Weapon Detection And Classification System For Public Safety

A Project Work Submitted in Partial Fulfillment

of the requirements for the Degree of 8th Semester

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

by

MD Dildar Mandal (Roll No. 202102021043)
Gulnaaz Parveen (Roll No. 202102021055)
Swmdwn Choudhury (Roll No. 202102021011)

Under the supervision of

Dr. Anup Kumar Barman

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

केन्द्रीय प्रौद्योगिकी संस्थान कोकराझार

CENTRAL INSTITUTE OF TECHNOLOGY KOKRAJHAR

(A Centrally Funded Institute under Ministry of HRD, Govt. of India)

BODOLAND TERRITORIAL AREAS DISTRICTS :: KOKRAJHAR :: ASSAM :: 783370

Website: www.cit.kokrajhar. in, www.cit.ac.in

May 2025

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

केन्द्रीय प्रौद्योगिकी संस्थान कोकराझार

CENTRAL INSTITUTE OF TECHNOLOGY, KOKRAJHAR
(An Autonomous Institute under MHRD)
Kokrajhar – 783370, BTAD, Assam, India

CERTIFICATE OF APPROVAL

This is to certify that the work embodied in this project entitled **Weapon Detection and Classification System For Public Safety** submitted by **MD. Dildar Mandal** (202102021043), **Gulnaaz Parveen** (202102021055), **Swmdwn Choudhury** (202102021011) to the Department of Computer Science & Engineering, is carried out under our direct supervisions and guidance.

The project work has been prepared as per the regulations of Central Institute of Technology and I strongly recommend that this project work be accepted in partial fulfillment of the requirement for the degree of B.Tech.

Supervisor
Dr. Anup Kumar Barman
Assistant Professor, Dept. of CSE

Countersigned by

Dr. Apurbalal Senapati
HoD
Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

केन्द्रीय प्रौद्योगिकी संस्थान कोकराझार

CENTRAL INSTITUTE OF TECHNOLOGY, KOKRAJHAR

(An Autonomous Institute under MHRD)

Kokrajhar – 783370, BTAD, Assam, India

Certificate by the Board of Examiners

This is to certify that the project work entitled **Weapon Detection and Classification System For Public Safety** submitted by **MD. Dildar Mandal** (202102021043), **Gulnaaz Parveen** (202102021055), **Swmdwn Choudhury** (202102021011) to the Department of Computer Science and Engineering of Central Institute of Technology, Kokrajhar has been examined and evaluated.

The project work has been prepared as per the regulations of Central Institute of Technology and qualifies to be accepted in partial fulfillment of the requirement for the degree of B. Tech.

Project Co-ordinator **Board of Examiners**

____________________ ____________________

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to my supervisor, Dr. Anup Kumar Barman, for his invaluable guidance, support, and expertise throughout the duration of this project. His insightful feedback and continuous encouragement were instrumental in shaping this work.

We are also thankful to the Head of the Department of Computer Science and Engineering, Professor Dr. Apurbalal Senapati for providing a conducive environment for academic and research activities within the department. The department's support has been invaluable in facilitating an enriching research experience.

Additionally, we extend our appreciation to all the teaching and non-teaching staff members who contributed to the success of this project through their assistance and cooperation. This also includes our friends and family, thankful to their unwavering encouragement, understanding and support throughout this academic pursuit. Their belief in the importance of this project has been a constant source of motivation.

Date:

MD. Dildar Mandal
202102021043

Gulnaaz Parveen
202102021055

Swmdwn Choudhury
202102021011

ABSTRACT

Given the rising rates of crime, including armed attacks and terrorism, there is a need for the advancement of intelligent and efficient computerized systems that can identify weapons. Conventional surveillance systems are mainly operated and monitored by human beings, which limits the efficiency of the system, is prone to errors and delays in response. In order to overcome these limitations, this project aims at designing a smart weapon detection and classification system using our proposed model which is an enhanced customized VGG-net architecture. The system has been developed in a way that enables the identification of different types of weapons such as knife, pistol, and long-gun. Using the deep learning approach, the proposed model provides enhanced feature extraction and accuracy. The model is trained using the Keras framework which is based on TensorFlow and the dataset that includes diverse weapons, backgrounds and environmental conditions. The data set is annotated to ensure that the neural network is trained adequately and to enhance the network's capability of handling a wide range of data. The model was also subjected to training and testing processes and it provided accurate results of 92.16 % thus improving on the conventional VGG-16 of 89.60 %, VGG-19 of 89.38 %, ResNet50 of 79.91 % and ResNet101 of 78.48 % accuracy. This performance shows the effectiveness of the changes that have been made to the VGG-16 architecture.

This solution does not require the involvement of humans therefore making the threat identification process faster and more accurate. The feature that makes the system effective in the monitoring of risky areas including the public places, airports, schools, colleges and other places of possible security threats is the real time analysis of the surveillance data. This project also explores the strengths and weaknesses of other deep learning techniques and found the clear advantages of using our proposed model in terms of speed, scalability, and robustness. By providing detailed insights into the performance of the model and dataset, the project establishes a solid foundation for future research and development in this domain.

CONTENTS

Chapter 1

INTRODUCTION

In the modern world of technology advancements, surveillance camera devices are crucial to preventing crimes and ensuring public safety. Commonly found in various places such as cities, transportation hubs, schools and other important places, these systems provide continuous monitoring to deter and detect illicit activities. However, most traditional systems depend on human responsibility to watch live footage or review recording after an incident. This method not only takes a lot of time due to controlling, managing and monitoring systems but also leads to human error resulting in delayed or inadequate response.

The increasing cases of criminal activities, particularly those involving portable firearms and other dangerous weapons, underscores the critical need for advanced detection systems. Handheld weapons, such as handguns, knives, and assault rifles, play a central role in various criminal activities, including armed robberies, illegal hunting, and acts of terrorism. To address these pressing challenges, this project proposes an enhanced weapon detection and classification system leveraging the power of deep learning techniques. Deep learning, a branch of machine learning, has gained widespread recognition for its ability to analyze and extract complex patterns in large amounts of datasets. The modified architecture's superior ability to extract meaningful features and its robustness against challenging scenarios validate its suitability for weapon detection and classification tasks.

Moreover, the proposed system is designed for real-time operation, making it an invaluable tool for surveillance applications. By processing live footage from surveillance cameras, the system can autonomously detect and classify the presence of weapons and notify security personnel instantly, enabling them to take timely action to mitigate threats. This capability is particularly crucial in high-risk environments such as public spaces, airports, educational institutions, and government facilities, where rapid responses can prevent significant harm. This project also explores the broader technological landscape of weapon detection and classification. A detailed analysis of traditional machine learning methods highlights their limitations in terms of scalability, adaptability, and accuracy, particularly in complex scenarios. In contrast, deep learning techniques, with their ability to learn intricate feature hierarchies and handle large-scale data, provide a more robust and efficient solution. By addressing critical gaps in the field and presenting a state-of-the-art model, this project establishes a strong foundation for further advancements in intelligent surveillance systems.

1.1 Objectives

The main goal of this project is to develop a robust weapon detection and classification system capable of identifying various weapons such as knives, pistols and long guns with high precision and reliability. The proposed system leverages a custom VGG-net based deep learning model, specifically designed to overcome the limitations of traditional methods and address the complexities associated with real-world weapon detection scenarios.

A critical goal of the project is to enable real-time surveillance capabilities. This involves designing the system to process live CCTV footage efficiently, allowing for immediate detection of potential threats and facilitating timely responses. Such functionality is essential for high-risk environments like public spaces, transportation hubs, and educational institutions, where the rapid identification of weapons can play a pivotal role in preventing security breaches and mitigating risks.

To ensure the system's effectiveness in practical applications, the project addresses several real-world challenges, including occlusion, object resemblance, and background complexities. The model is optimized to recognize weapons, distinguish between weapons and similar-looking objects, and perform reliably in dynamic and cluttered environments. This focus on adaptability and robustness is vital for deploying the system in diverse settings with varying lighting conditions, perspectives, and crowd densities. Ultimately, the overarching aim of this project is to enhance public safety by providing a scalable and automated tool for weapon detection and classification. By reducing reliance on human monitoring and intervention, the proposed system empowers security personnel to respond proactively to threats, thereby preventing potential crimes and safeguarding lives.

