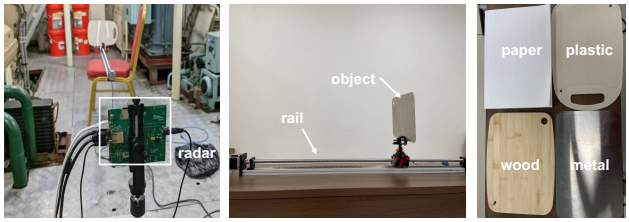


Fig. 2. Illustration of FMCW radar RF maps generating and object detection from raw signals. (a)FMCW Radar Signals. (b)Range-Doppler-Azimuth Cube. (c)Peak Detection.



(a) The intensity fitting curves (b) The Reflection itensity

Fig. 3. Reflection features of four common materials moving on slide rails.



(a) Experiment in engine room (b) Experiment in common room (c) Testing objects

Fig. 4. Experiment scenario.

ability to detect minute temperature differences on the surface of objects furthermore the ability to penetrate clothing [24]. Federico *et al.* [25] propose a system based on the active, passive mmWave, and IR camera and using IR sensors for noise-removing with image fusion techniques. Kowalski *et al.* [26] present comparative studies of imaging possibilities of the objects hidden under typical textiles using state-of-the-art THz (250 GHz) and Mid-Wavelength IR (3-8μm) devices and analyze both the power relationship between spectral bands and the thermal radiation model. Image fusion is an effective method for CTD, but the fusion of images captured by different devices may increase the cost. In real-world scenarios, the scalability of threat detection systems based on infrared cameras is limited by various factors, including differences in environments and detection distances. These limitations are particularly pronounced in scenarios characterized by high temperatures and low visibility.

C. CTD Using Commercial Single-Chip Radar

The application of miniaturized, lightweight, and low-cost mmWave radar sensors is gradually expanding; Single COTS mmWave radar-based human sensing solutions have attracted considerable attention recently. Existing studies propose to use the MD signatures [27] retrieved from the 77GHz FMCW radar to identify people carrying a concealed rifle based on their gait and motion characteristics, then, Omid *et al.*design a framework that analyzes MD signatures for individual anomaly detection among multiple individuals [28].However, the method based on MD signatures may not be effective when individuals deliberately conceal their posture or carry small objects. Additionally, Gao *et al.* [9] propose a carried object detection system by applying the range-azimuth-elevation (RAE) imaging results of COTS mmWave radar, which can detect three types of items hidden or open in real-time. Yanik *et al.* [7] present a two-dimensional imaging system based on the combination of low-cost FMCW mmWave radars and MIMO synthetic aperture radar (SAR) processing techniques that can image the target concealed in a paper envelope and a bag. However, SAR imaging methods have the disadvantage of requiring a long processing time and necessitating that the target be in close proximity to the radar.

III. Preliminary And Observation

Typically, different materials reflect RF signals in distinct ways. By analyzing radar echoes' features, it is possible to determine if people are carrying concealed threats. In this section, we describe how range, velocity, and angle of arrival can be measured to capture reflection signatures when performing CTD with mmWave radar sensors. At the same time, we propose a reflection intensity signature (RIF) related to the target material. Next, we describe some experiments and observations illustrating the challenges encountered in implementing mmCTD.

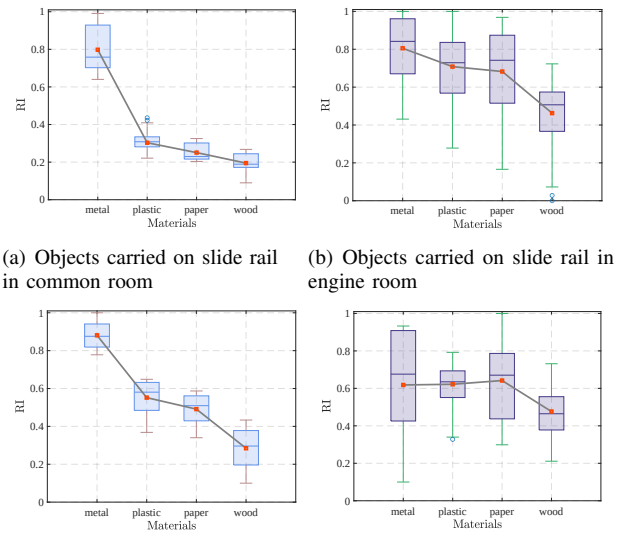
A. mmWave Radar Signal Model

The transmission signal of an FMCW radar is a continuous wave that is modulated with a linearly changing frequency over time. The RX signal and TX signal are mixed to generate an intermediate frequency (IF) signal. By performing signal processing and frequency domain analysis, the target can be estimated in terms of distance, speed, and angle. An example of the signal processing flow is illustrated in Fig. 2.

Range-Doppler-Azimuth estimation: To estimate the range-Doppler-azimuth of FMCW radar data, it is typically to use a 3D-FFT procedure. This involves performing a FFT on each dimension of the data separately, in a specific order. First, a FFT is applied to each range bin (called range-FFT) to obtain the range information of the target. Next, the Doppler dimension is varied by performing a FFT on the slow time dimension of the radar (called Doppler-FFT). Finally, the azimuth dimension is transformed by applying an FFT to each antenna channel dimension (called angle-FFT). The above process can be expressed as:

$$S_{\text{rva}} = \mathcal{F}_{\text{range}}\left\{\mathcal{F}_{\text{Doppler}}\left\{\mathcal{F}_{\text{angle}}\left\{S(n,k,m)\right\}\right\}\right\} \quad (1)$$

where $S(n,k,m)$ is the collected three-dimensional matrix data, with n, k, and m denote the sampling points, chirps, virtual antennas, respectively. The output of the 3D-FFT is



(a) Objects carried on slide rail in common room

(b) Objects carried on slide rail in engine room

(c) Objects carried by people in common room

(d) Objects carried by people in engine room

Fig. 5. Boxplots of RI for four objects under four situations.

denoted by S_{rva}, which provides an estimate of the range-Doppler azimuth.

B. Reflection Itensity Feature

The feasibility of material discrimination of objects based on radar signals can be verified by extracting features from the RF map. To this end, we propose an object reflection intensity feature, denoted as ρ. First, we perform static elimination to remove static objects in the Range-Doppler matrix (RDM). Next, we perform noncoherent accumulation of RDM of all transceiver antenna pairs to increase the signal-to-noise ratio of the obtained RDM maps. This adaptive algorithm is commonly used in radar target detection and adjusts the sensitivity according to the detection scene. After preprocessing, we apply CA-CFAR and OS-CFAR along the Doppler and range dimensions, respectively, and obtain the distance, velocity, and intensity information of the targets through peak detection. The target's reflection intensity I_j in the j-th frame can be represented as follows:

$$I_j = \sum_{i=1}^{N_p} A_i^2, \tag{2}$$

where N_P is the number of peaks in a frame and A_i is amplitude of the peak point. According to the electromagnetic wave transmission formula, the power captured at the RX antenna is modeled as:

$$P_r = \frac{P_t G_{Tx} G_{Rx} \lambda^2 \sigma}{(4\pi)^3 d^4} \propto \frac{\sigma}{d^4}, \tag{3}$$

where P_t denotes the output power if radar, G_{Tx} and G_{Rx} indicate the TX and RX antenna power gain, respectively, and λ is the wavelength, σ is the radar cross section of the target (RCS), which is related to the target's shape, size, material and incident angle.

We conduct the distance-intensity curve C_d based on the amplitude values in each frame using a polynomial fitting, as shown in Fig. 3. In particular, the C_d is related to P_r. According to the intensity curve and above model, the intensity increases as the object moves closer to the radar. So to remove the effect of propagation distance, we get the RIF feature that is only related to the reflection intensity of the object:

$$\rho = \gamma \cdot d^4 \cdot C_d, \tag{4}$$

where d represents the distance from the target to the radar obtained in each frame and γ is a parameter we set empirically. ρ is a function of the target reflection feature ,independent of the propagation distance, and expresses the material information of the object. By derivation, as long as we obtain the range d and the corresponding target strength C_d, we can estimate the ρ to judge the properties of the target.

