ICE-MILK: INTELLIGENT CROWD ENGINEERING USING MACHINE-BASED INTERNET OF THINGS LEARNING AND KNOWLEDGE BUILDING

A DISSERTATION
IN
Telecommunications and Computer Networking
and
Computer Science

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
KHALID JABER ALMALKI

M. Sc., University of Missouri-Kansas City, MO, USA, 2018
B. S., Jazan University, Saudi Arabia, 2014

Kansas City, Missouri
2022

© 2022
KHALID JABER ALMALKI
ALL RIGHTS RESERVED

ICE-MILK: INTELLIGENT CROWD ENGINEERING USING MACHINE-BASED INTERNET OF THINGS LEARNING AND KNOWLEDGE BUILDING

Khalid Jaber Almalki, Candidate for the Doctorate of Philosophy Degree

University of Missouri–Kansas City, 2022

ABSTRACT

The lack of proper crowd safety control and management often leads to spreading human casualties and infectious diseases (e.g., COVID-19). Many Machine Learning (ML) technologies inspired by computer vision and video surveillance systems have been developed for crowd counting and density estimation to prevent potential personal injuries and deaths at densely crowded political, entertaining, and religious events. However, existing crowd safety management systems have significant challenges and limitations on their accuracy, scalability, and capacity to identify crowd characterization among people in crowds in real-time, such as a group characterization, impact of occlusions, mobility and contact tracing, and distancing.

In this dissertation, we propose an Intelligent Crowd Engineering platform using Machine-based Internet of Things Learning, and Knowledge Building approaches (ICE-MILK) to enhance the accuracy, scalability, and crowd safety management capacity in

real-time. Specifically, we design an ICE-MILK structure with three critical layers: IoT-based mobility characterization, ML-based video surveillance, and semantic information-based application layers. We built an IoT-based mobility characterization system by predicting and preventing potential disasters through real-time Radio Frequency (RF) data characterization and analytics. We tackle object group identification, speed, direction detection, and density for the mobile group among the many crowd mobility characteristics. Also, we tackled an ML-based video surveillance approach for effective dense crowd counting by characterizing scattered occlusions, named CSONet. CSONet recognizes the implications of event-induced, scene-embedded, and multitudinous obstacles such as umbrellas and picket signs to achieve an accurate crowd analysis result. Finally, we developed a couple of group semantics to track and prevent crowd-caused infectious diseases. We introduce a novel COVID-19 tracing application named Crowd-based Alert and Tracing Services (CATS) and a novel face masking and social distancing monitoring system for Modeling Safety Index in Crowd (MOSAIC). CATS and MOSAIC apply privacy-aware contact tracing, social distancing, and calculate spatiotemporal Safety Index (SI) values for the individual community to provide higher privacy protection, efficient penetration of technology, greater accuracy, and effective practical policy assistance.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "ICE-MILK: Intelligent Crowd Engineering using Machine-based Internet of Things Learning and Knowledge Building," presented by Khalid Jaber Almalki, candidate for the Doctorate of Philosophy degree, and hereby certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Sejun Song, Ph.D., Committee Chair
Department of Computer Science & Electrical Engineering

Praveen Rao, Ph.D.
Department of Computer Science & Electrical Engineering

Baek-Young Choi, Ph.D.
Department of Computer Science & Electrical Engineering

Cory Beard, Ph.D.
Department of Computer Science & Electrical Engineering

Lein Harn, Ph.D.
Department of Computer Science & Electrical Engineering

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to offer my sincerest gratitude and appreciation to my advisor Dr. Sejun Song for his invaluable guidance and support throughout my doctoral research. I am especially indebted to him for the encouragement, valuable feedback, and expertise I needed during my Ph.D. journey. His patience and knowledge allowed me to learn new prospects and professional guidance in scientific research. Also, I am very thankful for my co-adviser and committee member, Beak-Young Choi, for her valuable feedback, support, and encouragement. This accomplishment would not have been possible without their invaluable support.

My sincere thanks are also extended to my committee members, Dr. Praveen Rao, Dr. Cory Beard, and Dr. Lein Harn, for their time, effort, and serving on my Ph.D. committee. I am gratefully indebted to them for their precious advice.

I want to express my profound appreciation and thanks to my parents, brothers, and sisters for their continuous encouragement and support throughout my Ph.D. journey. Also, a special thanks go to my supportive wife Khawlah and my daughter Kayan. I am incredibly thankful to them. They have always been my biggest inspiration for their patience, motivation, and prayers.

I also would like to extend my warmest appreciation to the School of Computing and Engineering and the School of Graduate Study for providing me the opportunity to complete my study at the University of Missouri-Kansas City in the United States.

Lastly, I would like to thank Saudi Electronic University for investing in me by

sponsoring me financially. Without their sponsorship, I would not be at this stage of my Ph.D. I am grateful to them for allowing me to get the opportunity to accomplish my educational journey.

CHAPTER 1

INTRODUCTION

1.1 Background

Due to civilization development and population growth, crowd safety control and management have become an increasingly important issue worldwide. In addition, crowded events are a significant part of modern human society, such as festivals, concerts, sports, political (protests), and religious events (Hajj or Kumbha Mela). Hence, the essential role of crowd management can be seen in ensuring the safety of the attendees at a crowded event. In many cases, the number of people gathering at the event location can create potential risks that lead to dangerous consequences if not controlled. Thus, cities and governments face the daunting task of containing such massive gatherings. These events require extensive attention from the planners to set up an appropriate crowd management system to ensure attendeesâ safety. Nevertheless, the lack of proper crowd safety control and management can often lead to human casualties and the spread of infectious diseases (i.e., COVID-19) at densely crowded events. Also, the crowd's density plays a critical role, and larger crowds at events are more complex, challenging to handle, and riskier, which can lead to dangerous incidents. Table 1 indicates numerous recent cases of crowd disasters that happened during events and caused human losses around the world. For example, in Madagascar in 2019, during a concert on their 59th Independence Day at the Mahamasina Municipal Stadium, at least 16 people were killed, and 101 were injured.

The show was about to start, and people believed they could enter the stadium, so they began to push each other [1]. A similar situation happened in Angola in 2017 where people tried to enter the stadium for a football match before the time. It ended up with a stampede with at least 17 people killed and 61 others injured [2]. Human crowd incidents are continuously occurring. In addition, the Hajj stampede in 2015 is another example of a crowd disaster. This happened during the annual Islamic pilgrimage called Hajj that occurs in Mecca, Saudi Arabia. Groups of pilgrims walked in the opposite direction to get to their destination faster. So, the crowd flow got blocked from the crowd flow in the opposite direction. It resulted in more than 2000 casualties, which is considered one of the most significant crowd disasters that caused human losses in previous years [3].

Table 1: Crowd disasters around the world.

Location (Country/City)	Human Casualties	Year	Event Type
Houston, USA	308	2021	Entertainment
Iran, Kerman	256	2020	Religious
Madagascar, Antananarivo	117	2019	Entertainment
Iraq, Karbala	130	2019	Religious
Bangladesh, Chittagong	60	2018	Religious
Angola, Uige	78	2017	Entertainment
Italy, Turin	1526	2017	Entertainment
Ethiopia, Addis Ababa	300	2016	Protests
Saudi Arabia, Mecca	2000	2015	Religious
Brazil, Santa Maria	242	2013	Entertainment
Egypt, Port Said	74	2012	Sport
Germany, Duisburg	361	2010	Entertainment
India, Jodhpur	224	2008	Religious

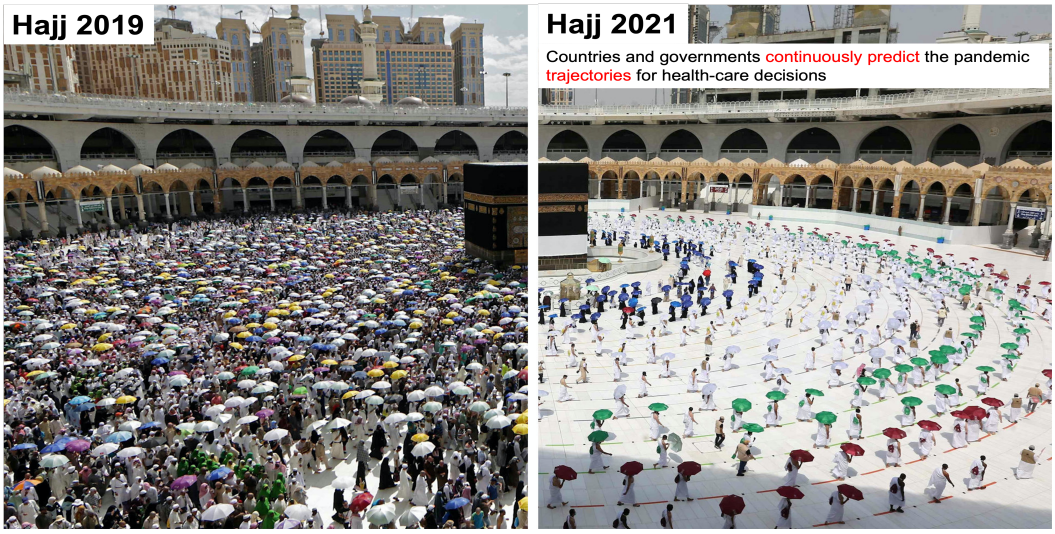

Figure 1: Hajj before Covid-19 vs. Hajj after Covid-19.

Therefore, crowd safety management focuses on potential crowd disaster prevention. Moreover, in early 2020, COVID-19 caused pandemic interventions all over the world. As of March 2022, the virus has infected more than 471 million people and caused more than 6 million deaths worldwide [4]. COVID-19 has changed life completely, including the ways of learning and working. The spread of the pandemic led to a halt to large social events due to the high contagiousness of the virus. Lockdowns, travel restrictions, and self-isolations due to the coronavirus have cleared public squares and travel destinations across the globe and left many public spaces deserted. As shown in Figure 1, governments and health systems around the world have made many policies regarding gathering restrictions such as social distancing (6 feet distancing between people), wearing a mask, and so on to prevent and protect people from this virus.

However, even with COVID-19 restrictions, a crowd disaster can still occur unless there is excellent crowd safety management to control the crowd. Otherwise, it can end

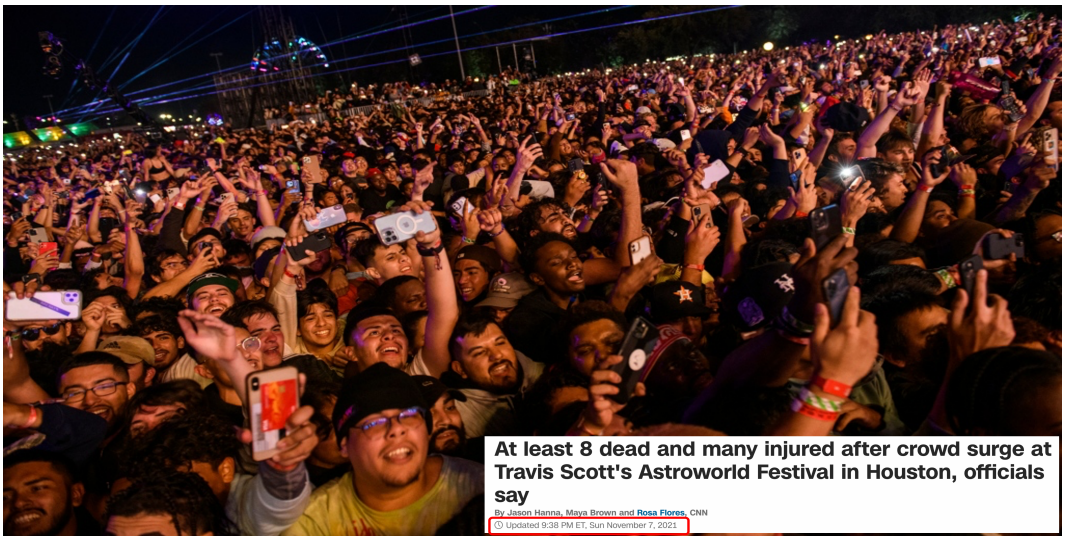

Figure 2: Astroworld Festival 2021 in Texas, USA.

up with human casualties. For example, in November 2021, a densely crowded music festival occurred in Houston, Texas. As shown in Figure 2, more than 50 thousand people attended the event. Due to the dense crowd, people compressed up against each other and pushed forward and back sides. Thus, the event ended with at least eight people killed and more than 300 injured [5].

On the other hand, crowd events have long been known to increase the risks of infectious disease outbreaks, specifically now with COVID-19, and the number of infection cases during densely populated events. For example, during the COVID-19 pandemic in April 2021, a massive gathering happened in India called Kumbh Mela, which is a major pilgrimage in Hinduism. As shown in Figure 3, the festival occurred with a huge number of attendees. Due to the lack of implementing the COVID-19 restrictions, the virus rapidly spread between the attendees, and over 386,000 new cases of COVID-19 and over 3,500 deaths were reportedly caused by the Kumbh Mela event [6].

Figure 3: Covid-19 and Kumbh Mela festival 2021.

Accordingly, crowd safety control and management are becoming increasingly vital elements to event planning. As part of that, crowd behavior analysis, including characterization, counting, and modeling the crowd's safety, are valuable instruments to ensure safety and avoid human disasters. However, we need an integrated and intelligent crowd engineering system to anticipate and control crowd disasters and infectious diseases in different contexts, which includes factors such as characterizing crowds, density, mobility, contact tracing, social distancing, and mask-wearing. This integrated system should enhance real-time accuracy, scalability, and crowd safety management capacity in addition to density and count detection.

1.2 Crowd Engineering

Crowding is an unavoidable occurrence, and it can occur anywhere and at any time. We can define the concept of the crowd as a large number of individuals gathered together in specified areas or for a particular event. Crowding is a potential hazard at entertainment events (sports), political events (protests, rallies), religious events (Hajj), shopping areas (malls), or transportation areas (train stations, airports). Usually, events generate a high crowd density, and there may be more significant risks to the attendees' safety unless there is an excellent crowd control system. To better understand this, many researchers have defined crowd in regards to this purpose. Duives et al. [7] defines the crowd as a large group of heterogeneous individuals who share a common goal in a common physical environment. Crowd members may behave differently when they are alone or in a small group.

In order to discuss the definition of crowd, we need to define crowd engineering and how it is essential to control the crowd. Therefore, crowd engineering is an integrated and intelligent science that aims to analyze and characterize crowds using various elements in different aspects, including density measures, mobility, contact tracing, social distancing, and mask-wearing to prevent human casualties and infectious diseases (i.e., COVID-19). However, crowd engineering is widely essential in smart cities, where it can plan, monitor, and manage crowds to enhance the accuracy, scalability, and crowd safety management capacity in real-time. Notably, artificial intelligence (AI) techniques such as crowd counting and mask-wearing detection in a crowd have been applied to the science of crowd engineering, which plays a critical role in promoting crowd safety and

management.

1.2.1 Types of Crowds

According to crowd safety and risk analysis, understanding the types of crowds possible in a space is paramount to enhancing the efficiency and accuracy of the crowd engineering techniques. Therefore, crowd density is considered a significant characteristic in determining the type of crowd, which can assess the efficiency of crowd movement, the capacity of a place, and it needs to understand the relative risks of each crowd density type. Therefore, we categorized the types of the crowd into three types, including *static*, *mobile*, and *hybrid crowds*.

Figure 4: Static Crowd (standing).

The **static crowd (standing)** usually gathers people in a specific location. It mainly stands during the event, such as musical events. As illustrated in Figure 4, people are standing in the event, and the crowd's density depends on the event's space. However,

these events could be dense or sparse, and the complexity of crowd management depends mainly on crowd density. Hence, the average space per person is 50 cm per square meter in a sparse crowd, and space decreases when the crowd's density increases [8].

However, in **mobile crowds (dynamic)** people move from one place to another, such as during protests. It requires more space and is harder to manage or control compared to the static crowd [2]. In addition, the risk of human casualties and the spread of infectious diseases' during mobile crowd events are more likely to occur because small incidents at the crowd are enough to cause panic throughout the whole crowd of attendees [9].

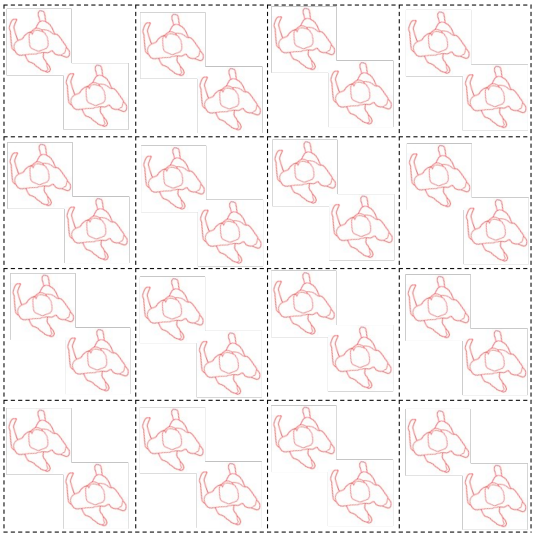

Figure 5: Mobile Crowd (unidirectional).

Mobile crowds can be *unidirectional*, as illustrated in Figure 5, where the crowd is moving in the same direction. As the density is related to the crowd type, when density becomes two people in a square meter, the flow speed is considered normal. However,

when density becomes more than three people per square meter, the flow speed decreases and becomes a high-risk crowd.

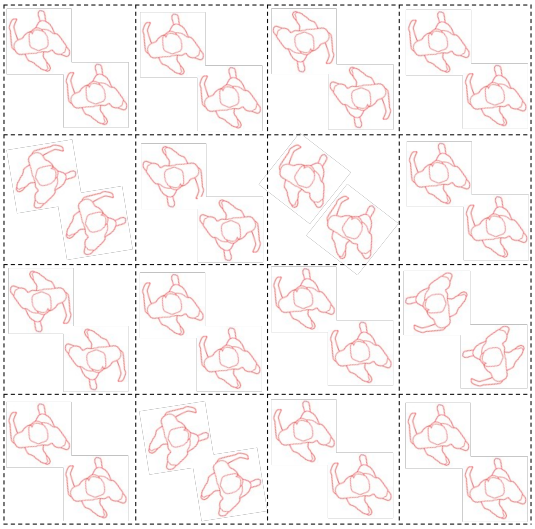

Figure 6: Mobile Crowd (non-unidirectional).

Furthermore, mobile crowds can be *non-unidirectional*. As illustrated in Figure 6, the crowd can move randomly in different directions. The flow speed decreases significantly, even at a lower density because to avoid collision crowd flow must halt. With this type of crowd, a high-density crowd is high-risk crowd. When the crowd force pushes forward, shock waves begin to ripple through the tightly packed mass, and it causes a crush risk that can quickly lead to crowd disasters.

The last type is the **hybrid crowd**, which includes all previous crowds types in the same space. These types of crowds happen where people need to move and stand in the same location, sharing a common goal or common interests, such as in shopping malls, train stations, and airports. As shown in Figure 7, the hybrid crowd is also considered a complex task to manage because it requires considering the density of each type of

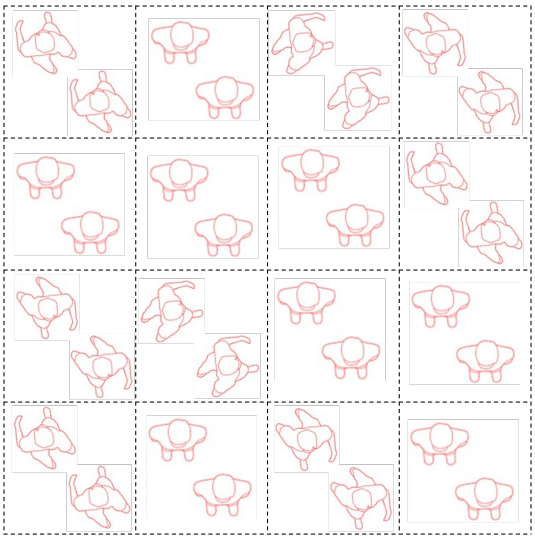

Figure 7: Hybrid Crowd.

crowd in the same place to control the crowd effectively. In addition, when people stand and move randomly in different directions in the crowd, the flow speed decreases significantly at the lower density, and collisions are avoided by stopping the flows momentarily. High-density creates a high-risk crowd, and it can lead to a crush and allow infectious diseases to spread. Therefore, risks of human casualties at crowded events are more likely to initially occur as a small incident which panics the other attendees to start hustling, collapsing, trampling, and stampeding each other. Inappropriate crowd management often results in disastrous repercussions such as injuries and losses of human life.

These types of crowds need to be anticipated by the planner to set up an appropriate crowd management system. According to crowd safety and risk analysis [10], understanding each crowd type's impact and consequence, including crowd density and relative risks, is critical for managing crowd safety.

1.3 IoT-based Crowd Engineering

The term Internet of Things (IoT) technology has been widely employed in smart cities to build innovative crowd management systems and control the crowd in particular contexts. Scientists and researchers have developed many applications and technologies in the field of crowd management. Nevertheless, these technologies enhance the understanding of crowd mobility, including characterization, density, and data analytics, by providing unique perspectives and solutions, which can promote safety management of urban communities. For example, CROMO [11] proposed an intelligent IoT framework that enhances safety management for mobile crowd events. It detects and prevents potential disasters using real-time Bluetooth low energy (BLE) signal characterization and analytics. The framework implementation consists of a BLE transmitter tag, a BLE signal scanner, and an analytic server. CROMO identifies the crowd density, object group location, flow direction, and speed by analyzing BLE beacon counts, radio strength signals received (RSSI) power, and variation patterns. Weppner et al. [12] proposed a procedure for estimating crowd density using a mobile device to scan for Bluetooth devices that appear in the environment. The device is used to scan and analyze social contexts and extend into advanced features, which will allow estimating the absolute number of devices in the environment. Notably, the recent advancement of the science of IoT technology has enhanced crowd surveillance systems' efficiency, scalability, and capability. Therefore, crowd management systems have incorporated IoT devices and sensors to provide valuable and accurate detection and analysis to control the crowd and ensure the safety of attendees. In addition, an intelligent crowd engineering system has employed IoT-based

crowd management strategies and integrated them with computer vision, ML, and deep learning approaches, which has resulted in an intelligent crowd system to control crowd disasters and infectious diseases, including crowd characterizing, density measurement, crowd mobility, contact tracing, social distancing, and mask-wearing.

1.4 AI-based Crowd Engineering

With the development of science and technology, AI plays a significant role in controlling and managing crowds. It is being used to simplify and improve how humans control crowds and enhance public safety and security during crowded events. Therefore, researchers in crowd management have proposed various techniques and strategies to control and manage the crowd. These techniques include crowd mobility, counting, and tracking. They have improved the performance and reduced the complexity of the crowd in practical contexts; however, crowds still have complexities and limited accuracy due to unforeseeable congested areas and perspectives, including image resolutions, occlusions, and non-uniform environments. This part will discuss the use of machine learning (ML) and the tools used to build intelligent crowd management techniques.

Machine Learning (ML) is a subfield of artificial intelligence, which is defined as the capability of a machine to imitate intelligent human behavior to solve complex problems. It provides self-learning and improves from experience without being explicitly programmed. The machine learning model consists of numerous layers on multiple levels to extract features from data. There are two types of learning in machine learning, *supervised* and *unsupervised*. Supervised learning trains the machine using labeled data

such as classification and recognition problems. However, unsupervised learning does not need these labeled inputs. It mainly uses unlabeled data to find patterns in the input data, such as clustering. In addition, *Convolutional Neural Networks (CNNs)* have become popular in crowd management applications in recent years. In practice, a CNNs algorithm can learn the mapping function of the input data and extract the features. It uses mathematical formulation operations to detect features from the input to allow the model to categorize detected objects [13]. Furthermore, machine learning methods are categorized into *regression* and *classification* based on detection purposes. For example, the regression methods used to detect the values and the classification used to detect a category or determine between true and false. These aspects of machine learning allow crowd management systems to be built and enhanced in smart cities to control crowd events in broad aspects.

On the other hand, **crowd counting** is becoming an increasingly important element of crowd management for AI techniques as well. It has many applications in smart cities, especially regarding public safety, such as surveillance, traffic monitoring, and urban planning [14]. Counting the number of people plays an essential role in the security and management of cities and facilities. It aims to develop AI tools that allow for a quick and accurate estimation of the count of people in a crowded environment. These tools can be used to perform real-time tracking of people counts in crowded areas. Hence, automated crowd interpretation using AI techniques [15] is becoming an increasingly critical task for many practical crowd safety applications [16]. Although many CNN-based methods have been proposed to improve the performance on complex crowd images to

deal with variations in scale, perspective, and image resolution [17, 18], they still have significant limitations in the face of occlusions that partially impede sight of individuals in a crowd scene.

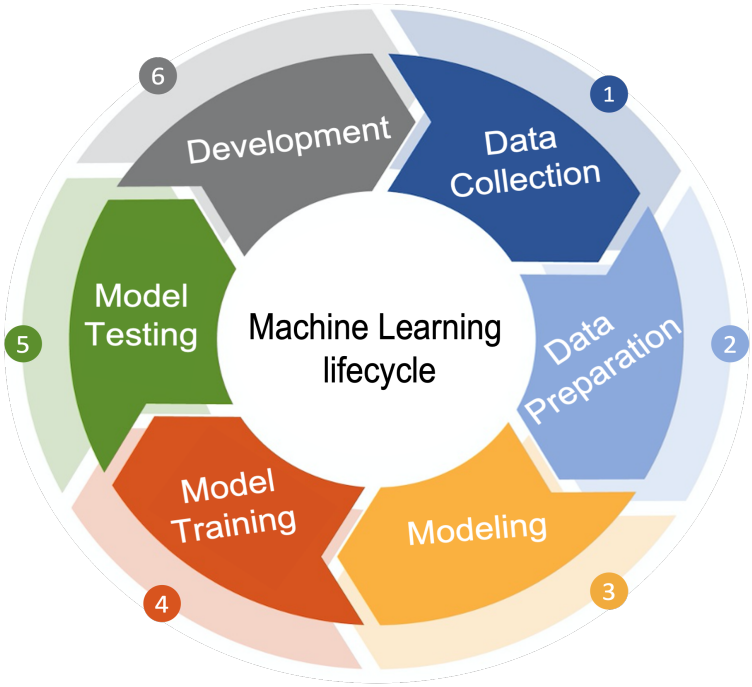

Figure 8: Machine Learning lifecycle.

As shown in Figure 8, building an intelligent crowd management system using machine learning requires several steps (lifecycle), which yeilds a robust system while reducing costs and time.

- **Data Collection:** Identifying relevant data for the implemented machine learning application is a major step required for efficient machine learning solutions [19]. There are two ways to collect this type of data. First is manual data collection, which is needed when applications may not have enough ready data to train the

model fully. Therefore, the developers manually collect data from various sources such as the internet by writing a programming script or generating the data from sensors (Raspberry Pi) or mobile devices. The second way is to find the required data from an existing resource such as Google and GitHub. This way will speed up the performance and save more time and effort.

- **Data Preparation:** This is an iterative process of structuring, cleaning, and transforming raw data into a desired format that can be used in modeling. It is considered the most critical step due to it being the most time-consuming effort to provide valuable data for further analysis and better decision-making in less time [20]. Nevertheless, the preparation process consists of significant activities, including structuring, cleaning, and labeling. For this task, the VGG Image Annotator (VIA) software [21] can be used to manually draw the annotations like bounding boxes, rectangles, circles, polygons, and points on the image and assign each labeled object to a class. This information generates a ground truth for each image, which is used to train the model.

- **Modeling:** Building the model is a core and the fundamental step of any ML project. The model can be created from scratch, which requires an understanding of the targeted problem's requirements, including data, parameters, configurations, and the model's features. Nevertheless, it is also possible to use a pre-trained CNN model as a backbone method to integrate into your model. The examples of the pre-trained architectures include VGG-16 [22], ResNet-152 [23], MobileNet-V2 [24], EfficientNet-B0 [25], DenseNet-121 [26], YOLO V3 [27], and YOLO V5

[28]. In addition, model selection is the process of writing and running the programs of a targeted problem with solid results using programming languages like Python, Matlab, C++, and R. Python programming language is considered the most popular language with tons of libraries and software such as TensorFlow [29], PyTorch [30], and Keras [31]. Specifically, TensorFlow is an open-source platform for ML released by Google. It uses efficient numerical computing tools to create deep learning architectures to assist developers in deploying their models across various platforms. PyTorch is an open-source framework (Torch library) released and developed by Facebook. It is mainly used in machine learning and deep learning applications. Furthermore, Keras is an open-source library written in Python. It provides a high-level neural network implementation. Keras is tightly integrated with TensorFlow to create ML models.

- **Model Training:** We train the model using the training set to check the model's performance and learning patterns for the best outcome of the problem. Usually, any ML problem dataset is characterized into three sets: training, testing, and validation. Furthermore, several parameters are used to evaluate the model performance, such as Adam optimizer [32], Stochastic Gradient Descent [33], learning rate, momentum, batch size, and epoch number.

- **Model Testing:** In this step, we test the model to check the model's performance and outcome accuracy. However, suppose we are not satisfied with the final models' accuracy and outcome. In that case, we can configure and train the model again,

which may improve the performance for a better outcome of the problem. The testing steps determine the model's percentage accuracy as per the targeted problem's requirement.

- **Development:** This is the final step of building an ML project. This step is considered the deployment step. It is required to deploy and integrate the model as a prototype or platform in real-world application in order to solve real-world problems.

To summarize, ML techniques, including deep learning, are widely used for crowd monitoring and associated tasks such as counting, density estimation, tracking, scene understanding, localization, and behavior detection to build an intelligent crowd management system. Furthermore, it provides various benefits such as efficiency, flexibility, and high-performance accuracy.

1.5 Research Domains

This dissertation aims to build an intelligent crowd engineering platform by enhancing accuracy, scalability, and capacity to support public safety and management in real-time. Specifically, three essential components will be addressed: IoT-based mobility characterization, ML-based video/image surveillance, and semantic knowledge as shown in Figure 9.

