Crowd Analytics with a Single mmWave Radar

Anurag Pallaprolu, Phillip Peng, Shaan Sandhu, Winston Hurst, Yasamin Mostofi
{apallaprolu,plp,s_sandhu,winstonhurst}@ucsb.edu,ymostofi@ece.ucsb.edu
University of California Santa Barbara
Santa Barbara, USA

ABSTRACT

This paper presents a novel approach for crowd analytics using a single monostatic mmWave radar. We propose a new mathematical model that infers the crowd size for dynamic and quasi-dynamic crowd behaviors. More specifically, we derive a novel closed-form mathematical expression that describes the statistical dynamics of undercounting due to crowd shadowing. This new methodical finding allows for significantly improved crowd density estimates. For spatially-patterned crowds where the mathematical solution does not extend, we then develop a Temporal Convolutional Network (TCN) which is purely trained on simulated data. We perform extensive testing over a total of 22 experiments, with up to (and including) 21 people and in 4 different areas, including indoors, and the proposed mathematical solution achieves a Mean Absolute Error (MAE) of 1.53. Lastly, we show how our framework can infer anomalies, bottlenecks, and crowd engagement level. Overall, the paper can have a significant impact on crowd management and urban planning.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Networks**; • **Mathematics of computing** → **Probability and statistics**;

KEYWORDS

mmWave Radar, Crowd Analytics, Crowd Counting, Occupancy Estimation

ACM Reference Format:
Anurag Pallaprolu, Phillip Peng, Shaan Sandhu, Winston Hurst, Yasamin Mostofi. 2024. Crowd Analytics with a Single mmWave Radar. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24), November 18–22, 2024, Washington D.C., DC, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3636534.3690664

ACM MobiCom '24, November 18–22, 2024, Washington D.C., DC, USA

ACM ISBN 979-8-4007-0489-5/24/11
https://doi.org/10.1145/3636534.3690664

1 INTRODUCTION

Crowd analytics brings insight into collective pedestrian behaviors by estimating crowd size, gauging arrival/departure rates, detecting anomalies, predicting behavior, and assessing interactions within an environment. This information plays a pivotal role for many applications. For instance, smart cities (e.g., a train station, traffic intersections) can utilize crowd analytics for applications such as traffic flow management and evacuation planning [45]. Crowd analytics is also important for safety planning, for instance during ritual or political gatherings [4, 44], as well as in the context of roadside safety and autonomous vehicles. Furthermore, crowd analytics is important for smart buildings, i.e., to optimize heating/cooling/lighting, and in retail, where it can be used to infer customers' shopping interests [12, 35]. Occupancy estimation can also be critical during a pandemic to evaluate whether crowd count limitations are being violated [26]. Learning attributes of collective behaviors, however, becomes challenging as an area gets more crowded.

On the other hand, recent years have witnessed rapid growth in the number of wirelessly-connected devices [2], including cost-effective off-the-shelf transceivers (*e.g.,* TI board AWR2243BOOST). This has motivated the use of wireless signals for tasks beyond communication, such as sensing and learning about the environment. This is particularly evident from the vision of the 6G cellular system, where Integrated Sensing and Communication (ISAC) is envisioned to be a foundational component [29]. In ISAC, transmitted mmWave signals are used for both sensing and communication. Specifically, these systems are expected to augment situational awareness in urban areas, complementing the capabilities of vision systems. This enhancement has the potential to deliver significant advantages for applications such as autonomous driving and roadside assistance systems.

In this paper, we focus on using off-the-shelf mmWave signals (*e.g.,* TI AWR2243BOOST) for crowd analytics. Most work on crowd analytics utilizes vision systems (see Section 2 on Related Work). While such systems undoubtedly play a role in this domain, their applicability may be limited by availability and potential privacy concerns, as evidenced by various surveys [11, 37]. As such, crowd analytics with off-the-shelf (or existing) RF transceivers provides a good alternative to vision systems.

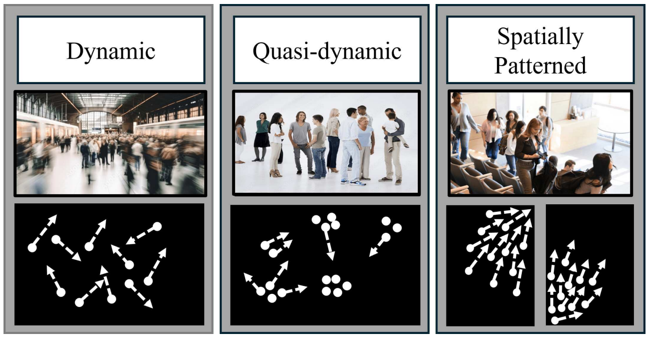

Fig. 1: Classification of crowd behaviors based on degree of correlation and extent of crowd mobility.

In recent years, great progress has been made in using RF signals to sense and learn about the environment for various applications. However, inferring collective crowd behaviors with RF signals is considerably challenging, especially as the area gets more crowded. As such, there is little work on crowd analytics with RF signals. For instance, there are a few existing works on using WiFi for crowd analytics. However, they rely on occupants crossing the line from the transmitter to the receiver to be counted. Furthermore, they suffer from the inherently low resolution of these signals.

mmWave signals, on the other hand, have been successfully used for many sensing applications. However, employing mmWave signals for crowd analytics presents a formidable and largely unexplored challenge. This complexity stems from the significant attenuation these signals experience in crowded areas, where individuals frequently obstruct one another, consequently diminishing the reflected signal's ability to convey information about obstructed individuals back to the transceiver. In other words, the total number of observed individuals can be far less than the true count due to blockage by the rest of the crowd, a phenomenon we describe as **crowd shadowing**. Section 2 provides a comprehensive review of the state of the art in crowd analytics.

In this paper, we provide a new foundation for using off-the-shelf mmWave signals for crowd analytics. We next discuss different crowd behaviors of interest to this paper.

Different Crowd Behaviors

Individuals in a crowd can utilize the space in many different ways. This has been studied extensively, and there are underlying crowd behaviors that have been observed and categorized in the literature [44, 56]. Here, we summarize three that are relevant to this paper, as shown in Fig. 1.

Dynamic Crowds: Individuals in a dynamic crowd are characterized by constant motion and uncorrelated movement patterns. They typically do not stop that often and traverse the area in an uncorrelated manner. Crowds at a train station or a public square fall in this category.

Quasi-dynamic Crowds: A quasi-dynamic crowd refers to a congregation of individuals exhibiting some degree of movement or fluctuation in their spatial arrangement, albeit at a slower, or, more irregular pace compared to dynamic crowds. Changes in spatial distribution tend to occur over longer time intervals. Examples include gatherings such as a mingling where individuals traverse the area, stopping occasionally to socialize or explore the environment.

Spatially-patterned Crowds: Spatially-patterned crowds move in a more structured way than dynamic or quasi-dynamic crowds, so that trajectories demonstrate greater spatiotemporal correlation. For instance, entering a classroom and finding a seat, *i.e*, a source-absorption pattern, falls in this category, where the traversed paths are predictable.

We next summarize our key contributions:

Statement of Contributions:

a) We propose a new mathematical model that infers the crowd size with mmWave signals. More specifically, we derive a novel closed-form mathematical expression describing the statistical dynamics of undercounting due to crowd shadowing. Our approach mathematically finds the probability of any given undercounting, i.e., the probability of observing k individuals given $N \geq k$ are present. By deriving a PDF that models the occurrence for all such k values, we then infer the total population count by comparing this theoretical PDF against the histogram of the observed data. This new methodical finding allows for significantly improved crowd density estimates. For instance, our mathematical characterization enables counting crowds of up to (and including) 21 people for both dynamic and quasi-dynamic crowds, resulting in a Mean Absolute Error (MAE) of 1.53 over 17 experiments in 4 different areas, with an emphasis on larger crowds.

b) While well-suited for dynamic and quasi-dynamic crowds, which already cover many scenarios, the proposed mathematical model does not extend to spatially-patterned crowds. Thus, we further develop a simple convolutional neural network (CNN), capable of learning temporal correlation, which is solely trained on a small set of simulated data. This pipeline can then provide crowd analytics for a larger set of crowd behaviors, including spatially-patterned crowds.

c) We extensively validate the proposed theories and algorithms with several experiments. When processing real data, we further propose a simple yet novel feedback-based approach to declare human presence, which properly strikes a balance between sensitivity and robustness. Overall, we test with a total of 22 experiments in 4 different areas, including indoors, and encompassing the three broad categories of crowd behavior shown in Fig. 1. Our experiments involve up to (and including) 21 people and are generally geared towards larger crowds, as mmWave signals can easily track a small number of people. Overall, our proposed mathematical model results in an MAE of 1.53 over all the experiments.

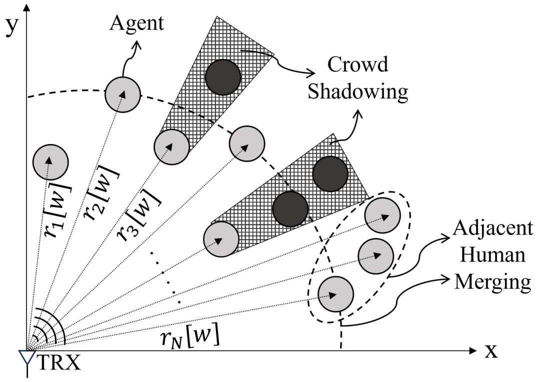

Fig. 2: A mmWave transceiver placed at origin emits chirps that are scattered by a crowd of moving agents.

d) In addition to crowd counting, we show how our framework can also infer crowd anomaly (*i.e.,* abnormal crowd behavior), as well as crowd bottlenecks, validated with a number of experiments. Lastly, we introduce a metric that well-reflects the level of engagement of the crowd with the space (e.g., to stop and look at a landmark, or to socialize).

Overall, the paper lays a comprehensive foundation for crowd analytics with mmWave signals, via 1) a novel mathematical solution with a broad applicability, 2) a complementary machine-learning pipeline (trained on a small simulated dataset) for spatially-correlated crowds, where mathematical modeling does not extend, and 3) extensive experimentation with crowd counting, in addition to anomaly detection, bottleneck inference, and space interaction analysis.

