e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

REAL-TIME ABNORMAL OBJECT DETECTION USING DEEP LEARNING

Md. Nahid Sultan*1, SM Sojib Ahamed*2, Hossain Imran*3,
Md. Arman Hossain*4, Md Nadiruzzaman Nahid*5

*1Department Of Computer Science And Technology, Southwest University Of Science And Technology, Mianyang, China.

*2School Of Electronic Science And Engineering ,Nanjing University Of Aeronautics And Astronauts, China.

*3School Of Energy Power And Mechanical Engineering, North China Electric Power University, China.

*4Research Institute Of Economics And Management, Southwestern University Of Finance And Economics, Sichuan, China.

*5Department Of Electrical Engineering & Automation, North China Electric Power University, China.

DOI : https://www.doi.org/10.56726/IRJMETS60191

ABSTRACT

The task of detecting abnormal objects is of utmost importance and finds applications in diverse sectors, ranging from security and surveillance to industrial quality control. Deep learning models have significantly transformed the computer vision field, demonstrating significant potential in real-world anomaly detection. This study utilizes YOLOv5, a cutting-edge real-time object detection model, to create a high-performing and precise abnormal item detection system. This paper discusses the challenges of detecting aberrant objects, emphasizing the need for real-time, resilient solutions, and introduces the YOLOv5 model for its speed and accuracy. Additionally, we outline its modification specifically tailored for the purpose of abnormal object detection. By doing fine-tuning on the YOLOv5 model using a specific dataset that consists of both normal and abnormal objects, we are able to customize the network to achieve superior performance in accurately identifying anomalies. In order to improve the efficacy of the model, we introduce a unique loss function that integrates both classification and localization losses. This approach aims to optimize the model's capacity to accurately identify and precisely determine the location of anomalous items. In addition, we investigate the potential of transfer learning to enhance the model's abilities in effectively addressing a wide range of scenarios and object categories. The empirical findings of our study provide evidence supporting the efficacy of the YOLOv5-based deep learning framework for the timely identification of anomalous items in both dynamic video streams and stationary photos. The model integrates YOLOv5 with deep learning techniques for accurate object detection, minimizing false alarms and ensuring prompt anomaly identification, making a significant contribution to computer vision. The proposed system, utilizing deep learning and YOLOv5, is suitable for various applications like security, surveillance, industrial automation, and healthcare. It addresses the need for efficient object detection in complex environments, contributing to cost-effective, digitally fortified facilities.

Keywords: Deep Learning, Object Detection, YOLOv5, Abnormal Object Detection, Computer Vision.

I. INTRODUCTION

1.1 Research Background and Significance

1.1.1 Research background

The realm of real-time abnormal object detection holds immense significance across various sectors, including security surveillance, industrial automation, and healthcare [1, 2, 3]. Ensuring timely identification of abnormal objects is paramount in averting potential threats [1], preventing equipment failures [2], and facilitating prompt medical intervention [3]. Traditional methods for abnormal object detection often grapple with the complexities posed by dynamic environments and diverse anomaly manifestations. This has fueled the adoption of deep learning techniques, particularly Convolutional Neural Networks (CNNs), which excel in learning hierarchical features directly from data [4, 5]. CNNs have demonstrated remarkable success in tasks such as image classification and object detection [4, 5]. However, abnormal object detection introduces unique challenges demanding tailored solutions. Anomalies can manifest subtly, fleetingly, or in intricate contexts,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

necessitating models adept at capturing multi-scale information for swift decision-making. This research endeavors to explore the effectiveness of the YOLOv5 model in abnormal object detection [6]. By leveraging its capacity to capture intricate features and contextual insights, this study aims to address the precision and efficiency challenges within anomaly detection systems.

1.1.2 Research Significance

The significance of this research lies in its endeavor to address the pressing challenge of real-time abnormal object detection, a crucial task with wide-ranging applications. In contexts such as security surveillance, the timely identification of abnormal objects is paramount for preempting potential threats and ensuring public safety [7]. Moreover, industries heavily reliant on automation, such as manufacturing, benefit from real-time anomaly detection systems that prevent costly equipment failures and production disruptions [8]. In the healthcare sector, the rapid detection of anomalies in medical images aids in timely diagnosis and treatment, thereby potentially saving lives [9].

Conventional methods often struggle to cope with the intricacies of dynamic environments and the myriad ways anomalies can manifest. The increasing adoption of deep learning techniques, particularly Convolutional Neural Networks (CNNs), underscores their potential to tackle such complexities [10, 11]. The versatility of CNNs extends from image classification to object detection, making them an attractive avenue for addressing real-time abnormal object detection challenges.

The utilization of the YOLOv5 model for aberrant object detection using deep learning holds significant research value. YOLOv5 epitomizes the forefront of real-time object identification technology, delivering unmatched levels of speed and precision. The utilization of abnormal object detection exhibits potential for significantly augmenting security in many areas. This technology facilitates the instantaneous recognition of irregular or suspicious items, hence fortifying security protocols and contributing to the deterrence of security breaches. Furthermore, the speedy and accurate detection capabilities of this technology have proven to be valuable in several industrial contexts. Its application in these settings has led to notable improvements in quality control, a reduction in production errors, and ultimately an enhancement in overall industrial efficiency [12]. In the field of healthcare, the utilization of YOLOv5, a deep learning model, has the potential to enhance the efficiency of anomaly detection in medical imaging. This, in turn, may result in expedited diagnoses and improved efficacy in treatment planning. The real-time analysis performed by the model also encompasses safety applications, including the detection of road dangers for autonomous vehicles and the maintenance of public safety during densely populated events. YOLOv5 optimizes efficiency and minimizes the probability of human error by implementing automated anomaly detection. This is particularly beneficial in situations when manual monitoring requires significant resources. Moreover, this study makes a valuable contribution to the forefront of deep learning and computer vision, propelling advancements in both practical applications and the wider domain of artificial intelligence research. It also paves the way for novel opportunities in innovation and exploration.

1.2 Research Aim and Objective

The primary aim of this thesis is to design, develop, and evaluate an efficient and robust real-time abnormal object detection system capable of accurately identifying and classifying abnormal objects in a variety of real-world scenarios. This research aims to contribute to the field of computer vision and machine learning by addressing the challenges associated with detecting and categorizing abnormal objects swiftly and accurately, with potential applications in security, surveillance, quality control, and safety across diverse domains.

This aim signifies that your research intends to create a practical, real-time solution for detecting abnormal objects and considers its significance in addressing real-world problems across multiple domains. It emphasizes the development of a system that balances accuracy, speed, and adaptability to various situations, which is a crucial aspect of real-time abnormal object detection.

This research seeks to create an efficient real-time abnormal object detection system for a broad spectrum of practical applications. It begins with an extensive review of relevant literature, followed by the collection and preprocessing of a diverse dataset. The research focuses on adapting and developing object detection algorithms to construct a system capable of swiftly identifying and categorizing abnormal objects in real-time. Performance evaluation and optimization aim to ensure both accuracy and efficiency. Additionally, user-

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

friendliness through interface design and ethical considerations are integral components. The research concludes with comprehensive documentation, contributing to the advancement of computer vision and machine learning solutions in various domains.

1.3 Motivation

The motivation behind the utilization of YOLOv5 for real-time abnormal item identification stems from the urgent requirement to enhance security and safety protocols in various industries. The prompt detection of atypical entities is crucial for taking proactive measures to minimize risks in modern settings characterized by the quick emergence of possible hazards. The real-time detection capabilities of YOLOv5 enable prompt responses to potential security threats, rendering it highly beneficial in security and surveillance contexts. Its application ensures the protection of the public in densely populated areas and helps mitigate the risks associated with violence or terrorism. Within industrial settings, this technology plays a crucial role in maintaining rigorous quality control standards, thereby minimizing the probability of defective items being released into the market and augmenting overall operational efficiency. Additionally, the expeditious detection capabilities of the model play a vital role in emergency response scenarios, assisting first responders in rapidly identifying potentially dangerous objects. Real-time anomalous item identification in the healthcare sector plays a crucial role in ensuring patient safety during medical procedures, effectively mitigating the risk of potential harm. The rationale for utilizing YOLOv5 stems from its ability to rapidly and precisely detect atypical objects, thereby augmenting security, safety, and operational effectiveness across a wide range of applications.

1.4 Structure

This dissertation will contain five chapters respectively identified below -

Introduction: Include information on the research's context, significance, purpose, inquiry, and motivation, as well as its structure.

Literature review: This literature review provides a comprehensive overview of current research in real-time abnormal object detection, outlining findings, methodologies, trends, challenges, and solutions, thereby advancing computer vision and security.

Methodology: Includes the many strategies and rules applied to achieve the goals of this dissertation, with a focus on the information and suggested models. The investigation's main experimental method is outlined in this section. The dataset is thoroughly detailed here. Learn more about the deep learning algorithm and YOLOv5

Results and Experiment: Where the development of the model is revealed, the accuracy of the various models is compared, and their applicability to various situations is investigated in greater depth.

Conclusion: A section that provides a simplified interpretation of the study's findings and how they relate to the primary objective of the dissertation. The model's shortcomings and the project's prospective applications are also discussed.

II. RELATED WORK

Abnormal object identification is a specialized domain within the subject of computer vision that focuses on the task of recognizing and localizing items that exhibit deviations from the anticipated or standard patterns within a given image or video. The approach of one-class classification involves the acquisition of a model exclusively from a single class of data, typically referred to as the normal class. Subsequently, this model is utilized to classify fresh instances of data as either normal or abnormal, employing a similarity or distance metric. Deep reinforcement learning is an approach that employs a neural network agent to acquire knowledge through its own actions and incentives in a given environment. Subsequently, this acquired policy is utilized to identify anomalies in novel scenarios. The utilization of deep neural networks involves the implementation of numerous layers of artificial neurons to acquire knowledge of intricate and nonlinear characteristics from extensive datasets. Subsequently, these acquired features are employed to identify anomalies within novel datasets.

2.1 Abnormality Data Idea, Characteristics, and Detection

The concept of abnormality data is a fundamental and adaptable one, with wide-ranging applications across various fields [7]. Fundamentally, this involves the crucial undertaking of identifying situations in which observed data deviates considerably from established standards or anticipated behavior [7]. The departure

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

from the established standard frequently results in atypical occurrences or inconsistencies, which can manifest in various ways and have significant consequences [8].

In the context of surveillance, abnormality data assumes significance in the detection of atypical behaviors in video recordings [8]. This phenomenon can involve a wide range of occurrences, including abrupt and unpredictable movements within a group of individuals as well as the positioning of things in unusual or unexpected positions. If left unnoticed, such atypical behaviors might present serious security hazards, underscoring the significance of comprehending abnormality data.

Similarly, within the complex realm of finance, the utilisation of abnormality data assumes a crucial function in identifying unconventional patterns or fraudulent behaviours inside financial data [9]. This may entail the detection of atypical trading patterns that could potentially indicate market manipulation or the identification of abnormal transactions suggestive of financial fraud [9].

The comprehension of abnormality data is not universally applicable and varies across different contexts. However, the extent to which this is true is heavily influenced by the environment and the particular field or use case being considered [7]. The perception of abnormality can vary depending on the specific context, leading to situations where behaviors or characteristics that are deemed abnormal in one setting may be judged normal or typical in another. In the healthcare field, it is important to note that aberrant vital signs for a baby may vary dramatically from those that are considered abnormal for an elderly patient [11]. The significance of domain knowledge and context-aware anomaly detection algorithms is highlighted by this contextual sensitivity.

