Image and Vision Computing 120 (2022) 104406

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A neural network aided attuned scheme for gun detection in video surveillance images

V.P. Manikandan [a], U. Rahamathunnisa [b,*]

[a] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India
[b] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

A R T I C L E I N F O

Article history:
Received 25 November 2021
Received in revised form 28 January 2022
Accepted 12 February 2022
Available online 16 February 2022

Keywords:
Classification
CNN
Feature extraction
Deep learning
Object detection

A B S T R A C T

Closed Circuit Television (CCTV) cameras are installed and monitored in private and open spaces for security purposes. The video and image footages are used for rapid actions, identity, and object detection in commercial and residential security. Object and human detection require different classifications based on the features exhibited from the static/ mobile footages. This article introduces an Attuned Object Detection Scheme (AODS) for harmful object detection from CCTV inputs. The proposed scheme relies on a convolution neural network (CNN) for object detection and classification. The classification is performed based on the Object's features extracted and analyzed using CNN. The hidden layer processes are split into different feature-constraint-based analyses for identifying the Object. In the classification process, feature attenuation between the dimensional representation and segmented input is performed. Based on this process, the input is classified for hazardous objects detection. The consecutive processing layer of CNN identifies deviations in dimensional feature representation, preventing multi-object errors. The proposed scheme's performance is verified using the metrics accuracy, precision, and F1-Score.External dataset training has improved accuracy by 8.08% and reduced error and complexity by 7.47 and 8.23 percentage points, respectively, in this process. Object classification based on labels is expected to be implemented in the future.

1. Introduction

Object classification is one of the main tasks to perform in computer vision. Object classification main goal is to predict the class of an image from the given image. Computer vision is widely used in many applications [1]. Computer vision helps the computer and software to get information about an image or video digitally. Computer vision trains the software to perform a particular task with the images' help in less time [2]. The main aim of object classification is to identify the features of an image with more accuracy. Object classification combines object localization and object detection to classify objects in an image or video [3]. Object localization is to locate the objects. Object detection is used to detect the details of an object in an image. Encoding objects play a vital role in computer vision technology. One of the most widely used features of object classification is a color histogram and scale-invariant feature transform. The histogram model is more robust than another model [4,5].

Hazardous detection identifies the people affected by a natural disaster and helps find damaged objects from a video or an image. Hazardous object detection is a challenging task to perform. Accurate object detection and locating a particular object are the two main detection features [6]. To ensure the safety of a particular place, CCTV is widely used in hazardous Object detecting processes in the whole world. With the help of CCTV, the quality of service, accuracy of detecting an object are increased [7]. A region-based convolutional neural network (RCNN) is a deep learning algorithm that detects objects after a natural disaster [8]. RCNN is more efficient in service, and the accuracy rate is high compared toanother deep learning algorithm. The following are some of the drawbacks of using R-CNN to detect objects: Even though the number of regions per image has been reduced to 2000, it still takes a long time to classify the regions. Real-time applications cannot be implemented because the prediction time for a new test image is around 47 s.Deep HAZMAT is a detecting algorithm used to detect objects by utilizing the images or video inputs with the help of the CNN network. HAZMAT is a challenging task to perform in computer vision technology to detect the details of objects from a video [9,10]. The CNN algorithm has been used in numerous real-world applications, including cancer detection and biometric authentication. CNN could be used for both image captioning and visual question answering. Input images are analyzed by CNN networks, generating natural language responses to the images.

* Corresponding author.
E-mail address: rahamathu.u@vit.ac.in (U. Rahamathunnisa).

https://doi.org/10.1016/j.imavis.2022.104406
0262-8856/

Object detection is the process of locating an object from an image or video with the help of computers and software. Deep learning is a part of the artificial neural network, which helps understand the system on its own [11]. Deep Neural Network (DNN) is one of the machine learning techniques used in object detection and classification. DNN helps to identify the objects without any hand designs features. For tasks like object recognition, DNN models perform at a human-equivalent level. Recent advancements have made it possible to test the hierarchical similarity of object representation between humans and neural networks. However, it is unclear how DNNs represent object examples in a single category. The geometric similarity between visual features extracted from high-level layers of various DNNs is computed to compare the ability of DNN models to explain invariant object representations across categories.

DNN based regression is used in object detection to detect low-resolution images and helps to capture detailed information about the objects [12]. A convolutional neural network (CNN) is a deep learning technique used to detect objects from an image and analyze the visual imagery of an image. CNN uses a pooling layer to reduce the size of the pixels and helps to find maintain the patch of the pixels [13]. Support Vector Machine (SVM) is used to classify objects in an image or video. An image-detection technique known as convolutional neural network (CNN) is used to analyze the visual imagery of an image. Pixels can be reduced using a pooling layer, which helps find and maintain the patch of pixels. The feature maps can be made smaller by combining multiple layers. That means fewer learning parameters, which means less work for the network. A convolution layer generates a feature map. The pooling layer sums up the features in that region. a firm grasp on the Support Vector Machine (SVM) is a tool for image and video classification. Classification can be linear or non-linear. Non-linear classification is implemented using a set of algorithms known as ‘Kernel methods. For pattern analysis, the kernel trick maps inputs into a higher-dimensional space. An algorithm called the Neural Support Vector Machine (NSVM) combines neural networks and support vector machines to learn (SVMs).

NSVMs use SVMs that take a central feature layer as an input to produce output. This makes them ideal for datasets with a large number of features.SVM is a learning algorithm that helps analyze the patterns and information of an object in images. SVM is more effective and efficient in dimensional spaces.When dealing with large datasets, SVM is out of place. SVM performs poorly when the data set contains a lot of noise, such as overlapping target classes. The SVM performs poorly when the number of features for each data point exceeds the number of training data samples. Classification, regression, and outlier detection can all be accomplished with the help of support vector machines (SVMs), a collection of supervised learning techniques. Support vector machines offer the following benefits: Effective in three-dimensional environments. It's still useful even if there are more dimensions than samples [14]. You only look at once deep learning technique, which combinesdeep learning and python language to detect Objects from an image. This technique is mostly used to classify the objects details or information of an image for efficient detection [15].

The main contributions of the paper are,

- Different classifications are needed for Object and human detection based on the characteristics of the static/mobile footages.
- To detect harmful objects from CCTV feeds, and Attuned Object Detection Scheme (AODS) is introduced to the system.
- Object features are extracted and analyzed using CNN to classify the subject matter.

2. Related works

CNN-based object detection models in general and dimensional feature detection in video images are briefly reviewed before a description of the performance metrics used to evaluate these detection models is given in this analysis.

To analyze the challenges in multi-object detection, Wang et al. [16] proposed a multi-scale convolutional neural network (MSCNN) approach, mainly used to eliminate 3D poses from an object. In MSCNN, rectilinear images are converted into panoramic images with the help of the CycleGAN network. MSCNN is more accurate in detecting panoramic images, and the detection speed is also high.

Bhatti et al. [18] proposed a new approach to detect illegal activities based on binary classification (BC) with deep learning (DL) approach. This is mainly used to detect harmful weapons in real-time surveillance cameras by implementing a deep learning algorithm. Compared with the existing model, the proposed method has high accuracy in detecting weapons.

Olmos et al. [19] proposed a new model using the image fusion approach to reduce the rate of false positives in object detection. It helps reduce the cost of the process and improve the accuracy rate of detecting an object. When experimented with another existing model, the proposed method reduces the false positive rate and increases the process's overall performance in detecting the objects in surveillance videos. A critical step in video surveillance is determining which cameras or groups can see a particular problem area. Analyzing past and future security trends for patterns can aid investigators in their work. Discouragement and documentation of crimes and injury claims by customers or employees can be achieved by using a good video surveillance system. As technology improves, many digital cameras now allow you to review footage from a smartphone or tablet even if it isn't present. It is easy to abuse surveillance systems, and there is a risk that cameras installed in public places will be abused. Public camera footage can be used by law enforcement to blackmail someone.

