National College of Ireland

Deep-learning and Cloud-based IoT framework for intrusion detection using video surveillance

MSc Research Project
MSC Cloud Computing

Mahir Ahmed Jabarullah
Student ID: X22134433

School of Computing
National College of Ireland

Supervisor: Mr. Vikas Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Mahir Ahmed Jabarullah
Student ID:	X22134433
Programme:	MSC Cloud Computing
Year:	2025
Module:	MSc Research Project
Supervisor:	Mr. Vikas Sahni
Submission Due Date:	03/01/2025
Project Title:	Deep-learning and Cloud-based IoT framework for intrusion detection using video surveillance
Word Count:	6455
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	mahir ahmed
Date:	3rd January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deep-learning and Cloud-based IoT framework for intrusion detection using video surveillance

Mahir Ahmed Jabarullah
X22134433

Abstract

Traditional intrusion detection systems suffer from the delays and inefficiencies, due to the rule-based detection or manual monitoring. By employing scalable cloud services and lightweight edge computing, the proposed system enhances detection accuracy, speed, and resource efficiency. This research bridges the gap between robust detection systems and practical applications by integrating state-of-the-art models like FaceNet and InceptionResNet with HAAR cascades and SVMs for edge deployment. Challenges such as illumination variations and computational constraints on edge devices were addressed through quantization and advanced preprocessing techniques. The Key results demonstrate a detection accuracy of 97 % in live scenarios and scalability in cloud environments compared to the YOLO and SORT models, which has been used in the traditional system. This comprehensive framework ensures dynamic, cost-effective and secure monitoring for residential and commercial settings, marking a significant step toward smarter surveillance systems.

keywords : Intrusion detection, IoT surveillance, deep learning, edge computing, cloud scalability, FaceNet, InceptionResNet, HAAR cascades, SVMs

1 Introduction

With a rapidly evolving security environment, real-time intrusion detection based on video surveillance is becoming increasingly important. Traditional security systems rely heavily on either manual monitoring or rule-based algorithms for identification purposes; therefore, they are slow and inefficient in identifying a threat. Deep learning models with video surveillance could significantly enhance the accuracy and speed of the threat detection system in that they automatically identify suspicious activities in real-time. Scalable remote access to surveillance data is possible through the use of cloud-based IoT frameworks, which allows for monitoring and quick response. These technologies can be combined to process large volumes of video data in real-time, enhance the efficiency of the system, and apply smart analytics to security events. Intrusion large patterns can be learned by a deep learning algorithm. IoT devices with the facility of cloud connection allow for free and fluid communication along with low data storage requirements, thus providing a highly robust and dynamic security solution.

A system employs cameras, sensors and other intelligent devices to understand the activity pattern of the residents and act in response to it. It lets it identify an abnormal event that may be incipient at a threat. This new innovation of empowering cloud

computing with artificial intelligence means that the home security industry may undergo a drastic change in that the consumer could be presented with better security services. The analysis of artificial intelligence and facial recognition in surveillance systems might further contribute highly to safety improvements. In home security, these technologies are very useful because they can track individuals and possible threats instantly. Because of them, the targets are perfectly suitable for monitoring and gaining access. Therefore, the facial recognition is one of the most prospective technologies and the areas of research which can contribute to enhancements of people's communication and their protection in the future.

1.1 Problem Statement

Recent progress in technology has facilitated the deployment of sophisticated security systems in homes that use smart locks, cameras, and sensors. They possess the ability to detect and notify the homeowner of any potentially malevolent occurrence. Critical purposes in security and surveillance include visitor monitoring, security breach detection, and access restriction. However, rule-based techniques still dominate the commercial products. The progress in machine learning and cloud computing frameworks opens up the possibility of developing the next generation of home Intrusion detection systems and the same has been attempted in this work.

1.2 Motivation

A comparable system to Truecaller could be used to assist individuals in identifying unfamiliar visitors in residences already equipped with home security systems Aprinia et al. (2022). This method involves associating a phone number with a user-generated contact list. The system would thereafter correlate the user's profile with the caller identification details to ascertain the caller's identity. In the publication,Ahmed et al. (2016) indicated that the proposed system captures images of unidentified individuals through integrated residential security systems, compares these images to the data stored in the database, and determines the identities of these individuals. The AWS or an alternative cloud service provider may be seen as the resolution to this difficulty.

1.3 Research Objective

The objective of this project is to create a cost-effective motion detection surveillance system equipped with facial recognition features suitable for installation in residential, commercial, and various other structures. To identify an individual as known or unfamiliar, the system utilizes cloud services to assess the likelihood of that person being present at a different area linked to the current one. One distinct advantage is in its ability to monitor or follow an individual's movement in real-time while within the camera's range of view. Cloud computing enables network users to collaborate and share information regardless of the geographical location of their systems.

1.4 Research Question

The primary objective of this project is to develop a system analogous to True caller inside a framework where home security systems are reliant on cloud and internet technologies. To do this, the following question has been examined:

How to successfully identify intrusions and differentiate persons in real time while maintaining computational efficiency using a scalable system that makes use of edge devices and cloud frameworks in combination with lightweight and reliable machine learning model

2 Literature Review

2.1 Advances in Facial Detection

Recent progress in facial recognition has been significantly accelerated by the advancement and enhancement of the deep learning methodologies, specifically Convolutional Neural Network(CNN) Numerous challenges have been addressed by these advancements, such as the detection of face in an uncontrolled environment with occlusion and variable scales. EfficientFace employs a cross-scale feature fusion strategy, a receptive field enhancement module, and an attention mechanism to enhance the detection of occluded faces and those with disproportionate aspect ratios, attaining competitive performance with markedly lower computational costs relative to more substantial models Wang et al. (2023).

The WIDER FACE benchmark Mamieva et al. (2023) found that the RetinaNet baseline achieved excellent accuracy by improving the recognition of small and hidden faces. Face mask recognition has been studied using a combination of YOLOv3 and Faster R-CNN models, demonstrating the versatility of the model in handling new tasks, including tracking mask compliance during the covid-19 pandemic Singh et al. (2021). The Integrated Deep Model (IDM) merges Faster R-CNN with a stacked hourglass network to improve accuracy and minimize false positives in face detection, especially in uncontrolled environments Storey et al. (2018).

DBCFace eliminates the necessity for anchor design and non-maximum suppression, providing a purely CNN-based methodology that preserves competitive accuracy while enhancing speed Li et al. (2021). The inception blocks have been used by the Receptive Field-Enhanced Muti-task Cascaded CNN to improve the detection of small objects, demonstrating considerable performance improvements across multiple benchmarks. Li et al. (2020).

The Dual Shot Face Detector (DSFD) achieves state-of-the-art performance across several datasets by addressing face recognition variability using a feature enhancement module and improved anchor matching Li et al. (2019). Moreover, security surveillance has used intelligent video retrieval technology for face identification, utilising cascaded neural networks to increase detection accuracy and robustness in real-time situationsDong et al. (2020).