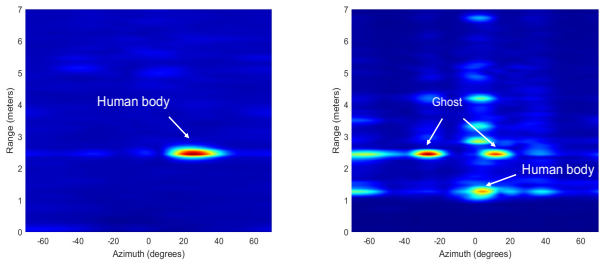
C. Experiments and Practical Challenge

In this section, we describe some experiments and observations to illustrate practical challenges when implementing mmCTD in real-world scenario.

We conducted experiments using an mmWave radar in both a standard office and a ship's cabin setting. Four similarly sized objects (paper, sheet metal, wood, and plastic, approximately 20cm $\times$ 10cm) were placed on stepper motor slides and moved towards the radar to collect signals. The experimental setup is depicted in Fig. 4. To simulate real-world scenarios where people carry concealed objects, we also conducted experiments where individuals carried different targets to the radar. The results of these experiments are shown in Fig. 5. Our findings demonstrate that in typical indoor scenes, objects made of metal and other materials can be easily distinguished. However, in the ship's cabin scene, the distribution of the reflection intensity is less clear. Whether it is the sliding rails or walking people, the upper and lower quartiles of the four objects show significant deviations.

Measured intensity varies with the object deflects: Similar to light, a wireless signal's reflection angle varies with the plane's deflection angle and the incident signal's direction. We place the object at different deflection angles to observe the change in its reflection intensity. As the deflection angle of the object increases, its RI value also decreases.In a real-world scenario, people with dangerous items can hide them around their bodies: front, back, and sides, and move freely. Therefore, the system implementation must be able to cover multiple directions, considering that the main application scenario is the ship cabin. We propose a dual-radar vertical view layout scheme, which can realize the front and side detection of the human torso and can easily be extended to more side detection. In practice, when a person walks in a particular direction, there are different patterns in the RF features of the two radars. We can achieve multi-side detection tasks by jointly estimating the feature information in the RF map.

Ghost targets in target detection: Unlike the characteristics of common indoor environments, wireless signals are susceptible to high noise interference in steel-enclosed spaces such as ship cabins [29]. On the one hand, the ship



(a) (b)

Fig. 6. comparison of RA maps in two scenarios. (a) RA map in ordinary rooms (b) RA map in ship engine room.

environment is time-varying due to the loads and external pressure. The ship's deformed hull, floor, etc. will cause the reflected signal's phase transition and frequency shift. On the other hand, there are metal walls, turbines, etc. Because metals reflect wireless signals more strongly, they experience higher-order and specular reflections in the cabin [30], and This challenges our radar-based feature detection. We deploy millimeter-wave radar inside the real-world ship cabin (engine room, wheelhouse access). Specifically, when we perform radar RF map feature extraction, the RA map after static object filtering is shown in Fig. 6, which contains multiple random virtual target clusters other than moving target objects, resulting in inaccurate estimates. Our observation on multi-view radars revealed that ghost targets exhibit independence in space, whereas real targets consistently manifest in similar positions across multi-view points.

IV. System Overview

Our study aims to implement a multi-view human concealed threat object carry and recognition system for ship cabins. The main architecture of mmCTD includes three components: data collection and preprocessing, target detection and ghost removal, and a concealed threat detection model, entitled MVCTDNet, combining DCNN and Multi-View Learning, as shown in Fig. 7.

1) *Data Pre-processing:* The preprocessing model is responsible for collecting the radar's raw data, performing static removal and background suppression on the data to reduce static clutter in the environment. The data is then generated into radar RF images, and the effect of distance on amplitude is removed from the imaging results for each frame.
2) *Ghost Removal and Feature Extraction:* We propose a random ghost removal method based on vertical viewing angle distribution dual radar to detect and localize pedestrian subjects in RF images automatically. A feature clipping algorithm based on the correlation of two radar coordinates is introduced to obtain the target object's features and reduce the model input's size.
3) *Multi-View Concealed Threat Detection Model:* Finally, to extract the human target features from the dual radar RF map and realize the detection of people concealed threatening objects. We propose a multi-view deep learning framework named MVCTDNet, which not only integrates data collected from different view radars but also removes specific information in each domain and enables detection across scenes and humans.

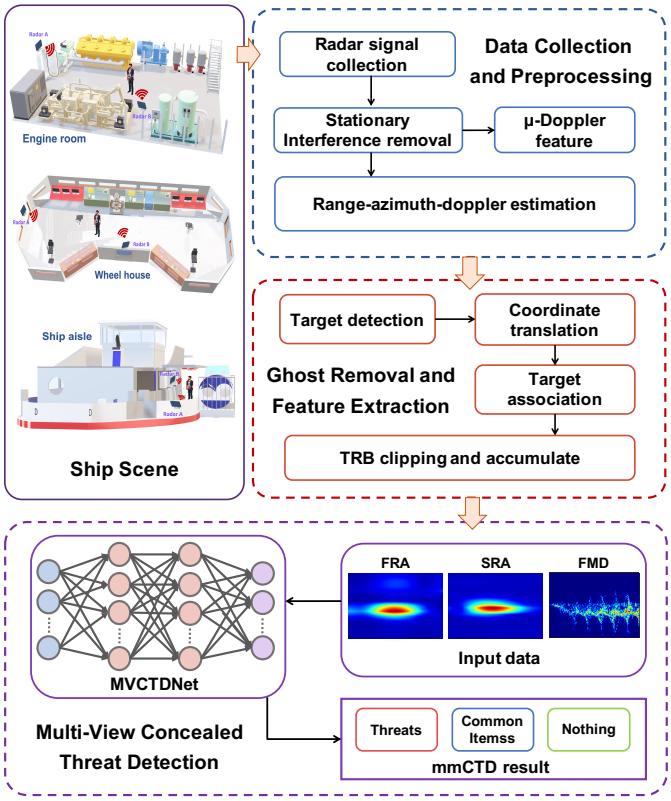


Fig. 7. Overview of mmCTD architecture.

V. Data Preprocessing And Feature Extraction

In this section, we introduce a random ghost elimination method in the target detection of the ship's room. First, we preprocess the collected radar data. Secondly, performing feature extraction and target detection in RF maps. Finally, we propose a ghost elimination algorithm based on a dual-radar coordinate correlation to complete the characteristic clipping.

A. Data Preprocessing

Stationary Interference Removal: The raw radar data is converted into an m×n matrix for each frame, where m denotes the fast time dimension index and n denotes the slow time dimension index. The radar data contains DC noise, which can be eliminated by performing mean subtraction on the slow time dimension index in the data matrix of each frame.

$$\overline{E}_{mn} = E_{mn} - \frac{1}{n}\sum_{i=1}^{n} E_{mi}, \tag{5}$$

where E_{mn} is the m_{th} row and n_{th} column element in the matrix, and the second term is the average of the slow time indices.

The radar echo contains many static objects and multi-path reflection information of the environment. We need to

eliminate stationary interference to get the target of interest in the RF map. The phase of the static obstacles in the scene does not change over time, and the difference between two adjacent chirps gets information from moving objects.

$$D_{m,n} = \overline{E}_{m,n+1} - \overline{E}_{m,n}, \tag{6}$$

where D_{mn} is the m_{th} range bin in n_{th} chirp after making difference between the adjacent chirps.

Range-azimuth Estimation: to get the information of human platform To get multiple views of people and concealed objects., RA maps are obtained by performing range-azimuth estimates, where the squared modulus (intensity) provides information on moving targets' power backscatter. Specifically, we perform range-FFT on the static interference filtered matrix to complete the distance estimation and angle-FFT along the antenna dimension on its output to complete the azimuth estimation. The imaging result of each frame has the target range index and azimuth-index spectra. Meanwhile, the user's spectrum appears at different positions in the RA map of each frame, while the static object response is reversed. mmctd further highlights the spectral characteristics of moving users by precomputing the average frequency response in unmanned scenes and performing static interference subtraction on the imaging results of each frame.