Castillo et al. [20] introduced a new model using a convolutional neural network (CNN) to detect cold steel weapons. The proposed model has a high inaccuracy rate in detecting cold steel weapons in surveillance video compared with the existing model.

Galab et al. [21] projected an automated knife detection approach based on a deep learning algorithm to enhance the frame's brightness in surveillance coverage areas. By enhancing the brightness of the frames, detecting knives will be increased efficiently. Experimental results increase the performance efficiency in detecting the weapons compared with the existing model. Using hazard detection techniques, video or image captures of natural disasters can locate victims and damaged objects. A difficult task is detecting potentially harmful objects, and it is essential to identify and locate specific objects to detect hazardous materials accurately. Our tests discovered that the system could detect stationary and moving hazards with precision. In addition, as part of our research, we performed experiments to detect potentially hazardous objects in motion.These include HAZOP, fault tree analysis, safety audit, failure mode and effect analysis, CHAZOP, and task analysis. There is a lot of interest in these methods for identifying potential hazards, but they can be time-consuming and difficult to perform correctly.

To reduce the false positives,Vallez et al. [22] introduced an autoencoder(AE)method based on a deep learning approach, which is used to detect handguns in surveillance videos. The proposed method reduces the false-positive rate in object detection compared with the existing models. An image or video can be searched for specific instances of an object using the computer vision technique known as “object detection.” Machine learning and deep learning are typically used for object detection algorithms. Because computers are programmed to mimic human intelligence, object detection is critical. Defining the target classes and training the model to recognize them using labeled examples is a supervised learning problem for image classification. Raw pixel data was the initial input for computer vision models. Supervised machine learning requires that models be trained on labeled examples in object detection. The boundaries and classes of the objects in each training dataset image must be recorded in a separate file.

Velasco-Mata et al. [23] projected a method to detect handguns by analyzing the poses of humans in public places. This method uses grayscale images to detect handguns, and the proposed method improves the efficiency of handgun detection.

Shahbaz et al. [24] proposed efficient foreground detection (EFDNet) algorithm based on a convolutional neural network (CNN). EFDNet is faster to train the objects. EFDNet is compared with the change detection dataset (CDNet); EFDNet increases the performance rate in foreground detection and reduces the time for processing.

Liu et al. [25] presented an anomaly detection of aerators based on computer vision in a real-time system. To build a feature classifier, a deep learning technique is used. Based on the region proposal idea, small object region detection is presented. Region detection and working state detection are the two modules of an expert system. The proposed method is robust, and anomaly detection of aerators is more accurate in the real-time expert system when compared with the existing model.

Park et al. [26] designed an object detection based on the estimation of multi-person poses. This method improves the performance of the overall service. Key point estimators and object detectors are widely used in real-time expert systems by analyzing multi-person poses. The proposed method is high in performance rate and accuracy rate of detecting an object compared with the state-of-the-art model.

Rodda et al. [27] proposed a new method to detect the suspicious activities of people in public places by using IoT technology. This proposed methodology is used to understand the activities of unknown persons and capture the images of people under the surveillance area. This methodology increases the accuracy rate of the suspicious activities of a particular person.

Mohtavipour et al. [28] proposed a new method named deep violence detection framework using handcrafted features based on a deep learning approach. The proposed deep violence detection method is more accurate in detecting violent activities, and time consumption is reduced compared with the state-of-the-art algorithm.

For efficient anomaly detection, Maqsood et al. [29] introduced a framework in a real-time system to detect the anomalies from videos. Detecting anomalies is based on handcrafted features and a deep learning approach. The proposed method is more effective and efficient in identifying videos' anomaliesthan the existing models.

The preceding review has shown a clear gap in the existing body of work in visual image gun detection. It is noteworthy that there is no way to obtain contextual information from CCTV images to represent earning in a surveillance context. The following sections describe object detection methods used for detection and the development of datasets that only include images from a surveillance perspective.

3. Attuned object detection scheme

A method using CNN Attuned Object Detection Scheme (AODS) is initiated for harmful object detection from CCTV inputs. The feature-constraint-based analysis extracts the feature of monitoring images and identifies the objects in that region. This work proposes the variation in dimensional feature representation and prevents multi-object errors. A CNN Attuned Object Detection Scheme detectsCCTV inputs (AODS) objects. Features from monitoring images can be extracted, and the objects in that area identified using a feature constraint-based analysis. To avoid multi-object errors, we present the idea of changing how dimensional features are represented. Onecould try a naive approach to solving this problem by using a CNN to classify the presence of an object in different regions of interest in a photograph. This method has some drawbacks because of the possibility of different aspect ratios and spatial distributions within the image. It is, therefore, necessary to select a large number of regions, which could lead to a huge computational burden. This has led to algorithms like R-CNN and YOLO that quickly find these occurrences. In Fig. 1, the AODS process is presented.

The proposed AODS method detects the dangerous objects in the CCTV footage. CCTV works under the conditions ofcapturing a constant sequence of images n transmitted over fiber cable or wirelessly to the recording equipment and then on to the monitor, which is independent of seeing the consequent images as video footage.To preserve the information in the original data set, feature extraction involves converting raw data into numerical features that can be processed. Compared to using machine learning directly on the raw data, this method produces superior results.Using a linear representation, topological groups can be represented in a finite-dimensional vector space. Finite-dimensional representations have seen the most development as a subfield of representation theory.Useful for image segmentation, adaptive compression, and regional image retrieval applications, detecting salient areas in images is an important step in these processes. Using this technique produces saliency maps with the same size and resolution as the input image quickly with minimal implementation hassles.As part of a feedback loop, predictions are made at the semantic level, and verification is sought at the image level. The system includes a detailed human body model, making it possible to conduct a model-driven human motion analysis. Every pixel in the image is decoded using an image-based constraint network based on the human model's internal structure.Images can be classified using rules for categorizing and labelling specific groups of pixels or vectors. Categorization laws can be derived from spectral or textural properties.

Detecting the edges of objects in images is a common image processing technique known as edge detection. Because it detects brightness variations, it's effective. Computer vision and machine learning applications use edge detection to segment and extract information from images.A Closed-circuit Televisioncontains acamera, lens, monitor and recorder. The camera captures a sequence of images, sending and receiving a signal to storethe device and monitor. In this manner, capturing a sequence of images at the same interval of time will lead to multi-object error. The above sequence of images is analyzed using restricted deep learning to overcome this scenario. In this scenario, the region detection is to be converted into controls that must be changed in appropriate and accurate time instances to improve the working of all the devices.

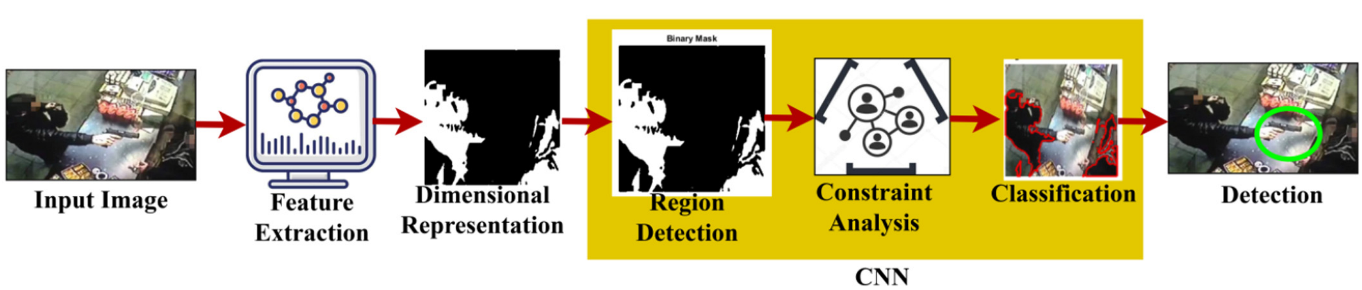


Fig. 1. AODS process.