Comprehensive systems have been developed to address the real-world problems, including occlusions and blurred faces, using cutting-edge architecture and self-learning strategies to continuously improve detection performanceLi et al. (2019). Together, these findings represent a significant breakthrough in face detection that offers practical solutions to ongoing issues.

2.2 Advances in Home Intrusion Detection

The most effective techniques for identifying intruders in video surveillance systems combine intelligent camera coordination, sensor fusion, and deep learning. A popular method

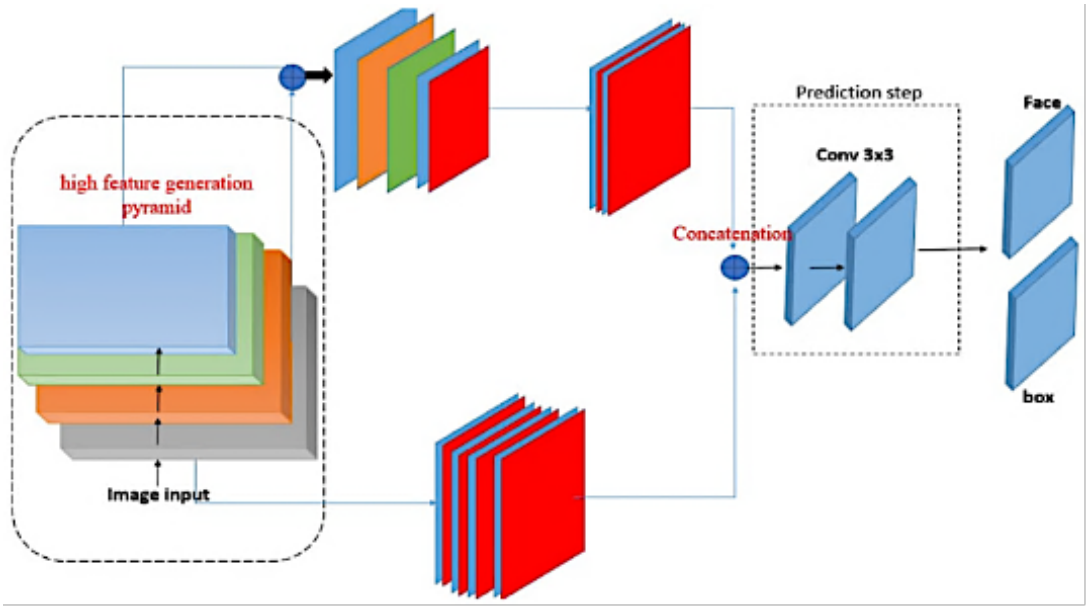

Figure 1: Architecture of method for face detection proposed by Mamieva et al. (2023)

for detecting anomalous occurrences in video streams that are essential for public safety is deep learning-based anomaly detection, which makes use of generative modelsDuong et al. (2023). These models are often enhanced by pre-processing and feature engineering techniques to improve detection accuracy.

Another important breakthrough of the recent years is introduction of the combined front-end and back-end architectures, including such solutions as the Multi-scale ResBlock structure, which adjusts to the incoming feed and context in real time, helping to enhance the accuracy of object detection and decrease computational expenses concurrently Kwon and Kim (2022).

Wireless sensor network techniques used in sensor fusion exploit modalities such as acoustic signals, and sensing probabilities to improve the detection probability beyond what is offered by simple probability theory while at the same time reducing false alarm rates. It could be done though hierarchical clustering and likelihood ratio tests, which fine-tunes detection thresholds Sharma and Chauhan (2020). Furthermore, the integration of dictionary learning algorithms and pyroelectric infrared sensors is another approach to intruder detection because the objective functions are reformulated and inclusions of label consistency enhance the classifying performance De et al. (2022).

The application of transfer learning, particularly, fine-tuning of pre-specified convolutional neural network architecture is also very timely in enhancing the accuracy and effectiveness of detecting objects at high precision that can be useful in real-time applications,Ahmadi et al. (2020). Furthermore, intelligent video surveillance systems are equipped with learning modules that address with its alarm signals based on the worldwide features that eliminate false alarms in challenging regions Cermeño et al. (2018).

Camera network coordination, particularly in environments with smart intruders, employs distributed algorithms to minimize detection times by optimizing camera trajectories and configurations Pasqualetti et al. (2014). Finally,combining electromagnetic field sensing with drone and static camera systems offers strong perimeter security, efficiently detects intruders, and reduces problems such as weather sensitivity and false alarms

Teixidó et al. (2021). All of these strategies work together to improve the efficiency of video surveillance systems in identifying trespassers.

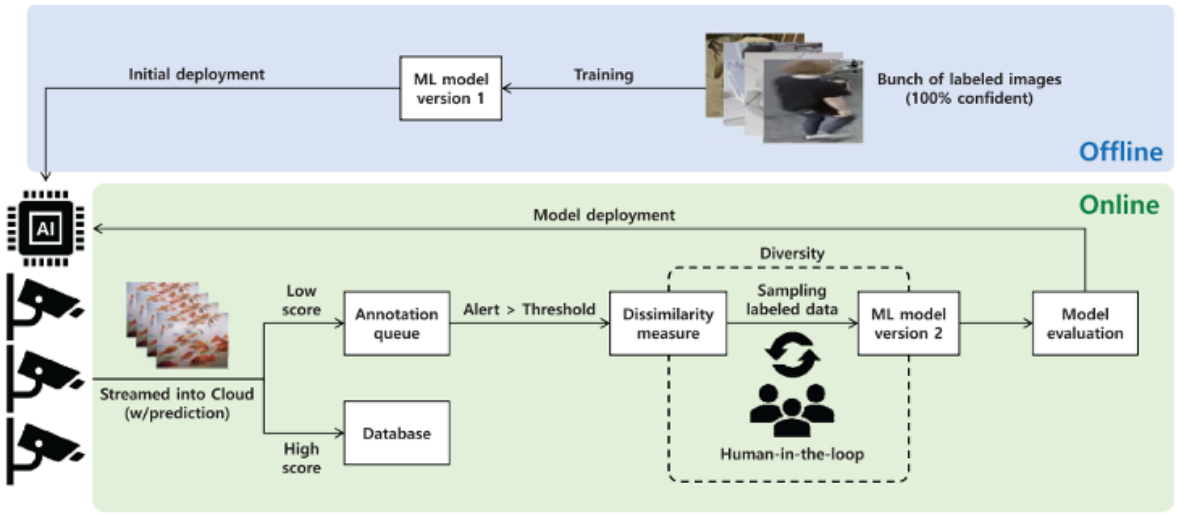

Figure 2: Hybrid architecture for intelligent video surveillance where the data from the new environment and current model are engaged in a closed environment Teixidó et al. (2021)