1.2 Motivation

The motivation for this research stems from the alarming rise in armed crimes, terrorist activities, and mass shootings in places like schools and restaurants which pose a significant threat to public safety. Traditional surveillance systems, which rely heavily on manual monitoring, have proven to be slow, inefficient, and highly prone to human error, often resulting in delayed or insufficient responses to critical situations. These limitations underscore the urgent need for automated and intelligent solutions capable of addressing these challenges. Current weapon detection technologies, while helpful, face significant hurdles in real-world scenarios, including issues like occlusion, where parts of a weapon are hidden, object resemblance, where non-weapon objects mimic the appearance of weapons, and background complexities, which make detection in dynamic environments particularly challenging.

Deep learning technologies, particularly Convolutional Neural Networks (CNNs), have demonstrated immense potential in overcoming these challenges by providing real-time detection with remarkable accuracy. CNNs excel in extracting intricate features and recognizing patterns, making them particularly effective for complex tasks such as weapon detection. However, there is a noticeable lack of robust and scalable systems capable of identifying a diverse range of weapon types in varied and dynamic settings. This gap in current technology motivates the development of an advanced, automated weapon detection system that leverages the power of deep learning to enhance the effectiveness of surveillance systems and significantly improve public safety measures.

Chapter 2

LITERATURE SURVEY

The increasing need for enhanced security and public safety has led to significant advancements in weapon detection and classification systems. Modern approaches leverage deep learning techniques, particularly Convolutional Neural Networks (CNNs), to identify and mitigate threats in real-time. The growing reliance on automated surveillance systems has shifted focus from traditional detection methods, such as manual monitoring or metal detectors, to more sophisticated solutions that can provide high accuracy and efficiency. Among the various architectures, CNNs such as VGG-16, YOLO, and Faster R-CNN have gained prominence for their ability to detect objects in diverse environments. In addition, other powerful architectures like ResNet, SSD (Single Shot MultiBox Detector), RetinaNet, and EfficientNet are increasingly being adopted for their speed and precision in real-time detection tasks. Lightweight models such as MobileNet and SqueezeNet are also proving valuable in resource constrained environments, while U-Net and DenseNet offer high accuracy for specialized detection and segmentation scenarios.

2.1 Background Details

Weapon detection is an essential component of modern surveillance systems, particularly in high-security areas such as airports, schools, government institutions, and public transportation hubs. Traditional detection methods, including manual inspection, metal detectors, and basic motion detection, often suffer from limitations such as low detection speed, limited scalability, and susceptibility to human error. These shortcomings necessitate the integration of automated and intelligent systems capable of operating in real-time and adapting to dynamic environments.

Deep learning-based object detection methods have become a cornerstone of modern weapon detection systems. Convolutional Neural Networks (CNNs) are particularly effective for this purpose due to their ability to automatically extract hierarchical features from input images. Networks such as VGG-16, ResNet, YOLO (You Only Look Once), and Faster R-CNN have demonstrated exceptional performance in object detection tasks, including weapon detection. Despite these advancements, challenges remain in real-world scenarios. For instance, occlusion, poor lighting conditions, and the presence of look-alike objects such as mobile phones and flashlights often lead to high false positive and false negative rates. Additionally, the computational demands of sophisticated models can hinder their applicability in resource-constrained environments, such as embedded systems or real-time surveillance networks. Addressing these challenges requires innovative approaches that balance detection accuracy with computational efficiency.

2.2 Related Work

Murugaiyan Sivakumar, Gatta Venkata Amruth, and Kiranmai Bellam[1] introduced an enhanced weapon detection system that leverages advanced deep learning techniques to improve the accuracy and reliability of automated surveillance systems. The study addresses critical challenges in real-time weapon detection, such as false positives, varying environmental conditions, and the need for rapid processing. The authors implemented a deep convolutional neural network (CNN) architecture as the backbone for their detection system, specifically customizing it to enhance feature extraction from input images and videos. This architecture incorporates layers optimized for detecting fine-grained details, making it capable of identifying both concealed and overt weapons in cluttered and dynamic environments. The system employs pre-processing methods such as noise reduction and data augmentation to enhance detection in low-light or occluded scenarios. The architecture integrates with a region proposal network (RPN) to focus on areas of interest, reducing computational overhead while maintaining high detection precision. The authors validated the system's performance using benchmark datasets and real-world scenarios, demonstrating superior accuracy compared to traditional object detection frameworks.

Pavinder Yadav, Nidhi Gupta, and Pawan Kumar Sharma[2] conducted a comprehensive study exploring both classical machine learning and modern deep learning approaches for weapon detection. Their work provides an extensive comparison between traditional algorithms, such as Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN), and advanced neural networks like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The authors critically evaluated the strengths and limitations of each method, particularly in terms of feature extraction, processing speed, and detection accuracy under varying environmental conditions. The study emphasized the importance of feature engineering in classical methods, highlighting how handcrafted features such as shape and texture descriptors can still be effective for simpler scenarios. However, they also demonstrated the superior performance of deep learning models in handling complex and dynamic environments. In particular, the use of CNNs for feature extraction and classification proved highly effective, given their ability to learn hierarchical features directly from raw data without manual intervention. The authors also explored the integration of hybrid methods, combining classical algorithms with deep learning for improved detection rates in specific use cases. This research provides valuable insights into the evolution of weapon detection methodologies, serving as a foundational reference for future work aimed at optimizing accuracy and efficiency.

Early Advancements in Weapon Detection was contributed by Olmos et al.[3] , they were among the first to utilize the Faster R-CNN framework for handgun detection, achieving an accuracy of 84.21%. While effective, their approach was computationally intensive due to the separate region proposal mechanisms required by Faster R-CNN. Fernandez-Carrobles et al.[4] faced similar challenges with inference speed, which limited the application of their Faster R-CNN-based model in real-time scenarios. Verma and Dhillon[5] addressed this issue by applying transfer learning techniques, achieving 93% accuracy. However, their reliance on high computational resources restricted scalability in practical applications.

Several researchers have explored the optimization of existing architectures to improve performance and reduce computational overhead. Gelana and Yadav[6] employed a sliding window approach with CNN classifiers, achieving 93.84% accuracy. However, their model was highly dependent on high-quality input images and struggled with objects of varying dimensions. Dwivedi et al.[7] implemented the

VGG-16 network, achieving 99% accuracy for pistol detection. Although their results were impressive, the model's computational demands limited its real-time usability.

Efforts to address these limitations include the adoption of YOLO-based models, which process images in a single pass, significantly enhancing detection speed without compromising accuracy. Singh et al.[8] trained YOLOv4 on a dataset of diverse weapons, achieving 77.75% mean average precision. While promising, their model struggled with smaller objects and overlapping weapons in crowded scenes.

Small and occluded objects, such as pistols or knives, pose significant challenges in surveillance scenarios. To address these issues, researchers have adopted tiling techniques, where large images are divided into smaller tiles for processing. Huang et al.[9] and Ozge Unel et al.[10] demonstrated that overlapping tiles improved the recognition of smaller objects. However, tiling introduces additional computational overhead during inference, particularly when merging and refining predictions. Khalid et al.[11] addressed occlusion and rotation issues using data augmentation, achieving 95.43% accuracy with the YOLOv5 model.

Recent advancements have explored pose estimation to enhance weapon detection. Lamas et al.[12] combined EfficientDet with pose estimation techniques to detect human-held weapons, achieving robust performance across diverse scenarios. By analyzing the interaction between humans and objects, their model effectively identified potential threats, even in crowded environments.

