Adversarial Attacks on Artificial Intelligence of Things-based Operational Technologies in Theme Parks

ARTICLE INFO

Keywords:
Artificial Intelligence of Things
Smart Surveillance Systems
Adversarial Attacks
Operational Technologies
Theme Parks Security

ABSTRACT

Theme parks represent a popular, yet vulnerable aspect of life, where large unsuspecting crowds gather and interact with technology. Artificial intelligence (AI), computer vision, and the Internet of Things (IoT) are transforming theme parks by revolutionizing various aspects. This research study is the first to identify critical components of theme parks that can be optimized, and comprehensively maps them onto emerging AI/IoT applications, often powered by machine learning or deep learning models. Additionally, the study sheds light on adversarial attacks targeting vulnerable smart surveillance systems, which generate a very large volume of video stream data. These systems serve as a prominent example of AIoT-based operational technologies (AIoT-OT) responsible for critical alerts and actions. Rigorous experimentation, involving a novel hybrid multi-pixel deception attack technique, demonstrates that advanced adversarial attack methods can significantly degrade the performance of detection systems. The performance metrics and attack success rate were measured by accuracy, precision, recall, F1-score, and AUC score. Before attack, the accuracy rates of 87. 45%, 83. 17% and 81. 40% were achieved for the EfficientNet, ResNet and MobileNet models, respectively. However, after applying the proposed MPD attack, the performance of each model declined significantly. The accuracy dropped to 61.23% for EfficientNet (with an attack success rate of 29.10%), 59.12% for ResNet (with success rate of 30.20%), and 55.17% for MobileNet (with success rate of 32.60%). This study signifies the need for a strategic plan of action and the development of robust methods for the proactive security of AIoT in theme parks.

1. Introduction

The diversity of the entertainment industry brings new and exciting avenues for all age groups. One of the most luring and thrilling adventures is theme parks. A range of fascinating, memorable, and immersive experiences encourage people to plan and book attractions bringing massive business opportunities and benefits to the investors. To maintain this hype, the industry competitors look for unique ideas, contemporary designs, state-of-the-art mechanics, and advanced technological solutions that can help improve the guest experience and enable theme park leaders to make better-informed business decisions. The quest to be the best has inspired the enablers to not only incorporate the latest existing innovations but also develop customized solutions that can process the classified nature of the physical, operational, technological, and people's information for operational success. These integrated solutions can be based on the Internet of Things (IoT), Artificial Intelligence (AI), Computer Vision (CV), and/or Cloud Computing (CC) along with the conventional technical corporate components. However, this integration brings a challenge of interoperability of diverse systems of assorted transformative technologies. Hence, the threat landscape of these interconnected components opens several opportunities for diverse attack vectors to evade/ poison the decision-making systems Habbal et al. (2024). This attack surface can range from identity management and access control systems, device memory, physical/mobile/cloud/web and/or administrative interfaces, device firmware, network services, data storage/processing centres, and most importantly data analytics algorithms/models.

IoT is not a new concept for theme parks, as it has been purposely utilized to analyse critical data for real-time decision-making. With the availability of state-of-the-art fast and low-cost IoT hardware, interactive software applications, augmented network capabilities, low-power wireless communication technologies, and advanced cloud computing services offering robust data analytics, the development of the IoT ecosystem has become indispensable Chithaluru et al. (2023). This paradigm of pervasive computing environment allows interconnecting smart objects including physical devices, vehicles, buildings, wearable gadgets, and other embedded electronics to communicate and exchange data using applications and network connectivity. Enabled technologies such as radio frequency identification (RFID), wireless sensor networks (WSN), IPv6, and wireless personal area networks (WPAN) establish the seamless integration of these smart components creating holistic cyber-physical systems (CPS) that collect, aggregate, exchange, and analyze data to extract useful information for intelligent informed decisions, better experience, and business enhancement. A comprehensive implementation of an IoT ecosystem (Internet of Everything - IoE) in theme parks ensures monitoring and controlling of the rides and autonomous vehicles, energy consumption, physical security, resource/ asset management, real-time decision-making, timely emergency responses, lost and found tracking, queues/ crowd management, incident warnings, and many other specific tasks with AI-enabled IoT (AIoT). The massive infrastructure of such CPS carries out machine-to-machine-cloud-application-human computations transferring and converting petabytes of data into

ORCID(s):

meaningful information. The operational success and reliability of the participating systems build the trust of consumers resulting in attraction/ business popularity. However, this fancy ecosystem is only desirable when a complete risk assessment and treatment plan are in place to counter adversarial behaviours and actions.

With the use of a combination of AIoT-based operational technologies (AIoT-OT), such as smart surveillance systems reliant on video data streams, it is vital to carefully address the distinct vulnerabilities that each of these environments can present. In particular, the exponential growth of IoT-based operation technologies (IoT-OT) has raised alarms about insecure device design by manufacturing vendors, interoperability challenges, firmware security updates, data analytics privacy, and exploitation of the communication network, resulting in the propagation of attacks disrupting CPS operation. The adversarial access to the systems can open doors to computational models that translate raw data into meaningful information for making informed decisions, leading to adversarial machine learning (AML) attacks. The principal abuse of AIoT-OT devices includes a coordinated attack against the CPS due to supply chain compromise, eroding dataset integrity, poisoning training data, publishing poisoned models, collecting and exfiltration of model artefacts, spamming models with Chaff data, and evading decision models resulting in a denial of reliable decision service. The proposed novel research study attempted to answer the following research questions:

- What are the potential AIoT-OT transformative technologies in theme parks?
- Which AIoT-OT operational technologies may have common machine/ deep learning models?
- What tactics, techniques, and procedures could be used for an AML attack on the AIoT-OT surveillance of theme parks?
- How the AIoT-OT efficiency can be affected by a successful AML attack?

Several users from diverse communities utilise distinct services/ facilities offered by the public-facing entertainment industries generating huge amounts of sensitive and personal information. It is critically important to acknowledge the susceptibility of this infrastructure and rigorously investigate associated risks, attack surfaces, attack vectors, security and privacy controls, and alignment with consumer, enterprise and industrial practices, regulations, and specifications. The researchers of this study rigorously looked into the potential assets/ components that could exist in theme parks, explored the autonomous operational flow of the connected assets, and adversarial behaviours and actions that could disrupt the functioning of systems and resilience to attacks. This study presents a proactive and strategic effort to explore the potentially secure use of AI and IoT in key operational activities of theme parks. It is a novel and rigorous investigation that brings attention to various components of theme parks that can be optimized through the transformative capabilities of AI and IoT. Additionally, the study examines the adversarial challenges associated with incorporating these technologies by devising attacks (AML) on critical features of decision-making systems reliant on video streams. By demonstrating successful performance degradation in smart visual monitoring systems, the study underscores the importance of the secure adoption of AIoT for operational activities in theme parks.

The significant contributions of this study are as follows:

- In-depth analysis of the operational landscape for smart theme parks.
- Mapping of theme park components onto the potential AIoT-OT applications.
- Rigorous experimentation on the UCF-Crime dataset leveraging a vision-language model for automatic frame trimming, dynamic sampling for balanced data selection, and training compute-efficient deep architectures.
- Development of novel AML attacks against a critical AIoT-OT application aiming to gauge deviation from normal functioning.

The structure of the research paper is designed to introduce the research problem and highlight the study's contribution to the field of AIoT-OT security. Section 2 provides a comprehensive review of significant and relevant research studies. Section 3 explains the datasets, methods, tools, and techniques used to implement the adversarial attacks. Section 4 outlines the experimental setup and presents the results of each devised attack, emphasizing their impact on performance degradation. Section 5 discusses the limitations of the study, while Section 6 offers suggestions for future research. Finally, the paper concludes in Section 7.

2. Related Studies

In recent years, the integration of transformative technologies has substantially revolutionized entertainment, recreation, and amusement public parks, providing creative solutions to improve guest experiences, maximize operational efficiency, and provide a safe and pleasurable environment. This literature review investigates the developing convergence of public places such as theme parks with AIoT-OT applications, looking at how different AI, IoT, and other IT/ OT technologies can be placed, and exploited with adversarial attacks.

2.1. Potential Operational Technologies (OT) for Theme Parks

In an interesting recent study, Asaithambi et al. (2023) researchers utilized big data analytics by using multimodal data from users including text mining and data mining from destination images, tourist activities, and weather forecasts to develop a hybrid system for thematic travel recommendations. The study utilised various ML algorithms such as naive bayes, CNNs, and sentiment analysis algorithms to develop a data-driven hybrid recommendation system. The system was capable of learning user preferences for personalised recommendations. The outcomes of this research study can be explored in the context of customer preferences for travel planning to theme parks. Rezapouraghdam et al. (2023) employed machine learning to predict the visitors' behaviour in marine protected areas. The study utilized a fuzzy set qualitative comparative analysis and neuro-fuzzy inference system to train and test visitors' behaviour. The outcomes of the study (information base) can be helpful for managers and policymakers to derive the content that can be helpful for the prospective environment. Utilizing ML-based algorithms can help in predicting and analyzing the complex behaviours of tourists and could be useful for decision-makers in theme parks as well. Joung and Kim (2023) developed an interpretable ML-based approach for customer segmentation. They specifically focused on customer segmentation for new product development to identify the opportunities in new product concepts. Such applications can be applied to theme parks where new services/ facilities might be introduced to the general public. For this study, Word2Vec was used to generate embeddings, and various ML-based algorithms such as decision trees, random forests, light gradient boosting, XGboost, and neural networks were used for segmentation. The researchers utilized the explainable AI method named shapley additive explanations (SHAP) to explain the predictions of ML algorithms.

Although the ticket management system can restrict the number of customers per day in a theme park, yet, crowd gathering at one spot within a specific location can be challenging to deal with. A research study Zhang et al. (2023) introduced an industrial-grade IoT system for crowd safety in coal mines. Unlike traditional LSTM-based approaches, this system was developed to improve the body movement and pressure values prediction using a hybrid absolute degree of incidence and LSTM network. This helped in utilizing the spatio-temporal features for crowd management. The system reduced the average cumulative prediction error and helped in workers' safety in coal mines. In a similar study, Zhu et al. (2023) focused on crowd analytics in railway stations using artificial intelligence. By incorporating a generalised AI framework, the researchers proposed crowd analytics by analyzing visual crowd data from video records. YOLO and Deep SORT were utilized to calculate flow volume, crowd density, and walking speed of people in the crowd. The proposed system was able to count pedestrians with a 95% accuracy. The proposed methodology can be applied to monitor crowds in dense public areas and inform decision-makers about any suspicious activities. Liao et al. (2023) worked on another crowd management task by formulating a fence layout problem as an optimization problem. The researchers proposed a bio-inspired ant colony-based approach paired with a congestion probability social force model (CP-SFM) for irrational pedestrians simulation. In the first phase, CP-SFM was utilized and then an ant colony crowd intervention algorithm was used to optimize the fence layout. The method was tested on 18 scenes in two railway stations.