2 RELATED WORK

Occupancy estimation via a variety of sensing modalities has attracted significant attention over the past several decades due to the broad set of related applications in domains ranging from commerce to public safety. Vision-based methods have benefited from the central place that image classification has long held in the development of machine learning, which has produced a number of sophisticated approaches for crowd analytics [10, 31, 43]. However, the use of vision systems in many common settings raises serious privacy and security concerns, promoting an attitude of cautious skepticism [11, 37], as we reviewed earlier.

In the area of RF sensing, researchers have also explored using RF signals for crowd analytics, albeit to a lesser extent than other RF sensing applications, and mainly with WiFi [13–16, 26, 57]. However, due to the poor resolution of WiFi, other assumptions were needed, such as requiring that occupants cross the link between the TX and RX to be counted [14, 57]. Other WiFi-based efforts have focused on counting a seated crowd [26].

Recently, mmWave systems have risen as an industry contestant for next-generation sensing [58]. As with WiFi, mmWave radars avoid privacy concerns present in vision systems, and they have proven effective in functioning within non-ideal environments such as through fog [19], offering an advantage over other sensing modalities, such as vision or LiDAR systems [54]. However, prior work on human motion sensing with mmWave radars has been predominantly restricted to identification and tracking of a small number of individuals [9, 24, 30, 38]. Similarly, while prior work has examined occupancy estimation, the maximum considered crowd size is relatively small, *i.e.,* 6 or less [6, 22, 23, 30, 41, 42, 53]. An exception is [27], in which the maximum reported crowd size is 10. However, in that work, all experiments are conducted in the same area, and minimal information is given on the clustering method and other aspects of the pipeline. Furthermore, much of the existing work on mmWave occupancy estimation utilizes radar boards with a wide field of view [23], a network of multiple radars [8], or a physical setup with the radar placed at an ideal vantage point [22], to circumvent the problem of crowd shadowing introduced in Sec. 1, while still only addressing lower counts.

In this paper, we propose a new foundation for crowd counting utilizing only a single commodity mmWave radar. Our key observation is that while several agents may not be observable to the radar board at any point in time, resulting in severe undercounting, the statistics of the undercounting carries crucial information on the crowd size, for which we then propose a new mathematical model. This novel solution effectively estimates the total number of people for both dynamic and quasi-dynamic categories, as we show with extensive experimentation with crowds of up to (and including) 21 people across 4 different areas. We additionally design a simple convolutional neural network to handle spatially-patterned crowds, while solely training it on simulated data, and further validating it experimentally. Finally, we show how our foundation can infer other crowd attributes, e.g., crowd anomalies, bottlenecks, and crowd engagement level.

3 PROBLEM STATEMENT

Consider the monostatic sensing scenario shown in Fig. 2, where a fixed mmWave transceiver (TRX) located at the origin of the coordinate system emits FMCW pulses (chirps) that interact with a crowd of moving agents. We define $\mathbf{r}[w] = [r_1[w], ..., r_N[w]\,]$ as the vector of radial distances between each agent in the crowd and the TRX at a discrete time step, w. In this paper, we aim to estimate the size of the crowd, N, with a single TRX, given noisy and incomplete observations of $\mathbf{r}[w]$ over a fixed time horizon. Denote the observation of $\mathbf{r}[w]$ as $\bar{\mathbf{r}}[w]$. We next discuss two major roadblocks to solving this problem.

Challenges in Estimating N **from** $\bar{\mathbf{r}}[w]$**:** Due to the quasi-optical nature of mmWave radiation [8, 9, 32], the line-of-sight (LOS) path for a single TRX is crucial for monostatic

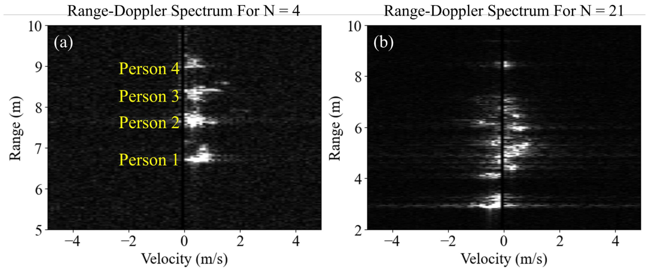

Fig. 3: Range-Doppler spectra for $N = 4$ and $N = 21$ person crowds: (a) Clear separability of clusters allows efficient occupancy estimation of $N = 4$ person crowd (b) Estimation for $N = 21$ person crowd, however, becomes considerably challenging.

sensing. As such, the likelihood of a direct path to all agents in a crowd decreases as N grows, leading to undercounting. We next identify two key culprits that can result in undercounting and a third that produces spurious entries in $\bar{\mathbf{r}}$.

1) **Crowd Shadowing Effect:** As an area gets more crowded, there is a higher chance that individuals block each other more frequently. In this paper, we refer to this recurring blockage as the **Crowd Shadowing Effect**, which results in severe undercounting and is quite challenging to address.

2) **Adjacent Human Merging:** If two individuals are too close to each other (either next to each other or at close enough ranges), they may be detected as one, resulting in undercounting. Thus, $\bar{\mathbf{r}}[w]$ may include the radial information of only a random subset of the agents.

3) **Background Noise:** The TRX is also more sensitive to background noise due to the lower wavelength of mmWave radiation, leading to overcounting. More specifically, effects such as secondary reflections off of static objects [9, 20] and environmental motion (e.g., wind-blown foliage [40]) may produce entries in $\bar{\mathbf{r}}$ that do not correspond to any agent.

To summarize, our goal is to design a robust occupancy analytic inference system using only noisy and incomplete observations of the radial information of moving agents, $\bar{\mathbf{r}}[w]$. We next provide a brief primer on the traditional range-Doppler analysis and then introduce a different representation that helps us capture fine-grained human activity across the sensing range. We leverage this representation in the later sections to develop a system for occupancy estimation of large crowds, using only a single mmWave radar.

4 FMCW RADAR SIGNAL PROCESSING

We start this section with a brief primer on FMCW signaling and the traditional range-Doppler analysis [46, 55]. We subsequently leverage this foundation to introduce our Phase Spectral Bandwidth Modeling approach, which offers a suitable framework for crowd analysis.

4.1 Range-Doppler Spectrum

Consider a monostatic mmWave FMCW TRX which periodically transmits a chirp sinusoid whose frequency grows linearly from f_0. Mathematically, $f_{TX}(t) = f_0 + St, 0 \le t < T_c$, where $S = B/T_c$ denotes the chirp slope, B is the bandwidth and T_c denotes the chirp duration. The complex baseband signal of the m^{th} chirp is then given by $s_{TX}(m,t) = \sqrt{P_{TX}}\exp(j\{2\pi f_0 t + \pi S t^2\})$, where P_{TX} denotes the TX power. Consider an object located at a distance $d[m]$ from the TRX, moving with a radial speed of v_0. The received signal after scattering is then given by $s_{RX}(m,t) = s_{TX}(m, t-\tau_d)/d[m]^2$, where $\tau_d = 2d[m]/c$ is the round-trip time-of-flight, and s_{RX} is mixed with s_{TX} to yield the IF signal [39, 46],

$$\begin{aligned} s_{IF}(m,t) &= s^*_{RX}(m,t)s_{TX}(m,t) \\ &\approx \frac{\sqrt{P_{TX}}}{d[m]^2}\exp\left(j\frac{4\pi}{c}\left(f_0(d[m-1]+v_0T_c)+t(Sd[m-1])\right)\right). \quad (1) \end{aligned}$$

Let $s_{IF}[m,n]$ denote the discretized IF signal, with T_r denoting the discretization time step, $N_r = T_c/T_r$ denoting the number of discrete range bins, and $T_f = N_c T_c$ representing the considered time duration. By performing a 2D DFT on $s_{IF}[m,n]$, we then obtain $Q_{IF}[v,r] = \mathcal{F}_{m,n}\{s_{IF}[m,n]\}$ as the traditional range-Doppler spectrum[1]:

$$|Q_{IF}[v,r]| = \frac{\sqrt{P_{TX}}}{d_0^2} D_{N_c}\left(\frac{v_0}{\Delta v} - v\right) D_{N_r}\left(\frac{d_0}{\Delta d} - r\right), \quad (2)$$

where $D_K(x) = \sum_{k=0}^{K-1}\exp\left(\frac{j2\pi kx}{K}\right)$ is the Dirichlet kernel, $\Delta d = c/2B$ is the range resolution and $\Delta v = c/2f_0T_f$ is the velocity resolution. The range and velocity of an object can then be estimated by locating the peaks of the spectrum.

Fig. 3 (a) shows a sample range-Doppler map for an $N = 4$ person experiment. Given the low number of occupants and their sufficient separation, we are able to accurately estimate occupancy through the identification of well-defined peaks. On the other hand, Fig. 3 (b) shows a range-Doppler map for an $N = 21$ person crowd. As can be seen, the information pertaining to the subjects is non-resolvable, due to detrimental factors such as crowd shadowing, coupled with inherent resolution constraints. Consequently, delivering accurate crowd analytics poses a significant challenge in this case, which is the main motivation for the work of this paper.

We next set forth a different representation that instead considers a 1D indicator of motion across the sensing range. This framework not only better illuminates the impact of factors such as blockage and resolution limitations, but also provides a suitable starting point for our proposed crowd analytics methodology.

[1] We ignore the variations in $1/d[m]^2$ over T_f and assume $d[m] \approx d_0$.

4.2 Phase-Bandwidth Analysis

In this paper, we propose to use the bandwidth of the wrapped phase as a fine-grained detector of the non-blocked motion in a dynamic scene. Traditional approaches rely on Doppler spectrum for distinguishing people from stationary objects, but in quasi-dynamic scenarios, relying solely on speed may result in false negative detections, as people may remain relatively still. By examining the bandwidth of velocity signals, even minor movements like breathing or fidgeting can be detected sensitively, enhancing human detection even at low motion speeds. We next outline a brief derivation and introduce the resulting Human Trace Map as an effective representation of the crowd's visible motion, which sets the stage for our proposed mathematical framework.

