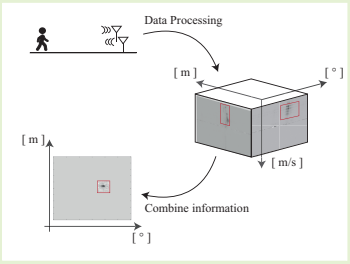
mmDetect: YOLO-based Processing of mm-Wave Radar Data for Detecting Moving People

mmDetect: YOLO-based Processing of mm-Wave Radar Data for Detecting Moving People

Michela Raimondi, Gianluca Ciattaglia, *Member, IEEE*, Antonio Nocera, *Student Member, IEEE*, Linda Senigagliesi, *Member, IEEE*, Susanna Spinsante, *Senior Member, IEEE*, and Ennio Gambi, *Senior Member, IEEE*

***Abstract*— The application of millimeter Wave (mmWave) Radar sensors for people monitoring raised a lot of interest in the context of Active Assisted Living (AAL), especially since the processing of Radar signals can provide interesting information about the observed subjects. Correct recognition of the ongoing behavior, however, cannot disregard the location of the subject. Detection approaches, based on Constant False Alarm Rate (CFAR) algorithms, sometimes fail to correctly identify the location of targets within the observed scenario, especially in complex environments such as indoor situations. This paper proposes the use of a mmWave Multiple Input Multiple Output (MIMO) Radar in combination with a You Only Look Once (YOLO) neural network-based algorithm for the detection and localization of moving targets in indoor environments by processing all the data cube information at the same time. Results are validated through experimental tests which involve subjects walking in linear or random mode, different Radar configurations, and different indoor environments. By exploiting at the same time information such as the angle, Doppler, and range distance of the target, the proposed approach proves to be very effective in the examined scenarios. Experimental results will be discussed in this work to demonstrate the effectiveness of the proposed method.**



***Index Terms*— Classification, FMCW Radar, people detection, target recognition, YOLO**

I. Introduction

The ability to identify and recognize moving people in a sequence of images (optical or derived from other sensors) is crucial to successfully tracking, and identifying them and their activities, and in general, understanding human behavior in indoor applications. Indoor surveillance is important not only for security reasons but also for health monitoring purposes. Collecting information about a moving target within a room, in fact, allows detecting possible falls [1]–[4], or monitoring activities of daily living for the purpose of assessing a person's level of autonomy in performing these actions independently.

The problem of indoor surveillance can be approached in different ways and with different types of sensors. The most common solutions make use of RGB cameras. These methods, however, present some problems and limitations; for example, they are affected by poor illumination of the detection area and may also present some privacy concerns [5]–[7], while adverse indoor conditions like smoke presence can affect their detection capability [8]–[10].

The problem of privacy and poor illumination can be overcome by using an RGB-D sensor, which provides a depth image less prone to release confidential information. The privacy issue is addressed but at the expense of a limitation on the maximum detectable distance. Sensors able to preserve privacy and reach longer detection distances are Lidars. Their usage is becoming common in vision applications of autonomous systems as they can provide depth images with very high resolution [11], [12], but they present some critical issues. In fact, Lidars rely on light emission, and in an indoor situation are unable to be used if smoke is present such as in case of fire. They are also very expensive with respect to other types of sensors.

Radar systems are a powerful alternative to the sensors previously cited. They can perform detection at greater distances than other sensors, reaching, for example in the case of automotive Radar, distances up to 30 m in short-range configuration, or greater than 200 m for the long-range one [13]–[15]. Compared with Lidars and RGB cameras, Radar sensors thus have several advantages, including greater achievable range, being less expensive than Lidars and high-resolution RGB cameras, and can work in the presence of smoke or in general in difficult optical vision situations. Thanks to these

The research activity presented in this paper was supported by the financial program DM MiSE 5 Marzo 2018 project "*ChAALenge*" - F/180016/01-05/X43.

M. Raimondi, G. Ciattaglia, A. Nocera, L. Senigagliesi, S.Spinsante and E. Gambi are with the Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy (e-mail: {m.raimondi@staff, g.ciattaglia@staff, a.nocera@pm, l.senigagliesi@staff, s.spinsante@staff, e.gambi@staff}.univpm.it).

characteristics, Radars can be used for indoor target detection and applications such as activity recognition and AAL [16]–[19].

A. State-of-the-art

Modern commercial automotive Radar sensors widely exploit the frequency range from 77 to 81 GHz, mainly because the automotive market uses the W-Band for their Radar systems [20]. Thanks to advances in this field, there are many development boards on the market designed for automotive, which can be used also for indoor applications [20], [21]. The detection, identification, and localization of targets is a relevant research field in Radar applications. To detect a target in general and also in indoor environments, a classical approach consists of the use of Constant False Alarm Rate (CFAR) thresholds on Radar processed signals [22]. However, their performance degrades rapidly in not homogeneous environments, which represent the most common situation, so, in most cases, the CFAR algorithm fails to give a correct solution to complex identification tasks [23]–[25]. CFAR algorithms have been modified and improved over time proposing different solutions, such as Cell Averaging (CA) CFAR or Ordered Sort (OS) CFAR [26]. Other algorithms have also been developed recently, such as Comp-CFAR [27], or CFAR based on log [28]. All current CFAR algorithms perform target detection via a reference window and process the data contained therein. The reference window is usually adopted to estimate the average interference power representing the range-Doppler side. This is used to obtain the detection threshold, which should be set high enough to limit the false alarm rate to an acceptable small percentage, but using reference windows reduces the efficiency of target detection. The main purpose of the CFAR is to define a threshold and all values above will be considered as a target. Many CFAR-based techniques have been proposed with the purpose of not only revealing a target but also classifying them [29], [30]. Unfortunately, these methods are very sensible to the configuration of the CFAR and must be calibrated. A more general approach must work without calibration of the algorithm, and this is one of the tasks of the methodology proposed in this work.