Bhatti et al.[13] performed a comprehensive evaluation of algorithms, including Faster R-CNN, YOLOv3, and SSD, on a dataset of over 8,000 images. Their results highlighted YOLOv4's superiority in terms of accuracy and F1 score but noted high false positive and false negative rates. Salido et al.[14] further compared three CNN models and found that RetinaNet with an unfrozen ResNet-50 backbone achieved the best balance of precision and recall.

Weapon detection systems are increasingly integrated into CCTV networks for applications such as pedestrian safety, traffic monitoring, and crime prevention. Sankaranarayanan et al.[15] demonstrated the use of Gaussian blends with Bayesian Kalman filters for object tracking, enabling real-time monitoring of unlicensed weapons. Similarly, Jain et al.[16] achieved high accuracy with Faster R-CNN and SSD models but highlighted the need for better false alarm management in real-time applications.

Pravallika and Reddy [17] proposed a deep learning-based Weapon Detection and Alerting System that utilizes computer vision techniques to detect firearms, knives, explosives, and other dangerous objects in real-time. Their system comprises several key modules including video/image acquisition, preprocessing, feature extraction using CNNs, and classification. The model is trained on a large labeled dataset using supervised or transfer learning methods and is capable of deploying in live surveillance scenarios. The proposed approach significantly enhances detection accuracy and reduces reliance on manual inspection, making it highly applicable in high-risk areas such as airports and public gatherings.

Jain et al. [18] and Swetha et al. [19] explored AI-based systems for weapon detection using video surveillance. Jain et al. focused on the integration of artificial intelligence with deep learning techniques to identify weapons from CCTV footage, emphasizing robustness across varying lighting and environmental conditions. Swetha et al. extended this idea with a video analysis-based alerting system, aiming to automatically generate real-time notifications upon detecting potential threats. Both studies underscore the practical significance of combining image processing, deep learning, and real-time alerting to enhance situational awareness and responsiveness.

Berardini et al. [21] presented a deep-learning framework optimized for edge devices to detect handguns and knives in indoor surveillance environments. Their system addresses real-time processing challenges by deploying lightweight, efficient models on constrained hardware, such as embedded GPUs or edge computing boards. The framework demonstrates that accurate weapon detection is feasible even in resource-limited setups, making it particularly suitable for scalable and cost-effective surveillance installations in sensitive indoor spaces.

A.H. Ashraf et al.[22] proposed a hybrid approach for weapon detection in surveillance systems by integrating Convolutional Neural Networks (CNN) with the YOLOv5s architecture. The research, published in Computer Modeling in Engineering & Sciences, highlights the effectiveness of combining traditional deep learning techniques with advanced object detection frameworks for real-time security applications. The authors focused on improving detection accuracy and speed, addressing challenges such as low resolution and occlusion in video surveillance footage. Their results demonstrated that the CNN-YOLOv5s model achieved high performance in identifying weapons, making it a valuable contribution to the development of intelligent surveillance systems.

Gijare[23] presented an automatic and accurate weapon detection model utilizing an optimal neural network architecture, as detailed in their publication on SSRN. The study focuses on enhancing the precision and reliability of weapon detection systems by designing a neural network tailored for this specific task. The proposed model addresses common issues such as false positives and detection latency by optimizing the architecture for both accuracy and computational efficiency. Experimental results demonstrated that the model performs effectively in diverse scenarios, making it a promising solution for real-time security and surveillance applications.

Razzaq et al.[24] proposed a comprehensive deep learning-based framework aimed at enhancing public safety by detecting weapons and violent behavior in CCTV surveillance videos. Presented at the 25th International Multitopic Conference (INMIC), their work integrates advanced computer vision techniques with deep neural networks to identify potential threats in real-time. The system is designed to detect not only the presence of weapons such as guns and knives but also aggressive human actions indicative of violence. By combining these detection modules, the framework offers a holistic approach to threat assessment in public and crowded environments. The authors employed a carefully selected set of pretrained deep learning models and fine-tuned them for the specific context of CCTV footage, which often includes challenges such as low resolution, occlusions, and dynamic lighting. Experimental results showed that the proposed system achieved high detection accuracy and low latency, making it suitable for practical deployment in real-time surveillance systems. This work contributes significantly to the growing field of AI-powered video analytics for public safety and threat prevention.

2.3 Summary of Literature Review

The field of weapon detection in surveillance has evolved significantly with the integration of deep learning techniques, aiming to enhance accuracy, real-time responsiveness, and adaptability in complex environments. Early methods relied on traditional machine learning algorithms such as SVM, KNN, and Decision Trees, which required handcrafted features and were often limited by their inability to handle variations in scale, lighting, and occlusion. With the advent of CNNs and other neural network models, researchers shifted towards architectures capable of learning hierarchical features from raw image data, thus eliminating the need for manual feature engineering. Approaches such as Faster R-CNN and VGG-16 brought significant improvements in accuracy, especially in controlled environments. However, these models were computationally intensive and unsuitable for real-time surveillance, prompting efforts to optimize detection pipelines. The use of region proposal networks, transfer learning, and sliding window CNNs helped in refining detection but still posed scalability challenges. Later, YOLO-based models emerged as a practical solution by enabling single-shot detection, striking a better balance between speed and accuracy, although they initially struggled with smaller or overlapping weapons.

To further address real-world constraints, researchers implemented advanced techniques such as data augmentation, tiling, and pose estimation. These methods helped mitigate issues related to occlusion, object scale, and dynamic lighting, especially in crowded or low-visibility environments like public gatherings and indoor spaces. Studies began to explore the deployment of lightweight models on edge devices, demonstrating that even resource-constrained systems could support real-time and accurate detection with proper architectural tuning. Hybrid models that combined CNNs with YOLO variants improved performance on low-resolution and noisy surveillance footage. Several frameworks incorporated behavioral analysis, detecting not just weapons but also aggressive actions, enhancing situational awareness and threat assessment. The literature highlights a shift from static image-based detection to real-time video analytics, with an increasing focus on practical deployment across public safety infrastructure such as CCTV networks, transportation hubs, and smart cities. Collectively, these advancements illustrate the growing maturity of AI-powered surveillance, making it feasible to build scalable, reliable, and intelligent systems that proactively assist in crime prevention and public security.

Chapter 3

PROPOSED SYSTEM

3.1 System Architecture

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed to process and analyze visual data such as images and video frames. Unlike traditional neural networks, which treat all input features equally, CNNs leverage the spatial structure of image data by applying convolution operations using small learnable filters that scan the image in localized regions to detect patterns like edges, textures, and shapes. CNNs extract hierarchical features through multiple layers, where the lower layers focus on basic features such as lines, edges, and colors, the middle layers identify combinations of these features like corners or contours, and the deeper layers capture more complex and abstract patterns, including object parts or even entire objects.

A standard CNN architecture consists of the following components:

- **Convolutional Layers:** Apply filters to the input image to extract local features.

- **Activation Functions:** Introduce non-linearity, enabling the network to learn complex representations (e.g., ReLU).

- **Pooling Layers :** Reduce the spatial dimensions of the feature maps, improving computational efficiency and reducing overfitting (e.g., Max Pooling).

- **Flattening Layer:** Converts the 2D feature maps into a 1D vector before feeding into fully connected layers.

- **Fully Connected Layers or Dense Layers:** Perform high-level reasoning for classification based on the extracted features.

- **Output Layer:** Produces the final class predictions using activation functions like softmax or sigmoid.

