International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Weapon Classification System using VGG16 and Inception ResNet Models

Yi Yi Aung[a]*, Kyi Zar Oo[b]

[a]*Myanmar Institute of Information Technology, Mandalay, Myanmar*
[b]*University of Technology (Yatanarpon Cyber city, Pyin Oo lwin, Myanmar)*
[a]*Email: yiyiaungcu007@gmail.com*
[b]*Email: kyizaroo1@gmail.com*

Abstract

Security forces urgently need to implement computerized systems in regard to the increasing number of criminal acts. In the battle against crime, the construction of modern weapon recognizing systems has become crucial. The nature and carefulness of the crime are determined by the type of weapon. In this study, distinct types of weapons classification using deep learning models is presented. The presented approach is developed using the Keras architecture, which is based on the TensorFlow framework, and makes use of the VGGNet and Inception ResNetV2 architecture. The classifier is trained using three classes: knife, gun, and background. The model uses the weapon images from Roboflow. The presented approach outperforms the VGG-16 model (96.25% accuracy) and Inception ResNet-V2 model (97.92% accuracy) in terms of classification accuracy. This study offers a crucial perspective on how well the presented deep learning models handle the challenging issue of weapon classification.

Keywords: Weapon Classification; Deep Learning; VGG16; Inception ResNet.

1. Introduction

In recent years, there has been a rapid increase in crime rates related to the usage of weapon. Gun violence is a critical worry in every country, but particularly in those where gun ownership is allowed. An average of 10.42 weaponry are possessed by 100 persons worldwide, according a survey [1]. There is a significant need for active monitoring systems to deal with the growing number of public crimes. Types of weapons determine the harshness of the crime. Identifying the possibility of any crime happening and determining the best course of action can be aided by ongoing surveillance including weapon classification.

Received: 4/30/2025
Accepted: 6/12/2025
Published: 6/23/2025

---

* Corresponding author.

Computer vision and artificial intelligence have made it easier to identify and categorize objects according to application requirements. Applications include security feeds, self-driving cars, and more. Therefore, this paper concentrates on classifying three classes of weapons. Weapons can be categorized using deep learning-based methods or conventional methods with machine learning classifiers. In conventional methods, the classifier is trained using manually derived features. Any new input image is then classified using the trained model. How accurate these kinds of methods is dependent on the diverse and robust the extracted features. Deep Convolutional Neural Networks (DCNN) are a greater option for overcoming these restrictions because they don't require any explicit input image features [2]. Several convolutional layers, pooling layers, and fully linked layers make up deep convolutional neural networks.

The convolutional layers in DCNN act as automatic feature extractors. Unlike traditional machine learning methods, where manual feature engineering is required, DCNNs learn to detect patterns and features that distinguish one from another. It outperforms the machine learning-based method in terms of accuracy. It can be accomplished by applying the transfer learning, which minimizes the model improvement time and eliminates the need for a huge dataset. According to these advantages, the weights of convolutional layers of new model with the weights of pre-trained Visual Geometry Group (VGG-16) model are initialized for the weapon classification. The system flow of the study is shown in Figure 1.