Fundamentally, comprehending the intricacies of abnormality data entails more than simply detecting deviations. It involves discerning the significance and meaningfulness of these deviations, as well as their possible indication of underlying difficulties or threats. The comprehension of this concept serves as the basis for the development of efficient anomaly detection systems, the enhancement of security measures, the improvement of quality control procedures, and the assurance of the dependability of systems and processes in many applications [7].

2.3 Gun and Knife Detection Using a Traditional Algorithm

Traditional abnormal object detection algorithms rely on several frame segmentation techniques to identify and extract important information from images. This section provides a detailed description of them.

2.1.1 Active Appearance Model (AAM)

The Active Appearance Model (AAM) is a computer vision technique utilised for the purpose of modeling and analyzing objects, commonly employed in tasks such as facial recognition and facial feature tracking. The system comprises a geometric model that represents the structure of the object and a model that captures the texture of the object. The aforementioned models are integrated in order to depict the visual representation of an object in diverse circumstances, encompassing changes in posture, facial expression, and illumination. A deformation model incorporates and explains these variances. Adaptive appearance models (AAMs) are frequently employed in the domain of computer vision for the purpose of accurately aligning and monitoring objects inside photos and videos. Although AAMs possess significant capabilities, they can be susceptible to initialization and may necessitate further data to enhance their resilience. Consequently, they hold considerable value in diverse computer vision applications.

The AAM employs a statistical model for the purpose of feature matching [14]. The primary use of this technology is in the detection of facial features. Upon the completion of annotating the image, AAM proceeds to transform it into a vector representation [15]. Principal component analysis (PCA) is employed to normalize the pictures. T. Rohit et al. [13] conducted a study where they aimed to optimize the detection of false positives in knife identification. To achieve this, they trained the Active Appearance Model (AAM) using a custom image dataset. In order to enhance precision, it is possible for the AAM (Automated Analysis Model) to accurately detect blades with a keen cutting edge inside an image, provided that the objects in question are distinctly discernible. One of the limitations of AAM is its failure to accurately recognise objects in photos with high levels of noise.

2.3.2 Harris Corner Detector (HCD)

The Harris Corner Detector (HCD) is a commonly used technique that is employed for the detection of corners in images. The procedure entails the examination of the fluctuations in image intensity across different

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

orientations. The algorithm entails the assessment of a corner response function that is derived from the gradient of the image. This function quantifies the likelihood that a specific pixel represents a corner. Following this, the system proceeds to detect pixels with high response rates as potential corner points. It then utilizes a non-maximum suppression strategy to guarantee that only the most noticeable and distinctive corner points are retained. The HCD algorithm demonstrates rotation invariance and computational efficiency. However, it is vulnerable to variations in lighting conditions and may have difficulties distinguishing between corners and conspicuous edges in some situations.

The extraction of features by HCD involves the utilization of the corners of images. The Harris detection procedure encompasses a series of sequential steps. The initial step is doing grayscale conversion on the image. The identification of the corners of the picture is accomplished by the utilization of spatial derivatives. The tensor structure of the detected item is established by the application of Harris computation, and the subsequent identification of the object is achieved by means of non-maximum suppression. By employing the AAM (Active Appearance Model) and HCD (Histogram of Color and Depth) methodologies in conjunction,[16] successfully achieved a higher level of effectiveness in detecting guns and knives compared to the work conducted [13]. In their study. utilized a customized picture dataset specifically designed for training purposes. However, the process of real-time detection is characterized by a significant delay due to the extensive time required for processing.

2.3.3 Color-Based Segmentation (CBS)

Color-Based Segmentation (CBS) is a computer vision technique employed to divide images or video frames into distinct regions based on their color properties. The procedure involves the establishment of color thresholds, the clustering of analogous colors, or the initiation of region expansion from seed pixels in order to gather pixels with similar hues. The application of CBS has been documented in various domains, including object identification, image segmentation, biomedical imaging, and industrial quality control. While this approach is characterized by its simplicity and computational efficiency, its effectiveness relies on its capacity to accurately perceive distinctions in color. Additionally, it may be vulnerable to fluctuations in lighting circumstances and require parameter adjustments.

In order to identify the cluster of the subset, the Cluster-Based Subset (CBS) method employs the k-means algorithm [17]. Once the reduction of undesired colors in the photographs has been performed, the item depicted in the image can be identified through the utilization of the HCD approach. In their study, researchers [13] and [14] utilized Human-Centered Design (HCD) and Cognitive Behavioral Science (CBS) methodologies, respectively, to discern and classify blades and weapons. The model underwent training using a distinct dataset of images. Human-Centered Design (HCD) was employed in order to identify the appropriate object following the removal of the unwanted color. The model underwent training using a dataset of lesser significance, resulting in limitations in its ability to accurately recognize X-ray images as it exhibited a higher rate of false positive identifications.

2.4 Gun and Knife Detection Using Deep Neural Networks (DNNs)

Deep neural network (DNN) learning techniques are constructed based on the foundation of neurons [18]. A deep neural network (DNN) consists of multiple levels, with each layer consisting of neurons. Each neuron is characterized by input points, hidden points, and output points. The layers in question are interconnected with one another on the basis of their interdependence. Regarding the weights of neurons, it is important to consider their significance in the context of neural networks. The output of the preceding neuron serves as the input for the subsequent neuron in the neural network. The layer is multiplied by its corresponding weight. The summation of all values is performed, and Incorporated into the established bias metric The sum that is acquired serves as an input for the subsequent neuron. The resulting value is subsequently sent to an activation function that performs a transformation on the parameters. The electrical signal is transmitted to the subsequent neuron.

Similarly, all input values are propagated throughout the system. The complete neural network As a result, the utilization of neural networks is employed for the purpose of predicting the outcome. The user's text is already academic. The discrepancy between the anticipated result and the actual outcome is referred to as an error, and it is computed by using a The error function is derived from the error value that arises during the process of weight updates. The technique is iteratively performed until the resulting error is minimized. A deep learning

algorithm is a computational model that utilizes multiple layers of artificial neural networks to extract and learn hierarchical representations of data. The utilization of object detection involves the identification of an entity by the analysis of its specific characteristics, which have been previously established as the basis for its recognition. The neural network is a computational model inspired by the structure and function of the human brain. The convolutional neural network (CNN), Overfeat, and region-based CNN are among the architectural approaches utilized for the detection of firearms and knives.The three object detection models under consideration are R-CNN, Fast R-CNN, and Faster R-CNN.

2.4.1 Neural Network (NN) Models

Neural network models have become a crucial cornerstone in the field of real-time anomalous item recognition. The current conceptual framework provides a thorough analysis of neural network models, including their structural configuration, learning mechanisms, practical applications, and the inherent challenges faced in real-time detection of abnormal items.

The primary element of neural network models is a network composed of artificial neurons arranged into discrete tiers, including the input layer, hidden layers, and output layer. Each individual neuron in a neural network is tasked with processing incoming data, which is subsequently passed to the next layer through weighted connections. The purpose of these connections is to introduce non-linearity by incorporating activation functions. The assignment of weights to the links within the network is of utmost importance in assessing their relevance and exerting an influence on the information flow.

The acquisition of knowledge and the development of skills are key components in the operation of neural network models. Throughout the training procedure, the neural network is exposed to annotated data and subsequently adjusts the weights of its connections with the objective of minimizing the disparity between its predictions and the actual target values. The methodology involves the application of forward propagation for forecast generation and backward propagation for weight adjustment, often incorporating optimization techniques like gradient descent. Neural network models provide the notable ability to autonomously learn relevant features from data, hence diminishing the necessity for manual feature engineering.

Deep learning is a subsection of neural networks that focuses on the application of models with multiple hidden layers, commonly known as deep neural networks. The architectural models showcased exhibit a remarkable level of expertise in the acquisition of intricate patterns and the systematic arrangement of data into hierarchical structures. Convolutional Neural Networks (CNNs) have exhibited competence in the processing of picture data for the objective of real-time identification of anomalous items. In contrast, Recurrent Neural Networks (RNNs) have demonstrated remarkable proficiency in efficiently handling sequential data, such as video streams and time-series data.

Neural network models provide a diverse array of applications in the domain of real-time detection and identification of abnormal items. Individuals are involved in a wide range of professional fields, including security and surveillance, manufacturing and quality control, healthcare, and finance. These models are of utmost importance in the identification of security threats, the detection of product failures, the monitoring of patient health data, and the prevention of fraudulent activities in real-time financial transactions.

Nevertheless, neural network models present many challenges. In order to attain optimal training outcomes in machine learning, it is imperative to take into account a multitude of aspects. Several variables must be considered in the context of deep learning. These factors encompass the need for a substantial quantity of labeled data, the meticulous selection of appropriate network architectures and hyper-parameters, and the imperative to tackle obstacles such as over-fitting. The comprehension of the capabilities of neural networks remains a significant challenge, particularly in sectors where safety is of utmost importance. This obstacle is derived from the inherent complexity of understanding the underlying reasoning behind a neural network's categorization of an object as abnormal, which may not always be immediately discernible.

In conclusion, it can be argued that neural network models have a significant impact on the prompt detection of atypical entities [18]. The ability to swiftly assess data and make immediate judgments has extensive consequences, enhancing security, ensuring quality assurance, and promoting safety across all domains.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

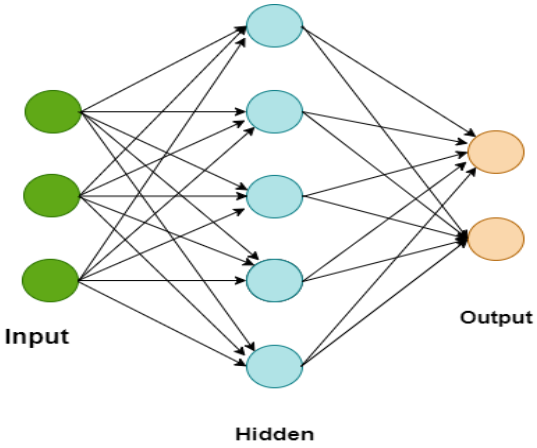

Figure 2-1 Artificial Neural Network (ANN)

2.4.2 Convolutional Neural Network (CNN) Model

A convolutional neural network (CNN) is a type of deep learning model that has been specifically engineered to evaluate and process data arranged in a grid-like structure, with a particular focus on data in the form of images. Convolutional layers are employed in order to extract features, whilst pooling layers are utilized to decrease the dimensionality of the data. Moreover, the implementation of completely connected layers is utilized with the explicit purpose of classification. Convolutional neural networks (CNNs) are extensively employed in the field of computer vision for many purposes, such as picture classification and object detection. Convolutional neural networks possess the ability to independently learn distinctive properties from input data, allowing them to effectively handle variations in scale and lighting conditions. Nevertheless, the utilization of these models necessitates a considerable quantity of annotated data and extensive computational resources for the purpose of training.