2.2. Potential IoT-OT Deployment in Theme Parks

Gao et al. (2023) focused on the analysis of the intersection of IoT and computer vision for crowd counting problems that was helpful in crowd management in public places. The study comprehensively provided an overview of computer vision and IoT-based techniques for crowd counting and pattern recognition that can be helpful in various commercial, medical, and surveillance applications. The focus of this research was on IoT environments and computer vision-based approaches such as object detection. The reviewed studies suggest that undoubtedly crowd management can be a critical issue in peak seasons and emergencies in theme parks and state-of-the-art AIoT-OT-based solutions can be adopted to counter the stampede threats.

Recently, Jarašūnienė et al. (2023) conducted a thorough investigation on the impact of IoT on warehouse management problems. The researchers argued that the use of IoT technology that can process large amounts of data is helpful for various operations involved in warehouse management. Additionally, the data collected can be used for informed decisions about inventory and automate various tedious processes. Salazar et al. (2023) worked on an interesting use case for inventory management using unmanned autonomous vehicles (UAVs) in smart city environments. According to the researchers, in large warehouses and inventories, manual management is challenging so UAVs were used to achieve the research goals. Researchers proposed a supervisory control and data acquisition (SCADA) system for inventory management of a well-known DHL Company. The proposed research can be integrated with any kind of inventory management in theme parks. In places such as theme/ amusement parks, there can be plenty of useful applications of warehouse management systems. The utilization can be made more useful with the integration of IoT and artificial intelligence.

Ramzan et al. (2023) developed a multi-objective approach for radio resource management in energy harvesting cooperative UAV-guided IoT networks. Researchers specifically worked on a cooperative cognitive IoT (CIoT) using various IoT devices where UAVs were used as relays. Researchers developed a multi-objective optimization method for the joint optimization of power systems, CIoT devices, and UAVs. Mixed Integer Non-Linear Programming (MINLP) and Outer Approximation Algorithm (OAA) were proposed to achieve optimization goals. The utilization of such a system in theme parks can bring effective business outcomes. Autonomous vehicles are well-known for the deployment of IoT components. The potential benefits of similar implementations are readily explored in industrial automation, robotics, theme park rides and trail health checks. There is a huge scope of extensive research in the domain of theme parks utilizing AIoT for operational technologies.

2.3. Potential AIoT-OT Implementation in Theme Parks

In another research study, Chithaluru et al. (2023) proposed an AI-based clustering method for green IoT in smart cities. The researchers tackled the problem with traditional IoT systems where monitoring, recognizing, and management of resources was challenging. They used a dynamic self-organizing neural network model to overcome these challenges. The experiments proved that neuro-fuzzy methods are effective for the sustainability of IoT devices in smart cities. The inferences from these studies can open new horizons for resource management in AIoT-OT-based theme parks.

Forkan et al. (2023) worked on the intersection of AI and IoT for road infrastructure management using a large city-scale sensing technique. The researchers highlighted the importance of roadside management and maintenance for smart cities which can be directly applied to large amusement parks as well. The proposed method AIoT-CitySense achieved an impressive 85% maintenance improvement in local municipal government in Australia. The solution was able to successfully detect path holes and line markings for pedestrians. This city-scale application of AI and IoT proved the effectiveness of data-driven infrastructure maintenance projects.

Chen et al. (2023) designed and implemented an AIoT-based autonomous mobile robot for cleaning garbage. This system has extensive applications in cleaning public places and can save time and cost for managing garbage. The entire system has an autonomous mobile robot (AMR), AI-based garbage recognition, and solar power with a mounted trash can to collect and manage garbage. The garbage recognition was done using a vision-based algorithm by processing images from the camera sensor mounted on a robot. The final system was able to automatically collect garbage without human supervision. Ijemaru et al. (2023) worked on an interesting concept of the Internet of Vehicles (IoV) for data collection and traffic engineering for waste management applications in smart cities. The researchers proposed a swarm intelligence-based IoV system for data collection for waste management. In addition, an analytics-based system was proposed for finding an optimal number of data collector vehicles and an optimal number of data collection points. The experiments were done via simulation using the proposed system to prove the efficacy of the innovative approach.

The AIoT-OT literature research shows that there are limited or near to non-existent studies that have explicitly explored the core components of theme parks where AIoT-OT can be beneficially and securely deployed. With the review of contemporary related studies, the researchers of this study aimed to reveal that theme parks have great potential for the strategic application/ placement of IoT devices generating real-time data from each entity residing in the infrastructure. To automate the operation technologies for decision-making, visualization, and predictions from the collected data, several AI models can play critical roles in enhancing user experience, security posture, and business revenue. However, the security of AI is critical in a massively integrated operational infrastructure.

Table 1 summarizes the mapping of significant theme park components onto the prospective AIoT-OT applications. The table also signifies the use of ML and DL models adopted by recent studies that can be common across different components of the theme park.

2.4. Theme Parks and Need for AIoT-OT Security

While it is important to perceive potential applications of AI, IoT, and OT in the entertainment industry, researchers have also emphasized diverse security mechanisms to meet the requirements of confidentiality, availability, integrity, authentication, non-repudiation, accountability, and privacy of the end-user devices, applications, network, servers, datastores, and backups in the eco-system Li et al. (2022). Asset inventory and control, identity and access management (authentication, authorization), continuous vulnerability management (with automated secure firmware updates), periodic configuration validations, cryptographic controls, secure configurations of wired and wireless network infrastructure, log keeping and auditing, etc., are some essential components to ensure the implementation of a zero-trust architecture for the AIoT-OT enabled secure theme parks. Habbal et al. (2024) stressed that it is essential to comprehend the critical needs of transformative technologies through the lenses of trust, risk, and security management regulations. In a recent study, researchers proposed the AI TRiSM framework which detailed several aspects of model monitoring, operations, AI application security, and privacy. The research also highlighted challenges and gaps in the reliable implementation of AI systems in smart use cases. The application of AI for security and security of AI, both notions are equally crucial in the critical infrastructures where all entities (systems, users, networks, applications) are of utmost concern Saied et al. (2024). Sánchez et al. (2024) utilized LSTM-CNN architecture for hardware-based device identification and compared the proposed architecture with other ML/DL techniques for time series. Researchers analyzed the model's robustness against potential context-based and ML/DL adversarial attacks highlighting the differing impacts of both attacks. This research also applied defence mechanisms to strengthen model resilience against evasion attacks.

Many critical territories of the theme parks rely on video streams that illustrate real-time happenings and demand prompt response in the event of escalating situations. Perturbations in these streams can benefit adversaries in carrying out diverse actions of disruptions. One of the crucial objectives of this study is the development and deployment of AML attacks against a critical AIoT-OT application, a smart surveillance system, aiming to gauge deviation from normal functioning. Achieving this objective will help the

Table 1
Mapping of Theme Park components with AI/ IoT Applications

Ref.	Potential Theme Park Components/ OT	AI/ IoT Applications	ML/DL Models
Gupta et al. (2023)	Entry and Visitor Management	RFID/NFC-enabled Wristbands, Facial Recognition Technology	CNN, SVM
Prandi et al. (2023)	Navigation and Wayfinding	Mobile App with Location Services, Bluetooth Beacons	CNN
Sheela et al. (2023)	Rides and Attractions	Predictive Maintenance Sensors, IoT-connected Queue Management	RL
Doğan and Niyet (2024)	Guest Services and Personalization	ML for Personalized Recommendations, Chatbots and Virtual Assistants	RS, NLP
Haghani et al. (2023)	Crowd Management and Safety	Crowd Density Sensors, Video Analytics and Surveillance	CNN, Anomaly
Bibri and Jagatheesaperumal (2023)	Entertainment and Shows	AR/VR Experiences, IoT-connected Stage Effects	RL
Chandan et al. (2023)	Food and Merchandise	IoT Inventory Management, Smart Payment Systems	RNN, GA
Pandiyan et al. (2023)	Energy and Resource Management	Smart Grid Technology, IoT Environmental Sensors	RL
Himeur et al. (2023)	Feedback and Analytics	IoT-based Surveys and Feedback Kiosks, Big Data Analytics	NLP, NN, RL, Ensemble
Arshi and Mondal (2023)	Parking and Transportation	IoT Parking Sensors, AI-Powered Traffic Management	HMM
Maleki Varnosfaderani et al. (2024)	Health and Safety	Health Monitoring Wearables, AI-Driven Emergency Response Systems	NN, Transformer
Alahi et al. (2023)	Environmental Conservation	IoT Environmental Sensors, AI-Powered Waste Management	NLP, CNN, RL, GA
Chengoden et al. (2023)	Accessibility and Inclusivity	IoT-Enabled Audio Guides, IoT-Assisted Accessibility Services	NLP, CNN, RL, XAI
Niksirat et al. (2024)	Lost and Found Tracking	IoT Tracking Tags for Locating Lost Belongings	SVM, RF
Anton Clavé et al. (2023)	Virtual Queuing Systems	IoT-Powered Virtual Queue Management	CNN
Lokman et al. (2023)	Maintenance and Cleaning	Predictive Cleaning and Maintenance Robots	ARIMA, LSTM, GA
Doğan and Niyet (2024)	Personalized Character Interactions	AI-Powered Character Interactions, Virtual Concierge Services	RS, NLP
Jo et al. (2024)	Dynamic Pricing and Offers	Dynamic Pricing Algorithms	RL
Ud Din et al. (2023)	Educational Experiences	IoT-Enhanced Educational Tours, IoT-Driven Storyline Customization	Adaptive ML, NLP, RL
Pang et al. (2024)	Smart Restroom Management	IoT Restroom Sensors (lighting/ humidity/ temperature/ occupancy)	XGBoost, K-Means
Perez et al. (2023)	Enhanced Accessibility Services	IoT-Enabled Audio Guides for Visitors with Visual Impairments	RNN
Murala et al. (2023)	Personalized Health and Wellness Recommendations	AI-Powered Health Tips, Hydration, and Movement Reminders	NLP, NN, RL, Yolo, XAI
Mansour et al. (2023)	Interactive Queue Entertainment	IoT-Connected Queue Displays with Interactive Games	RL, CNN, Ensemble, DT
Saad et al. (2023)	Smart Waste Collection Routing	Optimized Waste Collection Routes using IoT Devices	KNN
Doğan and Niyet (2024)	AI-Powered Dynamic Food Menus	Dynamic Menu Recommendations based on Visitor Preferences	RS, NLP
Patel et al. (2023)	IoT-Based Water Conservation	Water Usage Monitoring for Conservation Efforts	ANN, SVM, RF, CNN
Bibri (2023)	Virtual Character Interactions	IoT and AR/VR Technologies for Virtual Character Meet-ups	CNN, NLP
Battour et al. (2023)	AI-Guided Park Maintenance	AI-Scheduled Maintenance Tasks using Historical Data	ARIMA
Isaia and Michaelides (2023)	IoT-Enhanced Safety Measures	Safety Alert Bracelets for Emergency Situations	SVM, KNN, NN
Mathur and Sinha (2023)	Personalized Spectator Experiences	IoT-Linked Spectator Shows for Interactive Participation	Deep Q-Networks
Lee et al. (2023)	AI-Powered Queue Predictions	Predictive Queue Wait Times using AI Algorithms	CNN
Flavián et al. (2024)	Augmented Reality Scavenger Hunts	AR-Based Scavenger Hunts for Interactive Exploration	AR-RS, Cognitive Model
Errousso et al. (2024)	Smart Parking Assistance	IoT Parking Guidance Systems for Finding Available Spaces	Regression, Ensemble
Del Gallo et al. (2023)	Dynamic Attraction Scheduling	AI-Driven Dynamic Scheduling of Attraction Timings	RL, NN, SVM, DT, GA
Bernardes et al. (2023)	Weather Monitoring and Alerts	IoT Weather Stations for Real-time Weather Updates	ARIMA, Ensemble, LR
Wu and Hao (2023)	Personalized Multimedia Experiences	IoT-Integrated Displays for Personalized Content	RS, NLP
Alahi et al. (2023)	Language Translation Services	IoT/AI-Powered Translation Devices or Apps for Real-time Translation	Transformer, NN, RL, GA
Balducci et al. (2024)	Interactive Ride Storylines	IoT Sensors/Wearables to Influence Ride Story Elements	GenAI