With the progressive advancement in the performance of machine learning (ML) algorithms, data collected by the Radar can be processed as images and used to improve traditional Radar techniques. Experimental results [31], [32] demonstrate that the ML approach exhibits high robustness with respect to CFAR thresholds in noisy electromagnetic environments. Initially, the main ML techniques used for image classification were based on Convolutional Neural Network (CNN)s [31]. Today, real-time methods for target recognition such as YOLO, a network called the "single pass network", are preferred in conditions where the latency of the algorithm is relevant. YOLO reduces processing time compared to other ML techniques. The network was created for object recognition on images or videos, like in [33], and was later applied to Radar signals. In [34], a MIMO Radar able to exploit a bi-dimensional array is used to obtain an image similar to one obtained from an RGB sensor. Detection and classification of the target, which can be a person, a fence, or a road sign, are performed by applying a YOLO neural network using the combination of the sensors. The results show that the use of Radar helps the YOLO network in dark conditions or when the lens of the camera is dirty but the problem of privacy remains open. The obtained results also show that the Radar systems without the joint usage of the camera can reach 84% of accuracy in classification. Authors in [35] show how YOLO achieves good results on range-Doppler maps using vehicles as a target. These maps are obtained from the processing of the Radar signals and make it possible to measure the target's velocity and range distance from the sensor. Range-Doppler maps are also used in [36], where a YOLO neural network is applied for the classification of three different targets (pedestrians, vehicles, and bikes). In [37], [38], authors apply YOLO to range-angle Radar images and demonstrate the possibility of applying deep learning algorithms to high-resolution Radar sensor data, particularly in the range-angle domain. Furthermore, they show that the use of YOLO instead of CNN improves classification performance.

Range-azimuth maps are notoriously difficult to analyze because of noise, especially in indoor conditions [39]. These maps are subject to reflection problems and multipath, especially if the target of interest is a pedestrian that has a low Radar Cross Section (RCS). A method named Deep Image Prior (DIP) is proposed for denoising the range-azimuth map in [39]. This method is based on ML, but it is used to help the application of CFAR thresholds. In [40], the authors propose a method that exploits not only the range-azimuth map but also the range-Doppler map to improve the performance. Range and Doppler information is also considered in [41] to simplify detection on range-azimuth maps. In this article, Jiang et al. test the functioning of a CNN, by relating it to conventional methods. However, the data in this experiment are simulated and not tested in real environments.

B. Main work contribution

In this work, the proposed method makes use of a mmWave W-Band MIMO Radar together with a ML-based detection and localization technique for indoor target recognition and localization. By exploiting all the information from Radar signal processing (i.e., angle, Doppler, and range), it is possible to detect a moving person in an indoor scenario. The approach first uses the YOLO on range-Doppler maps, and then we apply YOLO to Doppler-azimuth maps. The information obtained from the dual use of the network will be combined to locate the pedestrian on the range-angle map. The main novelty compared to the current state of the art is the use of the three axes of the Radar data cube. Including the Doppler-azimuth map in the processing, make it possible to better detect and locate the target in the range-azimuth map. This is possible thanks to the MIMO capabilities of the Radar used, which increase the angular resolution performances. In addition, unlike CFAR, the proposed approach can be applied directly on Radar images and does not require the use of thresholds or any kind of pre-calibration.

The rest of the paper is organized as follows. Section II describes the Frequency Modulated Continuous Wave (FMCW)

Radar used and its basic operating principles. Section III introduces the main concepts concerning YOLO and the processing of the proposed method is explained. Section IV describes the experimental tests performed and the results derived from the application of the proposed algorithm. Conclusions are drawn in Section V.

II. Radar System and Signal Pre-Processing

A. FMCW Radar

The sensor used is a Texas Instruments mmWave Radar FMCW, equipped with 12 transmitters and 16 receivers [42]. A basic block scheme of an FMCW Radar is depicted in Fig. 1.

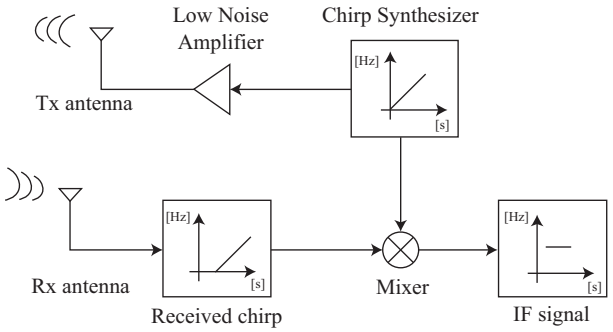


Fig. 1. Radar block scheme: the transmitted signal is generated by the Chirp synthesizer; the reflected back signal is collected by the receiver antenna and the mixer performs the mixing between them to obtain the IF signal.

Thanks to the MIMO technology, it is possible to obtain an azimuth virtual array of 86 elements with a field of view of 120°, leading to an angular resolution of 1.4°. The operating principle of the FMCW Radar is based on the transmission of chirp signals. In the case of the sensor used, the chirp is linearly modulated, and in particular, the device generates only chirps with a positive slope called "Up-Chirps". From Fig. 1 is possible to see that the transmitted signal is reflected back by any targets in the scene and received by the receiving antenna. At this point, the mixer performs the multiplication between the transmitted and the received signals which results in the intermediate frequency (IF) signal called also the beat signal. The transmitted and received signals and the IF signal are depicted in Fig. 2 for the simple case of a single target in the scene.

The chirp signal has a starting frequency f_{start} of 77 [GHz] and a stop frequency f_{stop}, which depends on the configuration of the Radar. The maximum usable Bandwidth is 4 [GHz] and the value depends on the slope of the chirp and the time t_{chirp}. The time that the transmitted chirp takes to go from the initial frequency f_{start} to the final frequency f_{stop} is called the chirp time and is indicated with t_{chirp}. The difference between the transmitted chirp and the received chirp is indicated as Δ_t. Only in the time window called $t_{overlap}$ the IF signal will be sampled by the Analog to Digital Converter (ADC). The velocity and distance of the target can be measured by processing the samples of IF signals. As mentioned above, the Radar used is equipped with MIMO technology, and to estimate the Angle of Arrival (AoA) θ it is necessary to use at least two receiver antennas.

