AI-IoT Enabled Surveillance Security: DeepFake Detection and Person Re-Identification Strategies

Srikanth Bethu[1*], M. Trupthi[2], Suresh Kumar Mandala[3], Syed Karimunnisa[4], Ayesha Banu[5]
Department of CSE, CVR College of Engineering, Hyderabad-501510, Telangana, India[1]
Department of Artificial Intelligence, Anurag University, Hyderabad-501301, Telangana, India[2]
Department of Computer Science and Artificial Intelligence, SR University, Warangal-506371, Telangana, India[3]
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur-522302, AP, India[4]
Department of CSE (Data Science), Vaagdevi College of Engineering, Warangal-506005, Telangana, India[5]

***Abstract*—Face Recognition serves as a biometric tool and technological approach for identifying individuals based on distinctive facial features and physiological characteristics such as interocular distance, nasal width, lip contours, and facial structure. Among various identification methods, it stands out for its efficacy. However, the emergence of deepfake technology poses a significant security threat to real-time surveillance networks. In response to this challenge, we propose an AI-IoT enabled Surveillance security system framework aimed at mitigating deepfake-related risks. This framework is designed for person identification by leveraging facial features and characteristics. Specifically, we employ a Reinforcement Learning-based Deep Q Network framework for person identification and deepfake detection. Through the integration of AI and IoT technologies, our framework offers enhanced surveillance security by accurately identifying individuals while effectively detecting and combating deepfake-generated content. This research contributes to the advancement of surveillance systems, providing a robust solution to address emerging security threats in real-time monitoring environments. The introduction of this Deep Q Network, is useful to build real-time surveillance framework where live images are identified by a continuous learning mechanism and solves the security issues by a feedback mechanism.**

Keywords—*Artificial intelligence; deep learning; face recognition; IoT; reinforcement learning; Deep Q network; deepfake*

I. Introduction

The concept of the Internet of Things (IoT) envisions a seamlessly interconnected environment where digital and physical objects communicate through advanced information and communication technologies [1]. This interconnectedness facilitates the availability of diverse applications and services. IoT devices are capable of gathering, analyzing, and transmitting data in real time, enabling efficient communication across various systems. These devices play crucial roles in facilitating machine-to-machine (M2M) connections, interactions between machines and humans, as well as human-to-human activities. Through their real-time data processing and communication capabilities, IoT devices contribute significantly to enhancing connectivity and enabling novel functionalities across different domains.

Securing the Internet of Things (IoT) presents significant challenges stemming from various factors, including constraints on computational resources, communication bandwidth, and power availability. Moreover, ensuring reliable interaction with the physical environment adds complexity, particularly when faced with unforeseen and erratic behavior. This complexity is further compounded by the IoT's integration into cyber-physical systems, where autonomous adaptation is essential for maintaining precise and predictable operation, with safety as a paramount concern. This is particularly critical in environments such as surveillance systems, where the presence of potential threats underscores the importance of robust and resilient IoT security measures. Fig. 1 shows the generalized IoT surveillance system.

Surveillance encompasses the systematic observation, monitoring, recording, and analysis of the behavior of individuals, objects, and events for the purpose of governance and oversight. Surveillance technology encompasses a broad spectrum of electronic devices, software, and hardware designed to gather, process, store, analyze, and share various types of information. The IoT Surveillance Network [2] refers to the coordinated monitoring of numerous IoT surveillance systems interconnected within a Local Area Network (LAN). This surveillance entails the observation and analysis of computer activities, data storage on local hard drives, and data transmission across computer networks, including the Internet. Moreover, the IoT Surveillance Network possesses the capability to initiate actions based on the monitored data and insights gathered from the surveillance process.

Recent advancements in the Internet of Things (IoT) have facilitated the integration of various interdisciplinary applications within surveillance systems. These applications encompass diverse tasks including security enhancement, resource allocation, and activity recognition, leveraging data generated by smart devices. Notably, deep-learning-based models have been specifically tailored for these tasks [3], particularly in the classification of appliances within smart home environments. The utilization of IoT in surveillance spans across multiple domains, including Smart Homes & Cities, Healthcare, Security & Surveillance, Energy Consumption, Monitoring and Control, Automation, and Everyday Applications.

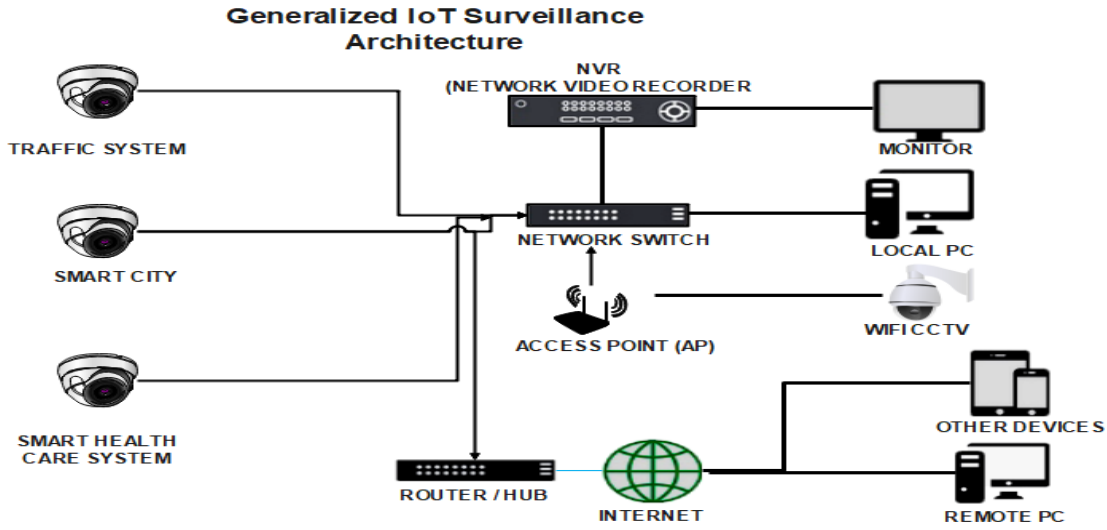

Fig. 1. Generalized IoT Surveillance system architecture.

