Overhead-Free People Counting in mmWave Networks Using IEEE 802.11bf Passive Sensing

Tanguy Ropitault[*,†], Anirudha Sahoo[‡], Steve Blandino[*,†], Nada Golmie[‡]
[*]Associate, Communications Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA
[†]Prometheus Computing LLC, Bethesda, MD, USA
[‡]Communications Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

Email: {tanguy.ropitault, anirudha.sahoo, steve.blandino, nada.golmie}@nist.gov

***Abstract*—Accurately assessing the number of people in a room is essential for enhancing operational efficiency, safety, and sustainability, enabling applications such as smart building management, energy conservation, and emergency evacuation planning. Wi-Fi sensing, favored for its privacy-preserving capabilities and the ubiquity of Wi-Fi infrastructure, has become a popular method for such sensing tasks, including people counting. Within this context, our paper introduces a Convolutional Neural Network (CNN) model for millimeter wave (mmWave) Wi-Fi people counting, capitalizing on the IEEE 802.11bf amendment, and in particular its passive sensing framework. By evaluating Range-Doppler (RD), Azimuth-Doppler (AD), and Azimuth-Range (AR) map representations as inputs, we establish AR maps' superiority, achieving up to 98.57 % accuracy for counting up to four individuals. Our solution, free of communication overhead, also minimizes energy consumption and computational requirements, offering a scalable and efficient sensing solution for people counting, which is particularly well-suited for Internet of Things (IoT) devices with limited resources.**

***Index Terms*—Wi-Fi Sensing, mmWave, IEEE 802.11bf**

I. Introduction

The ability to accurately count the number of people in a room has significant implications for a wide range of applications, from optimizing heating, ventilation, and air conditioning systems for energy efficiency to enhancing building security by detecting unauthorized access or facilitating the deployment of first responders during emergencies. Wi-Fi sensing, increasingly popular for tasks like people counting, offers a privacy-preserving alternative to cameras or Light Detection and Ranging (LiDAR) by analyzing signal variation from environmental movement. This method capitalizes on existing Wi-Fi infrastructure, avoiding the need for specialized hardware and reducing costs and complexity.

Recent progress in Wi-Fi-based people counting has seen the integration of Machine Learning (ML) techniques to interpret Wi-Fi signal characteristics for accurate occupancy estimation. Research has particularly focused on leveraging channel state information (CSI) and adopting Doppler-based methods. For example, Liu et al. showcased a Wi-Fi radar leveraging CSI to capture amplitude and phase information, feeding into a multi-layer neural network designed for counting individuals in indoor environments [1]. Additionally, Choi et al. utilized a compact Internet of Things (IoT) ESP32 node to collect CSI data, employing statistical features from CSI amplitudes to train ML models for people counting [2]. Similarly, Hanif et al. and Sharma et al. utilized CSI for occupancy estimation, applying various ML algorithms to prove CSI's effectiveness in accurate detection [3], [4]. Doppler-based techniques, especially Range-Doppler (RD) maps and micro-Doppler, have been investigated for their ability to identify dynamic movements indicative of human presence. In research conducted by El Amine et al., a 3D Convolutional Neural Network (CNN) model was employed to analyze RD maps generated by Wi-Fi pulse Doppler radar, emphasizing the significance of Doppler signatures in people counting [5]. Pegoraro et al. [6] and Stephan et al. [7] represent the only studies focusing on millimeter wave (mmWave) band people counting, despite not being directly related to Wi-Fi technologies. Both utilize Frequency-Modulated Continuous Wave (FMCW) radar for micro-Doppler representations: Pegoraro et al. for monitoring multiple individuals using CNN, and Stephan et al. for combining radar data with camera images for ML training.