Wrapped Phase Spectrum: Given a frame of chirp receptions $s_{IF}[m,n], m = 0,1,...,N_c-1$, we begin by computing the FFT across n to obtain the Range-FFT:

$$R_{IF}[m,r] = \sum_{n=0}^{N_r-1} s_{IF}[m,n] \exp\left(-j\frac{2\pi nr}{N_r}\right).$$

Unlike the traditional Doppler-FFT, we instead perform the Short-Time Fourier Transform (STFT) of $\text{Arg}(R_{IF}[m,r])$ over m, where $\text{Arg}(\beta e^{j\phi}) = \phi \bmod 2\pi$ denotes the wrapped phase of a complex number. We motivate this by observing that $\text{Arg}(\beta e^{j\phi})$ exhibits a jump discontinuity for every $\phi \geq 2K\pi, K \in \mathbb{Z}$, thus providing a higher sensitivity to changes in ϕ, compared to $\beta e^{j\phi}$. Using a moving window of W chirps and a shift of α chirps, we have the following STFT:

$$\Phi_{IF}[q,w,r] = \sum_{i=w}^{w+W} \text{Arg}(R_{IF}[i,r]) \exp\left(-j\frac{2\pi qi}{W}\right). \quad (3)$$

Human Trace Map: For each point (r_0, w_0) in space and time, we then obtain $\Phi_{IF}[q, w = w_0, r = r_0]$, which is a sensitive indicator of visible motion spectrum at that coordinate. However, Φ_{IF} is a tensor of rank 3, which complicates our subsequent analysis and visualization. We thus collapse the $q-$dimension by computing the bandwidth of $\Phi_{IF}[q, w = w_0, r = r_0]$ for each (r_0, w_0), which captures the intensity of motion at that coordinate. More specifically, we introduce the observed **Human Trace Map**, $\mathcal{H}$:

$$\mathcal{H}[w,r] = \Gamma_q(\Phi_{IF}[q,w,r]), \quad (4)$$

with $0 \leq w \leq \lfloor (N_c - W)/\alpha \rfloor$, $0 \leq r \leq N_r - 1$, and $\Gamma_q(f(q)) \in \mathbb{R}$ is a measure of the bandwidth of $f(q)$.

Fig. 4 shows a 25s sample of $\mathcal{H}$ (normalized by the window energy across depth) for an $N = 11$ person crowd. The figure shows the range of the visible individuals as a function of time. Furthermore, it explicitly highlights the underlying challenges in extracting crowd analytics as the size of the crowd increases, namely the Crowd Shadowing Effect and Adjacent Human Merging. In the subsequent sections, we then propose a new foundation for crowd analytics that can efficiently infer crowd information from $\mathcal{H}$, despite the underlying challenges. In order to simply indicate the presence/absence of inferred human motion, we then translate $\mathcal{H}$ to a binary version to generate $\mathcal{H}_{\text{b,vis}}[w,r]$, which we refer to as the observed Binary Trace Map.

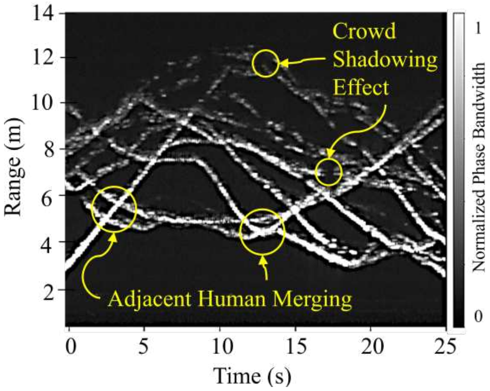

Fig. 4: Human Trace Map ($\mathcal{H}$) of an $N = 11$ sized crowd showing inferred range of visible individuals as a function of time, while also highlighting challenges such as Crowd Shadowing Effect and Adjacent Human Merging, which results in undercounting for large crowds.

Binary Trace Map: In this paper, observed Binary Trace Map refers to $\mathcal{H}_{\text{b,vis}}$, a binary map where, $\mathcal{H}_{\text{b,vis}}[w,r] = 1$ indicates that the presence of human is inferred for range r at time window w, while zero indicates otherwise. We emphasize that $\mathcal{H}_{\text{b,vis}}$ can only capture the presence for visible non-blocked motion. As such, it serves as our starting point for developing our proposed methodology for crowd analytics. Sec. 7 presents our novel and efficient algorithm for this binarization, based on a simple feedback design that strikes a balance between sensitivity and robustness.

We now formally pose our problem statement.

Problem Statement: Consider an area where a mmWave monostatic transceiver makes measurements in the vicinity of a crowd. Given the inferred Binary Trace Map, $\mathcal{H}_{\text{b,vis}}$, we are interested in finding the true crowd count, N. We note that the crowd count can be time-varying.

The aforementioned problem is considerably challenging, as severe undercounting can happen, for instance, due to recurring crowd shadowing.

Baseline Solution: A first-attempt solution would be to take the time average of the number of visible participants as our baseline for occupancy estimation, given by $\bar{K} = \mathbb{E}_w\{\sum_{r=0}^{r=N_r} \mathcal{H}_{\text{b,vis}}[w,r]\}$. While this can work for a small number of people, it results in severe undercounting as the size of the crowd increases. For instance, a crowd of 18 will be counted as 12 or 11 (see table in Fig. 11), necessitating a more fundamental approach, which is the main motivation for the proposed work of the subsequent sections.

5 A MATHEMATICAL SOLUTION FOR DYNAMIC/QUASI-DYNAMIC CROWDS

In this section, we propose a new mathematical foundation for estimating the true size of the crowd for the dynamic and quasi-dynamic cases (see Sec. 1, or Fig. 1). As discussed in Sec. 1, these scenarios have little spatial bias if observed over a sufficient period of time. Consider $\mathcal{H}_{\text{b,vis}}$. As discussed earlier, a column of $\mathcal{H}_{\text{b,vis}}$ indicates the ranges at which the corresponding human targets became visible at that time. As the size of the crowd, N, increases, the number of visible agents will be further from the true count due to the greater likelihood of crowd shadowing, leading to undercounting. Denote $n[w]$ as the number of visible individuals at time window w: $n[w] = \sum_{r=1}^{r=N_r} \mathcal{H}_{\text{b,vis}}[w, r]$. The histogram of $n[w]$ gives us the experimental density function representing how often a given number of people were observable throughout the observation period. We next show a novel mathematical approach which characterizes the probability mass function of $n[w]$ mathematically, for any given crowd size. By comparing this with the experimental histogram, we can estimate the true crowd size as the total number of people that minimizes the difference between the two. We next set forth the details of the proposed approach.

Consider the scenario shown in Fig. 2, where our goal is to estimate the crowd size, with the individuals modeled as discs of radius ρ. Without loss of generality, we restrict the crowd to a quadrant. We further assume no spatial biases, i.e., each agent is equally likely to be at any location in space at a given time. As discussed earlier, this is a very good model for dynamic crowds, and it serves as a good approximation for many quasi-dynamic cases, as we shall see. We then aim to evaluate the likelihood that exactly $n[w] = K$ agents are visible given a crowd of size N. To this end, we first derive the following lemma to establish the probability that a single agent is visible in a dynamic crowd of size N.

Lemma 5.1. *Consider a dynamic/quasi-dynamic crowd of N independent agents, each a disc of radius ρ, that are uniformly located in the area of Fig. 5 (a), with a maximum sensing depth r_{max}. Denote $\mathcal{V}$ as the event "an arbitrary agent is visible to the TRX" and $A = \pi r_{max}^2/4$. We then have the following for the probability of $\mathcal{V}$ given N:*

$$\mathbb{P}(\mathcal{V}|N) = 2\Big(A^{N-1}(r_{max}^2 - \rho^2)(N(N+1))\rho^2\Big)^{-1} \tag{5}$$
$$\times \left(A^{N+1} - (A - \rho\sqrt{r_{max}^2 - \rho^2})^N (N\rho\sqrt{r_{max}^2 - \rho^2} + A)\right).$$

Proof. Denote the agent being tested for visibility by D. We first find the probability that D is present at the depth r.

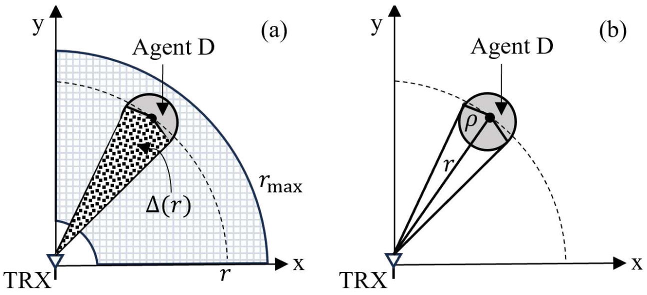

Fig. 5: Agent with radius ρ and at depth r is visible if the region $\Delta(r)$ contains no other agents.