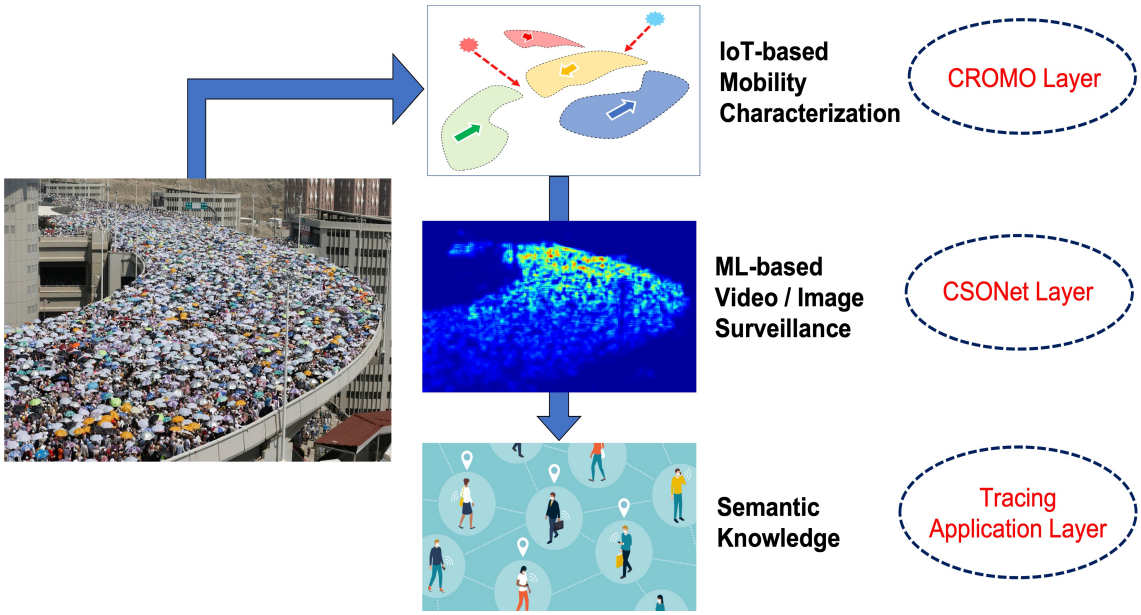

Figure 9: Intelligent Crowd Engineering (ICE) Platform

1.5.1 IoT-Based Mobility Characterization

Due to the unprecedented scale and speed of urbanization, cities are facing the daunting task of accommodating for urban dynamics. The concept of smart cities attract city planners and researchers as it facilitates many smart community services by combining cyber-physical systems and social entities through wireless, mobile, and intelligent information and communication technologies (ICT). One of the critical service requirements of future cities is safety management for citizens and communities [34]. Specifically, safety management during densely populated events such as religious, entertainment, and political gatherings becomes more significant as it happens more frequently and at large scale in modern cities. Unlike static crowd events where a crowd is formed

in a specific location, when a crowd is moving from a location to another (i.e., unidirectional), it requires more space (i.e., less density). If crowd mobility exhibits multiple non-unidirectional patterns, it would require even more space to be safe and is much harder to manage and control [35, 2]. Risk of human casualties with mobile crowd events are more likely to occur because small incidents at the crowd are enough to cause panic among the attendees to start hustling, collapsing, trampling, and stampeding each other. Any inappropriate crowd management in these situations often result in disastrous repercussions such as injuries and casualties [9].

The recent smart video surveillance inspired by advanced AI technologies and ML algorithms enables a broad spectrum of promising safety applications, including object detection and identification, behavior recognition and tracking, and anomalous event detection [36, 37]. However, video surveillance alone cannot identify and predict a particular crowdâs status. It cannot scale and lacks the capacity for providing appropriate mobile crowd safety management in real-time. For example, in Mecca, Saudi Arabia, during Hajj season, groups of pilgrims started going in the opposite direction to get to their destination faster. When the crowd flow got clogged from the colliding crowds, it resulted in more than 2000 casualties. Although there were 5000 video surveillance devices installed all around Mecca to monitor the Hajj season [34], the accident was not able to be prevented in time.

1.5.2 ML-Based Video/Image Surveillance

Crowd Counting: This is becoming an increasingly important issue of computer vision, as it has many applications in the context of smart cities, especially pertaining to public safety. The lack of proper crowd safety control and management often leads to human casualties and the spread of infectious disease (i.e., COVID-19) at densely crowded political, entertainment, and religious events. Hence, automated crowd interpretation using AI techniques [38, 39] is becoming an increasingly critical task for many practical crowd safety applications [40, 41, 42]. Although many CNN-based methods have been proposed to improve the performance of complex crowd imaging to deal with variations in scale, perspective, and image resolution [43, 44, 45, 46, 47], they still have significant limitations in the face of occlusions that partially impede sight of individuals in a crowd scene. Crowd images are often scattered with occlusions that makes it difficult to identify the total amount of people in a scene. As illustrated in Figure 31, the types of fixed environmental obstacles such as buildings, big trees, and walls are constrained to specific parts of an image, thus can be easily excluded from the crowd counting area. However, the interpretations of *scattered occlusions (SO)*, such as umbrellas and picket signs, are challenging, as they can obscure the sight of one or more individuals entirely or partially depending on crowd size, density and occlusion types [2]. Despite their commonness in many mass gathering scenes such as sport events, political rallies, or protests, existing approaches fail to accurately count people with the presence of SO in crowd images.

Safety Modeling in Crowd: In battling against the global pandemic caused by COVID-19, many countries and local governments continuously predict the pandemic

trajectories for healthcare-related policy decisions such as opening schools, limiting businesses, and planning infrastructure healthcare facilities and personnel. In addition, growing evidence shows that requiring masks in public areas in parallel with aggressive testing and rapid vaccinations are critical for achieving herd immunity [48, 49]. As illustrated in Figure 41, many scientists, the World Health Organization [50] and the CDC [51] confirms that masking in public places is one of the most effective health measures along with social distancing to break the coronavirus transmission chain. For example, the chance of coronavirus transmission reduces from 90% to 1.5% by wearing masks. Despite such safety recommendations, in reality, face mask-wearing and social distancing are practically practiced at a varied level depending on place and time.

Monitoring and assessing the conformity of mask-wearing and social distancing in public places would provide more accurate input to pandemic trajectory predictions and understand the safety level of communities. However, existing approaches do not directly address the issue and fall short of effective and scalable tasks. There are several ML-based object detection and classification approaches aimed for COVID-19 projections [51], diagnosis, social distancing [52], and contact tracing [53] applications; however, as the technologies to ensure masking in public places aim at surveillance and tracking individuals [54, 55], they are rendered less useful in *public areas*, such as *crowds* on the street, airports, and schools, due to scalability limitation and privacy concerns.

1.5.3 Semantic Knowledge Information-based Tracing Application

COVID-19 causes multiple pandemic waves worldwide for years due to the nature of its long incubation period, the aggressive asymptomatic transmission, and new mutations of the virus. Lockdown measures (shrinking the community at the elementary level, such as a family and limiting contact between communities) and social distancing is the last resort against it. In addition to vigorous testing and vaccination, very aggressive social distancing in public places is a vital part of the strategy for keeping such exponential infections from happening again after the first wave recedes and the society reopens. However, existing technologies, such as contact tracing apps, have not been adopted due to privacy and accuracy concerns. Without achieving a critical mass of individual users, these personal technologies have been rendered useless. Although large-scale policy efforts have been complicated, requiring the coordination of federal, state, and local governments and the logistics of regulation enforcement, applying technologies to small, focused communities can retain individual privacy, achieve wide user adoption, and allow easy implementation. Unfortunately, they are not helpful for the post-pandemic era (a new normal) to prevent another potential future pandemic wave.

1.6 Our Contributions

In this dissertation, we are proposing an intelligent crowd engineering platform using Machine-based Internet of Things Learning and Knowledge Building approaches (ICE-MILK) to enhance the accuracy, scalability, and safety management capacity for crowds in real-time. Specifically, a three-layer ICE-MILK structure, including IoT-based

mobility characterization, ML-based video surveillance, and semantic information-based application layers will be designed. Thus, we summarize the contribution of this dissertation work as follows:

- In the IoT-based mobility characterization, we propose CROMO that enhances crowd mobility characterization through real-time Radio Frequency (RF) data analytics. The system will be constructed by predicting and preventing potential disasters through real-time Radio Frequency (RF) data characterization and analytics. Object group identification, speed, direction detection, and density for the mobile group among the many crowd mobility characteristics will be anticipated for this step.

- In the ML-based video/image surveillance, we propose an architecture for scattered occlusion characterization called CSONet for efficient crowd counting and high-quality density heatmap generation. CSONet recognizes the implications of event-induced, scene-embedded, and multitudinous obstacles such as umbrellas and picket signs to achieve an accurate crowd analysis result. Two new scattered occlusion object datasets, which contain crowd images occluded with umbrellas (cso-umbrellas dataset) and picket signs (cso-pickets dataset), will be used. In addition, we propose a novel face masking detection system for Modeling Safety Index in Crowd (Mosaic), a Machine Learning (ML)-based approach for detecting masking in a crowd by building new dense mode crowd masking datasets. Mosaic detects, counts, and classifies the crowd's masking condition and calculates each community's spatiotemporal Safety Index (SI) values instead of monitoring individual

masking cases using a weighted neighbor relationship to ensure privacy protection.

- For the semantic knowledge information-based application, we applied the previous technologies to group semantics to track and prevent infectious diseases (i.e., COVID-19) and crowd-related accidents. We also propose a novel tracing application named crowd-based alert and tracing services (CATS). CATS applies privacy-aware contact tracing, social distancing, and ML-based mask-wearing principles to provide higher privacy protection, efficient penetration of technology, greater accuracy, and effective practical policy assistance. The system builds a sustainable, safe community cluster against COVID-19 and beyond using affordable Internet of Things (IoT) and edge-enabled technologies. We have integrated CATS components into an edge-based IoT system.

The primary purpose of this study is to improve crowd safety management methods through real-time RF to characterize and analyze the crowd mobility in speed, direction, and density to predict and prevent potential disasters. Also, building a ML-based video surveillance approach for effective dense crowd counting by characterizing SOs in crowd events. We integrated the crowd mobility characterization and the ML-based video surveillance approach to group semantics to track and prevent infectious disease (i.e., COVID-19) and the crowd-related accidents.

1.7 Dissertation Structure

The remainder of this dissertation is organized as follows. Chapter 2 discusses the related work in the IoT-based mobility characterization, ML-based video/image surveillance approaches, and semantic Knowledge application. Chapter 3 encompasses IoT-based mobility characterization and proposes CROMO to improve crowd mobility characterization through real-time RF data analytics. In Chapter 4, the Scattered Occlusions for Effective Dense-Mode Crowd Counting (CSONet) is introduced. Also, an ML-based face masking detection system for Modeling Safety Index in Crowd (Mosaic) will be included. Chapter 5 will elaborate on semantic information-based application layers, including crowd-based alert and tracing services for building a safe community cluster against COVID-19 (CATS) and crowd safety sensing (CroSS) for the post pandemic era. Finally, this dissertation will provide some pointers to future research work in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we discuss the state-of-art research work related to crowd Engineering in different aspects. As mentioned in Chapter 1, this dissertation aims to build an Intelligent Crowd Engineering (ICE) platform by enhancing the accuracy, scalability, and capacity to help ensure public safety and management in real-time. Thus, this related work chapter is organized into three sections, in which each section discusses previous works done at specific application domains.

2.1 IoT-based Mobility Characterization

Crowd monitoring and tracking topics are attracting the attention of researchers due to its importance for urbanizing cities. There have been several recent crowd management studies that address the issue of tracking the massive crowd using video cameras or wireless technologies . In this section, we review the most related works in Radio Frequency RF-based tracking.

2.1.1 RF-based tracking Approach

Radio Frequency (RF)-based tracking such as Wi-Fi, is another method of crowd managing which has many papers proposing solutions for crowd density estimation and tracking [56, 57, 58]. The authors in [56] used coordinated indoor Wi-Fi routers to collect data between TX and RX and used SVM model to train the data to count the number of

people in a room. The study in [57] attempts to localize people at TT festival in Assen, The Netherlands and improved accuracy by de-noising the collected Wi-Fi data. Both studies focus on limited crowd characteristics. Nunes et al. [58] analyzed MAC addresses and associated SSIDs to study the dynamism of tourists, but it is for a static posterior analysis rather than a real-time crowd management. K. Li et al. [59] proposed a framework to capture probe packets sent by smart-phones and use it to monitor crowd density in the indoors. Also, they used RSSI to indicate the closest sensor to smart-phones to collect data to reduce packet duplication. Our work is different where we have focused on tracking crowd mobility indoor and outdoor environments using multiple metrics such as RSSI, beacon count, and time-stamp. Also, our work uses bracelets instead of smart-phones due to data privacy issue. Patil et al. [60] suggested Wi-Fi to track the number of people at a massive event. They capture probe packets of attendees' smart-phone Wi-Fi to estimate the size of crowd. Their work is focused on estimating the number of attendees, while our work focuses on tracking crowd mobility at the events. Unlike the above, we use BLE instead of Wi-Fi due to its low power consumption, low cost, availability, flexibility in size and friendly-wearable. RFID has been one of the most common wireless technologies to identify and track objects with an active RFID tag. Yamin et al. [61] used GPS equipped RFID tags connected to a centralized database to track pilgrims. This method also applies to Al-Hashedi et al. who proposed using RFID connected to a data center to track pilgrims during Hajj [62] and Mitchell et al., who also mentioned the possibilities of using RFID along with smart-phones to track the pilgrims during Hajj [63]. In our work, we are using BLE instead of RFID because the RFID system cannot support any communication based

intelligent monitoring approaches [64]. GPS is a satellite-based system that has been used for navigating and tracking objects in outdoor environments. Blanke et al. [65] suggested using GPS for tracking crowd in large scale areas, but GPS has limitation coverage for indoor environments event, in which it does not support our study in this paper.

BLE is low power wireless technology that has been used to connect smart devices. There are a couple of papers used BLE in their solutions. Basalamah et al. [66] used an active mode Bluetooth Low Energy (BLE) tag. The beacon messages are scanned by smart-phones (detectors). However, the active BLE tags consume the battery power quickly. It also increased the chance of overhead and packet collisions at dense events. In facts, this approach decreased the data accuracy since the people carried the smart-phones (detectors) within the crowd. In contrast, our approach takes a passive mode tag that improves data accuracy, scalability, and power consumption. Weppner et al. [67] used smart-phones to scan for other Bluetooth devices to estimate the crowd density. However, this work did not provide any additional intelligent measurements for managing the crowds. Also, their work mainly focused on assessing the static crowd density, while our work tackles a mobile crowd to handle the crowd mobility and safety [68]. Alessandrini et al. [69] used RSSI in Wi-Fi for localization in the crowd and to track the flow. We are using RSSI in BLE [70] as a tool for localizing objects. Several papers studied RSSI in BLE for indoor localization. Wu et al. [71] used the BLE RSSI captured by three signal sniffers to classify if people are in a queue during the crowd at indoor. Our approach is different because we used BLE RSSI from the respond beacons, then we used RSSI average and variation to detect the crowd density and mobility indoor and outdoor events.

2.2 ML-based Video/Image Surveillance

2.2.1 Video Surveillance Approach

Video Surveillance is one of the most common and traditional ways for safety and security monitoring. Several papers addressed the issue of crowd density risks and proposed videos surveillance to estimate the density or monitor the crowd [36, 72]. However, video surveillance requires manual data analysis; it cannot respond in real-time. In addition, it is not accurate to track and estimate the high-density crowd using cameras because obstacles such as wall, tree and the human body can block the camera's vision from capturing objects [71]. Head detection using cameras is another approach because the human head is the most visible part of the human body at a crowd from the camera's tower. Shami et al. [73] proposed an algorithm that detects peoples' head at a crowd for counting the density using Convolutional Neural Network CNN. However, the accuracy of capturing the human heads can be affected in case if some pedestrians have umbrellas or small persons blocked by large persons from the cameraâs view. Bek et al. [9] proposed an approach to measure the crowd density flow for congestion risk assessment. The study used single camera tracking example without taking into consideration that large crowd required more than a camera, so in case of multiple cameras may have different measurements in tracking the moving crowd for risk assessment [74]. Alahi et al. [74] proposed unsupervised technique learning to match multiple camera single-view in tracking pedestrians by estimating the distance between pair cameras. Matching the view of multiple cameras to track pedestrians is a great effort. However, their work did not take into account object navigating. For example, in the case of out-view objects during moving from

one camera's angle to another or in the case of obstacles blocking a cameraâs view of an object. Therefore, due to the accuracy and capability limitation of the video surveillance approaches, it cannot manage crowd events alone.

2.2.2 Crowd Counting Approach

There have been significant studies and remarkable improvements made in crowd counting and density estimation. Traditional non-machine learning methods can be broadly classified into three categories, namely, detection-based, regression-based, and density estimation-based approaches [75]. Despite various advancements, those approaches have shortcomings of complexities and limited accuracy. In recent years, researchers mostly have adopted machine learning techniques to overcome those weaknesses. In this section, we briefly highlight noticeable prior studies.

2.2.2.1 Traditional Methods

A number of early methods have attempted to tackle the challenges of crowd counting and density estimation via implementing detection-based approaches. Generally, these methods use a detector or classifier to recognize a human's whole or body part to estimate the crowd count. Dollar et al. [76] applied a sliding window detector to extract the features from the input image and determine the human count. Most of the methods focused on extracting features, such as histograms of oriented gradients HOG [77], and Haar wavelets [78] from the crowd images to learn the density and the count. However, the counting results of the whole body methods perform poorly in highly crowded images. Although a part-based detector is proposed to detect the density of people in a

crowd [79, 80], these methods still face difficulties in locating people, especially when the crowd in a scene is highly occluded or densely populated, as it happens often in various events. Regression-based approaches have been proposed to tackle the limitations of the detection-based method, concentrating on the difficulty of detecting the count in a highly dense crowd scene. Regression methods aim to learn the mapping between extracted features from the image and the count number of objects [81, 82]. Regression methods typically have two main components: low-level feature extraction and regression modeling [83]. Density estimation-based methods are another approach for crowd counting and density estimation. Researchers have successfully addressed the issues of occlusion and clutter by using regression-based methods. Nevertheless, some of the existing techniques overlooked spatial information, which affected the result of counting. In contrast, Lempitsky et al. [84] proposed a supervised learning framework to estimate the count of objects in images. They used a linear mapping technique that focuses on the density through learning the relationship between the local image features and object density maps. However, Pham et al. [85] observed the limitation of the linear mapping and proposed a random forest framework to learn a non-linear mapping between local image features and density maps.

2.2.2.2 CNN-based Approaches

Several studies have proposed Convolutional Neural Networks (CNN) based approaches for crowd counting and density estimation. Those methods have obtained a significant improvement in crowd counting and density estimation addressing various kinds

of challenges, such as perspective, image resolutions, occlusions, and non-uniform environment. Here, we briefly summarize various of the recent methods for crowd counting and density estimation in terms of their CNN architectures. A multi-column CNN architecture called MCNN was proposed in [86] to estimate the crowd count and density map in an arbitrary crowd image. The CrowdNet method proposed in [87] is considered one of the early CNN based architectures inspired by VGG16 [22]. The CrowdNet combined convolutional networks and a shallow network to learn robust scale features and to generate the density maps. Cao et al. [88] present an encoder-decoder network called a scale aggregation network (SANet). The encoder layer extracts the multi-scale features, and the decoder will generate high-resolution density maps. Also, training loss is introduced by combining euclidean loss and local pattern consistency loss, which contributed to improving the final count. Li et al. [89] introduce the congested scene recognition network (CSRNet), which is one of the state-of-the-art in terms of performance among the ones inspired by VGG16. It consists of two essential components: CNN as the front-end layers and dilated convolution layers as the back-end. Zhang et al. [90] proposed a method that generates a probability map and presents the high expectations indicated in locations where heads are possible to be present. CANet [91] proposed a deep network architecture that performs multi-level feature comparison between the support and the query images and iterative refinements of the results. Chen et al. [92] proposed a scale pyramid network (SPN), which consists of a single column structure to extract multiple-scale features by dilated convolutions with various rates. ASNet [93] is also considered as a state-of-the-art method. It contains a density attention network that generates attention masks, and then

provides it to attention scaling network in order to generate scaling factors outputting attention-based density.

Despite improvements achieved by such recent approaches, the accuracy of crowd counting can significantly diminished in the presence of SOs in a crowd scene. Our method address the very issue of SOs and achieves a high accuracy of crowd counting in the presence of such SOs.

2.2.3 Modeling Safety Index in Crowd

While there are a number of image and video processing techniques that could be used for COVID-19 pandemic related analysis, they are limited to individual analyses rather than a scalable crowd analyses that can be practically used for crowded or big public areas such as streets, airports, and schools. ML methods have significantly improved the image classification networks in many aspects, especially the popular use of Convolutional Neural Networks (CNN) techniques. Since 2020, researchers have attempted to develop and deploy object detection and classification tools to prevent the spread of the COVID-19 virus by detecting the face masking in public places.

A face mask detection method is proposed that can classify the face masking conditions into three categories: correct face masking, incorrect face masking, and no face masking [94]. Meanwhile, the Principal Component Analysis (PCA) algorithm has been deployed to detect the masked and no-masked face [95] based on the observation that face recognition accuracy is extremity affected by wearing masks using the PCA algorithm. A real-time face emotion classification approach is proposed based on deep learning [96],

which adopted VGG16 to classify facial expressions. A face mask classifier has been implemented using You Only Look Once (YOLO) v3 [97], which uses Darknet-53 as the backbone and achieved a high accuracy in face mask classification. A hybrid deep learning model was proposed for face mask detection [98], which consists of two components, the first component used Resnet50 to extract the feature, and the second part used the concept of Support Vector Machine for classification. Most of the proposed face masking detection methods have used deep learning methods such as ResNets [99], Vgg-16 [22], Fast R-CNN [100], Yolo v3[27], and Yolo v5 [28]. While multiple face mask datasets have been constructed [101, 102, 103, 104, 105], none of them includes the scenarios of dense mode. This missing link and the compelling need inspired us to collect and annotate a new face mask dataset that contains crowd density models from low to high. In addition, face mask detection alone is not sufficient to prevent the transmission of the COVID-19 virus. An intelligent masking detection system is highly desired that can detect, count, and classify the crowd's masking conditions and generate SI values for each community instead of detecting individual masking cases.

Our work uniquely fills the current gap between high-level pandemic prediction model that does not take the reality of mask-wearing practices and individual level image processing that is not effective for scalable crowd analysis. Mosaic assesses the various types of mask-wearing practices from low resolution crowd images, and provide spatio-temporal SI in a scalable and effective manner.

2.3 Semantic Knowledge Information-based Tracing Application

In this section, we examine the existing solutions. We are well aware of the recent project activities of contact tracing in the U.S. (e.g., Apple and Googleâs contract tracing API [106], PACT [107] [108], PrivateKit [109], etc.) as well as the effective practices in countries in East Asia in combating this pandemic such as South Korea, Taiwan and Singapore [110]. While the success of contact tracing in East Asian countries is attributed to extensive testing and strong government coordination, the compromise of personal privacy was main weakness [111]. In contrast, US society highly values privacy and is distrustful of mass government surveillance. Policy decisions in the US are much more complex due to federal and state power divisions and the diverse populace. According to the recent polls [112], 71% of respondents said they have no plans to download and use a contact tracing app. Additionally, 44% expressed concern over digital privacy, 39% said the app gave a false sense of security, 37% believed the apps would not slow the spread of COVID-19, and 35% cited their distrust of the app providers. However, another survey shows 70 - 80% of Americans are willing to install an app if they are perfectly private and accurate, which is a significant increase. Many of RTLS companies, including Pozyx [113], Tsingoal [114], Localino [115], Iterate Labs [116], Arin [117], and RightCrowd [118], have already commenced COVID-19 contact tracing and social distancing application systems. Start-ups, enterprise/commercial GPS companies, and carriers are drawn to this space with wearables and the Internet of things apps. Several

commercial devices available for tracking such as Filip Technologies Inc. [119], Location Based Technologies, Inc. [120], Amber Alert GPS [121], Wonder Technology Solutions [122], hereO [123], Quattro [124], and Masternaut [125]. More examples include [126, 127, 128, 129, 130, 131, 132, 133, 134]. However, most of them are expensive and fundamentally rely on GPS that is not available in-doors, nor is it energy-inefficient. Furthermore, these devices use cellular communication for one-to-one communication. Thus, they incur high monthly charges, and the efficacy of monitoring is limited. Most of all, none of them addresses the issue of privacy that significantly impacts the crowd participation. According to ABI Research, the significant barriers of personal location devices and applications market have been expensive devices, cellular subscriptions, indoor locations, and severe regionalization and fragmentation of coverage [135]. In summary, the existing approaches have the following significant limitations.

- Privacy/Technology adoption: Many individuals in the US do not have smartphones or will avoid any kind of contact tracing such as children, people with disabilities, people who are undocumented or have family members who've been in trouble with the law, will deteriorate the value of the system.

- Accuracy/False positives: It is known that signal alone cannot clearly distinguish the existence of walls or barriers between contacts.

- Binary tracing information: Different types of contacts (thus different potential impacts of contacts) are ignored. Thus, it only acts as a less effective backtracking mechanism after a potentially long asymptomatic incubation period.

Likewise, once the technical issues such as privacy and accuracy on the contact tracing and social distancing are resolved effectively, the community's adaptation will be dramatically increased. The core of our idea is to efficiently utilize crowds/communities to protect people from pandemic outbreaks using innovative technologies.

CHAPTER 3

IOT-BASED MOBILITY CHARACTERIZATION

Human casualties at entertaining, religious, or political events often occur due to lack of proper crowd management. Notably, for the crowd in mobile, a minor accident can create a panic for the people to start stampeding and trampling others. Although many smart video surveillance technologies are recently proposed, it is still very challenging problems to predict a crash in real-time among the mobile crowd for preventing any potential disaster. In this work, we propose CROMO that enhances crowd mobility characterization through real-time Radio Frequency (RF) data analytics. Inspired by the recent advanced artificial intelligence (AI) technology and machine learning (ML) algorithms, traditional video surveillance technologies make object detection and identification possible in real-time. However, their scalability and capacity lack in a crowded mobile environment. CROMO propose to fill the gap via RF signal analytics. Among the many crowd mobility characteristics, we tackle object group identification, the speed, and direction detection for the mobile group. We also apply them to group semantics to track the crowd status and predict any potential accidents and disasters. Taking advantage of power-efficiency, cost-effectiveness, and ubiquitous availability, we specifically analyze a Bluetooth Low Energy (BLE) signal. We have tested CROMO in both a practical crowd event and the controlled indoor and outdoor lab environments. The results show that CROMO can detect the direction, the speed, and the density of the mobile crowd in

real-time. Therefore, it can help the crowd management in avoiding disasters possibilities at crowd events.

3.1 Introduction

Due to the unprecedented scale and speed of urbanization, cities are facing the daunting task of accommodating the urban dynamics. The concept of Smart Cities attracts city planners and researchers as it facilitates many smart community services by combining cyber-physical systems and social entities through the wireless, mobile, and intelligent information and communication technologies (ICT). One of the critical service requirements of future cities is the safety management for citizens and communities [34]. Specifically, the safety management during the densely populated events such as religious, entertainment (such as sport and music), and political gatherings becomes more significant as it happens more frequently and in a large scale in modern cities. Unlike static crowd events where a crowd is formed in a specific location when a crowd is moving from a location to another (i.e., unidirectional), it requires more space (i.e., less density). If crowd mobility exhibits multiple non-unidirectional patterns, it would require even more space to be safe and is much harder to manage or control them [35, 2]. Risks of human casualties at mobile crowd events are more likely to occur because small incidents at the crowd are enough to cause panic to the attendees to start hustling, collapsing, trampling, and stampeding each other. Any inappropriate crowd management often results in disastrous repercussions such as injuries and casualties [9]. Table 2 shows many recent cases of crowd disasters that cause human losses around the world [136]. The community

stakeholders need to provide their best efforts to maintain the crowds properly.

Table 2: Crowd caused disasters around the global.

Location (Country/City)	Human Casualties	Year	Event Type
Saudi Arabia, Mecca	> 2000	2015	Religious
Brazil, Santa Maria	> 242	2013	Entertainment
Egypt, Port Said	> 74	2012	Sport
India, Jodhpur	> 224	2008	Religious
Iraq, Baghdad	> 935	2005	Religious

The recent smart video surveillance inspired by the advanced Artificial Intelligence (AI) technologies and Machine Learning (ML) algorithms enables a broad spectrum of promising safety applications, including object detection and identification, behavior recognition and tracking, and anomalous event detection [36, 37]. However, video surveillance alone cannot identify and predict particular crowd status. It cannot scale and lacks the capacity for providing an appropriate mobile crowd safety management in real-time. As illustrated in Figure 10, the image shows very high density crowd that is located on a bridge. However, it does not reveal the group identity and location within the crowd and their moving direction and speed. For example, in Mecca, Saudi Arabia, during Hajj season, groups of pilgrims were taking the opposite road direction to get to their destination faster. When the crowd flow got clogged from the crowd flow from the opposite direction, it resulted in more than 2000 casualties. Although there were 5000 video surveillance installed all around Mecca to monitor the Hajj season [34], the accident was not able to be prevented in time.

Figure 10: Crowd Image Surveillance Illustration.

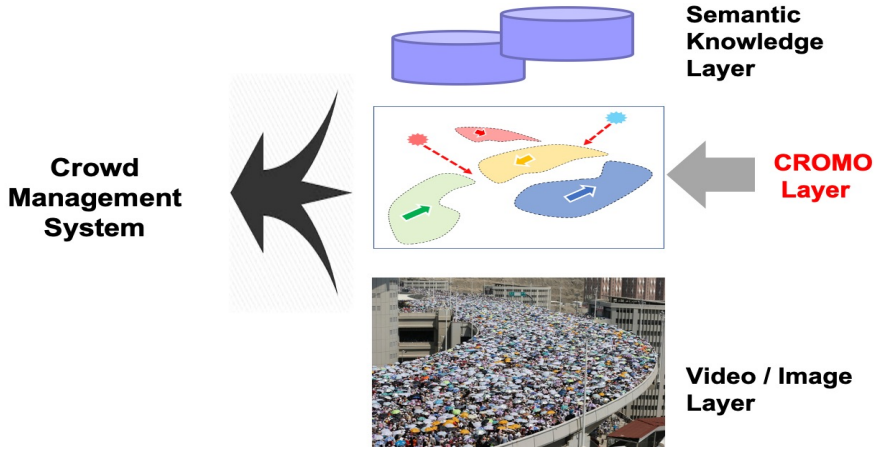

Figure 11: CROMO Layer Architecture.