Micro-doppler Signature: The small-scale rotation or swing of various parts of the human body during the movement of the human target will introduce Doppler modulation to the signal echo, which is called the micro-Doppler effect. If the pedestrian carries concealed objects, such as knives, the carrier objects will also rotate with the body swing during the movement process, resulting in Doppler frequency shift. Specifically, after calculating the RD map and implementing target detection, the range bin corresponding to the target location r is obtained.The MD signature is calculated by integrating along the range dimension to obtain the time-frequency representation of the target [31], and accumulate the features of multiple channels.

B. Ghost Targets Removal

mmCTD first performs target detection. However, some objects, such as metal walls and turbines in the cabin, will bring serious multipath reflection and specular reflection to the radar signal, resulting in numerous random ghosts in RF maps. They are close to or far away from the target cluster [32], so we cannot distinguish human targets.

We propose a dual radar target detection and ghost removal method. Specifically, Target detection and peak clustering are performed on the RF maps of the two views radar to obtain the location of the scene target on the map. Then the target position coordinates of the two views are converted to the same coordinate system to associate the human target and remove the ghost in the scene. First, the RF maps of multiple channels are integrated incoherently to increase the signal-to-noise ratio of the RF maps. Then the CFAR detection algorithm is performed on the accumulated result to detect the target. On this basis, we perform azimuth estimation, and

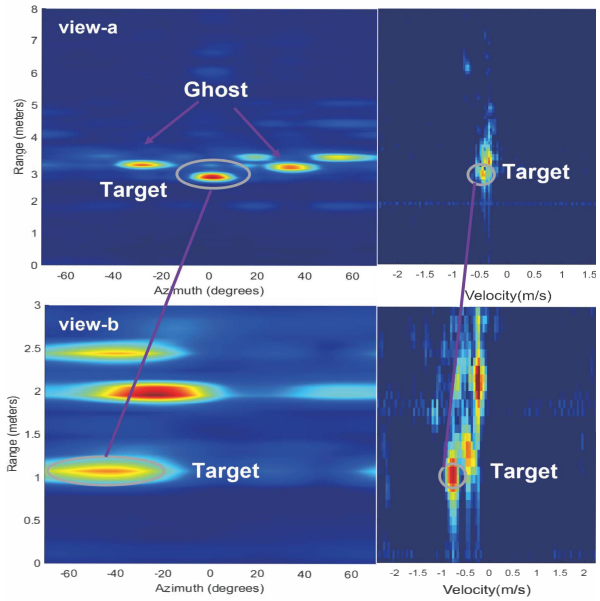


Fig. 8. Target association based on two-view radar RF signatures.

before it is necessary to go to Compensates for Doppler-induced phase shift. Next, the detection results of RF maps are combined to obtain the preliminary localization of the target(r, α, v).

Due to the influence of diverse scene characteristics of different radar views and target distance in the environment, the RF characteristics (area size, intensity) of the target will be pretty different, so the range covered by the target is not a constant feature. After performing object detection, we need post-processing to get the final object position, there may be multiple detection results from the same object. Inspired by the common keypoint detection evaluation metric object keypoint similarity (OKS) for pose estimation [33], we propose a metric called Target Point Similarity (TPS) to describe the correlation between two detection points and eliminate multiple detections from the same target.

$$TPS = exp\left\{-\left(\frac{\Delta d^2}{4s_1 s_2 {\delta_D}^2} + \frac{\Delta v^2}{4v_1 v_2 {\delta_V}^2}\right)\right\}, \tag{7}$$

where Δd is the distance between the two points, s_1 and s_2 being the distance from the radar at each point. Similarly, Δv is the difference in velocity between the two points, v_1 and v_2 being the velocities of the two points respectively. δ_D and δ_V are normalised factors to control attenuation, determined from the true values of known data sets, reflecting the degree of influence of the current point on the whole. The TPS formula is similar to a Gaussian distribution centred on the peak points, and we determine the δ_D and δ_V according to the actual scene so that the similarity is distributed in the approximate range of $0-1$. The proposed TPS takes into account the velocity and distance of the object and determines whether the points come from the same object by describing the similarity between the detected points.

We propose a TPS-based approach to target keypoint detection, the exact process is as follows: 1) The RF maps after

CFAR detection is divided into sub-regions with a size of $X \times Y$, and a set of detection points $\{k_{ni}|i = 1, 2, ..., N\}$ of subregion n is constructed, and the subregions with empty detection points are neglect. 2) Find the maximum amplitude point in each subregion set through peak search, and construct the keypoint set $\left\{\hat{k}_m|m = 1, 2, ..., M\right\}$. 3) Calculate the TPS of the keypoint $\hat{k}_m$ and the detected point k_{ni} in its subregion; if the value of TPS is less than the set threshold, the point k_{ni} is added to the peak point set $\left\{\hat{k}_m\right\}$. Otherwise, it is removed from the amplitude point set $\{k_{ni}\}$. Until all subregions are computed and the point set $\{k_{mi}\}$ is empty, the process ends.

After keypoint detection, the set of probable target point for view-i (including users and multipath ghosts) can be expressed as

$$P_i = \{p_{ij} = (r_{ij}, \alpha_{ij}, v_{ij})|j = 1, 2, 3..., n\}, \tag{8}$$

Where p_{ij} represents the positioning of the j_{th} target to be selected, including the distance, azimuth, and doppler, n is the total number of keypoint detected of view-i. Fig. 8 shows the target detection and association of two view radars.

In order eliminate the random ghosts and reduce the interference of clutter in the ship scene, we combine the radar keypoint sets P_a and P_b of two views. The real target has associated location features in both views, while ghosts and clutter do not have such features. To identify the real target in the RF maps and save computational costs, we transform the radar keypoint set of view-b into the coordinate system with view-a radar as the origin. First, the keypoint set is converted to a Cartesian coordinate system $\hat{P}_i = \{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), ..., (x_{iN}, y_{iN})\}$, and then the coordinates are transformed as following equation:

$$\begin{bmatrix} x_{BA} \\ y_{BA} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_B \\ y_B \end{bmatrix} + \begin{bmatrix} d_a \\ d_b \end{bmatrix}, \tag{9}$$

Where (x_B, y_B) is the keypoint in the view-B radar coordinate system and (x_{BA}, y_{BA}) is its corresponding coordinate in the view-A. d_a and d_b are the offset values of view-b concerning view-A, related to the placement of the radar, and θ indicates the rotation of the coordinate axis.

The correlation of coordinate points between $\hat{P}_{BA}$ and $\hat{P}_A$ is calculated, and the relative tolerance value is set according to the radar error and experience. If the correlation between two coordinates is more significant than the set threshold, the coordinate pair is considered the real target in the scene. Otherwise, it is deleted from the coordinate set. Finally, the point sets $\mathcal{P}_A$ and $\mathcal{P}_B$ containing only real targets are obtained.

C. Feature Extraction

In the radar RF maps (RFM), we focus only on the feature of the small area where the target object is located. To reduce the size of the input data to MVCTDNet, we identify the region of interest (RoI) through target detection and crop the RF map of each view radar accordingly. Since there is a difference in the shape of the target features in the RFM of the two views, we set the appropriate clipping dimension parameters $\Xi_A(\xi_r, \xi_\alpha)$ and $\Xi_B(\xi_r, \xi_\alpha)$ for their feature maps, respectively. by determining the true target location in each frame, the azimuthal dimension and the distance dimension are clipped to a rectangular block of a specific size, called the target feature rectangular block (TRB).