researchers answer the research question, "How the AIoT-OT efficiency can be affected by a successful AML attack", devised for this study. To serve this purpose, a comprehensive analysis of the related AML attack literature is also done to understand state-of-the-art adversarial attacks disrupting visual data streams.

2.4.1. *Adversarial Attacks and Defense of Visual AIoT-OT*

Studies have shown that well-crafted adversarial spatial and temporal perturbations can threaten the robustness of video recognition models. Although most of the work is done on image models, video models have been under study concernedly. In this section, recent work in video models is vigilantly discussed.

In a relevant study, Pony et al. (2021) introduced a video manipulation technique to fool the video classifier by introducing a flickering temporal perturbation attack which is normally unnoticeable by naked human eyes and could be implemented in the real-world setting. The method was first used in the single video and then generalized to the other dataset samples. The researchers used the Kinectics-400 dataset for experiments and proved that this method has a high fooling rate. Li et al. (2021) presented Geometric TRAnsformed Perturbations (Geo-TRAP) for attacking video classification models. The researchers illustrated the efficacy of parameterizing the temporal structure of the search space with geometric transformations to efficiently search for gradients, enabling query-efficient black-box attacks aimed at maximizing the probability of misclassifying the target video. This method employed a geometric transformation method to reduce the search space for effective gradients by searching for a small group of parameters that define these operations. This method had a success rate of 74% on the Jester dataset.

Cao et al. (2023) focused on the style transfer method to fool the video classifiers. Style transfer is the method in which the content of an image is projected on another to create stylized images. Researchers proposed StyleFool which is a black box video classifier attack method. It utilizes color proximity to select the best-stylized images to avoid unnatural details in resulting videos. Later, a gradient-free method was used to optimize the adversarial perturbations. The method was tested on UCF-101 and HMDB-51 datasets.

Object segmentation has applications in autonomous driving and action recognition which ultimately helps in activity recognition. Li et al. (2023) developed an object-agnostic adversarial attack by first frame attacking using hard region discovery. The gradients from the segmentation model were exploited to identify the confused regions to attack the segmentation model. These regions make it hard to identify the boundaries of the background and objects in the target frames.

In another study, Jiang et al. (2023a) explored the decision-based patch attacks on video classification models. They proposed a spatial-temporal differential evolution (STDE) framework. This method treated target videos as patch textures and added patches on only the most crucial key-frames of the videos that are selected by the temporal difference in an adaptive manner. The optimization goal is to minimize the patch areas without falling into the local minima. Experiments proved the good performance in threat, efficiency, and imperceptibility. The method was tested and validated using UCF-101 and Kinetics-400 datasets.

In another research, Jiang et al. (2023b) studied the sparse video attack and named the framework V-DSA. In this method, the threat model only returned the predicted hard labels. The method consisted of two modules named Cross Model Generator (CMG) and Optical Flow Grouping Evolution (OGE). CMG is used for query-free transfer attacks on individual frames and OGE for query efficient spatio-temporal attacks on target models.

Video compression is a crucial step in IoT devices for efficient storage and transmission of visual data using limited computing resources. Chang et al. (2023) attacked the deep learning-based video compression networks. This method can degrade the spatio-temporal correlation in successive key-frames by injecting flickering-based temporal perturbations. The results, validated through the Vimeo-90K dataset, showed that NetFlick can significantly downgrade the performance of video compressors in the physical and digital worlds.

Cao et al. (2024) proposed a modified StyleFool named as LocalStyleFool. This is an improved black box attack on video models which introduced regional style transfer-based perturbations on input videos to fool the target models. Segment anything (SAM) was used to extract the various regions to capture semantic information and tracking. Later, style transfer-based perturbations were added to selected regions. This resulted in inter-frame and intra-frame naturalness while maintaining the fooling rate. The experiments were performed using UCF-101 and HMDB-51 datasets due to their comprehensiveness and popularity in the research community.

Mu et al. (2024) proposed the sparse attacks on video classification models where fewer frames were perturbed to gain a high fooling rate. Researchers combined the additive and spatial perturbations to enhance attack performance. Structural similarity index (SSIM) was utilised instead of lp normalization for maintaining human perception. In addition, Bayesian Optimization was used to identify the most critical frames for perturbation. These techniques enabled the DeepSAVA to perform very sparse attacks on UCF-101 and HMDB51 datasets.

The video stream-based machine/ deep learning models in the theme parks can extend a new attack surface with increased security risks through the possibility of data manipulation, exploitation, and deformed decisions. The adoption of AI-based operational technologies must contemplate the possibility of potential risks with a strong defense strategy against data sets being corrupted, model theft, and adversarial samples. With this understanding, this research study aims to investigate state-of-the-art adversarial attacks on smart surveillance systems and measure the impact of the tactics, techniques, and procedures of adversaries.

3. Methodology

The researchers selected the UCF-Crime dataset Sultani et al. (2018) and used black box sparse spatial techniques to implement attacks on the video anomaly detection systems. In this section, the methodology is described in detail and an overview of the proposed methodology is presented in Figure 1.

3.1. Datasets

In the literature, various researchers have proposed multiple datasets related to anomaly detection in CCTV videos.

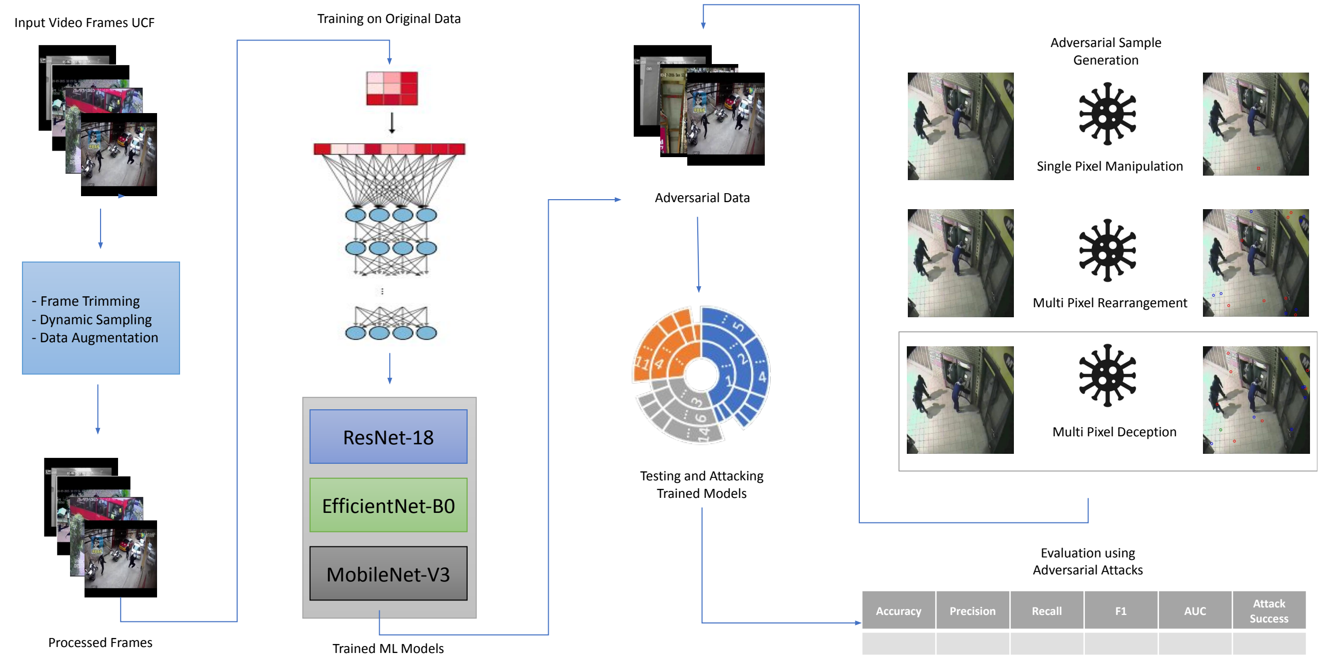


Figure 1: Overview of Proposed Methodology for Adversarial Attacks on AIOT

For a specific use case (theme parks) of this study, some of these datasets are not usable as they do not contain life-threatening scenarios such as explosions and accidents. Some well-known datasets are briefly presented in the following subsections.

3.1.1. UCSD Pedestrian

The University of California San Diego (UCSD) dataset by Mahadevan et al. (2010) is divided into two subsets: Ped1 and Ped2. It focuses on detecting anomalies in crowded scenes, such as people walking on the grass or riding bicycles in pedestrian zones. It consists of surveillance footage captured from a stationary camera overlooking pedestrian walkways on the UCSD campus. The dataset includes both normal pedestrian activity and anomalous events such as bikers, skaters, or people walking in unusual patterns.