Convolutional neural networks (CNNs) are a kind of deep learning algorithms that are frequently utilized for the examination of visual data, which includes both photographs and movies. The utilized methodology involves the application of convolution, a technique that enables the automated acquisition and extraction of characteristics from input data. The final outcome of a Convolutional Neural Network (CNN) is attained through the application of a sequence of convolutional, pooling, and fully connected layers to the input data. In order to accurately identify and extract pertinent attributes from the provided input, each layer inside the network executes distinct operations [19] [20]. The primary kinds or elements of CNN are as follows:

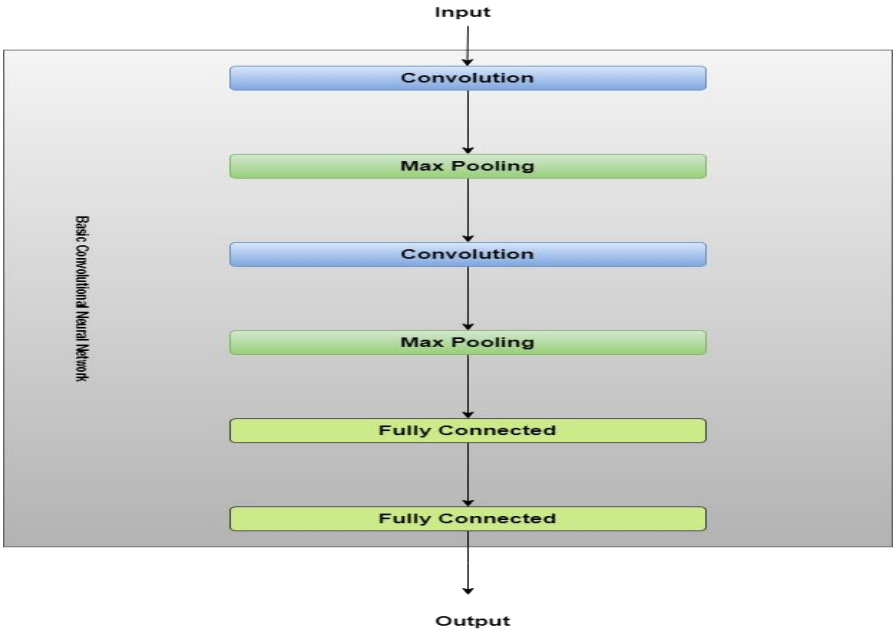

Figure Error! No text of specified style in document.**-1** Architecture of Basic CNN

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

- The convolutional layer performs the convolution function by utilizing a set of learnable filters, also known as kernels, to glide over the input data. This process aims to extract local patterns and characteristics from the input. The aforementioned layer generates a set of feature maps that depict different aspects of the supplied data.
- The pooling layer is a component that, after the convolution process, retains important information while reducing the spatial dimensions of the feature maps. One often employed pooling technique, known as "max pooling," aids in the extraction of the most significant features by preserving the maximum value inside each pooling region.
- The activation function is utilized in order to introduce nonlinearity into the network and enable the representation of complex interactions among features. Nonlinear activation functions, such as the Rectified Linear Unit (ReLU), are employed to process the output of convolutional and pooling layers.
- The fully connected layer(s) is employed to process the extracted high-level features and produce predictions subsequent to the convolutional and pooling layers. The inclusion of multiple layers in the network facilitates the acquisition of complex patterns and the generation of accurate predictions through the establishment of connections between neurons in one layer and neurons in all other layers.
- The softmax layer is commonly employed as the final layer in classification tasks to convert the output of the preceding layer into probabilities associated with different classes. Consequently, the network is capable of assigning a probability distribution across the classes.

2.4.3 Overfeat

The OverFeat architecture is a computational framework utilized in the domain of computer vision. It combines convolutional neural networks (CNNs) with traditional computer vision techniques to simultaneously accomplish object recognition, location, and classification tasks. The methodology being presented employs a sliding window approach to examine photos across different scales and positions. Furthermore, the approach integrates regression and class scoring techniques in order to improve the precision of item localization and classification. While OverFeat demonstrated impressive results during its time, its utilization of a sliding window approach can introduce substantial computational overhead, hence diminishing its effectiveness for real-time applications when compared to more modern object recognition methods.

The Overfeat algorithm is based on the convolutional neural network's (CNN) utilization of the sliding window technique. The central entity depicted in the image is the initial region of interest that the sliding window classifier is taught to identify. In their study, [21] successfully detected a firearm and achieved a favorable outcome as a result. Remarkably, despite the real-time identification of frames, the process continues to exhibit a significant degree of slowness.

2.4.4 R-CNN and Faster R-CNN

The Region-based Convolutional Neural Network (R-CNN) is a computer vision methodology utilized for the purpose of object detection. The process commences with the generation of region suggestions through the utilization of an external technique. Subsequently, features are extracted from these regions by employing a pre-trained Convolutional Neural Network (CNN). The application of object categorization and bounding box regression is utilized to ascertain the existence and precise positioning of things. Although the accuracy of R-CNN has been greatly enhanced, its computing demands are substantial as it heavily depends on external region recommendations. Consequently, this renders it unfeasible for real-time applications. The Region-based Convolutional Neural Network (RCNN) is a significant deep learning model utilised in the field of computer vision. It is specifically developed to address object detection and image segmentation applications. The year 2014 witnessed the emergence of RCNN, a pioneering development in the field of computer vision. Ross Girshick and his team successfully integrated deep learning techniques with object localization and classification, thereby achieving a notable advancement in this domain. The initial step involves the generation of region proposals, wherein a selective search method is employed to identify prospective locations of objects. This approach effectively reduces the computational burden. Subsequently, feature vectors are derived from the suggested regions through the utilisation of a pre-existing Convolutional Neural Network (CNN) that has undergone training. The feature vectors undergo individual processing by distinct neural networks to perform classification and localization tasks. The classification network is responsible for identifying the specific category of an object, such as a cat, dog, or car. On the other hand, the localization network is designed to

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

improve the accuracy of the region recommendations, enabling more precise detection of object boundaries. In order to reduce redundancy, the technique of non-maximum suppression is employed on the bounding boxes. While the RCNN model made significant advancements in object identification, later models such as Fast R-CNN and YOLO have further enhanced both speed and accuracy. These improvements have rendered them more suitable for real-time applications by incorporating innovative techniques like region of interest pooling and anchor-based region proposal networks. The following methods, such as Faster R-CNN, have endeavored to tackle these computational constraints.

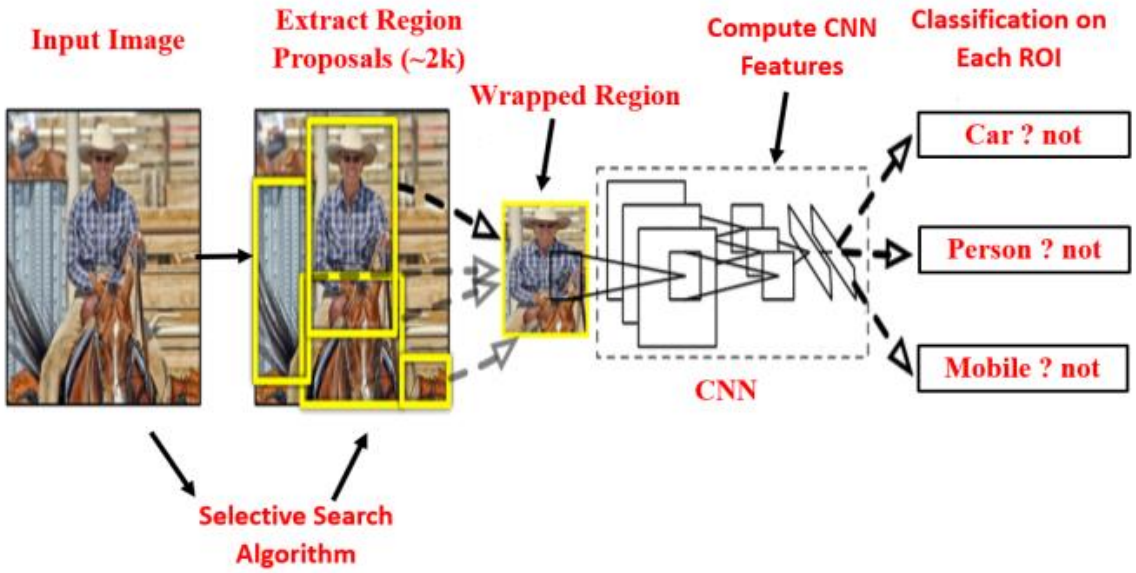

Figure 2-3 RCNN Architecture

The Faster R-CNN technique is a sophisticated approach in the field of computer vision that enhances the effectiveness of object detection by optimizing both the process of localizing objects and accurately classifying them. The proposed approach incorporates a Region Proposal Network (RPN) into the network architecture, allowing the model to create region proposals directly from feature maps and conduct object identification in a seamless way. This innovation optimizes the workflow and greatly improves computing efficiency in comparison to its predecessor, R-CNN. Although Faster R-CNN offers improved speed and practicality, it may not exhibit the same level of real-time performance as certain contemporary object recognition approaches. Additionally, it may possess certain limits when it comes to detecting small objects.

Dhillon et al. [22] suggested a handgun detection system by utilizing an R-CNN model with a classification head integrated into the VGG16 architecture. The model was trained using a dataset sourced from the internet movie firearm database (IMFDB). The authors employed an ensemble tree classifier and a support vector machine in their approach to perform classification, regression, and outlier identification tasks.

[23, 24] and [25] independently introduced a method for item detection and categorization that uses X-ray technology in their respective studies. The investigations conducted in this research explored various object detection algorithms, including sliding window CNN, You Only Look Once (YOLO), and Faster R-CNN. Based on the proposed approach, the object would be categorized into six distinct classifications, encompassing computers, firearms, their components, as well as knives. However, the proposed model was unable to detect objects. The R-CNN model served as a foundational framework for future object detectors, exerting a significant influence on their development. The R-CNN framework employs a greedy approach for idea collection, followed by a feature extraction backbone to extract supplementary features, and ultimately utilizes a Support Vector Machine (SVM) for classification. The integration of feature extraction techniques within a singular convolutional neural network (CNN) architecture resulted in a notable enhancement in the processing speed of the region-based convolutional neural network (R-CNN), as observed in the Fast R-CNN model. The Faster R-CNN algorithm was designed to incorporate a more intricate approach, which involved the integration of a distinct Region Proposal Network (RPN). This RPN effectively mitigated the need for a computationally intensive greedy algorithm, hence leading to notable improvements in processing performance. Notwithstanding these breakthroughs, the RPN's high computational cost continued to impede its progress. The

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

two steps of this process yield a high level of accuracy; however, this comes at the expense of detection speed due to the increased computational requirements.

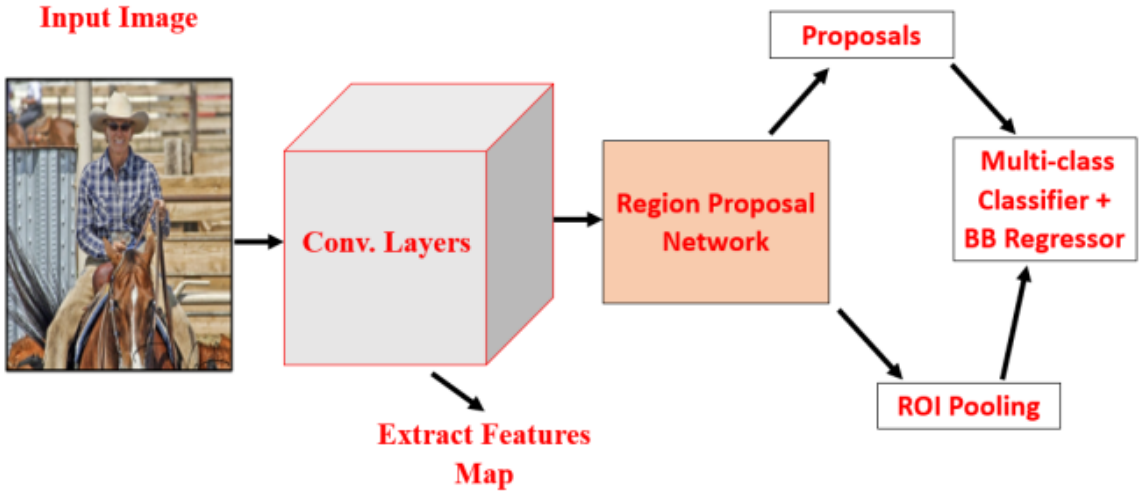

Figure 2-4 Faster R-CNN Architecture

2.4.5 Single Shot MultiBox Detector(SSD)

The SSD (Single Shot MultiBox Detector) is a computer vision algorithm specifically developed for the purpose of real-time object detection. The Faster R-CNN framework employs a region proposal network to generate bounding boxes, which are subsequently utilized for object classification. Although the precision of the technique is widely acknowledged, it operates at a rate of 7 frames per second, which may be considered a limitation. The current level of processing falls significantly short of meeting the demands of real-time applications. The utilization of SSD results in an acceleration of the process as it obviates the necessity for the region proposal network. In order to mitigate the decrease in accuracy, SSD incorporates several enhancements, such as the utilization of multi-scale features and default boxes. These enhancements enable SSD to achieve comparable accuracy to Faster R-CNN when utilizing lower resolution images, thereby significantly increasing its speed. According to the following comparison, it achieves the real-time processing speed and even beats the accuracy of the Faster R-CNN. The SSD approach utilizes a feed-forward convolutional network to generate a predetermined set of bounding boxes and corresponding scores to indicate the existence of object-class instances within those boxes. This is then followed by a non-maximum suppression phase to obtain the final detections. The first network layers are derived from a standardized design employed for the purpose of achieving accurate picture classification. This architecture, referred to as the base network, excludes any classification layers that may be present.