Algorithm 1 Unite Dual-radar Target detection

Input: Data frames from multi-view radar $\{S_{Ai}|i = 1, ..., N\}$, $\{S_{Bi}|i = 1, ..., N\}$;
Point correlation threshold ζ;
Rotation matrix $\mathbf{R}$ and offset vector $\mathbf{H}$;
Output: Target rectangular block (TRB) set $\left\{T\hat{R}B_{Aj}\right\}_{j=1}^{N_t}$,$\left\{T\hat{R}B_{Bj}\right\}_{j=1}^{N_t}$;

while $i < N + 1$ **do**
  Stationary interference removal on S_{Ai} and S_{Bi} using MTI algorithm and mean subtraction method;
  $RDA_i = \mathbf{FFT}_{azimuth}\mathbf{FFT}_{doppler}\mathbf{FFT}_{range}\{S_i\}$;
  Target detection using CFAR algorithm on RDA_i;
  Divide RF map into $X \times Y$ sub-regions;
  Build the point set $\{k_r\}$ for r_{th} sub-region;
  for $r = 1 \rightarrow X \times Y$ **do**
    $\hat{k}_{\mathbf{r}} \leftarrow \max$ amplitude $\mathbf{k}_r$;
    Pick $\hat{k}_{\mathbf{r}}$ to keypoint set $\left\{\hat{k}_m\right\}$;
    if $\mathbf{TPS}(\hat{k}_r, k_{ri}) \geq \zeta$ **then**
      remove k_{ri} from $\{k_r\}$;
    else
      Pick k_{ri} to $\left\{\hat{k}_m\right\}$;
    end if
  end for
  Determine the set of probable target points $\hat{P}_A$ and $\hat{P}_B$ for the $view - A$ and $view - B$, respectively;
  $\hat{P}_{BA} \leftarrow \mathbf{R}\hat{P}_B + \mathbf{H}$;
  Get the final target point sets $\{\mathbb{P}_\mathbb{A}|j = 1, ..., J\}$ and $\{\mathbb{P}_\mathbb{B}|j = 1, ..., J\}$ of the two perspectives by correlating $\hat{P}_{BA}$ and $\hat{P}_A$;
  $TRB_i \leftarrow$ clipping $\mathbf{RAM}_i$ by $\Xi(\xi_r, \xi_\alpha)$;
end while
while $j \leq J$ **do**
  $T\hat{R}B_j \leftarrow \sum_{m=1}^{N_f} TRB_{im} \cdot \gamma / N_f$
end while

Our preliminary observation shows that the TRB in RAM cannot be considered a time-invariant feature. Different speeds and walking styles during human movement, as well as changes in the distance from the radar sensor, can lead to specific changes in the region's intensity, size, and shape. To increase the robustness of the TRB, we need to eliminate some speed and distance effects. The average intensity I of TRB is related to the distance r as $I \propto \frac{\sigma}{r_4}$. To extract the reflection properties σ of the target in the RFM, We multiply the TRB of each frame with parameter related to the target distance. This is done to eliminate part of the effect of distance on feature intensity. Next, we choose to superimpose multiple consecutive frames to eliminate the effect of target's movement on TRB; according to the parameter setting of radar, the interval time of each frame is about $0.12s$, and the speed of human walking is $1.1m/s - 2.5m/s$; we make a reasonable

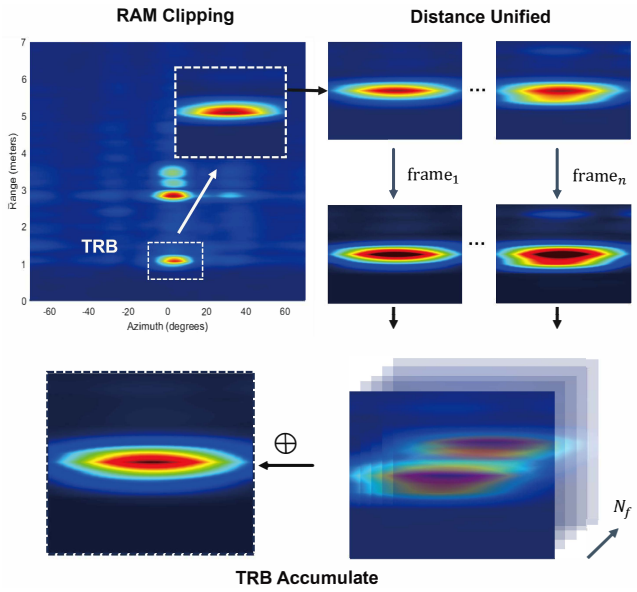


Fig. 9. Based on the location of multi-view detection, TRBs are clipped from the generated ra maps, the distance is united and the average accumulate is performed.

correlation between the number of accumulated frames N_f and the movement speed v_t, denoted as $N_f \propto v_t$. To calculate the mean value of the accumulation of multiple frames by expressing the above process as

$$T\hat{R}B = \frac{\sum_{j=1}^{N_f} TRB_{ij} \cdot \gamma}{N_f}, \tag{10}$$

Where number of accumulation frames N_f is determined by the target's motion speed v_t. TRB_{ij} is the region of target i obtained by our cropping in j_{th} frame, and γ is the parameter related to the distance r_i, which can unify the TRB_i at different distances. With the calculation of equation (10), we obtain the final $T\hat{R}B$, which contains the reflection features of the target object and the items it carries.

VI. Multi-View Concealed Threat Detection Model

In this section, we introduce the proposed MVCTDNet model, a multi-view domain adversarial training deep learning framework, for extracting meaningful information from three feature views to achieve environment-independent human concealed threat carry detection. The model contains three components: **View feature extractor**, **threat recognizer**, **environment discriminator**, as illustrated in Fig. 10.

To achieve the goal above, we first designed a view feature extractor whose inputs are RF feature maps from two radars, including the MD features of targrt movements and $T\hat{R}B$ obtained from Range-Angle map post-processing. The CNN-based feature extractor learns vector representations of multiple views. By layering the sets of different views into latent representations, they are fed to the threat recognizer and the environment discriminator, respectively. The threat recognizer aims to maximize the accuracy of the concealed item prediction, and the environment discriminator is designed to maximize the prediction of the environment (specifically, identifying the ship scene where the target is located). The features we need for classification should come from hidden items only, not from the environment and pedestrian subjects. Therefore, we introduce a gradient inversion layer between the feature extractor and the environment discriminator so that they play this minimax game. This process suppresses features unrelated to hidden items and prevents MVCTD-Net from being overfitted to a specific scene. Next, we elaborate on the details of the model.

A. Multi-View Feature Extractor

Since the input of MVCTDNet is the RF maps of three views, we propose a CNN-based feature extraction module to extract each view's unique representation. Data with various features are mapped to appropriate representation vectors. Specifically, our encoder consists of multiple convolutional layers, followed by batch normalization layers to fit the data to an approximately Gaussian distribution and rectified linear units (ReLu) to introduce non-linear activations.

The input of the i_{th} view is X_i, and the CNN-based feature extraction network G_{v_i} is applied to the i_{th} view to convert it into an N-dimensional latent representation $Z \in \mathbb{R}^N$.The corresponding parameters of the CNN block are denoted as θ_i, and we can obtain the latent feature F as follows.:

$$F_i = G_{v_i}(X_i; \Theta_i). \tag{11}$$

In our case, for the front-view TRB(FTRB) and side-view TRB(STRB), because they have the same characteristics, 3×3 and 5×5 filters with the same settings are used in each layer to extract different scale features of TRB. For the MD view, we choose three layers of 1×1 and 3×3 convolutions to extract the μD signature To capture the features of each view, first, we stack the feature representations F_f and F_s of the FTRB view and STRB view output by the encoder. Next, the stacked features $F_c = F_f \oplus F_s$ are sent to the convolution module G_{v_c} to adjust the number of features and reduce complexity. Before concatenating the representations of each view, we use global maximum pooling (GMP) to prevent overfitting. We concatenate the MD view output F_m with the above features, and the global representation vector F_v is expressed as follows:

$$F_v = G_{v_c}(F_c; \Theta_c) \oplus F_m. \tag{12}$$

B. Concealed Threat Recognizer

To predict the labels, the threat recognizer G_c maps the feature representation F_v output by the feature extractor G_v to the label space $\mathbb{R}^M$, where M is the number of partitioned threats. G_c consists of a two-layer linear network, each of these layers is followed by a ReLU activation function. Then a Softmax function is used to obtain the probability vector. For a given input F_v, the mapping function is defined as:

$$\hat{y}_i = G_c(F_v; \Theta_c), \tag{13}$$

where Θ_c are the learnable parameters in G_c and $\hat{y}_i$ denotes the predicted probability distribution of labeled data by G_c.