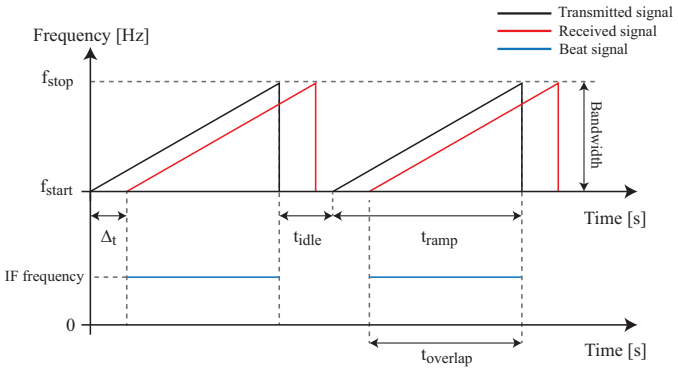


Fig. 2. Transmitted, received chirp and IF signal. In the case of only one target the IF signal is sinusoidal with only one frequency.

Each receiver antenna has a dedicated ADC that samples the related IF signal. This can be called "spatial sampling" and the angular information can be obtained from the processing of these samples. The signal transmitted by the Tx antenna is reflected by a target that has an angle θ with respect to the Radar antenna boresight and is received by the two antennas R_{x_1} and R_{x_2}; the relative signals are received from two different paths, which corresponds to a phase difference. At this point, the AoA θ can be estimated as

$$\theta = \frac{\arcsin(\Delta_\phi \lambda)}{2\pi d}, \tag{1}$$

where Δ_ϕ their phase difference between the two received signals, λ the wavelength and d the distance between the receiver antennas [43].

With the considered MIMO FMCW Radar, it is possible to calculate the position of the target with respect to the Radar position and also their relative velocity. The limits of the obtainable measurements depend on the device configuration. The configuration must be customized for the testing area where the measurements are conducted.

B. Radar signal pre-processing

The Radar system used in this work can be configured with a specific software provided by Texas Instruments, namely mmWave studio used to set the configuration parameters [44]. The main parameters to be set for this work are:

- t_{idle}: it is the time from one chirp transmission to another. Is used to restore the internal ramp generator from one transmission to another;
- t_{ramp}: chirp time duration. This parameter affects the used Radar Bandwidth;
- $f_{sampling}$: is the sampling frequency of the beat signal;
- f_{start}: is the starting frequency of the chirp;
- n_{ADC}: number of samples in each chirp;
- n_{chirp}: number of chirp in each frame;
- $t_{overlap}$: time over which the beat signal is sampled. Depends on the number of samples n_{ADC} and the sampling time;

The operating mode of the device is to group the transmissions in a frame. Each frame is composed of a certain number of chirp transmissions that can be set with the parameter

n_{chirp}. Each transmitted chirp will produce an IF signal sampled with a configurable number of samples indicated with n_{ADC}. Considering only one couple transmitter-receiver and one frame, the so-called Fast-Time/Slow-Time matrix can be obtained by placing side-by-side the sample vectors of each chirp. This process is depicted in Fig. 3.

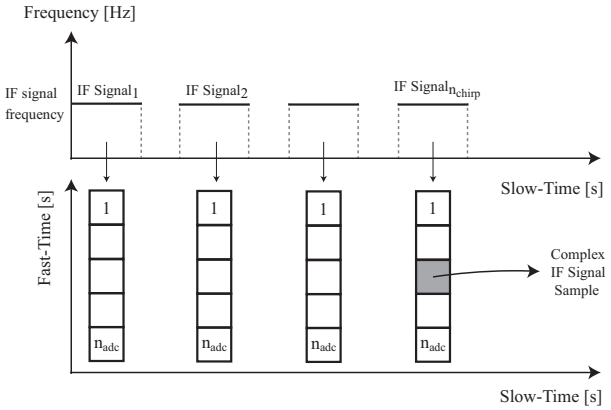


Fig. 3. Radar IF signal samples organization: Fast-Time/Slow-Time map.

The Fast-Time axis contains the samples of one chirp transmission and is composed of n_{ADC} samples, while the Slow-Time axis contains samples of different chirps and is composed of n_{chirp} elements. Extending the consideration to multiple couples transmitter-receiver it is possible to obtain a cube that is called Radar data cube, whose representation is reported in the left part of Fig. 4. The "Spatial Sampling" is the sampling along different receivers. All processing to extract the information about the target is based on this data organization. A data cube is obtained from each transmitted frame, so organizing the data in this form is an easy way to represent and manage the Radar IF signals' samples. To obtain the distance of the target from the Radar, its velocity and AoA, the most simple way is to compute a Fast Fourier Transform (FFT) along the different axes. For the purposes of this work, the computation is performed bi-dimensionally, resulting in the so-called detection maps. These maps are:

- Range-Doppler map, for the computation along the Fast-Time and the Slow-Time;
- Range-Angular map, for the computation along the Fast-Time and the Spatial Sampling;
- Doppler-Angular map, for the computation along the Slow-Time and the Spatial Sampling.

A schematic representation of how the maps are obtained from the Radar data cube is depicted in Fig. 4.

The conversion of the axis from FFT bins can be done with the classical Radar FMCW equations but with particular attention to the Doppler/Velocity axis. The transmission operating mode with the Radar is the Time Division Multiplexing (TDM), so this means that all the transmitters must transmit their signals before transmitting another chirp. Considering the couple T_{x_1} and R_{x_1}, if R_{x_1} receives the transmitted signal at the time $t = 0$, the next chirp from the same transmitter T_{x_1} will be received after $T = n_{T_x} \cdot (t_{idle} + t_{ramp})$. The about equation used for the conversion, on the basis of this

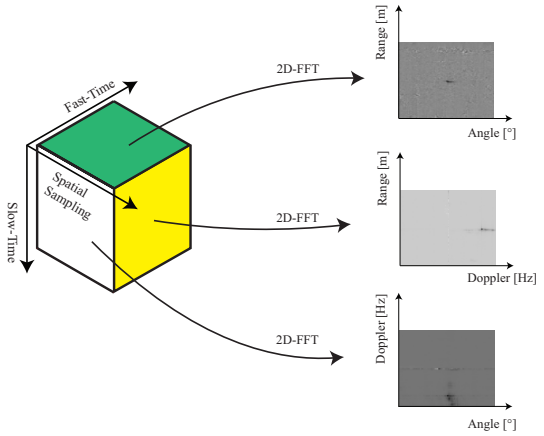


Fig. 4. Extraction of maps from the Radar data cube: range-angle (green), range-Doppler (yellow), Doppler-angle (white).

consideration of the maximum detectable velocity, is

$$v_{max} = \frac{\lambda}{4T} = \frac{\lambda}{4 \cdot n_{T_x} \cdot (t_{idle} + t_{ramp})}, \tag{2}$$

where v_{max} is the maximum detectable velocity, λ the center wavelength of the transmitted signal and n_{T_x} the number of active transmitters in the Radar configuration. The value n_{T_x} heavily affects the maximum detectable velocity and this must be taken into consideration when the configuration parameters are chosen.