While all of the above studies demonstrate ML's effectiveness with CSI and Doppler for people counting, their reliance on specialized testbeds limits scalability and deployment across commercial-off-the-shelf devices, primarily due to the absence of standardized Wi-Fi usage in their solutions. Moreover, the requirement for high-rate dedicated sensing traffic (from 0.5 ms to 1 ms) in existing methodologies reduces communication efficiency and increases energy usage, presenting significant drawbacks for network performance. Furthermore, the reliance on FMCW radar signals in the presented mmWave solutions hinders their seamless integration into a mmWave Integrated Sensing And Communication (ISAC) framework. Additionally, angular information's potential to enhance occupancy detection has yet to be fully explored; most current studies utilize only Range-Doppler or CSI, which represents a gap in current methodologies. To bridge these gaps, our study makes the following contributions:

1) we introduce a CNN model for mmWave band people counting within the IEEE 802.11bf framework, a 802.11 amendment currently developed for Wi-Fi sensing [8],

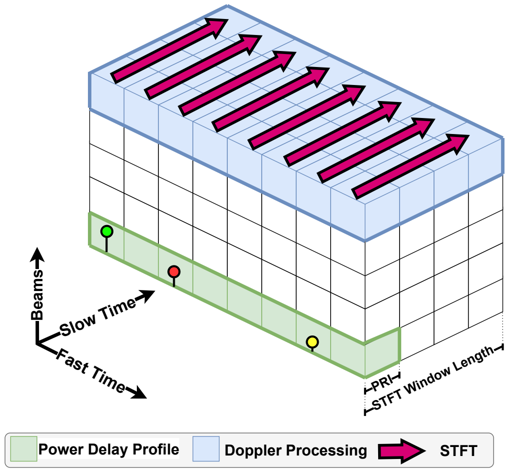


Fig. 1: Data Cube Representation in mmWave Sensings.

2) we eliminate the need for dedicated sensing traffic, i.e, or method performs passive sensing,
3) we leverage angular information, thus avoiding Doppler processing.

Our approach reduces network load, energy consumption, and computational demands, and using a standardized Wi-Fi solution like IEEE 802.11bf enhances interoperability and facilitates deployment, thus offering an efficient solution for mmWave people counting for IoT devices with limited resources.

The paper is structured as follows: Section II provides an overview of basic wireless sensing and introduces passive mmWave sensing. Section III details our framework for dataset collection, describes the dataset, and outlines our CNN model developed for the people counting task. In Section IV, we assess our model's effectiveness. In Section V, we summarize our findings and discuss the next steps in our work.

II. IEEE 802.11bf DMG passive sensing

Before exploring sensing in the mmWave band, we establish the basics of wireless sensing. This foundation is key for grasping passive sensing's mechanics within IEEE 802.11bf.

A. Wireless sensing in the mmWave band

The essence of wireless sensing is founded on the periodic transmission of pulses and the subsequent capture of echoes that bounce off objects and surfaces, enabling the estimation of their locations and movements. In the mmWave band, specific challenges arise mainly due to the significant propagation losses inherent at these frequencies. To counter these losses and sharpen signal clarity, beamforming techniques are deployed, which direct the transmitted and received signals to concentrate energy in certain directions. This fine-tuned directional control is critical not only for detailed spatial and temporal data collection but also for acquiring angular information about targets, thus significantly enhancing the richness of the sensing data.

The data collection process in mmWave sensing is three-dimensional, incorporating spatial (beams), temporal (slow-time), and echo delay (fast-time) aspects as depicted in Fig. 1. The spatial aspect is managed through beam configurations, with beamforming scanning the environment in different directions (azimuth and elevation) to increase target detection resolution. The temporal aspect is captured via slow-time sampling, with pulses transmitted and received at a predetermined pulse repetition interval (PRI), providing a timeline of the environmental dynamics. Each pulse transmission enables the Channel Impulse Response (CIR) estimation in the fast-time dimension, i.e., echo returns that map the distance to objects based on the time delay (the echo delay resolution in the fast-time dimension depends on the system's bandwidth; higher bandwidths enable finer resolution). The Power Delay Profile (PDP) in Fig. 1, visualizes the CIR, in relation to time delays, with each delay corresponding to a distinct range. This allows for an instantaneous range profile of the environment at the pulse transmission moment. The system's assembly of CIRs across various beam directions and multiple pulses results in a three-dimensional data cube. Each data slice within the slow-time dimension reflects a CIR for a certain beam direction, encapsulating a specific moment in time. For each beam, the variation of the CIR over slow time renders the temporal evolution of the spatial environment.