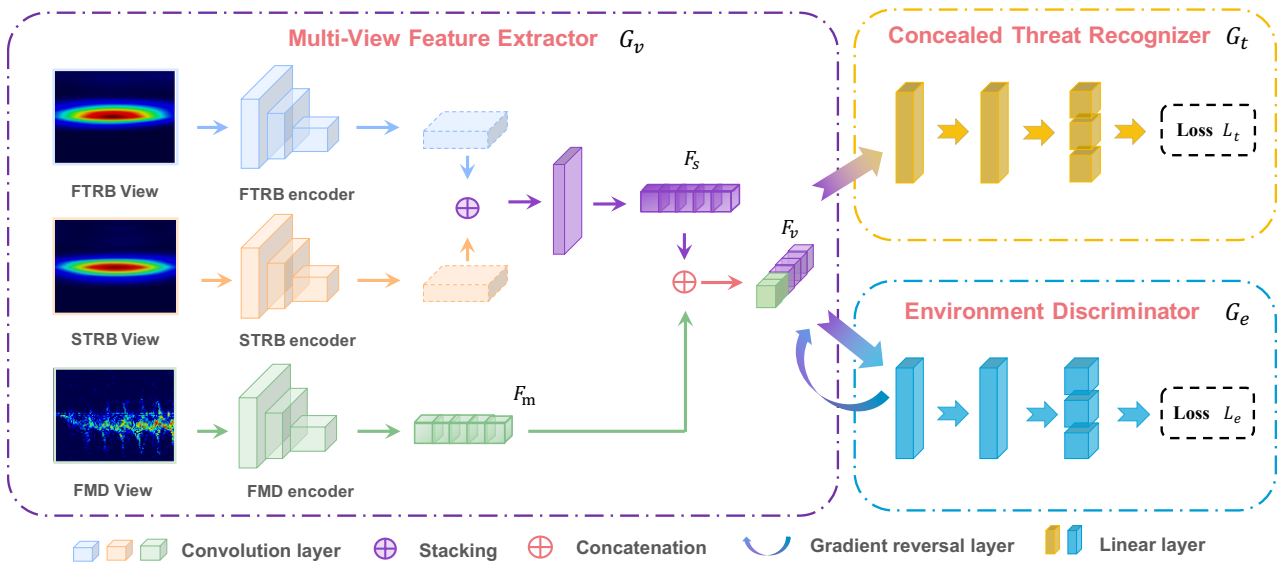


Fig. 10. The architecture of MVCTDNet with its adversarial learning strategy.

The Θ_c can be trained by minimizing the cross loss between the predictions $\hat{y}_i$ and ground truth labels y_i. The loss L_t of the threat recognizer is defined as:

$$L_t = -\frac{1}{|L|}\sum_{l=1}^{|L|}\sum_{m=1}^{|M|} y_{im}\log(\hat{y}_{im}), \tag{14}$$

where $|L|$ is the number of data with labels.

However, the above expressions ignore the impact of the environment; data from different regions of the ship contain specific environmental characteristics (e.g., equipment layout, cabin structure), and G_c may not be able to learn common characteristics of all scenarios.

C. Environment Discriminator

Considering that there are dense metals in ship areas and their size, structure, and layout vary greatly, it means that the RF features we extract contain many environmental features. To remove the influence of environment on the classification results of G_f, we employ a domain adversarial training [34] to derive mapping shared across different environments. MVCTDNet is trained to predict threat class labels using data collected from three scenarios. We introduce an environment discriminator G_e that maximizes the accuracy of environment label prediction during training. Specifically, our G_e consists of two fully connected layers and uses ReLu for activation, which projects the input F_v into the environment distribution $\hat{d}_i$:

$$\hat{d}_i = G_e(F_v;\Theta_e), \tag{15}$$

where Θ_e denotes the trainable parameters of G_e, we define the environment loss L_e as the cross-entropy between the environment distribution $\hat{d}_i$ and the ground truth label:

$$L_e = -\frac{1}{|L|}\sum_{l=1}^{|L|}\sum_{m=1}^{|E|} d_{ie}\log(\hat{d}_{ie}), \tag{16}$$

where d_i represents the set of ture labels and D is the number of labels.

Our ultimate goal is for the model to derive environment-independent representations. Therefore, we use the gradient reversal layer (GRL) to reverse the gradient during the back-propagation process to achieve loss negation and obtain the following loss function:

$$L = L_t - \lambda L_e, \tag{17}$$

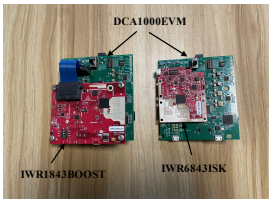
where λ is the weighting parameters.

VII. Implementation And Evaluation

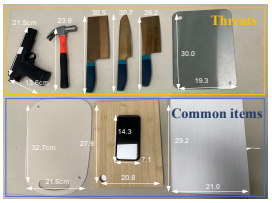
A. Experimental Methodology

Experimental Setup: We implement and verify mmCTD on COTS IWR1843 and IWR 6843 mmWave radars paired with DCA1000EVM data capture cards to acquire raw data. Specifically, IWR1843 operates in the frequency band from 77GHz to 81GHz, and IWR 6843 operates in the 60-64GHz frequency band. They all have 3 Tx antennas and 4 Rx antennas. IWR1843 provides FoV of 120° azimuth and 30° elevation with 15° angle resolution, and IWR6843 provides 120° azimuth FoV and 120° elevation FoV with 28° angle resolution. The radars are configured to sweeping bandwidth of $3.6GHz$, a sampling rate of $10Msps$, 128 chirps per frame, 256 sampling samples per chirp, and a frequency slop of $56.964MHz/\mu s$. The above configuration gives a range resolution of $0.04m$, and a velocity resolution of $0.25m/s$ (1843) and $0.32m/s$ (6843). We implemented mmCTD in MATLAB and Pytorch. After data collection, the model training and prediction is accomplished on a desktop with an Intel i5-10600KF CPU and an NVIDIA RTX 3080 GPU.

Date Collection: mmCTD is oriented towards threat detection in restricted areas of cruise ships, so we regard dangerous metal goods of a specific size as threats categories and mainly distinguish three target categories: **threats** (knives, metal gun

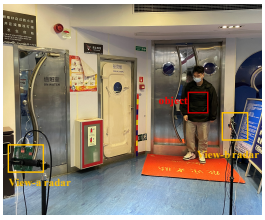


(a) Experimental device



(b) Objects involved in the experiment and their classification

(c) Experiment in engine room



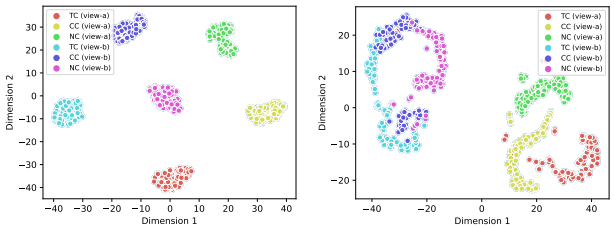
(d) Experiment in cabin aisle

Fig. 11. Illustration of the experimental setup.

model, hammer), **common objects** (book, mobile phone, plastics plate) and **carry nothing**. We chose to replace explosives with metal plates, and the common item category includes daily objects carried by people on board. Each class's target items have different sizes and shapes, as shown in Fig. 11(b).

We conducted experiments in three real ship scenarios, namely engine room, wheelhouse, and critical passage, which can represent typical restricted areas, as shown in Fig. 11. During the data collection, we placed two radars in the cruise ship area in a vertical orientation with a field of view covering the front and side of the human body, which can be easily extended to more directions. To increase the diversity of the data, we had 4 volunteers participate in the data collection, each wearing clothes of different thickness. During each data collection, the radar records for 6 seconds, and the user walks in a random motion pattern, including different speeds, directions, and starting points. In addition, the items and hidden parts carried by the user also changed, including front and side placement. For each class of items, we collected 100 sets of data in one scene. We also evaluated the performance in the case of state changes by having people walk in various distances and motion speeds. Overall, we collected a total of 91,800 frames of data.