4. Detection scheme

This detection scheme describes the object identification pursued in the proposed scheme. Itidentifies Objects and classification using CCTV camera inputs.This proposed work aims to obtain CNN by utilizing a feature-constraint-based analysis.Further improvement was achieved using our model, increasing the F1 score 1(c), precision 1(b) and metric accuracy 1(a) of the images. The following equation represents the proposed objective.

$$\left.\begin{array}{l} f(a,b)_{\ ext(i)} = \begin{cases} -x, f(a,b)_i > f_{\ det(i)} \\ x, x\geq 0 \end{cases}, \text{Metric accuracy } 1(a) \\ f(a,b) = c_0 + \sum_{i=1}^{\infty}\left(a_n \cos\frac{n\pi(-x)}{f_{\ det(i)}} + b_n \sin\frac{n\pi x}{f_{\ det(i)}}\right), \text{Precision } 1(b) \\ f_{\ det(i)} = \frac{-c_0 \pm \sqrt{f(a,b)^2 - a(i)}}{2\ \cos a_n}, F1\ score\ 1(c) \end{array}\right\}$$

The above Eq. 1(a) the aim of the proposed work is evaluated, $f(a,b)_{ext(i)}$. This equation allocates a value for each pixel in the present frame taken in the earlier frames at the exact pixel neighborhood. In this model, two different approaches towards CNN. Where the image 'a' video 'b' for the input image $f(a,b)$, and $f_{det(i)}$. Object detection is based on feature extraction, and the CNN is an image of either feature extraction or variable dimensional representation generated by the region detection in the earlier stage. Then the footage is resized to a predefined size and value into the CNN for classification.

The consecutive processing layer of CNN identifies deviations in dimensional feature representation, preventing multi-object errors. The proposed work addresses the multi-object error and increases by deriving $f(a,b)_{ext(i)}$ where $ext(i)$ denote extraction, and the classification is based on the Object's extracted and analyzed features. The extracted features' dimension representation is derived using the following Eq. (2).

$$\left.\begin{array}{c} \left(h(a,b)_{ext(i)} + h_{det(i)}\right)^{n(i)} = \sum_{c_0=0}^{h}\binom{ext(i)}{b} \times h^{(i)}a^{n-x} \\ \binom{ext(i)}{b} = \left\{h^{(i)}a^{n-x}/_{n(i)} \middle| h(a,b)_{ext(i-n)}\right\} \end{array}\right\} \quad (2)$$

The prediction $(h(a,b)_{ext(i)} + h_{det(i)})^{n(i)}$ is processed in the above Eq. (2) in this $\binom{ext(i)}{b}$ It performs the featured extraction and dimensional representation and retrieves the CNN region detection. This method applies to different image input processing stages without increasing the complexity. This equation analyzes the feature representation and dimension extraction of 2D video pixels and pixel-to-region estimation.Fig. 2 presents the dimension analysis process for sample input.

The input image is differentiated for different features based on the $\binom{ext(i)}{b}$, in identifying pre-dimension inputs. The identified inputs are further analyzed for the available dimensions preventing errors. The extracted dimensions are used for analyzing constraints.It is possible to determine the dimensions of an object and its volume by using conventional models for everyday objects and a prediction model based on regression analysis for human bodies using Haar Cascades. By comparing the measured values to the actual ones, a 93.69% accuracy is achieved. We introduce constraints analysis based on the region that does not provide explicit negatives. The common way is to differentiate one pixel fromitsearlier. The similarity between the pixels is compared for the identified regions without varying the properties. The different features are based on textural and color features exhibited by the image. The following Eq. (3) represents region detection from CCTV inputs.

$$\left.\begin{array}{c} f_{\ det(i)} = \prod_{sin^b}^{ni}\left(c_0 + x^i - a_0\right) * \left(\frac{ext(i)}{b}\right) + h(a,b)_{\ ext(i-n)} \\ h(a,b)_{\ ext(i-n)} = \prod_{sin^b}^{ni}\left(c_0 + x^i - a_0\right) \\ \prod_{sin^b}^{ni}\left(c_0 + x^i - a_0\right) = \begin{cases} -x, & f(a,b)_{\ ext(i-n)} > f_{\ det(i)} \\ x, & f(a,b)_i \geq 0 \end{cases} \end{array}\right\} \quad (3)$$

In this above Eq. (3) c_0 is denoted as region detection.The problem is finding and classifying variable objects on an image. If the input is represented in two dimensions, let the feature representation be denoted as '$-x$' and 'x', finding the image size and variables. In this evaluation $\prod_{sin^b}^{ni}(c_0 + x^i - a_0)$. The computation is determined by analyzing if the captured input image is a two-dimensional image or any other dimension; it will check and show the original size of the image by processing the feature extraction. Each image in the dataset will be in different sizes and variables. Only the necessary images are

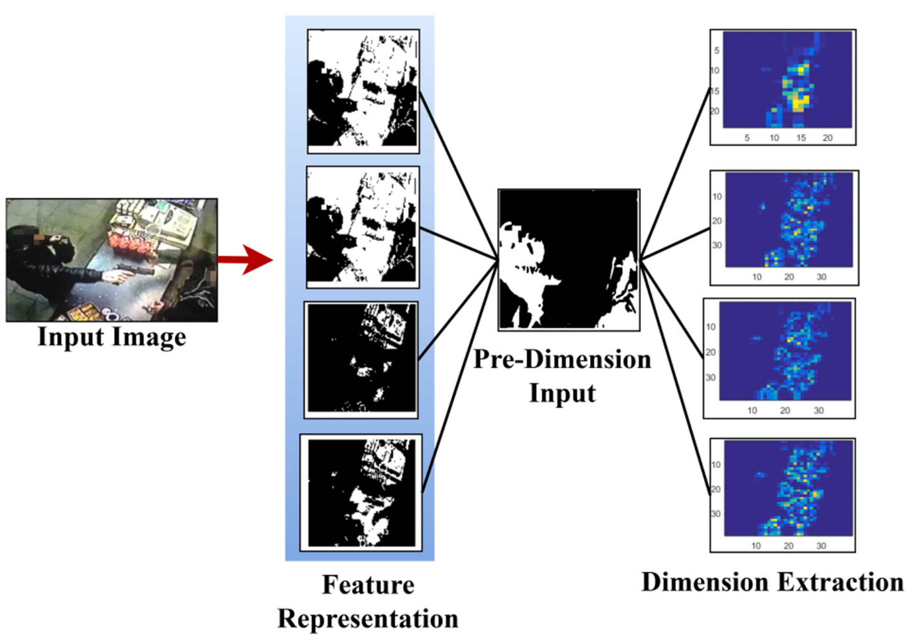


Fig. 2. Dimension analyses for a sample input.

observed and monitored in this manner. The time is less consumed and improves detection performance. The following Eq. (4) represents the limitations of the classification.

$$\left.\begin{array}{c}\left(x^{i}+a_{0}\right)^{n}=\sum_{x=0}^{-x}\left(\frac{f(a,b)_{i}}{(c_{0}+x^{i-n}-a_{0})}\right)x^{i}a^{n-c_{0}}\\ \text{such that}\\ \left(x^{i}a^{n-c_{0}}\right)+\binom{ext(i)}{b}=L^{c}\end{array}\right\} \quad (4)$$

The above Eq. (4) specifies the constraint analysis about the region detection. In image segmentation, adaptive compression, and region-based image retrieval, salient regions in an image can be identified and exploited. Saliency maps of the same size and resolution as the input image can be generated using this method.In this, the image from the region detection is a filter and process for pursuing data in CNN. In this equation where.'L^{c}' represented as limitations over the CNN. It filters out unwanted duplicate images and videos in the network. At this moment, computing $(x^{i}+a_{0})^{n}$ It examines the dataset in which it acquires the region detection over the constraint analysis. The duplicate images and videos will require classification produce constraints and difficult to proceed with the following process. The dataset consists of negative and positive images. The following Eq. (5) reduces the classification and time constraint.

