VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

Technical Seminar(18CSS84)
Report on

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

Submitted in partial fulfilment of the requirements for the award of the degree of
Bachelor of Engineering
in
Computer Science & Engineering

Submitted by

USN	Name
1BI19CS011	Akash Jain

Under the Guidance of
Dr. Maya B S
Assistant Professor
Department of CSE, BIT
Bengaluru-560004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
BANGALORE INSTITUTE OF TECHNOLOGY
K.R. Road, V.V. Pura, Bengaluru-560 004

2022-23

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

BANGALORE INSTITUTE OF TECHNOLOGY

Bengaluru-560 004

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Certificate

This is to certify that the Technical Seminar (18CSS84) entitled **"A neural network aided attuned scheme for gun detection in video surveillance images"** has been successfully presented by

USN	Name
1BI19CS011	**Akash Jain**

student of VIII semester B.E. for the partial fulfillment of the requirements for theBachelors Degree in Computer Science & Engineering of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY** during the academic year 2022-23.

Dr. Maya B S
Assistant Professor
Department of CS&E
Bangalore Institute of Technology

Dr. Girija J
Professor and Head
Dept. of Computer Science & Engineering
Bangalore Institute of Technology

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompanies the successful completion of any task would be incomplete without complementing those who made it possible and whose guidance and encouragement made my efforts successful. So, my sincere thanks to all those who have supported me in completing this technical Seminar successfully.

My sincere thanks to **Dr. M. U. Aswath**, Principal, BIT and **Dr. Girija J.**, HOD, Department of CS&E, BIT for their encouragement, support and guidance to the student community in all fields of education. I am grateful to our institution for providing us a congenial atmosphere to carry out the Technical Seminar successfully.

I would not forget to remember **Dr. Bhanushree K J**, Associate Professor and Technical Seminar Coordinator, for her encouragement and more over for her timely support and guidance till the completion of the Technical Seminar Coordinator.

I avail this opportunity to express my profound sense of deep gratitude to my esteemed guide **Dr. Maya B S**, Assistant Professor, Department of CS&E, BIT, for her moral support, encouragement and valuable suggestions throughout the Technical Seminar Coordinator.

I extend my sincere thanks to all the department faculty members and non-teaching staff for supporting me directly or indirectly in the completion of this Technical Seminar.

Akash Jain
1BI19CS011

ABSTRACT

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.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Chapter 1

INTRODUCTION

1.1 Overview

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. 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. 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. 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.

In this paper, a method using CNN Attuned Object Detection Scheme (AODS) is proposed 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 detects CCTV inputs (AODS) objects. Features from monitoring images can be extracted, and the objects in that area identified using a feature constraint-based analysis.

1.2 Insight to the domain

The global death toll from use of guns may be as high as 1,000 dead each day. From street crimes to an individual institution attack, many precious lives suffered. 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.

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.

DNN based regression is used in object detection to detect low-resolution images and helps to capture detailed information about the objects. CNN is a deep learning technique used to detect objects from an image and analyze the visual imagery of an image.

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.

The proposed scheme in this paper aims to address these challenges by providing an efficient method to search and detect guns in the in hidden areas using AODS method.

1.3 Objectives

The primary objective of this paper is to propose a efficient detection scheme for the guns using the AODS method. The proposed scheme aims toaddress the challenges associated with more efficient detection techniques, which even allow the detection of hidden guns. The proposed 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.. The authors identify the need for a scheme that can provide fast detection operations while ensuring the security of the people.

To achieve this objective, the authors propose a new region-based detection method and evaluate its performance using real-world datasets. The proposed scheme is compared with existing schemes to demonstrate its efficiency and performance.

In addition to proposing a new scheme, this paper aims to provide a comprehensive understanding of the proposed scheme and its potential use in the field of remote weapon detection. The authors highlight the benefits of using AODS method for detecting harmful weapons and identify the potential future research directions in this field.

Overall, the objectives of this paper is to propose a CNN Attuned Object Detection Scheme (AODS) for detection of harmful object from CCTV inputs. Once a CNN is built, it can be used to classify the contents of different images. CNNs can classify images by detecting features, like how the human brain detects features to identify objects. Images are made up of pixels.

.

1.4 Organization of report

The rest of the report is divided into 6 chapters as follows. Chapter 2 deals with the review of related works of the seminar topic. Chapter 3 describes the proposed system of the seminar topic. Chapter 4 describes the architecture of the proposed system. Chapter 5 gives the results of the proposed system. Chapter 6 gives the applications of the proposed system. Finally Chapter 7 gives the conclusion of the seminar topic.

CHAPTER 2

LITERATURE SURVEY

Chapter 2

LITERATURE SURVEY

2.1 Introduction

Literature Survey or Literature Review is a survey of previously existing scholarly resources such as books, journals, and articles related to specific topics or questions. It involves the search and evaluation of available literature in your given subject or chosen topic area. It documents the state of the art concerning the subject or topic you are writing about. Concerning the seminar, a literature survey was conducted to get a better picture of the idea, for idea development to understand the methodology that is currently being used to learn about the limitations present in the correct methodologies.

2.2 Summary of Papers

The below section provides a brief summary of the reference papers utilized for analysis of weapon detection scheme.

[1]Muhammad Tahir Bhatti, Muhammad Gufran Khan, Masood Aslam and Muhammad Junaid Fiaz 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.

The architecture of the proposed scheme consists of some of the different types of weapons like pistol, revolver and non-pistol. Non-pistol includes equipment other than pistol like knife or may be rod. The reason we choose pistol and revolver in the pistol class is because of our study and analysis after watching many robberies and shooting incident CCTV videos. We concluded that almost 95% of the weapon used in those cases were either pistol or revolver.

The methodology adopted in this work features the state of art deep learning, especially the convolutional neural networks due to their exceptional performance in this field. Therefore, mentioned techniques are used for both the classification as well as localizing the specific object in a frame so both the object classification and detection algorithms were used and because our object is small with other object in background so after experimentation, we found the best algorithm for our case. Sliding window/classification and region proposal/object detection algorithms were used.

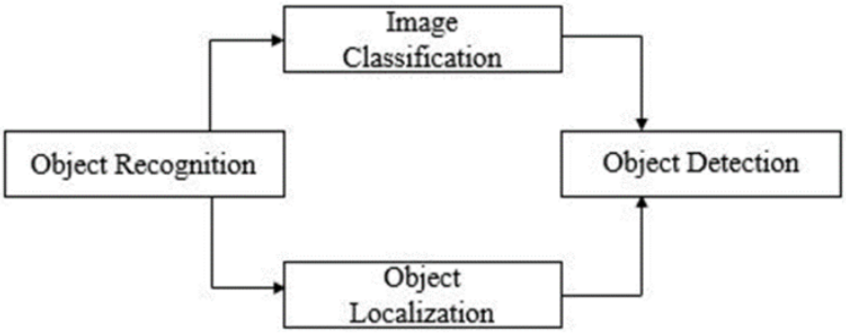


Figure 2.1: Object Recognition to detection Hierarchy

[2]Mai K. Galab, Ahmed Taha, Hala H. Zayed proposed 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.