III. Object Detection using YOLO

The purpose of the proposed method is to detect and localize a person within a room and to achieve this goal, is possible to exploit the YOLO network. YOLO is a detection algorithm that performs predictions on an image in a single run. It makes use of convolutional networks for the detection of multiple objects in a single image. This means that in addition to predicting an object's class, the neural network is also able to identify the position of the object inside the image. The first part of the YOLO network used to extract features is a convolutional neural network, i.e., a network mainly used for image classification. This is based on the main concepts of linear algebra such as matrices. In the first step, the image is divided into several grids of a certain size B; each grid cell will detect objects that appear within it with a certain confidence. To better understand how YOLO works, is possible to refer to Fig. 5; YOLO takes an input image and outputs a vector, which contains information about the position and class of the target to be identified. In particular, the first four elements of the vector determine the position given by the bounding box, and contain information about:

- vertex of the element at the top right of the box (x, y);
- width of the element (w);
- height of the element (h).

The remaining part of the output vector gives the probability that the subject belongs to certain classes $(p_0, \cdots, p_n)$.

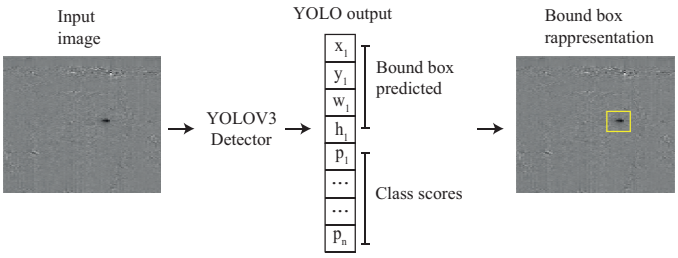


Fig. 5. Example application of the YOLO on the range-azimuth map.

Confidence is determined using the Intersection over Union (IoU) method, which is an evaluation metric used to measure the accuracy of an object detector on a particular dataset. To apply IoU, it is necessary to have:

- the ground-truth bounding boxes (i.e., the bounding boxes manually labeled in the test set which specifies where the object is in the image);
- the bounding boxes provided by the used model.

The IoU is given by the ratio between the overlap area and the union area; the overlap area is the area between the predicted bounding box and the ground-truth bounding box, while the union area corresponds to the area enclosed by both the predicted bounding box and the ground-truth bounding box. If the IoU is larger than 75% the prediction is considered good. When a new input arrives, the YOLO network estimates the position and class of the object for that input. Several bounding boxes may be generated for a single ground truth, and in order to choose which of these is the most important one, non-maximum suppression (NMS) is used, i.e., only the one with the highest value is extracted.

The YOLO network is trained on the basis of a dataset whose labels we already know or, a dataset to which we apply labels. The labels contain information about the class and the bound box. Typical parameters used when training a YOLO model include:

- *Learning Rate*: it adjusts how quickly the model updates its weights according to the gradient calculated during training;
- *Mini Batch Size*: fixed number of training examples that is less than the actual dataset;
- *Penalty Threshold*: detections that overlap by less than this value are penalized.
- *Warm-up period*: it represents the period during which the desired learning rate is to be achieved;
- *Augmentation*: it allows the model to be trained on different versions of the available data to avoid overfitting.

A. Proposed Method

The YOLO network is first applied to range-Doppler maps. Once obtained the bounding boxes, we can derive the position of the person, the assumed speed, and whether is moving away or towards the Radar. From these bounding boxes, in particular, we are interested in determining the portion of the range in which the target is located, as can be seen in Fig. 8(a), where the portion of interest is delimited by x_1 and x_2. If several bound boxes are predicted, and the case of a single pedestrian is analyzed, the one with the highest score is chosen. The process is then applied to azimuth-Doppler maps. In this case from the bounding boxes we obtain information about the portion of the angle, delimited by θ_1 and θ_2, where the target is located with respect to the Radar. An example is reported in Fig. 8(b). This way, we obtain information about the range and angle of the target, and we can combine them together in order to cut out the portion of no interest in the range-azimuth map, as done in Fig. 8(c). This allows us to clean up the range-azimuth map and improve target detection by eliminating noise on the map, due, for example, to multipath or reflections from other objects in the room. An example of what happens in practice is shown in Fig. 6. Considering a frame extracted from the Radar data cube, this can be represented using the "grey" colormap and work by eliminating the background on all sides of the Radar Data cube. In Fig. 6(a) is possible to see the range-Doppler image of a moving person. From this map, the resulting velocity of the target is about 1 $[\mathrm{m/s}]$, and approaching the Radar position. The elongated shape is due to the micro-Doppler effect [45].

As explained in Section III-A, we feed the network with the range-Doppler and the Doppler-azimuth maps to obtain the bounding boxes. To emphasize the results, we report the range-azimuth map in Fig. 6(c), which has not been processed, while Fig. 6(d) shows the post-processed image.

It is possible to observe that, by applying the proposed method, most of the noise is removed from the map. The example just discussed (Fig. 6) is based on the removal of the background. In order to generalize the proposed approach, in the following we will proceed using acquisitions without background removal. This is possible thanks to a colormap that is called "*Colorcube*". In this color scales a specific colormap is made, this is divided into equal steps and each one contains the red, green, and blue intensities of a specific color.

By changing the color scales in colorcube, even small differences in intensity are enhanced and the target can be visualized in a pronounced mode, furthermore using this colorbar is not necessary. This is not possible with other scales where there is a variation of color from maximum to minimum intensity value, or vice versa. The YOLO network is based on a CNN network, so having uniform images helps in increasing the performance, which is why this representation was adopted. An example of maps represented with "*Colorcube*" are shown in Fig. 7, where a range-Doppler image in 7(a) and an azimuth-Doppler map in 7(b) are reported, respectively. It can be seen that there are different speed scales and ranges between Fig.6 and Fig. 7 due to idle time value and chosen operating mode respectively.

