

# Model for Hidden Weapon Detection Using Deep Convolutional Neural Network

<sup>1</sup>Moradeke Grace Adewumi, <sup>2</sup>Olumide Sunday Adewale and <sup>3</sup>Bolanle A. Ojokoh

<sup>1</sup>Department of Computing and Information Science, Bamidele Olumilua University of Education Science and Technology, Ikere Ekiti State, Nigeria.

<sup>2,3</sup>Department of Computer Science, Federal University of Technology, Akure. Ondo State, Nigeria.

DOI : <https://doi.org/10.51583/IJLTEMAS.2024.1311013>

Received: 21 November 2024; Accepted: 26 November 2024; Published: 16 December 2024

**Abstract:** Insecurity has been a major threat to government and civilians in Nigeria for the past decade. Development of a security system is not yet enough to curb the situation. Hence, the need for weapon detection using Convolutional Neural Network. The researchers downloaded different images with guns and knives from the internet. Image labeler software was used to annotate each image separately and the results were saved as XML files. This was converted to CSV files which are represented in form of rows and columns. Rows are each element, while the column are the weight, height, Xmin, Ymin, Xmax and Ymax. Which represent the shape and location of the boxes. Extra files were created which was mapped to a particular number, and the label was represented in form of numbers such as 1 for knife and 0 for gun. TensorFlow API was used for the training. We trained 300epochs at 0.03 learning rate for Resnet50, Resnet101, InceptionV1 and the proposed model. The success rate of the training was determined, and the trained model was tested. The proposed model performed better than three other models when trained and tested with the same datasets.

**Keywords:** Weapon detection; Insecurity; Convolutional Neural Network.

## I. Introduction

Global rise in terrorism increases on daily basis despite the advancement in government munitions in combating terrorists. This has posed a serious threat to the government and civilian. The development of more security systems is required to curb this menace.

Terrorism involve weapons usually have significant impact on public, psychological and economic cost. Many people die yearly as a result of terrorist's violence. Psychological trauma is frequent among children who are exposed to high levels of terrorism in their communities (Sanam *et al.*, 2021). Children that witness terrorist activities or be a victims can experience a negative psychological effects for a very long term. Studies shows that handheld gun and knife are the primary weapons used for nefarious activities. These acts can be reduced by identifying the disruptive behaviour at early stage and monitoring the suspicious activities carefully so that law enforcement agencies can further take necessary actions (Velasin S. A. *et al.*, 2019).

The use of weapons such as improvised explosive devices (IED), guns and knives has been part of the major threat to Nigerian for the past decade which involved series of attacks by Nigerian terror organizations who make use of these dangerous weapons for their nefarious activities (Nalajala *et al.*, 2016). Unique among its sort occurred in Nigeria in 1986 when a letter bomb was sent to Dele Giwa in his house which led to his untimely death (Oluwasanmi, 1986). Nothing of such was heard until 2011 when series of blast occurred at the UN headquarter in Abuja that claimed many lives and destroyed lots of properties (Knoechelmann, 2014). These incidences have been occurring since then at different places such as markets, places of worship, bus stop and campuses where many people are involved (Pethó-Kiss, 2020). The most advanced and dangerous new bombs use mobile phones to enable terrorists to set off a device immediately (Vanimireddy and Kumari, 2012). This arose the development of weapon detection using Convolutional Neural Network (CNN).

## Convolutional Neural Network

This is an integral part of deep learning that is used for intelligence, processing, accuracy, and data improvement. It is composed of a multi-data processing layer that trains the data representation through an abstraction of the various levels. Convolutional Neural Network (CNN) is been used in various research studies such as: human pose segmentation, face recognition, image classification, image detection, speech recognition and so on (Lim, 2017). The convolutional neural networks are effective in applications such as image/video recognition, semantic parsing, natural language processing and detection. CNN are one of the best deep learning methods used in the computer vision domain. Prior to the deep learning era, traditional object detection techniques relied on hand-coded features that were not robust to changing lighting conditions and failed to detect objects when their orientations were changed. CNN's have been found to be more accurate and faster and it is a combination of layers in which each layers plays a distinct role in the network (Panthula, 2018).