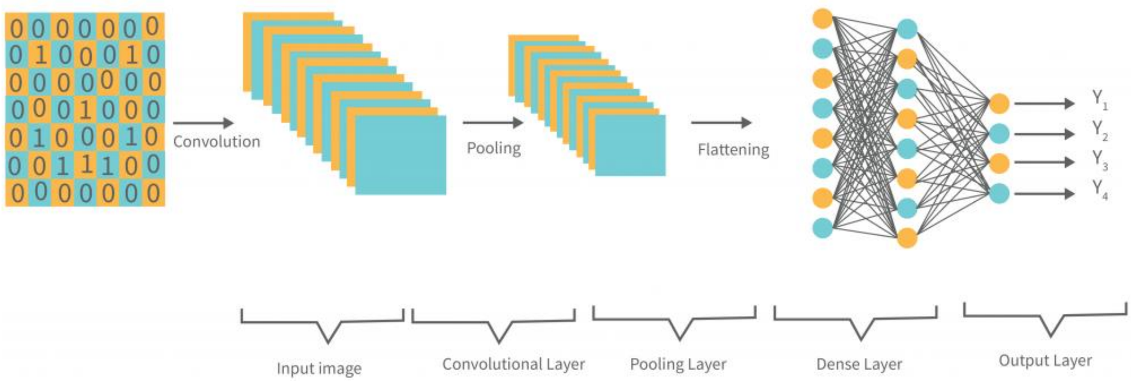

Figure 3.1: CNN architecture

The proposed system leverages a modified VGG-16 Convolutional Neural Network (CNN) architecture with 25 layers, specifically designed for weapon detection and classification in real-time surveillance scenarios. Before finalizing the 25-layer model, an intermediate 21-layer variant was developed and evaluated to understand the impact of increasing depth on detection performance. The standard VGG-16 network is widely known for its simplicity and effectiveness, consisting of 13 convolutional layers followed by 3 fully connected layers, making a total of 16 layers. It uses small 3×3 convolution filters and applies max pooling for spatial dimensionality reduction. However, while effective for general classification tasks, the standard VGG-16 architecture struggles with specific challenges in weapon detection, such as recognizing small objects, handling occlusion, ensuring real-time processing, and classification. To overcome these limitations, the proposed system introduces modifications to the VGG-net architecture. Additional convolutional layers are added to increase the network's depth, enhancing its ability to extract fine-grained features. Batch normalization layers are integrated to stabilize training and improve convergence speed. Leaky ReLU activation functions are used instead of standard ReLU to address the dying neuron problem, which is critical for detecting subtle features in complex scenarios. The modified architecture also incorporates dropout layers to mitigate overfitting and optimize generalization on unseen data.

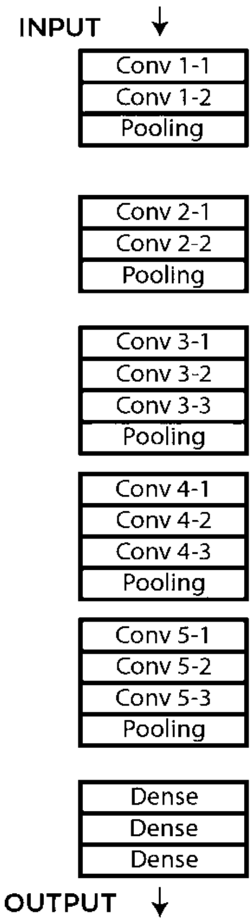

Figure 3.2: VGG-16 Model

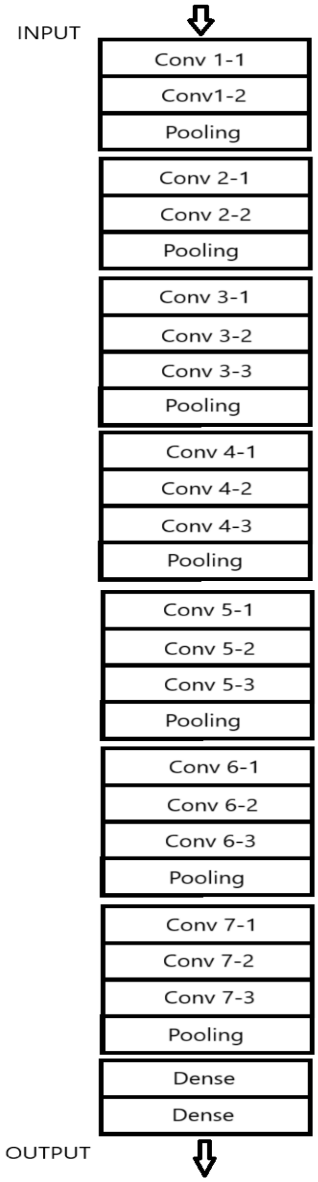

Figure 3.3:VGG-Net Model (21 layers)

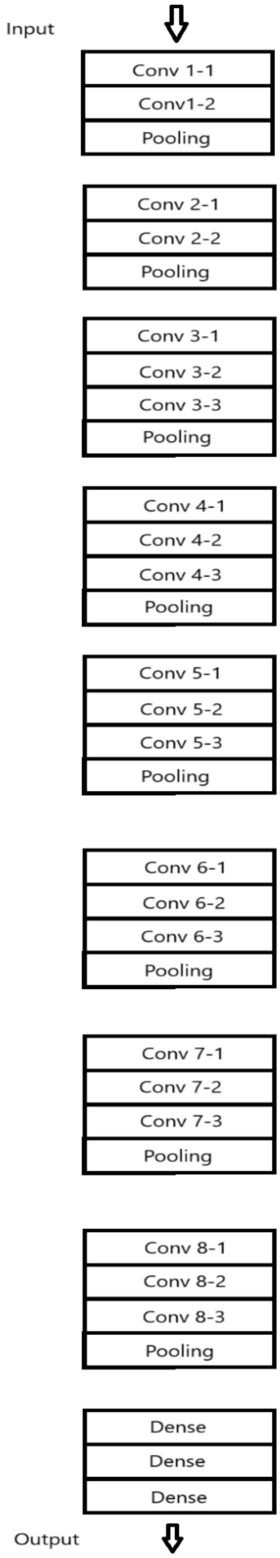

Figure 3.4: VGG-Net Model(25 layers)

The system processes input images captured from surveillance cameras, applies preprocessing techniques such as resizing and normalization, and passes them through the proposed VGG-net model for feature extraction and classification. Detected objects are annotated with bounding boxes and classified as one of the weapon types which includes pistol, knife, and long gun, with the results displayed in real-time.

3.2 Hardware Components

The proposed system utilizes CPU power essential for training and deploying neural network models. The key hardware components of the computer are as follows:

- Processor: AMD Ryzen 5 5600H (12) @ 4.25GHz
- Memory: 8 GiB
- Storage: 500 GiB SSD

3.3 Software Tools

The development and implementation of the proposed system involve the utilization of state-of-the-art software tools and frameworks.

- **Programming Language**: The entire system is primarily built using Python, a high-level programming language known for its versatility and rich ecosystem supporting deep learning and neural network development. Python was used to implement the proposed VGG-net architecture, handle data preprocessing, model training, evaluation, and backend server logic via Flask. Additionally, HTML, CSS, and JavaScript were used to design and develop the front-end interface of the web application. HTML and CSS provided the structure and styling, while JavaScript enabled real-time webcam access and dynamic interaction between the user interface and backend system, facilitating real-time weapon detection through a browser-based deployment.

- **Frameworks and Libraries**: The system employs a range of state-of-the-art libraries and frameworks that are widely adopted in the machine learning and deep learning domains:

    - **TensorFlow**: As the backbone of the system, TensorFlow provides a powerful platform for numerical computation and deep learning. It is instrumental in training the modified VGG-16 network, enabling efficient handling of large-scale computations and GPU acceleration for faster training. TensorFlow's flexibility also supports the implementation of custom layers and loss functions tailored to the needs of the modified architecture.

    - **Keras**: Serving as a high-level API running on TensorFlow, Keras simplifies the design, development, and experimentation of deep learning models. Its user-friendly interface facilitates rapid prototyping, while its modular nature enables easy integration of various layers and activation functions specific to the modified VGG-16 architecture.

    - **Pandas**: This data manipulation library is utilized extensively for preprocessing the input data. It simplifies tasks such as data cleaning, transformation, and organization, ensuring the dataset is well-structured and ready for model training. The ability to handle large datasets efficiently makes Pandas an indispensable tool for this project.

- **NumPy**: Essential for numerical computations, NumPy is employed to perform operations on multi-dimensional arrays and matrices. Its speed and efficiency make it ideal for handling the numerical data required for training and evaluating the neural network model.