3.1.2. ShanghaiTech Campus

ShanghaiTech Campus by Liu et al. (2018) has 13 scenes with complex light conditions and camera angles. It contains 130 abnormal events and over 270,000 training frames. Moreover, the pixel-level ground truth of abnormal events is also annotated in this dataset.

3.1.3. CHUK Avenue

CHUK Avenue Lu et al. (2013) contains 16 training videos and 21 testing videos with a total of 47 abnormal events, including throwing objects, loitering and running. The size of people may change because of the camera position and angle.

3.1.4. Subway

Subway Adam et al. (2008) is two hours long in total. There are two categories, i.e. ’Entrance’ and ’Exit’. Unusual events include walking in the wrong direction and loitering. More importantly, this dataset is recorded in an indoor environment while the above-mentioned are recorded in an outdoor environment.

3.1.5. UCF-Crime

The UCF Crime Sultani et al. (2018) is a large video dataset for anomaly detection in surveillance videos. The dataset has 1,900 untrimmed videos having 13 different classes including fighting, robbery, burglary, arson, and shooting, and normal activities like walking or traffic movements.

With observations, presented in Table 2, the UCF Crime dataset was deemed well-suited for detecting human-related life-threatening events due to its real-world complexity, covering diverse environments with untrimmed surveillance footage. It includes 13 classes of dangerous anomalies such as shootings, assaults, and explosions, which are critical for public safety. Its large scale and support for weakly supervised learning allow for effective training in identifying life-threatening anomalies, making it superior to more controlled or smaller datasets. Figure 2 shows input frames from the UCF-Crime dataset. Additionally, the dataset’s diversity makes it ideal for developing and evaluating adversarial attacks on smart surveillance systems in crowded public places such as theme parks, as the trained models can better

Table 2
Summary of Video Anomaly Detection Datasets

Dataset	Videos/Frames	Classes	Limitations
CHUK Avenue	37 (16 training, 21 testing)	2	Limited size, indoor setting, challenges with varying camera angles and perspectives
ShanghaiTech Campus	13 scenes (270,000 frames)	2	Complex lighting, camera angles, large training frames but limited scene variety
UCSD Pedestrian	2 subsets (Ped1 & Ped2)	2	Low-resolution, gray-scale, focused on pedestrian areas only
Subway	2 hours (Entrance, Exit)	2	Indoor environment only, focused on unusual behaviours like walking in wrong directions, limited scene diversity
UCF-Crime	1900	14	Highly imbalanced between normal and abnormal events, varying video lengths, camera angles, and limited temporal resolution

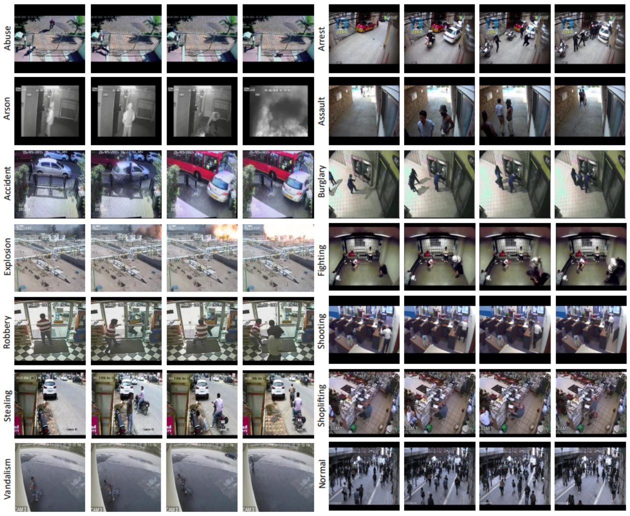


Figure 2: Sample video frames from UCF-Crime dataset

simulate real-world conditions and vulnerabilities, helping to assess the robustness of systems against malicious attacks.

3.2. Data Preprocessing

As discussed, the UCF-Crime dataset has some limitations and to address this, the researchers preprocessed the data by trimming videos and resizing the frames to a standard resolution to ensure consistency in input dimensions. A dynamic sampling method was employed to select representative frames from the input videos. Equation 1 presents the formula for dynamic sampling method.

Given a sequence of video frames $F = \{f_1, f_2, \dots, f_n\}$, where n is the total number of frames, and a desired number of samples k, the sampling interval I is computed as:

$$I = \max\left(1, \left\lfloor \frac{n}{k} \right\rfloor\right) \tag{1}$$

where $\lfloor \cdot \rfloor$ denotes the floor division operation. This ensures that:

- The interval is always at least 1 frame to prevent frame duplication,
- The sampling maintains approximately uniform temporal distribution across the video sequence,
- The method automatically adapts to videos of varying lengths while preserving the desired sampling density.

The dynamic interval sampling method provides an efficient approach for selecting representative frames from videos of varying lengths. This method automatically adjusts the sampling rate based on two key factors: the total number of frames in the video and the desired number of samples. Rather than using a fixed sampling rate, the method calculates an appropriate interval between sampled frames that scales with the video length. Importantly, the method guarantees that the interval between selected frames is never less than one frame, preventing redundant sampling while maintaining the video's temporal coherence. This approach offers a balance between computational efficiency and comprehensive video representation.

Given the highly imbalanced nature of the dataset, the original protocol, proposed by the authors, was followed by grouping all abnormal categories into a single "anomaly" category, resulting in a binary classification problem. For video trimming, this study utilized MoonDream Vikhyat (2024) a lightweight, open-source vision-language model combined with manual inspection. Frames were processed using a fixed prompt: "Valid image or noise" and any frame identified as noise was removed to create a cleaner, more balanced dataset. Additionally, various geometric data augmentation techniques, such as zoom, horizontal and vertical flips, and 45° and 90° rotations, were applied to decrease overfitting and improve the model's performance.

3.3. Deep Learning Models

To develop the proposed frame-based video classifiers, the researchers utilized state-of-the-art deep learning architectures for efficient performance. All the selected models were pre-trained on ImageNet Deng et al. (2009), so the researchers only added the new classifiers and trained those classifiers instead of training all the models from scratch. The choice of architecture is based on the fact that in AIoT environments large computing dedicated sources are not available. So, the models should have the capability to function on resource-constrained devices. Therefore, three different compute, parameter and resource-efficient architectures were selected. The researchers selected these based on the top-5 accuracy performance on the ImageNet benchmark dataset. In the following section, these deep architectures are briefly discussed.

3.3.1. EfficientNet-B0

EfficientNet-B0 Tan and Le (2019) is the smallest version of the EfficientNet model family, which is designed using a compound scaling method that uniformly balances network depth, width, and resolution. This approach allows EfficientNet-B0 to achieve high accuracy with fewer parameters and lower computational costs compared to traditional architectures. The model stands out for its ability to perform well on various tasks while being highly efficient in terms of both speed and memory usage. On the ImageNet benchmark, EfficientNet-B0 achieved a top-1 accuracy of 77.1%, making it one of the most efficient models for image classification. Its innovative scaling method and strong performance on standard datasets have led to its wide adoption in environments where computational resources are limited.

3.3.2. ResNet-18

ResNet-18 He et al. (2016) is a member of the Residual Networks (ResNet) family, which introduced the concept of residual learning to address the problem of vanishing gradients in deep neural networks. The key feature of ResNet-18 is the use of residual connections, which allow the network to bypass certain layers and directly pass information forward. This architecture enables deep models to learn more efficiently, even as depth increases. On the ImageNet benchmark, ResNet-18 has achieved impressive performance with a top-1 accuracy of around 69.8%, offering a powerful yet computationally efficient solution for image classification tasks. Its ability to handle deeper architectures while maintaining stable performance makes it a widely adopted model across various vision applications.

3.3.3. MobileNet-V3

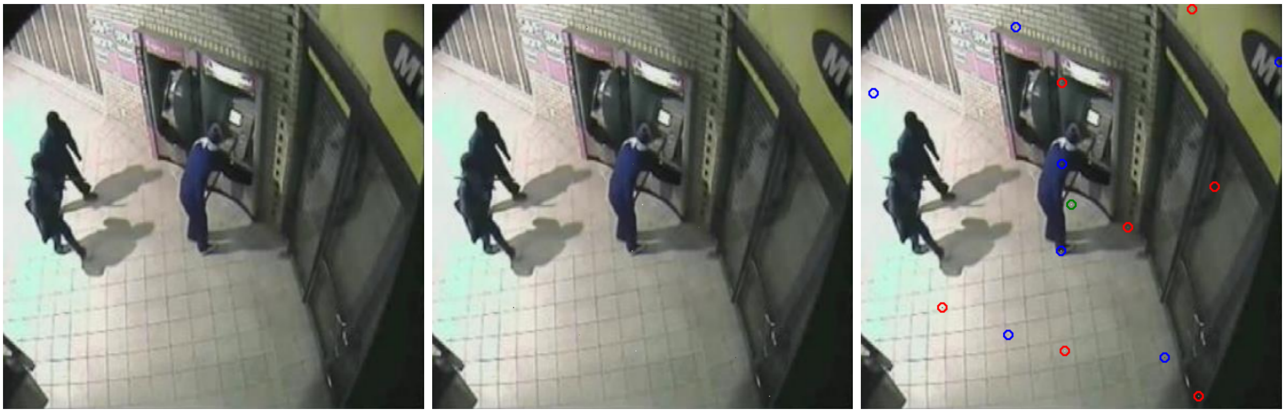
MobileNet-v3 Howard et al. (2019) is part of the MobileNet family. It was designed specifically for mobile and embedded devices where computational resources and power consumption are constrained. It incorporates advances from both MobileNetV2 Sandler et al. (2018) (inverted residuals and linear bottlenecks) and EfficientNet Tan and Le (2019) (neural architecture search) to create a highly optimized architecture. MobileNet-v3 uses a combination of depthwise separable convolutions and squeeze-and-excitation modules to maximize efficiency without significantly compromising accuracy. On the ImageNet benchmark, MobileNet-v3 Small achieved a top-1 accuracy of 67.4%, while maintaining a low computational footprint. Its ability to perform well under constrained conditions makes it a popular choice for real-time applications and devices with limited processing power.

3.4. Adversarial Attacks Implementation

This study implemented the sparse black box spatial adversarial methods to attack the violence detection models. The researchers ensured that sparse spatial perturbations are imperceptible by the human eyes but can successfully fool the state-of-the-art video classifiers. We used Pixle Pomponi et al. (2022) and One Pixel Su et al. (2019) attack as they only modify a few pixels in the input samples. The introduction to these methods is briefly presented in the following sections. A novel strong attack, based on the hybrid Multi Pixel Deception (MPD) attack, that utilizes the attacking powers of Pixel and OnePixel, is proposed. In one pixel, only one pixel is changed while Pixle aims to rearrange the minimum pixels to fool the models. In this approach, the researchers selected the pixels using Differential Evolution Price (2013) and rearranged and changed the target pixels. The results showed that this was a stronger attack which enhanced the attack success rate of the proposed adversarial attacks as compared to the individual spatial attacks. Figure 3 illustrates a single adversarial sample generated using the pixel manipulations of MPD attack while 4 shows a few samples generated using MPD and spatial perturbations are highlighted as well.