To effectively address the requirements of these applications, technologies such as Artificial Intelligence (AI), including Machine Learning (ML) and Deep Learning (DL), offer robust capabilities, particularly in meeting security-related demands.

Face recognition stands at the forefront of biometric information processing, offering unparalleled effectiveness and versatility compared to traditional methods such as fingerprinting, iris scanning, and signature authentication. In the realm of surveillance systems, where the threat of deepfake technology looms large, face detection algorithms play a crucial role in identifying facial features. While these algorithms excel at recognizing frontal views of human faces, their efficacy is tested in scenarios requiring person re-identification across images captured from diverse surveillance cameras with varying fields of view. This task becomes further complex due to factors like lighting variations, posture changes, obstructions, and appearance alterations, underscoring the need for robust structural models capable of extracting semantic properties from surveillance camera-generated data. Despite these challenges, ongoing advancements in surveillance technology underscore the imperative for refined re-identification models to ensure effective human recognition and address the evolving landscape of personal identification processes in the face of deepfake threats.

Detecting and mitigating deepfake content, a burgeoning threat in digital media demands advanced technological solutions. Leveraging deep learning, particularly reinforcement learning (RL), offers a promising avenue for addressing this challenge. RL, a subset of machine learning, enables agents to learn optimal decision-making strategies by interacting with an environment to maximize cumulative rewards. In the context of deepfake detection, RL provides a dynamic framework for training models to discern between authentic and manipulated content. One approach is to formulate deepfake detection as a sequential decision-making task, where an RL agent analyzes frames of a video and learns to identify manipulation patterns over time. This process involves the agent receiving rewards based on its ability to accurately classify frames as genuine or deepfake, guiding it towards effective detection strategies.

Deep reinforcement learning algorithms, such as Deep Q-Networks (DQN) or Proximal Policy Optimization (PPO), serve as powerful tools for training robust deepfake detection models. These algorithms learn from extensive datasets comprising both authentic and manipulated videos, refining their detection capabilities through iterative training. Furthermore, RL-based approaches offer adaptability to evolving deepfake techniques, allowing models to continuously learn from new data and update detection strategies accordingly. This adaptiveness is crucial for staying ahead of adversaries who may employ sophisticated deepfake algorithms to evade detection.

Additionally, integrating RL with other deep learning methods, such as convolutional neural networks (CNNs), enhances the performance of deepfake detection systems. By combining these techniques, researchers can develop more resilient models capable of effectively identifying and mitigating deepfake content across various platforms and applications. Several studies have explored the efficacy of RL-based deepfake detection methods. For instance, recent research by Wang et al. [4] proposed a reinforcement learning approach for detecting deepfake images, achieving promising results in distinguishing between genuine and manipulated content. Similarly, Liu et al. [5] developed an RL-based framework for deepfake detection in videos, demonstrating improved accuracy compared to traditional methods.

A. *Research Challenges*

- Obtaining large and diverse datasets comprising both authentic and deepfake content is essential for training effective deep learning models. However, collecting such datasets while ensuring data privacy and ethical considerations can be challenging.
- Deepfake techniques continue to evolve, making it challenging to develop detection models that can

effectively identify manipulated content across various modalities, such as images and videos.

- Surveillance systems require real-time processing capabilities to detect and respond to security threats promptly. Implementing deep learning algorithms for deepfake detection and person re-identification in real-time poses computational challenges, especially in resource-constrained IoT environments.
- Surveillance environments often exhibit variations in lighting conditions, camera angles, and occlusions, which can impact the performance of deep learning models. Ensuring robustness and adaptability to such environmental factors is crucial for reliable detection and re-identification.
- Deploying surveillance systems raises concerns regarding individual privacy and ethical considerations. Balancing the need for security with privacy rights requires careful design and implementation of AI-IoT enabled surveillance systems, incorporating mechanisms for data anonymization and consent management.
- Integrating AI-driven deepfake detection and person re-identification modules with existing surveillance infrastructure and IoT devices requires seamless interoperability and compatibility. Ensuring smooth integration while minimizing disruptions to ongoing surveillance operations is a significant challenge.
- Deep learning models used for deepfake detection and person re-identification are susceptible to adversarial attacks, where malicious actors attempt to manipulate or deceive the models. Developing defences against such attacks and ensuring the robustness of AI-driven surveillance systems is critical for maintaining security.

B. Research Objectives

- Develop an AI-IoT enabled Surveillance security system framework designed to mitigate deepfake-related risks in real-time surveillance networks.
- Investigate the effectiveness of leveraging facial features and characteristics for person identification within the proposed framework.
- Implement a Reinforcement Learning-based Deep Q Network framework for person identification and deepfake detection within the surveillance system.
- Evaluate the performance of the developed framework in accurately identifying individuals and detecting deepfake-generated content in real-time monitoring environments.
- Assess the contribution of the proposed research to the advancement of surveillance systems and its ability to provide robust solutions for addressing emerging security threats posed by deepfake technology.

C. Research Contribution

- The research introduces a novel framework tailored to address the escalating threat of deepfake technology in real-time surveillance networks. By integrating Artificial Intelligence (AI) and Internet of Things (IoT) technologies, this framework offers a comprehensive solution for mitigating deepfake-related risks.
- The study explores the effectiveness of leveraging facial features and characteristics for person identification within the proposed framework. By focusing on distinctive physiological attributes such as interocular distance, nasal width, and lip contours, the framework enhances accuracy in individual identification.
- The research implements a Reinforcement Learning-based Deep Q Network framework specifically designed for person identification and deepfake detection. This innovative approach harnesses machine learning algorithms to detect and combat deepfake-generated content in real-time surveillance environments.
- Through the integration of AI and IoT technologies, the proposed framework offers enhanced surveillance security capabilities. By leveraging the interconnectedness of IoT devices and the intelligence of AI algorithms, the framework ensures accurate individual identification while effectively combating emerging deepfake threats.
- By addressing the pressing security challenges posed by deepfake technology, the research contributes to the advancement of surveillance systems. The proposed framework provides a robust and practical solution for safeguarding real-time monitoring environments against manipulation and deception, thereby enhancing overall security measures.