AlexNet architecture is employed. It consists of eight layers: five convolutional layers followed by three fully connected (FC) layers. FC layers are the final layers of CNN that determine the result of the intended output. It sums the weights of the previous feature layers, indicating the right mix of "ingredients". The final FC generates 4096 neurons. The final step in CNN reduces the features vector from 4096 to 1000, for the nonlinear part.

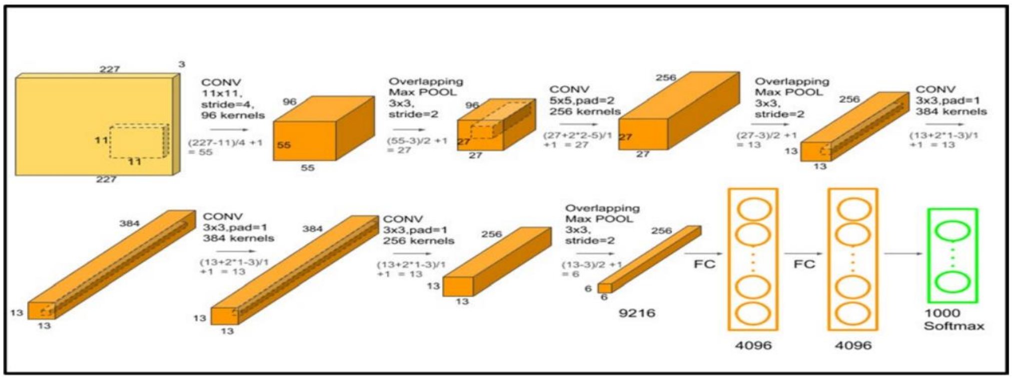


Figure 2.2: AlexNet Architecture

The proposed technique begins by applying pre-processing to the frames captured from the CCTV camera to make the frame suitable for the selected CNN. The images collected from the raw input data are resized to 227×227 to be suitable to work with AlexNet. In step 2, TL is applied after dividing the dataset into two groups: the first group is the learning setand the second group is the test set. Step 3 finds significant features using CNN. It applies a sequence of the convolutional layers and pooling layers. The classifier is composed of FC layers. Step 4 tests whether the frame exceeds the brightness level range or not. If the frame brightness exceeds this level, it must be improved by applying our "Brightness Handler Procedure" (BHp). Otherwise, go directly to the next step. Steps 5 and 6 classify the images and decide whether it is a knife or not.

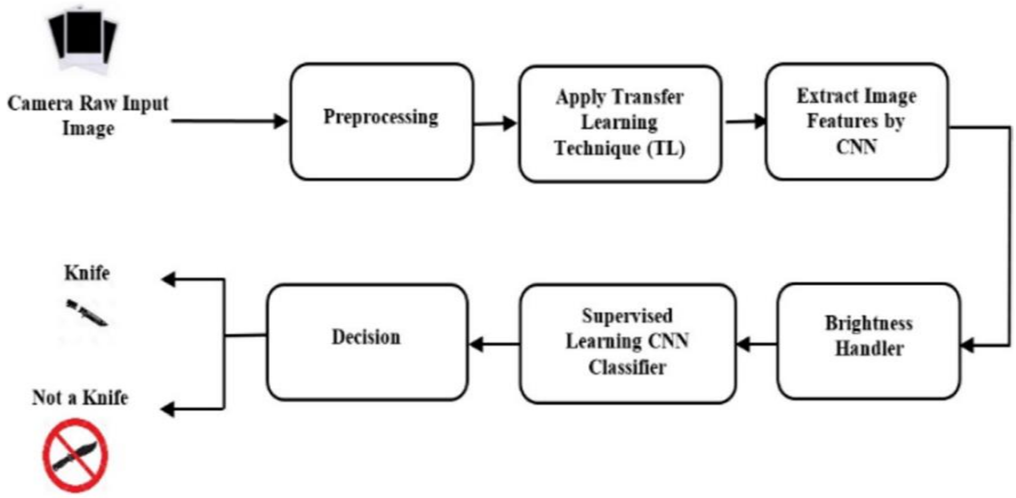


Figure 2.3: Block Diagram of Proposed Technique

[3]Alexander Egiazarov, Vasileios Mavroeidis, Fabio Massimo Zennaro, Kamer Vishie tackle the problem of detecting the presence of weapons within an image. They cast this problem as image detection/segmentation supervised learning problem: They want to learn a function that given an image as input returns the presence and the location of the weapon. Following the state of the art, they decided to use convolutional neural networks to learn such a function.

All the networks have the same architecture: three convolutional layers for feature extraction and three dense layers for binary. In the convolutional section, we use layers containing 32 or 64 filters with default stride of Ix I, with ReLU activation functions, and 2x2 max-pooling. In the dense section, we use fully connected layers. We use a ReLU activation function, except for the last layer where we rely on the softmax function compute output.

This proposed system has the use of multiple simple neural networks that could be cheaply trained to detect only specific components of a weapon and which could be easily aggregated to produce robust output. We tested our model on data from the same distribution as the training data (i.e., Google images), we examined it on synthetic sanitized data, and we evaluated it on out-of-sample (i.e., video frames) data. The results showed the reliability of the individual component part networks and the versatility of the overall system in integrating the outputs of the single networks.

[4]Lei Pang, Hui Liu, Yang Chen and Jungang Miao proposed real-time concealed various object detection under human dress in Metallic guns on human skeleton were used for passive millimeter wave imagery which relies on YOLO algorithm on dataset of small scale. Subsequently, comparison is undertaken between Single MultiBox Detector algorithm, state of the art, They decided to use convolutional neural networks to learn such a function. 4 YOLOv3-13, SSD-VGG16, and YOLOv3-53 on PMMW dataset. Moreover, the weapon detection accuracy computed 36 frames per second of detection speed and 95% mean average precision.

The architecture treats the target detection task as a regression problem, and directly obtains the target bounding box, the confidence Pc, and the probabilities of being a certain target by taking all pixel values of the image as the input. As shown in Figure 2, it uniformly samples the input image of a size of 416 × 416, and supposes the image is segmented by 3 × 3 grids. Each grid predicts B bounding boxes, involving seven predicted values (bx, by, bw, bh, and its corresponding Pc and class probability, c1 and c2), and all predicted values are output as a tensor with a shape of 3 × 3 × (B × 7). Where Pc indicates the confidence, defined as the target in the predicted bounding box; (bx,by) indicates the position of the center point of the target relative to its corresponding grid; (bw, bh) represents the category probabilities of two different targets.

This method divides one image into an S × S grid. For each grid cell, B bounding boxes with confidence are predicted. These predictions are encoded as an S x S x(B x 7) tensor. PMMW image contraband target detection by YOLOv3 requires the detection of metal gun and human body targets. The model was able to detect three different scale targets for grid sizes of 13 × 13, 26 × 26, and 52 × 52, respectively, and can predict simultaneously three bounding boxes of each scale.