$$\left.\begin{array}{c}\sum_{n_{x}}^{i-n}\left(\frac{n(i)(C_{p})*\beta}{det(i)+(c_{0})}\right)+\left(\frac{\alpha+n'}{\beta(x^{i})}\right)-t_{n}, \text{classification constraint}\\ \alpha=\frac{n\pi x}{f_{det(i)}}(\beta)\\ \sqrt{\frac{n'+C_{p}}{s_{0}/\alpha}}+\prod_{s_{0}}(C_{p}-L^{c})+\left(\frac{2\cos a_{n}}{\alpha}*ext(i-n)\right), \text{Time Complexity}\\ \beta=n'(det(i)+\alpha)*^{n\pi x}/_{C_{p}(s_{0})}\end{array}\right\} \quad (5)$$

In the above Eq. (5), constraint in the network is analyzed, where $\left(\frac{n(i)(C_{p})*\beta}{det(i)+(c_{0})}\right)$. It computes to find the inconvenience in the storage of the CCN. It is possible to compute the value of any one variable given the values of all the others in a constraint network, which represents a mathematical relationship between several variables. A constraint network has two types of nodes: cells and constraints. A cell can be used to store or remove a number. The image-level tags constrain the output labelling of a Convolutional Neural Network (CNN) classifier. Using a new loss function, we develop Constrained CNN (CCNN), a method that uses a new loss function to optimize for any set of linear constraints on a CNN's output space. The constraint analysis is a way to limit the region, and it follows some functions to search the selected feature. Thus the expected processing as $\left(\frac{\alpha+n'}{\beta(x^{i})}\right)$. Where α inconvenience of the feature is determined to access the region, the imageinput is dimensionally reduced and fed for classification using CNN. In this context, the predefined dimensions vary under different classifications, preventing overloaded analysis. The input images from the datasets are used for constraint-analysis-based region detection is used at the classification. In this manner $\sqrt{\frac{n'+C_{p}}{s_{0}/\alpha}}+\prod_{s_{0}}(C_{p}-L^{c})+\left(\frac{2\cos a_{n}}{\alpha}*ext(i-n)\right)$ Are derived from decreasing time limits in CCN and retrieving the problems. In Table 1, the region detection process for sample inputs is presented.

In Table 1, Region Detection for Sample Inputs figures 4th row 3rd column NA is because it has a single side dimension; therefore, dimension 1 analysis gives sufficient results while investigating the image.The action sequence is continuous concerning the available session time. The estimation of the above time limits is to be performed at t_{n} such that $\left(\frac{\alpha+n'}{\beta(x^{i})}\right)-t_{n}$. It is performed the varying instances. Therefore in this time, the remaining works are to be discharged to a machine with β such that the time for processing multiple works is coherent. This coherency in output achieving for process substitute works without lagging. Therefore, the rate of learning and processing instances, along with the classification of $n'(det(i)+\alpha)*^{n\pi x}/_{C_{p}(s_{0})}$. It is necessary to improve the rate of processing. The following Eq. (6) represents the process of classification.

$$\sum_{v=0}^{-u}a^{n-c_{0}}=\prod_{\beta}^{\alpha}(c_{0}+i-C_{p})*\left(\frac{n(i)}{b_{n}\sin\frac{n\pi x}{f_{det(i)}}(x^{i})}\right)+\left(\frac{L^{c}}{\alpha}*s_{0}\right), \text{if } i=n \quad (6a)$$

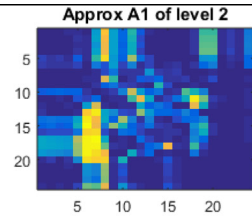
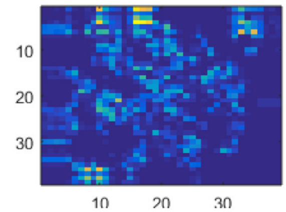
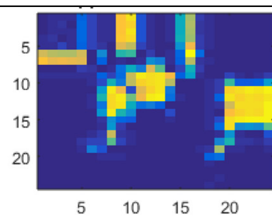
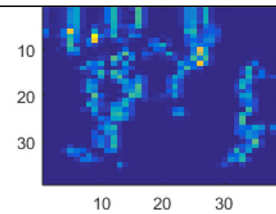
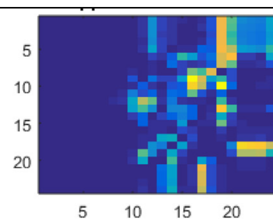
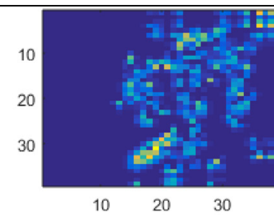
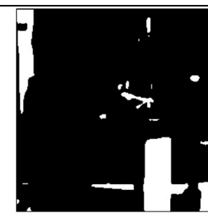
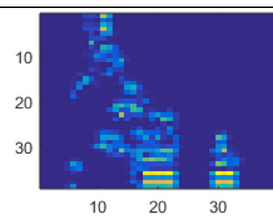
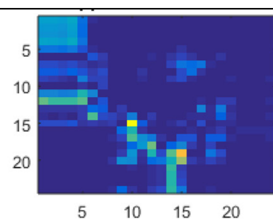
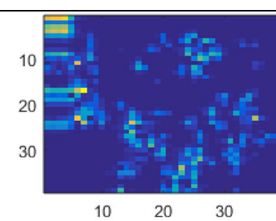
$$\left(\frac{n(i)}{b_{n}\sin\frac{n\pi x}{f_{det(i)}}(x^{i})}\right)+\left(\frac{L^{c}}{\alpha}*s_{0}\right)=\frac{n(i)(C_{p})}{\alpha-\beta}, \text{if } i=0 \quad (6b)$$

The classification process is analyzed in the above Eq. (6) series. Compared to other image classification algorithms, CNN's use a surprisingly low amount of preprocessing. There are several layers to a CNN, none of which can be seen by the end-user. Convolutional, ReLU, pooling, and fully connected layers are some of the more common ones found in deep learning algorithms. It is possible to categorize an image by applying specific rules to groups of pixels or vectors. Inferring categorization rules from spectral or textural properties is a viable option. There are both supervised and 'unsupervised' ways to classify data.The region can be split into two classified and unclassified objects in this process. According to the equation, $\sum_{v=0}^{-u}a^{n-c_{0}}$ where $-u$ is the classified objects, v is the unclassified Object, $n(i)$ is the total number of objects, C_{p} is the rate of processing and $\frac{n(i)(C_{p})}{\alpha-\beta}$. The proposed scheme relies on a convolution neural network (CNN) for object detection and classification. In the classification process, detection and extraction based deep learning methods are employed. In this classification based extraction learning, the probability of inequity $a^{n-c_{0}}$ and the errors $\frac{n(i)(C_{p})}{\alpha-\beta}$ are classified based on the output. This analysis presents the variation in $i=n$ and the stored data at the time of iteration. Therefore, at the initial state, the output of $i=1$ is independently classified.

Let the detection be performed concerning the value of'i', due to which the changes are unprocessed over there. This differential classification is performed by segregating the sequence and the instance through linear representation in Eq. (6) series. Classification of objects 'i' is to generate all possible solutions as in Eq. (6a). By satisfying the condition in Eq. (6b), the classification relies on $i=n$ and $i=0$ of the analyzing instance by computing the changes in the value of $i=n$ with respect to$i=0$. In this manner, the value of i will be 1, and then it processes the function and reach the detection stage. If the value of i will is 0, it will be terminated. Fig. 3 presents the classification using CNN.