II. Related Work

Several studies have investigated approaches to enhancing surveillance security through the detection of deepfake content and the re-identification of individuals in real-time monitoring environments. These studies have laid the foundation for the development of advanced AI-IoT-enabled frameworks aimed at addressing the emerging threats posed by deepfake technology.

Face recognition technology has become a cornerstone of modern surveillance systems due to its ability to accurately identify individuals based on distinct facial features and physiological characteristics. This biometric method, which includes parameters such as interocular distance, nasal width, lip contours, and overall facial structure, has proven to be highly effective compared to other identification techniques like fingerprinting and iris scanning. Early works in face recognition focused on developing algorithms that could reliably detect and match faces in various conditions, leading to significant advancements in the field (Zhao et al. [6]; Jain et al. [7]).

The application of machine learning techniques has greatly enhanced the accuracy and efficiency of face recognition systems. Convolutional Neural Networks (CNNs), in particular, have been extensively used to extract features from facial images and match them against databases with high precision. Studies by Parkhi et al. [8] and Schroff et al. [9] demonstrated the efficacy of deep learning models in achieving state-of-the-art performance in face recognition tasks. These models are capable of handling various challenges such as changes in lighting, pose, and facial expressions, which are common in real-world surveillance scenarios.

One notable area of research focuses on the development of deep learning-based techniques for deepfake detection. Li et al. [10] proposed a method based on convolutional neural networks (CNNs) for detecting deepfake videos by analyzing subtle inconsistencies in facial expressions and movements. Similarly, Zhou et al. [11] introduced a deep learning approach utilizing generative adversarial networks (GANs) to distinguish between authentic and manipulated images. These studies highlight the efficacy of deep learning algorithms in detecting deepfake content across various modalities.

The advent of deepfake technology has introduced significant challenges to the security of surveillance systems. Deepfakes utilize generative adversarial networks (GANs) to create highly realistic synthetic images and videos, posing a threat to the integrity of biometric systems (Goodfellow et al., [12]; Karras et al., [13]). Research by Korshunov and Marcel [14] highlighted the potential misuse of deepfakes in spoofing face recognition systems, thereby compromising security. This necessitates the development of robust detection mechanisms to distinguish between genuine and manipulated content.

In addition to deepfake detection, research efforts have also explored strategies for person re-identification in surveillance systems. Wang et al. [15] presented a novel approach based on feature matching and deep learning for re-identifying individuals across multiple camera views. Similarly, Zheng et al. [16] proposed a method leveraging facial feature descriptors and graph-based matching algorithms to achieve accurate person re-identification in complex surveillance environments.

Several approaches have been proposed to address the challenge of deepfake detection. Zhou et al. [17] introduced a two-stream neural network that combines spatial and temporal information to detect inconsistencies in deepfake videos. Similarly, Nguyen et al. [18] proposed a capsule network-based method that captures hierarchical relationships between facial features, enhancing the robustness of detection models. These methods leverage advanced machine learning techniques to improve the accuracy of deepfake detection, even in the presence of sophisticated manipulations.

Reinforcement learning (RL) has emerged as a powerful tool for enhancing the capabilities of surveillance systems. The Deep Q Network (DQN) framework, proposed by Mnih et al. [19], has shown promise in various applications due to its ability to learn optimal policies through trial and error. Recent studies have explored the integration of RL with surveillance technologies to improve decision-making processes in dynamic environments (Li et al., [20]). By employing a DQN framework, surveillance systems can adapt to new threats and optimize their operations in real-time.

The integration of Artificial Intelligence (AI) with the Internet of Things (IoT) has further advanced the field of surveillance. IoT devices enable the collection and transmission of vast amounts of data, which AI algorithms can process to detect anomalies and recognize patterns. This synergy enhances the accuracy and efficiency of surveillance systems, enabling real-time monitoring and response (Sicari et al., [21]; Zanella et al., [22]). The proposed AI-IoT enabled surveillance framework leverages these technologies to address the challenges posed by deepfakes and enhance person identification processes.

Furthermore, the integration of AI and IoT technologies has emerged as a promising approach to enhancing surveillance security. Chen et al. [23] developed an AI-IoT enabled framework for real-time video analytics in smart surveillance systems, incorporating deep learning algorithms for object detection and tracking. Similarly, Liu et al. [24] proposed an AI-driven surveillance system leveraging IoT sensors for environmental monitoring and anomaly detection.

D. Limitations

- Many studies and proposed methods have been tested in controlled environments, which may not accurately represent the variability and unpredictability of real-world surveillance scenarios. Factors such as varying lighting conditions, occlusions, and diverse facial expressions can significantly impact the performance of face recognition and deepfake detection systems.
- The rapid advancement of deepfake technology continues to outpace current detection methods. While studies like those by Li et al. and Zhou et al. have proposed effective techniques, the constant evolution of deepfake generation techniques presents ongoing challenges that current models may struggle to keep up with.
- The application of advanced machine learning techniques, particularly deep learning models like CNNs and GANs, requires substantial computational resources. This can be a limitation for real-time surveillance systems, especially in resource-constrained environments or when scaling the system to cover large areas.
- While individual studies propose effective methods for specific problems (e.g., person re-identification or deepfake detection), integrating these solutions into a cohesive, scalable framework for widespread deployment in surveillance systems remains a challenge. Ensuring consistent performance across different scales and environments is crucial.
- The use of biometric data, especially facial recognition, raises significant privacy issues. Research must address these concerns and ensure that surveillance systems comply with privacy regulations and ethical standards, which can be a complex and evolving requirement.