In this work, we propose CROMO that enhances crowd mobility characterization through real-time Radio Frequency (RF) data analytics [11, 137]. It enhances safety management for a mobile crowd events by predicting and preventing potential disasters

through real-time Radio Frequency (RF) data characterization and analytics. The motivation is to improve safety management method for the mobile crowd by filling up the scalability and capability gaps of the existing video surveillance via tightly integrating RF signal analytics (Figure 11). It implements a wireless-based, efficient and a scalable crowd/group tracking technology. Specifically, we exploit a tracking bracelet and monitoring infrastructure as well as a couple of abnormality scenarios and prediction algorithms. CROMO uses a Bluetooth Low Energy (BLE) [138] communication in this project as it is power-efficient, cost-effective, and ubiquitous [139]. Among the many crowd mobility characteristics, by using passive BLE scanners, we measure the number of beacons, the Radio Strength Signals Received (RSSI) value, and its variation pattern. Integrating the received beacon values, CROMO can identify the crowd density, the object group location, and the flow direction and speed. We have conducted various practical mobile crowd tests in both indoor and outdoor environments under different crowd conditions from low to high-density. We also apply them to group semantics to track the crowd status and predict any potential accidents and disasters. For example, participant's data such as names, ID numbers, group ID, destination locations, contacts, and necessary health information can be registered into the tracking bracelet. In case of an emergency, the data could be used to facilitate help by the public safety personnel. By using a couple of algorithms, CROMO is able to predict a few potential problems in the mobile crowd scenarios. CROMO detects speed among the moving groups as well as identify a potential collision by measuring the flow density. The proposed monitoring approach is explicitly designed for densely crowded environments.

3.2 System Architecture

Mobile crowd management is one of the hardest tasks because predicting human behavior during a crowded event is extremely difficult. In CROMO, we study the feasibility of using BLE beacon signals from various BLE transmitters for tracking the mobile crowd status.

3.2.1 CROMO Layer

According to the crowd safety and risk analysis [10], understanding the impact of crowd density (the number of people per square meter) for both a standing crowd and a mobile crowd is critical for managing crowd safety. For example, to assess the efficiency of crowd movement, a capacity of places, it needs to understand the relative risks of both standing crowd density and the moving crowd density. In some case, if a standing crowd becomes mobile and a unidirectional crowd becomes non-unidirectional, the planned capacity design fail. It can cause any unexpected disasters. As illustrated in Figure 12, when the crowd density (the number of people per square meter) increases, the comfort level of the crowd decreases and flow speed starts to decrease as people cannot take full paces forward. After a saturation point, crowd mobility becomes constrained and accumulated, and the flow rates dramatically drop. For the crowd moving in the same direction, when density becomes more than three people in a square meter, the flow speed starts to decrease, and when density becomes more than four people in a square meter, the flow speed drops to become a high-risk crowd. However, when the crowd moves randomly in different directions, the flow speed decreases significantly at the lower

density and even at the density of 3 or 2.5 people becomes a high-risk crowd. In a low-density case, a collision can be avoided by stopping the flows. However, in a high-density case, when crowd force pushes forward the people in front and shock waves began to ripple through the tightly packed mass, it causes a crush and crowd disasters. The crowd safety management should be able to predict the potential flow directions well before the crowd is getting into a high-crowd condition.

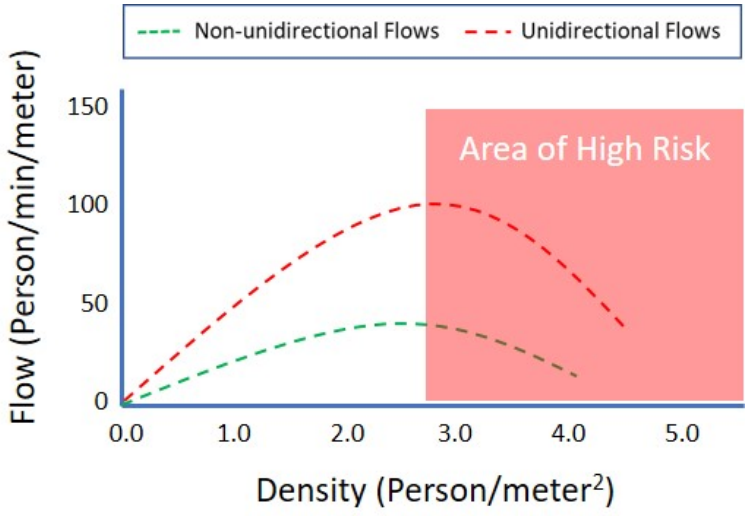

Figure 12: Mobile Crowd Density vs. Flows.

As shown in Figure 13 [37], using the advanced artificial intelligence (AI) technology and machine learning (ML) algorithms, the intelligent video surveillance enables to detect and track multiple moving objects. However, it cannot scale to monitor various objects in a high-density crowd due to the limitation of visual processing. Also, video surveillance cannot follow the moving objects if obstacles or another human block them. It is hard to handle the hand-over from one camera angle to another [74]. Furthermore,

the video surveillance alone cannot identify and predict particular crowd status such as group semantics. For example, it does not expose the group identity and location within the crowd and their moving direction and speed. Hence it alone lacks the capacity of providing an appropriate crowd safety management in real-time.

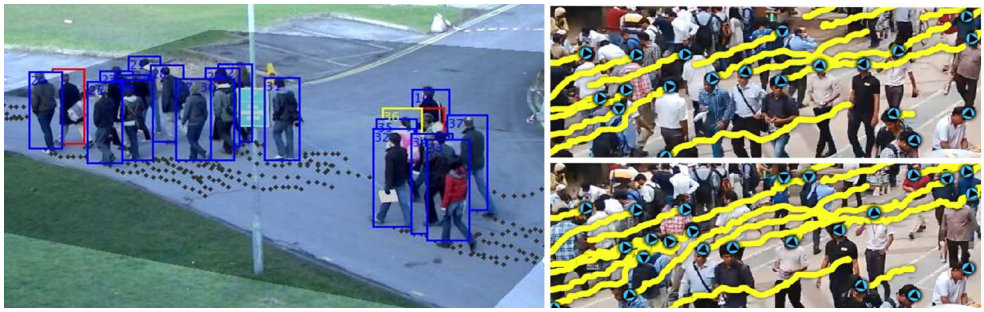

Figure 13: Smart Video Surveillance.

CROMO enhances safety management method for the mobile crowd by harnessing BLE signal data analytics layer (in Figure 11) over the existing video surveillance. Among the many crowd mobility characteristics, by using a BLE bracelet and BLE scanners, CROMO measures the beacon counts, the RSSI power, and its variation pattern. Although these metrics are used in various applications, their behavior in a high-density is not well known. By integrating the parameters and application-specific semantics over the video surveillance, CROMO can identify the crowd density, the object group location, and the flow direction and speed in both indoor and outdoor environments. CROMO also can predict any potential accidents and disasters. For example, participant's data such as names, ID numbers, group ID, destination locations, contacts, and necessary health information can be registered into the tracking bracelet. By using group speed and direction

detection algorithms, CROMO can predict a potential collision in various mobile crowd scenarios.

3.2.2 CROMO Design

CROMO uses a Bluetooth Low Energy (BLE) communication. As illustrated in Table 3, BLE is known to be more energy efficient than other wireless technologies such as classic Bluetooth and Wi-Fi [140]. The coverage range of BLE, over 100 m, is as good as others. It is enough to cover the densely populated crowd area. CROMO consists of the BLE tracker bracelets worn by a human, the BLE scanners, and the scanning algorithms. Each BLE tracker bracelet has a unique identification-ID to identify each bracelet.

Table 3: RF Transmission Approaches.

Protocol	Range	Mobility	Deployment
BLE	$\geq$ 100 m	$\leq$5 Mph	Ubiquitous, Low power usage, low association time
WiFi	$\geq$ 100 m	$\leq$5 Mph	Ubiquitous, Low power usage, high association time
Cellular	$\geq$ 10 Km	$\geq$60 Mph	Ubiquitous, Low power usage, high association time

3.2.2.1 BLE Scanning Approaches

In a wireless communication system, there are a couple of common messaging modes. In a passive mode, a node does not send any periodic message but scans incoming messages. In an active mode, a node periodically sends messages to indicate its existence. As illustrated in Figure 14, the combination of these two modes are used between the BLE scanners and the pedestrians' bracelets (BLE trackers) to communicate the crowd states. A tuple of the BLE tracker mode and the BLE scanner mode approaches including

an active-active approach (AAA), an active-passive approach (APA), a passive-active approach (PAA), and a passive-passive approach (PPA) are investigated to ensure efficiency in power usage, increase the scalability in message communications, and improve the accuracy in event detection.

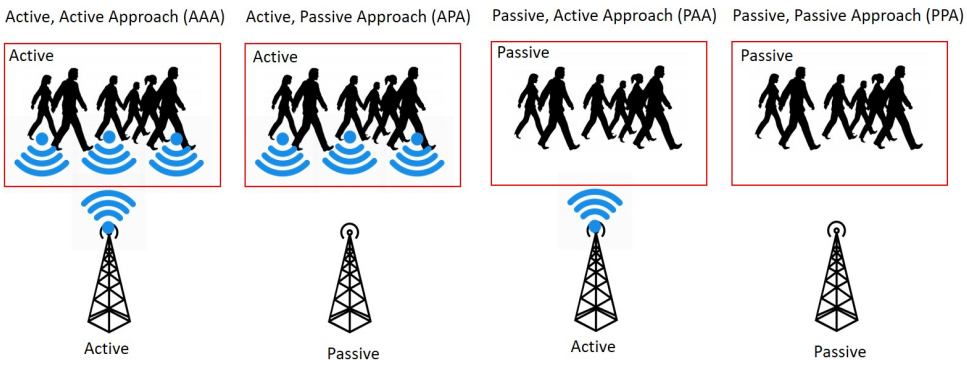

Figure 14: BLE Scanning Approaches.

- **An Active / Active Approach (AAA)** has a better chance of capturing most of the pedestrians' bracelets data because the BLE scanners point is sending a request and waiting for a response, in the same time the pedestrians' bracelets are on active mode, sending requests, and waiting for responses. However, this will cause higher power consumption on both sides, which is one of our significant concerns in improving battery consumption. That does not only increase battery consumption, but it also increases the number of requests and responses causing overhead and increasing the probabilities of data collisions.

- **An Active / Passive Approach (APA)** is a passive scanning approach for the BLE scanners. The pedestrians' bracelets keep on sending their locations. The BLE

scanners listen and respond upon receiving messages. The pedestrians' bracelets are always active, the power consumption on the bracelet is one of the main concerns, and one of the primary research focus is to reduce the power consumption on the pedestrians' bracelets. Also, the number of messages is proportional to the number of bracelets. Hence, for a densely populated environment, the approach may increase the message overhead and cause a high chance for message collisions.

- **A Passive / Active Approach (PAA)** is a BLE scanner driven approach. The BLE scanners are sending polling or probing requests. The pedestriansâ bracelets are listening and responding upon receiving the request messages. As the pedestrians' bracelets are passive, the power on the bracelet is efficiently utilized. Also, the number of messages is kept to a minimum as the bracelets are responding upon the requests. As the control is in the BLE scanner side, the responding messages from the bracelets can be efficiently controlled as well. The approach can decrease the message overhead and maintain a smaller chance of message collisions.

- **A Passive / Passive Approach (PPA)** does not perform any active probing. Both the BLE scanners and the pedestrians' bracelets are on listening mode. Although this approach can save power usage, it does not provide any meaningful information about the moving groups. A possible option is if the control room can detect a low traffic situation (during the off-peak times) by using other methods such as CCTV.

In a passive BLE scanner mode, BLE scanner does not perform any periodic active probing by assuming that BLE tracker periodically sends beacon messages. However,

if the BLE tracker bracelets sends beacon messages periodically, the power consumption on each bracelet is one of the primary concerns. Besides, the number of messages is proportional to the number of bracelets. Hence, in a densely populated crowd, the passive BLE scanner approach can significantly increase the message overhead and cause a high chance for message collisions. Meanwhile, in an active BLE scanner mode, a BLE scanner periodically sends polling or probing requests. The BLE tracker bracelets are listening and responding upon receiving a probe message. As the BLE tracker bracelets are in a listening mode without periodically sending beacon messages, it can maintain its power consumption efficiently. Besides, the number of beacon messages is kept to a minimum as the bracelets are responding only to the requests. As the control is in the BLE scanner side, it can adequately control the number and period of beacons according to the size and density of the crowd. The approach can decrease the message overhead and maintain a smaller chance of message collisions. CROMO uses **a passive BLE tracker mode and an active BLE scanner mode approach (PAA)**. An CROMO probing message includes a sampling factor by indicating a replying pattern. For example, a BLE scanner specifies the BLE IDs for a few specific groups. Using this methodology reduces the probabilities of collisions. It also helps to decrease the power consumption of the BLE tracker bracelet because only the bracelet with the specified ID will reply in response to the polling messages.

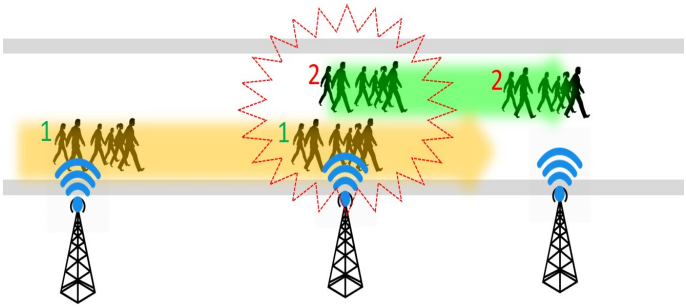

Figure 15: Group Speed Detection Illustration.

3.2.2.2 Crowd Detection Scenarios

Group Speed and Direction Detection: Even if the crowd is moving in the same direction, the moving speed can be different from one group to another. It may cause congestion and collision by the fast-moving groups. For example, Figure 15 illustrates the scenario where a couple of groups are walking in the same direction. However, a fast-moving group 1 in the back causes congestion by taking up the front group 2. CROMO can detect the moving speed of each group from the time-stamp and the distance from BLE scanners at point A to B. If the speed of each group can be detected earlier, the system could predict any potential collision. Assuming the average walking speed of a human is 3.1 mph (5 kph) [138], CROMO can identify the speed of each group and gives a warning to the fast-moving groups or members. Furthermore, if the group 2 is moving slower than the average speed, the system can alert the group 2 to speed up or give a slow down warning to the group 1. The BLE scanners also coordinate to detect the movement direction of each group. By comparing the time-stamps of each groups' passing position, CROMO can also detect the moving directions of each group. For example,

Figure 16 illustrates that a group 2 is moving from the BLE scanner point A to point B. However, assuming that group 1 is supposed to move in an opposite direction according to the schedule, CROMO can identify a potential wrong direction of the group 1 (or a temporary backward movement). The wrong movement shall be alerted to the group and other neighbor group as it may result in a collision with upcoming group 2.

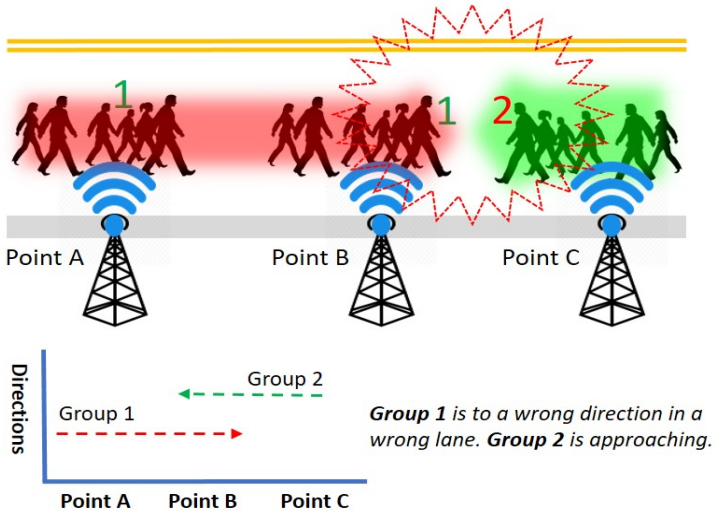

Figure 16: Wrong Lane Group Detection via Direction.

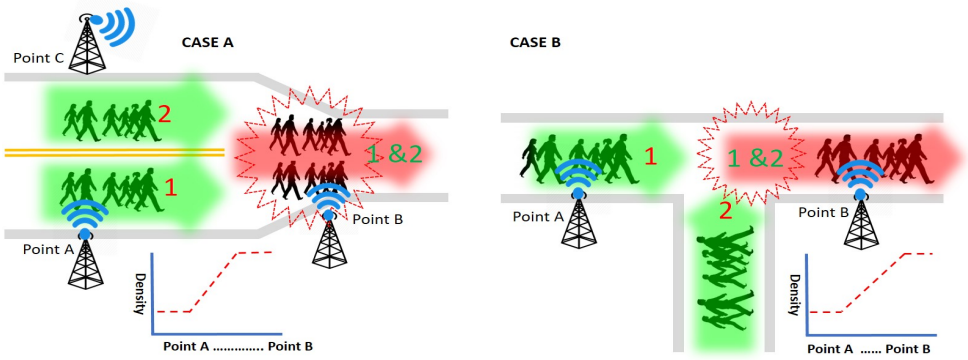

Figure 17: Lane Merge Detection via Density.

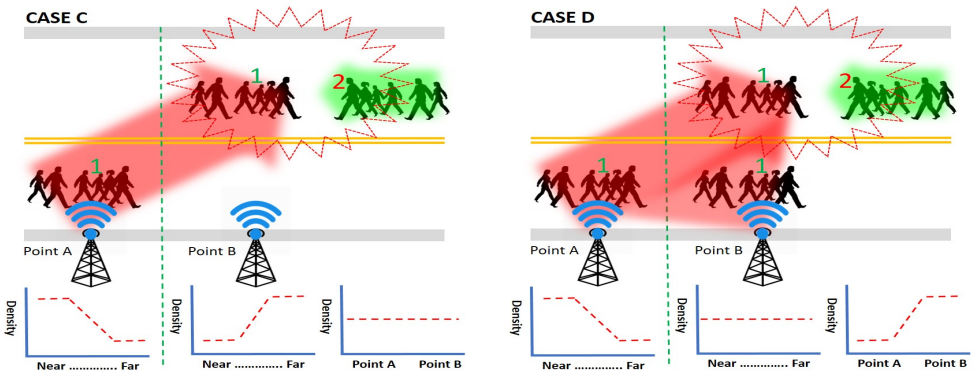

Figure 18: Wrong Lane Group Detection via Density.

Group Density Detection: CROMO can scan the density of each group within the monitoring range and shares the information with the neighbor scanners in real-time. In practice, a crowd collision (i.e., unexpected high density) often happens for various reasons. It can be due to the structure of the road as well as human errors. For example, as illustrated in Figure 17, two lanes are merged into one lane and various intersections mix flows into several directions. Also, shown in Figure 18, a group of people in a crowd may be veering off an opposite lane (when it is available) instead of using the slow and crowded path so that they can move faster. These cases can be identified by tracking the density changes on each point. Both CASE A and CASE B illustrate scenarios where group 1 and 2 are merging into one lane due to the road design either merging lanes or intersection. BLE scanners can detect both cases by checking the density distribution changes on a scanning point B. The density at the BLE scanner point B becomes higher than the density at the BLE scanner point A. CASE C illustrates a scenario where group 2 is, all of a sudden, changing its path in the middle of the road to veer off the opposite

lane. CASE D is a similar scenario, but group 2 takes both lanes. BLE scanners can detect CASE C by checking the density distribution changes on a scanning point. At the BLE scanner point A, the density near the scanner is high, but the far side is low. However, in the BLE scanner point B, although the total density is the same, the density near the scanner becomes low and the far side changes to high. BLE scanners can also detect CASE D by checking the density changes. At the BLE scanner point A, the density near the scanner is high, but the far side is low (like in CASE C). However, the density at the BLE scanner point B becomes higher than the density at the BLE scanner point A, as well as the density near the scanner and the far side of the scanner, become comparable. The density distribution can be measured by multiple scanners using the RSSI power and beacon counts (shown in experimental results).

3.3 Evaluations and Results

As CROMO is a mobile cyber-physical system in a densely populated environment, the communication feasibility issue should be evaluated in a real environment [9]. Hence, we verify how human object movements could affect the BLE signal in a crowded environment.

3.3.1 Experimental Setup

We conducted experiments at a couple of different settings. First, we experimented during one of the largest events at UMKC called âCulture Nightâ, where around 1300 people from different countries are presenting their cultures in a conference hall. The goal of this experiment is to study how significantly interference and crowd density

could affect CROMO. As the conducted experiments were mainly for the feasibility analysis, we wanted to see if the BLE signal from a BLE transmitter tag can indicate any crowd status. For the purpose, only one BLE transmitter within the crowd was carrying a smartphone with an application that advertises its BLE signal. Also, to explore multiple scenarios during the event to establish a better understanding of crowd status.

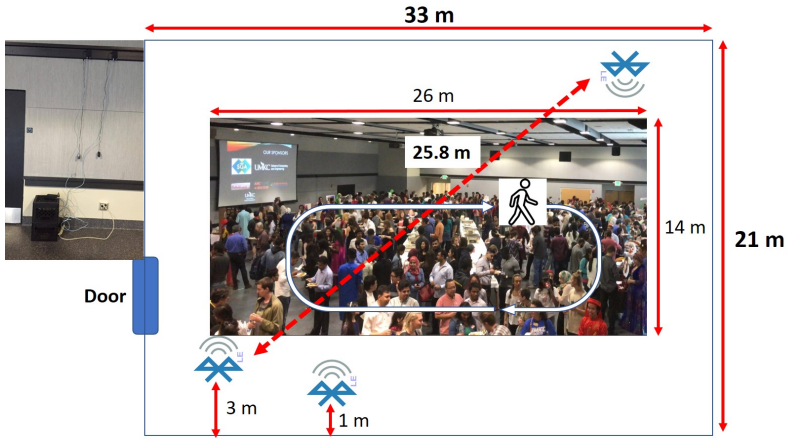

Figure 19: Culture Night Indoor Experiment.

As illustrated in Figure 19, a conference hall size is about 700 m^2 (33 m * 21 m), and people area is about 370 m^2 (26 m * 14 m). We posted a couple of BLE scanners on an opposite side placed at the height of 3 m. The distance between the two BLE scanners was about 25 m. We also put another BLE scanner placed at the height of 1 m next to the door. A moving human object walks around the hall in a circle.

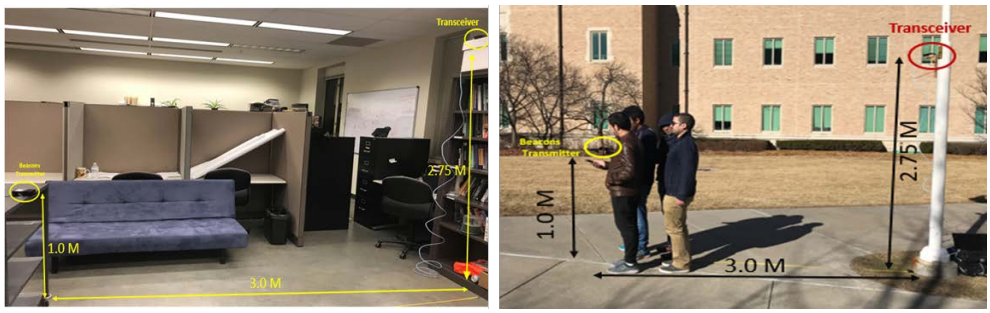

Figure 20: Indoor and Outdoor RSSI Experiment Setup.

Measurement Parameters	Beacon Count, RSSI power, and RSSI variation		
Computed Metrics	Density, Location, Speed, and Direction		
System Settings	**Type**		**Value**
	Environments		Indoor, Outdoor
	Scanner Position		1 and 3 m
Workload Settings	Crowd Density	No Crowd (NC)	Less than 10 (0 per 1 m^2)
		Medium Crowd (MC)	500 ~ 700 (1.5 per 1 m^2)
		High Crowd (HC)	Over 1000 (3 per 1 m^2)
	Human effect	No Human Interference (NHI)	0 human
		Single Human Interference (SHI)	1 human
		Multiple Human Interference (MHI)	3 ~ 4 human

Figure 21: Experimental Settings.

Second, we conducted RSSI experiments in both indoor and outdoor. As shown in Figure 20, we place a BLE transceiver at 2.75 m height and a BLE transmitter at 1 m height in 3 m apart. We put human interference from none to 3 or 4 people near the BLE transmitter. We set up the transceiver to scan for a minute each test, with a total of ten separate times for each test session.

As illustrated in Figure 21, we measure beacon count, RSSI power, and RSSI variation metrics to detect crowd density, location, speed, and direction. We test in both indoor and outdoor environment with different scanner positions (1 m and 3 m) as the system settings. The workload parameter, in Figure 21, consists of a couple of sets. The crowd density parameters characterized in No Crowd (NC), Medium Crowd (MC), and High Crowd (HC). For example, given 370 m^2 human area of the hall, when there are 1000 people, it is about three people in 1 m^2. We classified it HC as it belongs to the high-risk case for the free moving environment. In the RSSI experiment, the human effect parameter consists of No Human Interference (NHI), Single Human Interference (SHI), and Multiple Human Interference (MHI). As the RSSI testing environment is 3 m distance, a few human-objects can make a high crowd effect.

(they are in order now.)

3.3.2 Beacon Count Tests

Counting beacons for a given time to find a population seems to be a straightforward approach. However, when there is a mass of crowd, the result may not be the same due to the collision and interference.

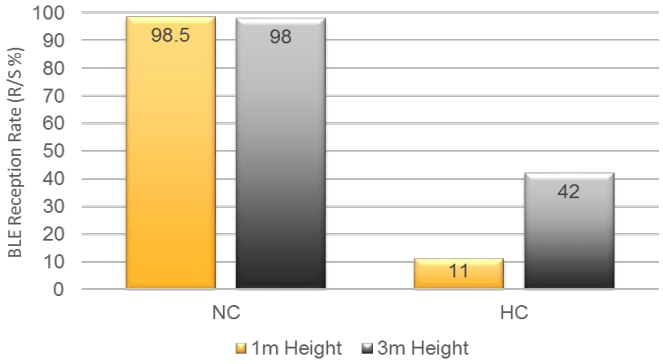

Figure 22: BLE Reception vs. Scanner Height in Crowd.

Figure 22 presents the results of the BLE reception rate (i.e., received (R) percentage of sent (S) beacons) for the different heights in both NC and HC environments. We setup the BLE transceiver heights for 1 meter (low) and 3 meters (high), respectively. In general, it shows that the higher crowd (HC) there were, the higher number of beacon messages were dropped. The result shows that the message-receiving ratio in HC environment was about 50% less than that in NC environment. Furthermore, the reception rate of the higher BLE detector (3 meters) was about 31% higher than that of the lower BLE detector (1 meter). In this result, we can see that the efficient BLE detector location (in height) is above every human height to avoid any signal absorption by the crowds.

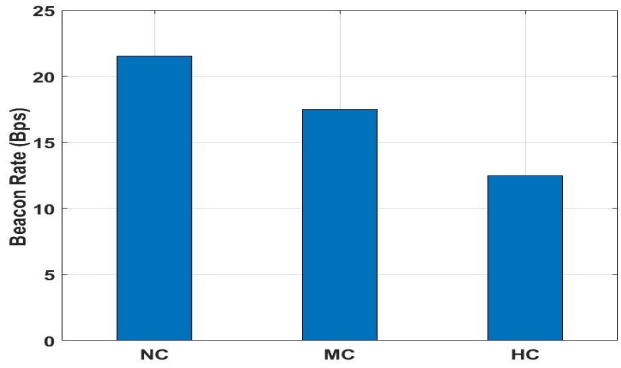

Figure 23: Beacon per Second (Bps).

Figure 23 shows the average received beacons per second (Bps). For the experiment, we configured the BLE transmitter's beacon advertising interval to 20 ms. Adding a random delay of 0–10 ms and a scan interval of 10 ms, a BLE scanner can get a beacon about every 40 ms, which means it receives around 24–25 Bps. The results show that in the NC environment the received beacons are around 23 Bps while in the HC environment it decreased to 13 Bps. The results indicate that human object certainly has an impact on the received beacon count. The effect is proportional to the amount of the crowd.

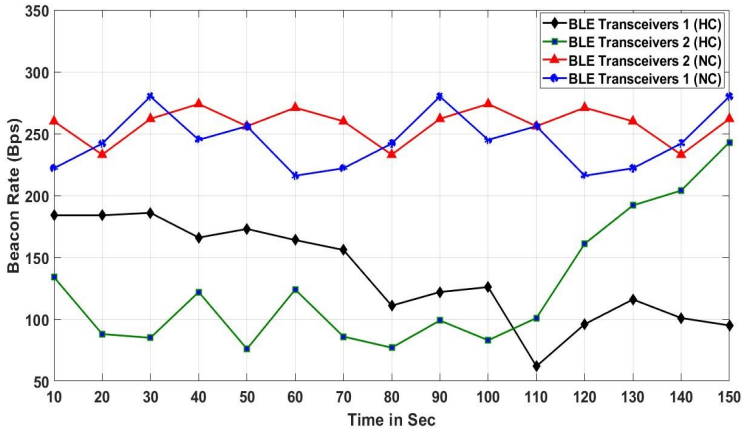

Figure 24: Beacon Counts for BLE Scanners.