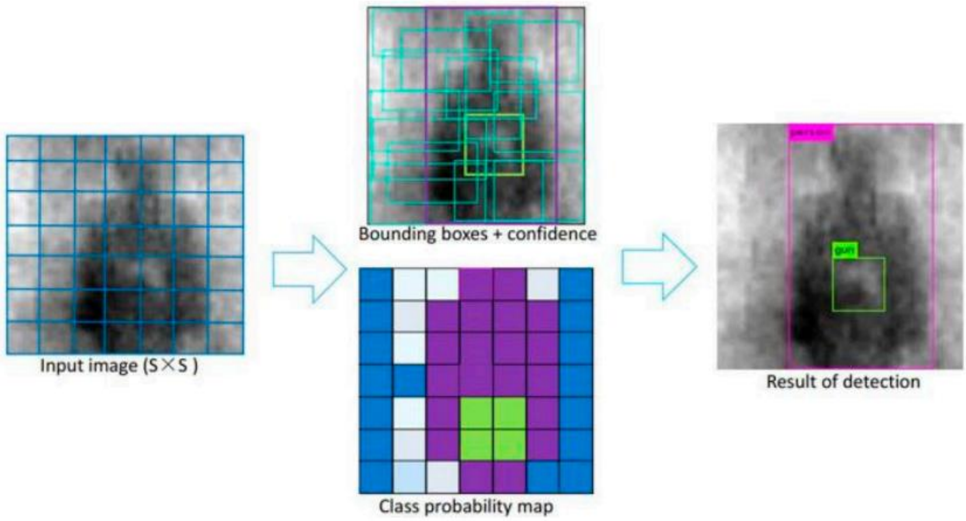


Figure 2.4: The YOLOv3 object detection method.

[5]Jesus Ruiz-Santaqueteria, Alberto Velasco-Mata, Noelia Vallez, Gloria Bueno, Juan A. Álvarez-G, and Oscarcia R Deniz proposed 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. For the detection of threats or dangerous objects such as firearms, most of the proposed methods are based only on the visual appearance of the objects, without taking into account the human pose or another additional information.

The whole network architecture, henceforth named as HRC+P (Hand Region Classifier + Pose data), is divided into two main branches. The first one is the hand region classifier (HRC). On the other hand, the processing of the pose image is carried out by another custom subnetwork. The last feature vectors of the two branches are then joined in a single feature vector connected to the final classification layer. During the training step of the whole framework, the HRC and pose branches are optimized simultaneously. In this way, the model is capable of learning the optimal combination of handgun appearance with human pose information to improve the classification performance.

The proposed method consists of generating the handgun predictions in the image. Each hand region of each detected person is passed through the classification network to obtain a class label (handgun vs no-handgun). Then, the bounding boxes of the regions classified as hand gun are included in the output list of predicted handguns.

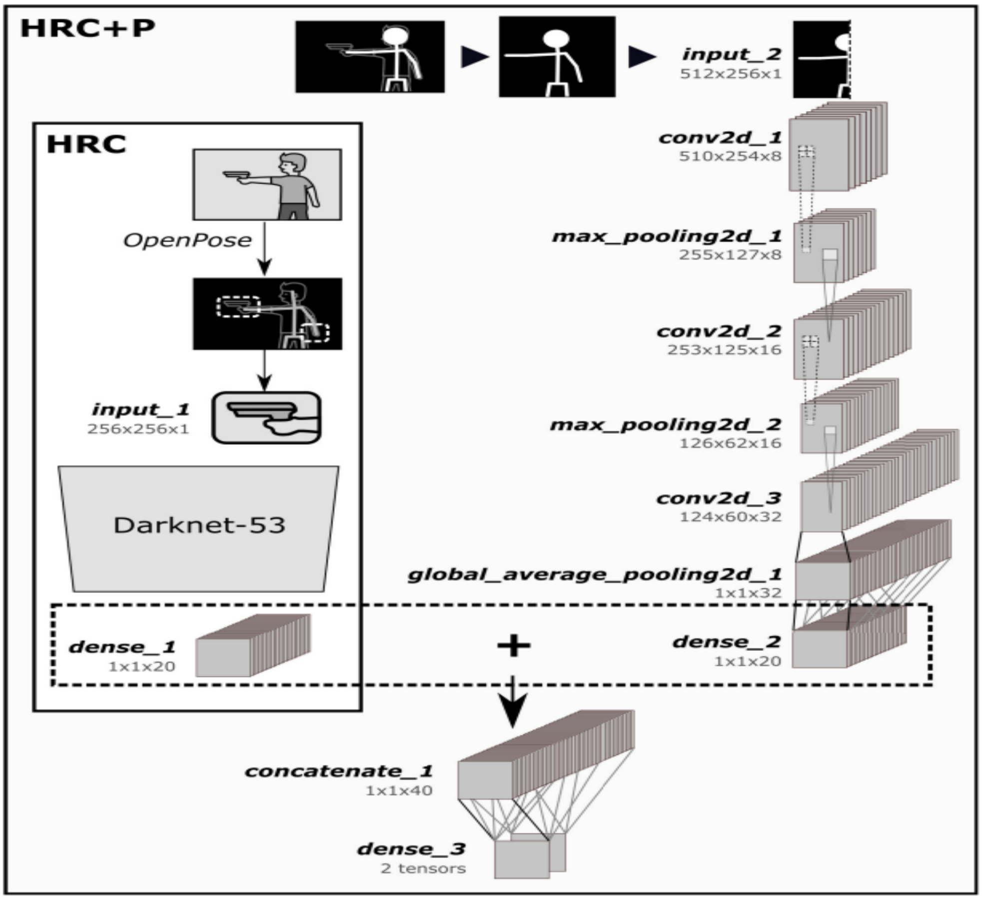


Figure 2.5: Network Architecture

[6]Seyed Mehdi Mohtavipour, Mahmoud Saeidi, Abouzar Arabsorkhi 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. d for feature extraction and data classification, respectively. DMOF and DMEI were two novel discriminative features that trained well the CNN network to predict all input frames of datasets. The experimental results showed that violence detection accuracy is obtained approximately 100% for both crowded and uncrowded environments.

A three-stream deep network has been proposed with three spatial, temporal, and spatiotemporal inputs. Spatial stream analyses the appearance in video sequences by focusing on a grayscale image. The temporal stream considers the speed of movement for moving targets by using a modified optical flow approach. In the spatiotemporal stream, we form the shape of actions by building a differential motion energy image (DMEI).