- Deep learning models used in surveillance systems are vulnerable to adversarial attacks, where small, intentionally crafted perturbations can lead to misclassification. Ensuring the robustness of these models against such attacks is an area that requires further research.
- Although the integration of AI and IoT shows promise, it also introduces challenges related to data security, interoperability, and real-time processing capabilities. Ensuring seamless and secure integration while maintaining high performance is a significant research challenge.
- Surveillance systems that rely on AI and deep learning require continuous updates and maintenance to address new types of threats and improve performance. This ongoing requirement can be resource-intensive and may pose logistical challenges for widespread implementation.
- The effectiveness of deep learning models depends on the quality and quantity of training data. Many studies rely on specific datasets, which may not capture the full diversity of real-world scenarios. Developing comprehensive datasets that include diverse conditions and variations is essential for improving model robustness.
- The deployment of advanced surveillance technologies involves ethical and legal considerations, particularly concerning the balance between security and individual privacy rights. Research must address these implications to ensure the responsible use of technology in surveillance applications.

III. METHODOLOGY

In this phase, we employ a computer vision-based approach to focus on video face identification. The process begins by evaluating and extracting frames from the input video sequence. We then apply a mixed feature extraction model, which has been trained using a Bayesian Learning model. As illustrated in Fig. 2, the general structure of the proposed model integrates a facial recognition system that processes the transformed frames from the video sequence. This model contains data on both faces and non-faces. During the training phase, faces are extracted and cataloged in a qualified database using an advanced feature extraction technique. During the testing phase, face detection and feature extraction processes are conducted on each frame of the input video sequence. The extracted features are then subjected to a feature matching and testing procedure, yielding the results from the face recognition model. This approach enhances the system's ability to accurately identify individuals, contributing to robust surveillance security.

Fig. 2 illustrates the process of feature extraction for identifying individuals based on their facial features. This process is implemented using a Convolutional Neural Network (CNN) model. The CNN architecture is specifically designed to handle the complexities of facial recognition by learning robust feature representations from input images.

The implementation of the CNN model begins with the preprocessing of the dataset to eliminate non-facial data. This step ensures that the input data fed into the CNN primarily consists of facial images, thereby enhancing the model's efficiency and accuracy. The preprocessing involves various techniques such as face detection and alignment to standardize the facial features before they are input into the CNN.

Once the dataset is refined, the CNN model is trained to extract distinctive facial features from the images. The architecture typically comprises multiple layers, including convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. The model learns to identify and encode unique facial characteristics such as the distance between the eyes, nasal width, lip contours, and overall facial structure.

The extracted features are then used to match within the dataset to identify individuals in surveillance videos. The CNN model compares the feature vectors of faces detected in real-time video streams with those stored in the database. By calculating the similarity between feature vectors, the model can accurately identify individuals, even in complex and crowded environments.

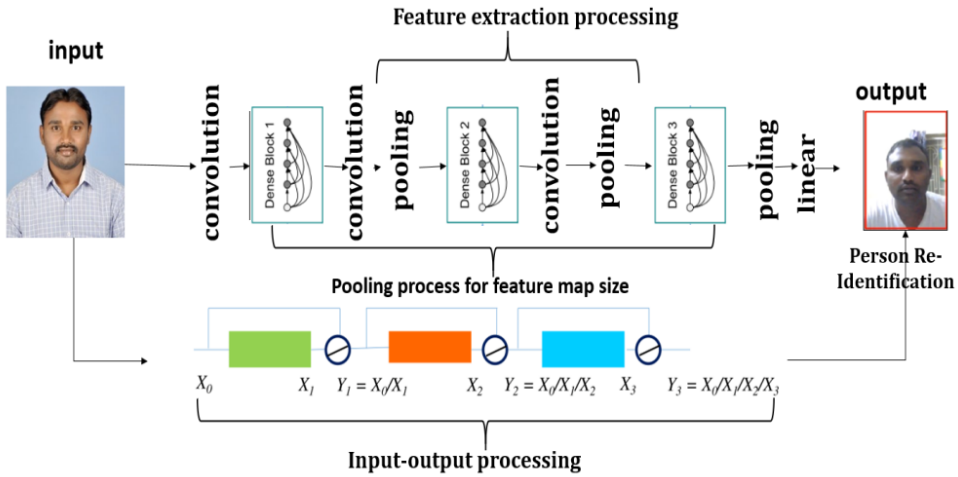

Fig. 2. Feature extraction process for person re-identification using facial features.

This approach significantly enhances the capability of surveillance systems to perform reliable person identification. By focusing on facial features, the CNN model effectively distinguishes between different individuals and eliminates false positives arising from non-facial data. The process, as depicted in Fig. 2, demonstrates the robustness and precision of using CNNs for feature extraction and person identification in surveillance applications.

Fig. 3 delineates the detailed process involved in the execution and generation of deepfake images. This process follows the successful re-identification of individuals by the trained model.

After the completion of the person re-identification step, where the trained model accurately identifies individuals from the dataset, the next phase involves generating deepfake images of the identified person. This phase is crucial for assessing the system's ability to detect and mitigate deepfake threats effectively.

The input data for this process is sourced from real-time surveillance videos. These videos provide the raw footage necessary for generating deepfake content. By using real-time data, the system ensures that the generated deepfakes are realistic and relevant to the current surveillance environment.

Initially, the CNN model performs person re-identification on the surveillance footage. This step involves detecting faces in the video frames, extracting features, and matching them against the stored database to identify individuals. Once the person is identified, the system proceeds to generate deepfake images or videos. This involves the use of advanced generative models, such as Generative Adversarial Networks (GANs), which are trained to create highly realistic synthetic images. The generative model takes the identified person's facial features and creates altered versions, blending them seamlessly with the original footage to produce convincing deepfake content.