Following the acquisition of spatial, temporal, and fast-time data, Doppler processing can be executed on the data cube. This step leverages the Short Time Fourrier Transform (STFT) to extract the frequency components of the time-varying CIR. Through STFT analysis, velocities are determined by evaluating the Doppler shift—the frequency variation resulting from the relative motion between radar and target. This analysis enhances the dynamic portrayal of the environment by disclosing the movement of objects, with the Doppler effect providing insights into the targets' velocity. Incorporating angular information via beamforming not only improves spatial resolution but also supports precise localization and tracking of targets, offering a three-dimensional environmental perspective that amalgamates range, velocity, and angular data.

Within this framework, the selection of the PRI emerges as a critical parameter that influences the radar system's ability to discern range and velocity accurately. The maximum unambiguous range, $R_{\max} = \frac{c \cdot PRI}{2}$, and velocity, $v_{\max} = \frac{\lambda}{4 \cdot PRI}$, are governed by the PRI, where c is the speed of light and λ the signal wavelength. A longer PRI increases $R_{\max}$ by allowing more time for echo returns, enhancing range detection but reducing velocity resolution due to lower sampling of Doppler shifts. Conversely, a shorter PRI improves velocity detection $v_{\max}$ at the expense of range resolution.

B. DMG Passive sensing in IEEE 802.11bf

The IEEE 802.11bf amendment, currently under development, is expected to be ratified by June 2025. It sets guidelines for sensing tasks using Wi-Fi in both the sub-7 GHz (Wireless Local Area Network (WLAN) Sensing) and mmWave bands (Directional Multi Gigabit (DMG) Sensing), which is our focus in this article, building upon the existing IEEE 802.11ad and IEEE 802.11ay amendments within the mmWave domain. Readers interested about WLAN sensing can refer to [9]. While the standard describes an active sensing mode, which offers more flexibility through the use of specific sensing frames, this article focuses on DMG passive sensing.

Passive sensing utilizes existing Wi-Fi signals for sensing. The advantage of passive sensing is its non-intrusive nature; it operates without requiring dedicated sensing transmissions, instead utilizing ongoing communication for the sensing task, thereby eliminating sensing overhead and reducing energy consumption by solely involving the reception of signals. However, a significant drawback is the lack of control over the timing of sensing, and in particular the PRI, as it is entirely dependent on the occurrence of traffic not dedicated to sensing, making it less flexible compared to active sensing. The IEEE 802.11bf draft currently outlines the operation of passive sensing using DMG beacon frames.

The operational framework for transmitting these DMG beacons is provided by IEEE 802.11ad/ay network protocols, which structure communication into distinct channel access periods within a Beacon Interval (BI). The BI is divided into the Beacon Header Interval (BHI) for network association tasks, including beamforming, and the Data Transmission Interval (DTI) for data exchange. During the BHI, and more precisely in the Beacon Transmission Interval (BTI), the access point (AP) sequentially emits beamformed DMG beacons through its various transmit sectors. Simultaneously, stations (STAs) listen with a quasi-omnidirectional antenna gain pattern to receive these beacons. This process is crucial as it enables STAs to evaluate and select the optimal sector for communication based on the received beacon's signal quality, integrating seamlessly with the passive sensing functionality to utilize these beacon transmissions for sensing purposes.

From a sensing perspective, each beacon transmitted with a given sector represents the spatial dimension of the data cube, as discussed previously. Given the scheduling of a beacon emission once per BI, the interval for each beacon transmission in a given direction or beam aligns with the BI duration, thereby setting the PRI for sensing to the BI length.