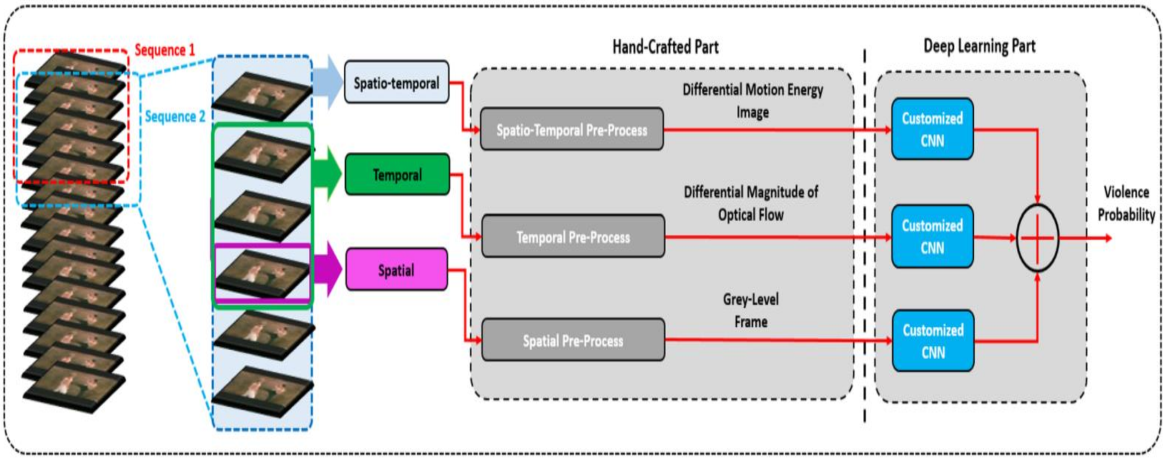


Figure 2.6: Proposed deep violence detection framework using handcrafted and deep learning parts

[7]Jesus Salido 1, Vanesa Lomas, Jesus Ruiz-Santaquiteria and Oscar Deniz proposed how including body pose information (skeleton keypoints and limbs retrieved by a pose detector) in the input images, as a preprocessing step, can improve the handgun detection performance.

It is common for object detection frameworks to organize their hierarchical architecture into three components:

1. Backbone network: Some of the networks used are Darknet-53, firstly employed for YOLO, VGG-16 and ResNet-50/101, which can be used with a setting obtained by training on a dataset elaborated for similar contexts;
2. Neck: This is the part of the network that strengthens the results by offering invariance to scale through a network that takes feature maps as the input at different scales. A very common implementation method is the feature pyramid network (FPN) and the multilevel feature pyramid network (MLFPN);
3. Detection head: This is the output layer that provides the location prediction of the bbox that delimits each object and the confidence score for a particular class prediction.

The Faster R-CNN was proposed to achieve the required processing speed for real-time applications. In the Faster R-CNN, the non-learning-based selective search algorithm is substituted by a region proposal network (RPN), which “learns” how to determine regions in which the objects are located. To propose the regions where each object is located, the RPN network slides an $n \times n$ spatial window of the input convolutional feature map obtained by the convolutional layers of a backbone network. The number of total proposals for each location is k. Therefore, as $k = n \times n$, $n = 3 \Rightarrow k = 9$.

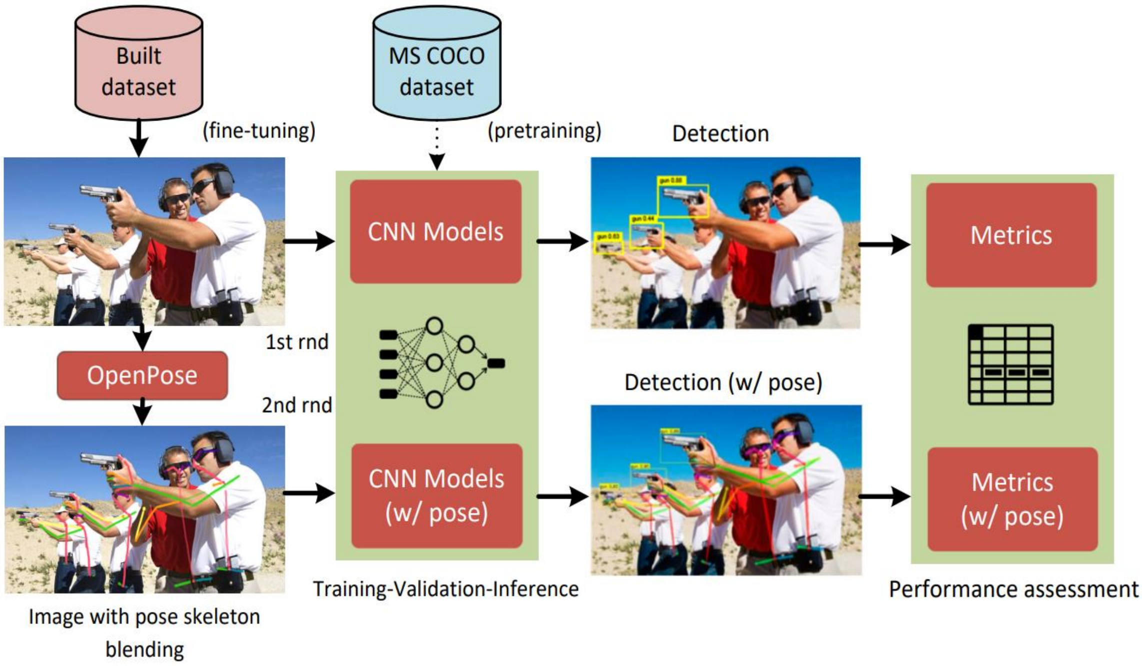


Figure 2.7: System block diagram

[8]Jaeseo Park, Junho Heo, Suk-Ju Kang proposed an object detection based on the early estimation of multi-person poses. This method improves the performance of the overall of service. Key point estimators and object detectors are widely used in real-time expert for the multiperson poses. The proposed method is high in performance rate and accuracy rate of an detecting an object compared with the state-of-the-art model.

The networks used as the baseline for this study were the human detector and SPPE developed using the PyTorch library. The details of each model and the performance of the baseline network are as follows. Two models were tested in this study. The first human pose estimator was HRNet-W48 , the input of which had a 384 × 288 high the pixel image resolution. This network has an average precision (AP) of 73.9 when using our baseline of human object detector with the COCO val2017 dataset. The second human pose estimator was Table 2 based on the original regional multi-person pose estimation (Rmpe). This model was a little less accurate, but faster than the first model. The final MPPE network of this model had an AP of 73.0 for the COCO val2017 dataset. Among the models shown in Tables II, Rmpe, the baseline MPPE, and our models used the same human object detector, i.e., Yolov3. Pre-trained weights were used to predict the images. This human object detector had a human AP of 55.5 for the COCO val2017 dataset. Table 2 shows the inference time and accuracy when using an image adjusted to 384 × 288 and 320 × 320 pixel resolutions as an input.