Evaluation Metrics: To quantify the performance of our system, we define the following evaluation metrics: *recognition accuracy* as the percentage of correctly predicted objects among all test samples; *Confusion matrix*, where each row corresponds to the actual target object, and each column represents the target label predicted by mmCTD; *precision* indicates the proportion of actual positive cases in all positive cases; *false alarm rate* (FAR) means the probability of false detection when the target does not exist; *missing alarm rate* (MAR) represents the probability of misidentifying the target, but it exists. The above metrics, $precious = tp/(tp + fp)$, $MAR = fn/(tp + fn)$, $FAR = fp/(fp + tn)$, are defined by true positive (TP), false positive (FP), true negative (TN), and false negative (FN) in confusion matrix. In the CTD task of the system, it is necessary to identify threats as much as possible, so the precision and missing rate are vital indicators.



(a) (b)

Fig. 12. Visualization of the feature distribution of our extracted TRB and original RF map with t-sne. (a) t-sne of TRB for three carrying cases. (b) t-sne of original RF map for three carrying cases.

B. Performance Evaluation

1) Performance Comparison with Different Views Input: The system's performance of different numbers of view inputs is evaluated without considering scene changes. The input is three-view (frontal TRB, side TRB, and frontal micro-Doppler), two-view (without frontal MD), and single-view (only frontal TRB) respectively. Considering the actual scenario, we randomly split the dataset (70% chest, 15% side, 15% leg) where users carry objects in different parts. The recognition results are shown in Table I. The multi-view feature extractor with three and two-view input has satisfactory recognition performance, and the average recognition accuracy reaches 90.27% and 86.16%. However, the single-view input has a lower recognition rate, only 68.44%. This is consistent with our validation that a single view is insufficient to identify features when objects are hidden from the side of the body or deflected. The results show that the three-view MVCTDNet can effectively identify hidden threats when items are hidden in different parts of the body, etc.

2) Performance Comparison with Different Environments : We also examine the threat recognition robustness of mmCTD in environment independence. We compare our solution with a 3-Views DCNN consisting only of multi-view feature extractor and threat recognizer without applying environment discriminator. We collect radar raw data in three different ship scenarios and process the raw data to obtain TRB features. Fig. 13 describes the threat recognition performance in three scenarios; MVCTDNet can achieve 87.3%, 90.1%, and 89.7% threat subject recognition accuracy, respectively, and the accuracy is 10.70% higher than the 3-Views DCNN baseline method on average. Furthermore, we observe that the threat identification accuracy in the cabin is slightly lower. Because more dense mechanical devices and metal bulkheads are in the cabin, which will cause more potent interference to the millimeter wave radar signal. The above shows that MVCTDNet can effectively extract environment-independent features and realize threat identification in multiple ship scenarios.

3) Overall Performance: We finally evaluate the performance of mmCTD in three scenarios of ships, and the corresponding evaluation results are shown in Fig. 14. As shown, the threats class obtained the highest average recognition precision of 89.1%, followed by The no-carry category was

TABLE I
CONFUSION MATRIX OF MMCTD UNDER VARIOUS VIEW INPUT

Network Instances	Class	Threats	Common items	Nothing	Recognition Rate (%)
3-Views	Threats	93.10	6.21	0.69	90.27
	Common items	3.33	86.67	10.00	
	Nothing	0.00	8.97	91.03	
2-Views (wo FMD View)	Threats	88.27	8.97	2.76	86.16
	Common items	4.00	83.33	12.67	
	Nothing	2.07	11.72	85.94	
1-View (w FTRB View)	Threats	67.60	19.31	12.41	68.44
	Common items	11.99	66.01	22.00	
	Nothing	7.59	20.69	71.72	

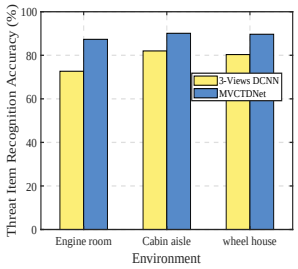


Fig. 13. Accuracy comparison in different environment.

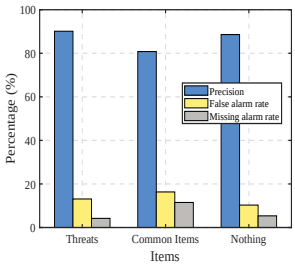
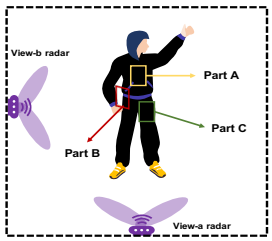
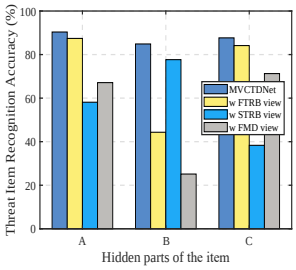


Fig. 14. Overall performance.



(a)



(b)

Fig. 15. Performance comparison for objects concealed in different parts. (a)Hide items in three body parts: chest, sides and legs. (b)Comparison in hidden parts robustness.

87.60%. In comparison, the common item type was slightly lower at 80.80%. Meanwhile, the no-carry class achieved the lowest FAR of 10.3%, 2.8% lower than the Threat class, and 5.0% lower than the Common Theme class. The Threat class performed the lowest MAR of 4.2%, which was 7.3% and 1.2% lower than the Common Theme and No Carry classes, respectively. The results show that mmCTD performs well in concealed threat identification with suitable settings in the ship scene.

C. Impact of Various Factors

1) Impact of Distance from Radar: To test the performance of mmCTD at different distances, we set the measurement range to 0-5m according to the application scenario, and the interval between each group is 1m. Fig. 17 shows the five sets' precision, FAR, and MAR versus distance. In the figure, we can observe that the farther the user is from the radar, the recognition accuracy of the system first increases and then decreases, the FAR decreases with the distance, and the MAR increases with the distance. This is in line with our common sense because when the distance between the user and the radar changes, the reflection amplitude of the target will change, thus affecting the performance of the system. When the user is 0-1m away from the radar, the system has poor performance; because the user is too close to the radar, there will be a more strong reflection, and the multipath effect will be more severe in the ship environment. The distance factor compensation is insufficient to overcome this influence and only valid within a certain range.

2) Impact of TRB Accumulate Frames: TRB is feature cropped from the original RA image and then accumulated over multiple frames. Compared with the original RA map, TRB contains more information about the target. We use t-SNE to visualize TRB and RA map features in 2D space Fig. 12. Regardless of the speed impact, we evaluate the performance of mmCTD under different number of TRB frames; the number of frames tested ranges from 2 to 10, and the interval between each group is two frames. Fig. 18 shows the relationship between the cumulative number of frames and the accuracy of hidden threat detection, FAR, and MAR. The figure shows that stacking a larger number of TRBs can improve accuracy and reduce the FAR and MAR, making the detection results more robust. At the same time, when the accumulated number of frames is greater than 6, the performance improvement of the system is not apparent, and more frames mean more computing time is required. Therefore, a trade-off between timeliness and accuracy is required to select an appropriate number of TRB accumulation frames.

3) Impact of Movement Speed: In this experiment, we aim to evaluate the robustness of mmCTD at different motion speeds. Specifically, we collect five sets of data in the range of 0.5m/s-2.0m/s according to the typical velocity of pedestrians. Fig. 16 shows the curves of threat recognition accuracy, FAR, and MAR corresponding to different target movement speeds. From the figure, we can observe that when the user walks at a speed of 0.5m/s-1.4m/s, the change in threat identification accuracy is not apparent. However, when the target moves in a running state, the accuracy decreases, and the FAR and MAR increase. Because the faster the moving speed, the greater the swing of the body, which can easily cause more serious multipath reflections, reducing the echo reception. The outline of target's area in heat map and MD signature will also

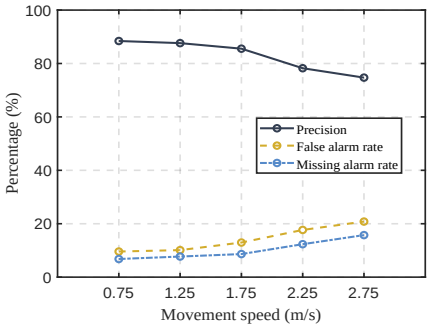


Fig. 16. Impact of movement speed.