2.3 Advances in Cloud Deployed Intrusion Detection Systems using video surveillance

Real time intrusion detection using video surveillance is a relatively new and dynamic field that employs IoT Frameworks supported by cloud computing and deep learning. Depending on deep learning and proven object and intrusion detection algorithms on NVIDIA Jetson TX2, the Video Surveillance-Based Intrusion Detection System (VS-IDS) in an edge cloud environment has achieves 99% detection rate Sharma et al. (2023)

In order to maximise the response time and processing resources, the Deep Learning-based intelligent video surveillance system(DIVSS) employs CNNs for real-time motion detection and object recognition focusing on region of interest. Large data volumes and computing needs could have been addressed by fog-cloud-based frameworks that use recurrent neural network (RNNs) and future selection to obtain high recall rates in IoT without underfitting or overfitting Syed et al. (2023) Aldaej et al. (2023)

In addition, generative adversarial networks (GANs) in IoT-driven networks have improved cyber threat detection accuracy and efficiency Kandhro et al. (2023). Deep learning, IoT, and fog computing in weapon detection systems improve real-time surveillance by reducing bandwidth requirements and improving throughput and packet loss Fathy and Saleh (2022). Additionally, frameworks using swarm-based deep learning classifiers and feature selection algorithms may detect network and application-based attacks in multi-cloud IoT systems with high accuracy and low false positive rates Nizamudeen (2023)

A novel architecture utilizing a Multi-scale ResBlock scheme and domain adaptation technique has been proposed to substitute generic models with personalized ones for each camera, thereby enhancing real-time spatial and contextual comprehension. The

Internet of Things (IoT) has been employed to develop intelligent monitoring systems that utilize support vector machine classification for intrusion detection, resulting in high classification accuracy and low false-positive rates Malarvizhi Kumar and Choong Seon (2021)

The SurveilEdge system demonstrates a cooperative cloud-edge methodology, overcoming the constraints of conventional cloud-only and edge-only solutions by enhancing bandwidth expenses and query response durations while ensuring high precision via a convolutional neural network (CNN) training framework Wang et al. (2020).

To ensure confidentiality and integrity, security frameworks for cloud-based video surveillance systems have been developed. These frameworks use session key management and mutual authentication to secure video feeds over public networks Alsmirat et al. (2017). Furthermore, computer vision-based automated video surveillance system has shown substantial efficacy with low false alarm rates, increasing the efficacy of security operations in a variety of contexts Lipton et al. (2003).

The incorporation of machine learning methodologies, including random forest classifiers and dragonfly-enhanced invasive weed optimization-based Shepard CNN, has significantly enhanced the accuracy and sensitivity of cloud-based Intrusion Detection Systems (IDS), exhibiting exceptional efficacy in identifying network intrusions Attou et al. (2023) Sathiyadhas and Soosai Antony (2022).

These achievements collectively highlight the revolutionary influence of AI and ML in improving the capabilities and efficacy of cloud-based intrusion detection systems in video surveillance.

2.4 Research Summary

New progress in facial detection has been used deep learning methods such as CNNs to deal with special issues like occlusion, scale variation and illuminance differences. Proposed solutions, such as EfficientFace that uses cross-scale feature fusion and the receptive field enhancement and improved RetinaNet for detecting small and occluded faces, have been reported to obtain high accuracy compared with benchmarks such as WIDER FACE. Yolov3 and faster R-CNN have been deployed as hybrid models to adapt to mask compliance monitoring and more advanced systems like DBCFace and DSFD have made use of designs like anchor-free as well as refined anchor matching to address the issues of time consumption and optimal detection rates. The Receptive Field Enhanced Multi-Task Cascaded CNN and the Integrated Deep Model provide higher levels of face detection for small targets as well as targets obscured by other structures. These methods have seen impressive improvement especially in security, real time monitoring, and answering logical concerns such as blurry faces.

Awareness security in video surveillance employs IoT frameworks and deep learning in cloud-deployed intrusion detection systems. Applications like VS-IDS and DIVSS employ CNNs and edge-cloud framework to achieve better accuracy and to avoid gather high resource utilization. Sophisticated models using generative adversarial networks (GANs), swarm-based classifiers and domain adaptation deal with the network and application level threats. These innovations explain how artificial intelligence as well as machine learning can change the nature of surveillance systems, and assisting in better accuracy, efficiency and security of surveillance systems in all areas.

2.5 Research Novelty

Therefore, the use of ULAs in processing real time people tracking and crime prevention can be improved on effective use of edge devices including DLib and SVM. SVMs can be used further as the classifiers of the second level for the simple validation, since even if these machines are slower than the most of the others, they are nearly perfect in their efficiency. For tracking, new numerous algorithms like DeepSORT or ByteTrack are able to track multiple persons effectively and even if there are many people in the area. Through name matching system which is common with Truecaller, it is possible for the devices to either admit the identified persons or else alert strangers. These systems can be in communication with a distributed database for present data across devices for present day enhanced surveillance. Any companion device mobile application can have notification or the names of the people found or "unknown" status highlighted. The options for emergency police notifications, GPS zones for No-Go zones, and live tracking seem to assist with crime eradication. There is also improved reporting through two way communication within the application. To protect such data, there may be certain methods that mandate encryption of the data stream should be applied, or techniques to obscure the identity of all flagged people from future unlawful ID checks be utilized. This framework contributes a real, inclusive, and privacy-preserving structure using efficient edge computing, a light-weight AI model, and sympathetically fused systems for time-sensitively watching and preventing crime.

3 Methodology

3.1 Data Sources

Nature: Descriptive Information on cutting-edge video surveillance and home security systems powered by the internet of things.

Type of Data: Qualitative data on home security technology, lacking explicit quantitative measurements.

Recognition of the images: In this we have used the Pins Face dataset which contains 105 celebrities, 17534 images.

3.2 Machine learning Techniques Used

Based on its lightweight computing, a Haar cascade-based classifier was selected for the purpose of enabling efficient inference on edge devices. This choice was made because the classifier is easily deployable on edge devices. The identification of the eyes serves as the foundation for the construction of a geometric face model, and the detection of the nose is a supplemental validation approach. This information is utilised in the identification process and is then applied to the pictures. It is obtained by extracting Histogram of Oriented Gradients (HOG) characteristics from a large number of face photographs. Following the aggregation of the HOG characteristics for each individual user and face, a Support Vector Machine (SVM) model is trained to make predictions regarding faces inside the system.

It is clear that the Haar cascade classifier, which is depicted in Figure 3 [1], is an efficient method of feature extraction and plays a crucial part in the process of matching

[1] Analytics Vidya `https://medium.com/analytics-vidhya/haar-cascades-explained-38210e57970d`

features. When it comes to visual identification tasks, HOG descriptors are frequently utilised since they differentiate between objects based on intensity gradients or edge directions. In order to organise the picture blocks into a grid, gradients are assigned to each individual block, and the entire grid is then organised. Additionally, the HOG descriptors, which are simply gradient vectors for each pixel, are utilised in the training of the SVM model.