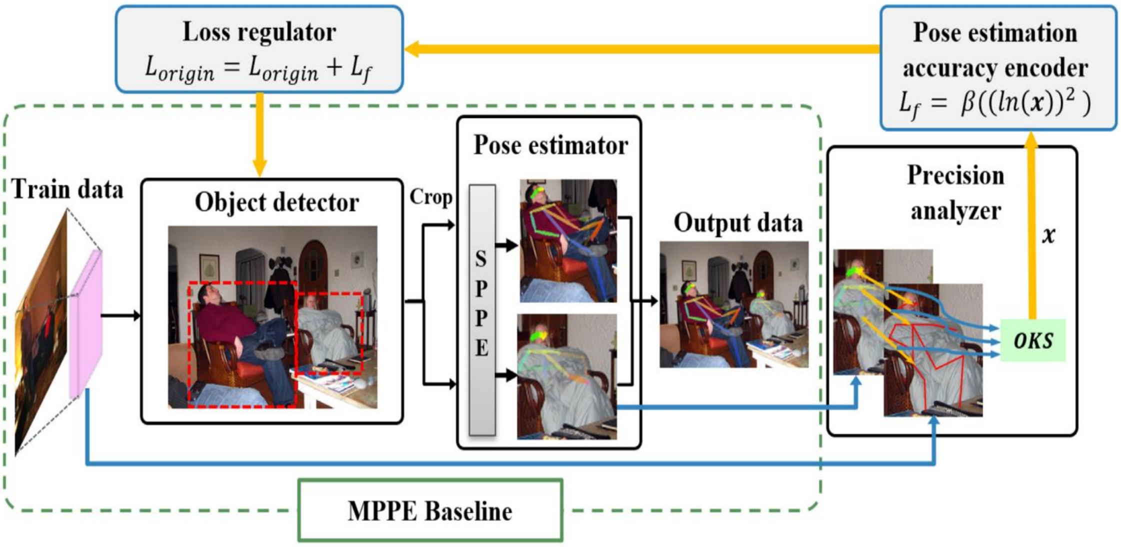


Figure 2.8 : Overall architecture

The overall architecture of the proposed system. The black arrows show the application of the proposed loss function to the object detector. In addition, the blue and yellow arrows show the operation of the proposed detection expander.

[9]Jose L. Salazar González a, Carlos Zaccaro a, Juan A. Álvarez-García a, Luis M. Soria Morillo a, Fernando Sancho Caparrini work was very much related to achieve real-time results. They did immense experimentation using different datasets and trained Faster–RCNN using 5 Feature Pyramid Network with Resnet50 and improves the previous state of the art by 3.91 %.

To train this architecture, we first used Fine-tuning with the frozen weights of the COCO dataset, one of the largest and most tested datasets to date, to take advantage of previous training on a large number of images. The object detection models used in this study were trained on the detectron2 framework (Wu, Kirillov, Massa, Lo, & Girshick, 2019) with the Faster R-CNN-FPN architecture with the resnet50 backbone. This configuration obtains an mAP of 40.210 over the COCO dataset, with an execution time of 90 ms per image of our CCTV dataset using an NVIDIA GeForce GTX-1080Ti card, providing a good balance between mAP and execution time

This variety of combinations allowed us to understand how the model performs with different training data and to analyze how the use of synthetic data impacts on it; also, the four test datasets allowed us to analyze and compare the performance of the different models in different contexts. Although the test set and the training set usually must satisfy the condition of the same distribution, in our case it is impossible to include synthetic images to the test set since they represent a real situation The results were obtained using the following detection COCO metrics11 and other standard metrics: Average Precision (AP), used as COCO primary challenge metric; AP 0.5, used also as PASCAL VOC metric; AP 0.75, strict metric; AP Across Scales, being this, APs, APm and APl; and True Positives (TP), False Positives (FP), False Negatives (FN), Precision, Recall, and F1-score, obtained with an IoU of 0.50 and a confidence limit of 0.95 or 0.99.

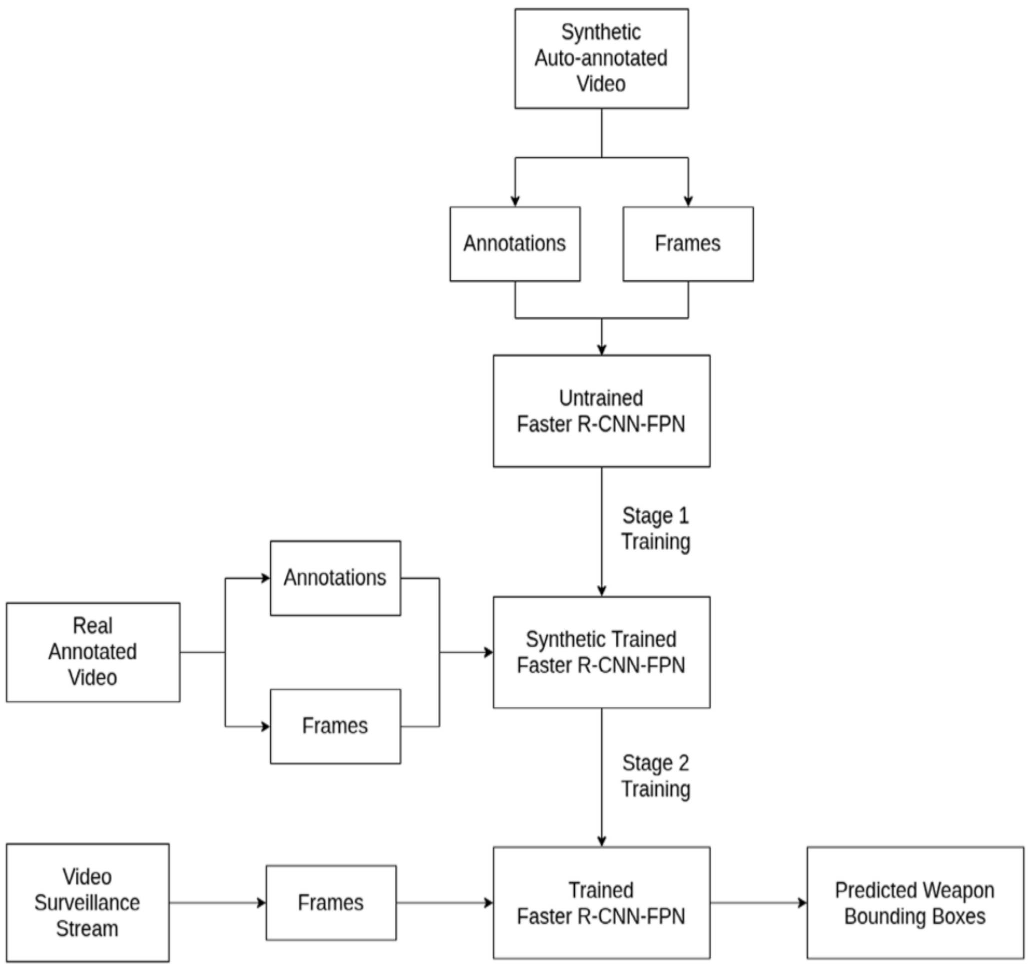


Figure 2.9: Block scheme of the complete procedure

2.3 Problem Statement

“To develop an efficient detection scheme for weapons hidden in the very narrow regions using Attuned Object Detection Scheme (AODS)”.

2.4 Objectives

- Automatically detect automatic guns and firearms using deep learning models.
- Detection of knives and sharp cold steel weapons in video surveillance footage.
- Real-world use in highly populated and complex scenes.
- Activation of an alarm in real-time with information about the weapon’s location.
- Weapon identification and classification to aid further investigation by security personnel.

CHAPTER 3

PROPOSED SYSTEM

Chapter 3

PROPOSED SYSTEM

The proposed AODS method detects the dangerous objects in the CCTV footage. CCTV works under the conditions of capturing 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 information in original data set, feature extraction involves converting raw data into numerical features that can be processed. Compared to using machine learning directly on raw data, this method produces superior result.