3.4.1. One Pixel Attack

The One Pixel attack Su et al. (2019) is an adversarial technique that aims to mislead convolutional neural networks (CNNs) by altering a single pixel in an image. Despite the minimal perturbation, this attack demonstrates the sensitivity of deep learning models to small, strategically placed changes in the input. The One Pixel attack is particularly effective in scenarios where imperceptibility is crucial, as it significantly alters the model's predictions while remaining visually undetectable to human observers. In the context of anomaly detection, this method exploits the network's reliance on local pixel dependencies, revealing vulnerabilities even in robust models. The simplicity of the One Pixel attack, combined with its effectiveness, makes it a valuable tool for evaluating the resilience of CNN-based anomaly detection systems.



(a) Original Input Frame

(b) Adversarial Frame

(c) Highlighted Spatial Perturbations

Figure 3: Adversarial sample generated using MPD. The green circle highlights the pixel manipulated using the One Pixel component of MPD. Red circles represent the position of original pixels while blue circles represent rearranged pixels via MPD.

3.4.2. *Pixle Attack*

Pixle Pomponi et al. (2022) is a targeted adversarial attack that modifies minimal pixel locations within an image to fool deep learning models, particularly convolutional neural networks (CNNs), by making imperceptible changes. Unlike traditional adversarial attacks that manipulate numerous pixels across an image, Pixle focuses on highly localized perturbations, which allows for more efficient attacks with reduced computational overhead. This characteristic is particularly advantageous when attacking anomaly detection models, as Pixle minimizes the likelihood of detection by human observers or automated defences. By exploiting the vulnerability of CNNs to small, concentrated changes, Pixle proves to be a potent adversarial technique in the context of visual anomaly detection.

3.4.3. *Multi Pixel Deception Attack*

The Multi-Pixel Deception (MPD) attack is a novel hybrid approach that combines the strengths of the Pixle and One Pixel attacks to create a more effective adversarial technique. By leveraging minimal perturbations across a small set of carefully selected pixels, this attack maintains the imperceptibility of the One Pixel attack while introducing the multi-pixel disruption capabilities of Pixle. This hybrid method targets convolutional neural networks (CNNs) used in anomaly detection tasks, exploiting their sensitivity to small, localized changes in the input. In experimental evaluations, the Multi-Pixel Deception attack achieved a higher attack success rate compared to individual attacks, demonstrating its effectiveness in misleading the model with minimal distortion to the original image. The combination of efficiency and potency makes this approach a powerful tool for testing the robustness of CNNs against adversarial inputs. Algorithm 1 demonstrates the adversarial sample generation process.

Algorithm 1 Multi-Pixel Deception (MPD) Attack

```
Input: Trained CNN model $f(x)$, Input image $x \in \mathbb{R}^{m \times n \times c}$,
       Number of pixels for modification $k_1 \in \{2, 3\}$, Number of pixels to rearrange $k_2 \in \{4, 6\}$, Perturbation
       bound $\epsilon$, Maximum iterations $T$
Output: Adversarial image $x'$
Initialize $x' \leftarrow x$ Randomly select $k_1$ pixels $P_1 = \{(i_1, j_1), \dots, (i_{k_1}, j_{k_1})\}$ for modifications Randomly select $k_2$ pixels $P_2 = \{(i'_1, j'_1), \dots, (i'_{k_2}, j'_{k_2})\}$ for rearrangements
for $t = 1$ to $T$ do
    /* Pixel Modification Phase */
    for each pixel $(i, j) \in P_1$ do
        Perturb pixel value: $x'[i, j, c] \leftarrow x[i, j, c] + \delta_{i,j,c}$
          where $\delta_{i,j,c} \in [-\epsilon, \epsilon]$
          Clip pixel values: $x'[i, j, c] \leftarrow \text{clip}(x'[i, j, c], 0, 1)$
    end
    /* Pixel Rearrangement Phase */
    for each pixel $(i', j') \in P_2$ do
        Rearrange pixel value: $x'[i', j', c] \leftarrow x[i', j', c] + \delta_{i',j',c}$ where $\delta_{i',j',c} \in [-\epsilon, \epsilon]$
          Clip pixel values: $x'[i', j', c] \leftarrow \text{clip}(x'[i', j', c], 0, 1)$
    end
    Evaluate $x'$ on model $f(x')$ if $\hat{y}' \neq \hat{y}$ then
        return $x'$ as adversarial image
    end
end
return final adversarial image $x'$ if no misclassification occurs after $T$ iterations
```

4. Experimental Setup and Results

In this section, the performance of the CNNs on the UCF-Crime dataset under adversarial attacks is presented. The model's behaviour on the original dataset and their

Table 3
Performance Metrics for Original Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC Score
ResNet-18	83.17	83.12	83.46	83.29	0.831
EfficientNet-B0	87.45	88.21	87.33	87.71	0.873
MobileNet-v3	81.40	82.15	81.36	81.74	0.813

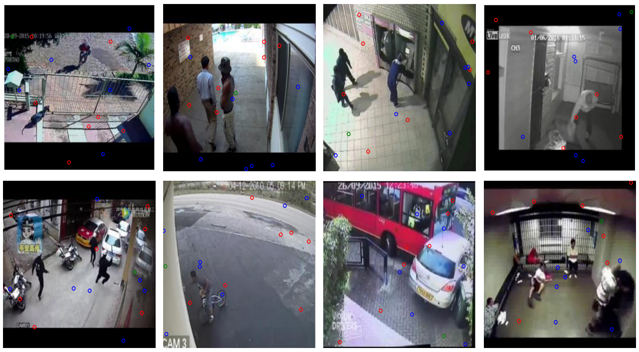


Figure 4: Adversarial Anomaly Samples Generated using MPD

degradation under three adversarial attacks: One Pixel Attack, Pixle Attack, and the proposed Multi-Pixel Deception (MPD) Attack was analyzed.

The experiments were conducted using an NVIDIA A100 GPU, which provides high computational power, allowing for efficient training of deep learning models. The A100, with its large memory capacity and high throughput, enabled us to handle large-scale datasets and complex models with accelerated performance. This study utilized PyTorch, and for adversarial attacks, the torchattacks library which provides a comprehensive collection of attack methods was employed. This library was integrated into the PyTorch pipeline to evaluate the robustness of the models under various adversarial scenarios.

4.1. Performance on Original Dataset

On the original dataset, without any adversarial manipulation, the models exhibited high classification performance, especially EfficientNet-B0, which outperformed ResNet-18 and MobileNet-v3 across all metrics. EfficientNet-B0 achieved the highest accuracy of 87.45%, precision of 88.21%, recall of 87.33%, F1 score of 87.71%, and AUC score of 0.873. These results reflect the model's superior ability to detect anomalies in video frames, attributed to its efficient architecture and feature extraction abilities. Table 3 provides a complete overview of models' performance on the original dataset.

ResNet-18 performed comparably, with an accuracy of 83.17% and an F1 score of 83.29%, making it a strong candidate for lightweight anomaly detection tasks. MobileNet-v3, designed for mobile and edge devices, slightly trailed with an accuracy of 81.40% and an F1 score of 81.74%. Despite these differences, all models demonstrated solid performance under normal conditions, highlighting their suitability for real-time AIoT-OT anomaly detection scenarios where computational resources may be limited.

4.2. Results for One Pixel Attack

Under the One Pixel Attack, the performance of all models dropped significantly, indicating their vulnerability to minimal pixel perturbations. EfficientNet-B0 remained the most robust, with an accuracy of 82.01%, precision of 82.32%, and F1 score of 82.42%. Although there was a noticeable drop in metrics compared to the original dataset, EfficientNet-B0 still maintained decent resilience against this attack. Table 4 presents all metrics for models under one pixel attack.

ResNet-18 and MobileNet-v3 suffered a more substantial decrease in accuracy, with ResNet-18 dropping to 78.17% and MobileNet-v3 to 76.40%. The One Pixel attack's success highlights the sensitivity of CNNs to even the smallest perturbations, especially in high-dimensional input spaces like video data. The degradation in performance across all models showed that adversarial attacks, despite their simplicity, can easily mislead CNN-based anomaly detection systems.

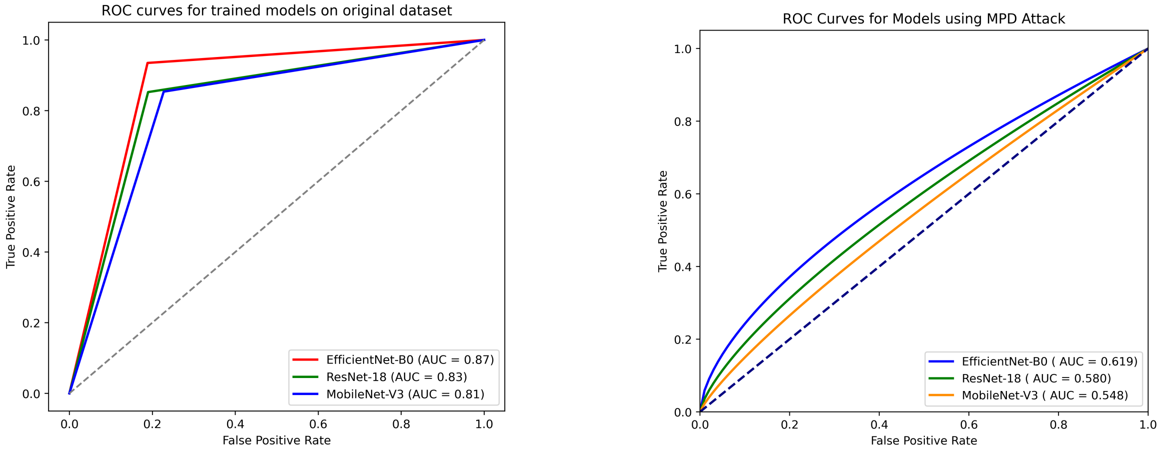
4.3. Results for Pixle Attack

The Pixle Attack, which manipulates multiple minimal pixel locations, had an even more profound impact on the models' performance. ResNet-18's accuracy dropped to 71.21%, and MobileNet-v3's performance fell sharply to 69.12%. EfficientNet-B0, while still leading, had its accuracy reduced to 76.12%, highlighting the effectiveness of this targeted attack in misguiding CNNs. Table 5 provides detailed performance metrics.