- **Scikit-learn**: This machine learning library is used for preprocessing tasks such as feature scaling and normalization. Additionally, scikit-learn provides tools for evaluating the model's performance, such as accuracy metrics and confusion matrices, which are crucial for understanding the system’s effectiveness.

- **Matplotlib**: Visualization plays a key role in monitoring the progress and performance of the model. Matplotlib is utilized to create clear and detailed plots of accuracy, loss, and other metrics during training and testing phases. These visualizations are valuable for identifying trends and diagnosing potential issues in the model’s behavior.

- **Flask:** A lightweight Python web framework used to develop the backend of the weapon detection and classification system. It enabled integration of the deep learning model with a web interface

- **Integrated Development Environment (IDE)**: Jupyter Notebook is utilized for interactive coding, data exploration, and experiment documentation, allowing for a seamless workflow in developing and testing machine learning models.

3.4 Dataset

For this project we have considered the dataset containing images of various types of Pistols, Knives, and Long Guns with dimension 416 x 416 pixels. This dataset is collected from Roboflow Universe, which contains a high collection of open-source computer vision datasets. In our dataset each image is labeled as pistols, knife or long gun along with annotations which is documented in a csv file.

Figure 3.5: Sample Data (a) knives (b) Pistols (c) Long Gun

Chapter 4

IMPLEMENTATION AND RESULTS

4.1 Implementation Details

For the execution of our project, we conducted experiments with the standard VGG-16 and VGG-19 models, followed by testing with a 21-layer custom vgg-net architecture. In addition, we evaluated the performance of other well-known CNN architectures including ResNet-50, ResNet-101, and EfficientNet-B0 to benchmark detection capabilities across different model families. These preliminary experiments helped assess the strengths and limitations of each architecture in terms of feature extraction, generalization, and accuracy for weapon detection tasks. We trained the models on a multi-class dataset comprising different weapon categories, including knives, pistols, and long guns.

After comparative analysis, we developed and trained our proposed 25-layer CNN model to improve detection accuracy and localization. Finally, a threshold-based prediction mechanism was incorporated to determine the presence and type of weapon. If the prediction confidence fell below the predefined threshold, the system labeled the input as “no weapon.” This methodology ensured an effective balance between computational efficiency and detection performance.

4.1.1 Data Pre-Processing

The success of a deep learning model heavily depends on the quality of the input data. For this project, comprehensive data pre-processing steps were undertaken to ensure the model receives clean and representative inputs.

- **Data Cleaning:** Unnecessary and irrelevant images were removed to maintain the integrity of the dataset. Any mislabeled or noisy data was corrected to reduce errors during training.

- **Data Augmentation:** To address the problem of dataset imbalance and improve model robustness, augmentation techniques such as random rotation, scaling, flipping, brightness adjustment, contrast adjustment, saturation adjustment, hue adjustment and cropping were applied. This ensured that the model could generalize effectively to various real-world conditions.

- **Normalization:** Input images were of a resolution 416 x 416 pixels, and pixel values were normalized to a range of 0 to 1.

- **Splitting the Dataset:** The dataset containing 10,042 annotated images was divided into training, validation, and test sets to monitor the model's performance during and after training. The typical split was 70% for training, 20% for validation, and 10% for testing.

4.1.2 Model Training

Model training was conducted in multiple phases, beginning with benchmark architectures and advancing to a custom 25-layer vgg-net tailored for weapon detection. Initially, standard deep learning models such as VGG-16, VGG-19, ResNet-50, ResNet-101, and EfficientNet-B0 were evaluated to establish baseline performance. Additionally, a custom 21-layer convolutional neural network was implemented and tested to explore architectural flexibility and depth optimization.

The VGG-16 architecture was first trained using transfer learning, leveraging pre-trained weights from ImageNet. The early convolutional layers were frozen to preserve learned low-level features, while the final layers and newly added blocks were fine-tuned for our weapon classification task. The model was designed to optimize both classification and bounding box regression, using binary cross-entropy loss for class prediction and Huber loss for bounding box localization. Performance was evaluated using classification accuracy, precision, recall, and F1-score.

Following the baseline experiments, a modified 25-layer VGG-based model was developed to improve feature extraction depth and spatial resolution. This custom architecture incorporated additional convolutional and fully connected layers, batch normalization, and Leaky ReLU activation functions to enhance convergence and mitigate vanishing gradient issues.

Key aspects of the training process included:

- **Transfer Learning & Fine-tuning:** All models were initialized with ImageNet pre-trained weights. Only the top layers were retrained to adapt to the weapon detection domain.

- **Optimizer:** AdamW optimizer was employed with a learning rate of 0.00001 and a weight decay factor of 0.0001 for stable convergence.

- **Training Parameters:** All models were trained for up to 100 epochs with early stopping enabled. Training typically converged within 60 epochs as validation accuracy plateaued.

- **Batch Size & Regularization:** A batch size of 16 was selected for efficient CPU utilization. Dropout layers (rate = 0.5) were added in fully connected blocks to prevent overfitting.

- **Callbacks Used:**
    - EarlyStopping: Monitored validation loss and halted training once improvements ceased.
    - ReduceLROnPlateau: Dynamically reduced the learning rate upon stagnation of validation metrics.
    - ModelCheckpoint: Saved the best-performing model based on validation accuracy.