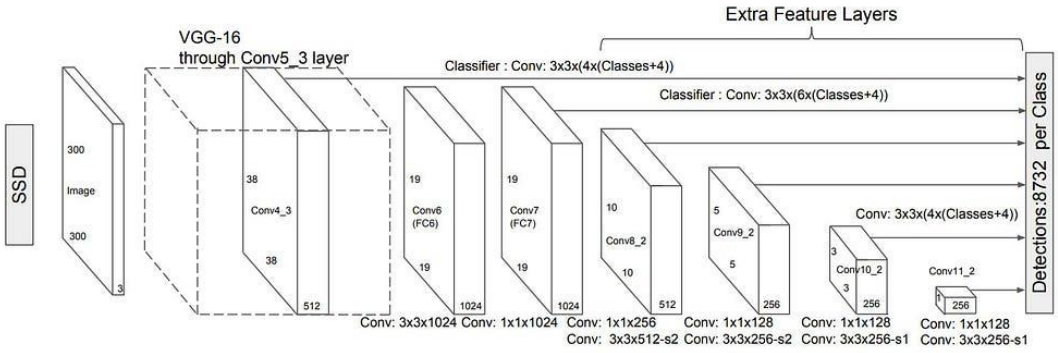

Figure 2-5 Single Shot MultiBox Detector (SSD)

The SSD object detection composes of 2 parts -

- Extract features map
- Apply convolution filters to detect object

The utilization of the VGG16 architecture is employed by SSD for the purpose of extracting feature maps. Subsequently, object detection is performed by using the Conv4_3 layer. To provide a visual representation, we

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

depict the Conv4_3 as having dimensions of 8 × 8 in terms of spatial extent, while it should actually be 38 × 38. Each individual cell, referred to as a location, generates four predictions for objects.

The utilization of a delegated region proposal network is not employed in the case of SSD. On the contrary, it can be concluded that it adheres to a straightforward approach. The computation involves the utilization of compact convolution filters to determine both the spatial coordinates and classification scores. Once the feature maps have been extracted, SSD utilizes 3 × 3 convolution filters for each cell in order to generate predictions. The aforementioned filters perform computations in a manner similar to conventional CNN filters. The output of each filter consists of 25 channels, which include 21 scores for each class and one boundary box.

In the Conv4_3 layer, a set of four 3 × 3 filters is utilized to transform the input with 512 channels into an output representation with 25 channels.

2.4.6 YOLO

The YOLO neural network is a singular neural network that produces bounding boxes, and the model is constructed in a solitary evaluation. The Deep Learning era represents the initial phase of object detection. In the YOLOv2 framework, several enhancements were incorporated, including batch normalization, customizable frequency input, anchor boxes, a better Dark-net backbone, and the utilization of ideal features. The enhancements made to the Common Objects in Context (COCO) dataset resulted in comparable Mean Average Precision (MAP) scores to those achieved by the Single-Shot Detector (SSD) method, while exhibiting a threefold increase in speed. The primary objective of the YOLOv5 model was to enhance several aspects, including accessibility, learning speed, inference speed, and deployment convenience. The current iteration of the YOLO algorithm is implemented using the PyTorch framework, in contrast to its previous iterations that were built using the Darknet framework. The user did not provide any text to rewrite [30].The YOLO algorithm for handgun identification was proposed by [26]. The dataset utilized for training purposes was the IMFDB dataset. The most relevant finding about the detection of a knife was derived from the Common Object in Context (COCO) challenges that were published in 2017. The object detection methodology employed in the COCO dataset [27] was founded upon a dataset of considerable magnitude. Developed a tailored dataset to train the YOLOv4 model and conducted a comparative analysis with the current leading approach [27]. A satisfying outcome was attained by the experimental evaluation of their methodologies on a limited number of videos. The primary emphasis of their research revolved around the identification and detection of a firearm, specifically a pistol and revolver, as well as common everyday items such as a wallet, metal detector, and cell phone. The researchers conducted a comparative analysis, examining the outcomes of the single shot multibox detector (SSD), R-CNN, and an alternative iteration of YOLO. While certain classification models demonstrated encouraging outcomes in static conditions, their performance was hindered in real-time scenarios, exhibiting reduced speed and accuracy when operating on a device with limited resources. The aforementioned studies shown a commendable F1 score when applied to the original dataset. However, it is important to note that these models may not be well-suited for scenarios including the presence of background objects.The Faster R-CNN model [28] demonstrated the most efficiency in pistol detection, whereas the CNN model [29] proved to be most effective in detecting knives. Due to the limitations of current algorithms, the detection of smaller items poses a challenge, particularly when applied to lightweight devices with limited computational resources.

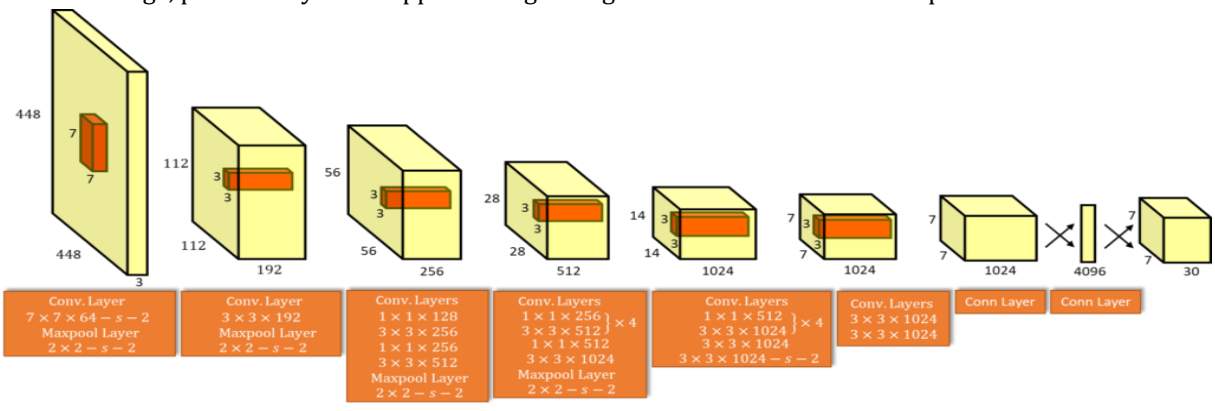

Figure 2-6 YOLO Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

2.4.7 YOLOv5

YOLOv5 [31] represents the most recent advancement in the YOLO series, serving as a cutting-edge single-stage algorithm for object detection. The YOLOv5 network is composed of three primary components, namely the Backbone, Neck, and Head. The Backbone comprises a convolutional neural network that aggregates and constructs picture representation characteristics at varying levels of detail. The neck of the architecture is comprised of multiple layers that combine and use visual features in order to advance prediction. In a similar manner, the head leverages characteristics from the neck to acquire both box and class prediction capabilities. The CSPDarknet53 backbone utilized in YOLOv5 consists of 29 convolutional layers with a size of 3 x 3. The receptive field of this backbone measures 725 × 725, and it encompasses a total of 27.6 million parameters. In addition, the integration of the SPP block into YOLO's CSPDarknet53 architecture enhances the coverage of receptive fields without compromising its computational efficiency. Similarly, the process of feature aggregation is executed using the PANet technique, which leverages several tiers of the backbone. YOLOv5 significantly advances the current state-of-the-art by incorporating several innovative features, including weighted residual connections, cross-stage partial connections, cross mini-batch normalization, and self-adversarial training. These features collectively contribute to the extraordinary efficiency of the YOLOv5 model. In the present investigation, the YOLOv5 model was trained and implemented on the PyTorch framework [32].

2.4.8 YOLOv5 Network Architecture

Convolutional Neural Network (CNN)-based object detectors have demonstrated significant potential in the domain of recommendation systems. YOLO (You Only Look Once) models are commonly employed for the purpose of object detection, exhibiting exceptional performance. The YOLO algorithm employs a grid-based approach to partitioning an image, wherein each grid cell is responsible for detecting items contained within its boundaries. Real-time object detection can be facilitated through the utilization of data streams. The computational resources required are few. The comprehensive architecture of YOLOv5 [33] is visually depicted in Figure 1.The YOLO series of models comprises three primary architectural components, namely Backbone, Neck, and Head.

- YOLOv5 Backbone: The proposed method utilizes CSPDarknet as the underlying framework for extracting features from photos, which are composed of cross-stage partial networks.
- YOLOv5 Neck: The system utilizes PANet for feature generation and the Pyramids network for feature aggregation, which is then passed to the Head for prediction.
- YOLOv5 Head: The system possesses multiple layers that create predictions based on the anchor boxes for the purpose of object detection.

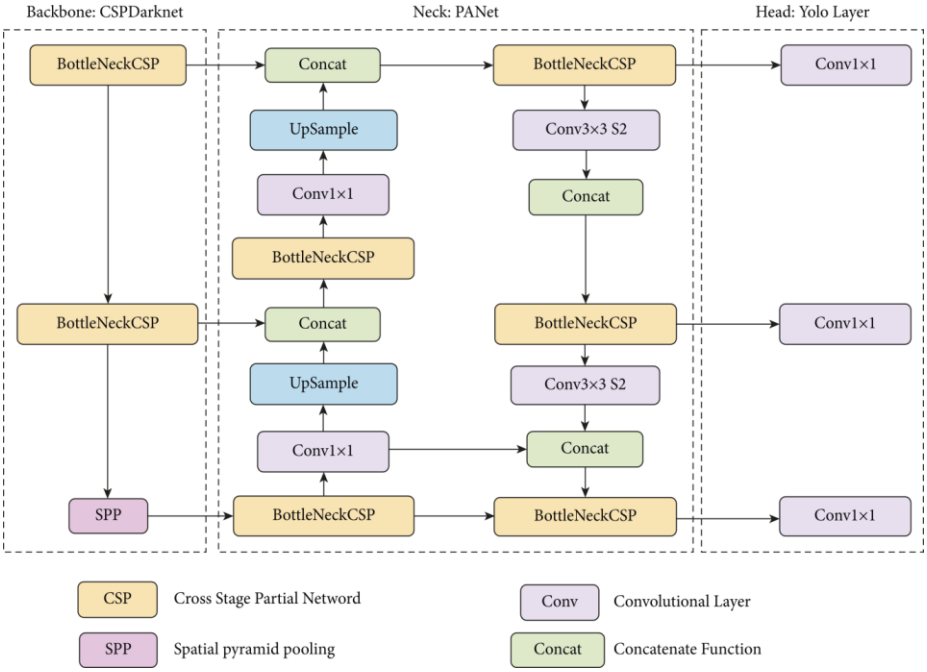

Figure 2-7 YOLOv5 Network Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

III. MATHEMATICAL EQUATIONS, SUBSECTIONS, TABLES, AND FIGURES

3.1 Overview

The proposed methodology for real-time abnormal object identification entails a systematic approach to formulating and executing experiments with the objective of promptly identifying and categorizing abnormal items or occurrences in real-time through the utilization of deep learning models. The following is a comprehensive outline of the fundamental procedures and factors to be taken into account within this particular technique. The example of abnormal object detection in figure 3-1.