B. YOLO parameters

For the tests, is considered a YOLOv3 and a SqueezeNet as backbone CNN, since it is the lightest in terms of weight and number of parameters. SqueezeNet is a pre-trained model on ImageNet [46], to which is applied layer freezing and transfer learning. Freezing a layer means that its weights cannot be changed further. The main advantage of transfer learning is that it mitigates the problem of insufficient training data limiting the number of parameters that can be updated

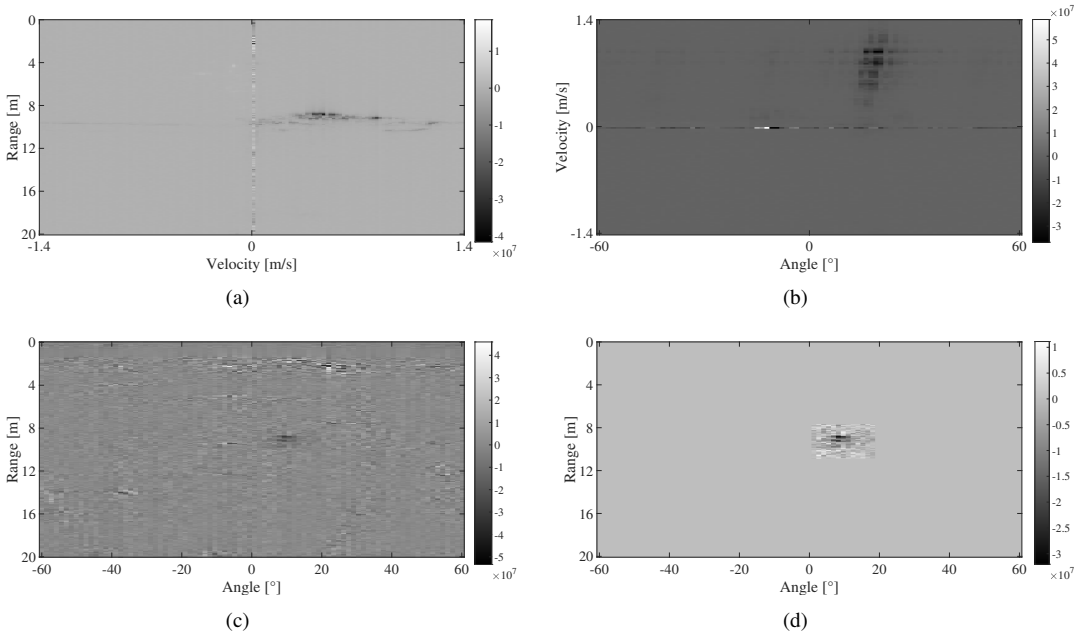


Fig. 6. Data processing for the single target case with frontal Radar: (a) range-Doppler map, (b) azimuth-Doppler map, (c) range-azimuth map pre-processing, (d) range-azimuth map after processing.

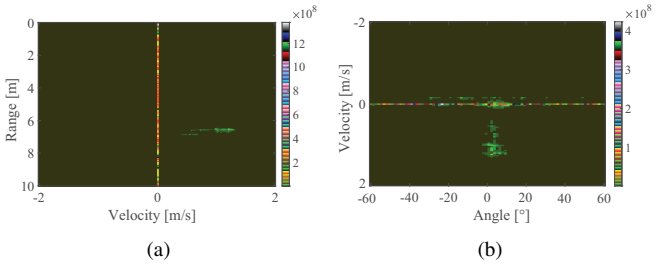


Fig. 7. Maps visualization with colorbar "*Colorcube*": (a) range-Doppler map, (b) azimuth-Doppler map.

during the training process. In this case, all the layers of the SqueezeNet are frozen and we train just two CNN-based "heads" of the YOLO starting from layer 27 to layer 34 and from layer 56 to the last layer of the SqueezeNet. This choice is made because layer 62 ("fire-9-concat") and layer 33 ("fire-5-concat") are detection network sources. The parameters chosen for the network are reported in Table I.

TABLE I
YOLOV3 PARAMETERS.

N_epochs	70
LearningRate	0.001
MiniBatchSize	8
Penalty threshold	0.5
Warm-up Period	500
Input size	227 x 227 x 3 pixel
Augmentation 1	Color jitter
Augmentation 2	Random horizontal flip
Augmentation 3	Random scaling by 10%

In order to avoid overfitting problems, the dataset is augmented before training. The augmentation process includes the application of random horizontal flipping, random x/y scaling, and finally the application of jitter color. Jitter color is a technique that allows to vary the brightness, contrast, hue, and saturation of the images.

IV. Experimental Results

A. Dataset realization

The experimental tests were conducted at the Department of Information Engineering (DII) of Università Politecnica delle Marche. The test targets are people moving within a room. Several settings and setups are considered, involving different Radar positions, different subjects, and different ways of walking, as discussed in this section. The chosen Radar parameters are shown in Table II.

TABLE II
RADAR PARAMETERS.

n_{ADC}	500
n_{chirp}	128
$f_{Bandwidth}$	3.8 [GHz]
$f_{sampling}$	20000 [kHz]
n_{frame}	78
f_{start}	77 [GHz]
t_{idle}	20 [µs]
t_{ramp}	52 [µs]
Ramp Slope	76 MHz [µs]
Range max	10 [m]

The Radar system is placed on a stand, and connected to the control computer and to the power supply. Two initial acquisitions were made to train the YOLO network: in the first, the Radar is placed at the height of 950 ± 2 [mm] above the ground, while during the second acquisition at 1200 ± 2 [mm] above the ground. The room in which the experiments were conducted is 8500 ± 2 [mm] long. The measurements are made with a DTAPE laser distance meter (model DT50). The analyzed cases involve a person walking with two walking modes during each acquisition:

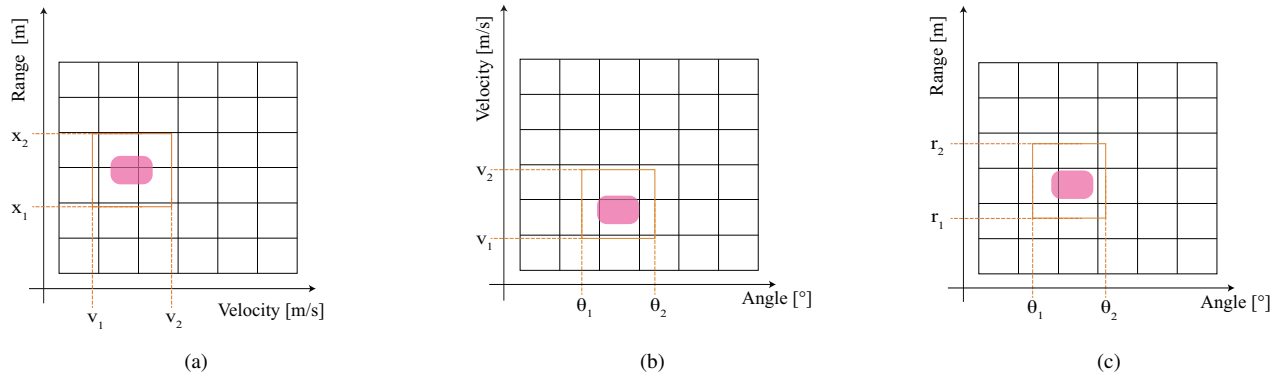


Fig. 8. Data processing: (a) detection on range-Doppler map, (b) detection on Doppler-azimuth map, (c) detection range-azimuth map.

- Linear mode, as shown in Fig. 9 (1);
- Random mode, as shown in Fig. 9 (2).

Two different subjects were involved in these training acquisitions.

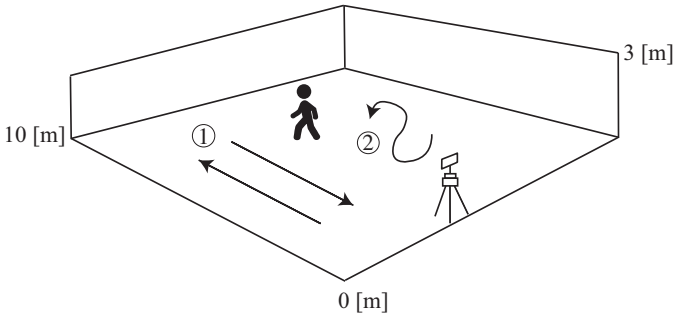


Fig. 9. Acquisition setup: linear walking mode (1), random walking mode (2).

As illustrated in Fig. 4, three different types of maps can be obtained from the Radar Data Cube. Therefore, for each acquisition obtained 78 images relating to the range-Doppler side, 78 relating to the azimuth-Doppler side and 78 relating to the range-Azimuth side. Based on the above, three different datasets were collected, from an acquisition in which the subjects move linearly and randomly; acquisitions are made considering one subject at a time. The realized datasets are used to train three different YOLOv3 Networks:

- YOLOv3 trained on images derived from range-azimuth maps;
- YOLOv3 trained on images derived from range-Doppler maps;
- YOLOv3 trained on images derived from azimuth-Doppler maps;

The first network listed is only used later to make a comparison, initially focusing on the realization of the algorithm with the other two networks.

These datasets are divided into two parts, 80% for training, 20% for validation. Before testing their functionality, the network is used on an acquisition performed under the same conditions considered during the training to verify the efficiency. The algorithm is tested on four different cases, considering other different subjects. The test set is composed of the following tests: Nella descrizione dei test va descritta meglio la differenza di setup rispetto la tabella 2. La question complex 1x o 2x è una cosa del nostro radar e sarebbe meglio metterla generica. Vanno inseriti i parametri modificati in relazione a quelli messi nella tabella. Gli altri li possiamo tralasciare essendo specifici per il nostro radar.

- Test 01: Test on a map obtained from an acquisition in which a person walks in a linear mode. The Radar is raised up to 180 [cm] and tilted by 15°.
- Test 02: The room remains unchanged, while the position and the height of the Radar vary. In addition, there is a change in the Radar configuration. With this configuration, two different acquisitions are performed, denoted in the following as 'Test02 a)' and 'Test02 b)';
- Test 03: Same setting of Tests 01 and 02, Radar positioned in front of the targets.
- Test 04: Change of environment and Radar configuration. In order to stress the algorithm more, the last acquisition is made in the hallway.

TABLE III
TEST RADAR PARAMETERS.

	Test01	Test02	Test03	Test04
n_{ADC}	500	500	500	500
n_{chirp}	128	128	128	128
$f_{Bandwidth}$	3.8 [GHz]	3.8 [GHz]	3.8 [GHz]	3.8 [GHz]
$f_{sampling}$	10000 [kHz]	10000 [kHz]	10000 [kHz]	10000 [kHz]
n_{frame}	78	78	78	78
f_{start}	77 [GHz]	77 [GHz]	77 [GHz]	77 [GHz]
t_{idle}	20 [µs]	5 [µs]	5 [µs]	5 [µs]
t_{ramp}	52 [µs]	52 [µs	52 [µs	52 [µs
Ramp Slope	76 MHz [µs]	76 MHz [µs]	76 MHz [µs]	76 MHz [µs]
Range max	10 [m]	20 [m]	20 [m]	20 [m]

B. Results and discussion

The accuracy of the network is evaluated by considering precision, recall, and average precision on the validation set. These metrics are the most widely used to evaluate the Object Detection algorithm [47]. Precision indicates the model's ability to avoid false positive predictions. Recall measures the ability of the model to avoid false negative predictions. It indicates the percentage of correctly identified positive

samples out of the total number of positive samples in the dataset.

Fig. 10 shows the trend of precision and recall related to a network trained on the range-Doppler dataset, azimuth-Doppler dataset, and range-azimuth dataset. From the figure it can be seen that the results of the network trained on range-azimuth images are not good, in fact, high precision values indicate that there are no false positives, while low recall values indicate the presence of many false negatives. While high levels of accuracy and precision are achieved in the other two networks, this means that they are able to make accurate predictions and have a good ability to correctly identify positive objects. These results suggest that the model has a good ability to generalize and is effective in the specific application for which it was trained.