The Pixle attack exploits the models' reliance on localized features, which are particularly vulnerable to subtle, spatially concentrated modifications. The attack led to a significant decrease in precision, recall, and F1 scores across all models. Notably, ResNet-18 and MobileNet-v3 struggled more than EfficientNet-B0 under this attack, reflecting the latter's better feature extraction capabilities and slightly enhanced robustness to multi-pixel perturbations.

4.4. Results for Multi-Pixel Deception Attack

The proposed Multi-Pixel Deception (MPD) Attack demonstrated the highest success rate among the tested adversarial methods (refer to Table 6), as evidenced by the substantial performance degradation it induces. ResNet-18's accuracy dropped to 59.12%, with an F1 score of



(a) ROC Curve for Original Models

(b) ROC Curve for Models exploited via MPD

Figure 5: ROC Curves for (a) Original Models and (b) Models exploited via MPD.

Table 4
Performance Metrics & Attack Success Rate for One Pixel Attack

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC Score	Success Rate (%)
ResNet-18	78.17	78.12	78.31	78.21	0.783	5.78
EfficientNet-B0	82.01	82.32	82.63	82.42	0.819	6.19
MobileNet-v3	76.40	76.13	75.92	76.07	0.764	6.03

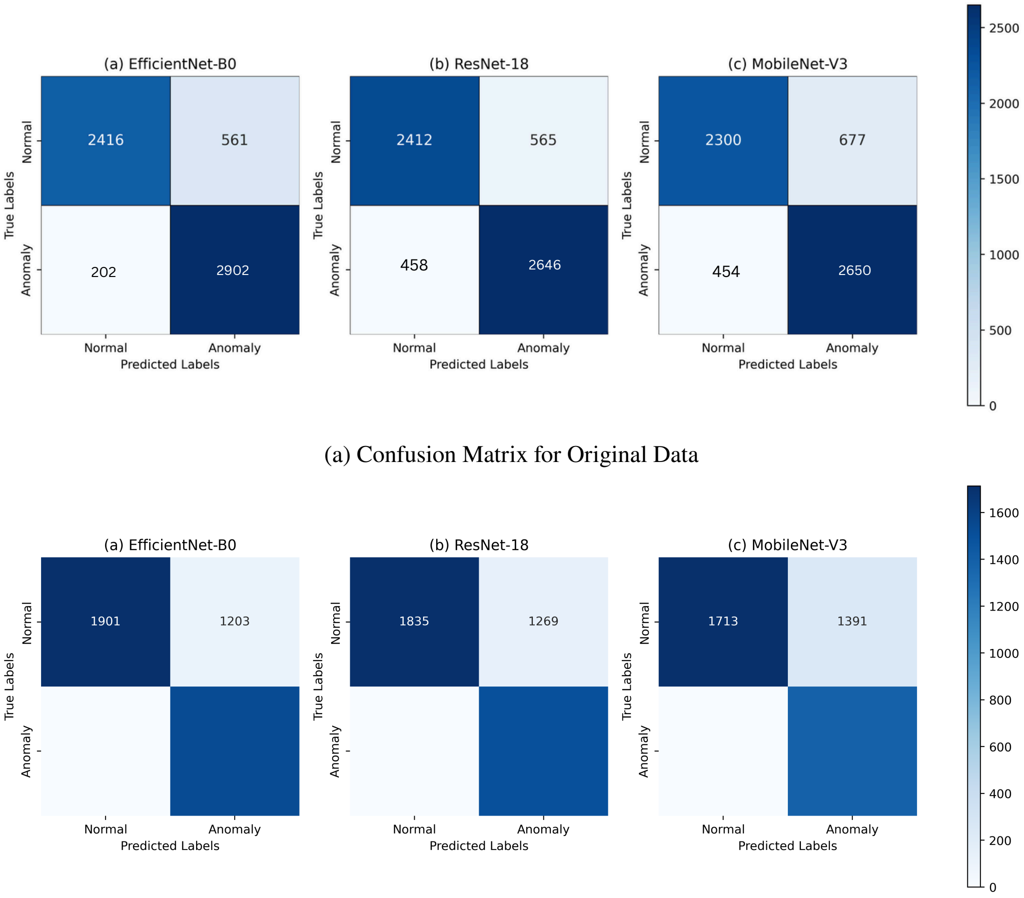
58.69%, marking a severe drop from its original performance. EfficientNet-B0, while more resilient, still experienced a significant reduction in accuracy (61.23%) and F1 score (61.76%). MobileNet-v3 suffered the most under this attack, with an accuracy of 55.17% and an F1 score of 54.71%. Figure 5 presents ROC Curves for models on original and adversarial data. It is evident that the MPD severely affects the predictive performance of models and increases the false positive rates (as shown in Figure 6) while affecting accuracy and recognition power.

The MPD Attack effectively combined the strengths of both the One Pixel and Pixle attacks, leveraging multi-pixel perturbations while maintaining minimal visual distortion. This hybrid attack proved to be the most disruptive, particularly for lightweight models like MobileNet-v3, which are designed for efficiency rather than adversarial robustness. The degradation across all models, especially in terms of AUC score and precision (see Figure 7 for comparison of models under adversarial attacks), underscores the effectiveness of the MPD Attack in exploiting CNN vulnerabilities in anomaly detection tasks. In addition, Figure 8 illustrates how the MPD attack shifts the model's focus to irrelevant manipulated image pixels using GradCAM Selvaraju et al. (2017), leading to incorrect predictions. Grad-CAM heatmaps highlight regions of the input image that the model considers important for making predictions. The figure is structured to demonstrate both original and adversarial effects of MPD attack on model attention patterns. The visualization is organized into three sections: column (a) shows the original input frames, columns (b)-(d) present paired comparisons of original versus adversarial GradCAM visualizations for each network: (b) EfficientNet-B0 (c) MobileNet-v3 (d) ResNet-18. Each GradCAM visualization uses a heat map overlay where warmer colours (red/yellow) indicate regions of higher attention, while cooler colours (blue) represent areas of lower attention. This representation allows for a direct comparison of how each model's attention mechanism responds to the adversarial MPD attack. This demonstrates the critical need for enhanced adversarial defence mechanisms in real-time applications, especially in resource-constrained environments like AIoT-OT systems.

While there are several video anomaly detection datasets available, this study opted not to use some widely known

Table 5
Performance Metrics & Attack Success Rate for Pixle Attack

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC Score	Success Rate (%)
ResNet-18	71.21	72.12	72.41	72.23	0.721	13.24
EfficientNet-B0	76.12	76.13	76.52	76.34	0.763	12.60
MobileNet-v3	69.12	68.16	69.33	68.78	0.681	16.24



(a) Confusion Matrix for Original Data

(b) Confusion Matrix for Adversarial Data

Figure 6: Confusion Matrices for (a) original data and (b) adversarial data.

Table 6
Performance Metrics & Attack Success Rate for MPD Attack

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**AUC Score**	**Success Rate (%)**
ResNet-18	59.12	59.29	58.14	58.69	0.580	30.20
EfficientNet-B0	61.23	62.36	61.23	61.76	0.619	29.10
MobileNet-v3	55.17	54.32	55.27	54.71	0.548	32.60

datasets for specific reasons. The XD-Violence dataset Wu et al. (2020), for example, integrates multimodal detection techniques, incorporating both audio and visual signals for detecting violent events. Since the focus of this work is scoped on attacking image-based models, leveraging a dataset that requires multimodal input would have been incompatible with this approach. Additionally, the RWF-2000 dataset Cheng et al. (2021), which focuses on violent activity detection in videos, was unavailable due to privacy concerns, making it impossible to integrate it into experiments for this study. These constraints led us to focus exclusively on the UCF-Crime dataset for its applicability to the research goals.

Although the UCF-Crime dataset is frequently used for video anomaly detection tasks, it presents certain limitations as well. The dataset's diversity in terms of video length, camera angles, and scene types poses challenges for image-based models, which may not effectively capture temporal dependencies or subtle anomalies. However, due to its wide acceptance in anomaly detection benchmarks, the researchers of this study chose to employ it as a primary dataset, despite its limitations. Additionally, only convolutional-based models, such as ResNet-18, EfficientNet-B0, and MobileNet-V3 Small were tested, to focus on models that are lightweight and suitable for AIoT applications. These models are known for their parameter efficiency and computational scalability, making them ideal for scenarios where large GPUs are not available, which is typical in edge computing environments.

During the experimentation of this study, larger models like DenseNet Huang et al. (2017) and other more complex architectures were deliberately avoided due to their propensity to overfit on smaller datasets like UCF-Crime. Larger

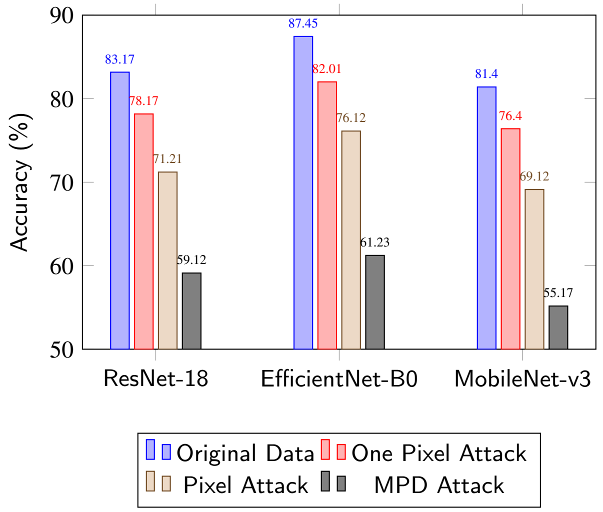


Figure 7: Comparison of Model Performance Across Different Attack Methods

models tend to perform well in environments with abundant labelled data. However, the limited size and diversity of the UCF-Crime dataset made overfitting a considerable risk, which could undermine the validity of the proposed adversarial attack results. Additionally, this study used only single-modality models focusing on visual input, as multimodal models incorporating audio or other sensor data were beyond the scope of this research. This study focused on exploring image-based convolutional models and how they respond to adversarial attacks in the specific context of anomaly detection.

5. Limitations

While the Multi-Pixel Deception (MPD) attack effectively deceived convolutional anomaly detection models with minimal perturbations, its evaluation was limited to a single dataset, UCF-Crime. Although this benchmark dataset enables reasonable generalization, its scalability in real-world scenarios remains uncertain. Additionally, experiments focused exclusively on CNNs due to their efficiency in AIoT applications, overlooking newer architectures like transformers Dosovitskiy (2020) and hybrid models Maaz et al. (2022), which may exhibit different vulnerabilities. Furthermore, the impact of MPD attacks on multimodal or higher-dimensional data, such as audio-visual models, remains unexplored, highlighting a crucial direction for future research.