We can easily confirm the following geometrically:

$$\mathbb{P}_R(r) = 2r\left(r_{\max}^2 - \rho^2\right)^{-1},\ \rho \le r \le r_{\max}. \tag{6}$$

We next condition on D's existence at depth r to evaluate the probability of its visibility. Consider the geometry of Fig. 5 (a), wherein we denote $\Delta(r)$ as the area in front of agent D that must not be occupied by any other agent, in order for D to be visible. From Fig. 5 (b), we can then see that $\Delta(r) = \rho\sqrt{r^2 - \rho^2}$. If another agent is outside of $\Delta(r)$, the chance that they block D is very small.[2] Consequently, for D to be visible, the rest of the $N - 1$ agents can lie anywhere in the region except over $\Delta(r)$, as that would entail crowd shadowing. Mathematically, we have the following expression

$$\mathbb{P}(\mathcal{V}|N, r) = A^{-(N-1)}\big(A - \Delta(r)\big)^{N-1}. \tag{7}$$

Combining Eq. 6 and Eq. 7, we then have the following for the probability that D is visible in a crowd of size N

$$\begin{aligned}\mathbb{P}(\mathcal{V}|N) &= \int_{\rho}^{r_{\max}} \mathbb{P}_R(r)\mathbb{P}(\mathcal{V}|N, r)dr \\ &= \int_{\rho}^{r_{\max}} 2r\left(r_{\max}^2 - \rho^2\right)^{-1} A^{-(N-1)}\big(A - \Delta(r)\big)^{N-1}dr \\ &= 2\big(A^{N-1}(r_{\max}^2 - \rho^2)\big)^{-1}\int_{\rho}^{r_{\max}} r(A - \rho\sqrt{r^2 - \rho^2})^{N-1}dr.\end{aligned}$$

By substituting $A - \rho\sqrt{r^2 - \rho^2} = u$, we can simplify the integral to the following expression

$$\mathbb{P}(\mathcal{V}|N) = 2\big(A^{N-1}\rho^2(r_{\max}^2 - \rho^2)\big)^{-1}\int_{A}^{\Omega} u^{N-1}(u - A)du,$$

where $\Omega = A - \rho\sqrt{r_{\max}^2 - \rho^2}$. The integrand is now a polynomial and can be directly evaluated to obtain Eq. 5. □

To summarize, Lemma 5.1 presents the likelihood of an arbitrary agent being visible in a crowd of N agents. As we consider the scenario of a dynamic/quasi-dynamic crowd, we assume that the locations of the N agents are independent.

[2] Note that this chance is not zero as other agents get close to the boundary of $\Delta(r)$. However, it is small, thus approximated by zero in our derivations.

We can then estimate the probability of exactly K agents being simultaneously visible, given N total agents, as a binomial distribution over K, as follows:

Theorem 5.2. *Consider a crowd of N independent agents, each a disc of radius ρ, that are uniformly located in the area of Fig. 5 (a), with a maximum sensing depth r_{max}. Let $A = \pi r_{max}^2/4$. Then, assuming independent visibility, the probability of exactly K agents being visible in the region is given by*

$$\mathbb{P}_a(K \mid N) = \binom{N}{K} \mathbb{P}(\mathcal{V}|N)^K \times (1 - \mathbb{P}(\mathcal{V}|N))^{N-K}. \quad (8)$$

Let $\mathbb{P}_e(K = \kappa)$ denote the empirical density given by the histogram of observed $n[w]$. We then estimate the true occupancy of a dynamic crowd by minimizing the Kullback-Leibler divergence between the analytical density of Eq. 8 and $\mathbb{P}_e$, over $\mathbb{N} = \{1, 2, .., N_{\max}\}$:

$$N^* = \underset{N \in \mathbb{N}}{\arg\min}\, D_{KL}\Big(\mathbb{P}_e(K = \kappa) \mid\mid \mathbb{P}_a(K = \kappa \mid N)\Big). \quad (9)$$

As we shall see, the proposed approach estimates the crowd size well for dynamic and quasi-dynamic cases. However, as the spatial biases get stronger in other scenarios, it may not serve as a good model. We next propose an ML-based approach that directly leverages $\mathcal{H}_{\text{b,vis}}$ for estimating N, complementing our analytical framework in cases such as sink/source of Fig. 14, where the spatial independence assumption is no longer a good approximation.

6 TEMPORAL NEURAL NETWORK

To estimate the true size, N, of a crowd, the analytical approach of Sec. 5 assumes that the visible count, $n[w]$, and observed depths, $\bar{\mathbf{r}}[w] = [r_1[w], ..., r_{n[w]}[w]]$, are independent across all w. More specifically, this approach ignores the temporal correlation among columns of $\mathcal{H}_{\text{b,vis}}$, which carries information about crowd motion. As an extension of Sec. 5, we instead directly utilize a contiguous block of columns from $\mathcal{H}_{\text{b,vis}}$ as an input for occupancy estimation. Such a pipeline not only estimates the size of a spatially-patterned crowd (see Fig. 1), but also demonstrates robustness to process noise that is spatiotemporally separable. We thus propose a simple Temporal Convolutional Network (TCN), trained purely on synthetic $\mathcal{H}_{\text{b,vis}}$ data generated by a lightweight crowd simulator, which we describe next.

Assembling a Simulation-based Dataset: Recall that the Binary Trace Map $\mathcal{H}_{\text{b,vis}}$, is a representation of human presence at a certain sensing depth when observed at a given time window. In Sec. 7, we show how to generate high-quality $\mathcal{H}_{\text{b,vis}}$ from $\mathcal{H}$ (Fig. 8) using our novel binarization algorithm to uncover depth-occupancy. Using the geometric model presented in Fig. 5 (a), we propose the generation of synthetic Binary Trace Maps $\mathcal{H}^{\text{syn}}_{\text{b,vis}}$, whose entries correspond to the depth-occupancy of visible agents in a simulated crowd of size N. To this end, we utilize the Markov motion model described in [14, 15] to independently evolve N randomly initialized agents. At each iteration, we collect the radial distances of all the agents, test for shadowing based on the criterion of Fig. 5 (a), and set the entries in $\mathcal{H}^{\text{syn}}_{\text{b,vis}} = 1$ if the corresponding agent is visible, to obtain a synthetic Binary Trace Map for each random initialization.

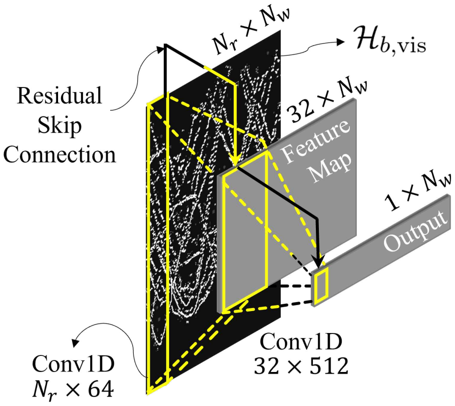

Fig. 6: Schematic of our proposed neural network.

To better model real $\mathcal{H}_{\text{b,vis}}$ and prevent overfitting, we add several non-idealities to $\mathcal{H}^{\text{syn}}_{\text{b,vis}}$. For instance, shadowing depends upon the extent of a person's body encroaching within $\Delta(r)$ of Fig. 5 (a). We first make shadowing probabilistic by sampling a Bernoulli random variable $\mathcal{B}(p_s)$ to decide if an agent is shadowed. We further add Gaussian depth uncertainty of $\mathcal{N}(0, \Delta r)$ to the radial distance of each simulated agent, and add depth-localized white noise to account for environmental artifacts such as foliage. Lastly, we allow the agents to pause intermittently, with their waiting time sampled from $Exp(\lambda)$. We set $p_s = 0.8, \Delta r = 3\text{cm}, \lambda = 1\text{s}$, and generate 500 such synthetic Binary Trace Maps for each $N \in \{1, 2, .., 30\}$, each with an evolution horizon of $N_w = 3000$ windows. We label each $\mathcal{H}^{\text{syn}}_{\text{b,vis}}$ in the dataset with a vector of residual errors denoted by $e[w] = N - n[w], w \leq N_w$, where $n[w] = \sum_{r=0}^{r=N_r} \mathcal{H}^{\text{syn}}_{\text{b,vis}}[w, r]$ represents the baseline estimate at w. Thus, the main goal of our proposed neural network is to predict $e[w]$ for unseen $\mathcal{H}_{\text{b,vis}}$ generated from real data.

Occupancy Estimation using the TCN: We utilize the labeled pairs of $(\mathcal{H}^{\text{syn}}_{\text{b,vis}}[w, r], e[w])$, as described above, to train a TCN [5] for the task of occupancy estimation on unseen $\mathcal{H}_{\text{b,vis}}$. TCNs have been well-studied in several multivariate forecasting contexts such as generative audio [36] and phoneme recognition [52], among others. Furthermore, they are better suited for processing $\mathcal{H}_{\text{b,vis}}$ than 2D-CNNs/RNNs, as they exploit spatiotemporal correlations and simplify setup and training. The block diagram of our neural network is shown in Fig. 6, and is based on the architecture of [1, 5]. More specifically, our network sequentially utilizes two `TemporalBlock` class instances (Fig. 6), each of which

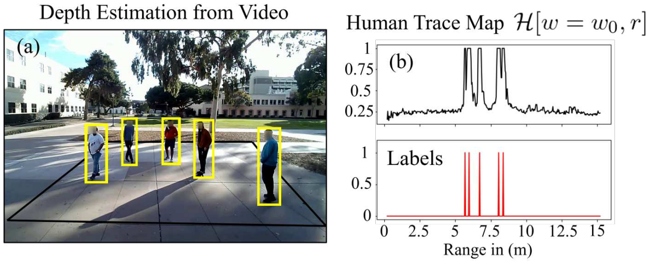

Fig. 7: Small training for depth-occupancy detection: (a) Depth estimation of humans in each frame of the concurrent video feed (b) Associating each video-based human depth to the nearest prominent peak in $\mathcal{H}$.

further comprises a Conv1D + ReLU layer, Batch Normalization ($\epsilon = 10^{-5}, \alpha = 0.1$) and Dropout ($p = 0.5$). We do not enable dilation for either Conv1D layer, leading to a highly-localized temporal neighborhood as the receptive field for the task of occupancy estimation. Due to the sparsity of $\mathcal{H}_{\text{b,vis}}$, we employ residual skip connections [21] to avoid padding-driven feature contamination, and train the network using Stochastic Gradient Descent on an NVIDIA RTX 3060Ti.