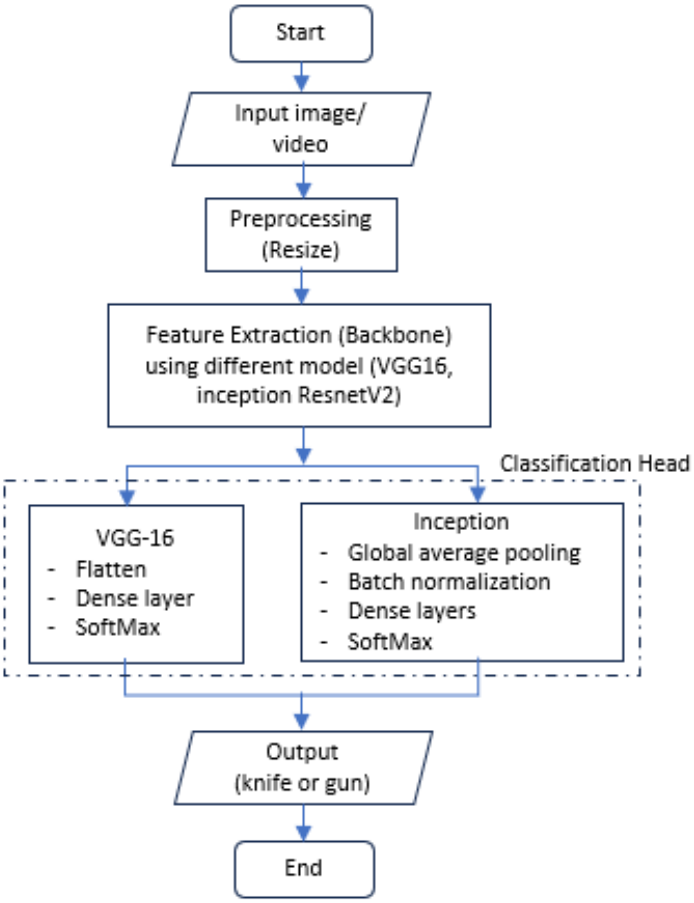

Figure 1: Flow diagram of the proposed weapon detection system

The aim of this study is to present a weapon classification system to classify different types of weapons to determine the method required to train, create layers, implement the training process, save training, determine the success rate of the training, and test the trained model. The proposed system tested on a robust dataset with high-quality weapon images from the public Roloflow100 to classify three types of weapons (Guns, Knives, and background) by employing deep learning approaches. The proposed system is employed with VGG16 and Inception ResNetV2 models to assess and compare the classification performance of proposed models.

The structure of the paper is as follows: Section 2 provides a review of the relevant literature. Section 3 outlines the materials and methods used in this study. Section 4 presents the results and analysis, while Section 5 offers a discussion of the findings. Finally, Section 6 concludes the study.

2. Literature Review

Currently, the surveillance cameras in different areas to prevent crime is the main interesting topic in computer vision. Many studies have been used different evaluation measure techniques for the realization of weapon detection and classification. The most useful deep learning models are You Only Look Once (YOLO) [3], Visual Geometry Group (VGG-16) [4], Residual Networks (ResNet) [5], and GoogleNet [6]. The classification of weapon may still have inaccuracy. By predicting bounding boxes and class probabilities simultaneously through a single neural network, YOLO achieves remarkable speed, processing images at 45 FPS with the base model and up to 155 FPS with its smaller variant, Fast YOLO. Despite a trade-off in localization accuracy, the model significantly reduces false positives and demonstrates strong generalization across domains, even outperforming established methods like DPM and R-CNN on non-natural image datasets. Most of the weapon classification and detection techniques try to solve this issue. In recent times, numerous deep learning-based approaches have been applied to various weapon detection and classification problems.

The two novel methods were introduced in a study on weapon categorization created with a deep CNN. The weights of the previously trained VGG-16 model were used in the proposed method. This model was used to examine how altering the completely connected layer neuron count affected categorization. It is also determined experimentally that when the dropout rate increases, average accuracy decreases. This finding concludes that just increasing the number of neurons would never increase accuracy. By leveraging the pre-trained VGG-16 model and fine-tuning it with a dataset comprising images of guns, knives, and no-weapon scenarios collected from various sources, the authors develop a robust classifier capable of distinguishing between weapon types with high accuracy. Training was conducted on an Nvidia GeForce GTX1050 Ti GPU to ensure efficient handling of the large image set, resulting in an impressive classification accuracy of 98.41%. This work demonstrates the potential of deep learning in enhancing real-time decision-making in surveillance systems by providing accurate weapon detection and classification, which is crucial for timely and appropriate threat response [7].

In 2020, the authors [4] used a variety of machine learning models, such as Single Shot Detection (SSD) and Region Convolutional Neural Network (RCNN), to solve the weapon detection problem. This paper presents a timely and relevant study on enhancing security through automated weapon detection in video surveillance

systems, leveraging advancements in computer vision. The proposed approach implemented on two types of datasets. The experimental results show that both algorithms attained good accuracy; however, the trade-off between speed and accuracy may determine how they are used in real-world situations. The SSD accuracy of 73.8% is unsatisfactory when compared to RCNN's speed. The higher RCNN offered better accuracy, but SSD higher speed allowed for real-time detection. By utilizing both pre-labelled and manually annotated datasets, the system is trained for improved adaptability and performance in diverse environments. The study demonstrates that both models yield strong detection accuracy, though their suitability for real-world deployment depends on the trade-off between inference speed and precision which SSD offering faster processing, while Faster R-CNN provides more accurate localization. Overall, this work underscores the critical role of deep learning in intelligent surveillance and contributes meaningfully to the development of responsive and efficient security monitoring systems.

Mane presented a deep learning based a robust and automatic weapon detection system based on four models including Faster RCNN with Inceptionv2, Faster RCNN with Resnet50, SSD with Inceptionv2 and SSD with Resnet50. The own created image dataset was divided in two parts: 80% was used for training and 20% was used for evaluation. The experimental study demonstrates that Faster-RCNN models outperform SSD models for weapon detection systems. This study concluded that the Faster R-CNN with InceptionV2 outperformed all the other models but the detection speed of Faster-RCNN is not quick [1].