The represented feature is split based on dimensions extracted, and the further classification occurs as u and v, for i and I detection. This identification ensures the quality of n, in finalizing a region. The end of a convolutional neural network contains one or more fully connected layers. When two layers are fully connected, every node in the first layer is connected to every node in the second. Classification of data is the sole purpose of the convolutions' features. Using the CNN algorithm for image processing is a wise decision. At the moment, these are the most effective algorithms for automating image processing that we have. Many businesses, including those specializing in image recognition, employ these algorithms. RGB data is used to create images.The final un-masked region in the n is identified as an object. As represented in Eq. (6a) holds lesser possibilities as the occurrence compared to Eq. (6b). The sequential occurrence of $i=n\neq i=0$ is not possible. Where $i=0$ is the terminating condition used to find the unclassified objects in the CCN, it will work differently. In this extraction analysis, to ensure that $i=n$ is validated based on the previously shared information in feature extraction is retrieved for analysis. Linear representation of Eq. (6) series separates the sequence and the instance in this

Table 1
Region detection for sample inputs.

Input	Dimension 1	Dimension 2	Detected Regions
			
			
			
		NA	
			

differential classification procedure. The change in i = I is subjected to a differential analysis method to improve the classification. Due to this procedure, increased accuracy and warnings are provided to the unclassified Object. To arrive at an estimate, the stored data specifies various possible videos. The filtered output is shown in Eq. (8).The analysis follows the determination of $i = I$ as

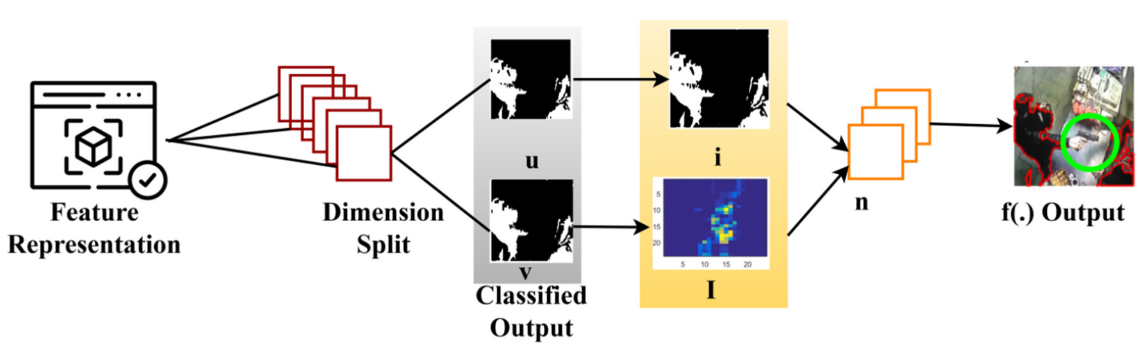


Fig. 3. Classification using CNN.

$$\left.\begin{array}{c}\sum_{C_p=0}^{a,b}\left(\frac{L^c}{\alpha}*s_0\right)a^{n-x}=\left(f(a,b)_{\ ext(i)}+L^c\right)^{n(i)}, i=1\\ such\ that\\ \sum_{C_p=0}^{a,b}L^c=\beta\left(\frac{\alpha+n'}{\beta(x^i)}\right)\end{array}\right\} \quad (7)$$

In the above Eq. (7), different training instances are classified in the previous time. The change in $i = I$ is analyzed in a differential manner to elevate the classification process. This helps increase detection accuracy and notify the unclassified Object with warnings. The estimation is preceded by defining different possible videos with the stored information. The filtered output is illustrated in Eq. (8);the consecutive processing layer identifies deviations in dimensional feature representation, preventing multi-object errors.The following Eq. (8) represents the filtered output is illustrated.

$$\left.\begin{array}{c}f\left(\prod_{s_0}\left(\complement_P-L^C\right)\right)=c_o+\sum_{n=1}^{i}\left(a_n\left(\frac{n(i)}{b_n\ \sin\frac{n\pi x}{f_{\ det(i)}}\left(x^i\right)}\right)\right)\\ such\ that\\ \left(\frac{n(i)}{b_n\ \sin\frac{n\pi x}{f_{\ det(i)}}\left(x^i\right)}\right)\alpha+\beta\left(\frac{\alpha+n'}{\beta(x^i)}\right)=2\frac{n'+C_p}{{}^{s_0}/_{\alpha}}\end{array}\right\} \quad (8)$$

In the above Eq. (8), the analysis is evaluated by $a_n\left(\frac{n(i)}{b_n\ \sin\frac{n\pi x}{f_{\ det(i)}}(x^i)}\right)$. In this, the analysis is obtained by the classification from the feature extraction, and it is evaluated as $\beta\left(\frac{\alpha+n'}{\beta(x^i)}\right)$It filters the unwanted duplicated information in the CNN network. The reliability of the architecture relies on the good quality videos and images delivered and multi-object errors without losing videos. This architecture is designed for improving the dimensional representation of the images and overcome the limitations in constraint analysis. Therefore, we automatically identify the classified images in this research by applying the data from CCTV camera image feeds. Object and human detection require different classifications based on the features exhibited from the static or mobile footages. In Table 2, the final output for sample inputs is presented.

An image database is employed for both training and testing. Various camera types and resolutions are used to reduce the feel of these pictures. Because the solution to this problem is unique, some metrics were created to reduce false positives. The following sections described object detection methods and datasets that only include images taken from a surveillance perspective.

In Table 3, the error for different frame intervals is analyzed.

Table 3 presents the classification possibility and error for different input from intervals. The proposed scheme relies on different frame intervals for maximizing detection accuracy. This is abrupt depending on the regions identified; the classification increases with the detected regions. As the classification distinguishes $-v$ and u, the error is comparatively less in different α mitigation instances. Table 4 presents the α for different Objects, and Classifications is presented.

The α observed for different object detection is tabulated in Table 4. The classifications are high for increasing object count for which accuracy is high. The α value has a direct impact on the accuracy; the high is the α, then the accuracy is less. Therefore, the classifications are reinstigated from $i = 0$ to $i = 1$ or I. This is pursued until $\alpha < 0$ is achieved. Contrarily, if $\alpha > 0$ is high, then accuracy is less, and the training iterate is required high. For different classifications, the segregation of error and α, reduces the complexity and augments accuracy. Classifications of computational problems are called “complexity classes.” Computational resources like time and memory are used to determine how difficult it is to solve the problems contained within these systems. Comparing F1 scores and accuracy, we can say that: F1-score (False Negative) and Accuracy (True Positive) are used when the True Positive and True Negative are more important. There are low false positives and low false negatives, which means that you can correctly identify the real threats and not be disturbed by false alarms. An F1 score of 1 is considered ideal, and an F1 score of 0 indicates that a model is an abject failure.

5. Performance assessment

The proposed scheme's performance is analyzed using MATLAB experiments incorporating the dataset from [30]. This dataset provides different videos and images of harmful Objects (guns) observed in CCTV footage. Besides, 4 videos and 850 images are used for training and classifying the inputs. An image is analyzed by extracting a maximum of 11 regions under 800 training instances. The metrics accuracy, precision, F1-Score, error, and classification complexity are compared for verifying the proposed scheme's efficiency. Comparative analyses are considered the existing AE + SVM and AE + KNN from [22] and BC-DL from [18].

6. Accuracy

In Fig. 4, the comparative analysis for accuracy for different regions and training instances. The proposed scheme achieves fair accuracy by addressing the constraints in region detection. In this validation, $\left(\frac{\alpha+n'}{\beta(x^i)}\right)$ is analyzed for different t_n for preventing duplicate inputs and inconvenience. Therefore, for the classification, u is predominant based on the input dataset. This case is different for the pursuing training instances, where $-v$ and $-u$ is identified and is matched with the previous region detected. The classification range is modeled for $\sum_{v=0}^{-u}a^{n-c_0}$ For preventing further constraints. This is performed until $i = 0$ or $i = n$ is observed; the contrary process verifies $i = I$ instances for preventing hiking errors. Therefore, the subsequent training instance is error-free until$i = 0$. The final pre-detected output is based on $f\left(\prod_{s_0}\left(\complement_P-L^C\right)\right)$ satisfying $\beta\left(\frac{\alpha+n'}{\beta(x^i)}\right)$ Indifferent iterations. If the region increases, then $\left(\frac{\alpha+n'}{\beta(x^i)}\right)-t_n$. It is validated for identifying the last known error in the identified region. The further error is jointly addressed based on the feature dimension, preventing accuracy failure. This is retained for multi-feature and multi-region inputs distributed between consecutive dimensions. Therefore from a dimension and feature perspective, the accuracy is retained through consecutive iterations.