Figure 3-1 Person with knife

The acronym YOLO, which stands for You Only Look Once, refers to a cutting-edge algorithm used for the purpose of real-time object detection. A single convolutional neural network (CNN) has the capability to detect and determine the positions of numerous objects inside an image or video. In contrast to alternative approaches for object identification that involve a series of sequential stages, such as region proposal, feature extraction, and classification, the You Only Look Once (YOLO) method does all of these tasks simultaneously within a single run through the Convolutional Neural Network (CNN). This feature enhances both speed and accuracy significantly. The fundamental concept underlying YOLO involves partitioning the input image into a grid comprised of cells. For each individual cell, the objective is to forecast a predetermined quantity of bounding boxes, together with their related confidence scores and class probabilities. The rectangular regions known as bounding boxes serve to enclose the objects, while the confidence scores provide an indication of the likelihood that the boxes contain an object. The class probabilities provide an indication of the type of object that is included within the box. The ultimate result of the YOLO algorithm is a collection of bounding boxes accompanied by their respective confidence scores and class labels. The Intersection over Union (IoU) metric quantifies the extent of intersection between two bounding boxes. Object detection jobs often include the utilization of a widely employed technique, wherein the objective is to accurately determine the location and classification of items within an image. The Intersection over Union (IoU) metric is computed by dividing the area of intersection between two bounding boxes by the area of their union. The formula for IoU is:

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad \text{...(3.1)}$$

Here is the more detailed breakdown –

Are of intersection($A \cap B$): The aforementioned region denotes the area of overlap between the bounding box predicted by the model and the ground truth bounding box.

$$A_{\text{intersection}} = \min(x_A, x_B) \times \min(y_A, y_B) \times \max(0, \min(x_A + w_A, x_B + w_B)) - \max(x_A, x_B)) \times \max(0, \min(y_A + h_A, h_B + h_B) - \max(y_A, y_b)) \quad \text{...(3.2)}$$

Area of Union ($A \cup B$): The aforementioned encompasses the entirety of the area encompassed by both the anticipated bounding box and the actual bounding box.

$$A_{\text{union}} = A_{\text{box}_1} + A_{\text{box}_2} - A_{\text{intersection}} \quad \text{....(3.3)}$$

Here A_{box_1} and A_{box_2} are the areas of the predicted and ground truth bounding boxes, respectively.

$$\text{IoU} = \frac{A_{\text{intersection}}}{A_{\text{union}}} \quad \text{...(3.4)}$$

The area of intersection refers to the region in which the two boxes overlap, whereas the area of union denotes the region encompassing both boxes. The IoU (Intersection over Union) metric is a numerical number that falls

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

within the range of 0 to 1. A value of 0 indicates that there is no overlap between two entities being compared, while a value of 1 signifies a complete and perfect overlap. A greater Intersection over Union (IoU) value signifies a stronger correspondence between the anticipated bounding box and the ground truth bounding box. In order to determine the area of intersection, it is necessary to ascertain the coordinates of both the upper left and lower right vertices of the intersected rectangle. The process involves conducting a comparison between the coordinates of the two boxes, followed by determining the maximum values for the top left corners and the minimum values for the bottom right corners. The calculation of the union area involves the summation of the areas of both boxes, followed by the subtraction of the area of their intersection.

3.2 Data Collection

I have dedicated almost two months to the analysis and refinement of my dataset. Due to the presence of unsuitable data and insufficient hardware support, I encountered limitations in conducting my experiments. I have captured videos and photographs within the international student apartment using a fixed distance and resolution. A dataset was created with the participation of over 20 students. I captured the movies and photographs from various perspectives. I utilized a mobile device to generate the dataset. The video resolution was 1440 x 1920 pixels. During this particular stage of the thesis, our focus has been on the identification and detection of objects that pose a potential hazard to human safety, including bladed weapons, cutting instruments, incendiary devices, and similar items. We procured images from online sources and subsequently categorized and provided annotations for them. In addition, we utilized tangible visual representations of accessible tools and recorded films through the utilization of our smartphone device.

3.3 Dataset

This thesis aims to identify and classify anomalous objects. This is a custom dataset. I have generated a proprietary dataset for the purpose of training and evaluating our object detection models. I have compiled data from multiple sources to create my dataset. The dataset comprises a total of eleven objects. The items under consideration include a gun, knife, a knife cover, normal person, person with gun, person with knife, person with scissor, person with stick, person with stick and knife, Scissor, stick. In the dataset 42 videos and 431 images. The video 30fps and resulation is 1440 × 1920 and image size is 640 × 640 Initially, we recorded recordings of ourselves in possession of atypical items within the confines of our hostel. I captured the videos from various perspectives. The custom dataset is generated from a university apartment for further experimental purposes. I ensured that a comprehensive collection of photos and videos was obtained from different vantage points, while also ensuring that all images adhered to consistent dimensions. Once the data was appropriately labeled, I proceeded to transform it into the proper forms. As an illustration, in the case of the TensorFlow Object Detection API, the data was transformed into the TFRecord format. An attempt was made to generate a dataset with the purpose of autonomously identifying potential dangers within an enclosed setting, such as the intrusion of an individual carrying a knife within the confines of a residential facility. The identification of these potential risks will be accomplished through the utilization of object detection techniques.

3.4 Data pre-processing

The processing of data is a crucial component in the detection of anomalous objects. The process entails converting unprocessed data into a structure that is appropriate for subsequent analysis and modeling. During the data processing I had to consider about some rule such as data augmentation, random cropping, scaling, flipping, contrast, lightness, etc.Data augmentation is a methodology employed in machine learning to expand the training set by generating modified replicas of a dataset through the utilization of existing data. This process encompasses the implementation of slight modifications to the dataset or the utilization of deep learning techniques to generate novel data instances. Augmented data is derived from the original dataset through the incorporation of slight modifications. Image augmentation involves using geometric and color space modifications to expand the training set, thereby enhancing its size and diversity. Synthetic data is produced using artificial means without reliance on the original dataset. The process of random cropping involves the creation of novel images through the random selection of a smaller section from an original image. This methodology proves to be advantageous in the realm of data processing, particularly in the context of training deep learning models that are designed to execute tasks such as object detection or image

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

categorization. The implementation of random cropping techniques serves to enhance the diversity and magnitude of our dataset, mitigate the potential for overfitting, and enhance the overall generalization capabilities of our models. Scaling refers to the act of altering the size of an image to achieve specific dimensions, such as width and height. The utilization of scaling techniques can effectively mitigate the computational burden and memory requirements associated with the processing of high-resolution photographs. Additionally, scaling contributes to enhancing the uniformity and consistency of images, hence facilitating their analysis.

Flopping refers to the act of horizontally or vertically flipping an image, hence generating a mirror image. The act of flopping can serve to enhance the dataset by introducing additional variances and diversity. Additionally, it can be utilized to rectify the orientation of some photos that may be inverted or rotated. Contrast refers to the variation in luminance levels observed between the brightest and darkest regions within an image. The visibility and clarity of picture features, such as edges, forms, and colors, can be influenced by contrast. Modifying the contrast has the potential to augment or diminish the level of detail in an image, contingent upon the intended outcome. The lightness of an image refers to its overall brightness and is subject to various factors such as illumination, exposure, and color. The luminosity of a picture has the potential to influence both its emotional and tonal qualities, as well as the way in which its content is perceived. Adjusting the lightness has the potential to achieve equilibrium or alter the visual representation of the image, contingent upon the desired objective.

Figure 3-2 Person with abnormal object

3.5 Training

The frequency at which the network is exposed to the complete training dataset. The augmentation of epochs has the potential to enhance the accuracy of the network; nevertheless, it concurrently amplifies the susceptibility to overfitting and necessitates a greater allocation of computational resources. The number of training epochs was set to 100. The quantity of training samples that are processed during a single iteration. Increasing the batch size has the potential to diminish the noise present in the gradient estimates and enhance the training process. However, this approach necessitates a greater amount of memory and may result in poorer generalization performance. During the training phase, the batch size used was 16. The momentum

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

component is associated with the Adam optimizer. Momentum is a beneficial factor that enhances the optimizer's ability to navigate across narrow valleys and noisy gradients within the loss landscape, resulting in improved efficiency and faster convergence towards the global minimum of the loss function. The range of normal momentum values is between the intervals of 0.9 and 0.999. The threshold for the intersection over union (IoU) utilized in the non-maximum suppression (NMS) algorithm. The NMS method effectively removes redundant and overlapping bounding boxes, retaining only those with the highest confidence. The IoU threshold is a parameter that establishes the degree of overlap permissible between two bounding boxes in order to classify them as representing the same item. Increasing the IoU threshold has the potential to decrease the occurrence of false positives, however, it also has the inherent drawback of potentially overlooking certain actual positives. Whereas the value of IoU is 0.5. We have undergone training utilizing the YOLOv5 model. Which exhibits exceptional performance. Initially, data was generated within the confines of our dormitory. Subsequently, proceed to categorize and provide explanatory notes for the given text. Subsequently, iterations are conducted on the aforementioned entity. Upon completion of 100 epochs, the mean Average Precision (mAP) score achieved by the YOLOv5 model is 94.8, indicating a commendable performance. The dimensions of the input images provided as input to the neural network. Increasing the scale of the image can enhance the network's capacity to identify objects that are both small in size and located at a considerable distance. However, this augmentation also results in elevated computational expenses and memory use. Where were images with a size of 640 located.