Another study on the automatic recognition of knives and guns suggested methods to alert human operators when the closed-circuit television system recognized the presence of knives or guns. This paper addresses the growing demand for intelligent CCTV surveillance systems by proposing automated algorithms capable of detecting firearms and knives in real-time video feeds. Recognizing the limitations of human operators in monitoring numerous camera feeds simultaneously, the study focuses on developing a practical solution that minimizes false alarms that it is an essential criterion for real-world deployment. The proposed knife detection algorithm demonstrates superior specificity and sensitivity compared to existing methods, while the firearm detection system achieves an impressive near-zero false alarm rate. These advancements suggest that the system can serve as an effective early warning tool, enhancing situational awareness, enabling quicker responses, and potentially reducing casualties in dangerous scenarios. The same study minimized the frequency of false alarms and built a system that could provide real-time warning when a dangerous situation was recognized in order to implement the system in real life [8]. In 2021, the authors offered a novel approach for weapon classification based on the VGGNet architecture. In this paper, a new constructed dataset consisting of seven different weapon types is used for classification. The suggested model achievement accuracy of 98.40% is superior to that of the VGG-16 model, which has an accuracy of 89.75%, the ResNet-50 model, which has an accuracy of 93.70%, and the ResNet-101 model, which has an accuracy of 83.33%. After the classification, the tested images contained the weapon coordinates. The weapon class and percentage success rate were calculated by placing these locations into a rectangular frame. By effectively identifying weapon images with various backdrops, the recently created model demonstrated excellent performance [9]. In order to automatically detect firearms in video surveillance, Salido and his colleagues [10] conducted a comparison analysis utilizing three CNN models. The paper presents an innovative approach to automatic handgun detection in videos, addressing the need for systems that minimize human supervision in surveillance and control applications. Additionally, the impact of

posture data on false alarms was assessed. The study further evaluates two classification strategies: the sliding window approach and the region proposal approach. Among the models tested, the Faster R-CNN-based model, trained on a newly curated database, demonstrated the most promising results, even in challenging conditions such as low-quality YouTube videos. The findings showed that the average precision and recall of RetinaNet tuned with the unfrozen ResNet-50 backbone were 96.36% and 97.23%, respectively. With accuracy and F1-score of 96.23% and 93.36%, respectively, the YOLOv3 obtained the highest results. However, for the limited dataset with low picture resolution, the 21 and 8 obtained as false negatives and false positives, respectively, were too high. The introduction of a new metric, Alarm Activation Time per Interval (AATpI), adds value to the evaluation process, providing a novel way to assess the efficacy of detection systems in video surveillance. Singh and his colleagues [11] also used YOLOv4 to train an image dataset containing swords, knives, machine guns, shotguns, pistols, and other weapons. The several weapon classes were grouped together and obtained mean average precision and average loss of 77.75% and 1.314, respectively. This paper presents a significant advancement in real-time weapon detection from CCTV surveillance videos, addressing the growing concern of weapon misuse due to increased accessibility. By employing a Scaled-YOLOv4 model trained on a custom weapons dataset, the authors achieve an impressive mean average precision (mAP) of 92.1 and an inference speed of 85.7 FPS on a high-performance RTX 2080TI GPU. The system's optimization using TensorRT for deployment on the Jetson Nano demonstrates a thoughtful approach to balancing high accuracy with low-latency and cost-effective edge computing solutions. However, real-time weapon detection cannot be adequately measured by mean average precision and mean average loss alone. Pose estimation was used by Lamas and his colleagues [12] to create a traceable and repeatable top-down weapon recognition method that could take advantage of the person using the weapon. Four distinct CNN architectures (Faster R-CNN, ResNet50, EfficientDet, and CenterNet) were used to train the system to identify knives and firearms. EfficientDet fared better than the other deep learning models when the Sohas weapon dataset was used for training. The technology performed well in identifying weapons used by humans. Data augmentation was used by Khalid and his colleagues [13] to address the intrinsic issues of rotation, affine, size, and occlusion. The extended dataset was trained using the YOLOv5 model, which produced an accuracy of 95.43. The requirement to prevent or reduce the use of illegal weaponry in society has continued to be a research bottleneck. A CCTV camera-based effective weapon detection method was proposed by Carrobles and his colleagues [14]. To identify knives and firearms, a Faster Region-based Convolutional Neural Network (Faster R-CNN) technology was created. The public weapon dataset, which contains photos of weapon objects in a range of sizes, was used to train the system. This paper aimed to detect weapon in public areas. The authors evaluate the effectiveness of two CNN architectures such as GoogleNet and SqueezeNet for the detection of guns and knives. The SqueezeNet-based model achieved an impressive 85.44% accuracy for gun detection, while GoogleNet showed relatively moderate performance in knife detection at 46.68%. Despite the disparity between the two detection tasks, both approaches demonstrate notable improvements over prior studies, highlighting the potential of lightweight and efficient deep learning models in enhancing real-time threat detection in surveillance systems. Idakwo and his colleagues [15] proposed an efficient weapon detection and classification system using the Capsule network. The tile numbers that need to be generated for each input image that is taken from the CCTV cameras must be determined. The CCTV captured width and height ratio size of image against the capsule network input size is used to detect knives and guns. It enhanced the surveillance image small-scale weapon image by employing a