7. Precision

The precision is estimated as computed in Eq. (1) using the different dimension analysis. The dimension segregated for different $f(a,b)_{ext(i)}$ such that the features are identified in different possible regions. Based on the representation, the $\binom{ext(i)}{b}$ based on prediction is first analyzed. In this analysis, the unclassified dimension features are assessed as $\prod_{sin^b}^{ni}(c_0 + x^i - a_0)$ for the first dimension observed. The further dimension analysis is presented as $(x^i + a_0)^n$ for reducing the classification constraint. Before the commencement of the classification, the identified constraints are mitigated. Therefore, a constraint-free classification is performed wherein the α is mitigated. If an inconvenience factor in the classified region is observed, then second computation of $\left(\frac{2\cos a_n}{\alpha}*ext(i-n)\right)$ is performed. This requires modified classification preventing lag in feature processing. Therefore the inequity is estimated by computing errors in different iterations. This is unanimous for different regions identified and training instances maximizing precision (Refer to Fig. 5).

Table 2
Sample output for inputs.

Input	Classified Output	Filtered Region	Output

Table 3
Error for different frame intervals.

Frame interval (s)	Regions	Classification possibility	Error
2	1	0.23	0.12
3	5	0.41	0.108
4	8	0.58	0.081
5	7	0.45	0.092
6	9	0.56	0.074
7	11	0.63	0.054

8. F1-score

The F1-score for different inputs is high under training instances and regions identified. This is achieved by performing classification based on prediction. The first α detection instance instigates the classification in two cases, i.e. high and less possibility. In the high classification possibility, $^{nnx}/_{C_p(s_0)}$ is analyzed for identifying $i = n$. This retains the comparative analysis for matching different dataset inputs for maximizing F1. In the less possibility induced classification, two cases are analyzed, i.e. $i = 0$ and $i = I$. This is based on$n'(det(i) + \alpha)$ for preventing further

Table 4
α for different objects and classifications.

Objects	Classifications	Accuracy	α
3	18	0.51	−0.06
4	25	0.58	−0.058
5	75	0.62	−0.042
6	69	0.57	−0.015
7	82	0.68	+0.26
8	92	0.71	+0.45
9	102	0.76	+0.054
10	116	0.82	+0.069
11	127	0.87	+0.078

classification constraints. Therefore, the proposed scheme requires further region detections. In the detected regions, first, α is identified for $\frac{n(i)(C_p)}{\alpha-\beta}$ classification. In this process, the odd constraint of $i = n \neq i = 0$ is identified for performing intermediate classifications as validated in Eq. (7). Therefore, the analysis $\left(\frac{n(i)}{b_n \sin\frac{n\pi x}{f_{det(i)}}(x^l)}\right)$ for all α_n. It is performed for validating a high F1 score. Therefore, the proposed scheme achieves a high F1-score for different regions and training instances, as presented in Fig. 6.

9. Error

The comparative analysis for errors for different regions and training instances is presented in Fig. 7. The proposed scheme reduces error in constraint mitigation and α analysis. In the first analysis, the $(C_p - L^c)$ computation reduces the errors in the identified region, preventing multiple classifications. Therefore, the error spotted is prevented from classification and remains in an unclassified instance, augmenting accuracy. This process is repeated until $i = 0$ is further observed. The changes are incorporated in the consecutive $\sqrt{\frac{n'+C_p}{s_0/\alpha}} + \prod_{s_0}(C_p - L^c)$. Reducing further constraint-based errors. The second way for error mitigation is the $\alpha = \frac{n\pi x}{f_{det(i)}}(\beta)$ estimation for the extracted features and represented dimensions. In this process, the $i = n$ is consistently verified post the classification iteration, improving the precision. Therefore, the classification filters $2\frac{n'+C_p}{s_0/\alpha}$ from the available representations, reducing errors. Both the processes are jointly used for preventing $\beta\left(\frac{\alpha+n'}{\beta(x^l)}\right)$ filters from the pre-detected output. Thus, for different regions and training instances, the errors with the region classification are diminished without compromising the precision.

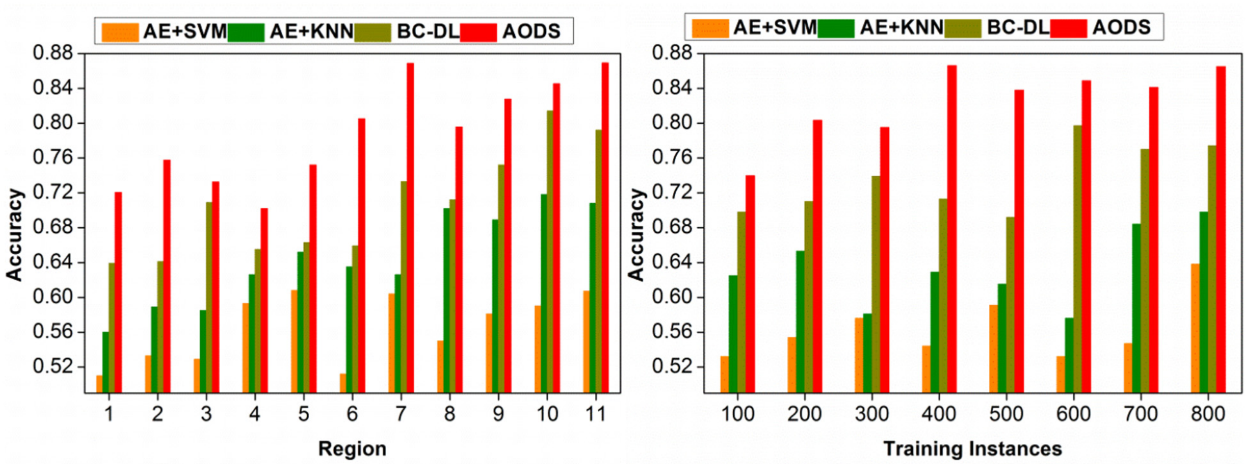


Fig. 4. Accuracy comparisons.

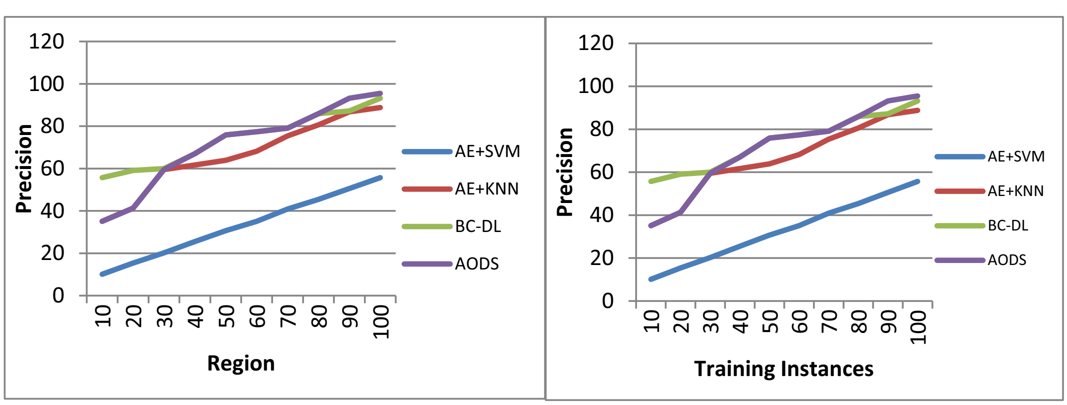


Fig. 5. Precision comparisons.