Our neural network corrects the residual errors caused by under/overcounting per time step, giving us a more agile estimate of the true crowd size N. More specifically, the TCN can be characterized as a function $\mathcal{G} : \{0,1\}^{N_r \times N_w} \to \mathbb{R}^{N_w}$, which maps a Binary Trace Map $\mathcal{H}_{\text{b,vis}}[w,r]$ to the residual error $e[w]$ across time. We thus estimate the occupancy by adding $e[w]$ to the baseline estimate: $N^*[w] = n[w] + \mathcal{G}(\mathcal{H}_{\text{b,vis}})[w]$. While the occupancy estimate of the network exhibits short-term variations due to the high chirp rate of the radar, a moving average is applied over the predictions to achieve a more robust and representative value. We utilize a sliding frame of columns of $\mathcal{H}_{\text{b,vis}}[w,r]$ over w for spatially-patterned crowds that do not achieve an equilibrium, and for the cases of dynamic and quasi-dynamic crowds of a fixed size, we instead convolve over an expanding frame of analysis. This prevents the network from being biased by outliers in $\mathcal{H}_{\text{b,vis}}$, enabling robust estimation of the equilibrium size. We next detail our adaptive thresholding approach for generating Binary Trace Maps ($\mathcal{H}_{\text{b,vis}}$) from Human Trace Maps ($\mathcal{H}$), by detecting human occupancy in $\mathcal{H}$ across space and time.

7 PRACTICAL DEPTH-OCCUPANCY DETECTION

The Human Trace Map, $\mathcal{H}$, includes a number of undesirable artifacts and requires careful processing before use in our prediction algorithms. Utilizing standard thresholding approaches such as CFAR [17] or DBSCAN [18] may be frustrated by conflicting requirements. On the one hand, Crowd Shadowing Effect and Adjacent Human Merging necessitate a high sensitivity to distinguish returns from nearly-collocated or partially occluded people. On the other hand, a single person spans multiple depth bins (see Fig. 4) and the same sensitivity that distinguishes individuals may also lead to counting the same person multiple times. Finally, detecting agents at farther depths is challenging due to the r^{-4} SNR decay of monostatic sensing, which may require depth-dependent thresholds. Thus, optimizing parameters for robust detection with standard approaches is challenging.

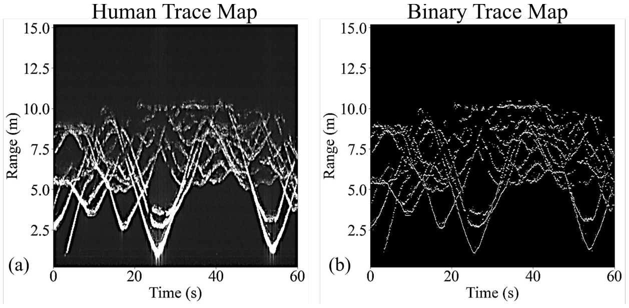

Fig. 8: (a) Sample Human Trace Map of an $N = 11$ crowd. (b) Corresponding Binary Trace Map as the output of the GBM ensemble. See PDF for optimal viewing.

To tackle these challenges, we introduce a machine learning approach that enables human range detection while also denoising $\mathcal{H}$. More specifically, we employ a Gradient Boosting Machine (GBM) [25], which is a lightweight and scalable supervised learning technique that can outperform even Deep Neural Networks in certain tasks [7], and utilizing an ensemble of multi-scale GBMs has been shown to further improve generalization capacity in recent work [28]. We thus propose an ensemble of GBMs trained on a small dataset of concurrently collected video and radar data. Importantly, the Human Trace Maps utilized for generating this dataset are distinct from those on which our approach is validated. To construct our dataset, we collect 15 minutes of concurrent video and mmWave data for $N = 1, 2, 5, 6$ agents in a single area, while the participants naturally traverse the area, in order to capture the impact of Crowd Shadowing and Adjacent Human Merging on the labeled trajectories. The video frame at window w is then used to produce ground-truth binary labels $\mathcal{D}[w,r]$ at each sensing depth r using the YOLO V5 network [50]. To achieve this, we use the predicted bounding boxes to perform a perspective projection [47], and generate a bird's eye view of the scene. As the quality of this step depends on the camera's viewpoint, we match each video depth estimate to the most prominent peak in $\mathcal{H}[w{=}w_0, r]$, within $r_{\text{tol}} = 1.5m$ of the estimated depth, as shown in Fig. 7.

We next detail the training of our GBM ensemble. Each GBM in the ensemble is trained independently using a standard supervised learning approach to learn the mapping from the Human Trace Map, $\mathcal{H}[w,r]$, at a single time step,

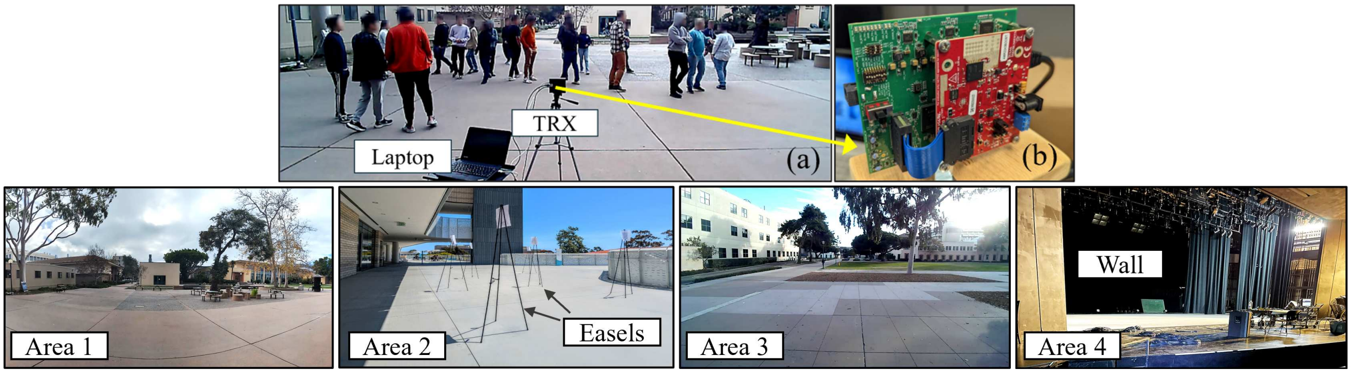

Fig. 9: Sample experimental setup and testing areas: (Top) (a) Monostatic radar faces the crowd during FMCW transmission, connected to a laptop via Ethernet (b) Front view of the radar. (Bottom) Open areas with foliage (Areas 1 and 3), rooftop of a building with multipath scattering (Area 2) and stage of an indoor theater (Area 4).

w, to the binary label vector $\mathcal{D}[w, r]$. To achieve this, the j^{th} GBM in the ensemble is parameterized by a depth-scan window of size R_j, and we slide this window along the r-axis of $\mathcal{H}[w, r]$ to obtain R_j-wide windows denoted by $\Lambda_j[w, r] = \mathcal{H}[w, r - \lfloor R_j/2 \rfloor : r + \lfloor R_j/2 \rfloor]$. For each w, the aim of the j^{th} GBM is to classify a window $\Lambda_j[w, r]$ as 1 if there exists a peak at its center. As motivated earlier, due to high levels of signal attenuation at the mmWave band, windows with peaks at far sensing depths can appear similar to noisy windows at nearer depths, which may lead to higher false negatives. Thus, it is important to incorporate the depth information when inferring the presence or absence of a peak at a given depth. To this end, we include the depth, r, along with the R_j-wide windows (of the Human Trace Map) in our training data. That is, we train the j^{th} GBM on

$$\textbf{Input: } (r, \Lambda_j[w, r]), \quad \textbf{Label: } \mathcal{D}[w, r]$$

for $0 \leq w \leq N_w$, $\lfloor R_j/2 \rfloor \leq r \leq N_r - \lfloor R_j/2 \rfloor$, where N_r is the total number of sensing depths (see Sec. 4). Assembling our training data this way yields a dataset of 5 million radar windows. Due to the low training time of 3 minutes, we tune the GBM hyperparameters [3] using 5-fold CV across 100 iterations to obtain a final validation accuracy of 98.7%.

In our implementation, we use an ensemble of 4 GBMs for multi-scale trajectory detection, with scan window sizes of $R_j = 4\rho j$ meters, $j = 1, 2, 3, 4$, in order to generate four predicted occupancy labels $\mathcal{D}_j$ from $\mathcal{H}$. The choice of the smallest window size is set to approximately twice the average diameter of a human torso, to avoid the detection of extremely narrow yet prominent peaks. The four predictions from the ensemble are then combined to create a quantized Human Trace Map, $\hat{\mathcal{H}} = (1/4) \sum_{j=1}^{4} \mathcal{D}_j$ whose entries belong to the set $\{0, 0.25, 0.5, 0.75, 1\}$. We select a detection threshold that is tied to an initial baseline estimate of the crowd size, to strike a balance between sensitivity and robustness. We observe from simulations that for $r_{\max} = 15m$, the baseline solution is less likely to undercount for $N < 10$, requiring lesser sensitivity. More specifically, we generate the Binary Trace Map, $\mathcal{H}_{\mathrm{b,vis}}$, from $\hat{\mathcal{H}}$, by first selecting the initial threshold as 0 and defining $\mathcal{H}_{\mathrm{b0}}$ as the output after thresholding. We evaluate $n_0[w] = \sum_{r=1}^{r=N_r} \mathcal{H}_{\mathrm{b0}}[w, r]$ and raise the threshold to 0.5 if $\mathbb{E}_w\{n_0[w]\} < 10$, i.e., when the crowd size is sufficiently low. We then re-threshold $\hat{\mathcal{H}}$ based on this heuristic to finally generate $\mathcal{H}_{\mathrm{b,vis}}$. Fig. 8 shows the result of our strategy when applied to an unseen $\mathcal{H}$.

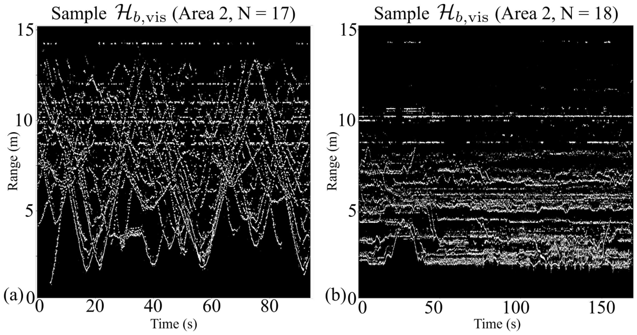

Fig. 10: Binary Trace Maps for (a) dynamic crowd of Fig. 11 (b) and (b) quasi-dynamic crowd of Fig. 11 (e). See PDF for optimal viewing.