distinct tiling processing technique based on attention and memory mechanisms. The system was trained using the public Mock Attack dataset which has varying sizes of weapon object images with their respective size ratio. The presented method can be successfully used in a surveillance system, as evidenced by average accuracy of 99.43%. In perspective it briefly, research has generally tried to categorize weapons in the literature; but no study has demonstrated the ability to classify and distinguish between various weapons. The current study suggests a weapon classification with extremely high accuracy and the ability to recognize weapon types than previous studies.

3. Material and Methods

3.1. Dataset

There are 1063 images that consist of background, gun, and knife were organized by downloading from the Roboflow, public object detection dataset. To successfully detect and recognize real-life weapons, the downloaded images must be of high-quality images with different angles. The Python programming language is used to develop the weapon classification system. The input images are resized into 224 × 224 × 3. The images are then categorized and labeled according to the weapon class to which they belonged. The sample images are shown in Figure 2.

Figure 2: Sample images of backgrounds, guns and knives from Roboflow dataset

3.2. Visual Geometry Group (VGG16) Model

Weapon classification using DCNNs involves training a neural network model to automatically recognize and categorize various types of weapons from images. This paper uses VGG16 and Inception ResNetV2 as deep

learning architecture. VGG16 is one of the most well-known and widely used Convolutional Neural Networks (CNNs) for image classification. It was developed by the Visual Geometry Group at the University of Oxford and won the 2014 ImageNet competition with high performance. It consists of 16 layers: 13 convolutional layers and 3 fully connected layers. The convolutional layers are arranged in blocks, with each block containing a series of convolutional layers followed by max-pooling layers. It uses small 3x3 convolutional filters and 2x2 max-pooling layers. By using small filters repeatedly, the model captures increasingly complex patterns while reducing the number of parameters and keeping the network deep.

After the convolutional layers, it has three fully connected layers with 4096, 4096, and 1000 neurons, respectively for classification. The output layer produces class probabilities. The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks. Rectification (ReLU) non-linearity is present in all hidden layers. Additionally, it should be mentioned that all but one network lack Local Response Normalization (LRN); this type of normalization increases computation time and memory usage rather than improving performance on the weapon dataset. The classification layer, which uses the softmax activation function, is the final layer of the model. In this layer, the output value is attained by labeling the number of classes in the dataset. The architecture of the VGG16 model is illustrated in Figure 3.