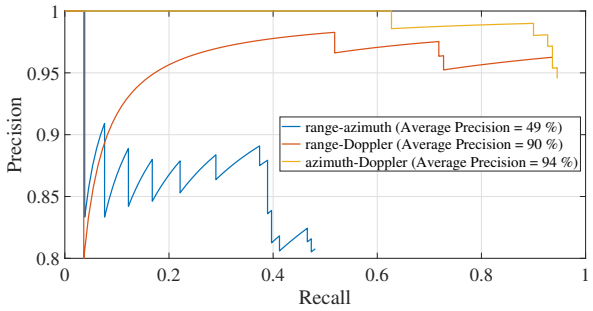


Fig. 10. YOLO performance relative to validation test: range-azimut (blue), range-Doppler (red), azimuth-Doppler (yellow).

Using the YOLO network on both sides for indirect detection on the range-azimuth map allows me to decrease the number of false negatives compared to the network applied directly on the range-azimuth maps.

To evaluate the performance of the algorithm of the test set we cannot exploit average precision because the bounding-boxes on range-azimuth are obtained indirectly using the YOLOv3 on the other maps. We then count the number of frames in which the bounding box is correctly detected. Considering that the total number of frames for all acquisitions is 78, Table IV shows the number of frames in which the bounding box is correctly detected in the range-Doppler map and the number of frames in which detection occurred in the azimuth-Doppler side in the four different tests under consideration. The last two columns refer to the number of frames in which detection is performed on the range-azimuth side, respectively, using the algorithm to cut range-azimuth maps and to the number of frames in which detection is performed.

As can be seen from the table, the application of the YOLO network on range-Doppler maps gives good results even in unfavorable situations such as a change of background. Range-azimuth maps in which only a portion of the range and not a portion of the angle is identified can be also considered as good results.

The obtained results testify to the good generalization ability of the network since the algorithm works properly even in situations that were not included in the training set initially used.

TABLE IV
NETWORK PERFORMANCE EVALUATION ON TESTS.

	range-Doppler	azimuth-Doppler	range-azimuth (indirectly)	range azimuth (directly)
Test01	78	77	77	33
Test02 a)	78	73	73	18
Test02 b)	76	63	63	25
Test03 a)	70	77	72	20
Test03 b)	78	73	73	23
Test03 c)	78	77	77	19
Test04	76	73	70	7

V. CONCLUSIONS

In this paper we have proposed an approach for target detection and localization based on the joint use of FMCW Radar and YOLO network. The results show that it is impossible to apply YOLO directly on range-azimuth maps, due to the numerous false negatives that are obtained; to increase resolution and detection accuracy, we exploited all three sides of the Radar cube, including range-Doppler, Doppler-azimuth and range-azimuth maps, withouth background removal. The focus was on generalizing the approach as much as possible, working on range-azimuth maps without pre-processing is difficult so a method is proposed that can use speed to locate the target. An important aspect of our proposed method is its superiority over Constant False Alarm Rate (CFAR) thresholds. Unlike CFAR, which requires calibration to determine the threshold and lacks classification capabilities, our YOLO-based approach inherently handles classification tasks due to its convolutional neural network architecture.

By considering experimental tests which included different Radar configurations, subjects involved and indoor environments, we have shown the feasibility of the proposed approach and proven that it is able to achieve a very good detection precision and a good generalization capability. Future works will include the application of a Kalman filter to predict the position of a moving target when the algorithm fails and skips detection in a frame. The multitarget case will also be analyzed.

REFERENCES

[1] S. Gasparrini, E. Cippitelli, S. Spinsante, and E. Gambi, “A depth-based fall detection system using a Kinect® sensor,” *Sensors*, vol. 14, no. 2, pp. 2756–2775, 2014.

[2] M.-L. Ge, E. M. Simonsick, B.-R. Dong, J. D. Kasper, and Q.-L. Xue, “Frailty, with or without cognitive impairment, is a strong predictor of recurrent falls in a US population-representative sample of older adults,” *The Journals of Gerontology: Series A*, vol. 76, no. 11, pp. e354–e360, 2021.

[3] B. Wang, Z. Zheng, and Y.-X. Guo, “Millimeter-wave frequency modulated continuous wave radar-based soft fall detection using pattern contour-confined doppler-time maps,” *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9824–9831, 2022.

[4] S. Li and X. Song, “Future frame prediction network for human fall detection in surveillance videos,” *IEEE Sensors Journal*, 2023.

[5] 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*, vol. 489, pp. 488–503, 2022.

[6] S. Saponara, A. Elhanashi, and A. Gagliardi, “Implementing a real-time, AI-based, people detection and social distancing measuring system for covid-19,” *Journal of Real-Time Image Processing*, pp. 1–11, 2021.

[7] P. K. Atrey, M. S. Kankanhalli, and A. Cavallaro, *Intelligent multimedia surveillance: current trends and research*. Springer, 2013.

[8] E. Cippitelli, S. Gasparrini, E. Gambi, and S. Spinsante, "A human activity recognition system using skeleton data from RGBD sensors," *Computational intelligence and neuroscience*, vol. 2016, 2016.

[9] M. M. Arzani, M. Fathy, A. A. Azirani, and E. Adeli, "Switching structured prediction for simple and complex human activity recognition," *IEEE transactions on cybernetics*, vol. 51, no. 12, pp. 5859–5870, 2020.

[10] S. S. Chaturvedi, L. Zhang, and X. Yuan, "Pay "attention" to adverse weather: Weather-aware attention-based object detection," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 4573–4579.

[11] N. Li, C. P. Ho, J. Xue, L. W. Lim, G. Chen, Y. H. Fu, and L. Y. T. Lee, "A progress review on solid-state lidar and nanophotonics-based lidar sensors," *Laser & Photonics Reviews*, vol. 16, no. 11, p. 2100511, 2022.

[12] Y. Li, Z. Li, Y. Wang, G. Xie, Y. Lin, W. Shen, and W. Jiang, "Improving the performance of RODnet for mmw radar target detection in dense pedestrian scene," *Mathematics*, vol. 11, no. 2, p. 361, 2023.

[13] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[14] V. Dahiya and A. Kumar, "Performance evaluation of millimeter-wave considering rain fade using nyusim model," in *2022 International Conference on Industry 4.0 Technology (I4Tech)*. IEEE, 2022, pp. 1–6.