In conclusion, we have designed a robust peak detection pipeline that efficiently detects trajectories in $\mathcal{H}$ and obtains consistent $\mathcal{H}_{\mathrm{b,vis}}$ for further analyses.

8 EXPERIMENTAL VALIDATION

We next extensively validate our proposed system for crowd occupancy estimation with several real-world experiments, wherein we generate $\mathcal{H}_{\mathrm{b,vis}}$ from real data using the adaptive depth-occupancy detection strategy of Sec. 7. We first

Experiment #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Area	1	1	3	2	2	4	1	1	1	3	1	1	1	1	3	1	3
Occupancy N	21	21	18	18	17	14	14	14	12	11	10	8	6	4	3	2	1
Analytical (Eq. 9)	18	18	17	16	17	14	16	17	15	11	11	9	9	5	4	3	2
Baseline $\bar{K}$	12	13	12	11	12	10	11	11	10	7	7	6	6	4	3	2	1
Type	Q	D	Q	Q	D	D	Q	D	D	Q	Q	D	Q	D	D	Q	D

Fig. 11: (Top) Three examples of dynamic crowds, (Middle) Three examples of quasi-dynamic crowds, (Table) True and estimated occupancy across several experiments. "D" indicates *Dynamic*, while "Q" indicates *Quasi-dynamic*.

introduce our experimental setup and testing areas, followed by several empirical results (*i.e.,* 22 experiments) to demonstrate occupancy estimation and crowd analytics for crowds of up to (and including) 21 people.[3] More specifically, we achieve a Mean Absolute Error (MAE) of 1.44 and 1.62 for dynamic and quasi-dynamic crowds, respectively, using only a single radar, and in environments that exhibit multipath and foliage, using the proposed mathematical solution of Sec. 5. Furthermore, we show the performance of our proposed neural network on several spatially-patterned crowds. Finally, we demonstrate how to infer features such as crowd engagement level, crowd anomalies, and bottlenecks.

8.1 Experimental Setup

We validate our proposed system with a TI AWR2243BOOST off-the-shelf mmWave radar board [48], as shown in Fig. 9 (b). We set the base frequency, f_0, as $76GHz$ and transmit FMCW pulses with a bandwidth $B = 5GHz$. We configure the chirp rate, f_c, at $400Hz$ and restrict our attention to estimating the occupancy within a maximum sensing depth of $r_{\max} = 15m$, by discarding any detection found beyond that range. While the radar possesses multiple antennas, we exploit the antenna diversity for denoising, thereby emulating a single TRX configuration. The Channel State Information (CSI) off of the radar is captured by a DCA1000EVM FPGA [49], which is connected to an external laptop via Ethernet. We collect radar data for a duration of up to 300s, but as we shall see in Sec. 8.2, we converge much faster, e.g., within 90s (average of 75.4s) when observing a dynamic crowd.

Experimental Areas: We conduct our experiments in three outdoor environments and one indoor environment, as shown in Fig. 9 (Bottom). Area 1 is a 12.8 m × 15 m open area with seating spaces to one side and a trailer on the other. Area 2 is a 15.2 m × 10.2 m open rooftop of a building that exhibits multipath activity, Area 3 is a 15.1 m × 12 m open area with considerable foliage in the vicinity, and Area 4 is a 10.8 m × 15 m stage of an indoor theater. We do not explicitly define the trajectories of our participants during data collection, ensuring natural crowd activity.

8.2 Experimental Results

We next discuss several empirical results that validate the performance of our system across four experimental areas, with crowds of up to (and including) 21 people. To generate $\mathcal{H}$, we use the 90^{th} percentile bandwidth for $\Gamma_q(\cdot)$ in Eq. 4. In the analytical approach of Eq. 9, the body radius, ρ, is conservatively set to $0.25m$, based on values of the biacromial width reported in [34, 51]. Furthermore, to avoid numerical issues, if $\mathbb{P}_e$ or $\mathbb{P}_a$ in Eq. 9 are 0 for any k, we set the probability to $\epsilon = 1\text{e}{-8}$ and then renormalize the distribution.

Occupancy Estimation of Dynamic Crowds Using Proposed Mathematical Model: We conduct a total of 9 dynamic crowd experiments across all four areas, with crowd sizes of up to (and including) 21 people. In these experiments, participants behave as if they were at the concourse of a train station or a busy city square, leading to a less cohesive crowd pattern. As can be seen in Fig. 10 (a), the sample Binary Trace Map, $\mathcal{H}_{\text{b,vis}}$, shows participant trajectories that span a wide range of sensing depths. Fig. 11 (Table) demonstrates the efficacy of our proposed analytical optimization of Eq. 9, which achieves an MAE of 1.44 across all crowd sizes, while the baseline approach (see Section 4.2) does not scale beyond $N = 10$, leading to an MAE of 4.4 at larger sizes due to increased Crowd Shadowing and Adjacent Human Merging.

[3] Our Institutional Review Board (IRB) committee has reviewed and approved this research.

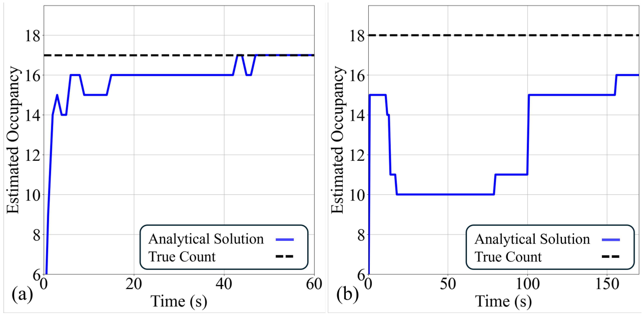

Fig. 12: Convergence of estimated occupancy given by Eq. 9 for (a) the dynamic crowd of Fig. 11 (b) and (b) the quasi-dynamic crowd of Fig. 11 (e).

In indoor areas, secondary bounce off of static objects makes counting large crowds difficult. However, simple calibration of $r_{\max}$ may be sufficient to overcome this challenge. In particular, for the indoor experiment in Area 4, we set $r_{\max} = 10.8m$ to account for the distance between the radar and the wall, and were able to accurately count 14 people.

Occupancy Estimation of Quasi-dynamic Crowds Using Proposed Mathematical Model: We conduct a total of 8 quasi-dynamic crowd experiments across Areas 1, 2, and 3, with crowd sizes of up to (and including) 21 people. In these experiments, the participants either visit a poster session or mingle at a party, leading to crowds with lower mobility. As can be seen in Fig. 10 (b), while $\mathcal{H}_{\text{b,vis}}$ shows localization of trajectories at distinct sensing depths, we do observe transitions between these levels as the participants move from one stable depth to another. Fig. 11 (Table) once again shows that our analytical optimization of Eq. 9 achieves an MAE of 1.62 across various crowd sizes, supporting the probabilistic model of Sec. 5. Notably, the baseline approach achieves an MAE of 5.3 for crowds of more than 10 people.

Convergence Analysis: Fig. 12 (a) demonstrates that our analytical approach converges to the final estimate within 80s of data collection, and we obtain a mean convergence time of 75.4s over all the dynamic crowd experiments. In stark contrast, Fig. 12 (b) shows a relatively longer convergence time of 160s for a quasi-dynamic crowd, due to frequent stopping of the participants requiring more temporal observations. We obtain a mean convergence time of 112.7s over all the quasi-dynamic crowd experiments, thus demonstrating the impact of lower crowd mobility.

Inferring Crowd Engagement Level: Dynamic and quasi-dynamic crowds differ in the extent of a participant's level of engagement with the occupied space. For instance, people exploring a landmark, or at a poster session of a conference, remain stationary for long durations, leading to low crowd mobility. Given that $n[w]$ represents the number of visible people at time w, we propose to exploit the first zero-crossing point τ_Z of its ACF, given by $K_{nn}[\tau] = \mathbb{E}_w\{n[w]n[w+\tau]\}$, to capture the level of crowd engagement.

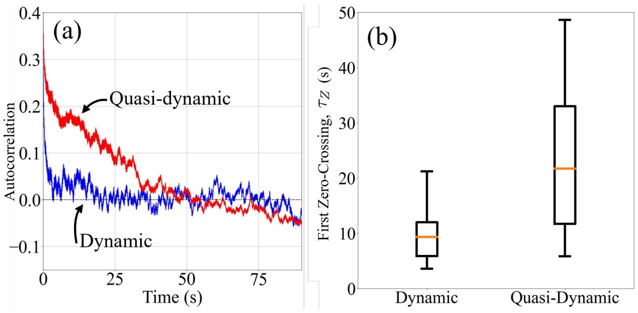

Fig. 13: Autocorrelation of $n[w]$ can reveal level of engagement of crowd with the space. (a) Quasi-dynamic crowds take longer to decorrelate compared to dynamic crowds (b) Variability of τ_z across all experiments.

We obtain an average τ_Z of 9.8s and 24.4s for all our dynamic and quasi-dynamic crowd experiments, respectively. In other words, quasi-dynamic crowds take longer to decorrelate on average, when compared to dynamic crowds, and we motivate τ_Z as a measure of "engagement" in a crowd. Fig. 13 (a) demonstrates a comparison of $K_{nn}[\tau]$ evaluated for a sample pair of dynamic and quasi-dynamic crowds. We observe a clear separability in the crowd engagement level, with $\tau_Z = 4.08s$ for the dynamic crowd and $\tau_Z = 46.08s$ for the quasi-dynamic crowd. Fig. 13 (b) further demonstrates the higher variability of τ_Z for quasi-dynamic crowds. Thus, we see that quasi-dynamic crowds show a higher engagement level, requiring longer observation times.