The proposed system also incorporates a method using CNN Attuned Object Detection Scheme (AODS) 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.

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.

The architecture diagram of the proposed system shows the different components of the system, including the user, cloud server, master index, and sub-indexes. The user interacts with the system through a client application, which generates the search token for the desired keyword using the user's secret key. The cloud server receives the encrypted search token and uses the hash index to retrieve the matching encrypted document IDs. The server returns the encrypted document IDs to the client, which decrypts them using the user's secret key to obtain the actual document IDs. The master index manages the sub-indexes and the revocation list, ensuring the integrity and security of the system.

CHAPTER 4

ARCHITECTURE

CHAPTER 4

ARCHITECTURE

4.1 Overview

An architecture diagram of the proposed system shows the different processes involved: Feature extraction, Dimension representation, Region detection, Constraint analysis and the Classification. 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 detects CCTV inputs (AODS) objects. Features from monitoring images can be extracted, and the objects in that area identified using a feature constraint-based analysis.

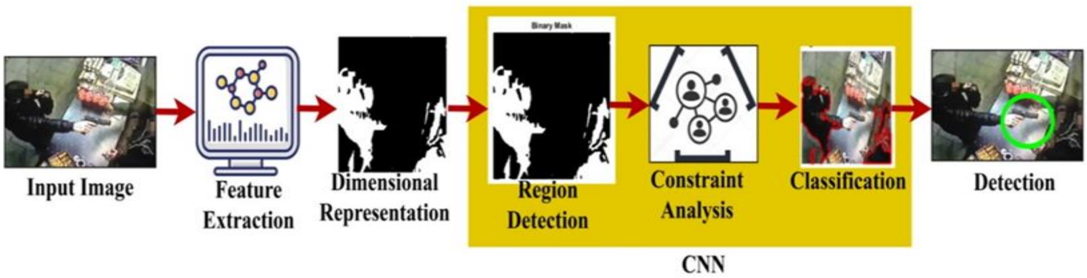


Figure 4.1.1: Attuned Object Detection System (AODS) process

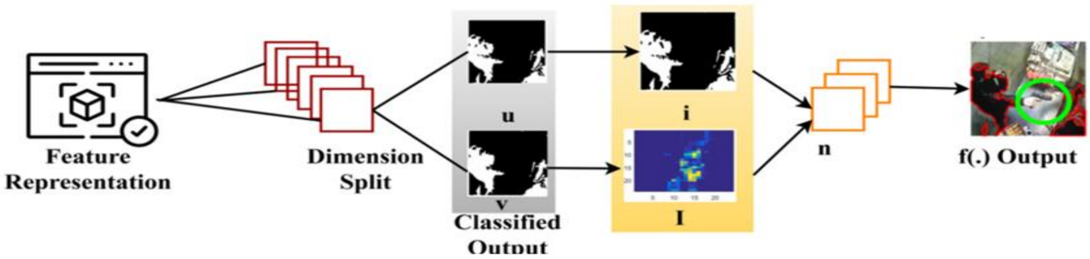


Figure 4.1.2: Classification of CNN

The proposed system uses attuned object detection scheme that provides detection of various weapons such as gun, knife, and other metal items. The system also provides a better and efficient search operation by utilizing the dimension split and region split techniques. The use of these techniques makes the system scalable and efficient for large datasets.

4.2 METHODOLOGY

4.2.1 Feature Extraction:

This detection scheme describes the object identification pursued in the proposed scheme. It identifies 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(a) , precision 1(b) and metric accuracy 1(c) of the images. The following equation represents the proposed objective-

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

The above Eq. 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. Where the image ‘a’ video ‘b’ for the input image f(a,b), and fdet(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.

4.2.2 : Dimension Representation

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) + hdet(i)) n(i) is processed in the above Eq. (2) in this ext ið Þ 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.

4.2.3 Region Detection:

The input image is differentiated for different features based on the ext ið Þ 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 from its earlier. 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) c0 is denoted as region detection.The problem is finding and classifying variable objects on an image.

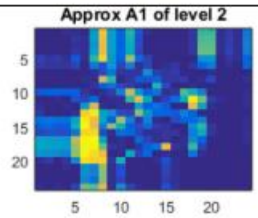
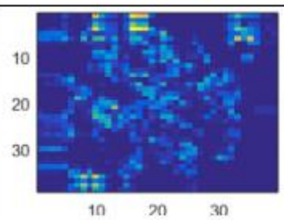
Input	Dimension 1	Dimension 2	Detected Regions
			

Figure 4.2.1: Region detection for sample inputs.

4.2.4 Constraint Analysis:

Each image in the dataset will be in different sizes and variables. Only the necessary images are 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 regionbased 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.'Lc ' represented as limitations over the CNN. It filters out unwanted duplicate images and videos in the network. At this moment, computing (xi + a0) n It examines the dataset in which it acquires the region detection over the constraint analysis.

4.2.5 Classification:

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 n ið Þð Þ Cp ∗β det ið Þþð Þ c0 . 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

αþn0 β xi ð Þ ————————————————————————————

. 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.

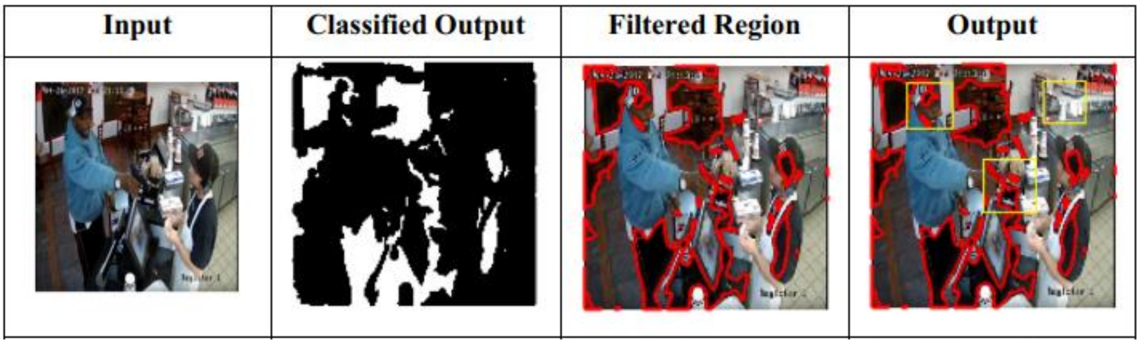


Figure 4.2.2: Classified output for sample inputs.

CHAPTER 5

RESULTS

Chapter 5

RESULTS

5.1 Accuracy

In Fig. 5.1, the comparative analysis for accuracy for different regions and training instances. The proposed scheme achieves fair accuracy by addressing the constraints in region detection. 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∑v=0 −u an−c0 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 untili = 0. The final pre-detected output is based on f ∏ s0 ƇP−LC satisfying β αþn0 β xi ð Þ Indifferent iterations. If the region increases, then αþn0 β xi ð Þ −tn. 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.