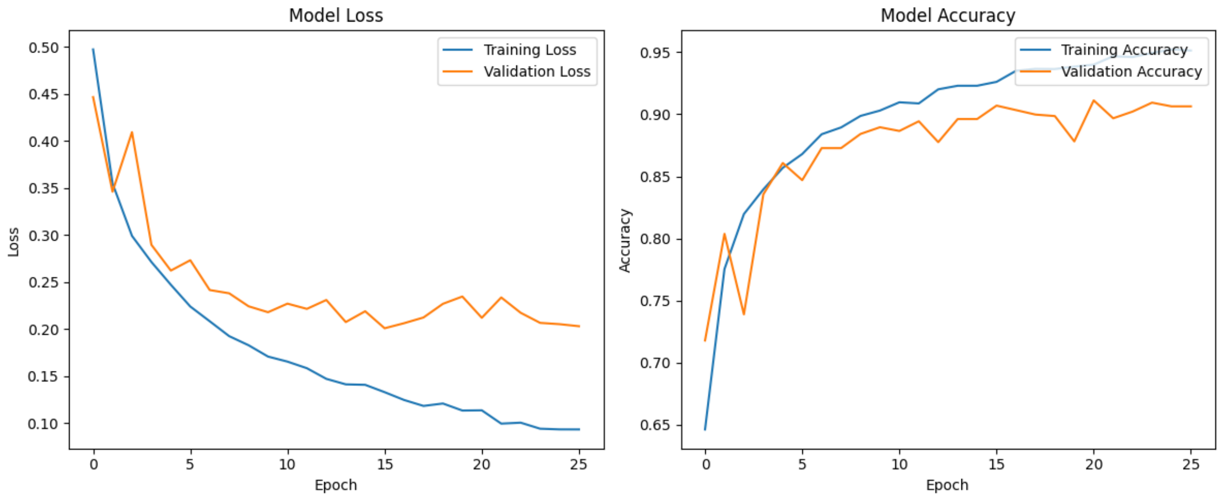

Figure 4.1: Variations observed during training for standard VGG 16

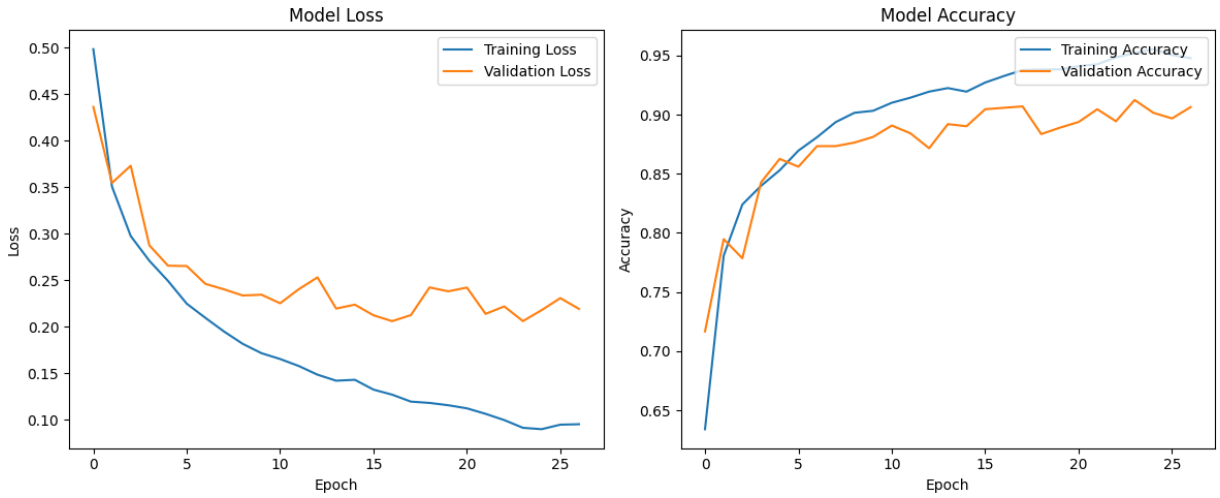

Figure 4.2: Variations observed during training for standard VGG 19

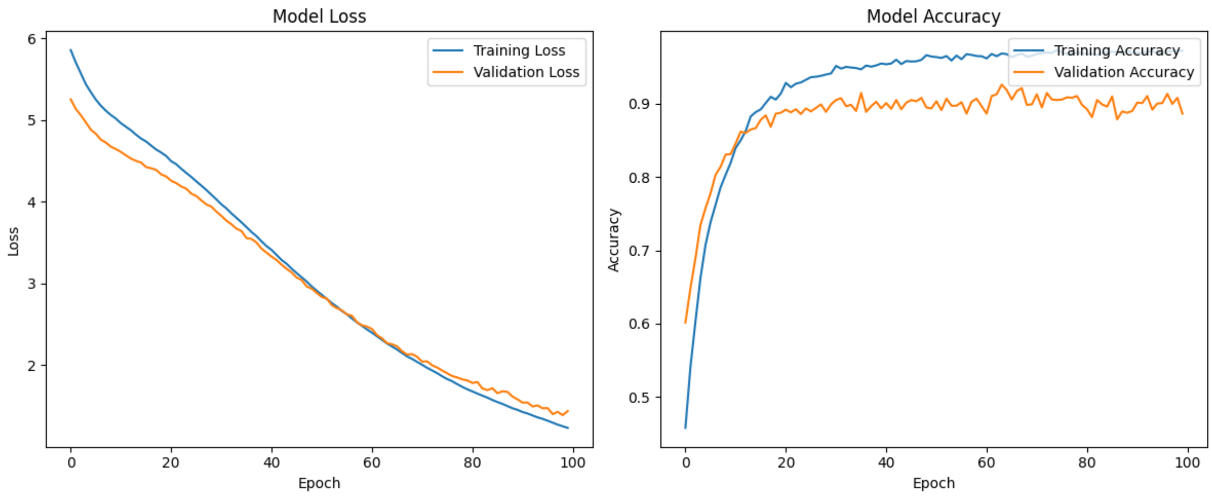

Figure 4.3: Variations observed during training for VGG-Net with 25 layers

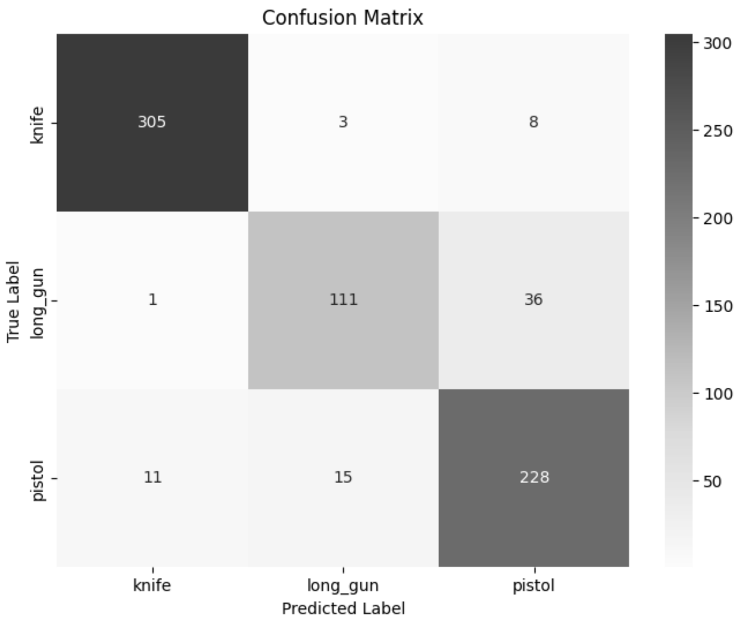

Figure 4.4: Confusion matrix of test data-set for VGG 16

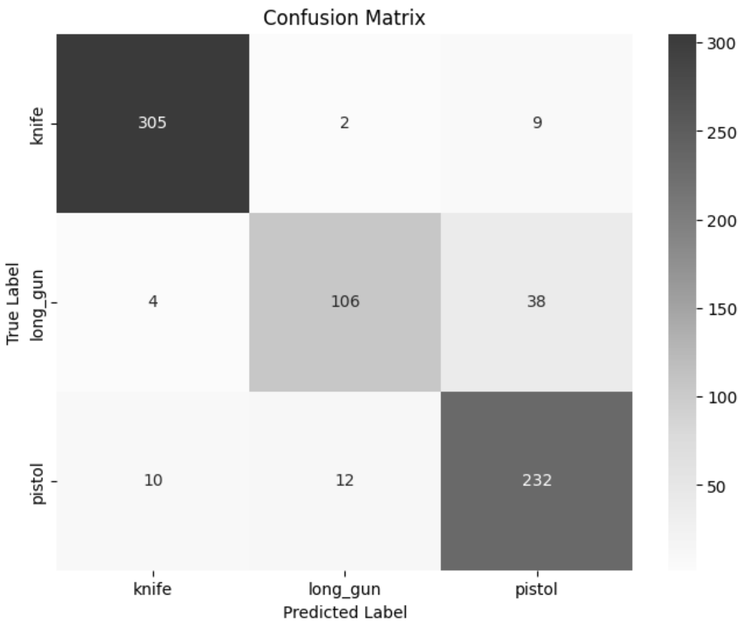

Figure 4.5: Confusion matrix of test data-set for VGG 19

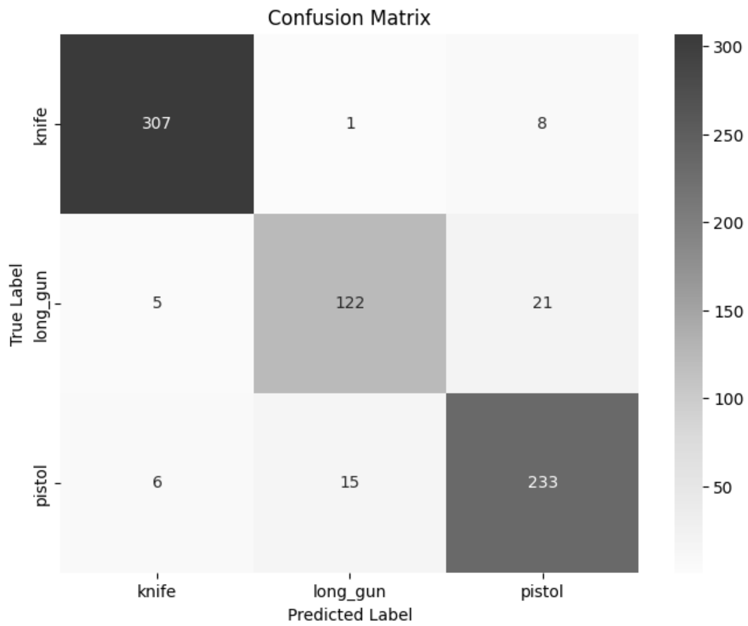

Figure 4.6: Confusion matrix of test data-set for proposed VGG-net (25 layer)

4.2 Results

The performance of various deep learning architectures was evaluated to identify the most effective model for weapon detection. Standard architectures such as VGG-16, VGG-19, ResNet-50, ResNet-101, and EfficientNet-B0 were trained using transfer learning with ImageNet pre-trained weights. Additionally, custom 21-layer and 25-layer VGG-based networks were developed to explore improvements through architectural modifications. Among these, the proposed 25-layer VGG-Net achieved the best overall performance, demonstrating strong generalization across weapon categories including pistols, knives, and long guns. While the 21-layer model slightly edged out in classification accuracy, the 25-layer model exhibited more stable training behavior and higher consistency in results across multiple runs.