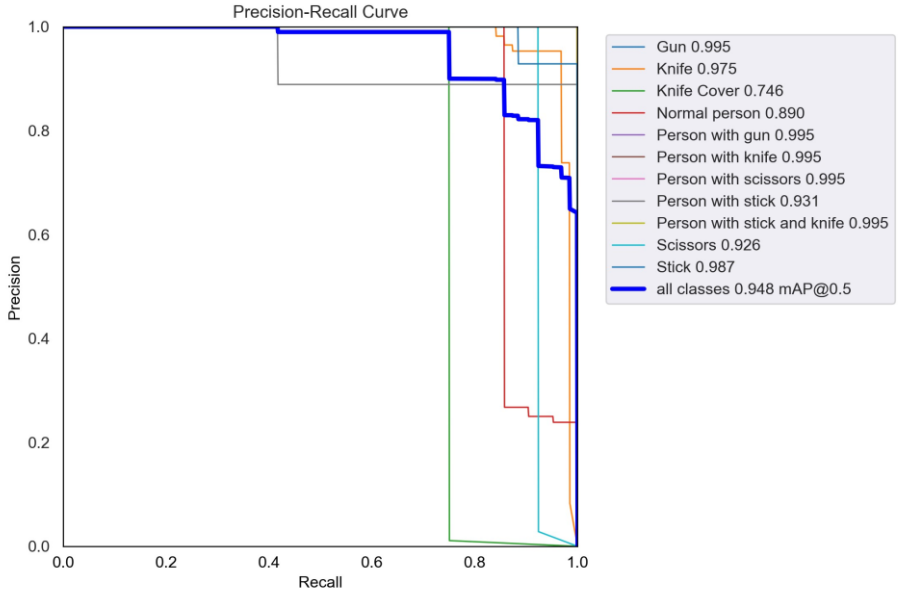

Figure 3-3 YOLOv5 MAP

3.6 Evaluation Metrics

In the assessment of object detection models, the evaluation metric chosen is mean average precision (mAP). Object recognition techniques like YOLO, SSD, and FRCNN commonly employ mean average precision (mAP) as a metric to assess the performance of their models before publishing their research. The problem of object detection is inherently complex. Consequently, it was necessary to employ precision, recall, and intersection over union (IOU) in our analysis. Precision is a metric that quantifies the degree of accuracy in a model's predictions by evaluating the proportion of correct predictions made by the model. In the context of this study, precision refers to the quantification of accurate identification and detection of potential threats, such as knife, or gun, in relation to the total number of detections made by the models under investigation. In this context, TP refers to the object detector successfully predicting a threat. The abbreviation "FP" denotes that the object detector model has performed a detection, but the detection outcome is deemed inaccurate.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad \ldots(3.5)$$

TP= True Positive (Predicted as positive as was correct)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

FP = False Positive (Predicted as positive but was incorrect)

The concept of recall pertains to the measurement of correct detections, including the inclusion of threats that were not able to be detected. A high recall rate indicates a greater likelihood of successfully detecting items, with a lower probability of failing to detect the intended target objects.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{...(3.6)}$$

FN = False Negetive

The Intersection over Union (IoU) metric is utilized to evaluate the precision of object detection. In this research study, an IoU threshold of 0.5 was employed, whereby any value below this threshold is classified as a false negative (FN), while those over it are considered true positives (TP). The calculation mostly involves determining the overlap ratio between the anticipated bounding box and the ground truth table.

$$\text{IoU} = \frac{\text{A}_{\text{intersection}}}{\text{A}_{\text{union}}} \qquad \text{...(3.7)}$$

The loss function employed in YOLOv5 encompasses a composite of terms that aim to tackle both object recognition and bounding box localization aspects. The loss function of YOLOv5 has been specifically built to optimize the model's performance in terms of both accurate classification and exact localization. The specific formulation may exhibit variations based on particular implementations and alterations; nonetheless, the fundamental framework encompasses elements pertaining to objectness confidence, categorization, and bounding box coordinates.

Objectness Confidence Loss (Binary Cross Entropy): The term in question quantifies the efficacy of the model in accurately predicting the existence or absence of an object within an anchor box. Optimizing this function is essential for YOLOv5's accuracy across different settings and datasets.

$$\begin{aligned}
&\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(x_i-x_i)^2+(y_i-y_i)^2] \\
&+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(\sqrt{w_i}-\sqrt{w_i})^2+(\sqrt{h_i}-\sqrt{h_i})^2] \\
&+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-C_i)^2 \qquad \text{....(3.8)} \\
&+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}(C_i-C_I)^2 \\
&+\sum_{i=0}^{S^2}1_{i}^{obj}\sum_{c\in clases}(p_i(c)-p_i(c))^2
\end{aligned}$$

The variable 1_i^{obj} represents whether an item occurs in cell i, while 1_{ij}^{obj} indicates that the jth bounding box predictor in cell i is responsible for that specific prediction. It is important to acknowledge that the loss function just imposes penalties on classification. An error occurs when an object is detected within the specified grid cell. Furthermore, the aforementioned approach just imposes penalties on bounding box coordinate errors when the predictor is deemed "responsible" for the ground truth box. This determination is based on the predictor having the highest Intersection over Union (IOU) value among all predictors in the corresponding grid cell.

IV. EXPERIMENT & RESULTS

4.1 Dataset Description

The dataset pertains to the detection of anomalous objects. The dataset provided is a customized dataset. I have constructed within our residential hall. The dataset is referred to as the International Student Apartment Dataset (Int_Std_Apt_dataset). The given subset of the dataset is utilized for the purpose of training the model designed for abnormal object detection. The dataset comprises a total of 431 photos, encompassing instances of both normal and aberrant items. Normal objects can be defined as the conventional or anticipated items that are commonly encountered in a given context. Conversely, abnormal objects can be characterized as deviations

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

or departures from the established norm. The inclusion of this dataset is of utmost importance for the model's acquisition and application of knowledge pertaining to the discernment and classification of both typical and atypical objects. The utilization of a validation dataset is essential in the process of refining and optimizing the model throughout the training phase. The dataset comprises a total of 123 photos that exhibit comparable features to the images present in the training dataset. These photographs encompass a variety of items, encompassing both normal and atypical instances. The purpose of this assessment is to evaluate the performance of the model and modify its parameters, including hyperparameters, in order to enhance its ability to accurately identify anomalous items.

The test dataset is reserved for the purpose of assessing the ultimate performance of the trained model. The dataset consists of 63 photos that were not included in the model's training or validation process. These photos are employed for the purpose of evaluating the model's capacity to generalize and identify anomalous objects in novel, unobserved data. The outcomes of the test dataset offer an evaluation of the model's accuracy, precision, recall, and other performance measures, which serve as indicators of its ability to identify anomalous items in real-life situations.

Table 4-1 Dataset values

Examine	image	labels
Train	**431**	**431**
Validation	**123**	**431**
Test	**63**	**-**

The integration of these training, validation, and test datasets is crucial in constructing and assessing a resilient anomalous object detection model. The process of data splitting is essential in enabling the model to effectively apply its learned knowledge from the training set to accurately predict outcomes on novel, unseen data. This capability is of utmost importance in practical scenarios involving abnormal item identification.

4.2 Hardware and Software Requirements

4.2.1 Hardware

The suggested hardware specifications, encompassing the NVIDIA RTX 4070 GPU, 16 GB of RAM, and a 12th generation Intel Core i7 processor, constitute a resilient and harmonious configuration that is exceptionally well-suited for a diverse array of research endeavors, particularly within the domains of machine learning, deep learning, and computer vision. The RTX 4070 GPU demonstrates exceptional capabilities in the field of high-performance computing, rendering it highly suitable for demanding computational activities such as building deep learning models and performing real-time object identification. The inclusion of 16 GB of RAM satisfies the minimum prerequisites for several research endeavors; however, larger datasets or intricate models may derive advantages from more memory. The 12th generation Intel Core i7 processor provides significant multi-core performance, which is crucial for CPU-intensive tasks involved in data preparation, model training, and inference. The integration of this particular hardware combination offers the requisite processing capacity to effectively address a diverse range of research objectives, hence establishing a robust framework for doing thesis work. However, it is possible to modify the hardware requirements based on the unique intricacies and requirements of the research endeavor.

4.2.2 Software

The development approach for real-time anomalous item identification can be streamlined by utilizing a software setup that involves Visual Studio Code (VS Code) and Anaconda. Visual Studio Code functions as the integrated development environment (IDE) utilized for the purpose of composing, evaluating, and troubleshooting code pertaining to the domain of object detection. The Anaconda distribution, which is a Python distribution, offers a streamlined approach to package management and offers a user-friendly environment for the management of Python programs, virtual environments, and tools related to data science. Jupyter Notebooks, when integrated with Anaconda, provide a comprehensive platform that facilitates interactive and visual data exploration, model prototyping, and documentation. Deep learning frameworks such as TensorFlow and PyTorch offer seamless integration for the purposes of model construction, training, and

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

deployment. OpenCV, frequently utilized in conjunction with Anaconda, offers fundamental functionalities for the analysis of images and videos, facilitating the real-time detection of objects from video streams and camera input. Data annotation tools such as Label Img or VGG Image Annotator (VIA) are utilized to facilitate the process of annotating data before model training. The integration of Git for version control and the inclusion of real-time code editing and debugging capabilities in VS Code significantly enhance the efficiency of developing real-time abnormal object detection models. This software combination facilitates the entire development process, encompassing data preprocessing, model creation, and deployment, while ensuring consistency in package management and environment configuration.

4.3 Package Requirements

In order to successfully execute the project, it is imperative that we get the necessary packages. For instance -

- Pandas: Pandas is a versatile open-source Python library for data manipulation and analysis, enabling tasks like data cleaning, transformation, and analysis. It is known for its ease of use and integration with other libraries.
- Numpy: NumPy is a key Python library for numerical computing, enabling efficient data manipulation and handling large datasets. It serves as the foundation for numerous scientific and data analysis libraries.
- Opencv-python: OpenCV-Python is a popular Python library for computer vision and image processing, offering tools for image and video analysis, object detection, and facial recognition. It's widely used in machine learning, robotics, and computer vision applications, offering versatility and extensive documentation.
- PyTorch: PyTorch is a popular Python-based machine learning framework, known for its flexibility and dynamic computation graph, ideal for deep learning research and development, offering tools for neural network training and multi-dimensional tensors efficiency.
- Seaborn is a Python data visualization library, built on Matplotlib, aimed at creating visually appealing statistical graphics for complex datasets, simplifying the process of creating various plots.
- Tensorflow: Google's TensorFlow is an open-source machine learning framework widely used for building and training deep learning models, offering versatility and robust community support.

4.4 Results

The mean Average Precision (mAP) score of 94.8% serves as a robust measure of the system's proficiency in reliably identifying objects within pictures or video frames. The mean Average Precision (mAP) is a commonly employed evaluation metric in the field of object identification. A substantial mAP score indicates that the system attains a notable degree of precision in both object recognition and localization. This is particularly evident when employing an IoU (Intersection over Union) threshold of 0.5. This accomplishment holds considerable significance, especially in domains where utmost accuracy and dependability are of utmost importance, such as surveillance, autonomous cars, and quality control.

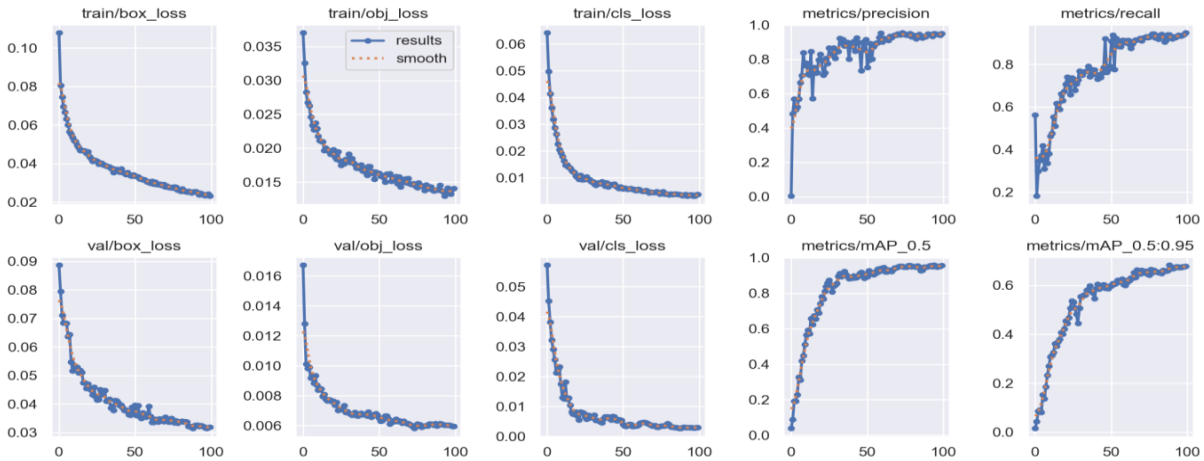

Figure 4-1 Labels correlogram

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

The measurement of 40.9 milliseconds represents the level of processing efficiency exhibited by the system. In the context of real-time object detection, achieving low latency is of utmost importance. The system's ability to analyze video frames or images at a speed of 40.9 ms indicates its capability to perform this task rapidly. This holds particular significance in scenarios that require prompt reactions, such as real-time security surveillance. The findings illustrate a remarkable equilibrium between precision and efficiency in the detection of objects in real-time. The system demonstrates a significant level of accuracy, achieving a mean average precision of 94.8% at a threshold of 0.5. Additionally, it maintains a rapid processing speed of 40.9 milliseconds per frame. These attributes render it highly suitable for various applications that prioritize both precision and real-time responsiveness.