Figure 24, compares the results of beacon counts on two separate BLE scanners (posted in an opposite side of a hall) for both NC and HC conditions while an object is moving in a circle as showing in Figure 19. The HC result exhibits a pattern that the beacon count increases when a moving object approaches a specific BLE transceiver, while the beacon count decreases in the other BLE transceiver. In addition to exposing the proximity of the moving object, the result also infers the moving speed and direction (i.e., for two BLE transceivers, it will display approaching and going.) We observed the pattern becomes evident when the crowd density increases. Meanwhile, the NC result shows that both BLE transceivers received almost the same number of beacons regardless of a moving object location. As a result, it indicates that when the environment is NC, object tracking is not possible. However, receiving the similar amount of beacons on both BLE scanners and the beacon counts are more than HC condition, we can observe that the crowd condition is at low risk.

3.3.3 RSSI Tests

In addition to beacon counting, we further measure a couple of BLE RSSI metrics including the RSSI power and variation. In both indoor and outdoor environments, we tested under three workload conditions including NHI, SHI, and MHI. The indoor test results in Figure 25 show that the average RSSI power in NHI is stronger than both SHI and MHI, while there is no significant difference in the signal power average between the SHI and MHI. On the other hand, the outdoor average RSSI power results in Figure 26 show that NHI receives stronger signal RSSI power than SHI, and SHI has stronger signal

RSSI power than MHI. As shown in Figure 27, the total indoor RSSI power of both SHI and MHI is the same, while the total outdoor shows each NHI, SHI, and MHI RSSI powers show a clear difference.

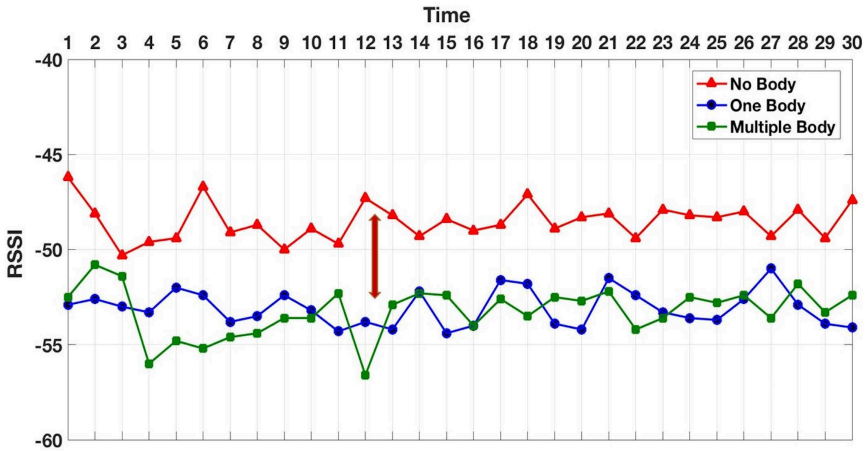

Figure 25: Indoor Average RSSI.

In summary, both indoor and outdoor results indicate that they can identify any human interference (i.e., NHI vs. non-NHI). However, the indoor case cannot discern the density level difference (i.e., SHI vs. MHI). Unlike the indoor case, the outdoor can distinguish the different density levels. It indicates that the RSSI power metric alone can identify a coarse level of the human interference, especially the indoor case.

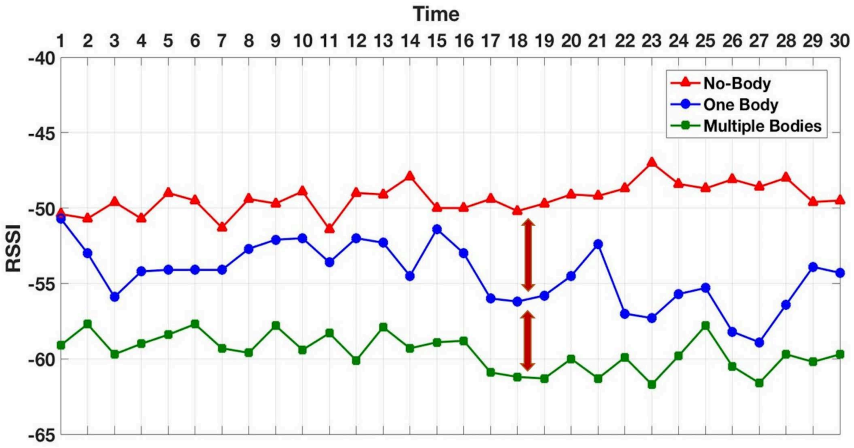

Figure 26: Outdoor Average RSSI.

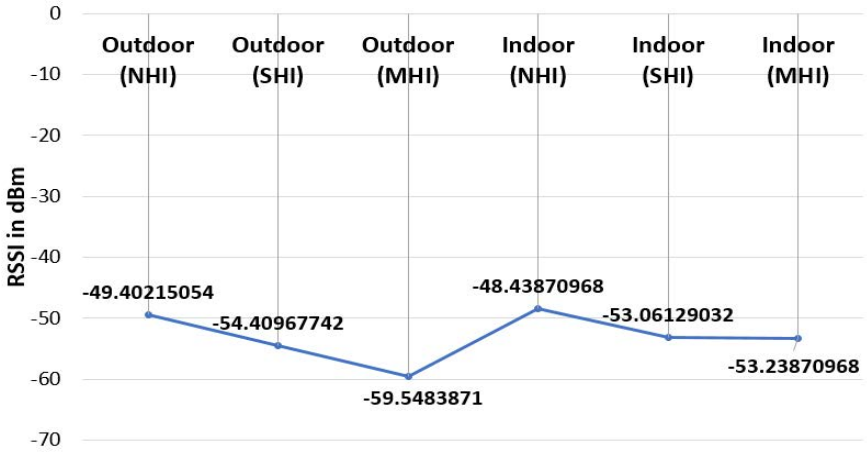

Figure 27: Average RSSI Comparison.

We test a variation of the received RSSI power. We pick the maximum, and minimum RSSI power values among the beacons received in a second (Vps), and use the difference as a variation value. The indoor RSSI variation test results in Figure 28 show that the RSSI variation in NHI is stable while the RSSI variation in other non-NHI (i.e., SHI and MHI) is randomly fluctuating. On the other hand, the outdoor RSSI variation

test results in Figure 29 show that the RSSI variation results in all workloads (NHI, SHI, and MHI) are unstably vibrating. As shown in Figure 30, the average indoor RSSI variation values are stable in all workload from 8.5 to 10 Vps. Although the average values are similar, the variation values in both SHI and MHI are fluctuating while NHI variation is stable. Meanwhile, the average outdoor RSSI variation values are much higher than the indoor variation average values and are also different between NHI and non-NHI. The non-NHI values are similar and higher than the NHI value. In summary, the indoor results indicate that they can identify any human interference (i.e., NHI vs. non-NHI). However, the outdoor case cannot discern any density. Hence, it indicates that the RSSI variation metric alone can identify a coarse level of indoor human interference. For example, it can be used to check if there is a person in the room or not, but it not possible to notice the crowd density.

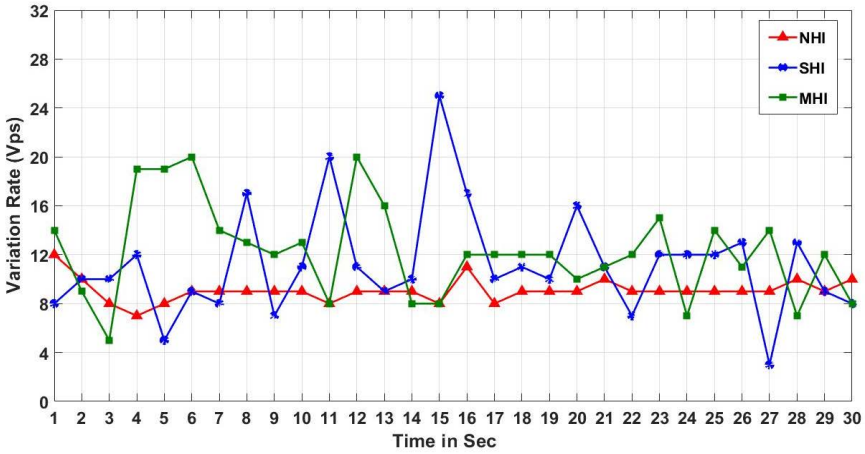

Figure 28: Indoor RSSI Variation.

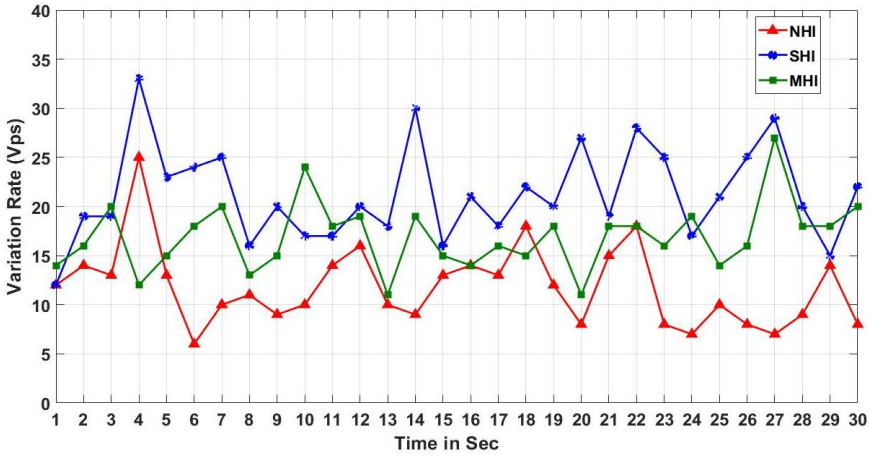

Figure 29: Outdoor RSSI Variation.

3.3.4 Discussion

The primary purpose of this study is to improve crowd safety management method through real-time Radio Frequency (RF) to predict and prevent potential disasters. Our approach focused on characterizing and analyze the crowd mobility in speed, direction, and density through BLE beacon count and RSSI power and variation. In summary, the findings from the experimental results include followings: BLE beacon count approach can be used to detect a location, direction and the speed of an object during the crowd by coordinating multiple scanners. The RSSI power average can be used to identify human interference in outdoor, while RSSI variation can check any human intervention in indoor, but it cannot evaluate the density. Therefore, by integrating those metrics, CROMO can identify the flow direction and speed, and the crowd density and object group location.

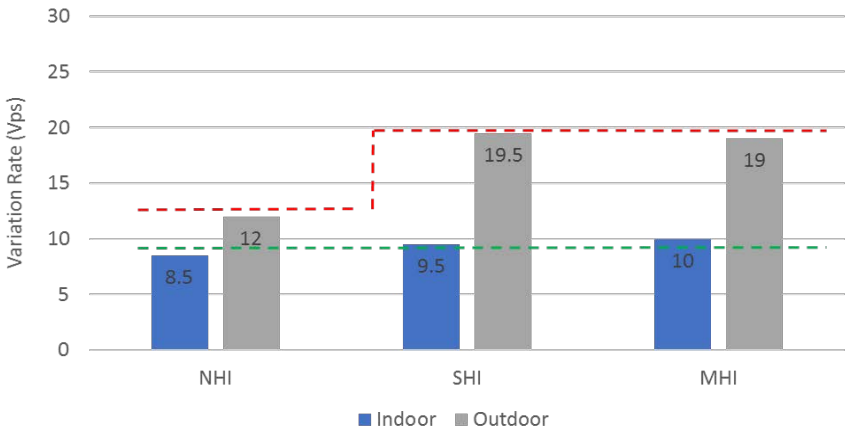

Figure 30: RSSI Variation Comparison (Indoor vs. Outdoor)

3.4 Conclusion

One of the critical services in smart cities is the safety management of urban communities. However, it is very challenging problems to predict a crash in real-time among the mobile crowd for preventing any potential disaster. In this work, we designed, implemented, and tested CROMO that enhances crowd mobility characterization through real-time BLE data analytics. We built a CROMO layer that fills the scalability and capability gaps of the smart video surveillance via BLE signal analytics. Among the many crowd mobility characteristics, we measure the beacon counts, the RSSI power, and its variation pattern by using a BLE bracelet and BLE scanners. We also apply them to group semantics to track the crowd status and predict any potential accidents and disasters. We have tested CROMO in both a practical crowd event and the controlled indoor and outdoor lab environments. The results show that by integrating the BLE data metrics, CROMO

can identify the crowd density, the object group location, and the flow direction and speed in real-time. It shows the feasibility of using the proposed CROMO layer enhancing the safety management system.

CHAPTER 4

ML-BASED VIDEO/IMAGE SURVEILLANCE

Overcrowding areas and exposure to large crowds often lead to human casualties and infectious diseases (i.e., COVID-19). Using video surveillance with intelligent techniques to analyze data is a growing trend in the field of computer vision for crowd analysis and management purposes. Various scenarios are considered in terms of analyzing a crowd, most of which are concerned with safety and health-related implications behind crowd formation. In this chapter, we propose two novel Machine Learning (ML) approaches in order to crowd management using intelligent techniques. We first propose a novel deep learning approach for effective dense crowd counting by characterizing scattered occlusions, called (CSONet). CSONet recognizes the implications of event-induced, scene-embedded, and multitudinous obstacles such as umbrellas and picket signs to achieve an accurate crowd analysis result.

Moreover, In addition to rapid vaccination, predicting possible trajectories of the COVID-19 pandemic is critical to health-care-related policy decisions and infrastructure planning. Growing evidence shows that face masks and social distancing can considerably reduce the spread of respiratory viruses like COVID-19. However, the current pandemic trajectory predictions take overly simplified policy input rather than actual observations of face masks and social distancing practices in a crowd. Thus, it is crucial to monitor and understand the extent of masking practices and assess the safety level in a scalable

manner. Thus we propose a novel face masking detection system for Modeling Safety Index in Crowd called (Mosaic), a Machine Learning (ML)-based approach for detecting masking in a crowd by building new dense mode crowd mask datasets. Mosaic detects, counts, and classifies the crowd's masking condition and calculates each community's spatiotemporal Safety Index (SI) values instead of detecting individual masking cases.

4.1 Part 1: Characterizing Scattered Occlusions for Effective Dense-Mode Crowd Counting

Crowd counting is becoming an increasingly important issue of computer vision, as it has many applications in the context of smart cities especially pertaining to public safety. The lack of proper crowd safety control and management often leads to human casualties and infectious disease (i.e., COVID-19) spreading at densely crowded political, entertaining, and religious events. Hence, automated crowd interpretation using AI techniques [38, 39] is becoming an increasingly critical task for many practical crowd safety applications [40, 41, 42]. Although many CNN-based methods have been proposed to improve the performance on complex crowd images to deal with variations in scale, perspective, and image resolution [43, 44, 45, 46, 47], they still have significant limitations in the face of occlusions that partially impede sight of individuals in a crowd scene. Crowd images are often scattered with occlusions that make it difficult to identify all human heads in the scene. As illustrated in Figure 31, the types of fixed environmental obstacles such as buildings, big trees, and walls are constrained to specific parts of a image, thus can be easily excluded from the crowd counting area. However, the interpretations of

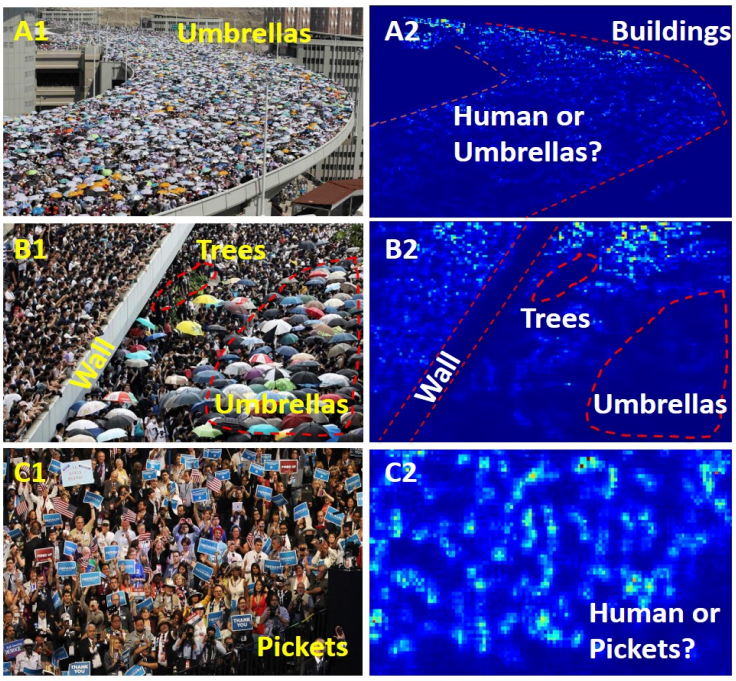

Figure 31: Crowd Map with Occlusion Objects (Overpass, Buildings, Walls, Fences, Trees, Umbrellas, and Pickets).

event-induced, scene-embedded, and multitudinous obstacles, namely *Scattered Occlusions (SO)*, such as umbrellas and picket signs are challenging, as they can obscure the sight of one or more individuals entirely or partially depending on crowd size and density as well as occlusion types [2]. Despite its commonness in many mass gathering scenes such as sport events, political rallies or protests, existing approaches fail to do accurate human counting in the presence of SO in crowd images.

In this work, we propose a novel deep learning approach for effective dense-mode crowd counting by characterizing scattered occlusions (CSONet). CSONet effectively

recognizes event-induced, scattered, and multitudinous occlusions and applies the effect to improve crowd counting accuracy and crowd density mapping quality. Specifically, CSONet tackles the dense-mode crowd scenarios such as people under umbrellas and behind pickets, which can hide people according to the event and recurring patterns in various ways. CSONet is an efficiently trained model using a simple convolutional structure comprised of three components. First, the Scattered Occlusion Datasets (SOD) component generates two new crowd counting datasets that contain diffused umbrella (cso-umbrellas dataset) and picket (cso-pickets dataset) occlusion objects in the crowd images. SOD also trains the model and outputs umbrella and picket heatmaps. Second, a network for Crowd Overfit Reduction (COR) is added on the well-trained VGG16-based CSRNet architecture [89] to reduce the Mean Absolute Error (MAE) and Mean Squared Error (MSE). We use the first ten layers of VGG16 to extract features from the crowd images. The extracted VGG16 features are grouped in a Spatial Pyramid Pooling (SPP) layer using average pooling in two different receptive fields (6×6 and 12×12) to soften overfitting. The Dilated Convolution Layers (DCL) outputs a predicted count and density map, which improves the crowd density prediction. Finally, a Scattered Occlusion Mapper (SOM) is implemented to combine the SO object heatmap with the human crowd heatmap to generate an accurate crowd density map and the crowd count. Using multiple datasets (cso-umbrellas dataset, cso-pickets dataset, and ShanghaiTech datasets A (dense-mode) and B (sparse-mode)), we demonstrate that CSONet's accuracy outperforms existing techniques such as SPN [92], ASNet [93] and CSRNet [89]. Our main objective is to achieve higher accuracy with the SO. CSONet reaches 100% better MAE and MSE

for cso-umbrellas(MAE-U and MSE-U) and 30% better MAE and MSE for cso-pickets (MAE-P and MSE-P) than CSRNet. CSONet also achieves 64% better MAE and 80% better MSE than SPN for umbrella dataset and 46% better MAE and MSE than ASNet for picket dataset. To the best of our knowledge, this is the first work that adaptively estimates the number of people occluded by objects scattered throughout a crowd scene to accurately quantify the total counts of people in a crowd image. The main contributions of this work include [141]:

- We have designed and developed a CSONet architecture, which is the first deep learning model for characterizing scattered occlusions of effective dense-mode crowd counting to the best of our knowledge.

- We have investigated the impact and challenges of SO in CNN crowd counting methods by collecting and annotating two new SO datasets, containing crowd images occluded with umbrellas (cso-umbrellas dataset) and picket signs (cso-pickets dataset).

- We have implemented COR by adding SPPL and DCL over modified VGG16 layers, which deploys a deeper CNN for capturing high-level features of extended receptive fields. COR was trained on the two new SO object datasets and the ShanghaiTech A and B datasets.

- We have built an algorithm that merges scattered object heatmaps and visible human heatmaps to generate a more accurate crowd density output.

4.1.1 Proposed Architecture

The proposed design aims to characterize scattered occlusions to improve the accuracy of crowd counting as well as the quality of crowd density mapping. In this section, we introduce the CSONet architecture that consists of a network for Scattered Occlusion Datasets (SOD), Crowd Overfit Reduction (COR), and Scattered Occlusion Mapper (SOM), as depicted in Figure 32. SOD creates two new scattered occlusion object datasets and trains on them. COR deploys a deeper CNN for capturing high-level features with larger receptive fields. SOM generates high-quality crowd density maps.

4.1.1.1 Scattered Occlusion Datasets (SOD)

In the Scattered Occlusion Datasets (SOD) component, we build two new datasets and perform CSONet training with these new datasets and two well-known public datasets.

- **Datasets and Experimental Settings:** Our goal is to investigate the impact and challenges of Scattered Occlusion (SO) objects in the CNN crowd counting methods. However, there has been no crowd image dataset available focusing on SO objects such as umbrellas and pickets. Hence, we have created new SO object datasets and trained our network CSONet on them. The generated dataset consists of the cso-umbrellas dataset and the cso-pickets dataset. They were collected from two resources. First, both umbrella and picket crowd images were mainly downloaded from Google images by running web search scripts with various keywords, including ”umbrellas” (”crowd with umbrellas” and ”crowd in the rain”) and ”pickets” (”demonstration” and ”protest”). Second, the cso-umbrellas dataset images

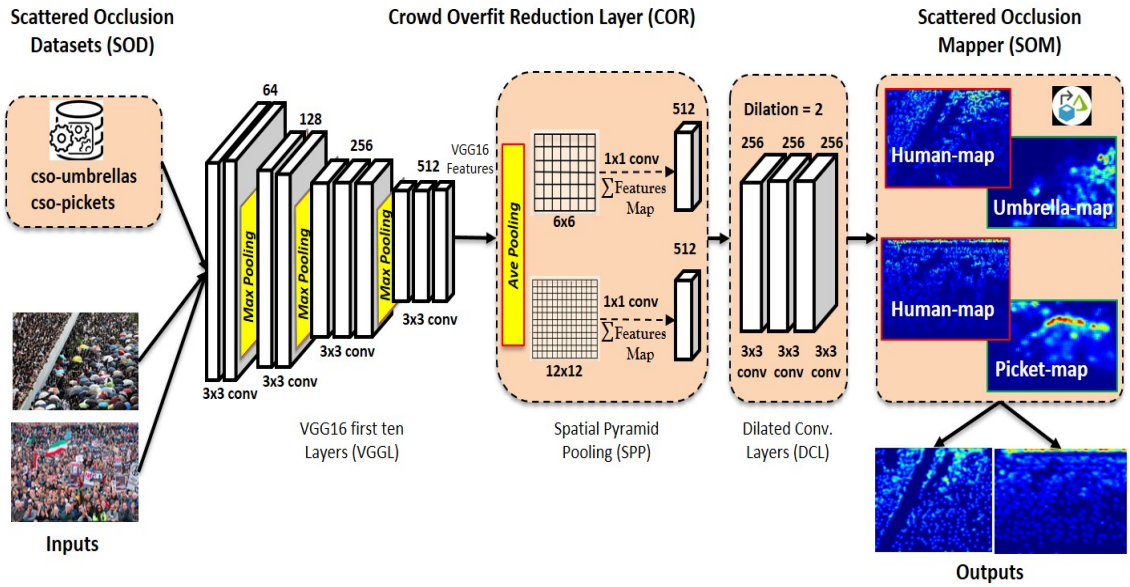

Figure 32: CSONet Architecture.

are partially converted from the Hajj event videos, an annual Islamic pilgrimage to Mecca, Saudi Arabia, during the summer, where the crowd holds umbrellas.

The **cso-umbrellas dataset** contains 250 crowd images and a total of 27,697 umbrella annotations. Among them, 170 images were used for training, and 80 images were used for testing. The **cso-pickets dataset** consists of 200 images and 9,681 picket annotations. 130 images were used for training, and 70 images were used for testing. To conduct comparisons with the existing state-of-the-art crowd counting method, we also train and test on the ShanghaiTech A and B datasets. The **ShanghaiTech dataset** is a large-scale crowd counting dataset containing 1198 images with 330,165 head annotations. It consists of two parts: ShanghaiTech A and ShanghaiTech B. Part A includes 482 dense-crowd images that have been collected randomly from the Internet. 300 images were used for training, and the remaining

182 images were used for testing. Part B has 716 sparse-crowd images, which were taken on busy streets in Shanghai. 400 images were used for training, and 316 were used for testing. Table 4 demonstrates a summary of the statistics of the datasets.

- **Ground-Truth Generation:** We have annotated all the images to generate the density map "ground-truth". We have applied the geometry-adaptive Gaussian kernels [86] as defined below to generate the density map for each crowd image. The labeled objects' locations in the original image are converted to the ground-truth density map $F(x)$ as follows:

$$F(x) = \sum_{t=1}^{N} \delta(x - x_i) * G_{\sigma i}(x), \text{ with } \sigma_i = \beta \overline{\mathrm{d_i}} \tag{4.1}$$

where N is the number of object annotations in the image, x_i is referring to each object in a given image, and $\overline{d}_i$ indicates the average distance of k-nearest neighbors. Also, the delta function $\delta(x - x_i)$ is convolved with a Gaussian kernel with the standard deviation parameter σ_i to generate density heatmaps.

- **Training details:** We have trained the CSONet structure in an end-to-end manner. Adam optimizer [32] is used as an optimization method to train CSONet with a learning rate of 1e-5 and a momentum of 0.9. Performing multiple experiments starting from 1e-4 to 1e-9, we found that 1e-5 is the ideal learning rate. In addition, we used other recommended training hyper-parameters, including a batch of size 32 and an epoch number of 100.

Table 4: Summary of statistics of the datasets.

Datasets	Images	Annotations	Avg. Count	Max. Count	Avg. Resolution
Shanghai A	482	241,677	501	3,139	589 x 868
Shanghai B	716	88,488	123	578	768 x 1024
cso-umbrellas	250	27,697	111	862	561 x 783
cso-pickets	200	9,681	48	386	728 x 969

4.1.1.2 Crowd Overfit Reduction (COR) Layer

In this subsection, we explain the network structure of the Crowd Overfit Reduction (COR) layer that consists of three components, including VGG16 Layers (VGGL) [22], Spatial Pyramid Pooling Layers (SPPL) [142], and Dilated Convolution Layers (DCL). We start with the VGG16 network, which was initially designed for large-scale natural image classification. VGG16 has thirteen convolutional layers and three fully connected layers. However, we have modified the VGG16 network, which learns ten convolutional layers with max-pooling, does two SPPs with average-pooling and applies three dense models in DCL.

- **Modified VGG16 Network Layer:** We apply the first ten convolutional layers and three max-pooling layers of VGG16 to extract the crowd features. The VGG16 is employed to ensure excellent learning performance in object classification and detection, which has been used by various practices such as CSRNet [89] and DADNet [143]. The input images commence with a fixed size by 224×224 pixel RGB image at the first convolutional layer. As illustrated in Figure 32, the images sequentially pass through a stack of 3×3 kernel convolutional layers with different filter depths

(64, 128, 256, and 512, respectively) and three max-pooling layers of 2×2 pixel windows in-between to create VGG16 features.

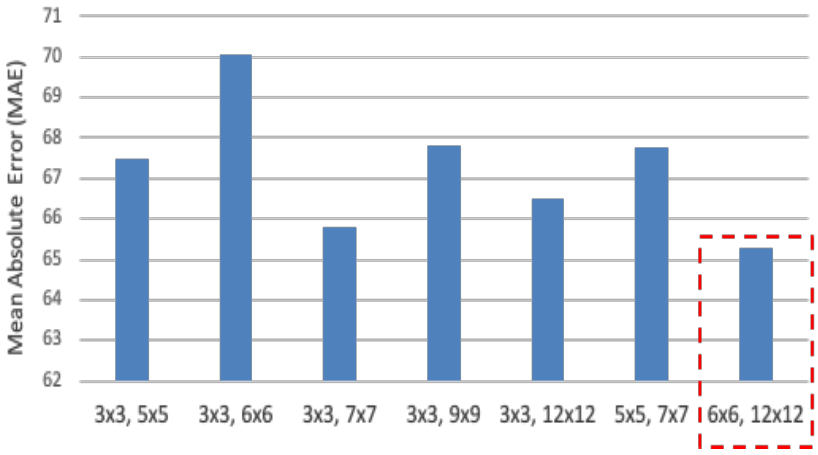

Figure 33: Various receptive fields performance.

- **Spatial Pyramid Pooling Layers (SPPL):** We have implemented Spatial Pyramid Pooling Layers (SPPL) to process the features extracted from the first ten VGG16 layers, which improves the semantic segmentation results [144] in the density map. We apply average pooling instead of max-pooling to reduce the overfitting and level the prediction results. In particular, the average pooling layer assigns the extracted VGG16 features into two different receptive fields (6×6 and 12×12), followed by a 1×1 convolutional layer. As presented in Figure 33, among the various sets of receptive field experimental results, the receptive fields of (6×6 and 12×12) achieve the lowest Mean Absolute Error (MAE) value.

- **Dilated Convolution Layers (DCL):** The Dilated Convolution Layers (DCL) is the last function of the COR structure. Generally, DCL is widely used in various

computer vision processes to promote crowd density predictions and improve semantic segmentation results. Moreover, DCL maintains the exponential expansion of the receptive field without reducing the resolution [145]. We implement DCL with three convolution layers using the same depth filter of 256 and a kernel size of 3×3. We set the dilation rate to two to gain better performance. However, we transfer the feature maps to these smooth layers to produce the CSONet outputs, predicted crowd count, and density heatmap. The motivation for utilizing the DCL in COR was to promote the dense prediction in congested images.