[15] S. Muckenhuber, E. Museljic, and G. Stettinger, "Performance evaluation of a state-of-the-art automotive radar and corresponding modeling approaches based on a large labeled dataset," *Journal of Intelligent Transportation Systems*, vol. 26, no. 6, pp. 655–674, 2022.

[16] A.-K. Seifert, M. Grimmer, and A. M. Zoubir, "Doppler radar for the extraction of biomechanical parameters in gait analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 547–558, 2020.

[17] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, J. Lorandel, and O. Romain, "Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 29–41, 2019.

[18] M. Amin, *Radar for indoor monitoring: Detection, classification, and assessment*. CRC Press, 2017.

[19] L. Senigagliesi, G. Ciattaglia, A. De Santis, and E. Gambi, "People walking classification using automotive radar," *Electronics*, vol. 9, no. 4, p. 588, 2020.

[20] M. Series, "Systems characteristics of automotive radars operating in the frequency band 76–81 ghz for intelligent transport. systems applications," *Recommendation ITU-R, M*, pp. 2057–1, 2014.

[21] H. H. Meinel, "Evolving automotive radar—from the very beginnings into the future," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*. IEEE, 2014, pp. 3107–3114.

[22] H. Rohling, "Radar cfar thresholding in clutter and multiple target situations," *IEEE transactions on aerospace and electronic systems*, no. 4, pp. 608–621, 1983.

[23] R. Srinivasan, "Robust radar detection using ensemble cfar processing," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 147, no. 6, pp. 291–297, 2000.

[24] M. A. Richards, *Fundamentals of radar signal processing*. McGraw-Hill Education, 2014.

[25] J. F. Tilly, F. Weishaupt, O. Schumann, J. Dickmann, and G. Waniliek, "Road user detection on polarimetric pre-cfar radar data level," *IEEE Robotics and Automation Letters*, 2023.

[26] M. Parker, *Digital Signal Processing 101: Everything you need to know to get started*. Newnes, 2017.

[27] Y. Liu, S. Zhang, J. Suo, J. Zhang, and T. Yao, "Research on a new comprehensive CFAR (comp-CFAR) processing method," *IEEE Access*, vol. 7, pp. 19 401–19 413, 2019.

[28] A. Gouri, A. Mezache, and H. Oudira, "Radar CFAR detection in weibull clutter based on zlog (z) estimator," *Remote Sensing Letters*, vol. 11, no. 6, pp. 581–589, 2020.

[29] X. Fang, J. Li, Z. Zhang, and G. Xiao, "Fmcw-mimo radar-based pedestrian trajectory tracking under low-observable environments," *IEEE Sensors Journal*, vol. 22, no. 20, pp. 19 675–19 687, 2022.

[30] L. Wang, J. Tang, and Q. Liao, "A study on radar target detection based on deep neural networks," *IEEE Sensors Letters*, vol. 3, no. 3, pp. 1–4, 2019.

[31] ——, "A study on radar target detection based on deep neural networks," *IEEE Sensors Letters*, vol. 3, no. 3, pp. 1–4, 2019.

[32] K. Lu, Z. Qian, J. Zhu, and M. Wang, "Cascaded object detection networks for FMCW radars," *Signal, Image and Video Processing*, vol. 15, no. 8, pp. 1731–1738, 2021.

[33] S. S. Gale Bagi, B. Moshiri, H. G. Garakani, M. Crowley, and P. Mehrannia, "Real-time pedestrian detection using enhanced representations from light-weight YOLO network," in *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1, 2022, pp. 1524–1529.

[34] A. Kosuge, S. Suehiro, M. Hamada, and T. Kuroda, "mmWave-YOLO: A mmWave imaging radar-based real-time multiclass object recognition system for ADAS applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[35] L. Zhou, S. Wei, Z. Cui, and W. Ding, "YOLO-RD: A lightweight object detection network for range Doppler radar images," in *IOP Conference Series: Materials Science and Engineering*, vol. 563, no. 4. IOP Publishing, 2019, p. 042027.

[36] R. Pérez, F. Schubert, R. Rasshofer, and E. Biebl, "Deep learning radar object detection and classification for urban automotive scenarios," in *2019 Kleinheubach Conference*, 2019, pp. 1–4.

[37] W. Kim, H. Cho, J. Kim, B. Kim, and S. Lee, "YOLO-based simultaneous target detection and classification in automotive FMCW radar systems," *Sensors*, vol. 20, no. 10, p. 2897, 2020.

[38] J.-C. Kim, H.-G. Jeong, and S. Lee, "Simultaneous target classification and moving direction estimation in millimeter-wave radar system," *Sensors*, vol. 21, no. 15, p. 5228, 2021.

[39] K. Endo, K. Yamamoto, and T. Ohtsuki, "A denoising method using deep image prior to human-target detection using MIMO FMCW radar," *Sensors*, vol. 22, no. 23, p. 9401, 2022.

[40] A. Zhang, F. E. Nowruzi, and R. Laganiere, "RADDet: Range-azimuth-doppler based radar object detection for dynamic road users," in *2021 18th Conference on Robots and Vision (CRV)*. IEEE, 2021, pp. 95–102.

[41] W. Jiang, Y. Ren, Y. Liu, and J. Leng, "A method of radar target detection based on convolutional neural network," *Neural Computing and Applications*, vol. 33, pp. 9835–9847, 2021.

[42] "TIDEP-01012, imaging radar using cascaded mmWave sensor reference design," https://www.ti.com/tool/TIDEP-01012, accessed on: 2023-05-25.

[43] J. J. Lin, Y. P. Li, W. C. Hsu, and T. S. Lee, "Design of an FMCW radar baseband signal processing system for automotive application," *SpringerPlus*, vol. 5, pp. 1–16, 12 2016.

[44] "mmWave studio," https://www.ti.com/tool/MMWAVE-STUDIO, accessed on: 2023-05-25.

[45] Y. Ding, Y. Sun, G. Huang, R. Liu, X. Yu, and X. Xu, "Human target localization using doppler through-wall radar based on micro-doppler frequency estimation," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8778–8788, 2020.

[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[47] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.