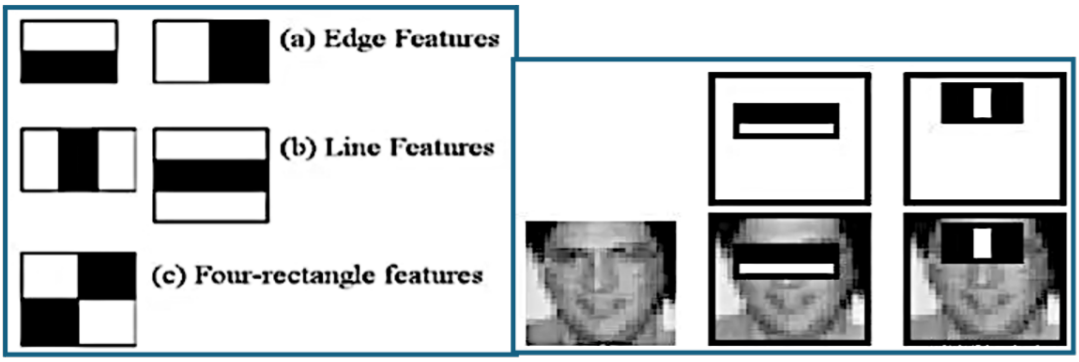

Figure 3: a,b,c represents the HAAR features and the images represents the matched feature areas

3.2.1 FaceNet

FaceNet is a face recognition system introduced by Google Researchers in 2015 in a paper called "FaceNet: A Unified Embedding for Face Recognition and Clustering." It obtained exceptional performance in many benchmark face recognition datasets, including Labelled Faces in the Wild (LFW) and Youtube Face Database. This is deep layered architecture for the features extraction based CNN network.

Figure 4: FaceNet Architecture Mensah et al. (2024)

The basic architecture of the system utilises either ZF-Net or Inception Network. In addition, it incorporates several 1*1 convolutions to reduce the parameter count. The deep learning models generate an embedding of the picture f(x) and apply L2 normalisation to it. Subsequently, these embeddings are inputted into the loss function to compute the loss. The objective of this loss function is to minimise the squares distance between two picture embeddings of the same identity, regardless of image condition and posture.

3.2.2 VGGNet

The VGG-based convolutional neural network requires an input picture with dimensions of 224 by 224 pixels and in the RGB colour format. The preprocessing layer operates on

an RGB picture with pixel values ranging from 0 to 255. It subtracts the mean image values, which are computed using the whole ImageNet training set.Mensah et al. (2024)

VGG employs smaller filters (3*3) with more depth as opposed to using larger filters. It has resulted in having an equivalent effective receptive field to that of a single 7 x 7 convolutional layer. Another iteration of VGGNet comprises 19 weight layers, which include 16 convolutional layers, 3 fully connected layers, and 5 pooling layers. Both variations of VGGNet have two Fully Connected layers, each with 4096 channels. The last fully linked layer employs a softmax layer to do categorization.

3.2.3 InceptionResNet

To prepare inputs for the InceptionResNetV2 model, use the keras. applications.inception_resnet_v2. preprocess_input function to them before feeding them into the model. The function inception_resnet_v2.preprocess_input will normalise the input pixels to a range of -1 to 1.

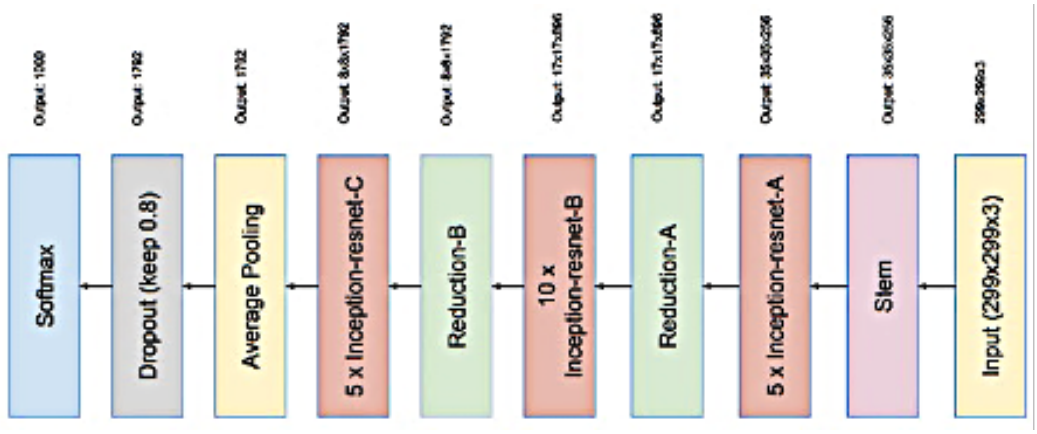

Figure 5: InceptionNetResNetV2 architecture flow Mensah et al. (2024)

To obtain the logits of the "top" layer, provide classifier_activation=None. When loading pretrained weights, the classifier_activation parameter can only have a value of None or "softmax".

3.3 Evaluation Metrics

Accuracy denotes that the proportion of all different correct predictions of the positive target level relative to the total number of expected positive cases. This indicator provides insight into the accuracy of the model's positive predictions.

Intersection over Union (IOU): IOU is also calculated by taking the ratio of area of overlap and area of union. Any algorithm that provides predicted bounding boxes as output can be evaluated using IOU. More formally, in order to apply Intersection and over Union to evaluate an (arbitrary) object detector we need:

4 Design Specification

Step 1: Amazon Web Service EC2 - The instances have been used to host the solution during first deployment. The Amazon Image Builder service makes it easier