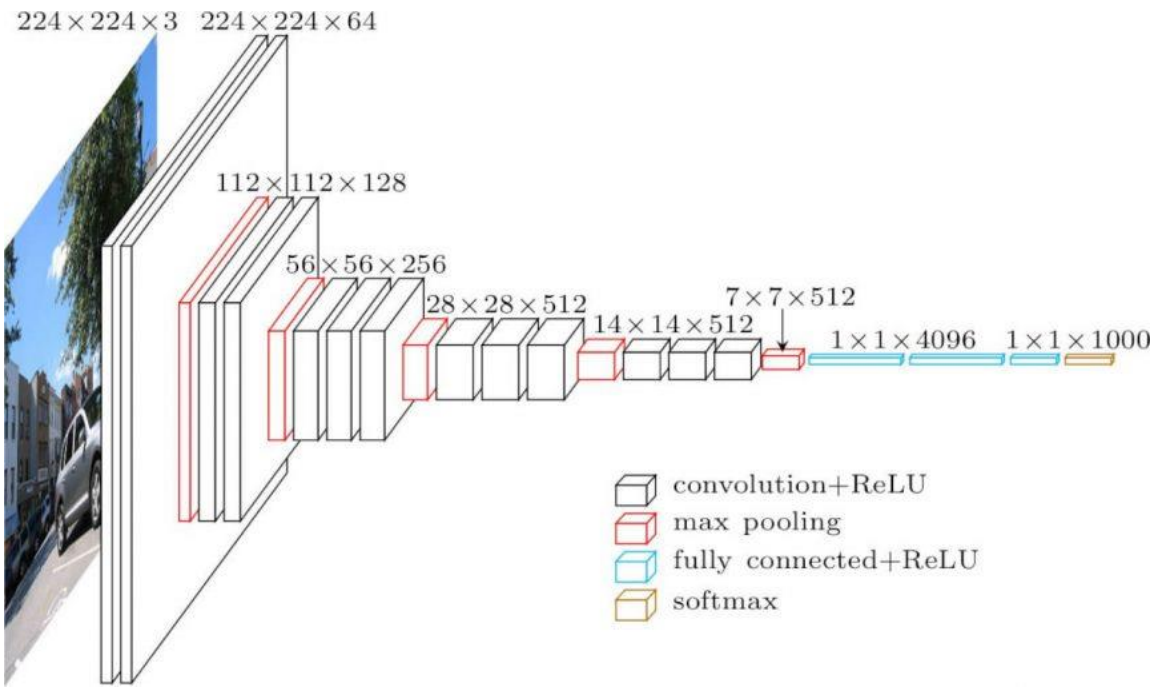

Figure 3: Architecture of VGG16

3.3. Inception ResNetV2 Model

Residual Networks (ResNet) was developed by Microsoft Research and won the 2015 ImageNet competition. The residual unit is shown in Figure 4. It introduced the idea of skip connections (also known as residual connections) to address the degradation problem that occurs when deep networks fail to improve as the number of layers increases. Inception is also another dominant architecture, and Inception V2 is known for its inception

modules, which allow the model to use multiple convolutional filter sizes in parallel within the same layer. The inception module uses multiple convolution filters with different sizes (e.g., 1x1, 3x3 and 5x5) along with pooling layers to capture features at multiple scales simultaneously. This reduces the need for deeper layers and more complex models while still allowing the network to learn complex features.

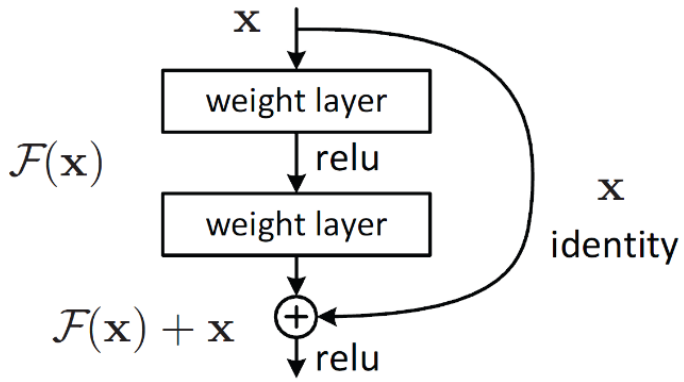

Figure 4: Inception architecture with residuals

The Inception-ResNetV2 architecture combines the strengths of both Inception Networks and ResNet to create an efficient and powerful model for image classification and other computer vision tasks. It optimizes convolutional layers by factorizing them, which reduces computational cost. For instance, a large 5 x 5 convolution can be factorized into two smaller convolutions (3 x 3), which reduces the number of parameters and the computational cost. It is great for capturing features at different scales, which is beneficial for weapons that come in various shapes and sizes. It is also more computationally efficient compared to deeper networks while still maintaining good performance.

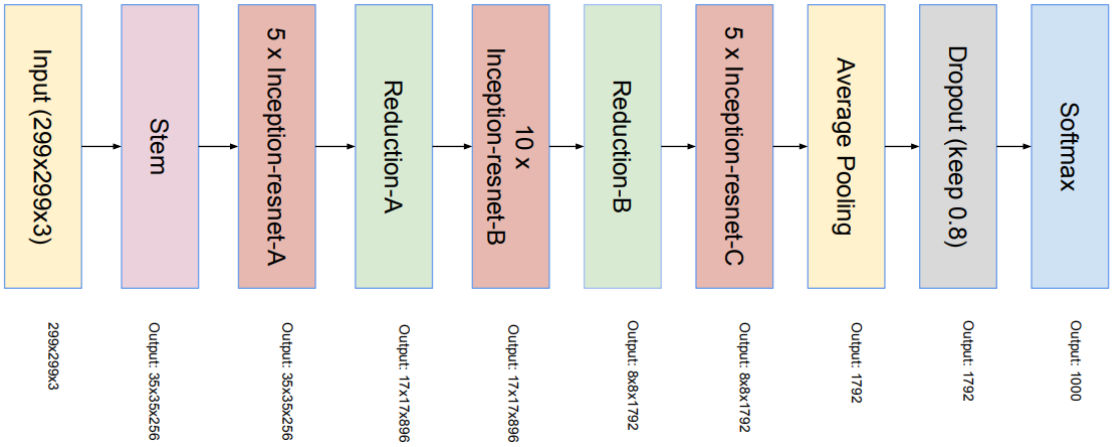

Figure 5: Overall Architectures of Inception ResNet V2

The Inception-ResNetV2 model has around 164 layers as shown in Figure 5. Despite the high number of layers, the use of residual connections helps prevent the degradation problem that often occurs in very deep networks. Each convolution block has 3 convolution layers and each identity block also has 3 convolution layers. It has

around 55 million trainable parameters which is significantly fewer than models like VGG16 (138 million). The efficient architecture allows to achieve high performance with fewer parameters. After the feature extraction layers, the final output is passed through a fully connected layer (typically with a SoftMax activation function for classification). Both are powerful architectures for feature extraction and classification tasks due to their specific designs and capabilities. Experimental setup and results have been explained in the section 4.