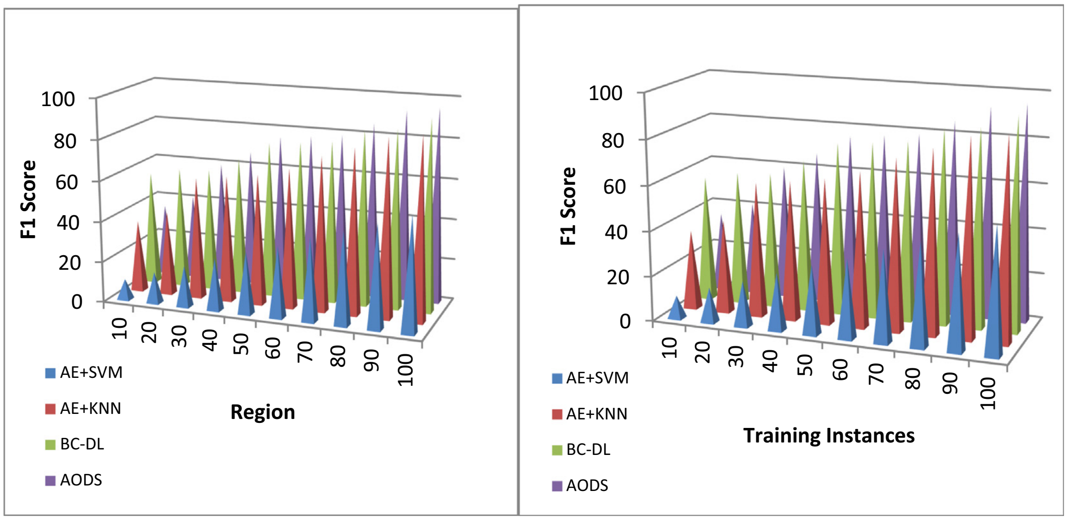


Fig. 6. F1-score comparisons.

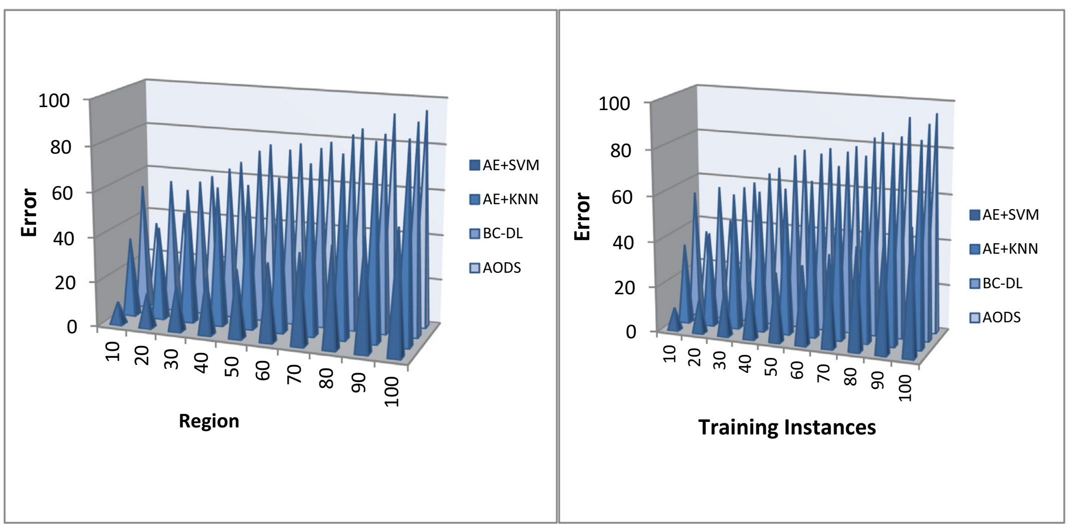


Fig. 7. Error comparisons.

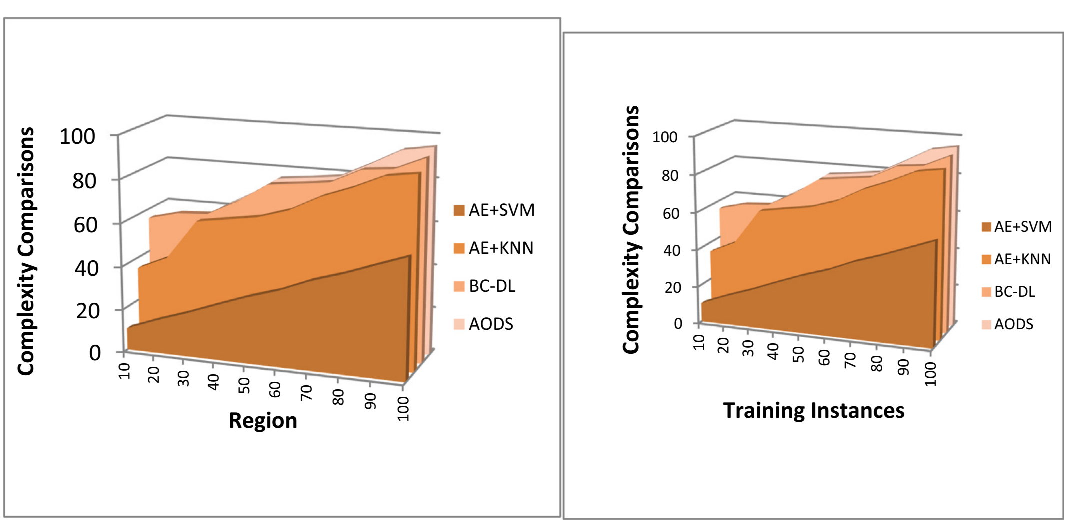


Fig. 8. Classification complexity comparisons.

Table 5
Object detection and training convergence for different methods.

Methods	Training convergence	Classification complexity (s)	Dimensional representation	Frame-to-object detection (%)
SVM	0.81	14.08	✓	68.37
AE + SVM	0.78	12.43	✓	75.69
KNN	0.76	10.29	✘	79.47
AE + KNN	0.68	9.14	✘	80.45
BC-DL	0.64	7.26	✘	83.69
AODS	0.53	5.17	✓	85.1

10. Classification complexity

In the proposed scheme, the $(x^i\alpha^{n-c_0}) + (b^{ext(i)}) = L^c$ violation classifications are identified through recurrent analysis. The analysis is performed and less and high classification possibilities. The α in multiple instances are identified as $ext\ (i-n)$ representation, preventing errors in computation. Therefore, $-u$ and v validations throughout $i=1$ and $i=n$ until $i=I$ is observed based on the classifications. In the error detection process, $\left(\frac{n(i)\complement_P}{\alpha-\beta}\right)$ relies on inequity, and hence the computations are prevented. The further classifications are pursued based on $-u$ and v independently without requiring high computations. In the proposed scheme, the classification constraints are mitigated in the occurring α instances. Therefore, the $i=1$ is sustained throughout the intervals and training instances, reducing the time for classification. This procedure is common for different training instances and regions detected (Refer to Fig. 8). In Table 5, the object detection and training convergence for different methods are analyzed.

In Table 5, the object detection ratio with convergence and complexity for different methods. The conventional SVM and sub encoder based $A\in$ ensures dimensional feature representation, whereas the complexity is high. This is due to the marginal variation as expressed by the classifier. The training is highly adaptable in different computation instances, maximizing the convergence. In a KNN, AE + KNN, and B-DL, dimensional representation is limited due to replicative features. Therefore, training convergence is high, whereas the detection is abruptly high. The proposed scheme achieves less convergence by reducing computation constraints, and hence $-v$ and u differentiation take place. Therefore training is performed for demandable classifications, reducing its convergence. The complex less computations reduce the frame-to-object detection accuracy. In Tables 6 and 7, the comparative analysis results are presented for different regions and training instances.