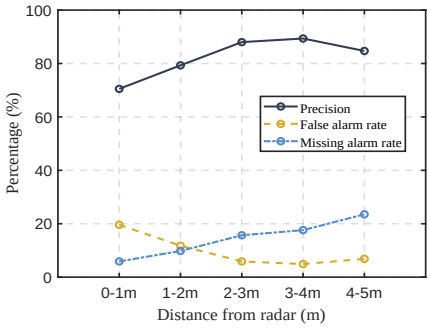


Fig. 17. Impact of distance from radar.

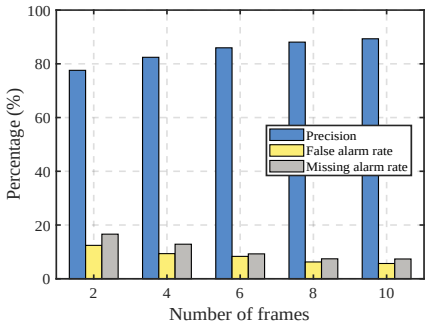


Fig. 18. Impact of TRB accumulate frames number.

TABLE II
COMPARISON OF MMCTD WITH THE STATE-OF-THE-ART-METHODS.

Method	Target	Device	Input	Performance	Acc. in Diff. Env	Range	Running time	Area	Multiple Direction
Liu *et al.* [8]	10 CAT. objects	77GHz mmWave radar	SAR images	Acc. 98.18%	-	90mm	$>$ 64.43s	-	N/A
Li *et al.*. [27]	w/wo shotgun	5.8GHz FMCW radar	Spectrums (2D)	Acc. 99.21%	-	-	$>$6s	Open spaces	N/A
COD [9]	phone, laptop, knife	77GHz mmWave radar	Spectrums (3D)	FAR 33.84%, MAR 23.85%	-	$<$ 6m	215.63ms	Lobby, room	N/A
Yang *et al.* [17]	gun, knife, phone, wo object	Passive mmWave imager	PMMW images	Precision 88.1%	-	1-4m	-	Lab	✓
Sun *et al.* [35]	w/wo object	Active mmWave imager	AMMW images	Precision 90.61%	-	-	-	-	✓
mmCTD (ours)	threat, common-item, wo object	77GHz mmWave radar	Multi-view Spectrums(2D)	Precision 89.33%, FAR 10.32%	89.03% (3)	1-4.5m	$>$ 0.76s	Ship cabin	✓

change as the speed changes. To improve the accuracy under fast motion, the superimposed frame of TRB can be adjusted accordingly.

4) Impact of Items Concealed Parts: All-round detection was our focus. We further tested the performance of mmCTD when threats were concealed in different parts of the user to ensure that the system could work stably. Our setup selects three common carrying sites: chest, side of the body, and legs, as shown in Fig. 15(a). At the same time, Fig. 15(b) depicts the threat identification accuracy of MVCTDNet and the models with three single-view (w FTRB-view, STRB-view, and MD-view) inputs. The test results show that the single-view input model declines in accuracy when the item is hidden in some parts. For example, when it is carried on the side, there is a missing alarm for the FTRB and MD single-view input, and the same is true for the STRB view when it is carried on the front. When item is concealed on legs, the recognition rate of the system is likely to decrease in comparison to the other scenarios due to the substantial movement of the legs. MVCTDNet has achieved good performance in the detection of hidden threats placed in different parts.

D. Comparison with state-of-the-art Methods

Owing to the device used, model input, detection targets, and application scenarios being different, we cannot fully reproduce the state-of-the-art methods of CWD on the ship scene and perform performance comparisons. We list their model and related parameters in Table II.

Compared to some of the methods mentioned in Table, our solution is designed for the cruise ship environment, using a low-cost mmWave radar (3T4R) under cost constraints. There are significant differences between ship cabins, we considers the system's application in multiple scenarios and discuss its accuracy in different scenarios. mmCTD makes a trade-off between detection time and accuracy and uses three-view features as input to achieve various-direction detection and overcome severe ghosts in ship scenes.

VIII. LIMITATIONS AND FUTURE WORKS

According to the advantages described above, our proposed mmCTD can meet some practical requirements of security check on cruise ship. However, our system also has some limitations:

Currently, our system divides objects into three categories without distinguishing between specific items. This is due to the resolution limitation of the 77GHz millimeter-wave radar and number of antennas, and our multi-view features is not enough to recognize the outline of the object. When the user carries small metal objects or many common items, it may cause missed and false alarms. For future research, we focus on the trade-off between efficiency and accuracy and explore more granular methods. Due to the equipment's limitations and the scene's complexity, the proposed mmCTD system requires the target to keep a relatively close distance $(1-4.5m)$ from the radar to maintain a good detection capability.

Our experiments only consider the detection of a single target. On the one hand, the deployment area is narrow passages and aisles. Second, when there are multiple users, the multipath effect in the scene will be more serious, and the complexity of system calculation will be reduced and

decreases with accuracy. Therefore, we include multi-user CTD covering more scenarios as part of our future work.

IX. Conclusion

In this paper, we investigate the feasibility of using mmWave radar for concealed threat detection, which can improve security screening services for ship restricted areas. To this end, we developed mmCTD, a system that uses multi-view COTS mmWave radars to detect three types of targets in real-time: threats, common items, and no-carry. To overcome severe multipath in cruise ships, mmCTD employs a method of correlating multiple radars for object detection. We then design a multi-view domain confrontation adaptive model to remove the environment-specific information in the features and effectively use the designed multi-view features for all-around detection. Furthermore, we conduct numerous experiments to analyze the impact of different factors and achieve satisfactory results, verifying that the system can be widely deployed on ships to identify concealed threats.

References

[1] M. Ahola and R. Mugge, "Safety in passenger ships: The influence of environmental design characteristics on people's perception of safety," *Applied Ergonomics*, vol. 59, pp. 143–152, Mar. 2017.

[2] K. Mhou and D. van der Haar, "A Review of Threat Profiling Techniques for Use in Concealed Weapon Detection Systems," in *Information Science and Applications 2018*, K. J. Kim and N. Baek, Eds. Singapore: Springer Singapore, 2019, vol. 514, pp. 201–209.

[3] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless Sensing for Human Activity: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2020.

[4] A. Hanif, M. S. Chughtai, A. A. Qureshi, A. Aleem, F. Munir, M. Tahir, and M. Uppal, "Non-Obtrusive Detection of Concealed Metallic Objects Using Commodity WiFi Radios," in *2018 IEEE Global Communications Conference (GLOBECOM)*. Abu Dhabi, United Arab Emirates: IEEE, Dec. 2018, pp. 1–6.

[5] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang, "Towards In-baggage Suspicious Object Detection Using Commodity WiFi," in *2018 IEEE Conference on Communications and Network Security (CNS)*. Beijing: IEEE, May 2018, pp. 1–9.

[6] Y. Zhao and W. Hu, "CentralNet Method for Human motion Recognition Based on Multi-feature Fusion of Millimeter Wave Radar," in *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. Xi'an, China: IEEE, Aug. 2021, pp. 1–6.

[7] M. E. Yanik and M. Torlak, "Near-Field 2-D SAR Imaging by Millimeter-Wave Radar for Concealed Item Detection," in *2019 IEEE Radio and Wireless Symposium (RWS)*. Orlando, FL, USA: IEEE, Jan. 2019, pp. 1–4.

[8] J. Liu, K. Zhang, Z. Sun, Q. Wu, W. He, and H. Wang, "Concealed Object Detection and Recognition System Based on Millimeter Wave FMCW Radar," *Applied Sciences*, vol. 11, no. 19, p. 8926, Sep. 2021.

[9] X. Gao, H. Liu, S. Roy, G. Xing, A. Alansari, and Y. Luo, "Learning to Detect Open Carry and Concealed Object With 77 GHz Radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 4, pp. 791–803, Jun. 2022.