IEEE 802.11bf does not specify how to perform the sensing processing using the DMG beacons. However, to aid STAs in effectively interpreting the beacon signals for passive sensing, IEEE 802.11bf introduces the DMG Beacon Sector Descriptor Element, a novel element field tailored for this purpose. This field is communicated between the AP and the STAs via information request/response exchanges, offering crucial details about the characteristics of each sector beam utilized during the beaconing process, including azimuth and elevation angles. This information is essential for processing and performing the sensing task.

III. Dataset Generation and CNN Model

A. Data generation Framework

Our framework comprises three distinct components: the Sensing Scenario Generator, the National Institute of Standards and Technology (NIST) Quasi-Deterministic (Q-D) Channel Realization Software [10], and the NIST ISAC Physical Layer Model (PLM) [11].

The Sensing Scenario Generator is designed to produce a specified number of scenarios, each with a certain number of moving human targets based on the Boulic model [12].

The NIST Q-D Channel Realization Software then processes these scenarios to generate the Double Directional Channel Impulse Response, employing Three-Dimensional ray tracing for accurate mmWave channel characterization and multipath components determination between the transmitter(s), target(s), and receiver(s).

The ISAC PLM, leveraging the channel data from the NIST Q-D Channel Realization Software, is an open-source simulation framework for mmWave wireless systems. It features an IEEE 802.11ay-compliant digital baseband transceiver, supporting extensive communication functions and introducing an ISAC mode, including passive sensing as anticipated in the forthcoming IEEE 802.11bf standard. This mode utilizes beamformed DMG beacon transmissions for sensing. Specifically, the ISAC PLM integrates a novel sensing processor that utilizes CSI obtained from the IEEE 802.11ay's Channel Estimation Field of the DMG beacon for environmental sensing. This processor includes a clutter removal filter to eliminate static background components, thereby isolating dynamic environmental changes indicative of moving targets. The ISAC-PLM allows us to collect the data cube specified previously for every scenario. Readers interested in exploring the features of ISAC-PLM are encouraged to refer to [13].

B. Sensing scenario

The simulation generated using our framework comprises a room measuring 12 m $\times$ 13 m $\times$ 3 m. The sensing setup is bistatic, operating in the 60 GHz band, where the transmitter (i.e., the AP in passive mode emitting beacons) is mounted on the ceiling at the midpoint of the room's y-axis (0 m, 6.5 m, 2.8 m), while the receiver (the non-AP STA) is positioned directly across on the opposite wall (12 m, 6.5 m, 2.8 m). Both the AP and the STA utilize a 4×16 element Phased Antenna Array (PAA), enhancing azimuthal steering capability. The employed codebook, i.e., the collection of sectors for beacon transmission, comprises 297 sectors, offering an azimuth resolution of 5.5 degrees and an elevation resolution of 20 degrees. The number of targets within the room varies from 0 (an empty room) to 4. Targets are assigned random initial positions and walking directions, maintaining a consistent speed of 0.90 m/s.

The sensing configuration sets the PRI at 100 ms for beacon transmission. The STFT window length is configured to 32, resulting in a Coherent Processing Interval (CPI) of 3.2 seconds (32 times the PRI duration). This setup enables Doppler analysis every 3.2 seconds, capturing target movements within these periods. We generated 10 000 instances of each target number scenarios, so our dataset size is 50 000.

For each scenario, the collected data include the PDP for each beam across both azimuth and elevation dimensions, as

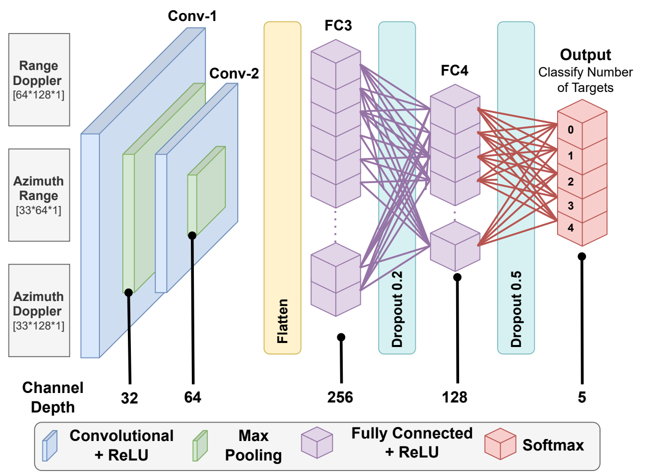


Fig. 2: Schematic of the CNN used for target classification with interchangeable input modalities.