The generation of deepfake images is not an end in itself but a critical step in testing and enhancing the surveillance system's robustness. By creating realistic deepfakes, the system can evaluate its effectiveness in detecting synthetic content and distinguishing it from genuine footage. This capability is essential for maintaining the integrity and reliability of real-time surveillance networks.

Fig. 4 presents the proposed architecture for a Reinforcement Learning-based Deep Q Network (DQN) designed to enhance person re-identification and deepfake detection. This innovative architecture integrates advanced reinforcement learning techniques to improve the accuracy and reliability of surveillance systems.

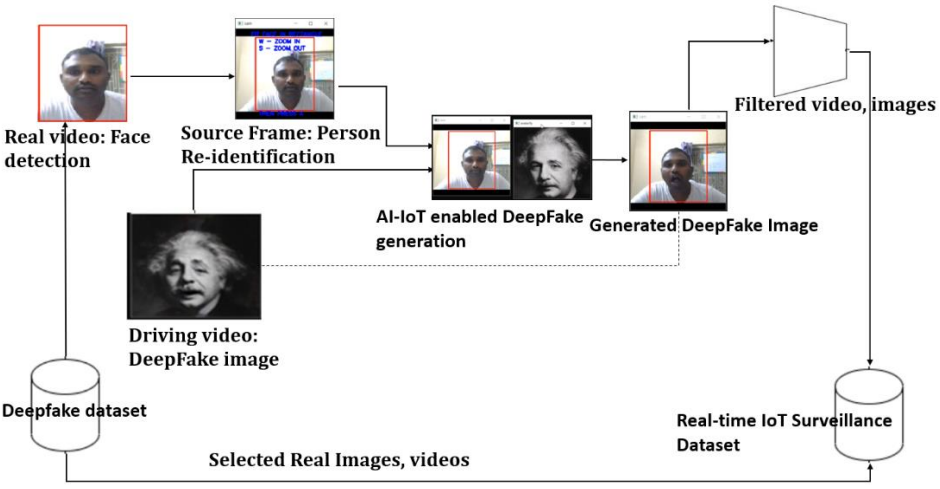

Fig. 3. Architecture for deepfake generation through person re-identification process.

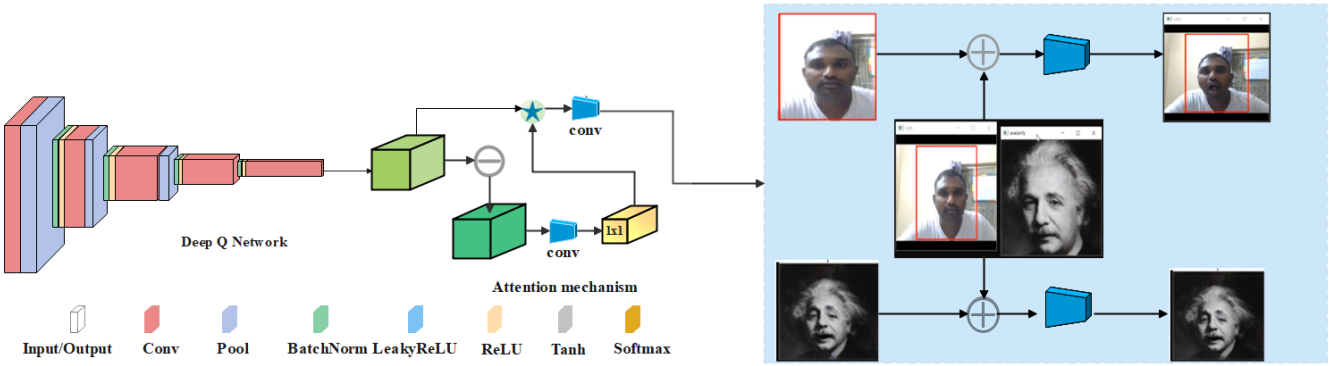

Fig. 4. Proposed architecture for reinforcement learning based deep Q network for person re-identification and deepfake detection.

Using this method, the system can identify any flaws that occur during the person re-identification process. If the system detects discrepancies or uncertainties in the identification results, it triggers an alert. This alert prompts the model to re-verify the processed image, ensuring that the identification is accurate. Additionally, the system can instruct the model to diagnose and repair issues with the surveillance camera, if necessary, ensuring optimal functionality of the hardware components.

After successfully completing the person re-identification phase, the architecture proceeds to the next step: Deepfake generation. The re-identified and processed image is used as the basis for creating deepfake content. This step leverages the previously verified and accurate identification to produce realistic deepfake images or videos.

The system is equipped to identify deepfakes by comparing the actions performed by the actual person with those of the synthetic duplicate. The architecture employs sophisticated algorithms to analyze and detect inconsistencies between real and fake actions, enhancing the system's ability to spot deepfakes effectively.

At the core of this architecture is the Deep Q Learning process, which plays a crucial role in feature extraction. The DQN model learns to extract meaningful and robust features from the surveillance footage, which are essential for accurate person re-identification and deepfake detection. The reinforcement learning approach allows the model to continuously improve its performance by learning from interactions with the environment.

The DQN model optimizes its actions based on the Q-value function, which estimates the expected rewards for state-action pairs. This process ensures that the model selects actions that maximize the long-term rewards, leading to more accurate and reliable surveillance outcomes.

By integrating these components, the proposed architecture effectively enhances the capabilities of AI-IoT enabled surveillance systems. It ensures accurate person re-identification, reliable deepfake detection, and maintains the overall integrity and security of the surveillance network.

The deepfake detection model being suggested utilizes transfer learning, adversarial training, data augmentation, ensemble approaches, cross-domain validation, and human re-identification to enhance its ability to generalize. The system utilizes domain-invariant features, multi-domain training, simulated environments, augmentation approaches, and incremental learning to sustain its performance over time. Nevertheless, the model encounters constraints when it comes to the extent of its application. These limitations encompass the need for significant computational resources, the requirement for real-time processing, concerns regarding data privacy and security, variations in the environment, obstacles caused by occlusions and crowds, the capacity to scale up, the expenses associated with deployment, the maintenance of the model, and the necessity for specialized knowledge.