[10] R. Feng, E. D. Greef, M. Rykunov, H. Sahli, S. Pollin, and A. Bourdoux, "Multipath Ghost Recognition for Indoor MIMO Radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[11] Z. Wei, B. Li, T. Feng, Y. Tao, and C. Zhao, "Area-Based CFAR Target Detection for Automotive Millimeter-Wave Radar," *IEEE Transactions on Vehicular Technology*, pp. 1–15, 2022.

[12] S. Li and S. Wu, "Low-Cost Millimeter Wave Frequency Scanning Based Synthesis Aperture Imaging System for Concealed Weapon Detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 7, pp. 3688–3699, Jul. 2022.

[13] H. Yang, Z. Yang, A. Hu, C. Liu, T. J. Cui, and J. Miao, "Unifying Convolution and Transformer for Efficient Concealed Object Detection in Passive Millimeter-Wave Images," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2023.

[14] B. Kapilevich, B. Litvak, A. Shulzinger, and M. Einat, "Portable Passive Millimeter-Wave Sensor for Detecting Concealed Weapons and Explosives Hidden on a Human Body," *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4224–4228, Nov. 2013.

[15] M. U. Malik, W. Ahmed, A. Majid, Z. Yaqoob, and A. Rafique, "Learning augmented standoff Concealed Weapon Detection," in *2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. Islamabad, Pakistan: IEEE, Jan. 2017, pp. 830–837.

[16] T. Liu, Y. Zhao, Y. Wei, Y. Zhao, and S. Wei, "Concealed Object Detection for Activate Millimeter Wave Image," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9909–9917, Dec. 2019.

[17] H. Yang, D. Zhang, A. Hu, C. Liu, T. J. Cui, and J. Miao, "Transformer-Based Anchor-Free Detection of Concealed Objects in Passive Millimeter Wave Images," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–16, 2022.

[18] Hua-Mei Chen, Seungsin Lee, R. Rao, M.-A. Slamani, and P. Varshney, "Imaging for concealed weapon detection: A tutorial overview of development in imaging sensors and processing," *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 52–61, Mar. 2005.

[19] B. Kapilevich and M. Einat, "Detecting Hidden Objects on Human Body Using Active Millimeter Wave Sensor," *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1746–1752, Nov. 2010.

[20] G. Y. Tian, A. Al-Qubaa, and J. Wilson, "Design of an electromagnetic imaging system for weapon detection based on GMR sensor arrays," *Sensors and Actuators A: Physical*, vol. 174, pp. 75–84, Feb. 2012.

[21] A. Lamas, S. Tabik, A. C. Montes, F. Pérez-Hernández, J. García, R. Olmos, and F. Herrera, "Human pose estimation for mitigating false negatives in weapon detection in video-surveillance," *Neurocomputing*, p. S0925231221019159, Jan. 2022.

[22] R. Mahajan and D. Padha, "Detection Of Concealed Weapons Using Image Processing Techniques: A Review," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. Jalandhar, India: IEEE, Dec. 2018, pp. 375–378.

[23] N. Jasim Hussein, F. Hu, and F. He, "Multisensor of thermal and visual images to detect concealed weapon using harmony search image fusion approach," *Pattern Recognition Letters*, vol. 94, pp. 219–227, Jul. 2017.

[24] M. Kowalski, M. Kastek, M. Walczakowski, N. Palka, and M. Szustakowski, "Passive imaging of concealed objects in terahertz and long-wavelength infrared," *Applied Optics*, vol. 54, no. 13, p. 3826, May 2015.

[25] F. Garcia-Rial, D. Montesano, I. Gomez, C. Callejero, F. Bazus, and J. Grajal, "Combining Commercially Available Active and Passive Sensors Into a Millimeter-Wave Imager for Concealed Weapon Detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 3, pp. 1167–1183, Mar. 2019.

[26] M. Kowalski and M. Kastek, "Comparative Studies of Passive Imaging in Terahertz and Mid-Wavelength Infrared Ranges for Object Detection," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 2028–2035, Sep. 2016.

[27] Y. Li, Z. Peng, R. Pal, and C. Li, "Potential Active Shooter Detection Based on Radar Micro-Doppler and Range-Doppler Analysis Using Artificial Neural Network," *IEEE Sensors Journal*, vol. 19, no. 3, pp. 1052–1063, Feb. 2019.

[28] O. Bazgir, D. Nolte, S. R. Dhruba, Y. Li, C. Li, S. Ghosh, and R. Pal, "Active Shooter Detection in Multiple-Person Scenario Using RF-Based Machine Vision," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3609–3622, Feb. 2021.

[29] M. Chen, K. Liu, J. Ma, Y. Gu, Z. Dong, and C. Liu, "SWIM: Speed-Aware WiFi-Based Passive Indoor Localization for Mobile Ship Environment," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 765–779, Feb. 2021.

[30] K. Liu, W. Yang, M. Chen, K. Zheng, X. Zeng, S. Zhang, and C. Liu, "Deep-Learning-Based Wireless Human Motion Tracking for Mobile Ship Environments," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 24 186–24 198, Dec. 2022.

[31] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor Person Identification Using a Low-Power FMCW Radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018.

[32] J. Chen, S. Guo, H. Luo, N. Li, and G. Cui, "Non-Line-of-Sight Multi-Target Localization Algorithm for Driver-Assistance Radar System," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2022.

[33] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, "DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 4929–4937.

[34] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-Adversarial Training of Neural Networks," in *Domain Adaptation in Computer Vision Applications*, G. Csurka, Ed. Cham: Springer International Publishing, 2017, pp. 189–209.
[35] P. Sun, T. Liu, X. Chen, S. Zhang, Y. Zhao, and S. Wei, "Multi-Source Aggregation Transformer for Concealed Object Detection in Millimeter-Wave Images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 6148–6159, Sep. 2022.

Danei Gong received the B.S. degree in communication engineering from Wuhan University of Technology (WUT), Wuhan, China, in 2020. He is currently pursuing the master's degree with WUT, Wuhan, China. His research interests are in wireless sensing.

Kezhong Liu received the B.S. and M.S. degrees in marine navigation from the Wuhan University of Technology(WUT), Wuhan, China, in 1998 and 2001, respectively. He received the Ph.D. degree in communication and information engineering from the Huazhong University of Science and Techonogy, Wuhan, China, in 2006. He is currently a professor with School of Navigation, WUT. His active research interests include indoor localization technology and data mining for ship navigation.

Xuming Zeng received the Ph.D. degree in Earth Exploration and Information Technology from the China University of Geosciences, Wuhan, China, in 2018, and the Post-Doctoral Fellow in Traffic and Transportation Engineering with the School of Navigation, Wuhan University of Technology (WUT), Wuhan. He is currently an associate researcher with WUT where he has been involved in the study of wireless sensing networks for shipboard environments. His research interests include routing protocols, media access control (MAC), quality of service (QoS), and reliable wireless transmission.

Shengkai Zhang received his Ph.D. degree from the EIC Department, Huazhong University of Science and Technology (HUST) in 2021. He received his M.Sc and M.Phil degrees from HUST and Hong Kong University of Science and Technology in 2012 and 2014, respectively. He is currently an associate professor at Wuhan University of Technology (WUT). His recent research interests include state estimation, wireless sensing, mobile computing, multi-sensor fusion, robot control and planning.

Mozi Chen received the B.S. degree in electric engineering from the Hubei University of Technology, China, in 2013. He received the M.S. and Ph.D.degree in navigation engineering from the Wuhan University of Technology (WUT), China, in 2016 and 2020. He is currently an associate researcher in WUT. His research work has been focusing on wireless sensing techniques and machine learning algorithms for human localization, emergency navigation and activity recognition in mobile environment, i.e., cruise ships.

Kai Zheng received the Ph.D. degree with the School of Geodesy and Geomatics, Wuhan University, in 2020. He is currently an associate professor with the Wuhan University of Technology. His research interests are GNSS precise positioning techniques.

Dashuai Pei received the BS degree in Automation from the Northwest Minzu University, Lanzhou, China, in 2018, and the MS degree in traffic and transportation engineering, in 2022, from the Wuhan University of Technology, Wuhan, China, where he is currently working toward the Ph.D. degree. His research interests are in wireless sensing.