4.1.1.3 Scattered Occlusion Mapper (SOM)

Scattered Occlusion Mapper (SOM) is the last component of CSONet architecture. It generates a high-quality crowd density heatmap and an accurate crowd count by merging Scattered Occlusion (SO) object data with human crowd data. Most of the existing crowd counting methods require a visible head to detect and count the number of people, which cannot delimit individuals under umbrellas or behind pickets. A simple one-to-one mapping won't work as the SO object's impact on visual saliency for an image depends on the size, density, mobility type, flow direction, and velocity. As illustrated in Figure 34, an umbrella's effect is different in the sparse and dense crowd scenarios. An earlier study proposed an illustration for crowd counting per unit [2]. We propose a procedure for estimating the number of people under umbrellas or behind pickets in a particular crowd event. Our analysis shows that each umbrella covers zero to three people, and each picket occludes zero to two people corresponding to the Occlusion Object to Human Ratio

(OHR). Also, we assume that the number of SO objects cannot be more than the original human count. However, those effects converge into similar values in the high-density images. As shown in Figure 35, the average number of people under an SO object mainly depends on the OHR. Therefore, a formula is proposed to count the total number of people in an SO image:

$$T_{human} = D_{human} + (D_{so} * \alpha) \tag{4.2}$$

where T_{human} is the total predicted crowd count in an image, D_{human} is the detected human count. D_{so} indicates the number of predicted objects in an image (umbrellas D_u or pickets D_p). An α can be measured by using the ground truth values named MSOI (Measured SO Impact). Also, it is estimated as an SO Impact (SOI) value. According to Figure 35, an α value is chosen from the SOI value according to the OHR. For example, if OHR is 40 % (i.e., human count : SO object count = 100 : 40), α is 2.

4.1.2 Experiments

We test the proposed CSONet using multiple different datasets, including two new SO object datasets (cso-umbrellas and cso-pickets), along with two public crowd datasets, ShanghaiTech A [86] and ShanghaiTech B [86]. In this section, the evaluation metrics are introduced and then SO evaluations are conducted to analyze the efficacy of the proposed model. We evaluate and compare the performance of CSONet to various crowd counting methods including SPN, ASNet, and CSRNet. The CSONet prototype was implemented using the Pytorch framework [30]. All of the experiments were conducted on an NVIDIA

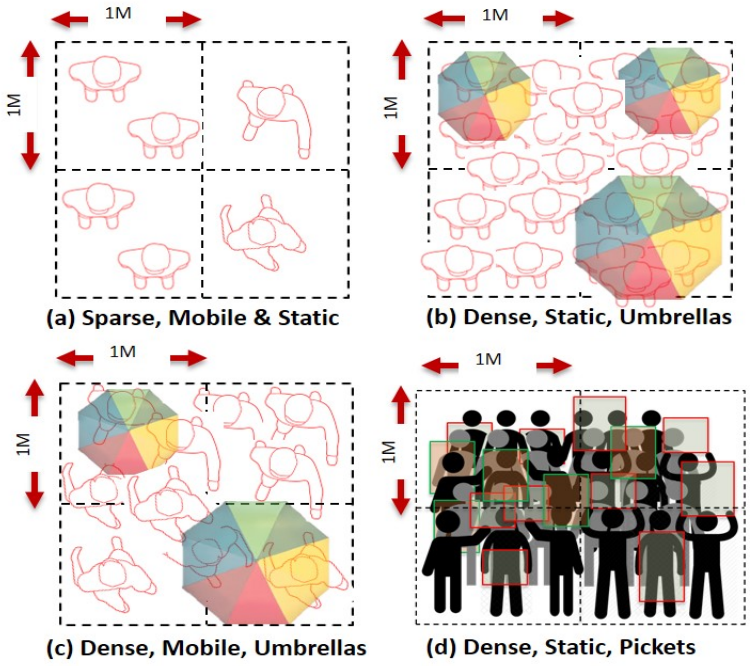

Figure 34: Crowd image annotations with different mode, type, and object.

GeForce GTX 1080 Ti.

4.1.2.1 Evaluation Metrics

Both Mean Absolute Error (MAE) and Mean Squared Error (MSE) are adopted in our performance testing. These metrics are broadly used in crowd counting to evaluate the accuracy of the measurement performance.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|Y_i - Y_i^G| \tag{4.3}$$

$$MSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y_i - Y_i^G)^2} \tag{4.4}$$

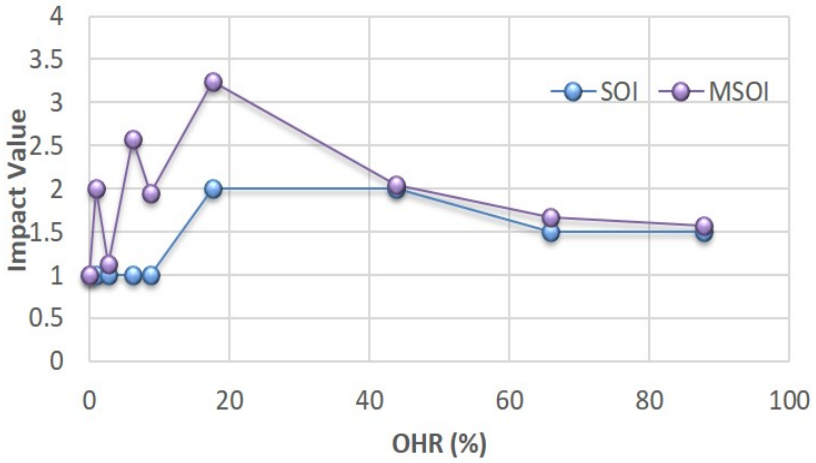

Figure 35: Scattered Occlusion (SO) impact values.

where N is the number of test images, Y_i is the predicted number, and the Y^G is the ground-truth counts of the test image i.

We also compute the Structural Similarity Index (SSIM) [146], which is a metric used to measure the similarity between two images. The SSIM value ranges from 0 to 1, equaling 1 if the two images are identical.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.5}$$

Following the preprocessing method given by [146], Eq. (4.5) measures the similarity between two images. Where x is the estimated density-map, and y is the ground-truth. C_1 and C_2 are small constants, used to avoid division by zero. In addition, μ_x and σ_x^2 are the local mean and variance estimations of x, and σ_{xy} is the local covariance estimation. μ_y and σ_y^2 are computed similarly.

For SO evaluation, we use ERT (error against real ground truth) in Eq. (4.6).

$$ERT = |V_i - OH_i^G| \tag{4.6}$$

where V_i refers to the detected human count, the OH_i^G presents the real ground-truth (RGT) of the test image i.

4.1.2.2 Scattered Occlusion Evaluations

The experiment is designed to evaluate SO object detection's performance and the accuracy of the estimated number of people occluded by the SO. We investigate how significantly umbrellas and pickets impact the accuracy of crowd counting and density estimation. For this purpose, we use an original crowd image with 114 people and continuously increase SO object annotations over the image from 0 to 87.7 % (Occlusion Object vs. Human Ratio (OHR)), as shown in Table **??** and Figure 36. CSRNet [89] is selected as a baseline to compare and evaluate the prediction accuracy of the proposed work. Table **??** shows the statistics of the experimental results with various scenarios and methods. The Real Ground Truth (RGT) value is the original number of people in the crowd image (i.e., 114 people) and the known number of SO objects (umbrellas (U) or Pickets (P)) placed on the image (i.e., from 0 to 100 umbrellas or pickets). We also use a Detected Ground Truth (DGT) of the number of SO objects (U/P) and the number of remaining visible humans heads (H), based on manual Matlab-based annotation analysis. DGT-U is the DGT after umbrella annotation, and DGT-P is the DGT after picket annotation. As the number of SO object annotation increases, the number of visible heads decreases due to occlusion. Also, the SO object count accuracy reduces due to many overlaps (i.e.,

only 72 umbrellas are detected after applying 100 umbrellas). We run both CSRNet and CSONet to find the number of humans and SO objects in a crowd image. As presented in Figure 36, after applying 75 SO object annotations (65.8% of OHR), there are almost no visible human heads. However, as CSONet applies the SO object impacts (SOI) for its final crowd counting according to Eq. (5.3), its prediction results are as good as RGT.

Table 5: Experimental results with SO objects.

OHR(%)	**RGT**	**DGT-U**		**DGT-P**		**CSRNet**		**CSONet**	
U/P	U/P	H	U	H	P	U	P	U	P
0	0	114	0	114	0	119	119	113	113
0.9	1	112	1	111	2	107	107	113	115
2.6	3	108	4	106	3	105	105	115	112
6.1	7	97	7	100	7	85	85	102	106
8.8	10	94	11	94	9	88	88	102	105
17.5	20	57	17	68	19	68	86	95	97
43.9	50	15	49	30	46	18	55	110	106
65.8	75	6	64	9	70	14	13	105	110
87.7	100	0	72	0	73	7	4	110	110

Figures 37 and 38 compare the crowd counting performance of CSRNet and CSONet in the aspect of ERT in Eq. (4.6) for umbrella and picket annotations, respectively. The ERT of CSONet is much lower than the ERT of CSRNet. The ERTs of CSRNet significantly increase when OHR increases. However, the ERT of CSONet does not increase for all OHR. Therefore, CSONet's crowd counting performance is much more stable and accurate than CSRNet, indicating that merging the human and SO density heatmaps is critical for better crowd count accuracy.

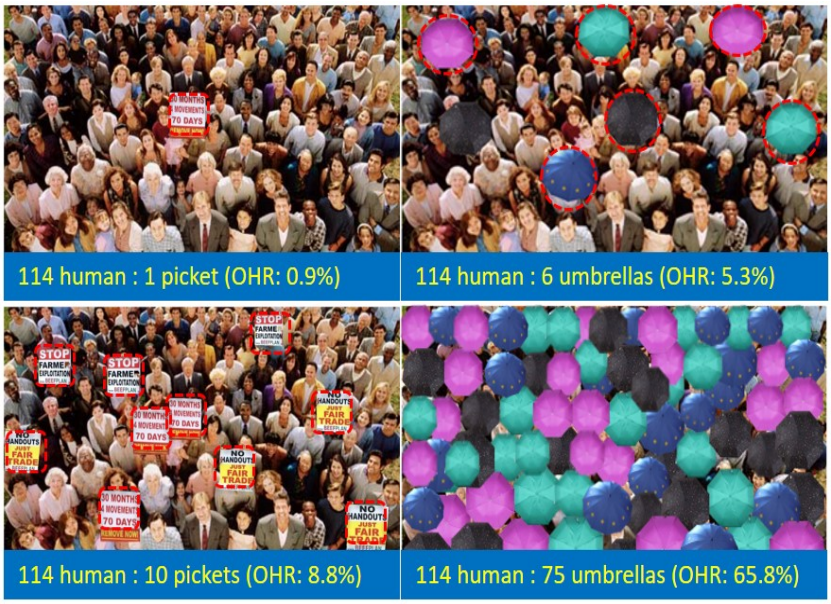

Figure 36: Crowd images with SO object annotations.

4.1.2.3 Performance Comparison

The performance in MAE and MSE metrics with ShanghaiTech datasets (i.e., MAE-A means MAE with ShanghaiTech A) of the existing state-of-the-art crowd counting solutions, including CP-CNN [45], CSRNet [89], PCC Net [147], SPN [92], and ASNet [93] are compared in Table 6. It shows that the most recent ASNet achieves the least MAE-A and MSE-A. SPN is as good as ASNet, which is 27% better than other earlier approaches such as CSRNet. ASNet did not evaluate ShanghaiTech B dataset (sparse-mode), as they are interested in counting densely populated crowd with ShanghaiTech A (dense-mode) dataset.

Table 7 presents the crowd counting accuracy results with the new SO object

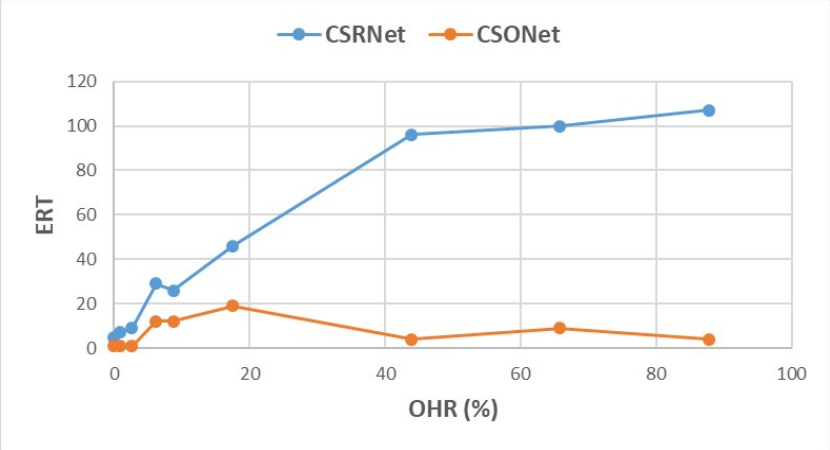

Figure 37: Error against RGT umbrella annotations.

Table 6: Performance comparisons of different methods on ShanghaiTech A (dense-mode) and B (sparse-mode).

Method	**MAE-A**	**MSE-A**	MAE-B	MSE-B
CP-CNN [45]	**73.6**	**106.4**	20.1	30.1
CSRNet [89]	**68.2**	**115.0**	10.6	16.0
PCC Net [147]	**73.5**	**124.0**	11.0	19.0
SPN [92]	**61.7**	**99.5**	9.4	14.4
ASNet [93]	**57.78**	**90.13**	-	-

datasets. We choose 80 umbrella and 70 picket images from cso-umbrellas and cso-pickets datasets, respectively, and tested them with SPN, ASNet, CSRNet, and CSONet to obtain the MAE, MSE, and SSIM values (i.e., MAE-U means MAE for the umbrella images). According to the density of cso-umbrellas and cso-pickets datasets in Table 4, the MAE and MSE with the ShanghaiTech dataset in Table 6 align with the results in Table 7. For example, MAE-P and MSE-P maintain lower values due to the cso-picket images are sparse. Also, MAE-U and MSE-U of CSRNet are slightly higher than SPN and ASNet. CSONet's performance in terms of accuracy is significantly better than the

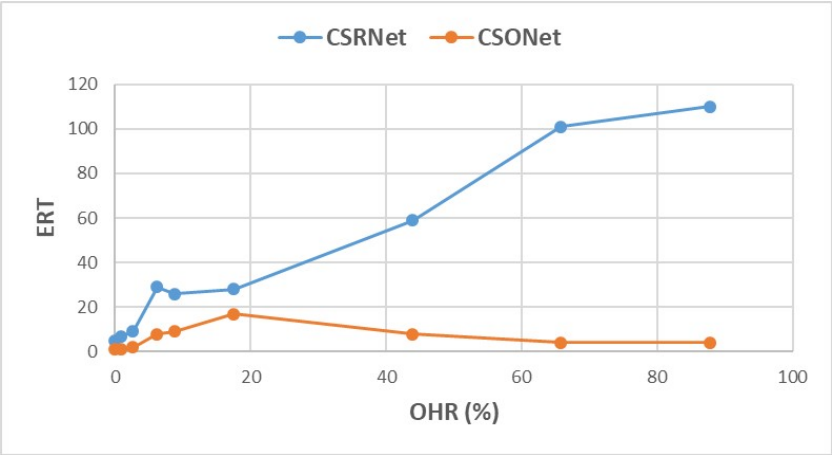

Figure 38: Error against RGT picket annotations.

Table 7: Performance comparisons of CSRNet, SPN, ASNet, and CSONet with cso-umbrellas and cso-pickets datasets.

Method	MAE-U	MAE-P	MSE-U	MSE-P	SSIM-U	SSIM-P	SSIM-A
CSRNet [89]	73.6	19.1	135.9	35.5	0.83	0.92	0.76
SPN [92]	59.9	24.2	120.1	42.3	0.85	0.90	-
ASNet [93]	71.5	21.3	133.7	40.5	0.81	0.88	-
CSONet	**36.5**	**14.6**	**66.5**	**28.4**	**0.87**	**0.94**	**0.91**

other methods. For example, CSONet achieves 100% better MAE and MSE for cso-umbrellas(MAE-U and MSE-U) and 30% better MAE and MSE for cso-pickets (MAE-P and MSE-P) than CSRNet. CSONet also achieves 64% better MAE and 80% better MSE than SPN for umbrella dataset and 46% better MAE and MSE than ASNet for picket dataset. The SSIM measures the similarity between the ground-truth and the estimated density-map images. Although SOs already impact the DGT images, the CSONet still creates a higher structural similarity than the other methods. Figure 39 presents overall performance results of crowd density heatmaps and crowd counts (human, umbrella, and

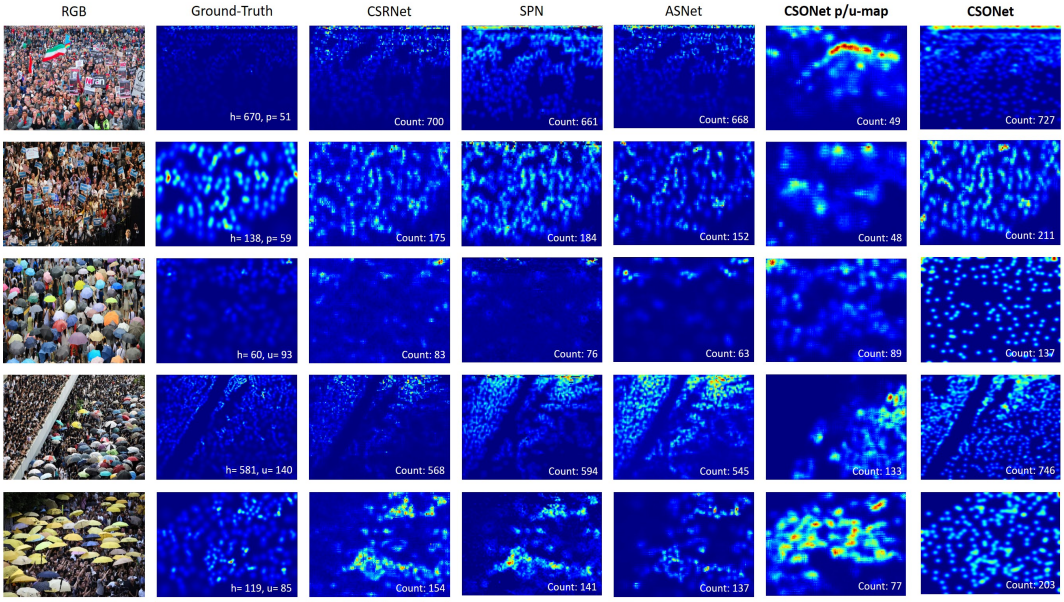

Figure 39: Five crowd image samples are randomly selected from the SO object datasets and evaluated with CSRNet, SPN, ASNet, and CSONet. We display the density maps and counts of each sample. For a given RGB image, from the left column, the detected ground-truth (DGT) density map shows the heads (H) and umbrellas (U)/ pickets(P) counts. The CSRNet, SPN, and ASNet prediction results (count and density map) are in columns 3, 4, and 5, respectively. The last two columns present CSONet results. The h-map is the human map and count, and p/u-maps are the detected SO objects (pickets/umbrellas). Finally, the CSONet map and count demonstrate the estimation of human count and density map, which applies the SOI (i.e., under umbrellas or behind pickets) in a particular crowd event.

picket) with five SO image samples. According to the heatmaps of DGT, CSRNet, SPN, and ASNet, the area covered by SOs are shown by low density. However, the CSONet adjusts those areas by identifying p/u heatmaps and overlaying them to human heatmaps, which results in more accurate crowd counting. As showen in Figure 40, the similarity (between the ground-truth and the estimated density-map images) of CSONet creates a higher structural similarity than the other methods with accuracy of 95.44%

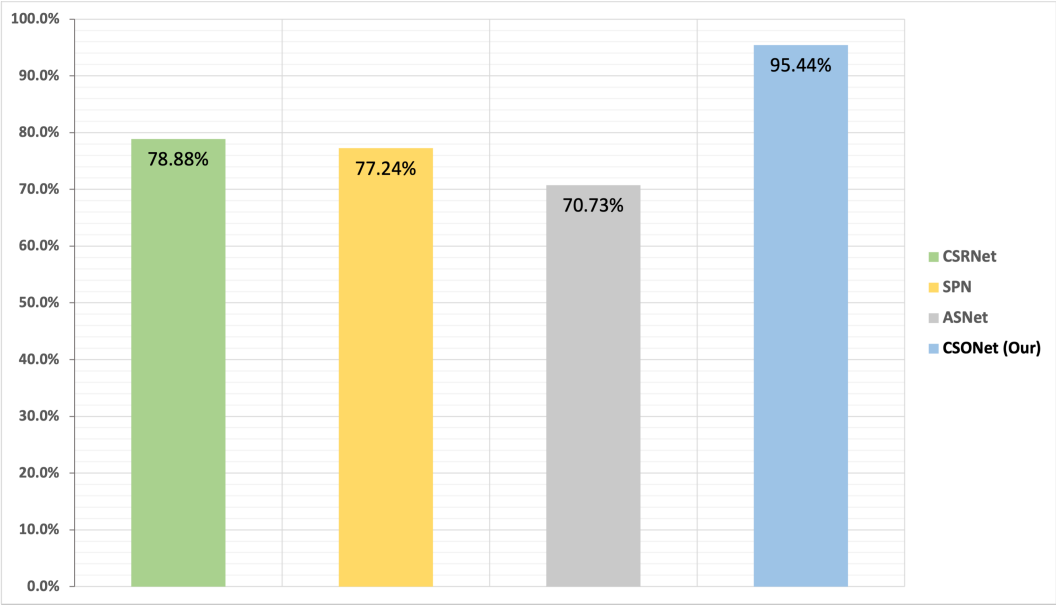

Figure 40: Overall Performance Accuracy.

4.2 Part 2: Modeling Safety Index in Crowd by Detecting Face Masks against COVID-19 and Beyond

In battling against the global pandemic caused by COVID-19, many countries and local governments continuously predict the pandemic trajectories for health-care-related policy decisions such as school opening and limiting businesses and infrastructure planning such as preparing and planning healthcare facilities and personnel. In addition, growing evidence shows that requiring masking in public areas in parallel with aggressive testings and rapid vaccinations are critical for achieving herd immunity [48, 49]. As illustrated in Figure 41, many scientists and the health organization guidelines from WHO [50] and CDC [51] confirm that masking in public places is one of the most effective health measures along with social distancing to break the coronavirus transmission chain [148].

For example, the chance of coronavirus transmission reduces from 90% up to 1.5% by wearing masks. Despite such safety recommendations, in reality, face mask-wearing and social distancing are practiced at a varied level depending on places and times.

Monitoring and assessing the conformity of mask-wearing and social distancing in public places would provide more accurate input to pandemic trajectory predictions and understand the safety level of communities. However, existing approaches do not directly address the issue and fall short of effectively and scalable tasks. There are several Machine Learning (ML)-based object detection and classification approaches aimed for COVID-19 projections [51], diagnosis, social distancing [52], and contact tracing [53] applications. However, as the technologies to ensure the masking in public places aim at surveillance and tracking individuals [54, 55], they are rendered less useful in *public areas*, such as *crowds* on the street, airports, and schools, due to the scalability limitation and privacy concerns.

Table 8: Comparison with existing datasets.

Dataset	Mask Type	# Images	# Faces	Ave # of face in an image	less than 5 people in an image
Kaggle FMD [104]	Real-world	853	4072	4.77	527 **(61.7%)**
MOXA3k [105]	Real-world	3000	12176	4.05	2028 **(67.6%)**
Mosaic (Proposed)	Real-world	530	5309	10	53 **(10%)**

For example, Kaggle [104] has a Face Mask Detection (FDM) dataset with 853 images, and 66 (as of 6/1/2021) published mask detection codes using YOLO versions and Faster RCNN models. Also, MOXA3k [105] with 3000 masks dataset has been published for monitoring of people wearing medical masks. However, as described in Table 8, around 60% to 70% of the datasets are less than five people (4 to 5 people in each image on average), which do not represent realistic crowd scenarios.

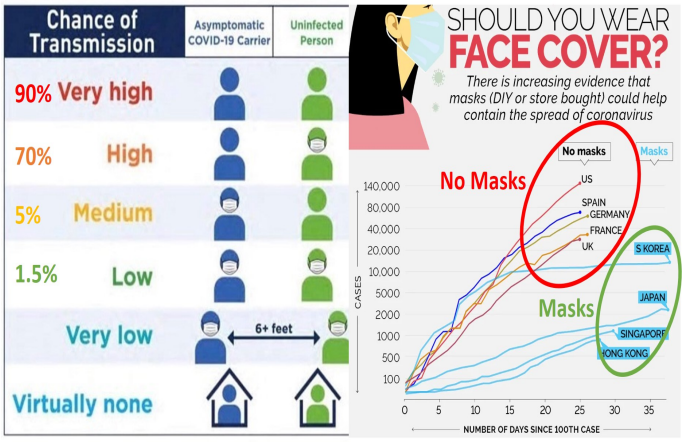

Figure 41: Effectiveness of masking against COVID-19

In this work, we propose an ML-based face masking detection system for **Mo**deling **Sa**fety **I**ndex in **C**rowd (Mosaic). Mosaic builds **a new dense mode crowd masking dataset** to detect, count, and classify the crowd's masking condition in addition to monitoring social distancing. It calculates spatio-temporal Safety Index (SI) values for each community instead of detecting individual mask-wearing conditions to ensure an individual's privacy. Mosaic calculates the weighted degree of each person's masking to obtain a more sophisticated SI model. The SI data can be shared or published to calculate the area-based SI maps (as opt-in data) for assisting effective policy decisions and relief plans against COVID-19. The evaluation results show that Mosaic detects various conditions and types of masking states and calculates SI values of a crowd effectively. Our goal is to investigate the impact and challenges of masking and social distancing the crowd to develop each community SI. To the best of our knowledge, this is the first work that effectively and scalably gauges the safety level from dense mode crowd image analyses for mask-wearing practices and social distancing in a crowd.

The main contributions of this work include [149]:

- A new crowd face-mask dataset (dense mode) is collected, annotated, and augmented to investigate the impact and challenges of mask-wearing. As shown in Table 41, Mosaic dataset maintains only 10% of images less than five people in an image (more than 10 people in each image on average);
- An advanced ML-method is proposed to detect the various type of mask-wearing features and social-distancing values in a dense mode corwd; and
- A new SI calculation approach is introduced using weighted neighbor relationship to ensure privacy protection.

4.2.1 Mosaic Architecture

To tackle the open challenges mentioned above, in this paper, we propose an ML-based **Mosaic** system, which calculates SI values by detecting and classifying the crowd's masking conditions. To serve this purpose, Mosaic consists of three modules, namely Data Training (DT), Feature Extraction (FE), and Safety Index (SI) Modeling.

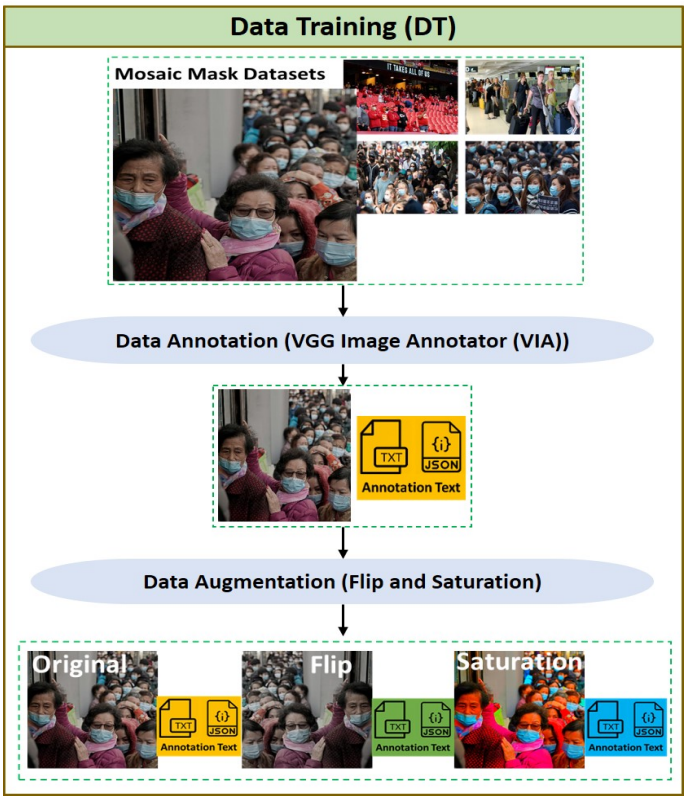

Figure 42: Data Training (DT) in Mosaic System

4.2.1.1 Data Training (DT)

The DT module is in charge of annotating and augmenting crowd masking images to create datasets, as shown in Figure 42. Besides, DT module also trains the collected datasets to enable various ML procedures.

- **Mosaic Mask Datasets:** We create and train a new dense mode masking dataset for the crowded environment. Mask images were collected from Google images by running web search scripts with various keywords, including ”mask” (’crowd with mask’ and ’COVID-19 face mask’). The dataset contains 530 images with either a

Figure 43: Existing Face Mask Detection (Kaggle FMD and Moxa3k) Vs. Dense Mode Crowd (Mosaic) Datasets

mask or no mask in a crowd. The dataset has various degrees of high-density crowd images. To the best of our knowledge, the 'Mosaic' dataset is the first dense mode face mask images in a crowded environment.

As shown in Figure 43, the existing mask datasets such as Kaggle FMD [104] and Moxa3k [105] usually contain images with a small number of faces for monitoring mask-wearing conditions. As illustrated in Table 8, Kaggle FMD has 4.77 faces in a picture on average, and Moxa3k has 4.05 faces in an image on average, but Mosaic has 10 faces in an image on average. Furthermore, most of the existing dataset images have less than 5 people in an image (61.7 % in Kaggle FMD and 67.6 % in Moxa3k), which is not suitable for using mask-wearing detection in a crowded environment. However, in Mosaic dataset, only 10% of images is less than 5 people in an image.

We divided the Mosaic datasets into three categories, training set (70%), validation

set (20%), and test set (10%).

- **Data Annotation:** The VGG Image Annotator (VIA) software [21] is used to manually draw bounding boxes on the images and assign each labeled object to its class. Our dataset has two classes: *Mask* that includes all people wearing a mask, and *No-Mask* that refers to all people without any mask or incorrect masking. We also generate the ground-truth density map. All fully or partially front-facing human faces are labeled *Mask* and *No-Mask*. Every labeled object and the information are exported to a .json file for training. We converted the .json file information to the .txt format to utilize it in YOLO v3 [150].
- **Data Augmentation:** Image augmentation techniques, including flip and saturation, are applied to increase the training set (70% of 530 images) to 1116 training images (372×3). It improves model performance by reducing the chance of model overfitting. Eventually, the dataset size becomes 1274 by adding 158 images (validation set (20%) and test set (10%)).
- **Training Configuration:** Adam optimizer [32] is used as an optimization method for the training of Mosaic by configuring a learning rate (1e-4), and a momentum (0.9). Performing multiple experiments starting from 1e-3 to 1e-9, we found that 1e-4 is the most optimal learning rate. Other recommended training hyper-parameters are adopted, including a batch of 16 and an epoch number of 100. The Mosaic structure is implemented using the Pytorch framework [30]. All of the experiments were conducted on an NVIDIA GeForce GTX 1080 Ti.