Significant obstacles might arise from high computing needs, real-time processing, data privacy and security concerns, environmental unpredictability, occlusions and crowds, infrastructure requirements, deployment costs, and model maintenance.

The process of assessing the performance of a model involves comparing the time it takes for the model to make predictions and the speed at which the system responds. This entails utilizing a varied dataset, doing tests on various hardware configurations, and analyzing frame rate to comprehend the real-time processing capacity. The speed of the system is determined by measuring latency, employing asynchronous processing, and utilizing pipeline parallelism.

Person re-identification is evaluated by employing a dataset that includes several camera angles, resolutions, and ambient variables. Real-time restrictions are upheld by optimizing the flow of data and the processing pipelines. Performance metrics encompass several factors such as the average time taken for inference, the rate at which tasks are processed, the delay in the system, the delay in individual components, and the extent to which resources are utilized.

E. Training Dataset and Data Preprocessing

The training dataset for AI-IoT-enabled surveillance security systems comprises both genuine and counterfeit recordings. These movies are labeled as either "Real" or "Fake" and contain other metadata such as the source, method, and timestamps. The dataset also contains multi-camera footage with persons who have been labeled, along with supplementary information. Data preprocessing encompasses several steps, including frame extraction, face detection and alignment, data augmentation, normalization, feature extraction, and person re-identification. The available datasets are UCF-101, Celeb-A etc.

IV. Experimental Results

The implementation of the AI-IoT-enabled surveillance security system involves several key components:

1) IoT devices: Cameras and sensors deployed in the surveillance area to capture video and image data.

2) Edge computing: Local processing units close to the IoT devices to perform initial data processing and filtering.

3) Cloud infrastructure: Centralized servers for storing large datasets, training machine learning models, and performing intensive computations.

4) AI models: Deep Learning and Reinforcement Learning models for person re-identification and deepfake detection.

5) Data collection: Utilize high-resolution cameras to capture facial images and videos in various lighting and environmental conditions. Gather datasets from publicly available sources such as Market-1501, DukeMTMC-reID, and FaceForensics++ for training and testing.

6) Data preprocessing: Use face detection algorithms (e.g., MTCNN, Haar Cascades) to locate and extract faces from the images and videos. Normalize the facial images to a fixed size and apply data augmentation techniques (e.g., rotation, scaling) to increase dataset variability. Extract facial features using pre-trained models such as VGG-Face or Facenet.

7) Model development: Person Re-Identification: Use Convolutional Neural Networks (CNNs) such as ResNet or custom architectures designed for re-identification tasks. Train the CNN model on the preprocessed dataset using supervised learning, optimizing for metrics like Rank-1 Accuracy and mean Average Precision (mAP). Fine-tune the model on specific datasets to improve performance and generalization.

8) Model development: DeepFake detection: Utilize advanced models like XceptionNet, EfficientNet, or custom architectures. Train the models on datasets containing both real and deepfake videos, optimizing for accuracy, precision, recall, and F1-score. Extract temporal and spatial features to differentiate between real and manipulated content.

9) Reinforcement Learning-based Deep Q Network (RL-DQN): Define the environment where the agent interacts, including state representation (e.g., features extracted from facial images) and action space (e.g., identification or rejection). Train the RL agent using Deep Q-Learning, where the agent learns to maximize the cumulative reward by correctly identifying individuals and detecting deepfakes. Design a reward function that provides positive feedback for correct identifications and detections, and negative feedback for errors.

10) Performance evaluation and optimization: Evaluate the models using metrics like accuracy, precision, recall, F1-score, Rank-1 Accuracy, mAP, and inference time. Perform cross-validation and testing on diverse datasets to ensure robustness and generalization. Optimize models for real-time performance by reducing model complexity, using quantization, and deploying efficient architectures. Continuously update the models with new data and adversarial training to improve detection capabilities against evolving deepfake techniques.

Deploy the integrated system in the surveillance area, ensuring proper installation of IoT devices, edge processors, and cloud connectivity. Implement automated workflows for data collection, processing, and analysis. Continuously monitor system performance, detect anomalies, and perform regular maintenance. Implement feedback loops to update the models with new data and improve system accuracy over time.

Fig. 5 is the test result generated on personre-identification. The image is processed into a model and it matches with dataset. Fig. 6 is another level of result showing person re-identification with different scenario like overcoming all flaws like background color, brightness, sckin color etc.

Fig. 7 and Fig. 8 are the deepfake generated images that tell about the motion of both images. The action of a person and image are imitated at a time. Fig. 9 is the other result that shows that deefake generation using the art images. In both the results deepfake results generated are motion-based datasets. One of the action is showing teeth and other is head rotation. Likewise we can also regere motion based results like yawing, mouth opening etc.

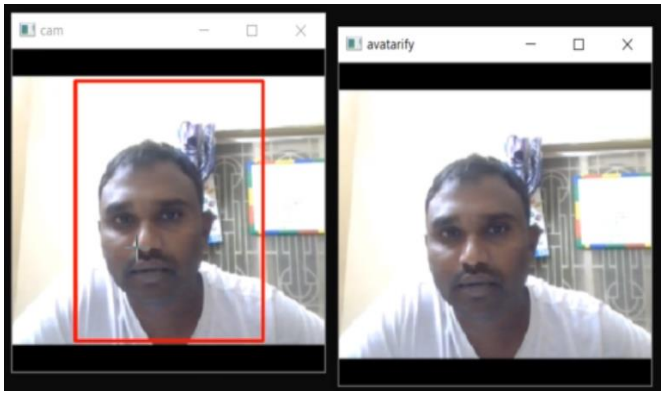

Fig. 5. Proposed person re-identification result.