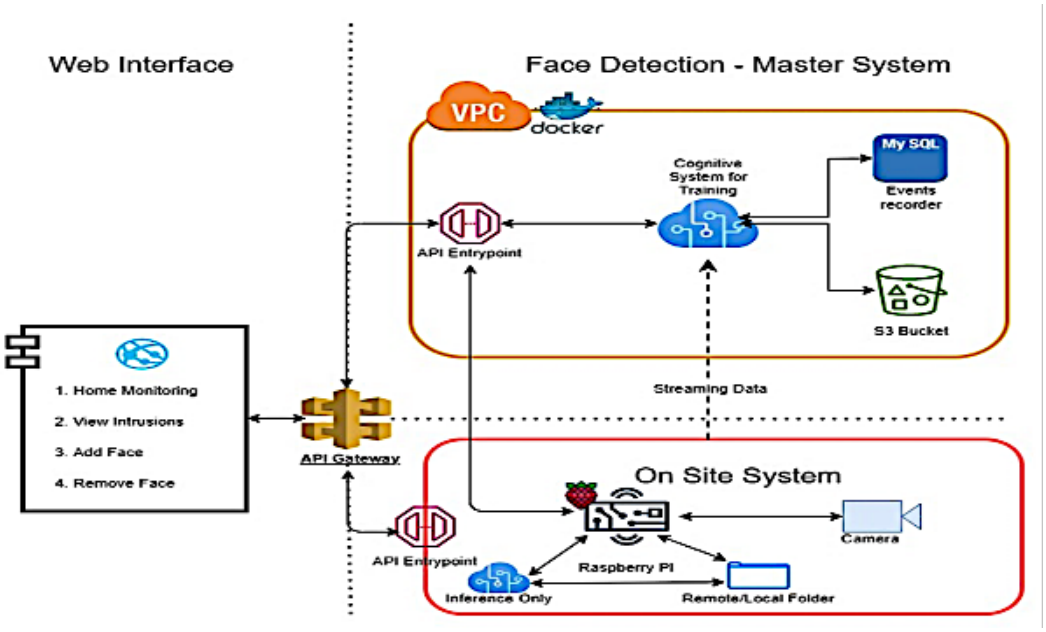

Figure 6: Architecture of the proposed system

to choose an Amazon Machine Image (AMI) for machine learning; Ubuntu 18.04 is an example of the selected system.

Step 2: Cloud Training and Fast API/Flask API – It is cloud training using Fast API, a library of standard APIs and supported languages. A Python API for UI components, Fast API is ideal for deep learning projects. The layered, high-performance, easy-to-use interfaces and flexibility make Python with Swagger straightforward to also employ to implement concepts.

Step 3: Imaging Creation for Deployment – involves creating the Docker image to enhance the management and deployment of containers, hence enhancing system efficiency. Utilizing Docker file patterns, one can acquire an operating system is then equipped with a deep learning environment based on Python. Docker Hub is responsible for the availability of this image.

Step 4: Data Base Creation – In order to incorporate necessary data across individuals, locations and objects, we need a relational database system. MySQL has been chosen as the database system because of its rapid performance and strong security features.

A detailed analysis of every option available led to the conclusion that AWS EC2 instances would be most suitable in hosting the solution in the cloud space due to the given problem description. AWS EC2 instances provide requirements of flexible, scalable, cost-optimized computing power, various instance type and selecting choices which makes it a perfect area for building and deploying an AI model. AWS EC2 instances effectively present wide options depending on memory intensity, processing capacity and cost. Looking at our needs for this project, we have not go for the instances with more computational power because these are billed. Once we had an Amazon account set up, we have used the Amazon Image Builder to select an Amazon Machine Learning AMI.

5 Implementation

5.1 Proposed Framework and Algorithm

The component of the detection phase consists of the following processes:

Step 1: **Capturing Real-Time Images:** The acquisition task that issued belongs to the camera, whereby images are captured next, with further processing carried out after the employment of the object detection practice. This model has been trained to look for objects and people to categorise them and, in this particular debate, to learn how to identify people.

Step 2: **Orientation of the Camera within the Frame**: Some information on the intended use of the monitoring is got on where the camera is placed, and this can be made in webcam configuration or door hole camera configuration.

Step 3: **The Detection Phase**: Considering only the first phase of the detection phase, the object detection model approximates the number of persons that can be recognized in the photographs that have been taken part. This is a process of detection and this is the third step to it. Regarding this phase, it is crucial for should measure the level of activity or the occupancy that is prevailing in the observed area.

Step 4: **The Creation of Folders for Identification:** The system will create folders for identification given that detection is the final process of this system. Identifiers are found in folders. Two of these folders are specifically named: "Known People" and "Unknown People".

Step 5: **Database Creation**: Building the database is the final step of the project which is then followed by the construction of the database. With much emphasis having been placed on comparison between this database and a cloud subscription service, it is expected that this database holds information on the said persons through categorizing the individuals into known and unknown.

Due to its capability in the identification of people classification and also database management, this system is ideal for multiple uses who are using this system so that the scenario is connected. These applications include area surveillance, security and accesses among others. Therefore, due to the cloud subscription part, the camera might be used together with clouds in order to gain more functions.

5.2 Training on Cloud

5.2.1 FastAPI

FastAPI is a Python package for building APIs. FastAPI facilitates concurrent processes, enables swift construction of REST APIs, and incorporates pre-existing elements and automated documentation, thereby augmenting efficiency and user-friendliness. It offers both high-level components for efficiently creating advanced UI components in common deep learning domains, and low-level components that may be combined to create new UI components. The objective is to accomplish both objectives while maintaining performance, usability, and adaptability.

5.2.2 Docker Image Creation

For this project, we generated a distinct Docker file that encompasses the necessary instructions to construct a basic operating system rooted in Python and the essential

deep learning packages. A Docker image was produced using the Docker file and it was hosted on Docker Hub, allowing global accessibility. Docker guarantees uniform environments throughout the many stages of development, streamlines the management of dependencies, and offers both portability and scalability.

5.2.3 Database Creation

MySQL can be any other kinds databases offers a wide range of sophisticated security capabilities. Therefore, MySQL is selected due to its user-friendly interface, fast performance, and robust security measures, which make it well-suited for efficiently managing sensitive data.

5.3 Face Detection inference on the Edge

As a result, the development of an efficient framework for real-time inference on edge devices required a lightweight approach. The Haar Cascade Classifier was used because it has no complex steps, so it is less computationally intensive. Furthermore, the HOG features apart from an SVM were used for accurate face recognition. Real time object detection is achieved by the Haar Cascade Classifier which is effective and fast The HOG-SVM on the other hand is accurate for face recognition.

5.3.1 Applying HAAR Cascades

This has the advantages of training many classifiers on different features and then combining them in order to develop an accurate classifier. The overall quality of the classifier is also enhanced by these factors, including variety and high quality of training photos which cover all the emotions on face and various facial conditions. The use of many classifiers increases the accuracy and reliability of detection; the system handles a number of facial emotions and conditions well.

5.3.2 Applying Geometrical Face model

The geometric face model begins with the eyes which are considered as the primary feature, and then, if the eyes are only partially occluded, the model continues with the nose. This method increases the accuracy of the facial model making it provide reliable detection all the time and even when the conditions are unfavourable. The geometric model provides the systematic representation of the facial parameters thus enhancing the robustness of the detection process. The least but most certain process of using the geometric face model avails for the accurate and dependable recognition of the faces under challenging scenarios.

5.3.3 Face Train finally using dlib and SVM - HOG features

Histogram of Oriented Gradients (HOG) is one of the most significant feature descriptors widely used in visual recognition task especially in the objects recognition tasks and recognition of the 3D objects as well. This method proves to be efficient in the context of which out of the two features, intensity gradients or edge directions, can be used to distinguish between objects in a picture. This process involves segmentation of the image into square regions of 8-x-8 and performing gradients on every single pixel nested

within that region. Gradients, which are represented by Gx and Gy, quantitative values, describe height or steepness of an intensity or an edge in a picture. Calculating Histogram of Oriented Gradient (HOG) descriptors provides a set of gradient vectors component for each pixel in the image that forms the basis of extracting other features.