well as slow time. To analyze this data, we employ three representations for input into our CNN: the Range-Doppler (RD), the Azimuth-Doppler (AD), and the Azimuth-Range (AR) map. The RD map is derived by performing Doppler processing across the slow time dimension and summing the results across the beam dimension. Our analysis predominantly focuses on the azimuth dimension, given its higher angular resolution resulting from the specific design of the codebook and PAA. Consequently, the AD representation is generated by aggregating data over the fast time and elevation dimensions, emphasizing azimuth information. Lastly, the AR map is created by summing over slow time and elevation dimensions but notably does not involve Doppler processing.

C. Convolutional Neural Network for Target Classification

The proposed ML architecture, depicted in Fig. 2, employs a CNN tailored for the task of classifying the number of targets within a monitored environment. The network is designed to operate with different input modalities, including RD, AD, and AR maps, to evaluate their respective performance.

The network comprises two convolutional layers followed by max-pooling, a flattening step, and two fully connected layers. Specifically, the first convolutional layer (Conv-1) consists of 32 filters of size 3×3, employing the ReLU activation function and same padding, regularized using L2 regularization to mitigate overfitting. Subsequently, a max-pooling layer with a 2×2 window reduces spatial dimensions. The second convolutional layer (Conv-2) increases the channel depth to 64 while retaining the previous layer's characteristics. Following another max-pooling operation, the feature maps are flattened to form a one-dimensional vector.

The flattened output serves as input to the first fully connected layer (FC3) with 256 neurons, coupled with ReLU activation and L2 regularization. A dropout layer with a rate of 0.2 follows to prevent overfitting. This is succeeded by a second fully connected layer (FC4) with 128 neurons and a subsequent dropout layer with a rate of 0.5 to further enhance generalization. The final layer is a softmax classifier that outputs the probability distribution over the target classes, ranging from 0 to 4 targets.

The model is compiled using the Nadam optimizer, with a learning rate empirically set to 0.0001 to leverage the convergence properties, and utilizes sparse categorical crossentropy as the loss function. The architecture is designed to harness the spatial hierarchies present in the input data, effectively learning discriminative features conducive for the classification task at hand. Training involved a dataset of 50 000 scenarios, with 10 000 per target class. The dataset was split such that 70 % of the scenarios for each target number were used for training, with an 80-20 division for training and validation, respectively. The remaining 30 % formed the test set. The number of epochs for the training was set to 200.

IV. Results and Discussion

A. Consequences of Passive Sensing

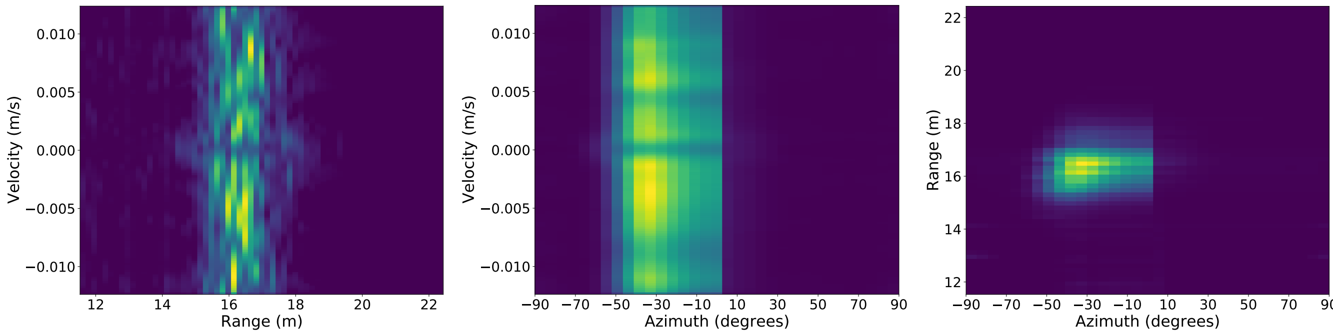
Passive sensing with DMG beacons limits PRI to 100 ms, constraining the system's ability to track fast movements accurately. At 60 GHz with a 100 ms PRI, the maximum unambiguous velocity limit is 0.025 m/s, making the accurate tracking of quick actions, such as human walking (the typical walking speed for an adult is from 1.2 m/s to 1.4 m/s), challenging due to the likelihood of aliasing in the Doppler data.