Occupancy Estimation in Spatially-patterned Crowds Using Temporal Convolutional Network: We now demonstrate the performance of our proposed Temporal Neural Network in providing on-line occupancy estimates of spatially-patterned crowds. Fig. 14 (a) shows participants following a source-absorption-sink configuration enter one-by-one from a narrow opening, sit at random locations, wait for 15s, and exit through the same entry point. Fig. 14 (b) shows a scenario where participants exit an auditorium after a lecture and leave the radar's field-of-view. We process $\mathcal{H}_{\text{b,vis}}$ for both cases by applying our network over sliding frames of 30s to estimate the occupancy. Fig. 14 (right) shows that our network provides estimates that are comparable to vision-based ground truth, thus enabling occupancy estimation of complex crowd topologies. Notably, our network also accurately estimates the sizes of dynamic and quasi-dynamic crowds, with an MAE of 0.88 and 1.5, respectively.

Inferring Crowd Anomaly: While the examples of Fig. 14 involve a gradual change in the crowd size, the identification of an abrupt shift in occupancy holds practical significance in the context of panic event response. To this

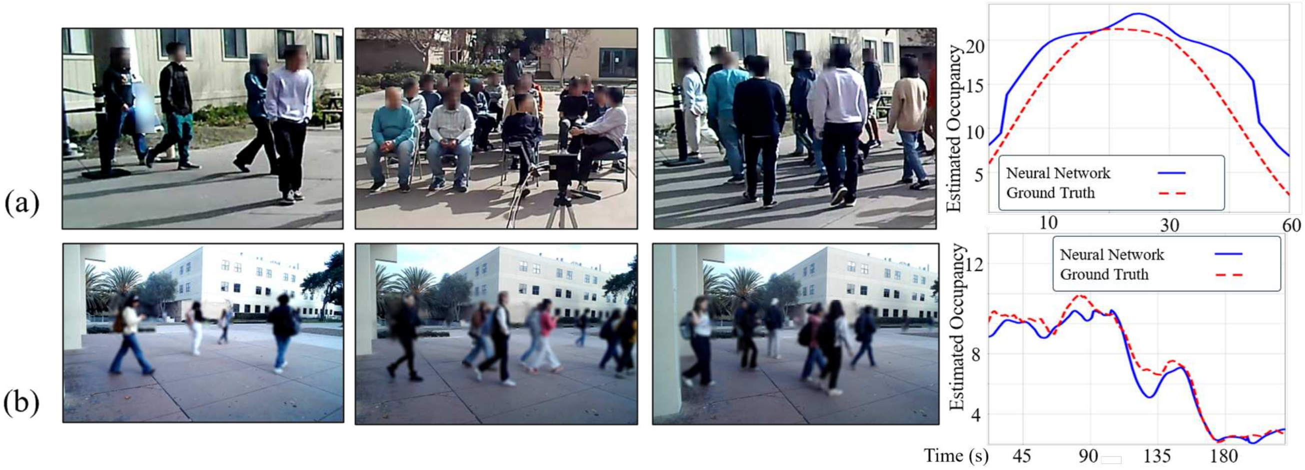

Fig. 14: Occupancy estimation of spatially-patterned crowds: (a) Participants enter one-by-one to randomly sit in chairs placed in an open area, followed by an orderly exit (b) Participants exiting an auditorium after a lecture.

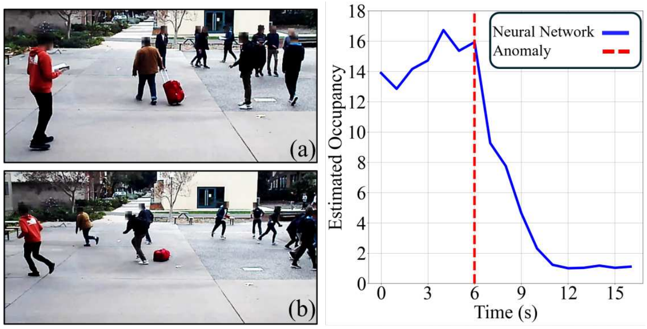

Fig. 15: Crowd anomaly detection: A dynamic crowd rapidly evacuates upon receiving a signal. Neural network is able to precisely detect this event.

end, we conduct an experiment where the participants of a dynamic crowd receive an external signal to rapidly evacuate the experimental area. We exploit the causal nature of our network to scan over 5 second wide frames (without having to re-train the network) as the event itself occurs within a span of < 10 seconds. As can be seen from Fig. 15, the slope of the occupancy curve is ∼ 3 people/s, whereas it is < 1 person/s for the case of Fig. 14 (a). Thus, our neural network is able to precisely detect a sudden shift in the occupancy, identifying such crowd panic events in real-time.

Inferring Crowd Bottleneck: We now demonstrate our neural network's ability to infer the size of a bottleneck. Figs. 16 (a) and (b) depict scenarios in which participants enter an area through a variable-size opening and emulate a dynamic crowd. We employ our neural network to process $\mathcal{H}_{\text{b,vis}}$ by averaging over an expanding time frame. As seen in Fig. 16, the occupancy estimates show distinct profiles while converging to the true size of the subsequent dynamic crowd. Specifically, the faster crowd flow due to the wider bottleneck is clearly demonstrated by the larger slope of the estimated occupancy, compared to the slope for the narrow bottleneck. Thus, we indirectly establish the size of the bottleneck by observing the slope of the occupancy estimates over time.

9 DISCUSSION & FUTURE WORK

Execution Time: Our proposed solution of Sec. 5 takes $1ms$, while our neural network takes $0.5s$, to estimate occupancy from $180s$ of FMCW data on a 13th Gen Intel Core i7 CPU.

Impact of crowd density/experimental area: As N increases, the difference between the distributions $\mathbb{P}_a(K|N)$ and $\mathbb{P}_a(K|N+1)$ shrinks, making it harder to correctly estimate N. This saturation occurs more rapidly in smaller areas. Further characterization is a direction for future work.

Evaluation in More Complex Environments: The quasi-optical nature of mmWave [32] leads to pronounced Ghost Multipath Reflections (GMRs) [20], i.e., secondary bounces off of static objects in the vicinity of human motion [8, 9]. Some of our presented results already have nearby objects that can cause such multipath (e.g., Areas 2 and 4 of Fig. 9). If the secondary reflector is not a very strong reflector, then its impact may not affect the performance. However, for stronger secondary reflectors, the performance can degrade if GMRs are not addressed. One way to address GMRs is based on the geometry of the area. For instance, the GMRs due to the wall in Area 4 of Fig. 9 lie beyond the sensing range of the radar and were naturally filtered, thus not affecting our performance. However, if a strong reflector is not beyond the sensing range of the board, these GMRs can appear as highly-correlated trajectories with human motion. Fig. 17 shows such a case where a person is walking near a large lateral wall. Addressing the impact of GMRs, using recent work such as [33], is thus an interesting future work direction.

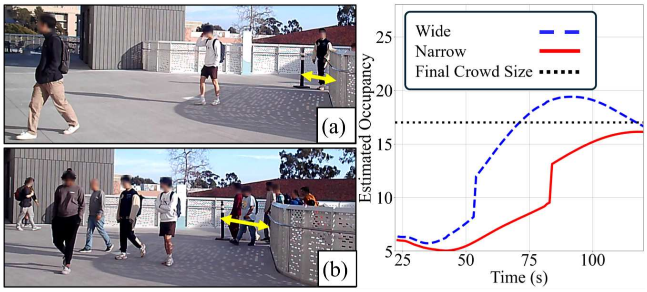

Fig. 16: Crowd bottleneck inference: A dynamic crowd enters through a narrow (a) and wide (b) opening. Our approach detects bottlenecks based on flow rate.

Comparison with Existing Techniques: Our framework enables counting large crowd sizes. As extensively discussed in Sec. 2, most existing work counts based on tracking, which limits their capability to small crowds. Further, they typically require more processing/information (*e.g.,* AoA). Since tracking individuals in a large crowd is an open problem, a straightforward baseline is to count based on the number of visible participants, as done in our paper (i.e., time-average of number of visible participants in Table of Fig. 11), which did much worse than the proposed method. Notably, counting based on the maximum number of visible individuals does even worse than the average. In our pre-processing module, which takes raw radar returns and produces the Binary Trace Maps, the use of Wrapped Phase Spectrum instead of traditional range-Doppler analysis (Fig. 3) is also novel for producing high-quality $\mathcal{H}_{\text{b,vis}}$, especially for denser crowds.

Challenges and Extensions: We identify three key directions for future work. First, NLOS scenarios add complexity not considered in our framework, which can be addressed by increasing TX power to overcome high penetration loss or using environmental knowledge to geometrically resolve multipath effects. Second, challenges in counting very large crowds (e.g., over 20) could be mitigated by incorporating reliable angle of arrival (AoA) information, enabling counting in specific angular sectors. Finally, our current approach does not account for all crowd dynamics, such as transient scenarios or spatial bias in crowd density, which could be addressed by tailoring the analytical model to the scenario. This opens an intriguing avenue for future research, potentially using vision-based methods to automatically learn these priors and improve occupancy estimation accuracy.

10 CONCLUSION

In this paper, we perform occupancy estimation over a diverse set of crowd behaviors using only a single mmWave radar. We introduce a novel mathematical model for crowd size inference using mmWave signals, addressing the problem of crowd shadowing by statistically quantifying its impact. This forms the basis for a fast, simple predictor which achieves a Mean Absolute Error of 1.44 and 1.62 for dynamic and quasi-dynamic crowds, respectively, across 17 experiments in 4 distinct locations. While excelling in dynamic and quasi-dynamic scenarios, the applicability of our model is further extended to spatially-patterned crowds through a temporal convolutional neural network. Extensive real-world validations show robust and accurate crowd counting, and its application in crowd dynamics inference.

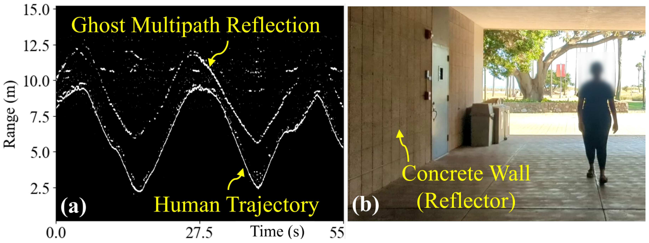

Fig. 17: GMR's impact: Large wall causes Ghost Multipath Reflections that mirrors human trajectory.