6. Future Directions

While video streams are critical to many surveillance and monitoring systems, including those in theme parks, other types of data streams are also susceptible to adversarial attacks. These include audio streams, sensor data, and textual data from real-time communication systems. Each of these modalities presents unique vulnerabilities and challenges for adversarial attacks, and research in these areas has been gaining momentum.

For instance, adversarial attacks on audio streams often aim to fool speech recognition models through techniques like hidden voice commands or imperceptible perturbations that disrupt transcription. Such methods have demonstrated the vulnerability of systems like virtual assistants. Similarly, IoT devices and sensor networks have become increasingly targeted by adversarial attacks. Perturbations in sensor readings, such as GPS spoofing or the manipulation of environmental sensor inputs, have been shown to cause incorrect decisions in critical systems, such as autonomous vehicles or smart grids.

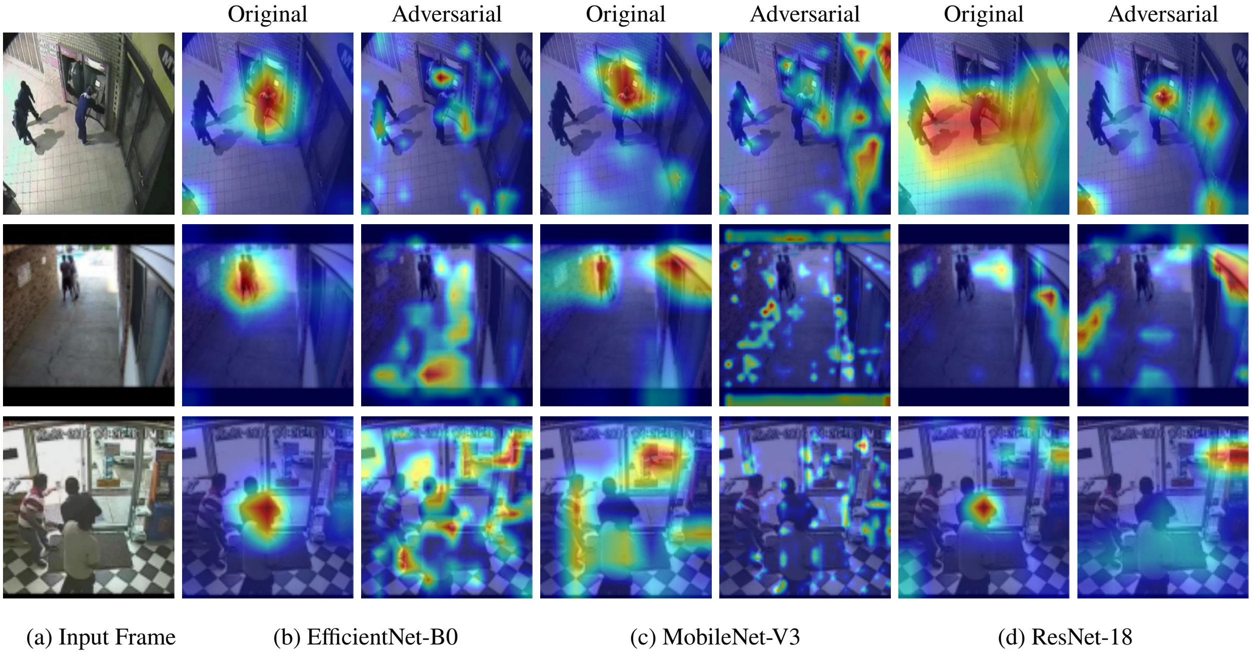
In the domain of natural language processing, adversarial examples have been crafted by substituting words or phrases to mislead models performing tasks like sentiment analysis or spam detection. Research techniques in this area include synonym substitution, paraphrasing, and even character-level modifications.

Highlighting these areas of research could inspire further exploration of cross-modal adversarial attacks or investigations into how vulnerabilities in one data stream could propagate to others in multimodal systems. Adopting such an interdisciplinary perspective could lead to the development of more robust defenses for systems that rely on multiple data streams.

Finally, integrating real-time adversarial testing within AIoT systems and developing defence mechanisms to counter these attacks would be critical steps toward improving the security of anomaly detection systems in edge computing environments.

7. Conclusion

This pioneering study highlights theme parks as critical infrastructure with significant potential for large-scale AIoT integration. While AIoT-based operational technologies offer transformative applications in public-facing venues, their adoption demands heightened vigilance against security and privacy threats. This research identifies key AIoT technologies in theme parks, maps their applications to AI and IoT systems, and examines the security risks they pose. Additionally, it demonstrates the vulnerability of smart surveillance systems to adversarial machine learning (AML) attacks, which can disrupt critical video-based models used for monitoring, maintenance, and anomaly detection. Findings from the proposed hybrid multi-pixel deception (MPD) adversarial attack technique reveal that such attacks can degrade model performance, mislead security systems, and hinder or delay responsive measures, ultimately compromising safety and security and potentially endangering visitors. The study urges theme park operators, technology vendors, and security professionals to recognize the evolving threat landscape and adopt proactive strategies to prevent service disruptions and mitigate attack transferability across similar AI models.



(a) Input Frame (b) EfficientNet-B0 (c) MobileNet-V3 (d) ResNet-18

Figure 8: GradCAM Selvaraju et al. (2017) visualization demonstrating adversarial effects of MPD attack on models' attention. (a) Input frame samples (b) EfficientNet-B0 results (c) MobileNet-v3 results (d) ResNet-18 results

References

Adam, A., Rivlin, E., Shimshoni, I., and Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE transactions on pattern analysis and machine intelligence*, 30(3):555–560.

Alahi, M. E. E., Sukkuea, A., Tina, F. W., Nag, A., Kurdthongmee, W., Suwannarat, K., and Mukhopadhyay, S. C. (2023). Integration of iot-enabled technologies and artificial intelligence (ai) for smart city scenario: Recent advancements and future trends. *Sensors*, 23(11).

Anton Clavé, S., Carlà-Uhink, F., and Freitag, F. (2023). *Time: Represented, Experienced, and Managed Temporalities in Theme Parks*, pages 309–322. Springer International Publishing, Cham.

Arshi, O. and Mondal, S. (2023). Advancements in sensors and actuators technologies for smart cities: a comprehensive review. *Smart Construction and Sustainable Cities*, 1(1):18.

Asaithambi, S. P. R., Venkatraman, R., and Venkatraman, S. (2023). A thematic travel recommendation system using an augmented big data analytical model. *Technologies*, 11(1):28.

Balducci, F., Buono, P., Desolda, G., Lanzilotti, R., Piccinno, A., and Costabile, M. F. (2024). Exploring the impact of an iot-based game on the experience of visitors at a natural science museum. *J. Comput. Cult. Herit.* Just Accepted.

Battour, M., Mady, K., Salaheldeen, M., Elsotouhy, M., Elbendary, I., and Boğan, E. (2023). Ai-enabled technologies to assist muslim tourists in halal-friendly tourism. *Journal of Islamic Marketing*, 14(5):1291–1309.

Bernardes, G. F., Ishibashi, R., Ivo, A. A., Rosset, V., and Kimura, B. Y. (2023). Prototyping low-cost automatic weather stations for natural disaster monitoring. *Digital Communications and Networks*, 9(4):941–956.

Bibri, S. E. (2023). The metaverse as a virtual model of platform urbanism: Its converging aiot, xreality, neurotech, and nanobiotech and their applications, challenges, and risks. *Smart Cities*, 6(3):1345–1384.

Bibri, S. E. and Jagatheesaperumal, S. K. (2023). Harnessing the potential of the metaverse and artificial intelligence for the internet of city things: Cost-effective xreality and synergistic aiot technologies. *Smart Cities*, 6(5):2397–2429.

Cao, Y., Li, J., Xiao, X., Wang, D., Xue, M., Ge, H., Liu, W., and Hu, G. (2024). Localstylefool: Regional video style transfer attack using segment anything model. *arXiv preprint arXiv:2403.11656*.

Cao, Y., Xiao, X., Sun, R., Wang, D., Xue, M., and Wen, S. (2023). Stylefool: Fooling video classification systems via style transfer. In *2023 IEEE symposium on security and privacy (SP)*, pages 1631–1648. IEEE.

Chandan, A., John, M., and Potdar, V. (2023). Achieving un sdgs in food supply chain using blockchain technology. *Sustainability*, 15(3).

Chang, J.-W., Sheybani, N., Hussain, S. S., Javaheripi, M., Hidano, S., and Koushanfar, F. (2023). Netflick: Adversarial flickering attacks on deep learning based video compression. *arXiv preprint arXiv:2304.01441*.

Chen, L.-B., Huang, X.-R., Chen, W.-H., Pai, W.-Y., Huang, G.-Z., and Wang, W.-C. (2023). Design and implementation of an artificial intelligence of things-based autonomous mobile robot system for cleaning garbage. *IEEE Sensors Journal*.

Cheng, M., Cai, K., and Li, M. (2021). Rwf-2000: An open large scale video database for violence detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 4183–4190.

Chengoden, R., Victor, N., Huynh-The, T., Yenduri, G., Jhaveri, R. H., Alazab, M., Bhattacharya, S., Hegde, P., Maddikunta, P. K. R., and Gadekallu, T. R. (2023). Metaverse for healthcare: A survey on potential applications, challenges and future directions. *IEEE Access*, 11:12765–12795.

Chithaluru, P., Al-Turjman, F., Kumar, M., and Stephan, T. (2023). Energy-balanced neuro-fuzzy dynamic clustering scheme for green & sustainable iot based smart cities. *Sustainable Cities and Society*, 90:104366.

Del Gallo, M., Mazzuto, G., Ciarapica, F. E., and Bevilacqua, M. (2023). Artificial intelligence to solve production scheduling problems in real industrial settings: Systematic literature review. *Electronics*, 12(23).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Doğan, S. and Niyet, İ. Z. (2024). *Artificial Intelligence (AI) in Tourism*, pages 3–21. Future Tourism Trends Volume 2. Emerald Publishing Limited.

Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Errousso, H., Alaoui, E. A. A., Benhadou, S., and Nayyar, A. (2024). Intelligent parking space management: a binary classification approach for detecting vacant spots. *Multimedia Tools and Applications*.

Flavián, C., Ibáñez-Sánchez, S., Orús, C., and Barta, S. (2024). The dark side of the metaverse: The role of gamification in event virtualization. *International Journal of Information Management*, 75:102726.

Forkan, A. R. M., Kang, Y.-B., Marti, F., Banerjee, A., McCarthy, C., Ghaderi, H., Costa, B., Dawod, A., Georgakopolous, D., and Jayaraman, P. P. (2023). Aiot-citysense: Ai and iot-driven city-scale sensing for roadside infrastructure maintenance. *Data Science and Engineering*, pages 1–15.