Fig. 3, displaying the three maps for a single target case, illustrates these limitations. The RD map is scattered, with aliasing across the velocity spectrum, complicating the interpretation of movement and the counting of a single target. The AD map, despite similar aliasing issues, consolidates azimuth readings into a distinct cluster (from -50 degrees to 5 degrees), indicating a single object's presence. Yet, this method's efficacy for multiple target scenarios remains questionable. Conversely, the AR map, by discarding velocity, provides a clearer indication of a singular target, enhancing the potential for accurate target counting.

ML models, particularly CNN, have shown resilience in handling aliased datasets [14]. This capability stems from their ability to learn hierarchical representations, extracting patterns and features that are robust against distortions like aliasing. In the context of RD and AD maps where aliasing is present, a CNN's architecture can still identify relevant features for successful analysis, although positionally accurate and cleaner AR maps may present fewer complexities for the model to learn and thus could offer a more straightforward and efficient solution for tasks such as people counting. This premise leads us to the next section where we explore the differences of the RD, AD, and AR maps' performance using our CNN model.

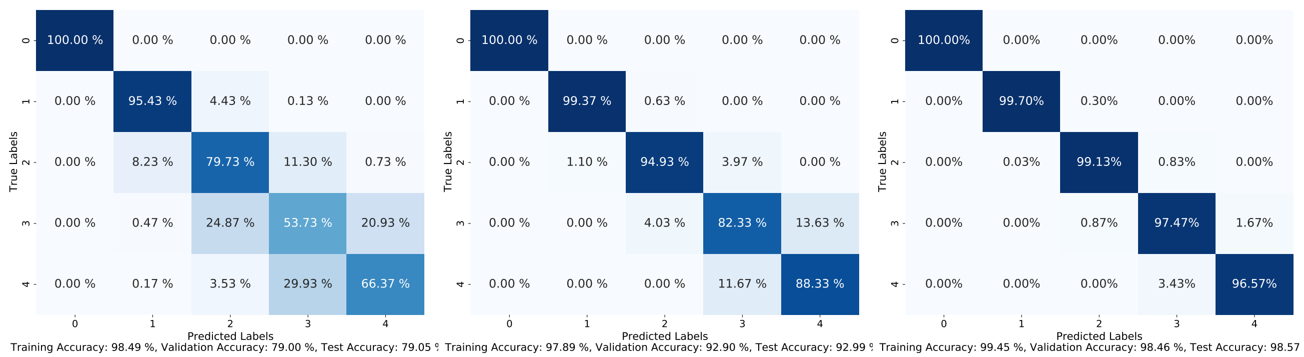
B. CNN Results

The training and validation accuracies (displayed at the bottom of Fig. 4) for the three different input types helps to understand the model ability to generalize from the training data to unseen data. For the RD input, the model achieves a high training accuracy of 98.49 %, indicating that it can



(a) Range-Doppler Map (b) Azimuth-Doppler Map (c) Azimuth-Range Map

Fig. 3: An example of the three representations used as an input of the CNN. One target is present in the room.



(a) RD Confusion Matrix. (b) AD Confusion Matrix. (c) AR Confusion Matrix.

Fig. 4: Comparative confusion matrices for the three radar map representations used in target classification.

effectively learn and fit to the training dataset. However, the validation accuracy is notably lower at 79.00 %, which suggests that the model may not be generalizing as well to new, unseen data and thus overfits.