11 ACKNOWLEDGEMENTS

We thank the anonymous reviewers and Shepherd for valuable comments and suggestions. This work is funded in part by ONR award N00014-23-1-2715, in part by NSF CNS award 2215646, and in part by NSF CNS award 2226255.

REFERENCES

[1] CMU LocusLab TCN Implementation. https://github.com/locuslab/TCN. Accessed: 2024-06-24.

[2] Cellular IoT connections expected to reach 3 billion in 2023. In *Ericsson Mobility Report November 2023*. Ericsson, 2023.

[3] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

[4] H. J. Alshalani, N. I. Alnaghaimshi, and S. M. Eljack. ICT System for Crowd Management: Hajj as a Case Study. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)*, pages 1–5, 2020.

[5] S. Bai, J. Z. Kolter, and V. Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[6] V. Barral, T. Domínguez-Bolaño, C. J. Escudero, and J. A. García-Naya. An IoT system for smart building combining multiple mmWave FMCW radars applied to people counting. *arXiv preprint arXiv:2401.17949*, 2024.

[7] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci. Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[8] M. Canil, J. Pegoraro, A. Shastri, P. Casari, and M. Rossi. ORACLE: Occlusion-Resilient and Self-Calibrating mmWave Radar Network for People Tracking. *IEEE Sensors Journal*, 2023.

[9] W. Chen, H. Yang, X. Bi, R. Zheng, F. Zhang, P. Bao, Z. Chang, X. Ma, and D. Zhang. Environment-aware Multi-person Tracking in Indoor

Environments with mmWave Radars. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(3):1–29, 2023.

[10] Z.-Q. Cheng, Q. Dai, H. Li, J. Song, X. Wu, and A. G. Hauptmann. Rethinking Spatial Invariance of Convolutional Networks for Object Counting. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19606–19616, 2022.

[11] D. Commisso. Concerns Grow Over Consumer Privacy and Facial Recognition Tech. Technical report, civicscience.com, 2021.

[12] M. Cruz, J. J. Keh, R. Deticio, C. V. Tan, J. A. Jose, and E. Dadios. A People Counting System for Use in CCTV Cameras in Retail. In *2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pages 1–6, 2020.

[13] F. Demrozi, C. Turetta, F. Chiarani, P. H. Kindt, and G. Pravadelli. Estimating indoor occupancy through low-cost BLE devices. *IEEE Sensors Journal*, 21(15):17053–17063, 2021.

[14] S. Depatla and Y. Mostofi. Crowd counting through walls using WiFi. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE, 2018.

[15] S. Depatla, A. Muralidharan, and Y. Mostofi. Occupancy estimation using only WiFi power measurements. *IEEE Journal on Selected Areas in Communications*, 33(7):1381–1393, 2015.

[16] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi. A trained-once crowd counting method using differential wifi channel state information. In *Proceedings of the 3rd International on Workshop on Physical Analytics*, pages 37–42, 2016.

[17] J. Eaves and E. Reedy. *Principles of Modern Radar*. Springer Science & Business Media, 2012.

[18] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, number 34, pages 226–231, 1996.

[19] Y. Golovachev, A. Etinger, G. A. Pinhasi, and Y. Pinhasi. Millimeter wave high resolution radar accuracy in fog conditions—theory and experimental verification. *Sensors*, 18(7):2148, 2018.

[20] J. Guo, M. Jin, Y. He, W. Wang, and Y. Liu. Dancing waltz with ghosts: measuring sub-mm-level 2d rotor orbit with a single mmWave radar. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*, pages 77–92, 2021.

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[22] X. Huang, H. Cheena, A. Thomas, and J. K. Tsoi. Indoor detection and tracking of people using mmwave sensor. *Journal of Sensors*, 2021:1–14, 2021.

[23] T. Instruments. People tracking and counting reference design using mmwave radar sensor. *Design Guide: TIDEP-01000*, 2020.

[24] M. Jiang, S. Guo, H. Luo, Y. Yao, and G. Cui. A robust target tracking method for crowded indoor environments using mmWave radar. *Remote Sensing*, 15(9):2425, 2023.

[25] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[26] B. Korany and Y. Mostofi. Counting a stationary crowd using off-the-shelf WiFi. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 202–214, 2021.

[27] S. Li and R. Hishiyama. An indoor people counting and tracking system using mmwave sensor and sub-sensors. *IFAC-PapersOnLine*, 56(2):7096–7101, 2023.

[28] W. Li, J. He, H. Lin, R. Huang, G. He, and Z. Chen. A lightgbm-based multi-scale weighted ensemble model for few-shot fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2023.

[29] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi. Integrated Sensing and Communications: Toward Dual-Functional Wireless Networks for 6G and Beyond. *IEEE Journal on Selected Areas in Communications*, 40(6):1728–1767, 2022.

[30] H. Liu, X. Liu, X. Xie, X. Tong, and K. Li. PmTrack: Enabling Personalized mmWave-based Human Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(4):1–30, 2024.

[31] J. Liu, C. Gao, D. Meng, and A. G. Hauptmann. DecideNet: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2018.

[32] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Khoryaev, and A. Lomayev. Experimental investigations of 60 GHz WLAN systems in office environment. *IEEE Journal on Selected Areas in Communications*, 27(8):1488–1499, 2009.

[33] N. Mehrotra, D. Pandey, U. Madhow, Y. Mostofi, and A. Sabharwal. Instantaneous Velocity Vector Estimation Using a Single MIMO Radar Via Multi-Bounce Scattering. In *2024 IEEE Conference on Computational Imaging Using Synthetic Apertures (CISA)*, pages 1–5. IEEE, 2024.

[34] M. S. Mesa, V. Fuster, A. Sanchez-Adnres, and D. Marrodan. Secular changes in stature and biacromial and bicristal diameters of young adult Spanish males. *American Journal of Human Biology*, 1993.

[35] V. Nogueira, H. Oliveira, J. Augusto Silva, T. Vieira, and K. Oliveira. RetailNet: A Deep Learning Approach for People Counting and Hot Spots Detection in Retail Stores. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 155–162, 2019.

[36] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[37] E. Pantano, D. Marikyan, and V. Vannucci. Generational attitudes to facial recognition: How retailers and policymakers should approach the use of developing facial recognition technology. Technical report, University of Bristol, 2023.

[38] J. Pegoraro and M. Rossi. Real-time people tracking and identification from sparse mm-wave radar point-clouds. *IEEE Access*, 9:78504–78520, 2021.

[39] S. Rao. Introduction to mmWave sensing: FMCW radars. *Texas Instruments (TI) mmWave Training Series*, pages 1–11, 2017.

[40] T. S. Rappaport and S. Deng. 73 GHz wideband millimeter-wave foliage and ground reflection measurements and models. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1238–1243. IEEE, 2015.

[41] L. Ren, A. Yarovoy, and F. Fioranelli. Grouped People Counting Using mm-wave FMCW MIMO Radar. *IEEE Internet of Things Journal*, 2023.

[42] A. Santra, R. V. Ulaganathan, and T. Finke. Short-range millimetric-wave radar system for occupancy sensing application. *IEEE sensors letters*, 2(3):1–4, 2018.

[43] P. Sivaprakash, M. Sankar, R. Chithambaramani, and D. Marichamy. A Convolutional Neural Network Approach for Crowd Counting. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1515–1520, 2023.

[44] B. Solmaz, B. E. Moore, and M. Shah. Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2064–2070, 2012.

[45] G. Solmaz, F.-J. Wu, F. Cirillo, E. Kovacs, J. R. Santana, L. Sanchez, P. Sotres, and L. Munoz. Toward Understanding Crowd Mobility in Smart Cities through the Internet of Things. *IEEE Communications*

Magazine, 57(4):40–46, 2019.

[46] A. G. Stove. Linear FMCW radar techniques. In *IEE Proceedings F (Radar and Signal Processing)*, number 5, pages 343–350. IET, 1992.

[47] R. Szeliski. *Computer vision: algorithms and applications*. Springer Nature, 2022.

[48] Texas Instruments. *AWR2243 Evaluation Module (AWR2243BOOST) mmWave Sensing Solution*. Available online at https://www.ti.com/lit/ug/spruit8d/spruit8d.pdf.

[49] Texas Instruments. *DCA1000EVM Quick Start Guide*. Available online at https://www.ti.com/lit/ml/spruik7/spruik7.pdf.

[50] Ultralytics. YOLOv5: A state-of-the-art real-time object detection system. https://docs.ultralytics.com, 2021. Accessed: 1st March 2024.

[51] E. US Department of Health and Welfare. Skinfolds, Body Girths, Biacromial Diameter, and SelectedAnthropometric Indices of Adults. Technical report, National Center for Health Statistics, 1970.

[52] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang. Phoneme recognition using time-delay neural networks. In *Backpropagation*, pages 35–61. Psychology Press, 2013.

[53] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu. Respiration tracking for people counting and recognition. *IEEE Internet of Things Journal*, 7(6):5233–5245, 2020.

[54] Z. Wei, F. Zhang, S. Chang, Y. Liu, H. Wu, and Z. Feng. mmWwave radar and vision fusion for object detection in autonomous driving: A review. *Sensors*, 22(7):2542, 2022.

[55] V. Winkler. Range Doppler detection for automotive FMCW radars. In *2007 European Radar Conference*, pages 166–169. IEEE, 2007.

[56] S. Yang, T. Li, X. Gong, B. Peng, and J. Hu. A review on crowd simulation and modeling. *Graphical Models*, 111:101081, 2020.

[57] Y. Yang, J. Cao, X. Liu, and X. Liu. Wi-Count: Passing people counting with COTS WiFi devices. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9. IEEE, 2018.

[58] J. Zhang, R. Xi, Y. He, Y. Sun, X. Guo, W. Wang, X. Na, Y. Liu, Z. Shi, and T. Gu. A survey of mmWave-based human sensing: Technology, platforms and applications. *IEEE Communications Surveys & Tutorials*, 2023.