Figure 4-2 Person with knife

Table 4-1 dataset information

Parameter	Value
Batch size	16
Image size	640 * 640
Epoch	100

4.4.1 Labels

In Figure (4-3), a comprehensive analysis has been presented regarding the various designations assigned to objects. In my experiment, I have a total of eleven objects that are utilised for the purpose of detecting abnormalities. Presented below is the figure that will be elucidated.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

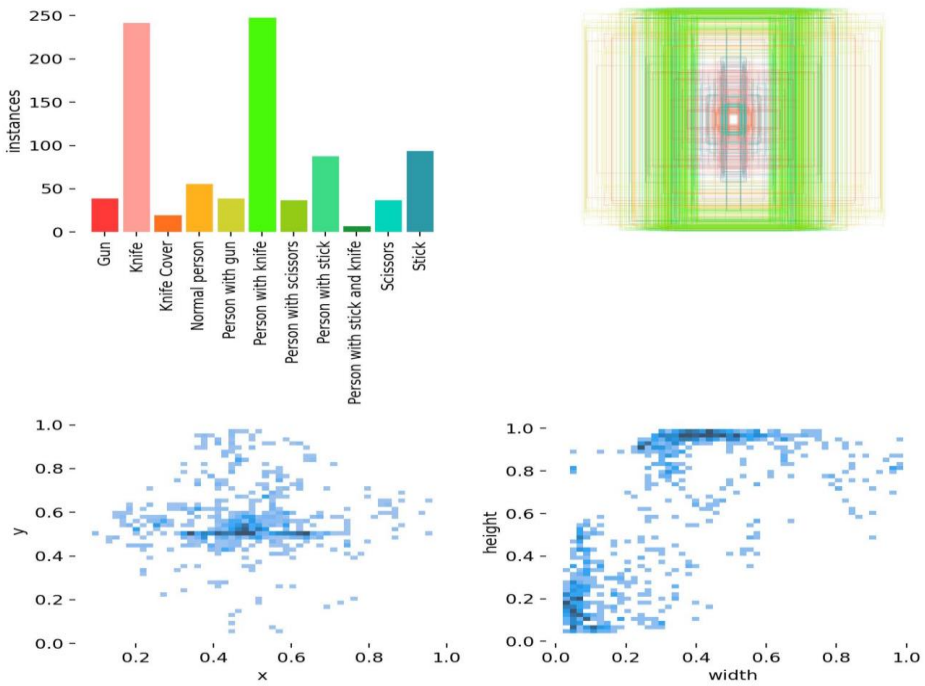

Figure 4-3 Instances Label

In this context, a variety of object classifications can be observed. The object's highest label is "Person with knife," followed by the label "Knife." Furthermore, the detection of the position of a stick and a person holding a stick was observed. The lowest rank is held by an individual equipped with a stick and a knife.

4.4.2 Precision

Precision is a quantitative measure that assesses the level of accuracy in the predictions made by a model. It is determined by calculating the ratio of true predictions made by the model to the total number of predictions. In the present study, accuracy is defined as the measurement of the correct identification and detection of potential threats, such as fire, knife, or gun, relative to the total number of detections made by the models being examined. In the present context, TP denotes the accurate identification of a potential danger by the object detection system. The acronym "FP" signifies that the object detector model has executed a detection, however, the result of the detection is considered to be imprecise.

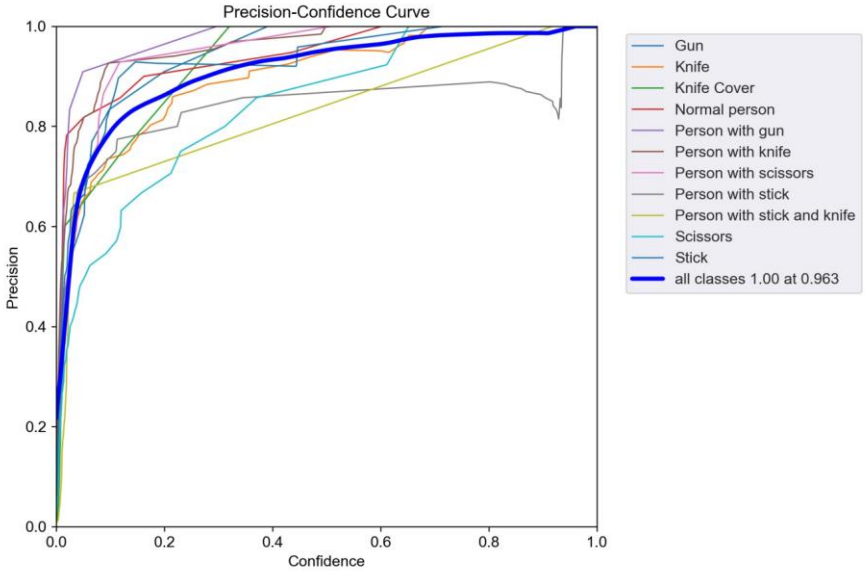

Figure 4-4 Precision

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

4.4.4 Recall

The notion of recall is concerned with quantifying accurate identifications, encompassing the incorporation of undetected dangers. A higher recall rate is indicative of an increased probability of successfully detecting items while reducing the likelihood of failing to detect the intended target objects. The outcome yielded a percentage . Figure (4-4) as depicted below.

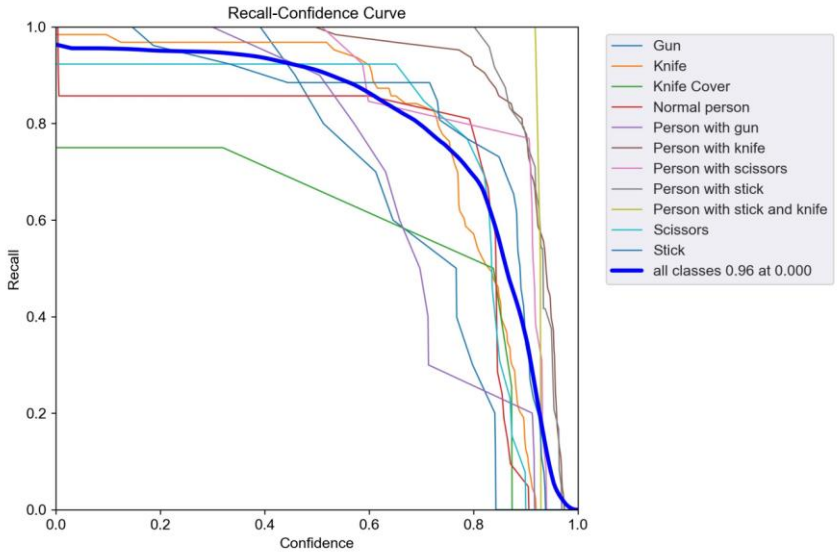

Figure 4-5 Recall-Confidence Curve

4.4.4 Precision-Recall Curve

The Precision-Recall metric serves as a valuable indicator of prediction performance, particularly in scenarios when there is a significant imbalance between classes. In the field of information retrieval, precision is a metric used to assess the relevance of retrieved results, whereas recall is a metric used to evaluate the extent to which all relevant results are successfully retrieved.

The accuracy-recall curve illustrates the inherent tradeoff between precision and recall across various threshold values. In the conducted experiment, the precision-recall accuracy was determined to be 94.8%. A substantial area under the curve signifies the presence of both elevated recall and precision, with high accuracy being associated with a reduced false positive rate, and high recall being associated with a diminished false negative rate. The attainment of high scores for both metrics indicates that the classifier is producing precise outcomes (high precision) and capturing a significant proportion of all positive outcomes (high recall).

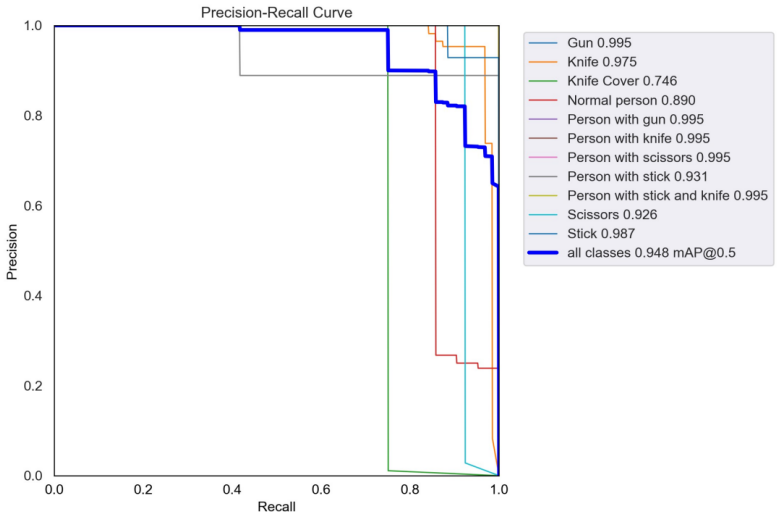

Figure 4-6 Precision-Recall

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

4.4.5 Confusion Matrix

The confusion matrix is a matrix that provides a concise summary of the performance of a machine learning model when evaluated on a specific dataset for testing purposes. The measurement of classification model performance is frequently employed to assess the accuracy of predicting categorical labels for input instances. The matrix presents the quantities of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) generated by the model when applied to the test dataset.