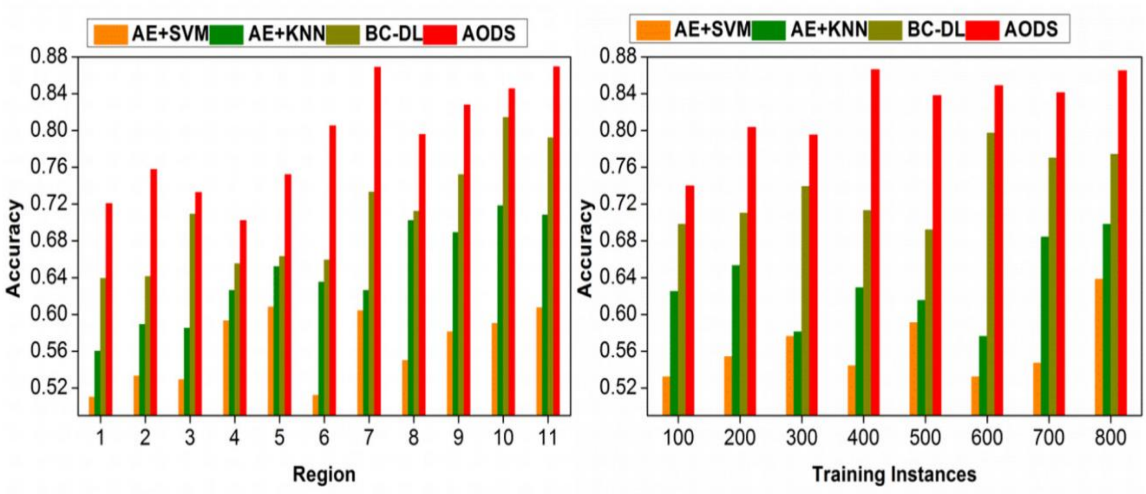


Figure 5.1: Accuracy Comparison

5.2 Precision

The precision is estimated as computed 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 ext ið Þ b based on prediction is first analyzed. In this analysis, the unclassified dimension features are assessed as ∏sinb ni (c0 + xi − a0) for the first dimension observed. The further dimension analysis is presented as (xi + a0) 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 2 cos an α ∗ext ið Þ −n 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).

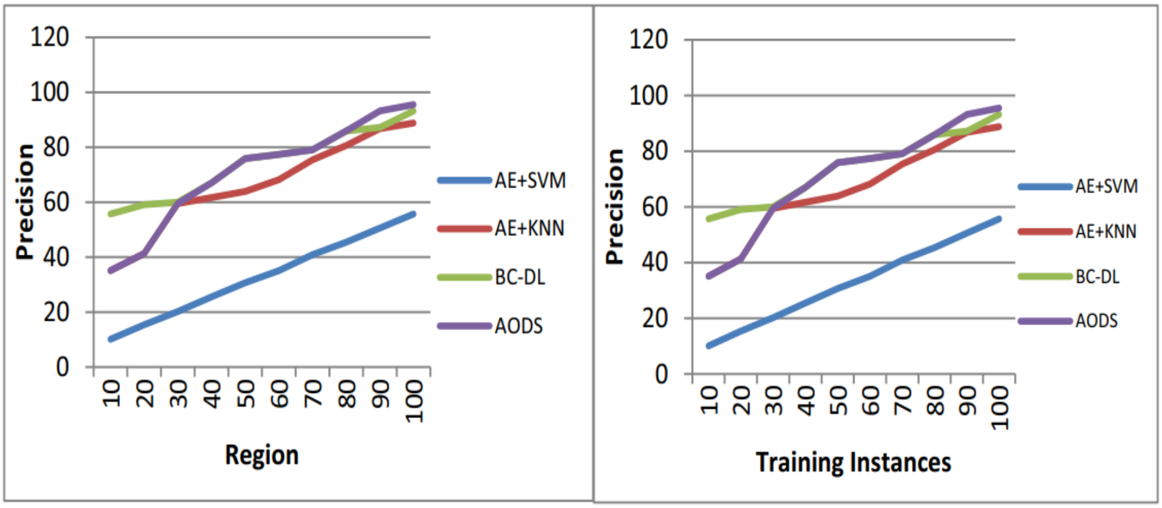


Figure 5.2: Precision Comparison

5.3 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. 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 classification constraints. Therefore, the proposed scheme requires further region detections.

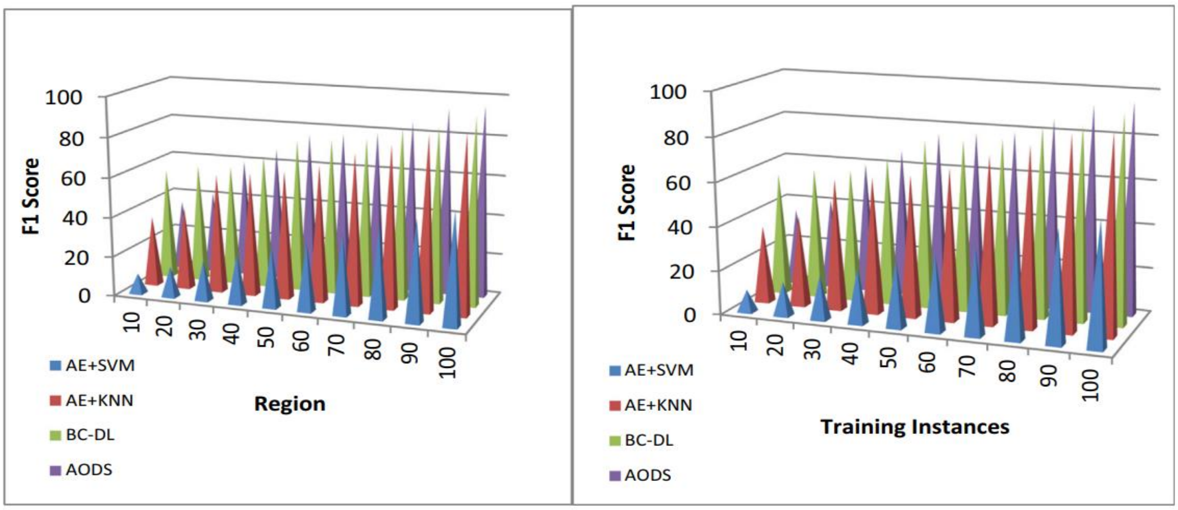


Figure 5.3: F1-Score Comparison

5.4 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 $(Cp - Lc)$ 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.