11. Concluding remarks

This article introduces an attuned object detection scheme for high accuracy hazardous object detection from CCTV images. The input image is analyzed for its feasible dimensions using the associated features. Based on the dimension representation, the regions are detected and classified. A convolutional neural network with constraint mitigation is employed in the classification process. This process is recurrent until the detected region can represent error fewer dimensions.This is the first time handgun appearance has been used with pose information to solve this problem. Next, be looking at the pose's variation over time to see if that provides any additional insight. The Attuned Object Detection Scheme and other methods will be considered for this several have been proposed for the problem of action recognition. Computational time is an important consideration in this scenario. The limitations in each region are analyzed using the hidden layer computations of the learning process. Therefore, the classifications are pursued based on less or high possibilities observed in the different training instances. The segmented regions are classified post the error and variation mitigation in detecting hazardous objects like guns. In this process, the training using an external dataset improves the accuracy by 8.08%, reducing error by 7.47% and complexity by 8.23%. In the future, the object classification based on labels is planned to be incorporated for identifying the object category. Besides, this would improve the object detection for high speed framed inputs.

Credit authorship contribution statement

V.P. Manikandan: Visualization, Investigation, Supervision, Software, Validation, Writing – review & editing. **U. Rahamathunnisa:** Conceptualization, Methodology, Software, Data curation, Writing – original draft.

Table 6
Comparative analysis for regions.

Metrics	AE + SVM	AE + KNN	BC-DL	AODS	Findings
Accuracy	0.607	0.708	0.792	0.8692	8.34% High
Precision	0.597	0.702	0.819	0.891	8.95% High
F1-Score	0.726	0.758	0.819	0.9105	7.53% High
Error	0.119	0.109	0.0982	0.0847	8.01% Less
Classification Complexity (s)	11.42	9.22	7.83	4.54	8.69% Less

Table 7
Comparative analysis for training instances.

Metrics	AE + SVM	AE + KNN	BC-DL	AODS	Findings
Accuracy	0.638	0.698	0.774	0.8649	8.08% High
Precision	0.592	0.713	0.829	0.883	8.58% High
F1-Score	0.719	0.782	0.848	0.9057	6.14% High
Error	0.12	0.1	0.0882	0.0794	7.47 Less
Classification Complexity (s)	11.16	9.07	7.92	4.75	8.23% Less

Declaration of Competing Interest

None.

References

[1] C. Premachandra, M. Tamaki, A hybrid camera system for high-resolutionization of target objects in omnidirectional images, IEEE Sensors J. 21 (9) (2021) 10752–10760.
[2] R. Jin, D. Lin, Adaptive anchor for fast object detection in the aerial image, IEEE Geosci. Remote Sens. Lett. 17 (5) (2019) 839–843.
[3] E. López-Rubio, M.A. Molina-Cabello, F.M. Castro, R.M. Luque-Baena, M.J. Marín-Jiménez, N. Guil, Anomalous object detection by active search with PTZ cameras, Expert Syst. Appl. 181 (2021), 115150.
[4] ChunleiHuo NuoXu, Xin Zhang, Yong Cao, Gaofeng Meng, Chunhong Pan, Dynamic camera configuration learning for high-confidence active object detection, Neurocomputing 466 (2021) 113–127 (ISSN 0925-2312).
[5] D. Giveki, Robust moving object detection based on fusing Atanassov's intuitionistic 3D fuzzy Histon roughness index and texture features, Int. J. Approx. Reason. 135 (2021) 1–20.
[6] J.L.S. González, C. Zaccaro, J.A. Álvarez-García, L.M.S. Morillo, F.S. Caparrini, Real-time gun detection in CCTV: an open problem, Neural Netw. 132 (2020) 297–308.
[7] A.A. Ahmed, M. Echi, Hawk-eye: an AI-powered threat detector for intelligent surveillance cameras, IEEE Access 9 (2021) 63283–63293.
[8] M.J. Stainer, P.V. Raj, B.M. Aitken, S. Bandarian-Balooch, M.J. Boschen, Decision-making in single and multiple-screen CCTV surveillance, Appl. Ergon. 93 (2021), 103383.
[9] J. Cameron, M.E. Kaye, E. Scheme, Dynamic prioritization of surveillance video data in real-time automated detection systems, Expert Syst. Appl. 161 (2020), 113672.
[10] T. Sikandar, K.H. Ghazali, M.F. Rabbi, ATM crime detection using image processing integrated video surveillance: a systematic review, Multimedia Systems 25 (3) (2019) 229–251.
[11] K.M. Bellazi, R. Marino, J.M. Lanza-Gutierrez, T. Riesgo, Towards a machine learning-based edge computing oriented monitoring system for the desert border surveillance use case, IEEE Access 8 (2020) 218304–218322.
[12] J.G. Choi, C.W. Kong, G. Kim, S. Lim, Car crash detection using ensemble deep learning and multimodal data from dashboard cameras, Expert Syst. Appl. (2021) 115400.
[13] I. Ahmed, M. Ahmad, J.J. Rodrigues, G. Jeon, Edge computing-based person detection system for top view surveillance: using CenterNet with transfer learning, Appl. Soft Comput. 107 (2021), 107489.
[14] J. Xu, A deep learning approach to building an intelligent video surveillance system, Multimed. Tools Appl. 80 (4) (2021) 5495–5515.
[15] C.S. Sung, J.Y. Park, Design of an intelligent video surveillance system for crime prevention: applying deep learning technology, Multimed. Tools Appl. (2021) 1–13.
[16] D. Wang, Y. He, Y. Liu, D. Li, S. Wu, Y. Qin, Z. Xu, 3D object detection algorithm for panoramic images with multi-scale convolutional neural network, IEEE Access 7 (2019) 171461–171470.
[18] M.T. Bhatti, M.G. Khan, M. Aslam, M.J. Fiaz, Weapon detection in real-time CCTV videos using deep learning, IEEE Access 9 (2021) 34366–34382.
[19] R. Olmos, S. Tabik, A. Lamas, F. Perez-Hernandez, F. Herrera, A binocular image fusion approach for minimizing false positives in handgun detection with deep learning, Inform. Fusion 49 (2019) 271–280.
[20] A. Castillo, S. Tabik, F. Pérez, R. Olmos, F. Herrera, Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning, Neurocomputing 330 (2019) 151–161.
[21] M.K. Galab, A. Taha, H.H. Zayed, Adaptive technique for brightness enhancement of automated knife detection in surveillance video with deep learning, Arab. J. Sci. Eng. 46 (4) (2021) 4049–4058.
[22] N. Vallez, A. Velasco-Mata, O. Deniz, Deep autoencoder for a false-positive reduction in handgun detection, Neural Comput. & Applic. 33 (11) (2021) 5885–5895.
[23] A. Velasco-Mata, J. Ruiz-Santaquiteria, N. Vallez, O. Deniz, Using human pose information for handgun detection, Neural Comput. Applic. (2021) 1–14.
[24] A. Shahbaz, K.H. Jo, Dual camera-based supervised foreground detection for low-end video surveillance systems, IEEE Sensors J. 21 (7) (2021) 9359–9366.
[25] Y. Liu, H. Yu, C. Gong, Y. Chen, A real-time expert system for anomaly detection of aerators based on computer vision and surveillance cameras, J. Vis. Commun. Image Represent. 68 (2020), 102767.
[26] J. Park, J. Heo, S.J. Kang, Feedback-based object detection for multi-person pose estimation, Signal Process. Image Commun. 116508 (2021).
[27] A. Rodda, S.S. Mallika, M. Jaiganesh, M. Mrunalini, Suspect detection system–an architecture based on visual surveillance analytics using IoT, Mater. Today Proc. (2020) https://doi.org/10.1016/j.matpr.2020.09.738.
[28] S.M. Mohtavipour, M. Saeidi, A. Arabsorkhi, A multi-stream CNN for deep violence detection in video sequences using handcrafted features, Vis. Comput. (2021) 1–16.
[29] R. Maqsood, U.I. Bajwa, G. Saleem, R.H. Raza, M.W. Anwar, Anomaly recognition from surveillance videos using a 3D convolution neural network, Multimed. Tools Appl. 80 (12) (2021) 18693–18716.
[30] https://www.kaggle.com/parthmehta15/fire-gun.