4. Experimental Results and Analysis

In experiments, three image classes (knife, gun, and background class) have been conducted. Numerous knife and handgun photos have been downloaded from the public available Roboflow. Images of people, automobiles, chairs, and other objects are included in the background class. The dataset is split into training and test sets at random. For VGG16, the training set contains 691, 192 and 180 images of knives, guns and background class respectively. Test set contains 200, 20 and 20 images of knives, guns and background classes respectively. For Inception RestNetV2, there are 1063 images of knives, guns, and the background class in the training set. In the test set, 240 images are included. For all of the experiments, 30 epochs have been used for the training of these networks. Batch size of 16 is used for these experiments in the training phase. To update the weights of the network, Adaptive Moment Estimation (Adam) optimizer is used.

An input size is set at 224 × 224. To extract the initial features, 64 kernels of size 3 × 3 are applied to the input image in two successive convolutional procedures. An input image size is decreased by applying a pooling layer of size 2× 2. The completed image is subjected to two successive convolutional procedures using 128 kernels and a max pool operation in order to extract the deeper features. This is done twice again, each time employing 512 kernels in sequence.

To analyze the efficiency of the proposed approach, confusion matrix and classification accuracy are used as evaluation parameters. Figure 6 and 7 shows the training and validation performance metrics of the Inception ResNetV2 and VGG 16 models across 30 epochs. According to Figure 6 and 7, blue line represents training accuracy, and rapidly increases, fluctuating significantly between epochs but generally improving, reaching close to 1.0. The orange line represents validation accuracy, which steadily increases and stabilizes around 0.975. The fluctuations in the training accuracy could indicate that the model is learning well but may have some variance or overfitting tendencies. The smoother validation loss curve suggests the model generalizes well on unseen data, though the fluctuating training loss may indicate some noise or instability during training.

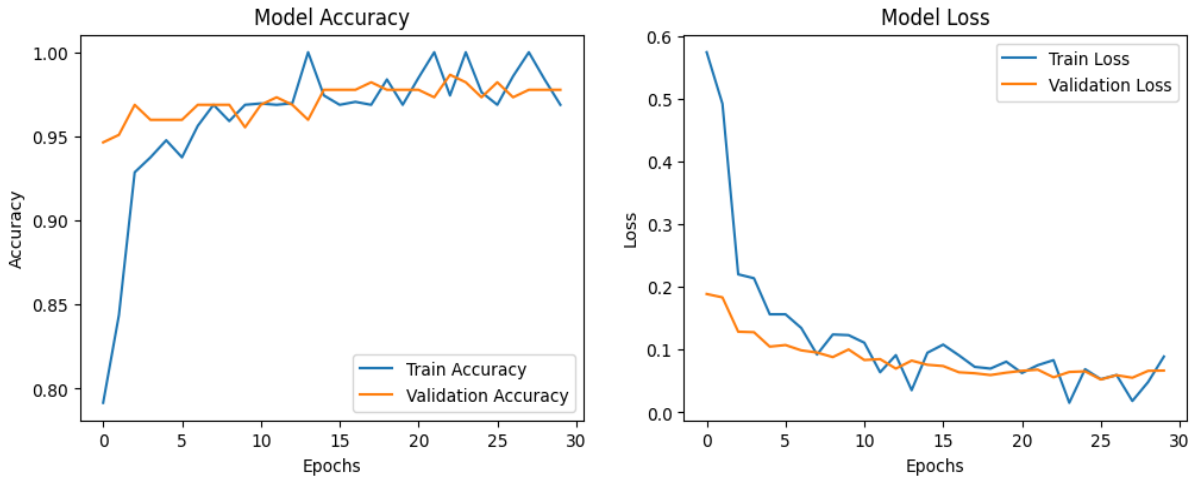

Figure 6: Accuracy and Loss of the Model while Training and Validation using Inception ResNetV2