4.2.1.2 Feature Extraction (FE)

As illustrated in Figure 44, the FE module uses a YOLO v3 [150] to extract features. The Super-Resolution (SR) technique reconstructs a High-Resolution (HR) image from the Lower-Resolution (LR) images before applying the YOLO. After the feature extraction, FE classifies features as *Mask* and *No-Mask*.

- **Image SR:** An SRCNN [151] is used for single image SR to recover HR images from LR images and to enhance images' perceptual quality. SRCNN learns an end-to-end mapping between the LR and HR images. The mapping is represented as a CNN that takes the LR image as the input and outputs the HR one. We investigate the challenges of facemask-wearing conditions in the proposed crowd face-mask dataset, low-quality images and congested images are prevalent in the dataset. We integrated the Super Resolution (SR) technique to identify the facemask-wearing conditions and improve the classification accuracy. Practically, we have added the SRCNN network before image classification to reconstruct HR images and restore image details. SRCNN has a simple architecture that consists of three convolutional layers, including the patch extraction and representation layer, the non-linear mapping layer, and the reconstruction layer. The LR images are upscale using bicubic interpolation before inputting to SRCNN. The bicubic interpolation performs cubic interpolation on each of the two axes. It takes 4×4 pixels into account and produces smoother results. It is also similar to other interpolation-based upsampling methods that improve the input image resolution based on its image signals without bringing any more information.

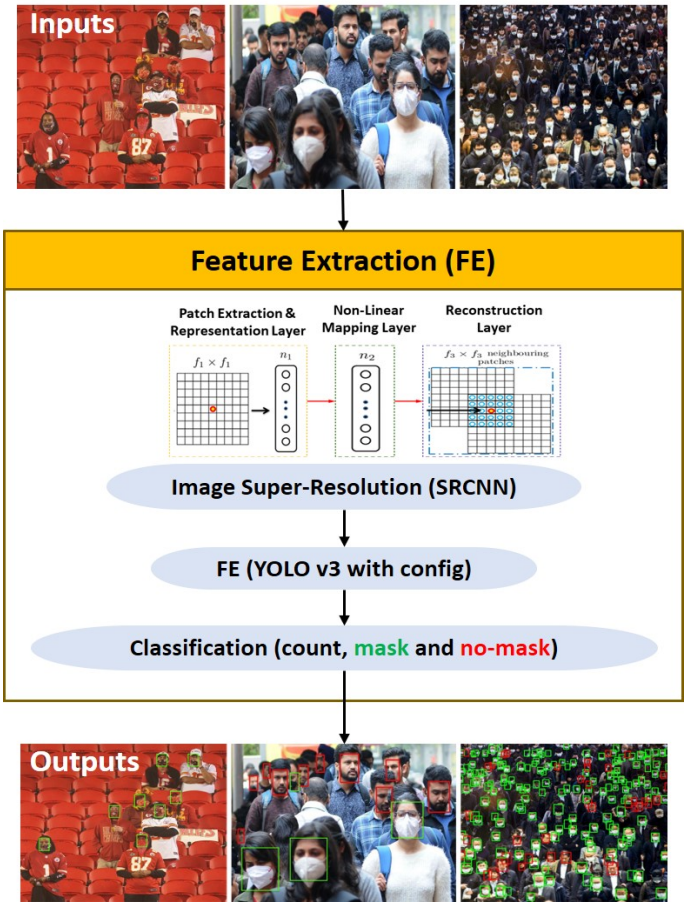

Figure 44: Feature Extraction (FE) in Mosaic System

The network settings are identical to settings that are explained in [151], such as $f_1 = 9, f_2 = 1, f_3 = 5, n_1 = 64$, and $n_2 = 32$. We also define the optimizer as Adam and the learning rate to be (0.0003). The patch extraction and representation layer extract patches from the interpolated image Y and represent each patch as a high-dimensional vector. The non-linear mapping layer maps each feature map into the HR patch representation. The reconstruction layer aggregates the HR patch representation to generate the HR image $F(Y)$.

- **Feature Extractor (YOLO):** YOLO v3 extracts the target objects' features of masking from an input image. The architecture is composed of 53 convolutional layers (i.e., Darknet-53). It continuously passes the extracted features to a masking check function, which detects and evaluates the masking status in various aspects, including mask, no-mask (incorrect masking). We have changed the hyperparameters of Yolo, such as batch size, learning rate, and epochs number.
- **Classification:** We have added a classification layer at the end of the feature extractor (YOLOv3). It classifies the masking features like a red box for *No-Mask* and a green box for *Mask*. It also counts the total number of people with *Mask* and *No-Mask*. Also, the classification layer can detect incorrect masking and various styles and colors. This information has been used to generate the Safety Index (SI) value for each community.

4.2.1.3 Safety Index (SI) Modeling

The SI modeling module calculates the safety index of each image by applying a weighted degree algorithm as presented in Algorithm 1.

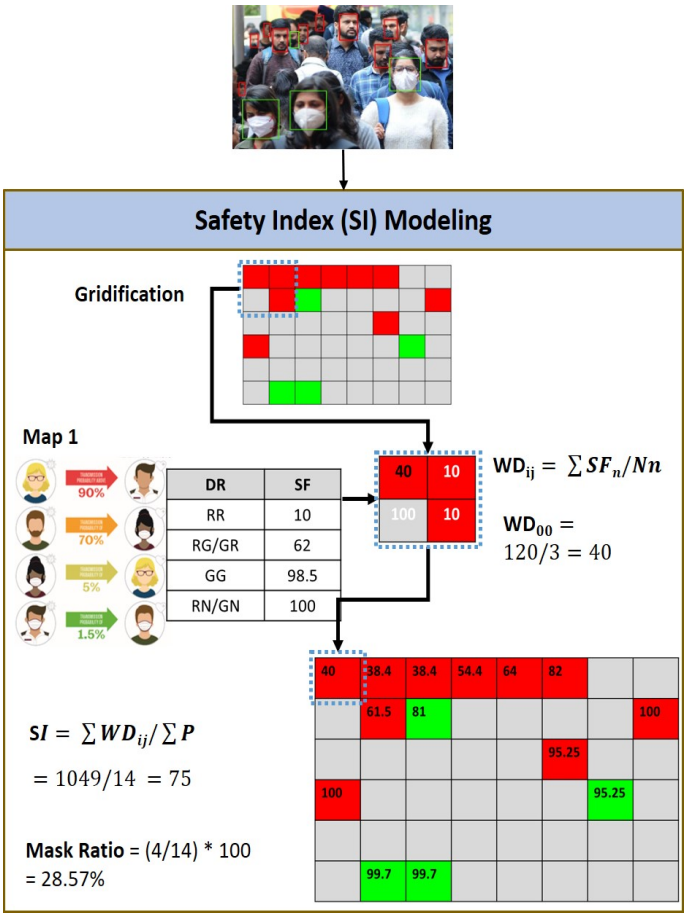

Figure 45: Safety Index (SI) Modeling in Mosaic System

As illustrated in Figure 45, a classified output image is converted into a grid (Gridification) by applying perspective effects. We then fill a red color for *No-Mask*, a green color for *Mask*, and a gray color for space (no people). For each grid, we apply the following SI equations (4.7 and 4.8).

$$SI = \frac{\sum WD_{ij}}{\sum P} \tag{4.7}$$

$$WD_{ij} = \frac{\sum SF_n}{N_n} \tag{4.8}$$

WD_{ij} presents a Weighted Degree (WD) value of a person on a grid i and j, which is the sum of neighbors' Safety Factor (SF) value, $\sum SF_n$, divided by the number of neighbors, N_n. For example, in Figure 45, for WD_{11}, the sum of neighbors' Safety Factor (SF) value is 120 (i.e., 100 + 10 + 10), and the number of neighbors is 3. Hence, WD_{11} value is 40.

SI value of an image is calculated by dividing the total number of detected people in the image, $\sum P$, from the total WD values for the entire grids, $\sum WD_{ij}$. For example, in Figure 45, the total WD values for the entire grids is 1049, and the total number of detected people in the image is 14. Hence, the SI value of an image is 75.

Map 1 represents the SF values according to degree relation (DR), which represents R as *No-Mask* (Red), G as *Mask* (Green), and N as no people (Gray). SF values are drawn from the infection values (%) in Figure 41. For example, as RR (*No-Masks* without distancing) has a very high transmission chance (90%), the SF value is 10. Also, for RG or GR (*Mask* to *No-Mask*), as it is unclear which side asymptomatic COVID-19 carrier is, we average out the transmission chance (38%). The SF value becomes 62.

The mask ratio (%) is the masked people divided by the total number of detected people in the image. In Figure 45, it is 28.57%.

Algorithm 1 SI Modeling Algorithm

Result: Safety Index value

Grids = [(G_0, $class_0$), (G_1, $class_1$),....(G_i, $class_i$)] SI.list = [ ] Green.count = 0 Red.count = 0

for *G in Grids* **do**
    G.SF.list = [ ] N = Neighbors (Grids, G) **for** *n in N* **do**
        if *G.class is (Green or Red)* **then**
            if *n.class is Gray* **then**
                G.SF.list.Append(100)
            end
        end
        else if *G.class is Green* **then**
            Green.count = Green.count+1 **if** *n.class is Red* **then**
                G.SF.list.Append(62)
            end
            if *n.class is Green* **then**
                G.SF.list.Append(98.5)
            end
        end
        else if *G.class is Red* **then**
            Red.count = Red.count+1 **if** *n.class is Red* **then**
                G.SF.list.Append(10)
            end
            if *n.class is Green* **then**
                G.SF.list.Append(62)
            end
        end
        else if *G.class is Gray* **then**
            continue
        end
    end
    WD_ij = Avg(G.SF.list) SI.list.Append (WD_ij)
end

SI = Avg(SI.list) Mask Ratio = (Green.count / $\sum$ (Green.count, Red.count)) $\times$ 100 **return** SI, Mask Ratio

4.2.2 Evaluations

Experimental study has been conducted using the trained Mosaic prototype and the testing dataset. Figure 46 presents the detection results of incorrect masking (e.g., the neckbeard, the sniffer, the stache, and the nose plug) and wearing various mask styles and colors. For example, the red box detects both *No-Mask* and *incorrect mask*.

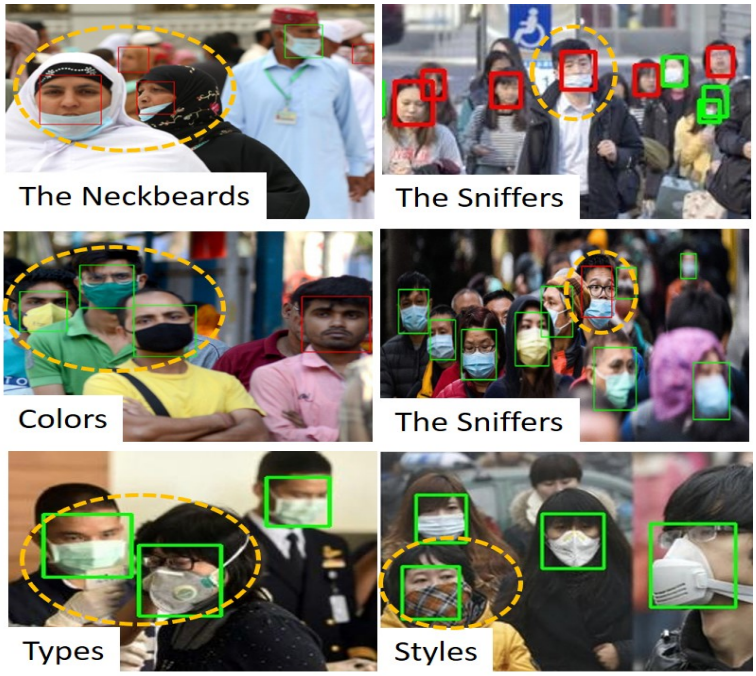

Figure 46: Detecting incorrect masking and various styles and colors.

$$Recall = \frac{TP}{TP + FN} \tag{4.9}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.10}$$

The recall rate and precision metrics defined by equations (5.2 and 4.10) with true-positive (TP), false-negative (FN), and false-positive (FP) are used to evaluate the

Table 9: Classification performance.

Mask Recall	Mask Precision	No Mask Recall	No Mask Precision	mAP
90%	88%	79%	76%	82%

classification detection performance. A mean Average Precision (mAP) in object classification tasks is calculated. mAP is in a range from 0% to 100%, and the higher value means better accuracy. As shown in Table 9, mask recall of 90%, mask precision of 88%, no mask recall of 79%, no mask precision of 76%, and mAP (average) of 82% are high enough for detection tasks in crowd. Both recall and precision results in *No Mask* are relatively low because most of the labeled objects in the dataset are with *Mask*. We observed that adding the SR network outperforms an excellent result and significantly improves the classification accuracy.

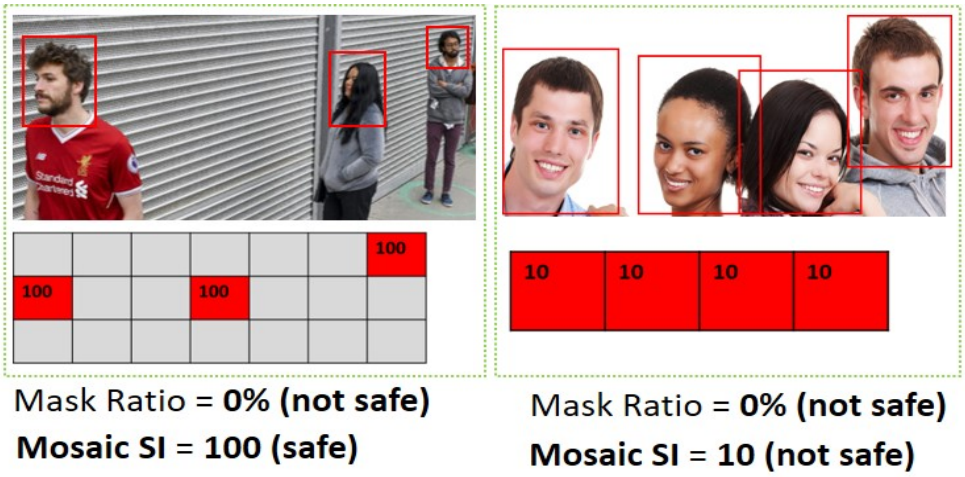

Figure 47: Mosaic SI Range from 100 (safest) to 10 (not safe).

Figure 47 presents the Mosaic SI value range from 100 (the safest) to 10 (the

least secure) with a couple of extreme examples. Both images' mask ratios are 0% (no mask-wearing), which indicates that the scene has the highest transmission probability. However, the people in the left image keep social distancing very well, but the people in the right image do not. Hence, the left image indeed has less transmission chance. The Mosaic SI value differentiates the transmission chance using both mask-wearing relationships and social distancing. The left image's Mosaic SI value is 100 (very safe), while the right image's Mosaic SI value is 10 (very unsafe).

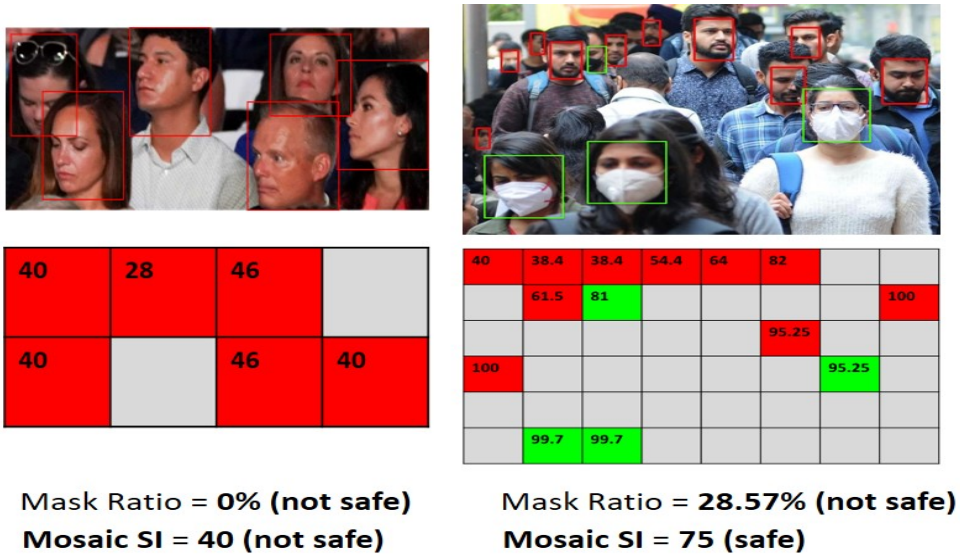

Figure 48: Mosaic SI vs. Mask Ratio.

Figure 48 shows a couple of crowd image examples. The left image shows that its mask ratio is 0% (no mask-wearing), which indicates that the scene has the highest transmission probability. With some social distancing, its Mosaic SI value is 40, which is not safe. The right image shows that its mask ratio is 28.57%, which indicates that the scene is not safe. However, its Mosaic SI value is 75 with social distancing, which is safe.

4.3 Conclusion

In this chapter, we introduced two novel intelligent approaches in order to crowd management using intelligent techniques. We proposed architecture for scattered occlusion characterization called CSONet for efficient crowd counting and high-quality density heatmap generation. CSONet recognizes event-induced, scattered, and multitudinous occlusions and applies the effect to a human crowd map to generate an accurate crowd count and high-quality density map. We also proposed a novel face masking detection system for Modeling Safety Index in Crowd (Mosaic), a Machine Learning (ML)-based approach for detecting masking in a crowd by building new dense mode crowd masking datasets. Mosaic detects, counts, and classifies the crowd's masking condition and calculates each community's spatiotemporal Safety Index (SI) values instead of monitoring individual masking cases. Mosaic can provide an accurate assessment of SI of a community based on real measurement of masking and social distancing practices. Through extensive evaluations, we demonstrated that the accuracy of these methods outperforms over the state-of-art existing approaches.

CHAPTER 5

SEMANTIC KNOWLEDGE INFORMATION-BASED TRACING APPLICATION

Due to its long incubation period, aggressive asymptomatic transmission, and new mutations of the virus, COVID-19 is causing multiple pandemic waves worldwide. Despite recent vaccination, social distancing, and social restriction efforts, false negatives and dormant positives can make pandemics challenging to restrain. In addition to rapid vaccination, effective contact tracing, mask-wearing, and social distancing are critical for outbreak containment and for achieving herd immunity. However, the existing technology solutions, such as contact tracing apps and social-distance sensing, have been met with suspicion due to privacy and accuracy concerns and have not been widely adopted. Without achieving a critical mass of individual users, these personal technologies have been rendered useless.

On the other hand, large-scale policy efforts have been complicated, requiring the coordination of federal, state, and local governments and regulation enforcement logistics. However, local communities balance these approaches and are an unrealized, powerful resource to prevent future outbreaks.

This chapter proposes two novel infection management systems: Crowd-based Alert and Tracing Services (CATS) to build a safe community cluster against COVID -19. CATS applies social distancing and masking principles to small, focused communities to provide higher privacy protection, efficient penetration of technology, and greater

accuracy. We have designed an intelligent tag for managing social distancing. We also implemented a Machine Learning (ML)-based face mask tracking system to build non-binary Safety Impact Values (SIV). Second, we propose Crowd Safety Sensing named (CroSS) for building a sustainable safe community cluster against COVID-19 and beyond using affordable Internet of Things (IoT) technologies. CroSS monitors social distancing policies to small, focused communities for accommodating efficient technology penetration, greater accuracy, effective practices, and privacy policy assistance. We implemented a social distancing method and integrated it into an edge-based IoT system.

5.1 Part 1: Crowd-based Alert and Tracing Services for building a Safe Community Cluster against COVID-19

Fighting against the global pandemic caused by COVID-19, many countries make a mask-wearing and social distancing in public areas compulsory in parallel with aggressive testings and vaccinations to achieve herd immunity. Many scientists support that they are the most effective health measures to break the coronavirus transmission chain. However, the technologies ensuring those health measures have not been broadly adopted due to privacy and accuracy concerns. Without gaining a critical mass of individual users, these personal technologies have been rendered useless. Although large-scale policy efforts have been made aggressively, the technologies cannot effectively support federal, state, and local governments' coordination and regulation enforcement logistics.

In this work, we propose Crowd-based Alert and Tracing Services (CATS) to build a safe community cluster, which provides higher privacy protection, efficient penetration

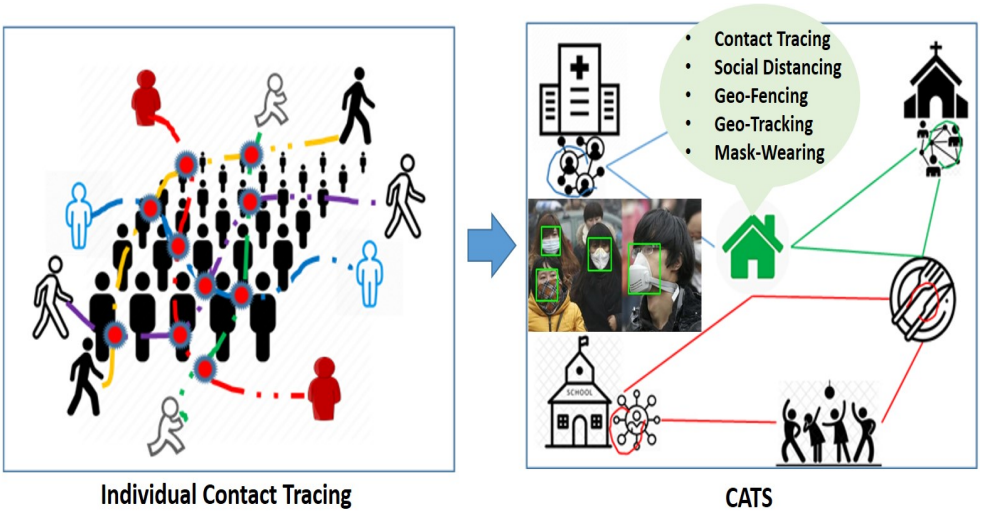

Figure 49: CATS system concept

of technology, greater accuracy, and effective practical policy assistance [152]. As illustrated in Figure 49, CATS enhances the technology-based tracing capacity by transforming the task from personal tracking to community gatekeeping and from binary to multi-context of contact information and policy assistance. First, as society gradually reopens, each community, such as schools, churches, businesses, and events, needs to be evaluated for appropriate gatekeeping methods such as masks and sanitization requirements, and temperature checks to ensure the members' safety. It is critical to deploy tracing and social distancing methods among the members. CATS facilitates tracing at a community or a facility level using multiple form-factors (i.e., a smartphone app, plug-in, or a smart tag) rather than an individual level to bear the characteristics of contacts according to the adoption choices of specific communities. Second, CATS enables public authorities to efficiently and dynamically assess their social distancing policies using the area-based

safety value maps (opt-in data), various duration and distance alerts, and actively informing others via direct covert communication non-binary Safety Impact Values (SIV). As of the first step, we have designed and developed a Machine Learning (ML)-based face mask tracking method to find SIV of the community by measuring the % of mask-wearing and the % of no or wrong mask-wearing people. It can adequately educate policymakers about the pandemic's meaningful status at the broader level and assist in effective policy decisions and relief plans. Community-based safety spectrum data such as SIV from the masking status (crowdsourcing data from each community) will become a novel dataset that would augment existing biological COVID-19 data with sociological data.

5.1.1 Crowd-based Alert and Tracing Services (CATS)

As shown in Figure 50, during the pandemic, almost all the communities are gatekeeping by putting up signs to wear masks and keep social distancing. However, putting up posters is not practical since some may intentionally challenge those rules, and others neglect them. Also, students who stay extended hours in school may not follow the social distancing and mask-wearing rules unintentionally. A more effective and less invasive nudge is required. However, surveys show that many people hesitate to adopt those technologies due to privacy and accuracy concerns.

Figure 50: Typical signs for pandemic rules

5.1.1.1 Social Distancing Measurement

As illustrated in Figure 51, we have implemented a proof-of-concept tag with WiFi and BLE beacon stuffing and RSSI-based distance measurement functionalities using ESP 32 chipset. An Android smartphone app is developed to control the tag, and Google Cloud Messaging (GCM) is used for server communication. Both WiFi and BLE beacons should work in real-world environments, which may have hundreds of tags. When the beacon signal becomes prevalent in a crowded intersection, there could be a chance of beacon collision among the tags. Although WiFi and BLE beacons support spatiotemporal frequency isolation methods for mitigating the potential interference, it happens due to hidden nodes, periodic delivery, and broadcast. Hence, we look for an efficient beacon technology that can control and mitigate the beacon collision. A few theoretical studies show that periodic beacon's collision probability in practice is high. CATS tag uses relative location (distance) to measure social distancing. The length of two BLE tags can be calculated using the RSSI power using a simplified formula in Eq. (5.3) [153].

The RSSI signal strength depends on distance and broadcasting power value. BLE works with broadcasting power value (N) around 2â4 dBm, which depends on the environmental factor. The signal RSSI strength will be around -26 (a few inches) to -100 (40â50 m distance). The Measured Power is a factory-calibrated constant of expected RSSI at a distance of 1 meter.

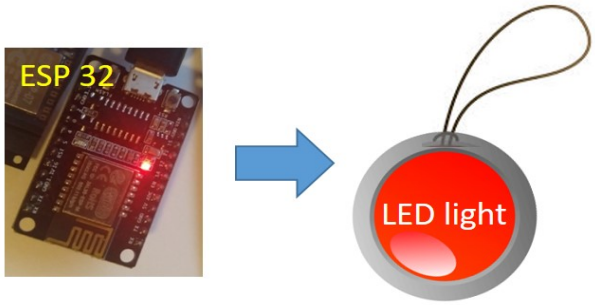

Figure 51: Conceptual Smart Tag form-factor illustration

$$Distance = 10^{(MeasuredPower - RSSI)/(10*N)} \tag{5.1}$$

The RSSI value would be less accurate and not stable for distance measurement in a densely populated area such as shopping malls, grocery stores, and office buildings. However, considering the nature of social distancing problems, the environment is less populated, and sub-meter distance accuracy is not significant. The Ultra Wide Band (UWB) can measure distance and location to an accuracy of 5 to 10 cm with minimal noise interference due to the short pulse width, unlike WiFi, Bluetooth, and other narrowband radio systems that only reach several meters. UWB also consumes less power than WiFi,

although Bluetooth 4.0 also uses significantly low power. However, UWB is not as pervasive as Bluetooth and WiFi. None of the current smartphones and mobile devices support UWB except the recent Apple's iPhone 11. As we choose to use both BLE and WiFi beacons, we have conducted various ways to enhance the accuracy, including contextual information, characterization of occlusion materials for different signals, and investigate device positions (height and movement) accurate distance analysis. CATS supports different configuration and operation options to cope with social distancing's various privacy and security requirements, which differ among communities such as schools, churches, industries, and government facilities. As monitoring and reporting are contained within an organization, any public system (server or other tags) does not read personal data, including personal identification and SIV inputs. They are kept in private within the personal smartphone app. The SIV scores are shared only after the ID anonymization. Any log data only stays within the community server and will be removed within a couple of weeks without any reviews.

5.1.1.2 Mask-Wearing Detection (MWD) for SIV

To reduce the spread of COVID-19, CDC urges individuals to cover their mouths and noses with a mask when they around others. Wearing a mask is meant to protect other people in case someone in the group is infected. Many people do not wear masks, even in public areas, and when social distancing measures are challenging to maintain. Such cases can be seen when crowds walk on the street, in airports, and schools. As presented in Figure 52, MWD system collects data, detects face masks, classifies the

masking condition in a crowd, and counts people. To serve this purpose, MWD consists of data training and feature extraction modules. Eventually, we will deploy the model into the CCTVs. MWD can report SIV for each community and provide a sense of which area is safe from the spread of COVID-19. The system only creates SIV of the community and eliminates any personal identities to ensure privacy.

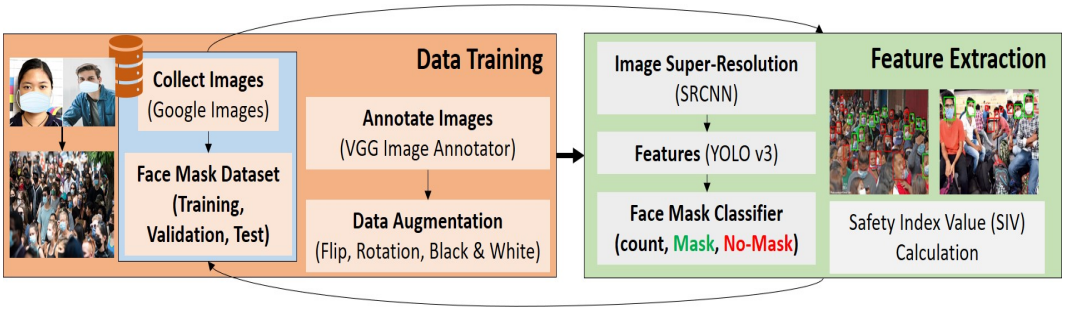

Figure 52: MWD system architecture

- **Data Training** module is in charge of annotating and augmenting crowd masking images to create datasets and trains the collected datasets to enable various ML procedures. First, we create and train new **MWD Datasets** for the crowded environment. Mask images were collected from Google images by running web search scripts with various keywords, including "mask." The dataset contains 526 images with either a mask or no mask in a crowd. As illustrated in Figure 52, the images have various crowd density, unlike the existing mask datasets [101, 102], which only contain images with a small number of faces for surveillance purposes. We divided our datasets into three categories, training set (70%), validation set (20%), and test set (10%). Second, the VGG Image Annotator (VIA) software [21] is used for **data annotation**. It manually draws bounding boxes on the images and assigns each labeled object to *Mask* that includes

all people wearing a mask, and *No-Mask* that refers to all people without any mask. We also generate the ground-truth density map. All fully or partially front-facing human faces are labeled *Mask* and *No-Mask*. Every labeled object and the information are exported to a .json file for training. We converted the .json file information to the .txt format. Third, **data augmentation** techniques, including scaling, flipping, rotation, and converting images color scale to black and white scale, are applied to increase the training set. We use these techniques to grow the dataset to 1262 images. It improves model performance by reducing the chance of model overfitting. Finally, for **training configuration**, Adam optimizer [32] is used as an optimization method for the training of Mosaic by configuring a learning rate (1e-4), and a momentum (0.9). Performing multiple experiments starting from 1e-3 to 1e-9, we found that 1e-4 is the most optimal learning rate. Other recommended training hyper-parameters are adopted, including a batch of 16 and an epoch number of 100. It is implemented using the Pytorch framework [30]. All of the experiments were conducted on an NVIDIA GeForce GTX 1080 Ti.