Gao, M., Souri, A., Zaker, M., Zhai, W., Guo, X., and Li, Q. (2023). A comprehensive analysis for crowd counting methodologies and algorithms in internet of things. *Cluster Computing*, pages 1–15.

Gupta, S., Modgil, S., Lee, C.-K., and Sivarajah, U. (2023). The future is yesterday: Use of ai-driven facial recognition to enhance value in the travel and tourism industry. *Information Systems Frontiers*, 25(3):1179–1195.

Habbal, A., Ali, M. K., and Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (ai trism): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications*, 240:122442.

Haghani, M., Coughlan, M., Crabb, B., Dierickx, A., Feliciani, C., van Gelder, R., Geoerg, P., Hocaoglu, N., Laws, S., Lovreglio, R., Miles, Z., Nicolas, A., O'Toole, W. J., Schaap, S., Semmens, T., Shahhoseini, Z., Spaaij, R., Tatrai, A., Webster, J., and Wilson, A. (2023). A roadmap for the future of crowd safety research and practice: Introducing the swiss cheese model of crowd safety and the imperative of a vision zero target. *Safety Science*, 168:106292.

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

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F., and Amira, A. (2023). Ai-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. *Artificial Intelligence Review*, 56(6):4929–5021.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.

Ijemaru, G. K., Ang, L.-M., and Seng, K. P. (2023). Swarm intelligence internet of vehicles approaches for opportunistic data collection and traffic engineering in smart city waste management. *Sensors*, 23(5):2860.

Isaia, C. and Michaelides, M. P. (2023). A review of wireless positioning techniques and technologies: From smart sensors to 6g. *Signals*, 4(1):90–136.

Jarašūnienė, A., Čižiūnienė, K., and Čereška, A. (2023). Research on impact of iot on warehouse management. *Sensors*, 23(4):2213.

Jiang, K., Chen, Z., Huang, H., Wang, J., Yang, D., Li, B., Wang, Y., and Zhang, W. (2023a). Efficient decision-based black-box patch attacks on video recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4379–4389.

Jiang, K., Chen, Z., Zhou, X., Zhang, J., Hong, L., Wang, J., Li, B., Wang, Y., and Zhang, W. (2023b). Towards decision-based sparse attacks on video recognition. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 1443–1454.

Jo, S., Lee, G. M., and Moon, I. (2024). Airline dynamic pricing with patient customers using deep exploration-based reinforcement learning. *Engineering Applications of Artificial Intelligence*, 133:108073.

Joung, J. and Kim, H. (2023). Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *International Journal of Information Management*, 70:102641.

Lee, H., Chatterjee, I., and Cho, G. (2023). Ai-powered intelligent seaport mobility: Enhancing container drayage efficiency through computer vision and deep learning. *Applied Sciences*, 13(22).

Li, P., Zhang, Y., Yuan, L., Zhao, J., Xu, X., and Zhang, X. (2023). Adversarial attacks on video object segmentation with hard region discovery. *IEEE Transactions on Circuits and Systems for Video Technology*.

Li, S., Aich, A., Zhu, S., Asif, S., Song, C., Roy-Chowdhury, A., and Krishnamurthy, S. (2021). Adversarial attacks on black box video classifiers: Leveraging the power of geometric transformations. *Advances in Neural Information Processing Systems*, 34:2085–2096.

Li, S., Iqbal, M., and Saxena, N. (2022). Future industry internet of things with zero-trust security. *Information Systems Frontiers*, pages 1–14.

Liao, X.-C., Chen, W.-N., Guo, X.-Q., Zhong, J., and Hu, X.-M. (2023). Crowd management through optimal layout of fences: An ant colony approach based on crowd simulation. *IEEE Transactions on Intelligent Transportation Systems*.

Liu, W., W. Luo, D. L., and Gao, S. (2018). Future frame prediction for anomaly detection – a new baseline. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lokman, A., Ramasamy, R. K., and Ting, C.-Y. (2023). Scheduling and predictive maintenance for smart toilet. *IEEE Access*, 11:17983–17999.

Lu, C., Shi, J., and Jia, J. (2013). Abnormal event detection at 150 fps in matlab. In *Proceedings of the IEEE international conference on computer vision*, pages 2720–2727.

Maaz, M., Shaker, A., Cholakkal, H., Khan, S., Zamir, S. W., Anwer, R. M., and Shahbaz Khan, F. (2022). Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications. In *European conference on computer vision*, pages 3–20. Springer.

Mahadevan, V., LI, W.-X., Bhalodia, V., and Vasconcelos, N. (2010). Anomaly detection in crowded scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1975–1981.

Maleki Varnosfaderani, S., Forouzanfar, M., and . (2024). The role of ai in hospitals and clinics: Transforming healthcare in the 21st century. *Bioengineering*, 11(4).

Mansour, D.-E. A., Numair, M., Zalhaf, A. S., Ramadan, R., Darwish, M. M. F., Huang, Q., Hussien, M. G., and Abdel-Rahim, O. (2023). Applications of iot and digital twin in electrical power systems: A comprehensive survey. *IET Generation, Transmission & Distribution*, 17(20):4457–4479.

Mathur, N. and Sinha, S. (2023). Security model and access control mechanisms for attack mitigation in ioe. In *Blockchain Technology for IoE*, pages 67–99. CRC Press.

Mu, R., Marcolino, L., Ni, Q., and Ruan, W. (2024). Enhancing robustness in video recognition models: Sparse adversarial attacks and beyond. *Neural Networks*, 171:127–143.

Murala, D. K., Panda, S. K., and Dash, S. P. (2023). Medmetaverse: Medical care of chronic disease patients and managing data using artificial intelligence, blockchain, and wearable devices state-of-the-art methodology. *IEEE Access*, 11:138954–138985.

Niksirat, K. S., Velykoivanenko, L., Zufferey, N., Cherubini, M., Huguenin, K., and Humbert, M. (2024). Wearable activity trackers: A survey on utility, privacy, and security. *ACM Comput. Surv.* Just Accepted.

Pandiyan, P., Saravanan, S., Usha, K., Kannadasan, R., Alsharif, M. H., and Kim, M.-K. (2023). Technological advancements toward smart energy management in smart cities. *Energy Reports*, 10:648–677.

Pang, Z., Guo, M., Smith-Cortez, B., O'Neill, Z., Yang, Z., Liu, M., and Dong, B. (2024). Quantification of hvac energy savings through occupancy presence sensors in an apartment setting: Field testing and inverse modeling approach. *Energy and Buildings*, 302:113752.

Patel, A., Kethavath, A., Kushwaha, N., Naorem, A., Jagadale, M., K.R., S., and P.S., R. (2023). Review of artificial intelligence and internet of things technologies in land and water management research during 1991–2021: A bibliometric analysis. *Engineering Applications of Artificial Intelligence*, 123:106335.

Perez, A. J., Siddiqui, F., Zeadally, S., and Lane, D. (2023). A review of iot systems to enable independence for the elderly and disabled individuals. *Internet of Things*, 21:100653.

Pomponi, J., Scardapane, S., and Uncini, A. (2022). Pixle: a fast and effective black-box attack based on rearranging pixels. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE.

Pony, R., Naeh, I., and Mannor, S. (2021). Over-the-air adversarial flickering attacks against video recognition networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 515–524.

Prandi, C., Barricelli, B. R., Mirri, S., and Fogli, D. (2023). Accessible wayfinding and navigation: a systematic mapping study. *Universal Access in the Information Society*, 22(1):185–212.

Price, K. V. (2013). Differential evolution. In *Handbook of optimization: From classical to modern approach*, pages 187–214. Springer.

Ramzan, M. R., Naeem, M., Chughtai, O., Ejaz, W., and Altaf, M. (2023). Radio resource management in energy harvesting cooperative cognitive uav assisted iot networks: A multi-objective approach. *Digital Communications and Networks*.

Rezapouraghdam, H., Akhshik, A., and Ramkissoon, H. (2023). Application of machine learning to predict visitors' green behavior in marine protected areas: Evidence from cyprus. *Journal of Sustainable Tourism*, 31(11):2479–2505.

Saad, M., Ahmad, M. B., Asif, M., Khan, M. K., Mahmood, T., and Mahmood, M. T. (2023). Blockchain-enabled vanet for smart solid waste management. *IEEE Access*, 11:5679–5700.

Saied, M., Guirguis, S., and Madbouly, M. (2024). Review of artificial intelligence for enhancing intrusion detection in the internet of things. *Engineering Applications of Artificial Intelligence*, 127:107231.

Salazar, F., Martínez-García, M. S., de Castro, A., Chávez-Fuentes, C., Cazorla, M., Ureña-Aguirre, J. d. P., and Altamirano, S. (2023). Uavs for business adoptions in smart city environments: Inventory management system. *Electronics*, 12(9):2090.

Sánchez, P. M. S., Celdrán, A. H., Bovet, G., and Pérez, G. M. (2024). Adversarial attacks and defenses on ml-and hardware-based iot device fingerprinting and identification. *Future Generation Computer Systems*, 152:30–42.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

Sheela, A. S., Ahamed, B., Poornima, D., and C, C. S. A. (2023). Integration of wireless sensor networks with iot in smart transportation systems and traffic management. In *2023 International Conference on Emerging Research in Computational Science (ICERCS)*, pages 1–6.

Su, J., Vargas, D. V., and Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841.

Sultani, W., Chen, C., and Shah, M. (2018). Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6479–6488.

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6105–6114. PMLR.

Ud Din, I., Awan, K. A., Almogren, A., and Rodrigues, J. J. P. C. (2023). Integration of iot and blockchain for decentralized management and ownership in the metaverse. *International Journal of Communication Systems*, 36(18):e5612.

Vikhyat (2024). Moondream. `https://github.com/vikhyat/moondream`. Accessed: 2024-09-30.

Wu, P., Liu, J., Shi, Y., Sun, Y., Shao, F., Wu, Z., and Yang, Z. (2020). Not only look, but also listen: Learning multimodal violence detection under weak supervision. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, pages 322–339. Springer.

Wu, T. and Hao, F. (2023). Edu-metaverse: concept, architecture, and applications. *Interactive Learning Environments*, 0(0):1–28.

Zhang, J., Yan, Q., Zhu, X., and Yu, K. (2023). Smart industrial iot empowered crowd sensing for safety monitoring in coal mine. *Digital Communications and Networks*, 9(2):296–305.

Zhu, Y., Ni, K., Li, X., Zaman, A., Liu, X., and Bai, Y. (2023). Artificial intelligence aided crowd analytics in rail transit station. *Transportation Research Record*, page 03611981231175156.