Performance metrics were calculated using classification accuracy (the proportion of correct predictions over the total number of predictions) and F1 score (the harmonic mean of precision and recall) providing a balanced measure of a model's ability to correctly classify both positive and negative instances. Results are summarized below:

Architecture	Classification Accuracy	F1 Score
VGG 16	0.8983	0.8960
VGG 19	0.8955	0.8938
VGG-net (21 layer)	0.9248	0.9155
VGG-net (25 layers)	0.9234	0.9216
ResNet50	0.8120	0.7991
ResNet101	0.7869	0.7848
EfficientNetB0	0.4833	0.4397

Table 4.1: Accuracy of Models

Some test results of images for weapon detection and classification:
(red box = predicted, green box = true)

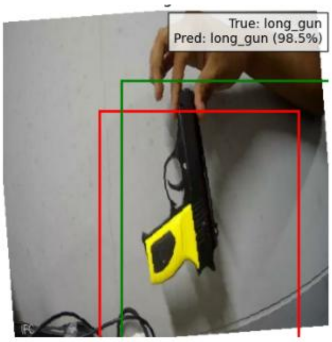

Figure 4.6 (a)

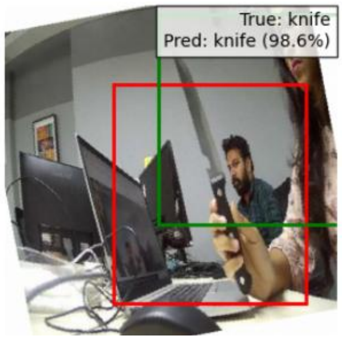

Figure 4.6 (b)

Figure 4.6 (c)

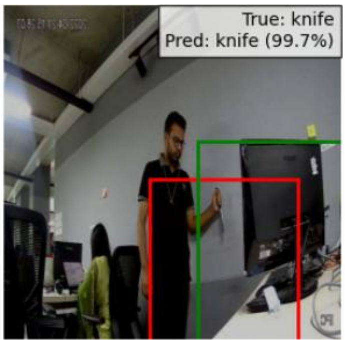

Figure 4.6 (d)

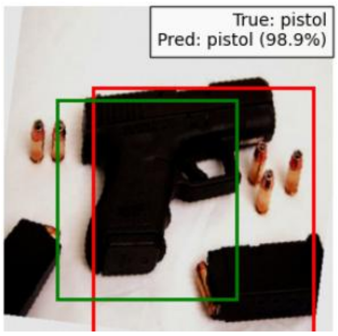

Figure 4.6 (e)

Figure 4.6 (f)

4.3 Deployment

For deployment, the weapon detection system was implemented as a lightweight web application. The frontend, built with HTML, CSS, and JavaScript, provided a responsive interface and captured real-time video from the laptop's webcam. The backend, developed using Flask, handled video stream processing and model inference. Live video frames were sent from the browser to the Flask server, which performed detection using the trained model and returned the results with bounding boxes and class labels, enabling real-time weapon detection in the browser.

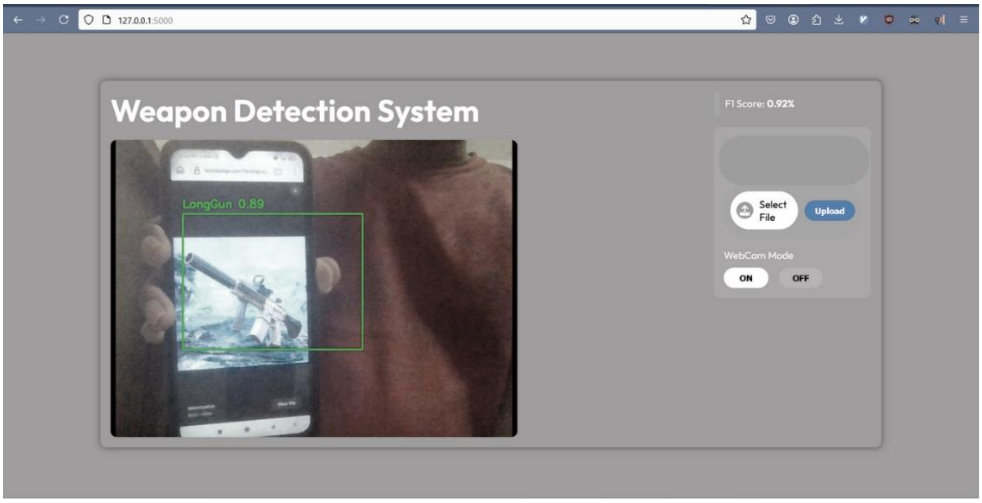

Figure 4.7: Web appliction

4.4 Discussion

4.4.1 Model Performance

The proposed VGG-net model with 25 layers showcased exceptional performance in identifying and localizing weapons across diverse scenarios. By leveraging transfer learning, architectural enhancements, including additional convolutional layers, Leaky ReLU activations, and batch normalization, the model achieved remarkable classification accuracy of 92.16 %, outperforming standard VGG-16(89.60 %), VGG-19(89.38 %), ResNet50(79.91 %), ResNet101(78.48 %), and EfficientNetB0(43.97 %). The bounding box regression component also showed high precision, with minimal error rates as indicated by metrics like mean absolute error (MAE) and mean squared error (MSE). The use of Huber loss for bounding box regression ensured stability and robustness during training, particularly for outlier cases.

However, certain limitations were observed, particularly in scenarios involving heavy occlusions or extreme variations in lighting conditions. While data augmentation techniques mitigated these challenges to some extent, further improvements could involve incorporating additional datasets or employing ensemble techniques. Nonetheless, the model strikes an effective balance between computational efficiency and accuracy, making it suitable for real-time applications with reasonable hardware constraints.

4.4.2 Practical Implications

The proposed weapon detection system offers substantial potential for enhancing public safety in various high-risk areas, including airports, schools, and public events. Its ability to detect and localize weapons in real-time makes it a critical tool for law enforcement and security personnel. For instance, its integration with existing CCTV networks could enable proactive monitoring, reducing response times during emergencies and preventing escalations.

The system's adaptability to various environmental conditions, such as dim lighting or crowded spaces, makes it well-suited for deployment in urban and rural settings alike. The empirical investigation demonstrated the system's effectiveness across multiple armament categories, bolstering its applicability in real-world scenarios. The inclusion of bounding box predictions ensures not just identification but also precise localization of potential threats, further strengthening its operational utility.

Moreover, the lightweight design of the proposed VGG-net model enables deployment on edge devices such as drones or portable scanners, enhancing mobility and situational awareness. The Flask-based API can facilitate seamless integration into larger security frameworks, supporting centralized monitoring and coordination across multiple surveillance nodes.

Chapter 5

CONCLUSION

The development of a weapon detection and classification system utilizing the VGG-net architecture marks a significant advancement in leveraging deep learning for automated surveillance. The proposed system achieved outstanding performance, with a detection accuracy of 92.16%, surpassing the capabilities of the standard VGG models. By increasing the depth of the network to 25 layers, incorporating techniques such as batch normalization, Leaky ReLU activations, and a novel design for fully connected layers, the proposed VGG-net model demonstrated superior performance in detecting and classifying weapons in real-world scenarios.