- **Feature Extraction** module uses a YOLO v3 [150] to extract features. It continuously passes the extracted features to a masking check function, which detects and evaluates the masking status in various aspects, including mask, no-mask (incorrect masking). We have changed the hyperparameters of Yolo, such as batch size, learning rate, and epochs number. The Super-Resolution (SRCNN) technique reconstructs a High-Resolution (HR) image from the Lower-Resolution (LR) images before applying the YOLO. After the feature extraction, we have added a classification layer at the

end of YOLOv3. It classifies the masking features like a red box for *No-Mask* and a green box for *Mask*. It also counts the total number of people with *Mask* and *No-Mask*. Eventually, SIV is calculated for each image by applying the mask ratio, which is $SIV = Total_{mask}/Total_{people}$.

5.1.2 Evaluations

We conducted an ML-based MWD experiment using the trained MWD prototype with the testing dataset. Figure 53 presents the detection results of incorrect masking (e.g., the neckbeard, the sniffer, the stache, and the nose plug) and various face mask styles and colors. For example, the mask should cover the nose and mouth to stop the spread of infection. The red box detects both *No-Mask* and *incorrect mask*.

Figure 53: Detection of various masks and incorrect masking

Figure 54 shows the calculated SIV in % from a target image. While the target image is going through the MWD model, it identifies the number of detected people ($Total_{people}$), total number of people with Mask ($Total_{mask}$), the number of people with No-Mask ($Total_{no-mask}$), and the original count of people in an image. SIV is calculated

for each image by applying the mask ratio, which is $SIV = Total_{mask}/Total_{people}$. For example, the leaft image of Figure 54 has 82 of $Total_{people}$ and 9 of $Total_{mask}$ that results 11% of SIV (very unsafe as only 11% is masking).

Figure 54: SIV calculations

$$Recall = \frac{TP}{TP + FN}, \; Precision = \frac{TP}{TP + FP} \tag{5.2}$$

Table 10: Classification performance

Mask Recall	Mask Precision	No Mask Recall	No Mask Precision	mAP
82%	79%	65%	63%	71%

The recall rate and precision metrics defined by Eq. (5.2) with true-positive (TP), false-negative (FN), and false-positive (FP) are used to evaluate the classification detection performance. A mean Average Precision (mAP) in object classification tasks is also

calculated. mAP is in a range from 0% to 100%, and the higher value means better accuracy. As presented in Table 10, mask recall of 82%, mask precision of 79%, no mask recall of 65%, no mask precision of 63%, and mAP (average) of 71% are pretty accurate for detection tasks in a crowd. Both recall and precision results in *No Mask* are relatively low because most of the labeled objects in the dataset are with *Mask*.

5.2 Part 2: Crowd Safety Sensing (CroSS) for the Post Pandemic Era

COVID-19 causes multiple pandemic waves worldwide for years due to the nature of the long incubation period, the aggressive asymptomatic transmission, and new mutations of the virus. The lockdown (shrinking the community at the elementary level, such as a family and cutting the strings between the communities) for social distancing is the last resort against it. In addition to vigorous testing and vaccination, ”very aggressive” social distancing in a public place is a vital part of the strategy for keeping such exponential infection from happening again after the first wave recedes and the society reopens. However, the existing technologies, such as contact tracing apps, have not been adopted due to privacy and accuracy concerns. Without achieving a critical mass of individual users, these personal technologies have been rendered useless. Although large-scale policy efforts have been complicated, requiring the coordination of federal, state, and local governments and the logistics of regulation enforcement, applying technologies to small, focused communities can retain individual privacy, achieve wide user adoption, and allow easy implementation. They are not helpful for the post-pandemic era (a new normal) to prevent another potential future pandemic wave.

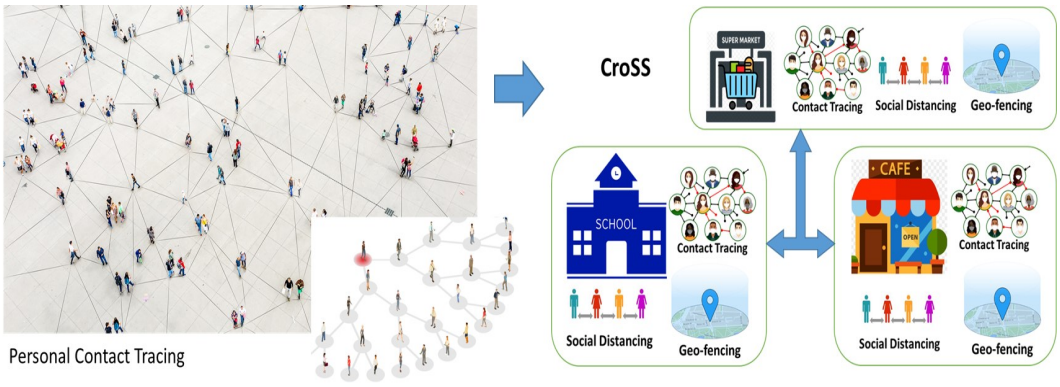

Figure 55: CroSS system concept vs. personal contact tracing

In this work, we propose a novel sensing strategy and system, named **Crowd Safety Sensing (CroSS) to build a sustainable safe community cluster**, which provides efficient penetration of technology, greater accuracy, effective practices, and privacy policy assistance [154]. CroSS uses affordable Internet of Things (IoT) technologies to transfer to small communities without any heavy facility investment. As illustrated in Fig 55, CroSS system facilitates effective social distancing technologies at a community or a facility level rather than a personal level. Limiting the tracing scope from an unbounded area (e.g., city-wide or nation-wide tracing scope) to each community, CroSS can improve scalability and effectiveness and supports the characteristics of contacts to the adoption choices of specific organizations. CroSS enhances the technology-based tracing capacity by transforming the task from personal tracking to community gatekeeping, from binary to multi-context of contact information and policy assistance, and from backtracking to forward-tracing.

CroSS enables innovative capabilities and transforms the technology-based tracing as follows:

- **From a personal tracking tool to a community gatekeeper:** As society gradually reopens, each community, such as schools, churches, businesses, and events, need to be evaluated for appropriate gatekeeping methods such as social distancing, mask-wearing and sanitization requirements, and temperature checks to ensure the members' safety. Among them, it is critical to deploy alerting and social distancing methods among the members. CroSS facilitates alerting at a community or a facility level using multiple form-factors (i.e., a smartphone app, plug-in, or a smart tag) rather than an individual level to bear the characteristics of contacts to the adoption choices specific communities. Furthermore, the anonymized and encrypted tracing data remains within the involved community for a limited duration. It enhances privacy and adoption and penetration of technology by incentivizing communities to opt-in for their activities and embracing people without smartphones or reluctant to contact tracking,

- **From binary to multi-context of contact information and policy assistance:** The existing contact tracing strategy uses binary information (a confirmed case or not) only after a long potential asymptomatic period. CroSS enables public authorities to efficiently and dynamically assess their social distancing policies using the area-based safety value maps (opt-in data), various duration and distance alerts, and actively informing others via direct covert communication non-binary Safety Impact Values (SIV). It can adequately educate policymakers about the pandemic's meaningful status at the broader level and assist effective policy decisions and relief plans.

- **From backtracking to forward-tracing:** All existing contact tracing approaches are designed to trace retrospectively for the confirmed case only. Hence, they are not scalable to tackle the pandemic, which has prolonged incubation and asymptomatic transmission. CroSS can proactively track potential incubation case contacts by exchanging the safety spectrum between close contacts instead of waiting for a confirmed case.

As illustrated in Fig 56, many ongoing privacy-aware smartphone-based contact tracing approaches by Apple and Google [106], and PACT [107] take a similar concept at its foundation, which consists of exchanging a key code stage (steps 1 and 2), sharing the key code by an infected stage (steps 3 and 4), and alerting (opt-in) the contacts stage (steps 5 and 6) [155]. However, existing methods have significant limitations in technology adoption, accuracy, and efficacy.

CroSS (1) effectively incentivize technology adoption among communities to provide (2) meaningful assistance to policymakers and public health authorities. Tracing boundaries and details are set clearly by adopting communities focusing only on their stakeholders and based on their unique tracing context (e.g., teacher vs. student, pastor vs. parishioner), thus (3) providing a significant reduction in false negatives while (4) enriching contact safety information. The effective penetration and the meaningful and enriched contact tracing information enabled by our approach will allow other researchers in various disciplines to model and understand the spread of COVID-19, inform and educate policymakers and public health authorities about the science of virus transmission and prevention, and encourage the communities and policymakers on the development of

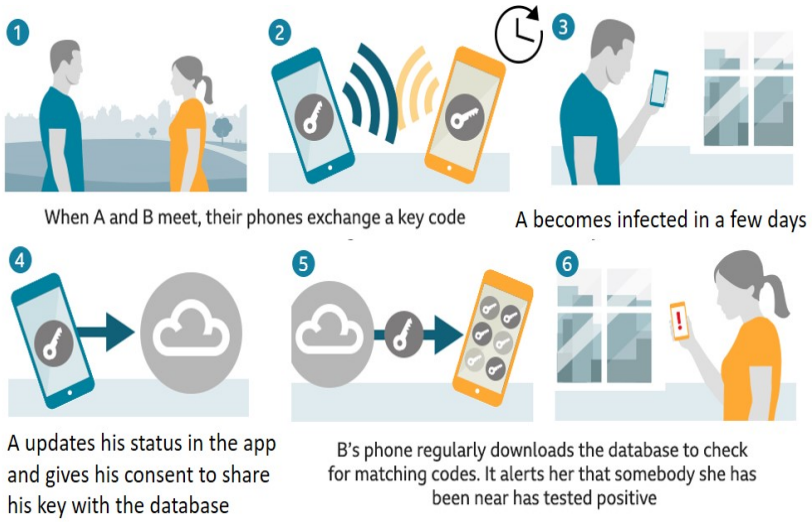

Figure 56: Privacy-aware contact tracing by Apple and Google

processes and actions to address this global challenge. Community-based safety spectrum data such as Safety Impact Values (SIV) crowd-sourcing data from each community will be a novel dataset that would augment existing biological COVID-19 data with sociological data. New community-specific contact patterns would emerge, unique in duration, distance, and the number of contacts. The impact of each contact can be disparate, according to the community.

5.2.1 Crowd Safety Sensing (CroSS)

As shown in Fig 57, almost all the communities are gatekeeping by putting up signs to wear masks and keep social distancing. However, just putting up posters is insufficient, since some are challenging the value of those rules, and others are neglecting

them. For example, some customers may intentionally refuse to wear masks. Also, students who stay extended hours in school may not follow the social distancing and mask-wearing rules unintentionally. A more effective and less invasive nudge is required. A growing number of contact tracing apps are being developed and released to complement manual contact tracing and ensure social distancing using location data or BLE beacons to automatically detect if a user may have been exposed to risk. However, surveys show that many people hesitate to install those apps due to the app's privacy and accuracy concerns.

Figure 57: Signs for pandemic rules

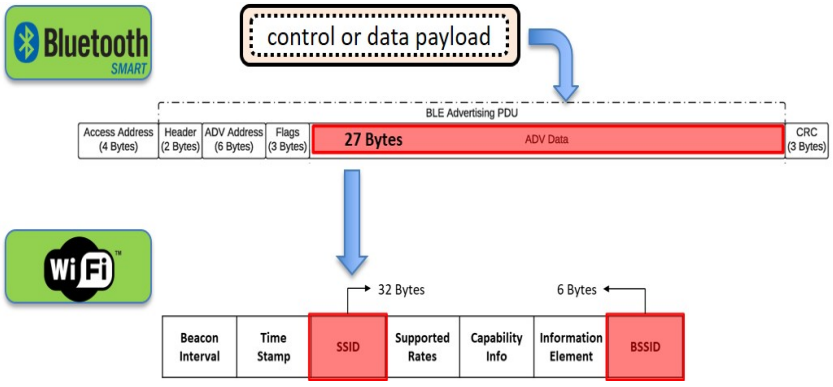

Figure 59: WiFi and BLE Beacon stuffing

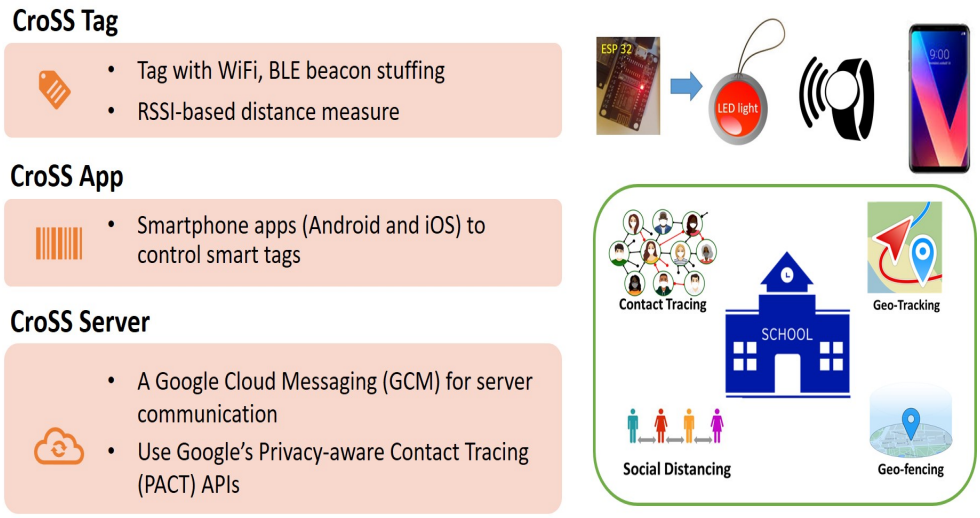

Figure 58: CroSS system architecture

5.2.1.1 Technology Overview

As illustrated in Fig 58, CroSS system consists of three main components, including a CroSS tag (hardware and software), a CroSS app, and a CroSS server. **A CroSS tag** uses both WiFi and BLE beacons. BLE beacon is used between tags to do physical distancing and contact tracing, using a periodic active beacon. WiFi beacon is used between the community facilities (i.e., WiFi access points) and tags for fencing and tracking the tag holders, which uses a passive beacon with probe request and response. As illustrated in Fig 59, a tag identity and supplementary information (i.e., SIV) are embedded into a beacon frame by using a beacon stuffing technique [119]. A CroSS tag prototype was built using ESP32, a series of low-cost, low-power systems on a chip microcontroller with integrated WiFi and dual-mode Bluetooth 4.2 communications, built-in antenna switches, and 520 KB SRAM. It comprises an LED light to indicate a social distancing alert, a

battery (charging is optional), a microSD card for extended logging, and a power switch (toggle to turn on and off). As of future work, it will be wrapped by a water-resistant cover with a harness (i.e., neckless and wristband). **A CroSS app** is for both Apple and Google by using a cross-platform IDE (i.e., Flutter). When a member checks in a community, a CroSS tag ID is associated with the app and registered to the CroSS server. The app remains disabled most of the time while the tags are exchanging signals. CroSS app facilitates a Safety Impact Value (SIV) calculation before any registration. The app calculates an SIV (score) with the age, race/ethnicity, gender, some medical conditions, use of certain medications, poverty and crowding, certain occupations, pregnancy, and recent travel history information. It enables the non-binary tracing model. The SIV score will be readjusted automatically according to the exposure duration, distance, and environment. It also supports the check-out process to initiate the contact tracing data upload and tag reset. **A CroSS server** maintains the encrypted contact tracing information and SIVs from social distancing measurement for a limited time (i.e., for two weeks). The server will delete the contact tracing data and associated SIV data within the community service in a couple of weeks without any review. The server also monitors the occupancy density (to identify any high-risk area) in all community parts to ensure an environment capable of social distancing according to the configuration. It ensures compliance by checking alerts if a tag is still functioning.

CroSS system supports four conceptual functions, including social distancing, contact tracing, geo-fencing, and geo-tracking.

- **Social Distancing:** Each member wears a CroSS tag that continuously monitors

and stores beacon messages from nearby tags. If the received signal strength (i.e., RSSI) is approximately 6 feet for a certain duration, the tag indicates a visual warning by flashing a red LED light. It also logs each tag's unique identifier, along with a timestamp and the approximate distance. They will be transferred to the CroSS server, when the tag enters the gateway.

- **Contact Tracing:** The concept of contact is different according to the community system configuration, such as distance, duration, and SIV information (i.e., wearing masks, travel, and health history). Contact tracing provides a dynamic reporting application to identify members who have been in contact with a confirmed case. The tag logs the unique identifier, timestamp, and approximate distance when there is a contact. A CroSS server database renders all contact points and allows for further analysis and reporting. If a member is tested positive later, all community contacts for the last two weeks are tracked and notified. The application provides an analysis of high-risk areas and people to enable the development of proactive mitigation strategies. The data captured is called through an API, or optionally, enriched with social distancing monitoring data or peopleâs COVID-status information. The community can communicate with affected members to implement their management processes.

- **Geo-Fencing:** The fencing technology keeps a tag within a predefined set of community boundaries, which looks for weakening and losing the âheartbeatâ from a tag. For example, if a tag holder tries to leave the community without returning the tag (brought it to home or other places), it will give a visual warning. The process

starts with the WiFi APs, which broadcast WiFi beacon (a probe request) periodically. If a tag does not receive any WiFi beacon signal for a long duration(over a threshold), it turns an LED light on to complete the check-out process by returning the tag.

- **Geo-Tracking:** The tracking technology searches a specific heartbeat signal from a missing tag to identify the current location or the entire location history. For example, a tag can be missing because a tag holder leaves the tag somewhere within the community without adequately checking out (a zombie tag). The process starts with the CroSS server when it receives an inoperative or missing tag report. It enables the missing tag area by sending a service request message via a WiFi beacon and sends a probe request to trigger a probe response from the missing tag. When a nearby smartphone app receives a probe response from a (lost) tag, it relays its location.

5.2.1.2 IoT-based CroSS Design and Development

We have implemented a proof-of-concept tag with WiFi and BLE beacon stuffing and RSSI-based distance measurement functionalities using ESP 32 chipset. An Android smartphone app is developed to control the tag, and a Google Cloud Messaging (GCM) is used for server communication. CroSS system operates massive scale deployment and practical experiments.

Both WiFi and BLE beacons have to work in real-world environments, which may

have hundreds of tags. When the beacon signal becomes prevalent in a crowded intersection, there could be a chance of beacon collision among the tags. Although WiFi and BLE beacons support spatiotemporal frequency isolation methods for mitigating the potential interference, it happens due to hidden nodes, periodic delivery, and broadcast. Hence, we look for an efficient beacon technology that can control and mitigate the beacon collision. We address the critical and practical issues of power usage and beacon collision at the same time by eliminating the periodic beacon message broadcast. A few theoretical studies show that periodic beacon's collision probability in practice is high. We design a specific SSID-based beacon probe request and probe response mechanism. An 802.11 probe request or response frame format is similar to a WiFi beacon frame format. However, the probe response is sent in response to a probe request. The probe responses don't carry the Traffic Indication Map (TIM) that identifies stations using a power-saving mode. For example, WiFi Direct enables devices to connect without requiring a wireless access point uses a probe request and response when a station searches a peer WiFi Direct station. It continues to send probe request frames (with an SSID named DIRECT) and listen for a while. When another WiFi Direct station receives a probe request, it triggers a probe response. Instead of configuring tags to send periodic WiFi beacon frames, we enable a listening mode for a tag that only responds upon a probe request. Instead, WiFi APs periodically initiates a probe request that triggers probe responses from the tags in the range. The second innovative approach is to let WiFi APs send a probe request with a specific SSID (i.e., a group of tags or a missing tag). The probe request triggers only one probe response from the matching group of tags instead of triggering multiple concurrent probe

responses. It controls the concurrent degree of the probe responses using the proposed approach. This opportunistic trigger mechanism reduces the number of beacon messages in the network and significantly enhances a tag's power usage.

CroSS tag uses relative location (distance) to detect both contact and social distancing. The distance of two BLE tags can be calculated using the RSSI power using a simplified formula in Eq. (5.3) [153]. The RSSI signal strength depends on distance and broadcasting power value. BLE works with broadcasting power value (N) around 2â4 dBm, depending on the environmental factor. The signal RSSI strength will be around -26 (a few inches) to -100 (40â50 m distance). The Measured Power is a factory-calibrated constant of expected RSSI at a distance of 1 meter.

$$Distance = 10^{(MeasuredPower - RSSI)/(10*N)} \tag{5.3}$$

The RSSI value would be less accurate and not stable for distance measurement in densely populated areas such as shopping malls, grocery stores, and office buildings. An alternative to RSSI is the Bluetooth direction-finding technologies such as angle-of-arrival (AoA) and angle-of-departure (AoD). They provide sub-meter location accuracy and are ideal for dense indoor areas. Also, they can offer better granular data on person-to-person contacts. However, considering the nature of social distancing and contact tracing problems, the environment is less populated, and sub-meter distance accuracy is not significant. Rather than distance accuracy, as illustrated in Fig 60, there are numerous false-positive cases caused by various occlusions. For example, people can be counted

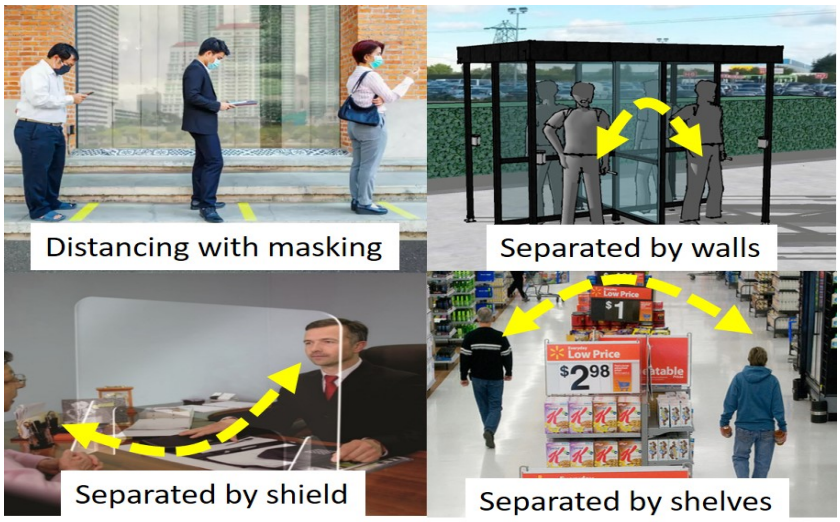

Figure 60: Erroneous scenarios due to occlusions

as the social distancing violation cases if they are within 6 feet but are (a) staying in different spaces separated by a wall, (b) talking over the anti-spray transparent glass, or (c) shopping in different aisles separated by high shelves. Barriers (i.e., desks, bags, etc.) can vary the distance measurement by absorbing signals. Also, when a family member with separate tags goes together, the result can be erroneous. We use the Channel Impulse Response (CIR) to detect false positive, false negative, and erroneous scenarios caused by occlusions. We tested the channel sounding effect of various RF signals (WiFi and BLE) in the presence of multiple occlusions and a large number of concurrent tags. It is known that the Ultra Wide Band (UWB) can measure distance and location to an accuracy of 5 to 10 cm with minimal noise interference due to the short pulse width. At the same time, WiFi, Bluetooth, and other narrowband radio systems can only reach several meters. UWB also consumes less power than WiFi, although Bluetooth 4.0 also uses significantly low power. However, UWB is not as pervasive as Bluetooth and WiFi. None of the

current smartphones and mobile devices support UWB except the recent Apple's iPhone 11. As we choose to use both BLE and WiFi beacons, we have conducted various ways to enhance the accuracy, including contextual information, characterization of occlusion materials for different signals, and investigate device positions (height and movement), and accurate distance analysis. Specifically, we use the RSSI power interruption and the RSSI variations to detect any tags' interference.

CroSS supports different configuration and operation options to cope with various privacy and security requirements of social distancing and contact tracing, which differ among communities such as schools, churches, industries, and government facilities. As monitoring and reporting are contained within a community, any public system (server or other tags) does not read personal data, including personal identification and SIV inputs. The SIV scores are shared only after the ID anonymization. The contact identities will be resolved only after opt-in for confirmed contact case tracing is conceded as other privacy-aware contact tracing approaches. Any log data only stays within the community server and will be removed within a couple of weeks without any reviews.

5.2.2 Evaluations

We first explain RSSI (power and variation) experiments for detecting false-positive social distancing cases due to an occlusion. As shown in Fig 61, we place a smartphone (Samsung Galaxy 8) based BLE transceiver at 1 m height and an ESP 32 based BLE transmitter at 1 m height 2.3 m apart in an indoor environment. We put an object interference from none to one near the BLE transmitter. We set up the transceiver to scan each test for

a minute, with ten separate times for each test session.

Figure 61: Smart tag experiment setup

5.2.2.1 False-positive social distancing

As presented in Fig 62, we measure the beacon's RSSI power metric to detect the impact of the interfering object. We test the RSSI power both with interference (with shield (WS) and no interference (no shield (NS)). The test results show that the average RSSI power in NS is steadily stronger than in WS. It indicates that we can detect the existence of an object between the beacon transmitter and receiver using normalized average power values in addition to the distance.

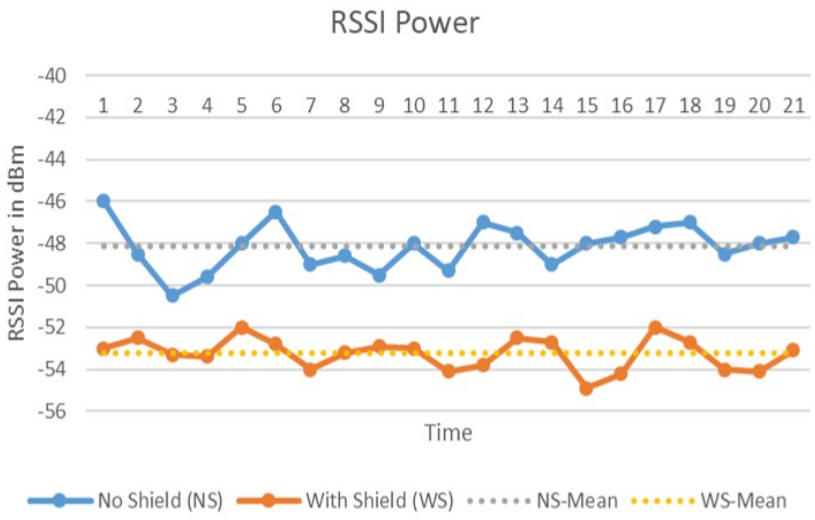

Figure 62: Occlusion detection with RSSI power

As presented in Fig 63, we also tested a variation of the received RSSI power. We pick the maximum and minimum RSSI power values among the beacons received in a second (Vps), and use the difference as a variation value. The indoor RSSI variation test results show that the RSSI variation in NS is stable for all workloads from 8.5 to 10 Vps while the RSSI variation WS is dynamically fluctuating, which can be seen in their standard deviation values. In summary, the results indicate that both RSSI average power and variation metrics can identify any interference between the beacon transceiver and receiver.

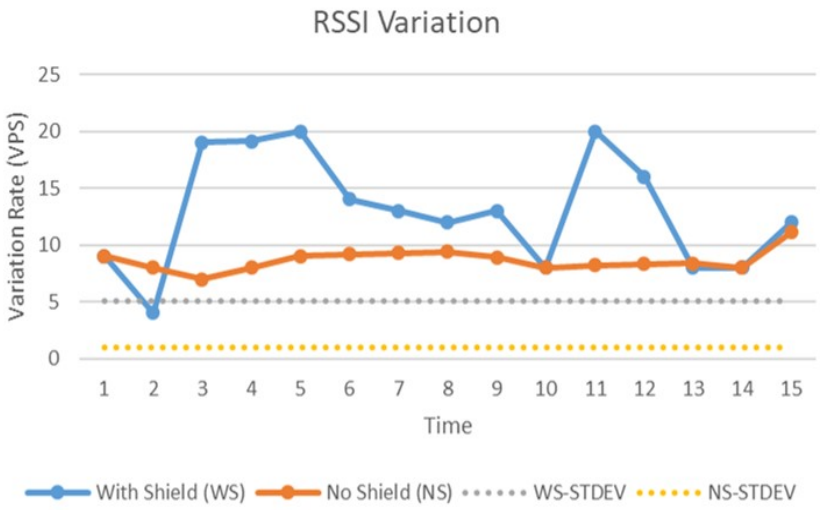

Figure 63: Occlusion detection with RSSI power

5.3 Conclusion

In This chapter, we proposed two novel infection management systems in order to build a safe community cluster against COVID-19 and beyond.

CATS applies social distancing and mask-wearing principles to small, focused communities. We also implemented an ML-based mask tracking system to build non-binary Safety Impact Values (SIV). CroSS proposed Crowd Safety Sensing, which is a novel tracing strategy and system to build a sustainable, safe community using affordable Internet of Things (IoT) and edge-enabled technologies. We have integrated CroSS components into an edge-based IoT system.

These approaches provide higher privacy protection, efficient penetration of technology, greater accuracy, and effective practical policy assistance. The feasibility test results show that CroSS can identify false-positive social distancing cases.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, we presented and discussed an Intelligent Crowd Engineering platform using Machine-based Internet of Things Learning and Knowledge Building approaches (ICE-MILK) to enhance the accuracy, scalability, and crowd safety management capacity in real-time. Specifically, we design a three-layer ICE-MILK structure, including IoT-based mobility characterization layers, ML-based video/image surveillance, and the semantic knowledge information-based application layer.