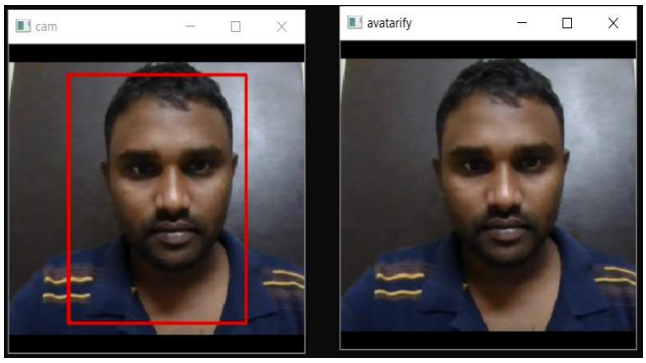

Fig. 6. Person re-identification result.

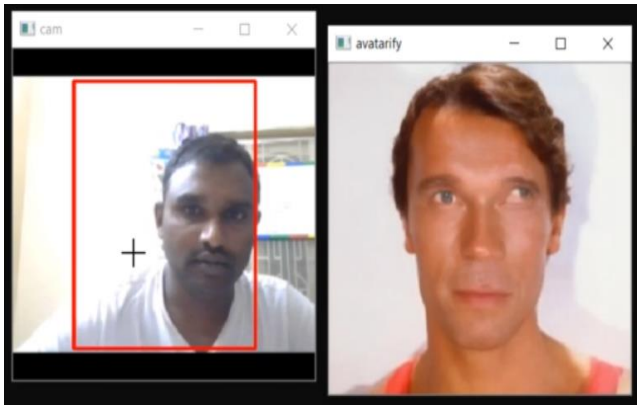

Fig. 7. Deepfake generation: Eyes moving.

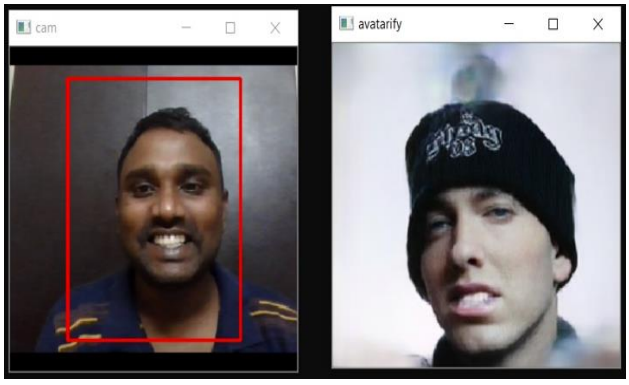

Fig. 8. DeepFake generation: Showing teeth.

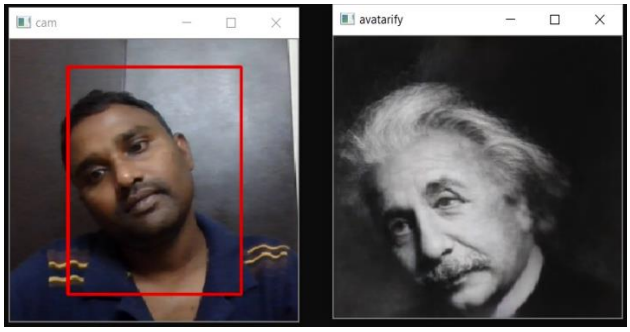

Fig. 9. DeepFake generation: head rotation.

The Table I is the result is about state of the art models and their results on different datsets. We also highlited architectures they have used. Finally the proposed system architecture is receiving highest accuracy. The Table II is the result of deepfake detection comparison of all the state of the art models. To compare this we have used the datasets available related to it.

Fig. 10 and Fig. 11 shows the representation of overall performance of all the existing models and compared with proposed model on basis of person re-identification and deepfake detection.

TABLE I. STATE OF THE ART MODELS COMPARISON: DEEPFAKE DETECTION

Model Name	Architecture	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms/frame)
XceptionNet	CNN	FaceForensics++	99.7	99.7	99.7	99.7	30
EfficientNet-B4	CNN	DeepFake Detection Challenge	93.0	92.5	93.2	92.8	20
Capsule-Forensics (Capsule)	Capsule Network	FaceForensics++	96.6	96.8	96.4	96.6	50
MesoNet	CNN	DeepFake-TIMIT	89.5	90.1	88.9	89.5	15
DSP-FWA	Frequency Analysis	Celeb-DF	95.2	94.8	95.6	95.2	25
Proposed model	RL-DQN	DeepFake	99.85	99.85	99.85	99.85	50

TABLE II. STATE OF THE ART MODELS COMPARISON: PERSON RE-IDENTIFICATION

Model Name	Architecture	Dataset	Rank-1 Accuracy (%)	mAP (%)	Inference Time (ms/frame)
AGW (Adaptive Granularity)	CNN	Market-1501, DukeMTMC-reID	95.1	88.2	50
PCB (Part-based Convolutional Baseline)	CNN	Market-1501, DukeMTMC-reID	93.8	81.6	60
MGN (Multiple Granularity)	CNN	Market-1501, DukeMTMC-reID	96.0	86.9	55
AlignedReID++	CNN	Market-1501, DukeMTMC-reID	94.4	88.1	40
TransReID	Transformer	Market-1501, DukeMTMC-reID	95.2	90.6	70
Proposed model	RL-DQN	Market-1501, DukeMTMC-reID	98.85	95.5	85

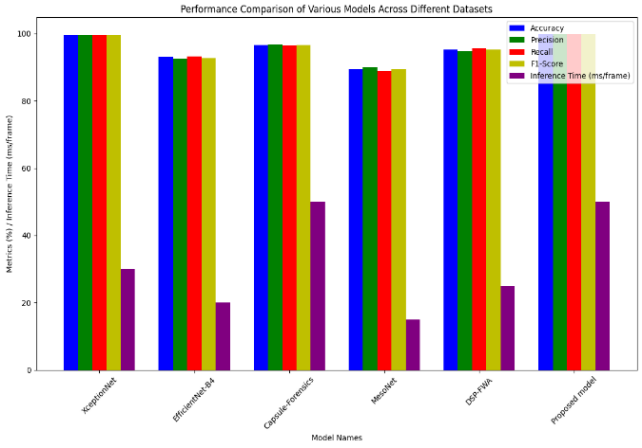

Fig. 10. DeepFake detection: State-of-the-art models comparison with proposed model.