In contrast, the AD input shows a training accuracy of 97.89 % and a validation accuracy of 92.90 %. The closer alignment of training and validation accuracies in this case suggests a better generalization capability of the model when using AD inputs.

The AR input yields the best results, with a training accuracy of 99.45 % and a validation accuracy of 98.46 %. These numbers suggest that the model, when trained with AR inputs, is highly effective in recognizing and learning patterns that are generalizable to unseen data. The slight difference between the training and validation accuracies points to a well-fitted model that is not overfitting. Overall, the AR input not only provides the best training performance but also demonstrates the greatest reliability in validation, indicating its superiority in this application both in terms of learning and generalization. It is worth mentioning that we tried CNN models using the three different inputs simultaneously but accuracy results were not better than the AR-only representation (94 %).

The confusion matrices for the three input types RD, AD, and AR, shown in Fig. 4, provide a comprehensive overview of the model's classification capabilities. As displayed in Fig. 4a, The RD maps yield an overall test accuracy of 79.05 %. The model performed well in classifying cases with no targets, achieving a 100 % success rate. The model's remarkable performance in identifying scenarios without targets can be attributed to a key feature of how ISAC-PLM operates. It eliminates the static clutter originating from the surroundings. Consequently, the only reflection power observable in the data cube emanates from Gaussian noise introduced during channel transmission. This refinement ensures that the presence of targets is the sole variable, simplifying the classification process for no-target scenarios. However, as the number of targets increases, the accuracy decreases, with a notable drop observed for four targets where the model showed a 29.93 % confusion with three targets. This suggests that the model may struggle to differentiate between higher numbers of targets when trained on RD maps. Despite this, it is noteworthy that misclassifications predominantly occurred with adjacent target classes, indicating that errors are generally within a margin of one target.

For AD maps, as seen on Fig. 4b the model's test accuracy improves to 92.99 %. Similar to RD maps, the model has perfect accuracy in predicting the absence of targets. Classification for one target is almost perfect (99.37 %), with a minor confusion of 0.63 % with two targets. The accuracy remains high for two targets but drops for three to four targets. The trend of the model confusing target numbers primarily with adjacent classes persists, which shows that when the model makes errors, the guessed number of targets is typically within one of the actual value.

The AR maps depicted on Fig. 4c produces the highest test accuracy at 98.57 %, indicating a robust performance across all target numbers. The accuracy for detecting zero targets is flawless at 100 %. The model also demonstrates high precision in distinguishing between one, two, and three targets with minimal confusion. Notably, the accuracy for four targets is remarkable at 96.57 %, with very little confusion with three targets (3.43 %). This showcases the model's strong discriminative power when using AR maps, providing evidence that AR maps enable the model to maintain high accuracy even as the number of targets increases.

In general, across all three input types, our model shows a consistent trend of accurately classifying zero targets and demonstrated a strong ability to distinguish between adjacent target classes. The rarity of the model confusing a target class with another that is more than one class apart indicates a high degree of precision in the classification task. The superiority of AR maps in accuracy stems from mitigating the Doppler aliasing effect and leveraging the mmWave band's high azimuth resolution with PAA, which enables precise target discrimination. However, RD and AD can still be used to detect the presence/absence of persons in a room, but are less able to accurately count the number of persons.

The remarkable performance achieved with AR representations in passive sensing, as evidenced by the high training, validation, and testing accuracies, has significant implications for practical applications. It indicates that passive sensing can effectively address practical tasks such as people counting without the traditional communication overheads common to sensing technologies, saving both energy and network resources, enabling addition of sensing capabilities while simultaneously meeting the critical needs of future networks: ultra-high reliability, low latency, and efficient bandwidth utilization. Furthermore, the elimination of Doppler processing with AR maps greatly lessens the computational load, rendering this strategy feasible for devices with limited processing power such as IoT devices. This computational efficiency also translates into reduced memory requirements, as people counting with AR maps necessitates analyzing data from only a single CPI, unlike micro-Doppler techniques that demand data from multiple CPIs. Additionally, the independence from Doppler effects widens the scope for detecting stationary objects or individuals, expanding the applications of passive sensing technologies.