In the IoT-based mobility characterization layer, we designed, implemented, and tested CROMO that enhances crowd mobility characterization through real-time BLE data analytics. We built a CROMO layer that fills the scalability and capability gaps of the smart video surveillance via BLE signal analytics. Among the many crowd mobility characteristics, we measure the beacon counts, the RSSI power, and variation pattern using a BLE bracelet and BLE scanners. We also apply them to group semantics to track the crowd status and predict potential accidents and disasters. We have tested CROMO in both a practical crowd event and controlled indoor and outdoor lab environments. The results show that by integrating the BLE data metrics, CROMO can identify the crowd density, the object group location, and the flow direction and speed in real-time. It shows the feasibility of using the proposed CROMO layer to enhance the safety management system.

In addition, In the ML-based video/image surveillance layer, we proposed an architecture for scattered occlusion characterization called CSONet for efficient crowd counting and high-quality density heatmap generation. We first generated, annotated, and trained two new scatter occlusion object datasets: the cso-umbrellas and the cso-pickets datasets. We then implemented CSONet using spatial pyramid pooling and dilated convolutional layers to expand the receptive field without losing resolution in the congested scenes. CSONet recognizes event-induced, scattered, and multitudinous occlusions and applies the effect to a human crowd map to generate an accurate crowd count and high-quality density map. Through extensive evaluations, we demonstrated that the accuracy of CSONet outperforms the state-of-art existing crowd counting approaches. Also, we have proposed a novel face masking detection system for Modeling Safety Index in Crowd (Mosaic), a Machine Learning (ML)-based approach for detecting masking in a crowd by building new dense mode crowd masking datasets. Mosaic detects, counts, and classifies the crowd's masking condition and calculates each community's spatiotemporal Safety Index (SI) values instead of monitoring individual masking cases. SI data can be shared or published to calculate the area-based SI maps (as opt-in data) for assisting effective policy decisions and relief plans against COVID-19. Through variable evaluations, we have shown that Mosaic scalably detects various conditions and types of masking states and calculates SI values of a crowd effectively. Mosaic can provide an accurate assessment of SI of a community based on real measurement of masking and social distancing practices. It can also support pandemic trajectory predictions and input policy planning with better precision in a scalable manner. Future work includes incorporating Mosaic into pandemic

prediction modeling for improved ramification and accuracy of predictions.

Finally, In the semantic knowledge information-based application layer, we proposed a novel tracing strategy and system named Crowd-based Alert and Tracing Services to build a safe community cluster using affordable Internet of Things (IoT) and edge-enabled technologies. It applies social distancing and mask-wearing principles to small, focused communities to provide higher privacy protection, efficient penetration of technology, greater accuracy, and effective practical policy assistance. We have implemented a smart tag to support social distancing. We also implemented an ML-based mask tracking system to build non-binary Safety Impact Values (SIV).

In our future work, we will integrate the ICE-MILK components to enhance the accuracy, scalability, and crowd safety management capacity in real-time. In addition to density and count detection, we will identify group characteristics such as mobility type, directions by using ML technologies. For example, Group can be clustered according to people facing the same direction or moving in the same way.

CHAPTER 7

APPENDIX

The following papers have been published as a direct result of the research discussed in this dissertation:

- K. J. Almalki, M. Mohzary, B. -Y. Choi, S. Song and Y. Chen, "Crowd Safety Sensing (CroSS) for the Post Pandemic Era," 2021 IEEE Globecom Workshops (GC Wkshps), 2021, pp. 1-6, doi: 10.1109/GCWkshps52748.2021.9682120.

- K. J. Almalki, M. Mohzary, B. -Y. Choi, S. Song and Y. Chen, "Mosaic: Modeling Safety Index in Crowd by Detecting Face Masks against COVID-19 and Beyond," 2021 IEEE International Smart Cities Conference (ISC2), 2021, pp. 1-7, doi: 10.1109/ISC253183.2021.9562953.

- K. J. Almalki, B. -Y. Choi, Y. Chen and S. Song, "Characterizing Scattered Occlusions for Effective Dense-Mode Crowd Counting," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021, pp. 3833-3842, doi: 10.1109/ICCVW54120.2021.00428.

- K. J. Almalki, S. Song, M. Mohzary and B. -Y. Choi, "CATS: Crowd-based Alert and Tracing Services for building a Safe Community Cluster against COVID-19," 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2021, pp. 697-701.

- A. Jabbari, K. J. Almalki, B. -Y. Choi and S. Song, ”ICE-MoCha: Intelligent Crowd Engineering using Mobility Characterization and Analytics,” 2019 Sensors, 19, 1025. https://doi.org/10.3390/s19051025.

- A. Jabbari, K. J. Almalki, B. -Y. Choi and S. Song, ”CROMO: Enhancing Crowd Mobility Characterization through Real-time Radio Frequency Data Analytics,” 2018 IEEE International Smart Cities Conference (ISC2), 2018, pp. 1-8, doi: 10.1109/ISC2.2018.8656977.

REFERENCE LIST

[1] Keith Still, Marina Papalexi, Yiyi Fan, and David Bamford. Place crowd safety, crowd science? case studies and application. *Journal of Place Management and Development*, 2020.

[2] G Keith Still. *Introduction To Crowd Science*. CRC Press, 2014.

[3] Wikipedia. Hajj, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Hajj.

[4] Google. Coronavirus (covid-19), 2017. [Online]. Available: https://news.google.com/covid19/map?hl=en-USmid=%2F%2F02j71gl=USceid=US%3Aen.

[5] CNN. At least 8 dead and many injured after crowd surge at travis scott's astroworld festival in houston, 2021. [Online]. Available: https://www.cnn.com/2021/11/06/us/houston-astroworld-festival/index.html.

[6] BBC News. India's kumbh festival attracts big crowds amid devastating second covid wave, 2021. [Online]. Available: https://www.bbc.com/news/world-asia-india-56770460.

[7] Dorine C Duives, Winnie Daamen, and Serge P Hoogendoorn. State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies*, 37:193–209, 2013.

[8] Yasushi Nakauchi and Reid Simmons. A social robot that stands in line. *Autonomous Robots*, 12(3):313–324, 2002.

[9] Sebastian Bek and Eduardo Monari. The crowd congestion levelâa new measure for risk assessment in video-based crowd monitoring. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1212–1217. IEEE, 2016.

[10] Aniket Bera and Dinesh Manocha. Realtime multilevel crowd tracking using reciprocal velocity obstacles. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 4164–4169. IEEE, 2014.

[11] Abdoh Jabbari, Khalid J Almalki, Baek-Young Choi, and Sejun Song. Cromo: Enhancing crowd mobility characterization through real-time radio frequency data analytics. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–8. IEEE, 2018.

[12] Jens Weppner and Paul Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 193–200, 2013.

[13] Jianxin Wu. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23):495, 2017.

[14] Daniel Onoro-Rubio and Roberto J López-Sastre. Towards perspective-free object counting with deep learning. In *European conference on computer vision*, pages 615–629. Springer, 2016.

[15] Ke Chen, Shaogang Gong, Tao Xiang, and Chen Change Loy. Cumulative attribute space for age and crowd density estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2467–2474, Portland, OR, USA, 2013. IEEE.

[16] Tal Hassner, Yossi Itcher, and Orit Kliper-Gross. Violent flows: Real-time detection of violent crowd behavior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, Providence, RI, USA, 2012. IEEE.

[17] Lei Liu, Wenjing Jia, Jie Jiang, Saeed Amirgholipour, Yi Wang, Michelle Zeibots, and Xiangjian He. Denet: A universal network for counting crowd with varying densities and scales. *IEEE Transactions on Multimedia*, 2020.

[18] Deepak Babu Sam, Neeraj N Sajjan, R Venkatesh Babu, and Mukundhan Srinivasan. Divide and grow: Capturing huge diversity in crowd images with incrementally growing cnn. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3618–3626, Salt Lake City, UT, USA, 2018. IEEE.

[19] Yuji Roh, Geon Heo, and Steven Euijong Whang. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1328–1347, 2019.

[20] Jason Brownlee. *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery, 2020.

[21] Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA, 2019. ACM.

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *In Proceedings of the IEEE International Conference on Learning Representations (ICLR)*, May 2015.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016.

[24] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[25] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[26] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[27] Eugenio Grant. Covid Mask Detection with Yolo v3., 2021. [Online]. Available: https://www.kaggle.com/eugeniogrant/covid-mask-detection-with-yolo-v3.

[28] Chinnatip Taemkaeo. YOLOv5 Facemask., 2021. [Online]. Available: https://www.kaggle.com/chinnatiptaemkaeo/yolov5-facemask-2021.

[29] TensorFlow Community. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. [Online]. Available:https://www.tensorflow.org/.

[30] Adam Paszke, Sam Gross, Soumith Chintala, and Gregory Chanan. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration. *PyTorch: Tensors and Dynamic Neural Networks in Python With Strong GPU Acceleration*, 6, 2017.

[31] François Chollet et al. Keras, 2015. [Online]. Available: https://keras.io.

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *In Proceedings of the IEEE International Conference on Learning Representations (ICLR)*, May 2015.

[33] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

[34] Express News Service. Requirements and future of smart cities, 2014. [Online]. Available: https://www.expresscomputer.in/columns/icts-and-smart-cities/2434/.

[35] Dirk Oberhagemann. Static and dynamic crowd densities at major public events. *Vereinigung zur FÄrderung des Deutschen Brandschutzes e. V.(vfdb), German Fire Protection Association, Technical-Scientific Advisory Board (TWB), Department*, 13:13–01, 2012.

[36] Sohei Kojima, Akira Uchiyama, Masumi Shirakawa, Akihito Hiromori, Hirozumi Yamaguchi, and Teruo Higashino. Crowd and event detection by fusion of camera images and micro blogs. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 213–218. IEEE, 2017.

[37] Pritikana Das, M Parida, and VK Katiyar. Analysis of interrelationship between pedestrian flow parameters using artificial neural network. *Journal of Modern Transportation*, 23(4):298–309, 2015.

[38] Jiang Liu, Chenqiang Gao, Deyu Meng, and Alexander G Hauptmann. Decidenet: Counting varying density crowds through attention guided detection and density estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, UT, USA, 2018. IEEE.

[39] Tao Zhao, Ram Nevatia, and Bo Wu. Segmentation and tracking of multiple humans in crowded environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1198–1211, 2008.

[40] Yu-Jen Ma, Hong-Han Shuai, and Wen-Huang Cheng. Spatiotemporal dilated convolution with uncertain matching for video-based crowd estimation. *IEEE Transactions on Multimedia*, 2021.

[41] Jing Shao, Chen-Change Loy, Kai Kang, and Xiaogang Wang. Slicing convolutional neural network for crowd video understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5620–5628, Las Vegas, NV, USA, 2016. IEEE.

[42] HY Swathi, G Shivakumar, and HS Mohana. Crowd behavior analysis: a survey. In *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, pages 169–178, Bangalore, India, 2017. IEEE, IEEE.

[43] Lingbo Liu, Zhilin Qiu, Guanbin Li, Shufan Liu, Wanli Ouyang, and Liang Lin. Crowd counting with deep structured scale integration network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1774–1783, Seoul, Korea (South), Korea (South), 2019. IEEE.

[44] Deepak Babu Sam, Shiv Surya, and R Venkatesh Babu. Switching convolutional neural network for crowd counting. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4031–4039, Honolulu, HI, USA, 2017. IEEE.

[45] Vishwanath A Sindagi and Vishal M Patel. Generating high-quality crowd density maps using contextual pyramid cnns. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1879–1888, Venice, Italy, 2017. IEEE.

[46] Cong Zhang, Hongsheng Li, Xiaogang Wang, and Xiaokang Yang. Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 833–841, Boston, MA, USA, 2015. IEEE.

[47] Lu Zhang, Miaojing Shi, and Qiaobo Chen. Crowd counting via scale-adaptive convolutional neural network. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1113–1121, Lake Tahoe, NV, USA, 2018. IEEE.

[48] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, and Benjamin J Cowling. Rational use of face masks in the covid-19 pandemic. *The Lancet Respiratory Medicine*, 8(5):434–436, 2020.

[49] R.C. Reiner IHME COVID-19 Forecasting Team. and et al. R.M. Barber. Modeling covid-19 scenarios for the united states. In *Nature*, volume 27, pages 94–â105, 2021. [Online]. Available: https://doi.org/10.1038/s41591-020-1132-9.

[50] World Health Organization (WHO). Who coronavirus disease (covid-19) dashboard, 2021. [Online]. Available: https://covid19.who.int/.

[51] Centers for Disease Control and Prevention (CDC). How to protect yourself others, 2021. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.

[52] Sonali Agarwal, Narinder Singh Punn, Sanjay Kumar Sonbhadra, P Nagabhushan, KK Pandian, and Praveer Saxena. Unleashing the power of disruptive and emerging technologies amid covid 2019: A detailed review. *arXiv Preprint arXiv:2005.11507*, 2020.

[53] World Health Organization (WHO). Contact tracing in the context of covid-19, 2021. [Online]. Available: https://apps.who.int/iris/handle/10665/332049.

[54] Chandrika. Deb. Face Mask Detection, 2021. [Online]. Available: https://dev.to/chandrikadeb7/face-mask-detection-my-major-project-3fj3.

[55] Mei L Ngan, Patrick J Grother, and Kayee K Hanaoka. Ongoing face recognition vendor test (frvt) part 6a: Face recognition accuracy with masks using pre-covid-19 algorithms. *NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology*, 2020. [Online]. Available: https://doi.org/10.6028/NIST.IR.8311.

[56] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas Spanos. Freecount: Device-free crowd counting with commodity wifi. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[57] Cristian Chilipirea, Andreea-Cristina Petre, Ciprian Dobre, and Maarten van Steen. Presumably simple: Monitoring crowds using wifi. In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, volume 1, pages 220–225. IEEE, 2016.

[58] Nuno Nunes, Miguel Ribeiro, Catia Prandi, and Valentina Nisi. Beanstalk: a community based passive wi-fi tracking system for analyzing tourism dynamics. *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems-EICS 17*, 2017.

[59] K. Li, C. Yuen, S. S. Kanhere, K. Hu, W. Zhang, F. Jiang, and X. Liu. Understanding crowd density with a smartphone sensing system. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pages 517–522, Feb 2018.

[60] PH Patil and AA Kokil. Wifipi-tracking at mass events. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–4. IEEE, 2015.

[61] Mohammad Yamin and Yasser Ades. Crowd management with rfid and wireless technologies. In *2009 First International Conference on Networks & Communications*, pages 439–442. IEEE, 2009.

[62] Abdullah Hussein Al-Hashedi, Muhammad Rafie Mohd Arshad, Hasimah Hj Mohamed, and Ahmad Suhaimi Baharuddin. Identifying the determinants of rfid adoption intention in hajj organizations. In *Research and Innovation in Information Systems (ICRIIS), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[63] Ricardo O Mitchell, Hammad Rashid, Fakir Dawood, and Ali AlKhalidi. Hajj crowd management and navigation system: People tracking and location based services via integrated mobile and rfid systems. In *2013 International Conference on Computer Applications Technology (ICCAT)*, pages 1–7. IEEE, 2013.

[64] Mandeep Kaur, Manjeet Sandhu, Neeraj Mohan, and Parvinder S Sandhu. Rfid technology principles, advantages, limitations & its applications. *International Journal of Computer and Electrical Engineering*, 3(1):151, 2011.

[65] Ulf Blanke, Gerhard Troster, Tobias Franke, and Paul Lukowicz. Capturing crowd dynamics at large scale events using participatory gps-localization. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–7. IEEE, 2014.

[66] Anas Basalamah. Sensing the crowds using bluetooth low energy tags. *IEEE Access*, 4:4225–4233, 2016.

[67] Jens Weppner and Paul Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In *Pervasive Computing and Communications (PerCom), 2013 IEEE International Conference On*, pages 193–200. IEEE, 2013.

[68] Jakub Neburka, Zdenek Tlamsa, Vlastimil Benes, Ladislav Polak, Ondrej Kaller, Libor Bolecek, Jiri Sebesta, and Tomas Kratochvil. Study of the performance of rssi based bluetooth smart indoor positioning. In *Radioelektronika (RADIOELEKTRONIKA), 2016 26th International Conference*, pages 121–125. IEEE, 2016.

[69] Alfredo Alessandrini, Ciro Gioia, Francesco Sermi, Ioannis Sofos, Dario Tarchi, and Michele Vespe. Wifi positioning and big data to monitor flows of people on a wide scale. In *Navigation Conference (ENC), 2017 European*, pages 322–328. IEEE, 2017.

[70] Mohamad Omar Al Kalaa, Walid Balid, Naim Bitar, and Hazem H Refai. Evaluating bluetooth low energy in realistic wireless environments. In *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*, pages 1–6. IEEE, 2016.

[71] Fang-Jing Wu and Gürkan Solmaz. Are you in the line? rssi-based queue detection in crowds. In *Communications (ICC), 2017 IEEE International Conference on*, pages 1–7. IEEE, 2017.

[72] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detection-by-tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[73] Mamoona Shami, Salman Maqbool, Hasan Sajid, Yasar Ayaz, and Sen-Ching Samson Cheung. People counting in dense crowd images using sparse head detections. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[74] Alexandre Alahi, Judson Wilson, Li Fei-Fei, and Silvio Savarese. Unsupervised camera localization in crowded spaces. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 2666–2673. IEEE, 2017.

[75] Chen Change Loy, Ke Chen, Shaogang Gong, and Tao Xiang. Crowd counting and profiling: Methodology and evaluation. In *Modeling, Simulation and Visual Analysis of Crowds*, pages 347–382. Springer, 2013.

[76] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, 2011.

[77] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893, San Diego, CA, USA, 2005. IEEE.

[78] Paul Viola and Michael J Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. Springer.

[79] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2009.

[80] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, Tampa, FL, USA, 2008. IEEE.

[81] Ke Chen, Chen Change Loy, Shaogang Gong, and Tony Xiang. Feature mining for localised crowd counting. In *BMVC*, volume 1, page 3, 2012.

[82] Antoni B Chan and Nuno Vasconcelos. Bayesian poisson regression for crowd counting. In *2009 IEEE 12th International Conference on Computer Vision*, pages 545–551, Kyoto, Japan, 2009. IEEE.

[83] David Ryan, Simon Denman, Clinton Fookes, and Sridha Sridharan. Crowd counting using multiple local features. In *2009 Digital Image Computing: Techniques and Applications*, pages 81–88, Melbourne, VIC, Australia, 2009. IEEE.

[84] Victor Lempitsky and Andrew Zisserman. Learning to count objects in images. In *Advances in Neural Information Processing Systems*, pages 1324–1332, Red Hook, NY, 2010.

[85] Viet-Quoc Pham, Tatsuo Kozakaya, Osamu Yamaguchi, and Ryuzo Okada. Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3253–3261, Santiago, Chile, 2015. IEEE.

[86] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 589–597, Las Vegas, NV, USA, 2016. IEEE.

[87] Lokesh Boominathan, Srinivas SS Kruthiventi, and R Venkatesh Babu. Crowdnet: A deep convolutional network for dense crowd counting. In *Proceedings of the 24th ACM International Conference on Multimedia*, pages 640–644, Amsterdam, The Netherlands, 2016. Association for Computing Machinery.

[88] Xinkun Cao, Zhipeng Wang, Yanyun Zhao, and Fei Su. Scale aggregation network for accurate and efficient crowd counting. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 757–773, Cham, 2018. Springer.

[89] Yuhong Li, Xiaofan Zhang, and Deming Chen. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1091–1100, Salt Lake City, UT, USA, 2018. IEEE.

[90] Youmei Zhang, Chunluan Zhou, Faliang Chang, Alex C Kot, and Wei Zhang. Attention to head locations for crowd counting. In *International Conference on Image and Graphics*, pages 727–737, Cham, 2019. Springer, Springer.

[91] Weizhe Liu, Mathieu Salzmann, and Pascal Fua. Context-aware crowd counting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5099–5108, Long Beach, CA, USA, 2019. IEEE.

[92] Xinya Chen, Yanrui Bin, Nong Sang, and Changxin Gao. Scale pyramid network for crowd counting. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1941–1950. IEEE, 2019.

[93] Xiaoheng Jiang, Li Zhang, Mingliang Xu, Tianzhu Zhang, Pei Lv, Bing Zhou, Xin Yang, and Yanwei Pang. Attention scaling for crowd counting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4706–4715, 2020.

[94] Bosheng Qin and Dongxiao Li. Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19. 2020.

[95] Md Sabbir Ejaz, Md Rabiul Islam, Md Sifatullah, and Ananya Sarker. Implementation of principal component analysis on masked and non-masked face recognition. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1–5. IEEE, 2019.

[96] Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi. A real time face emotion classification and recognition using deep learning model. In *Journal of Physics: Conference Series*, volume 1432, page 012087. IOP Publishing, 2020.

[97] Chong Li, Rong Wang, Jinze Li, and Linyu Fei. Face detection based on yolov3. In *Recent Trends in Intelligent Computing, Communication and Devices*, pages 277–284. Springer, 2020.

[98] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N Taha, and Nour Eldeen M Khalifa. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. *Measurement*, 167:108288, 2021.

[99] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[100] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448, 2015.

[101] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application. *arXiv Preprint arXiv:2003.09093*, 2020.

[102] Prajna Bhandary. Observations., 2021. [Online]. Available: https://github.com/prajnasb/observations.

[103] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. Maskedface-net–a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144, 2020.

[104] Larxel. Face Mask Detection, 853 Images Belonging to 3 Classes., 2021. [Online]. Available: https://www.kaggle.com/andrewmvd/face-mask-detection.

[105] Biparnak Roy, Subhadip Nandy, Debojit Ghosh, Debarghya Dutta, Pritam Biswas, and Tamodip Das. Moxa: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks. *Transactions of the Indian National Academy of Engineering*, 5(3):509–518, 2020.

[106] Apple. Apple And Google Partner On COVID-19 Contact Tracing Technology, 2021. [Online]. Available: https://nr.apple.com/d2s6h070n0/.

[107] PACT: Private Automatic Contact Tracing, 2021. [Online]. Available: https://pact.mit.edu/.

[108] CoEpi: Community Epidemiology in Action, 2021. [Online]. Available: https://www.coepi.org/.

[109] Private Kit: Safe Paths; Privacy-by-Design Covid19 Solutions using GPS+Bluetooth for Citizens and Public Health Officials, 2021. [Online]. Available: http://safepaths.mit.edu/.

[110] TraceTogether, Safer Together, 2021. [Online]. Available: https://www.tracetogether.gov.sg/.

[111] Contact Tracing by WHO, 2021. [Online]. Available: https://www.who.int/features/qa/contact-tracing/en/.

[112] Samuel Altmann, Luke Milsom, Hannah Zillessen, Raffaele Blasone, Frederic Gerdon, Ruben Bach, Frauke Kreuter, Daniele Nosenzo, Severine Toussaert, and Johannes Abeler. Acceptability of app-based contact tracing for covid-19: Cross-country survey evidence. *Available at SSRN 3590505*, 2020.

[113] Pozyx". Ultra-wideband RTLS Solutions — Pozyx, 2021. [Online]. Available: https://www.pozyx.io/.

[114] Tsingoal, 2021. [Online]. Available: http://www.tsingoal.com/.

[115] Localino, 2021. [Online]. Available: https://www.localino.net/en/.

[116] Iterate Labs, 2021. [Online]. Available: https://iteratelabs.co/.

[117] Arin, 2021. [Online]. Available: https://teamarin.net/tag/iot/.

[118] Rightcrowd, 2021. [Online]. Available: https://www.rightcrowd.com/social-distancing-monitoring/.

[119] Filip technologies, 2021. [Online]. Available: http://www.myfilip.com/.

[120] Tracking solutions, location based technologies, 2021. [Online]. Available: https://locationbasedtech.com/.

[121] Amber alert gps, 2021. [Online]. Available: https://amberalertgps.com/.

[122] Wonder technology solutions, closed, 2021. [Online]. Available: https://www.crunchbase.com/organization/wonder-technology-solutions.

[123] HereO. Hereo, the first gps watch designed for kids, 2021. [Online]. Available: https://www.hereofamily.com/.

[124] Location of - samsung galaxy grand quattro gt-i8552 real-time gps tracking, 2021. [Online]. Available: http://www.locationof.com/phones/samsung-galaxy-grand-quattro-gt-i8552/.

[125] Masternaut,the smarter fleet platform, 2021. [Online]. Available: https://www.masternaut.com/.

[126] National Association for Child Abduction Prevention Meltzer, Andrew Mitchell. System, method of portable usb key interfaced to computer system for facilitating the recovery and/or identification of a missing person having person's unique identification, biological information, December 6 2005. US Patent 6,973,449.

[127] Sherry Smith. Location system for bluetooth enabled devices, December 7 2010. US Patent 7,848,704.

[128] Don Gunasekara. System and method for displaying time-sensitive alerts in a mobile device, February 10 2015. US Patent 8,954,514.

[129] Daniel G Farley and John R Haynes. Emergency warning system integrated with building hazard alarm notification system, May 11 2010. US Patent 7,714,733.

[130] AA LLC Warner, Chris J. Locality based alert method and apparatus, October 13 2009. US Patent 7,603,432.

[131] Scott A Riggins and David P Sharp. Missing child reporting, tracking and recovery method and system, August 7 2014. US Patent App. 14/059,049.

[132] Charles Byers, Stuart Goldman, Richard Krock, Karl Rauscher, and Lucent Technologies Inc Runyon, James. Method for calling multimedia ip units about an impending emergency situation, March 1 2007. US Patent App. 11/217,014.

[133] GreatCall Inc Gonzalez, Thomas A. Child alert system, August 1 2006. US Patent 7,084,771.

[134] Catalina Marketing Corp Simrell, Roy. Amber alert system and method for locating missing children, July 1 2014. US Patent 8,768,856.

[135] Abi research, tracking device market projection, 2021. [Online]. Available: short-url.at/dmJZ1/.

[136] Asgary. Arcgis web application, yorku.maps.arcgis.com,, 2018. [Online]. Available: https://yorku.maps.arcgis.com/apps/webappviewer/index.html?id=e7c52856187642e19bd227865393432c.

[137] Abdoh Jabbari, Khalid J. Almalki, Baek-Young Choi, and Sejun Song. Ice-mocha: Intelligent crowd engineering using mobility characterization and analytics. *Sensors*, 19(5), 2019.

[138] Robin Heydon. *Bluetooth low energy: the developer's handbook*, volume 1. Prentice Hall Upper Saddle River, NJ, 2013.

[139] Joakim Lindh, Christin Lee, and Marie Hernes. Measuring bluetooth low energy power consumption. *Application Note AN092*, 2017.

[140] Guntur Dharma Putra, Azkario Rizky Pratama, Alexander Lazovik, and Marco Aiello. Comparison of energy consumption in wi-fi and bluetooth communication in a smart building. In *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*, pages 1–6. IEEE, 2017.

[141] Khalid J Almalki, Baek-Young Choi, Yu Chen, and Sejun Song. Characterizing scattered occlusions for effective dense-mode crowd counting. In *Proceedings of*

the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, pages 3840–3849, October 2021.

[142] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916, 2015.

[143] Dan Guo, Kun Li, Zheng-Jun Zha, and Meng Wang. Dadnet: Dilated-attention-deformable convnet for crowd counting. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1823–1832, Nice, France, 2019. Association for Computing Machinery.

[144] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, Honolulu, HI, USA, 2017. IEEE.

[145] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *In Proceedings of the IEEE International Conference on Learning Representations (ICLR)*, 2016.

[146] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[147] Junyu Gao, Qi Wang, and Xuelong Li. Pcc net: Perspective crowd counting via spatial convolutional network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[148] Texas AM University at Galveston. Effectiveness of mask-wearing against COVID-19, 2021. [Online]. Available: https://www.tamug.edu/covid-19/resource-files/print/WearingMask-SelflessService/8-5x11.pdf.

[149] Khalid J Almalki, Muhammad Mohzary, Baek-Young Choi, Sejun Song, and Yu Chen. Mosaic: Modeling safety index in crowd by detecting face masks against covid-19 and beyond. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–7, 2021.

[150] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv Preprint arXiv:1804.02767*, 2018.

[151] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014.

[152] Khalid J Almalki, Sejun Song, Muhammad Mohzary, and Baek-Young Choi. Cats: Crowd-based alert and tracing services for building a safe community cluster against covid-19. In *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 697–701, 2021.

[153] Abhishek Ghosh. rssi distance, 2021. [Online]. Available: shorturl.at/mwDH8/.

[154] Khalid J Almalki, Muhammad Mohzary, Baek-Young Choi, Sejun Song, and Yu Chen. Crowd safety sensing (cross) for the post pandemic era. In *2021 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6, 2021.

[155] Leo Kelion. Coronavirus: UK confirms plan for its own contact tracing app, 2021. [Online]. Available: https://www.bbc.com/news/technology-52263244.

VITA

Khalid Jaber Almalki was born on June 1, 1991, in Jazan, Saudi Arabia. He received his MS degree in Computer Science from the University of Missouri-Kansas City, Kansas City, MO, USA, in 2018 and his BS degree in Computer Networking from Jazan University, Jazan, Saudi Arabia in 2014.

He is currently working as a Lecturer at the College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia. In his Interdisciplinary Ph.D. at the University of Missouri-Kansas City, Kansas City, MO, USA, he specialized in Telecommunications and Computer Networking as a discipline and Computer Science as a co-discipline. Mr. Almalki's research interests include machine learning, crowd sciences, smart city, and IoT technologies.

Mr. Almalki published several papers in world-ranked conferences and journals such as International Conference on Computer Vision (ICCV), International Conference on Mobile Systems, Applications, and Services (MobiSys), IEEE Global Communications Conference (GLOBECOM), and IEEE Sensors Journals.

During his Ph.D., Mr. Almalki received the Best Paper Award at IEEE International Smart Cities Conference, 2018. He also earned several grants, including UMKC Graduate Student Travel Grant 2021, Balaji Travel Grant 2021, and IM 2021 Student Travel.