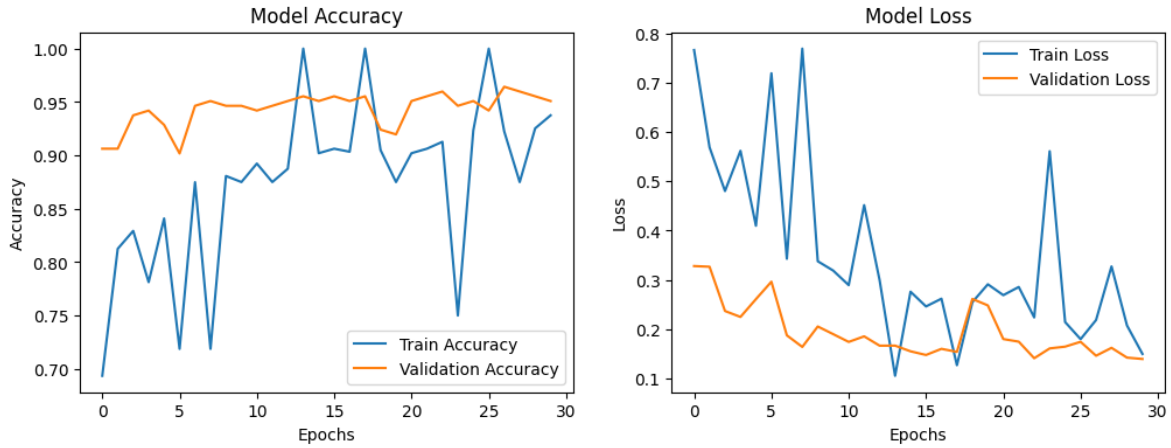

Figure 7: Accuracy and Loss of the Model while Training and Validation using VGG16

In brief, both models perform well with both high accuracy and low loss on the validation set. The accuracy raised ground and stable validation loss indicate the model has converged successfully. Table 1 shows the classification accuracy and processing time in minutes for both models, VGG16 and Inception ResNetV2. From this table, it can be concluded that the average accuracy with 30 epochs achieves good results for Inception ResNetV2. For VGG16, the accuracy was less than ResNetV2 and the training took more time. Notably, unless there is a lot of data, these two models overfit quickly, making it impossible to learn more.

Table 1: Classification Accuracy and Execution Time

Model name	Accuracy (%)	Time (min)
VGG16	96.67	25
Inception ResNet V2	97.92	15

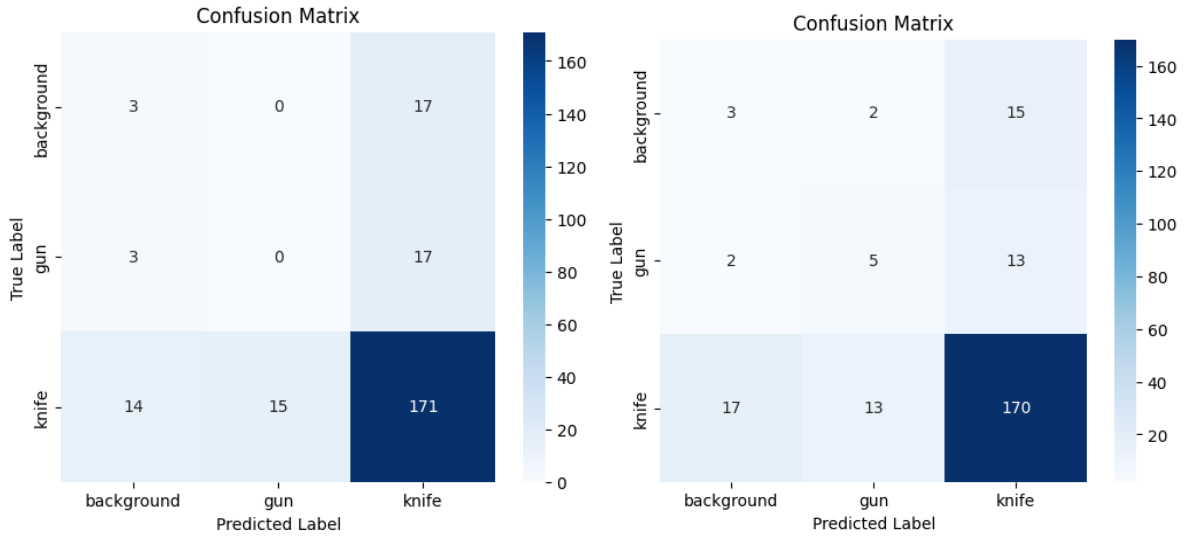

Figure 8: Confusion matrices of the performance classification using Inception ResNetV2 and VGG16

Figure 8 illustrates the confusion matrix for Inception ResNetV2 and VGG16 through three classes: knife, gun, and background. The matrices show how well the model predicts each class by comparing the true labels to the predicted labels. The first figure shows the classification performance of InceptionResNetV2 model. The confusion matrix reveals that the model performs well in classifying "knife" images, correctly identifying 171 instances, but struggles significantly with "gun" and "background" classes. It fails to correctly classify any "gun" images, often mislabeling them as "knife," and misclassifies most "background" instances in a similar manner. This suggests a strong model bias toward the "knife" class, possibly due to imbalanced training data or overlapping features between weapons. To enhance performance, especially for gun detection, strategies such as data augmentation, class rebalancing, and incorporating attention mechanisms should be considered.