V. Conclusion

The CNN model presented in this paper, utilized with AR maps, has proven to be effective for people counting in passive sensing contexts, outperforming the RD and AD maps representations in both accuracy and computational efficiency. Our CNN model saves energy and computational resources, an approach particularly beneficial for devices with limited processing capabilities. Additionally, it facilitates the seamless integration of sensing functionalities without imposing any communication overhead, thus ensuring compliance with future network requirements for ultra-high reliability, low latency, low-energy consumption and efficient bandwidth utilization. It is however important to acknowledge the limitations posed by the setting of PRI to the BI duration as imposed by IEEE 802.11bf, which may restrict the ability to perform certain sensing tasks. Despite this, for applications like people counting, passive sensing is a viable and effective solution. Looking ahead, we intend to further validate our algorithms with real-world data collected from a specialized measurements environment, a critical step in ensuring the robustness and real-world applicability of our models.

References

[1] S. Liu, Y. Zhao, and B. Chen, "WiCount: A deep learning approach for crowd counting using WiFi signals," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*. IEEE, 2017, pp. 967–974.

[2] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-cal: Wifi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, pp. 24 395–24 410, 2022.

[3] A. Hanif, M. Iqbal, and F. Munir, "WiSpy: Through-wall movement sensing and person counting using commodity WiFi signals," in *2018 IEEE SENSORS*. IEEE, 2018, pp. 1–4.

[4] A. Sharma, J. Li, D. Mishra, G. Batista, and A. Seneviratne, "Passive WiFi CSI sensing based machine learning framework for COVID-Safeoccupancy monitoring," in *2021 IEEE international conference on Communications workshops (ICC workshops)*. IEEE, 2021, pp. 1–6.

[5] A. El Amine and V. Guillet, "Device-free people counting using 5 ghz wi-fi radar in indoor environment with deep learning," in *2020 IEEE Globecom Workshops (GC Wkshps*. IEEE, 2020, pp. 1–6.

[6] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 2994–3009, 2021.

[7] M. Stephan, S. Hazra, A. Santra, R. Weigel, and G. Fischer, "People counting solution using an fmcw radar with knowledge distillation from camera data," in *2021 IEEE Sensors*. IEEE, 2021, pp. 1–4.

[8] "IEEE P802.11bf™/D4.0 Draft Standard for Information technology— Telecommunications and information exchange between systems Local and metropolitan area networks— Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Enhancements for Wireless LAN Sensing," 2024.

[9] T. Ropitault, C. R. C. M. da Silva, S. Blandino, A. Sahoo, N. Golmie, K. Yoon, C. Aldana, and C. Hu, "IEEE 802.11bf WLAN Sensing Procedure: Enabling the Widespread Adoption of WiFi Sensing," *IEEE Communications Standards Magazine*, Mar. 2024.

[10] NIST CTL, "Q-D Realization Software," https://github.com/wigig-tools/qd-realization, Accessed: 2024.

[11] S. Blandino, "Integrated Sensing and Communication Physical Layer Model," https://github.com/wigig-tools/isac-plm, Accessed: 2024.

[12] R. Boulic, N. Magnenat-Thalmann, and D. Thalmann, "A Global Human Walking Model with Real-Time Kinematic Personification," *Vis. Comput.*, vol. 6, no. 6, p. 344–358, nov 1990. [Online]. Available: https://doi.org/10.1007/BF01901021

[13] S. Blandino, T. Ropitault, C. R. C. M. da Silva, A. Sahoo, and N. Golmie, "IEEE 802.11bf DMG Sensing: Enabling High-Resolution mmWave Wi-Fi Sensing," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 342–355, 2023.

[14] A. H. Ribeiro and T. B. Schön, "How Convolutional Neural Networks Deal with Aliasing," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2755–2759.