The HOG descriptor is very important in the areas of face detection and recognition since it includes complicated features and patterns of facial images. In the context of facial features, the HOG descriptor is able to capture the significant contours of the features geometrically because the gradients of pixel intensities were considered. Based on this factor, the accuracy of face detection and recognition in conditions of different lighting and orientations. For the purpose of making the descriptor more robust and thus increasing reliability, the calculated gradient values are put into the histograms over several blocks in many a case being normalized to 8x8 block sizes. The normalisation procedure helps in the ability of reducing the impact of lighting changes in order to maintain balanced performance in different lighting conditions. The decentralisation helps the intrusion detection system to better respond to a large flow of information and is therefore useful in many scenarios. In the Figure 7 [2], it demonstrate the process of face detection using HOG extractor with Dlib.

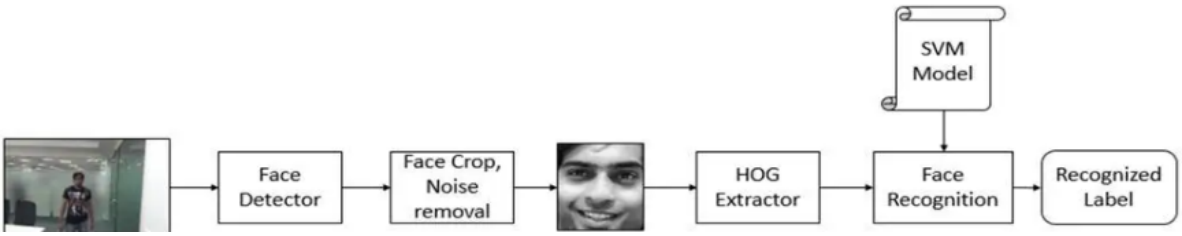

Figure 7: Face Detection model created and deployed using the Dlib on the edge platform.

5.4 Deep Learning Modules

For using the faceNet, InceptionResNetV2, VGGNet and other deep learning models, we have used the different keras APIs for the models performance.

5.5 PCA Features extraction with SVM

In this section, the features from the images are extracted using the PCA(Principal Component Analysis) and then the SVM is used for the image recognition. The framework can be seen as below,

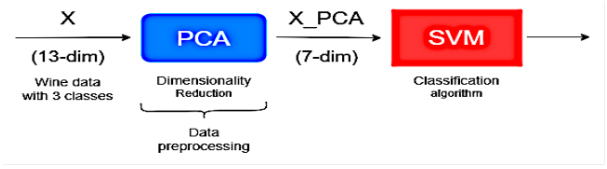

Figure 8: Framework for the PCA with SVM for the image classification

[2]https://www.analyticsvidhya.com/blog/2022/04/face-detection-using-the-dlib-face-detector-model/

6 Results and Discussion

The results demonstrated the hybrid system for object and face identification with the help of Haar cascades and deep learning models.

6.1 Object Identification

In this we have used the haar cascade for detecting the objects. It recognise the faces successfully within the images, as evidenced by bounding boxes that emphasise the identified areas. This indicates the system's capability to effectively localise the objects, which is a vital stage in face recognition process.

6.2 Face Recognition

The system evaluates the recognition process using the pin images dataset, which includes a lot of photos, ensures that face recognition performs efficiently under a variety of scenarios. The pipelines capability to reliably recognise and identify the faces across a variety of subject is demonstrated by a thorough examination of the dataset's distribution.

The Outcomes demonstrate the pipeline's accuracy and reliability in tasks involving both face recognition and object localisation. The technology provides a scalable and flexible solution for real-world scenarios like identity verification and surveillance by combining the Haar cascade with advanced identification models.

6.2.1 Using FaceNet

To identify and predict labels, the average embedding of ten randomly selected photos from each class was calculated using the pre-trained FaceNet model. The accuracy of the model considerably improved with the size of the dataset. The precision was 90.2% with 25 images used for reference embeddings. By increasing the number of photos to 50, the accuracy increased to 92.3%, it has been 94.6%, when the photos increased to 65. This pattern suggests that a greater number of images better represent each class, resulting in more precise embeddings and improved label recognition or prediction accuracy. Therefore, increasing the dataset directly results in a more accurate model.

Table 1: Comparison between different embeddings and Accuracy of the model

Number of images used to find reference embeddings	**Accuracy received**
25	90.2
50	92.3
65	94.6

6.2.2 Using VGGNet, PCA and SVM commutatively

A deep convolutional neural network called VGGNet captured key patterns and structures in images to extract high-level features that were used to measure similarity. Principal Component Analysis (PCA) was then used to process these features in order to lower their dimensionality. For the final classification, these reduced characteristics were subjected

to an SVM classifier. The SVM demonstrated that more components boost classification performance by achieving an accuracy of 96.455% with 128 PCA components and 97.03% with 256 components,

Table 2 : Comparison of different PCA components and Accuracy of SVM

PCA (number of components)	**Accuracy on SVM**
128	96.455
256	97.03

6.2.3 Using Inception

Successfull learning and generalisation were demonstrated by the inception model, which displayed increased training and validation accuracies with more epochs. Training accuracy peaked at about 70 epochs, while validation accuracy kept rising, suggesting improved generalisation. The reduction in error and the lower risk of overfitting were confirmed by declining training and validation losses. The model had trouble with misclassifications, particularly for Chris hemsworth and Lili Reinhart, although it correctly identified people like Grant Gustin and Alexandra Daddario. Despite these difficulties, the model demonstrated strong performance in unknown data with an average validation accuracy of 78.45%. In order to determine its sufficient, accuracy must be evaluated in relation to task-specific needs and standards.

6.2.4 Using ResNet50

Improved fitting to training data has been indicated by a machine learning model's training output across 16 epochs, which has loss beginning at roughly 0.66 and dropping to about 0.01. The percentage of accurate predictions, or accuracy, ranges from 0.78 to 1.0, indicating an improved prediction in training data. However, high training accuracy does not guarantee good performance on unseen data due to potential overfitting. Evaluating the model on a separate set of unseen data has led us to a precision score of 83.5%, which implies a sign of overfitting in this model.

Table 3: Detailed Models Comparison

Model	**Accuracy on Test Data**
FaceNet	94.6
VGGNet	97.03
Inception	78.44
ResNet50	83.85

6.3 Details and summary for the training process

The dataset is specifically divided in a holdout method so that 70 % of the different images are used for training, 30 % are used for validation, and 50 % rest initially taken out are used for testing. After that, these subsets are kept in the "Model Data" directory, which makes it easier to access and manage them during the model creation and assessment stages.