The second matrix is the performance of VGG16 classification model. The model correctly 5 gun images and 3 background images, indication better differentiation, although misclassification remains an issue, especially with knife, which continues to dominate prediction. While knife images are still well-classified with 170 correct predictions. Many background and gun images are still being incorrectly labeled as knife. The confusion between weapon classes and background suggests overlapping features and possible data imbalance, highlighting the need for further refinement through techniques such as data augmentation, feature enhancement, or class weighting to improve accuracy and reduce bias toward the knife class.

5. Conclusion

Video surveillance is necessary for weapon identification and classification in order to lower the number of crimes that occur in public areas such as malls and schools. This paper uses VGG16 and Inception ResNetV2 as a basis model to offer deep CNN architectures. The convolutional layer weights are initialized using the weights of the previously trained VGG16 model on the Roboflow datasets in order to train the presented networks, but the fully connected layer weights are initialized at random. Training the proposed network using images of knives, guns, and background classes allows the weights to be fine-tuned. The usefulness of the presented model

is demonstrated by the maximum accuracy of 96.67% achieved for VGG16 and 97.92% for Inception ResNetV2, respectively. Inception ResNetV2 achieves high performance while maintaining a relatively low number of parameters compared to VGG16. Additionally, it may be used to larger datasets by training with GPUs and studies should be developed to detect coated guns. Future studies could explore the integration of attention mechanisms to enhance the model's focus on weapon-relevant regions, potentially improving detection accuracy for smaller or partially obscured objects. Expanding the dataset with more diverse scenarios and weapon types, including occlusions and varying lighting conditions, would help improve model generalization and robustness in real-world environments.

References

[1] S. B. Mane, "Weapon Detection and Classification Using Deep Learning," *ITEGAM- J. Eng. Technol. Ind. Appl. ITEGAM-JETIA*, vol. 10, no. 47, 2024, doi: 10.5935/jetia.v10i47.1039.

[2] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 10, 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[4] H. Jain, A. Vikram, Mohana, A. Kashyap, and A. Jain, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India: IEEE, Jul. 2020, pp. 193–198. doi: 10.1109/ICESC48915.2020.9155832.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[6] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[7] N. Dwivedi, D. K. Singh, and D. S. Kushwaha, "Weapon Classification using Deep Convolutional Neural Network," in *2019 IEEE Conference on Information and Communication Technology*, Allahabad, India: IEEE, Dec. 2019, pp. 1–5. doi: 10.1109/CICT48419.2019.9066227.

[8] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, "Automated Detection of Firearms and Knives in a CCTV Image," *Sensors*, vol. 16, no. 1, p. 47, Jan. 2016, doi: 10.3390/s16010047.

[9] V. Kaya, S. Tuncer, and A. Baran, "Detection and Classification of Different Weapon Types Using

Deep Learning," *Appl. Sci.*, vol. 11, no. 16, p. 7535, Aug. 2021, doi: 10.3390/app11167535.

[10] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, pp. 66–72, Jan. 2018, doi: 10.1016/j.neucom.2017.05.012.

[11] S. Ahmed, M. T. Bhatti, M. G. Khan, B. Lövström, and M. Shahid, "Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos," *Appl. Sci.*, vol. 12, no. 12, p. 5772, Jun. 2022, doi: 10.3390/app12125772.

[12] A. Lamas *et al.*, "Human pose estimation for mitigating false negatives in weapon detection in video-surveillance," *Neurocomputing*, vol. 489, pp. 488–503, Jun. 2022, doi: 10.1016/j.neucom.2021.12.059.

[13] S. Khalid, A. Waqar, H. U. Ain Tahir, O. C. Edo, and I. T. Tenebe, "Weapon detection system for surveillance and security," in *2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, Manama, Bahrain: IEEE, Mar. 2023, pp. 1–7. doi: 10.1109/ITIKD56332.2023.10099733.

[14] M. M. Fernandez-Carrobles, O. Deniz, and F. Maroto, "Gun and Knife Detection Based on Faster R-CNN for Video Surveillance," in *Pattern Recognition and Image Analysis*, vol. 11868, A. Morales, J. Fierrez, J. S. Sánchez, and B. Ribeiro, Eds., in Lecture Notes in Computer Science, vol. 11868. , Cham: Springer International Publishing, 2019, pp. 441–452. doi: 10.1007/978-3-030-31321-0_38.

[15] M. A. Idakwo, R. E. Yoro, P. Achimugu, and O. Achimugu, "An Improved Weapons Detection and Classification System".