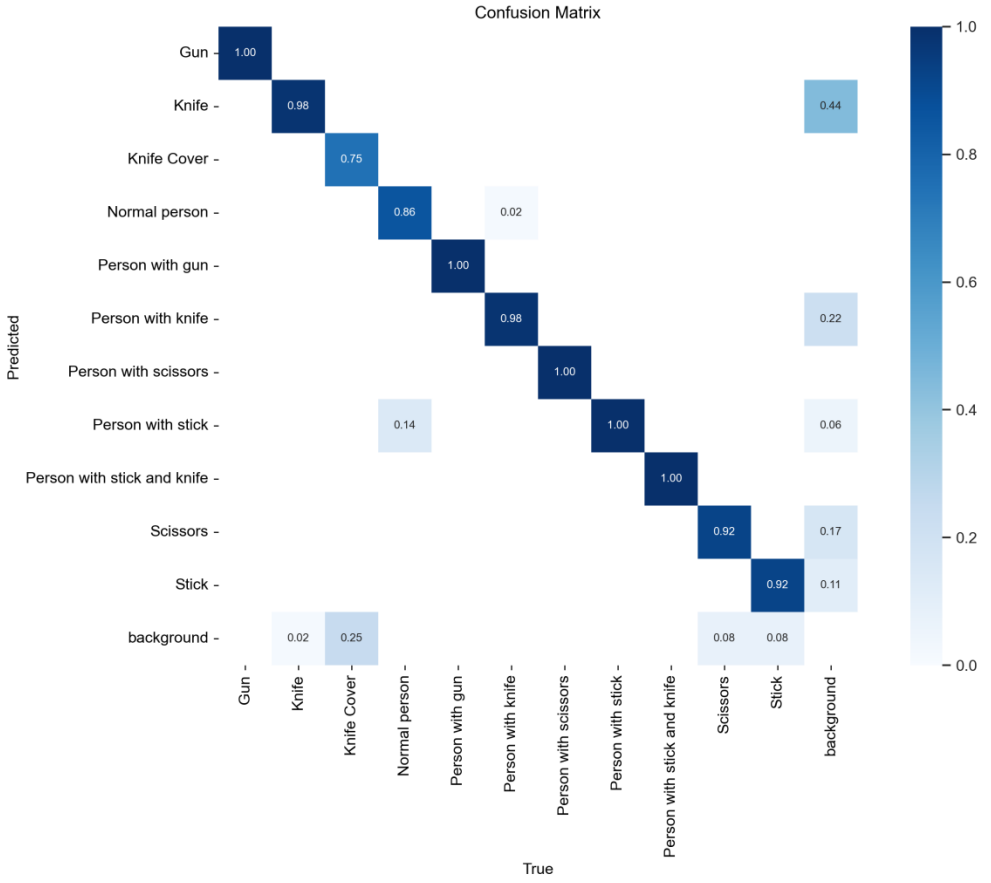

Figure 4-7 Confusion Matrix

Accuracy: The performance of the model is assessed by means of accuracy. The metric being referred to is the ratio of the total number of correct instances to the total number of occurrences.

$$\textbf{Accuracy} = \frac{\textbf{TP+TN}}{\textbf{TP+FP+TN+FN}} \qquad \textbf{...(4.1)}$$

4.4.6 F1 score, Precision, and Recall

The assessment of an object detection model through the utilization of precision and recall metrics can yield significant insights about the model's performance across different confidence thresholds. The F1 score is a valuable metric for finding the optimal confidence threshold that strikes a compromise between precision and recall in a given model. It should be noted that the F1 score ranges from 0 to 1, covering a domain of confidence values. A comprehensive assessment of the overall performance of a given model can be obtained by deriving a singular evaluation metric from the collection of F1 scores.

The combination of these metrics proves effective in assessing a model's performance across different confidence levels, thereby offering useful insights into its performance and identifying optimal values that align with the design specifications. In general, when the confidence threshold is raised, it is observed that the precision increases while the recall decreases. This trend is evident in the outcomes obtained from a bespoke YOLO v5 model, as illustrated below.

$$\textbf{F1} = 2 \cdot \frac{\textbf{precision} \cdot \textbf{recall}}{\textbf{precision+recall}} \qquad \textbf{...(4.2)}$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

4.4.7 Performance of YOLOv5

Table 4-1 YOLOv5 Performance

Model	Precision	Recall	PR Curve	mAP@.50
COCO	87.50	86.80	87.70	88.90
VOC	89.40	83.70	86.10	89.30
Proposed	96.30	96	94.80	94.80

4.4.8 Comparison

Table 4-2 Comparison

Model name	Accuracy
Suthar et al.[35]	55.6
Patrikar et al [35]	78.43
Wen et al [37]	93.52
proposed	94.80

V. CONCLUSION

The YOLOv5 architecture represents a progressive advancement of the YOLO object detection framework, providing a proficient and precise approach to identifying aberrant objects in real-time scenarios. By utilizing several components, such as a deep convolutional backbone, scaled YOLO heads, anchor boxes, and advanced data augmentation techniques, the system is capable of accurately detecting objects that exhibit deviations from the expected patterns within a specific environment. In order to effectively execute real-time abnormal item detection using YOLOv5.The YOLOv5 model has been specifically designed to achieve high efficiency in real-time scenarios, making it well-suited for applications that necessitate rapid detection and timely response to anomalous items. The model possesses the capability to undergo fine-tuning and customization in order to cater to unique use cases and datasets. This enables the model to perform exceptionally well in diverse settings where the definition of abnormality may vary. The presence of a comprehensive and diversified dataset, consisting of properly annotated instances of both normal and abnormal objects, is of utmost importance in the training process of YOLOv5. This dataset plays a crucial role in enabling the model to effectively identify and differentiate deviations from the established norm. Post-processing techniques, such as the utilization of non-maximum suppression (NMS), can be implemented to improve the accuracy of object localization and decrease the occurrence of false positives. The fields of object detection and abnormality detection are always getting better, so it's important to stay up to date on the latest academic studies and progress in order to get the best results. In brief, YOLOv5 demonstrates efficacy as a robust instrument for real-time identification of anomalous objects, presenting a harmonious equilibrium between computational efficiency and precision. The versatility and adaptability of this technology make it a desirable option for activities that require the identification of items or patterns that deviate from the anticipated standard. Nevertheless, in light of technological progress, it is imperative to remain up-to-date with the most recent advancements in this domain in order to maximize the potential of YOLOv5 and its associated models.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

VI. REFERENCES

[1] Dutta, A., Gupta, A., & Kumaraguru, P. (2020). Abnormal event detection in videos using deep learning: A survey. arXiv preprint arXiv:2004.05861.

[2] Hassan, N. U., Akram, T., & Kim, D. (2018). A survey of industrial vision systems, applications and tools. Computers & Industrial Engineering, 125, 362-391.

[3] Li, H., Pang, Y., Song, W., & Song, G. (2019). Application of deep learning methods in medical image processing. Computational and Mathematical Methods in Medicine, 2019.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E，. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, Association for Computing Machinery (ACM), 2017, 60(6): 84–90.

[5] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[6] YU F，. Multi-Scale Context Aggregation by Dilated Convolutions[EB/OL]. arXiv.org, 2015-11-23. (2015-11-23). https://arxiv.org/abs/1511.07122.

[7] CHANDOLA V, BANERJEE A, KUMAR V，. Anomaly detection[J]. ACM Computing Surveys, Association for Computing Machinery (ACM), 2009, 41(3): 1–58.

[8] SABOKROU M, FAYYAZ M, FATHY M, et al.，. Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes[J]. IEEE Transactions on Image Processing, Institute of Electrical and Electronics Engineers (IEEE), 2017, 26(4): 1992–2004.

[9] SORNETTE D，. Critical market crashes[J]. Physics Reports, Elsevier BV, 2003, 378(1): 1–98.

[10] CHAWLA N V, BOWYER K W, HALL L O, et al.，. SMOTE: Synthetic Minority Over-sampling Technique[J]. Journal of Artificial Intelligence Research, AI Access Foundation, 2002, 16: 321–357.

[11] CHURPEK M M, YUEN T C, HUBER M T, et al.，. Predicting Cardiac Arrest on the Wards[J]. Chest, Elsevier BV, 2012, 141(5): 1170–1176.

[12] SARKAR, M. R. (2023). Artificial intelligence in healthcare.

[13] FUKUYAMA F, AXELROD R, JERVIS R，. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration[J]. Foreign Affairs, JSTOR, 1998, 77(2): 142.

[14] TIWARI R K, VERMA G K，. A Computer Vision based Framework for Visual Gun Detection Using Harris Interest Point Detector[J]. Procedia Computer Science, Elsevier BV, 2015, 54: 703–712.

[15] LI S, WU S，. Low-Cost Millimeter Wave Frequency Scanning Based Synthesis Aperture Imaging System for Concealed Weapon Detection[J]. IEEE Transactions on Microwave Theory and Techniques, Institute of Electrical and Electronics Engineers (IEEE), 2022, 70(7): 3688–3699.

[16] COOTES T F, EDWARDS G J, TAYLOR C J，. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, Institute of Electrical and Electronics Engineers (IEEE), 2001, 23(6): 681–685.

[17] GLOWACZ A, KMIEĆ M, DZIECH A，. Visual detection of knives in security applications using Active Appearance Models[J]. Multimedia Tools and Applications, Springer Science and Business Media LLC, 2013, 74(12): 4253–4267.

[18] Sasikaladevi, V.; Mangai, V. Colour Based Image Segmentation Using Hybrid Kmeans with Watershed Segmentation. Int. J. Mech.Eng. Technol. 2018, 9, 1367–1377.

[19] BAI X, WANG X, LIU X, et al.，. Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments[J]. Pattern Recognition, Elsevier BV, 2021, 120: 108102.

[20] KIRANYAZ S, AVCI O, ABDELJABER O, et al.，. 1D convolutional neural networks and applications: A survey[J]. Mechanical Systems and Signal Processing, Elsevier BV, 2021, 151: 107398.

[21] BOROVYKH A, BOHTE S, OOSTERLEE C W. Conditional time series forecasting with convolutional neural networks[C]. International Conference on Artificial Neural Networks, ICANN 2017, 10614 LNCS: 729-730.

[22] Lai, J.; Maples, S. Developing a Real-Time Gun Detection Classifier. Tech. Rep. Available online: http://vision.stanford.edu/teaching/cs231n/reports/2017/pdfs/716.pdf (accessed on 1 March 2022).

[23] Verma, G.K.; Dhillon, A. A handheld gun detection using faster r-cnn deep learning. In Proceedings of the 7th International Conference on Computer and Communication Technology, Allahabad, India, 24–26 November 2017; pp. 84–88.

[24] Kundegorski, M.E.; Akçay, S.; Devereux, M.; Mouton, A.; Breckon, T.P. On using feature descriptors as visual words for object detection within X-ray baggage security screening. In Proceedings of the 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), Madrid, Spain, 23–25 November 2016.

[25] Zhang, J.; Xing, W.; Xing, M.; Sun, G. Terahertz image detection with the improved faster region-based convolutional neural network. Sensors 2018, 18, 2327.

[26] AKCAY S, KUNDEGORSKI M E, WILLCOCKS C G, et al.， . Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery[J]. IEEE Transactions on Information Forensics and Security, Institute of Electrical and Electronics Engineers (IEEE), 2018, 13(9): 2203–2215.

[27] de Azevedo Kanehisa, R.F.; de Almeida Neto, A. Firearm Detection using Convolutional Neural Networks. ICAART 2019, 2,707–714

[28] Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. September. Microsoft coco: Common objects in context. In European Conference on Computer Vision; Springer: Cham, Switzerland, 2014; pp. 740–755.

[29] Bhatti, M.T.; Khan, M.G.; Aslam, M.; Fiaz, M.J. Weapon detection in real-time cctv videos using deep learning. IEEE Access 2021, 9, 34366–34382. Olmos, R.; Tabik, S.; Herrera, F. Automatic Handgun Detection Alarm in Videos Using Deep Learning. Neurocomputing 2018, 275,66–72.

[30] Nakib, M.; Khan, R.T.; Hasan, M.S.; Uddin, J. February. Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network. In Proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–4.

[31] L. Tan, T. Huangfu, L. Wu, et al., “Comparison of YOLO v3, Faster R-CNN, and SSD for Real-Time Pill Identification,” 30, Jul. 2021

[32] GitHub link: https://github.com/ultralytics/yolov5.

[33] GitHub link: https://github.com/pytorch/pytorch.

[34] A. Bochkovskiy, C.-Y. Wang, H. Yuan, and M. Liao, “Yolov4: optimal speed and accuracy of object detection,” 2020, https:// arxiv.org/abs/2004.10934.

[35] Suthar, Anjali. Abnormal Activity Recognition in Private Places Using Deep Learning: A Survey. International Journal for Research in Applied Science and Engineering Technology 11(6) 2023, 2753–61.

[36] Patrikar, Devashree R. and Mayur Rajaram Parate. Anomaly detection using edge computing in video surveillance system: review. International Journal of Multimedia Information Retrieval 11(2) 2022, 85–110.

[37] Wen, JianFeng, YiHai Qin, and Shan Hu. Abnormal behavior identification of examinees based on improved YOLOv5. in Ruishi Liang and Jing Wang (eds). International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022) 2023.