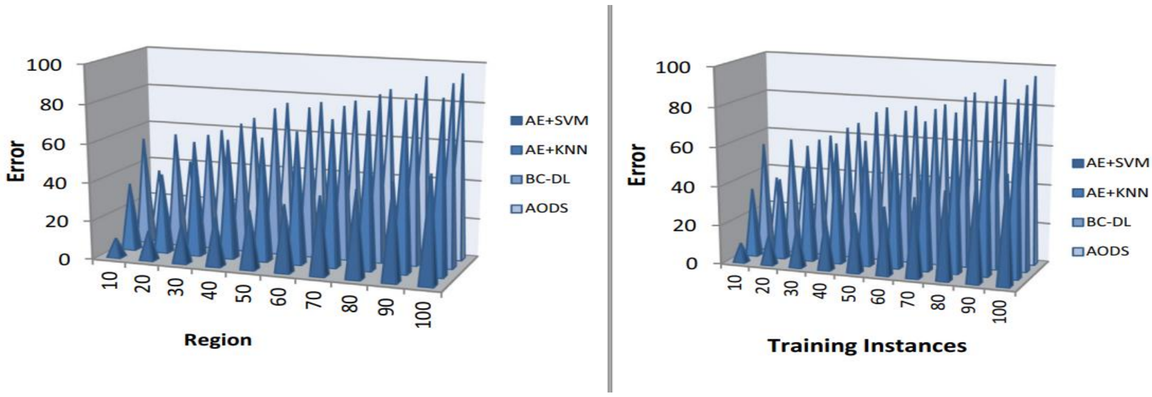


Figure 5.4: Error Comparison

5.5 Classification Complexity

The analysis is per formed 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. 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.

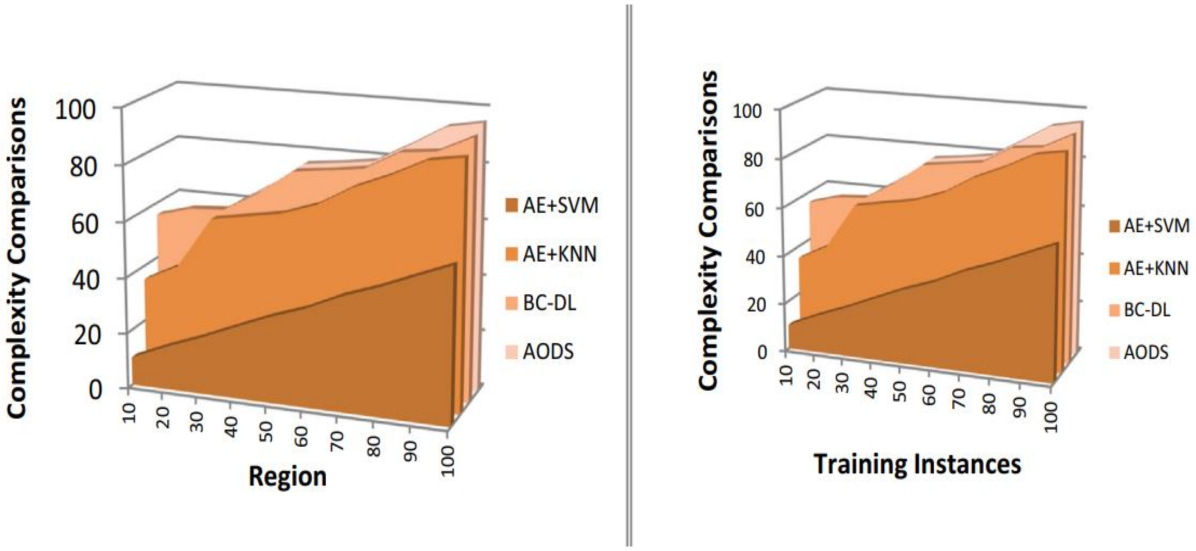


Figure 5.5: Classification Complexity Comparison

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

Figure 5.6: Comparative analysis for regions.

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

Figure 5.7: Comparative analysis for training instances

CHAPTER 6

APPLICATIONS

Chapter 6

APPLICATIONS

Secure searchable encryption scheme can be applied in various scenarios where sensitive data needs to be stored in the cloud, but still needs to be searchable with efficient retrieval times while preserving the confidentiality of the data. Some potential applications of the scheme are as follows:

- **Healthcare**: It was found that hospitals with metal detectors were more than 5 times as likely to frequently confiscate weapons, suggesting this intervention is effective. It was also found that hospitals with psychiatric units were more likely to have frequent confiscation of weapons, likely due to the standard procedure of searching patients before admission to the psychiatric unit.
- **Public sectors:** As on serious note, the weapon detection is most likely to be used in public areas where there is lot of crowd and the people have an advantage of conjusted areas to hide there weapons.
- **Government organizations**: Government organizations are the places where people are most likely to be targeted to be harmed or killed due to many reasons and can include severe harm to the individual or the group of people.
- **Reduce manpower:** Easily detect pistols, long weapons in running video frame and reduces the dependency and manpower required for it.
- **Eye protection:** Due to continuous eyes on screen, it may lead to sluggishness and human errors, which can be reduced drastically with this system.

CHAPTER 7

CONCLUSION

Chapter 7

CONCLUSION

7.1 Summary of the proposed system

The proposed system providess 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. Experimental results show that the proposed scheme achieves high efficiency and performance in practical scenarios.

7.2 Limitations of the proposed system

- The proposed scheme assumes as with most imaging, we confront the problems of noise and resolution limitation in CWD imaging.
- Equipment limitations and Real-world implementation
- The proposed method involves modifying the object detector, which increases the training time up to 2.2 times per epoch.
- Addition of pose information to the dataset slightly worsened the handgun detection

7.3 Future enhancement

Future work consists of increasing the efficiency of the detection algorithm byoptimizing the various functions or using parallel computing techniques and incorporating object classification based on labels for identifying object categories and improving object detection for high-speed framed inputs.

REFERENCES

[1] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning", in IEEE Access, vol. 9, pp. 34366-34382, 2021.

[2] Galab, M.K., Taha, A. & Zayed, H.H, “Adaptive Technique for Brightness Enhancement of Automated Knife Detection in Surveillance Video with Deep Learning,” Arab J Sci Eng 46, 4049–4058, 2021.

[3] Alexander Egiazarov, Vasileios Mavroeidis, Fabio Massimo Zennaro, Kamer Vishie, "Global study on homicide 2019, data: UNODC homicide statistics 2019”, July 2019.

[4] L. Pang, H. Liu, Y. Chen, and J. Miao, “Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm”, Sensors, vol. 20, no. 6, p. 1678, 2020.

[5] Mohtavipour, S.M., Saeidi, M. & Arabsorkhi, A,”A multi-stream CNN for deep violence detection in video sequences using handcrafted features”, Vis Comput 38, 2057–2072, 2022.

[6] González, Jose L. Salazar, Carlos Zaccaro, Juan A. Álvarez-García, Luis M. Soria Morillo, and Fernando Sancho Caparrini, "Real-time gun detection in CCTV: An open problem", Neural networks 132, 297-308, 2020.

[7] Velasco-Mata, A, Ruiz-Santaquiteria, J., Vallez, N. et al “Using human pose information for handgun detection”, Neural Comput & Applic 33, 17273–17286, 2021.

[8] J. Salido, V. Lomas, J. Ruiz-Santaquiteria, and O. Deniz, ‘‘Automatic handgun detection with deep learning in video surveillance images’’, Appl. Sci., vol. 11, no. 13, p. 6085, Jun. 2021.

[9] Jaeseo Park, Junho Heo, Suk-Ju Kan, “Feedback-based object detection for multi-person pose estimation,Signal Processing”, Image Communication, Volume 99, 2021.

[10] Manikandan, V. P., and U. Rahamathunnisa, "A neural network aided attuned scheme for gun detection in video surveillance images", Image and Vision Computing 120, 2022.