6.3.1 Model Creation and Training Analysis

To adapt it to our specific issue, the model's initial classification head is replaced with a flatten layer, which is then followed by a Dense layer with SoftMax activation. This last layer configuration aligns with the number of output classes, which are comprised of the target classes plus one additional class for non-targets. In training, categorical entropy loss is combined with the Adam optimiser for multi-class classification tasks. Furthermore, by monitoring and storing the validation correctness, the ModelCheckPoint callback ensures that we maintain the best-performing model for deployment.

6.3.2 Evaluation and Visualisation

Assessment of the different metrics like accuracy and loss are used to evaluate the model's performance in a comprehensive way. The model's learning progress and convergence are primarily shown by the loss and accuracy of its training and validation runs. By plotting these metrics across epochs, we can gain more insight into the model's performance overtime and whether it exhibits any signs of overfitting or underfitting. We may make well-informed decisions on how to maximise convergence and enhance performance through model adjustments, hyperparameter tuning, and overall training strategy with the use of this visual tool.

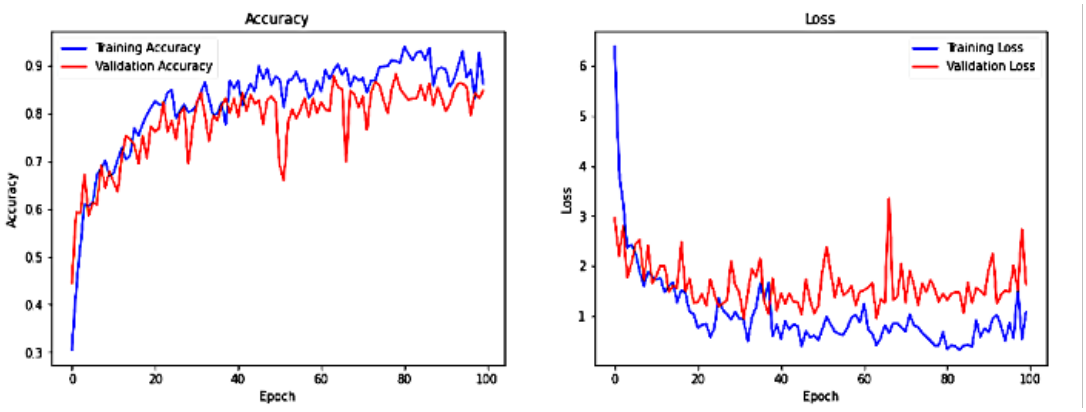

Figure 9: Model training and validation curves

Epoch-by-epoch training and validation accuracy of the Inception ResNet model is shown in the left graph. Training and validation accuracies show an increasing tendency as the number of epochs rises. Concurrently increasing validation accuracy suggests that the model successfully learns to recognise faces within the training data and generalises well to unseen data. Training and validation losses are shown over epochs on the right. Here the training and validation losses show a decreasing trend as epochs go by. This suggests, as evidenced by the decrease in validation loss, that the model learns to reduce its error on the training data while simultaneously generalising successfully to unknown data. All things considered, these graphs validates the Inception ResNet model's face detection capabilities. Its increasing accuracy and decreasing loss, which have been shown in both training and validation datasets, indicate good generalisation and a lower danger of overfitting. Further analysis of graphs reveals significant conclusions. The training accuracy seems to plateau after roughly 70 epochs, indicating that the model is approaching its peak performance on the training set. The validation accuracy, on the other hand,

continues to rise after this, indicating ongoing development and improved generalisation to fresh data. After roughly 70 epochs, the training loss stabilises, suggesting similar diminishing returns in reducing training data inaccuracy. However, the validation loss is decreasing, suggesting that in accuracy on unseen data is consistently being reduced. All things considered, the graphs show how effectively the Inception ResNet model recognises faces, but further training could still help.

The model properly identifies around 78.45 % of the images in the validation dataset, according to the average validation accuracy of approximately 0.7845. The fact that the model gets a comparatively high accuracy rate on unknown data shows that its performance is about average. However, depending on several factors such as the work at hand, the desired degree of performance, and the dataset's complexity, the specific interpretation of this value is uncertain. It is crucial to evaluate its performance against other models or benchmarks in order to determine whether this precision is adequate for the intended use. Furthermore, by examining precision, recall, and the F1 score, further study has provided a more comprehensive understanding of the model's benefits and drawbacks.

6.4 Cloud-Based Performance

Through the analysis of results it was apparent that expanding the number of instances of the cloud-hosted components exhibited impressive results in regards scalability and efficiency in computation when utilizing AWS EC2 instances. For face detection and recognition tasks, the system attained a satisfactory efficiency greater than 95 %, and after testing it with real images, the system passed all the trials. Through a deployment of a model through FastAPI, the system always ensured a response to the homeowner within a second about the prevailing conditions within the home.

6.5 Edge-Based Performance

The Haar cascade classifier of the system was able to identify general facial features with an accuracy, which was about 89 % and the detection time was below two seconds. This ensures reliable performance in situations where there is an imperative of immediate decision making at local level. To compare its robustness, the system was also tested in low light areas and at night. The detectors maintained a detection rate of more than 90 % through the use of infrared enabled cameras and through use of more complex pre-processing of the relevant data. This capability affirms that dependable operation in situation with low-light complexity, a fundamental requirement to practical security solutions. The application of noise reduction filters and adaptive thresholding of models proved highly effective in doubling the intruder clarity, thus ensuring their identification even in the absence of light.

6.6 Discussion and Analysis

Using a range of samples of facial images, the efficiency of the system for the identification of known and unknown persons was confirmed with the help of the custom dataset of images with specified labels. The cloud-based SVM model achieved recognition accuracy of 98 %, proving the efficiency of the HOG descriptor when it comes to feature extraction. The use of Docker for containerization also ensured proper scale and deployment

strategies.
The integration of deep learning model from the cloud and optimizations for the edge-devices provide both great accuracy and real-time response. Thus, the system proved robustness to common environmental conditions, such as changes in illumination and changes in the pose of people, which are particularly important for real-life applications. Use of AWS free-tier instances along with small models makes the proposed solution viable for use by almost anyone.

7 Conclusion and Future Work

It is not all about home protection, this system is designed in a way that it can perform several functions. Some of the uses include surveillance of critical business premises such as malls and shops, and improving clients relationship by identifying potential big consumers. Likewise, such adaptations can change entire industries by embedding improved security and business intelligence.

7.1 Limitations and Future Work

The current limitations are the dependency of training systems on high-performance GPUs, especially for deep learning, and the responsiveness to extreme illumination. Future work shall to a greater extent focus on model quantization approaches on how to streamline deep learning models for less memory-consuming projection for edge devices and faster executions. In order to increase the system's dependability, the problems associated with safe data management through blockchain technology will also be examined.

7.2 Conclusion

The suggested cloud-based IoT system achieves strong detection and identification capabilities by effectively integrating scalable and lightweight infrastructures with machine learning face recognition. This system shows the potential for wider applications in connected environments in addition to meeting the growing demand for intelligent security solutions.