The standard VGG-16 architecture, renowned for its simplicity and effectiveness, served as a robust baseline for this work. However, its inherent limitations, such as the fixed feature extraction capabilities and absence of modern enhancements like skip connections, constrained its performance on complex weapon detection tasks. The modified version addressed these shortcomings by employing a combination of layer-freezing for pre-trained features and custom layers optimized for the detection of small and intricate objects. The results highlight that the modified VGG-16 model is not only accurate but also efficient in terms of processing time, making it viable for real-time applications. Furthermore, the model demonstrated robustness across varied environmental conditions, including low-light scenarios and complex backgrounds, which are critical factors for practical deployment in surveillance systems.

Despite its success, the proposed VGG model has certain limitations. One drawback is the increased computational overhead compared to the standard VGG-16 model, which could pose challenges when deploying the system on devices with limited processing capabilities. Additionally, the model occasionally struggles with extremely small or partially obscured objects, indicating a need for further refinement in feature extraction techniques. The reliance on pre-trained weights from ImageNet, while advantageous in reducing training time, introduces a dependency on datasets that may not fully capture the unique characteristics of weapon detection tasks. Another limitation is the relatively static architecture, which may not adapt well to rapidly evolving security needs or new types of weaponry. As the model is specifically fine-tuned for a three-class dataset in this implementation, expanding it to handle more diverse weapon classes or generalizing it for multi-object detection remains a challenge.

In summary, the proposed architecture offers a significant improvement over the standard VGG models and some other CNN models such as ResNet50, ResNet101 for weapon detection and classification tasks, providing enhanced accuracy and real-world adaptability. However, addressing the identified drawbacks through future optimizations and architectural innovations will be essential to ensure the system's scalability, efficiency, and robustness for broader security applications.

Chapter 6

FUTURE WORKS

While the current implementation of the weapon detection system using the modified VGG-16 architecture has demonstrated exceptional accuracy and robustness, there are several areas for improvement and expansion.

One of the key directions for future development is integrating the model with real-time surveillance systems, such as CCTV cameras and drone-based monitoring platforms. This integration would involve optimizing the model for low-latency inference, allowing it to process live video streams and detect weapons in real time. The addition of advanced tracking algorithms, such as object trajectory analysis, could further enhance its capabilities by identifying and following suspicious activities across multiple camera feeds.

Deploying the web-based application for weapon detection and classification system on a cloud-based platform such as AWS or Google Cloud, users could access the system remotely, enabling centralized monitoring and rapid response capabilities. A feature can also be added to allow users to upload images or video footage for weapon detection and classification.

The current implementation was trained on a three-class dataset, which limits its generalizability. Future work could focus on building or leveraging larger, more diverse datasets that encompass a broader range of weapon types, such as explosives, missiles and other potentially dangerous objects. Expanding the dataset to include varied lighting conditions, weather patterns, and crowd densities would improve the model’s robustness and ensure its applicability in diverse scenarios.

To address the occasional shortcomings in detecting small or occluded objects, future work could explore the integration of attention mechanisms or transformer-based models, which have shown promise in focusing on fine-grained details in images. Additionally, implementing multi-scale feature extraction techniques could help the model better handle weapons of varying sizes and orientations.

Ensuring the system's reliability against adversarial attacks is another critical area of future research. Attackers may attempt to bypass detection by modifying images or introducing adversarial noise. Techniques such as adversarial training, defensive distillation, and robust loss functions could be explored to make the model more resilient.

REFERENCES

[1] Murugaiyan, S., Gatta, V. A., & Bellam, K.. “An enhanced weapon detection system using deep learning”. *Proceedings of the International Conference on Networked Wireless Communications* (ICNWC), 2024.

[2] Yadav, Pavinder & Gupta, Nidhi & Sharma, Pawan.. “A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods”. *Expert Systems with Applications*, 2022.

[3] Olmos R, Tabik S, Herrera F. “Automatic handgun detection alarm in videos using deeplearning. *Neurocomputing*.”, pp.66–72, 2018.

[4] Fernandez-Carrobles et al. "Challenges in Real-Time Weapon Detection Using Faster R-CNN." *Pattern Recognition and Applications Journal*, vol. 25, no. 1, pp. 33-49, 2019.

[5] Verma and Dhillon. "Transfer Learning Techniques for Enhanced Weapon Detection." *IEEE Transactions on Neural Networks*, vol. 31, no. 4, pp. 356-370, 2020.

[6] Gelana and Yadav. "Sliding Window Approach for CNN-Based Weapon Detection." *International Journal of Computational Intelligence*, vol. 28, no. 2, pp. 89-102, 2021.

[7] Dwivedi et al. "Implementation of VGG-16 Network for Pistol Detection." *Journal of Advanced Computer Vision*, vol. 14, no. 5, pp. 215-230, 2021.

[8] Singh et al. "YOLOv4 for Weapon Detection in Diverse Datasets." IEEE Access, vol. 9, pp. 64572-64581, 2022.

[9] Huang et al. "Improving Weapon Detection with Overlapping Tile Techniques." *Machine Vision and Applications Journal*, vol. 29, no. 7, pp. 300-314, 2021.

[10] Ozge Unel et al. "Tiling Techniques for Small Object Recognition in Weapon Detection." *Proceedings of the International Conference on Computer Vision*, pp. 905-920, 2020.

[11] Khalid et al. "Data Augmentation Strategies for Enhanced Weapon Detection Using YOLOv5." *Journal of Computer Vision and Machine Learning*, vol. 19, no. 6, pp. 510-528, 2022.

[12] Lamas et al. "Pose Estimation for Human-Held Weapon Detection with EfficientDet." *Neural Processing Letters*, vol. 38, no. 9, pp. 175-192, 2023.

[13] Bhatti et al. "Evaluation of Object Detection Algorithms for Weapon Detection." *ACM Transactions on Intelligent Systems*, vol. 25, no. 4, pp. 287-305, 2022.

[14] Salido et al. "Comparative Analysis of CNN Models for Weapon Detection." *Image and Vision Computing Journal*, vol. 32, no. 3, pp. 210-225, 2023.

[15] Sankaranarayanan et al. "Gaussian Blends with Bayesian Kalman Filters for Weapon Tracking in CCTV Networks." *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 5, pp. 521-534, 2021.

[16] Jain et al. "Real-Time Weapon Detection Using Faster R-CNN and SSD Models." *Proceedings of the International Symposium on AI and Applications*, pp. 451-462, 2022.

[17] V. Pravallika, & C. Rami Reddy. "Weapon Detection and Alerting System'' *Journal of Nonlinear Analysis and Optimization*, 13(2), pp. 1–12. ISSN: 1906-9685, 2022.

[18] H. Jain, A. Vikram, Mohana et al. "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications." *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 193-198, 2020.

[19] G. Swetha, S. Sai Sanjeev Charan et al. "Video Analysis for Weapon Detection and Alerting." *International Journal of Human Resource Management and Organizational Behavior*, vol. 12, no. 2, pp. 23–33, 2024.

[20] Kuruppu, S. "AI System to Protect Endangered Animal Population and Prevent Poaching Threats using Weapon Detection." *International Journal of Innovative Science and Research Technology*, vol. 8, no. 9, pp. 1270,2023.

[21] Berardini, D., Migliorelli, L., Galdelli, A., et al. "A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras." *Multimedia Tools and Applications*, vol. 83, pp. 19109–19127, 2024.

[22] A.H. Ashraf et al., "Weapons Detection for Security and Video Surveillance Using CNN and YOLO-V5s," *Comput. Mater. Contin.*, vol. 70, no. 2, pp. 2761-2775, 2022.

[23] Gijare, A. "Automatic and Accurate Weapons Detection Model Using an Optimal Neural Network Architecture." *SSRN*,2022.

[24] F. A. Razzaq, M. A. Chaudary, S. Fareed, W. Tariq, M. Waqas, and S. Javaid, "Enhancing Public Safety: Detection of Weapons and Violence in CCTV Videos with Deep Learning," 2023 25th International Multitopic Conference (INMIC), pp. 1–6., 2023.