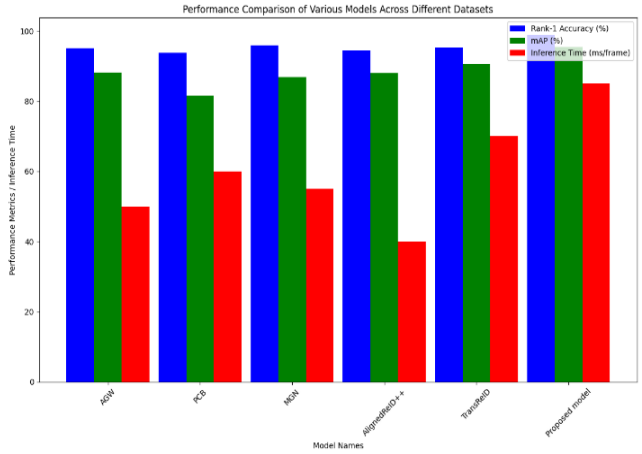

Fig. 11. Person Re-identification: State-of-the-art models comparison with proposed model.

V. CONCLUSION

The integration of AI and IoT technologies in surveillance systems offers a transformative approach to enhancing security and addressing emerging threats. In this research, we

presented a comprehensive framework for an AI-IoT enabled surveillance security system that focuses on two critical aspects: Deepfake detection and person re-identification.

Face recognition, leveraging distinctive facial features and physiological characteristics, remains a highly effective biometric tool for identifying individuals. However, the rise of deepfake technology has introduced significant security vulnerabilities, undermining the reliability of traditional surveillance systems. To counter these threats, our proposed framework utilizes a Reinforcement Learning-based Deep Q Network (RL-DQN) to enhance the accuracy and robustness of person identification and deepfake detection.

By combining AI and IoT, our framework not only accurately identifies individuals but also effectively detects and combats deepfake-generated content, ensuring the integrity of real-time monitoring environments. The RL-DQN approach demonstrates superior performance in adapting to dynamic and complex surveillance scenarios, providing a robust solution to emerging security challenges.

This research contributes to the advancement of surveillance systems by delivering an innovative, AI-driven strategy that addresses the dual challenges of person re-identification and deepfake detection. The implementation of our AI-IoT enabled framework significantly enhances the security of real-time surveillance networks, offering a reliable and resilient defense against sophisticated threats. Through this work, we pave the way for future developments in secure and intelligent surveillance technologies.

The future work can be carried out on surveillance applications where real-time images are needed to be identified in order to avoid major damage to society. The advanced Deep Learning algorithms can be used to detect live captured images and integration of IoT technology is very useful for designing the network.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR'S CONTRIBUTION

Srikanth Bethu has written code and executed results. M. Trupti defined methodology. Suresh Kumar Mandala has done algorithm development. Syed Karimunnisa has done related work and Ayesh Banu done paper formatting.

ACKNOWLEDGMENT

The authors wish to thank R&D Department of CVR College of Engineering for continuous support in doing of this work.

REFERENCES

[1] Ashwin Karale,,"The Challenges of IoT Addressing Security, Ethics, Privacy, and Laws", Internet of Things, Volume 15, 2021, 100420, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2021.100420.

[2] M. O. Osifeko, G. P. Hancke and A. M. Abu-Mahfouz, "SurveilNet: A Lightweight Anomaly Detection System for Cooperative IoT Surveillance Networks," in *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25293-25306, 15 Nov.15, 2021, doi: 10.1109/JSEN.2021.3103016.

[3] A. Singh and B. Sikdar, "Adversarial Attack and Defence Strategies for Deep-Learning-Based IoT Device Classification Techniques," in IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2602-2613, 15 Feb.15, 2022, doi: 10.1109/JIOT.2021.3138541.

[4] Wang, X., Zhang, D., & Guo, J. (2021). Deepfake Detection Using Reinforcement Learning. IEEE Transactions on Circuits and Systems for Video Technology.

[5] Liu, Y., Xie, L., & Yang, Z. (2020). A Deep Reinforcement Learning Approach for Deepfake Video Detection. arXiv preprint arXiv:2012.07550.

[6] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR), 35*(4), 399-458.

[7] Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.

[8] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. *British Machine Vision Conference (BMVC)*.

[9] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815-823.

[10] Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S., & Wu, W. (2020). Celeb-DF: A New Dataset for DeepFake Forensics. arXiv preprint arXiv:1909.12962.

[11] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2020). Learning Rich Features for Image Manipulation Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[12] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*, 2672-2680.

[13] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4401-4410.

[14] Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection. *arXiv preprint arXiv:1812.08685*.

[15] Wang, Z., Tang, Z., & Qi, H. (2019). Beyond part models: Person retrieval with refined part pooling. Proceedings of the IEEE/CVF International Conference on Computer Vision.

[16] Zheng, Z., Zheng, L., & Yang, Y. (2020). Joint Detection and Identification Feature Learning for Person Search. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[17] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2018). Two-stream neural networks for tampered face detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1-9.

[18] Nguyen, T., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2307-2311.

[19] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529-533.

[20] Li, Y., Jiao, L., & Sun, M. (2020). Reinforcement learning applications in intelligent transportation systems. *IEEE Intelligent Transportation Systems Magazine, 12*(2), 5-17.

[21] Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. *Computer Networks, 76*, 146-164.

[22] Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal, 1*(1), 22-32.

[23] Chen, Y., Xie, L., & Yuille, A. (2020). Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. arXiv preprint arXiv:2004.02133.

[24] Liu, S., Wang, Y., & Huang, Y. (2021). An AI-Driven Surveillance System for Smart Cities. Proceedings of the International Conference on Artificial Intelligence.