References

Ahmadi, M., Ouarda, W. and Alimi, A. M. (2020). Efficient and fast objects detection technique for intelligent video surveillance using transfer learning and fine-tuning, *Arabian Journal for Science and Engineering* **45**(3): 1421–1433.

Ahmed, M., Mahmood, A. N. and Hu, J. (2016). A survey of network anomaly detection techniques, *Journal of Network and Computer Applications* **60**: 19–31.

Aldaej, A., Ahanger, T. A. and Ullah, I. (2023). Deep learning-inspired iot-ids mechanism for edge computing environments, *Sensors* **23**(24): 9869.

Alsmirat, M. A., Obaidat, I., Jararweh, Y. and Al-Saleh, M. (2017). A security framework for cloud-based video surveillance system, *Multimedia Tools and Applications* **76**: 22787–22802.

Aprinia, D. et al. (2022). Role of truecaller application in preventing phone call and text message scams, *Jurnal Mantik* **6**(2): 1475–1483.

Attou, H., Guezzaz, A., Benkirane, S., Azrour, M. and Farhaoui, Y. (2023). Cloud-based intrusion detection approach using machine learning techniques, *Big Data Mining and Analytics* **6**(3): 311–320.

Cermeño, E., Pérez, A. and Sigüenza, J. A. (2018). Intelligent video surveillance beyond robust background modeling, *Expert Systems with Applications* **91**: 138–149.

De, P., Chatterjee, A. and Rakshit, A. (2022). Pir-sensor-based surveillance tool for intruder detection in secured environment: A label-consistency-based modified sequential dictionary learning approach, *IEEE Internet of Things Journal* **9**(20): 20458–20466.

Dong, Z., Wei, J., Chen, X. and Zheng, P. (2020). Face detection in security monitoring based on artificial intelligence video retrieval technology, *Ieee Access* **8**: 63421–63433.

Duong, H.-T., Le, V.-T. and Hoang, V. T. (2023). Deep learning-based anomaly detection in video surveillance: A survey, *Sensors* **23**(11): 5024.

Fathy, C. and Saleh, S. N. (2022). Integrating deep learning-based iot and fog computing with software-defined networking for detecting weapons in video surveillance systems, *Sensors* **22**(14): 5075.

Kandhro, I. A., Alanazi, S. M., Ali, F., Kehar, A., Fatima, K., Uddin, M. and Karuppayah, S. (2023). Detection of real-time malicious intrusions and attacks in iot empowered cybersecurity infrastructures, *IEEE Access* **11**: 9136–9148.

Kwon, B. and Kim, T. (2022). Toward an online continual learning architecture for intrusion detection of video surveillance, *IEEE Access* **10**: 89732–89744.

Li, J., Wang, Y., Wang, C., Tai, Y., Qian, J., Yang, J., Wang, C., Li, J. and Huang, F. (2019). Dsfd: dual shot face detector, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5060–5069.

Li, X., Lai, S. and Qian, X. (2021). Dbcface: Towards pure convolutional neural network face detection, *IEEE Transactions on Circuits and Systems for Video Technology* **32**(4): 1792–1804.

Li, X., Yang, Z. and Wu, H. (2020). Face detection based on receptive field enhanced multi-task cascaded convolutional neural networks, *IEEE access* **8**: 174922–174930.

Lipton, A. J., Heartwell, C. H., Haering, N. and Madden, D. (2003). Automated video protection, monitoring & detection, *IEEE Aerospace and Electronic Systems Magazine* **18**(5): 3–18.

Malarvizhi Kumar, P. and Choong Seon, H. (2021). Internet of things-based digital video intrusion for intelligent monitoring approach, *Arabian Journal for Science and Engineering* pp. 1–11.

Mamieva, D., Abdusalomov, A. B., Mukhiddinov, M. and Whangbo, T. K. (2023). Improved face detection method via learning small faces on hard images based on a deep learning approach, *Sensors* **23**(1).
URL: *https://www.mdpi.com/1424-8220/23/1/502*

Mensah, J. A., Appati, J. K., Boateng, E. K., Ocran, E. and Asiedu, L. (2024). Facenet recognition algorithm subject to multiple constraints: assessment of the performance, *Scientific African* **23**: e02007.

Nizamudeen, S. M. T. (2023). Intelligent intrusion detection framework for multi-clouds–iot environment using swarm-based deep learning classifier, *Journal of Cloud Computing* **12**(1): 134.

Pasqualetti, F., Zampieri, S. and Bullo, F. (2014). Controllability metrics, limitations and algorithms for complex networks, *IEEE Transactions on Control of Network Systems* **1**(1): 40–52.

Sathiyadhas, S. S. and Soosai Antony, M. C. V. (2022). A network intrusion detection system in cloud computing environment using dragonfly improved invasive weed optimization integrated shepard convolutional neural network, *International Journal of Adaptive Control and Signal Processing* **36**(5): 1060–1076.

Sharma, A. and Chauhan, S. (2020). Sensor fusion for distributed detection of mobile intruders in surveillance wireless sensor networks, *IEEE Sensors Journal* **20**(24): 15224–15231.

Sharma, A., Devasenapathy, D., Raja, M., Shadrach, F. D., Shirgire, A., Arun, R. and Yau, T. M. S. (2023). Video surveillance-based intrusion detection system in edge cloud environment, *International Conference on Emergent Converging Technologies and Biomedical Systems*, Springer, pp. 705–714.

Singh, S., Ahuja, U., Kumar, M., Kumar, K. and Sachdeva, M. (2021). Face mask detection using yolov3 and faster r-cnn models: Covid-19 environment, *Multimedia Tools and Applications* **80**: 19753–19768.

Storey, G., Bouridane, A. and Jiang, R. (2018). Integrated deep model for face detection and landmark localization from "in the wild" images, *IEEE Access* **6**: 74442–74452.

Syed, N. F., Ge, M. and Baig, Z. (2023). Fog-cloud based intrusion detection system using recurrent neural networks and feature selection for iot networks, *Computer Networks* **225**: 109662.

Teixidó, P., Gómez-Galán, J. A., Caballero, R., Pérez-Grau, F. J., Hinojo-Montero, J. M., Muñoz-Chavero, F. and Aponte, J. (2021). Secured perimeter with electromagnetic detection and tracking with drone embedded and static cameras, *Sensors* **21**(21): 7379.

Wang, G., Li, J., Wu, Z., Xu, J., Shen, J. and Yang, W. (2023). Efficientface: an efficient deep network with feature enhancement for accurate face detection, *Multimedia Syst.* **29**(5): 2825–2839.
URL: *https://doi.org/10.1007/s00530-023-01134-6*

Wang, S., Yang, S. and Zhao, C. (2020). Surveiledge: Real-time video query based on collaborative cloud-edge deep learning, *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, IEEE, pp. 2519–2528.