## Object Detection

Object detection according to (Zhao and Zheng, 2012) is a term used to describe the process of locating things inside an image. Face detection, pedestrian detection, and skeleton detection are some of the most common subtasks. Object detection is one of the most fundamental computer vision issues, and it can provide useful information for semantic comprehension of images and videos. It's used in different applications such as: image categorization, human behaviour analysis, face recognition, and autonomous driving. Object detection is the process of finding and classifying objects in an image. This can be approached using deep learning like regions with convolutional neural networks (RCNN) that combines rectangular region proposals with convolutional neural network features. Object detection is a supervised machine learning problem, which involve training models on labelled examples. Each image in the training dataset must be accompanied with a file that includes the boundaries and classes of the objects it contains. Object detection's main goal is to identify and find one or more effective targets in still image or video data. It covers a wide range of critical techniques, including image processing, pattern recognition, AI, and machine learning. Image classification is the prediction of object in an image. Object localization is the identification of the location of one or more objects in an image and drawing a bounding box around their context. Object detection is the combination of localization and classifications. Object detection uses feature extraction and learning algorithms or models to recognize instances of various category of objects (Pallavi, *et al.*, 2019). Object detection is the estimation of the class and location of objects contained within an image. It is basically an instance-wise vision task. Prior to the rise of deep learning, object detection in computer vision was accomplished using manually created machine learning features including shift-invariant feature transform, histogram of directed gradients, and many more (Sultana *et al.*, 2019).

(Adewumi *et al.*, 2022) identified some variables that can be used to identify the terrorists most especially in a gathering, this can be furthered strength by developing a model for detecting such weapons in order to eradicate such menace in our society. The use of weapons to perform evil acts has been posing a serious threat to Nigerians, as seen by a number of attacks carried out by the Boko Haram terrorist group in the country for the past decades. On the 5<sup>th</sup> June, 2022, a mass shooting and bomb attack occurred at a Catholic church in the city of Owo in Ondo State, Nigeria, 41 corpses were recorded while many were injured. This kind of occurrence had claimed numerous lives and left numerous buildings and businesses in ruins.

Indeed, Nigeria in particular in Africa has shown the global expression of terrorism. The deployment of improvised explosive devices, targeted killings, ambushes, drive-by shootings, suicide bombers, and kidnappings are some of the terrorists' tactics that call for urgent attention (Kingdom *et al.*, 2015).

## Research Motivation

Insecurity has been major challenge confronting our societies this days, rumor of wars from different quarters of the world almost every day with the use of weapons. This is worrisome and called for concern. Weapons are harmful objects that are used by some sets of people most especially the terrorists to injure governments, civilians and the military. Most of these weapon objects are not easily identified by necked eyes, on this basis, there is a need to develop a model that can be used to identify and detect these objects on the human body most especially while in a crowd to save people from being injured.

## Aim and Objectives of the study

To develop a Region- Based Convolutional Neural Network Model that can identify and detect any form of weapon on an image.

## Related Works

(Kaya *et al.*, 2021) observed that with increased number of criminal activities, automatic control systems seems to becoming the primary need for security measure. He proposed a model to detect seven different weapon types using the deep learning method. Qiang *et al.*, (2020) proposed an object detection algorithm by jointing semantic segmentation (SSOD) for images. Wei, (2019) improved convolutional pose machines for estimating human pose using image sensor data. The goal was to create a new system that uses Google Neural Network and convolutional pose machines to estimate human position. Lim, (2017) worked on the design of a training network based on a convolutional neural network for the classification of objects. The goal was to create a convolutional neural network-based training network and train the picture data set for object classification in a limited number of class problems. (Jong Hyun Kim, 2017) utilized Visible Light Camera Sensors for Nighttime Images with Convolutional Neural Network-Based Human Detection. The goal was to use a convolutional neural network to detect humans in a range of situations. Fox *et al.*, (2017). Worked on simulation and mathematical modeling for the identification of suicide bombers. The plan was to use radar to find people wearing suicide bomb vests with wires for detonation. Rafi, (2016) Explored an effective convolutional network for estimating human poses. He created a network architecture with a minimal memory footprint that is effective for estimating human position, and he trained it with components that follow best practices for effective learning. His objective was to learn features at various scales and in various levels. (Akçay *et al.*, 2020). Used several CNN driven detection paradigms, including sliding window based CNN, to work on deep convolutional neural network architectures for object categorization and detection within X-ray baggage security footage. In all the literature that was evaluated, no researchers investigated the use of convolutional neural networks to detect hidden objects. Therefore in order to protect civilians from the threat of insecurity in the society, there is need to identify all instances of weapons on human body, especially when in a crowded setting. Hence this study.

## II. Methodology

The proposed algorithm used regions with convolutional neural networks (RCNN), to find weapon objects on an image. RCNN is a deep learning approach that is used to detect various object in an image. The R-CNN detector first generate region proposals, that is the regions that might contain an object using a selective searches algorithm that based on computing hierarchical grouping of similar regions based on color similarity, fitness, shape, size and texture.

Color Similarity: This group the most similar color regions until the whole image becomes a unique region. This is obtained using equation 1

$$S_{colour}(R_i, R_j) = \sum_{k=1}^n \min(C_i^k, C_j^k) \quad (1)$$

where:  $S$  is the similarity,  $C_i^k, C_j^k = K^{th}$  the values of histogram bin of region  $R_i$  and  $R_j$  of the object.

Texture Similarity: This is calculated by generating gaussian derivatives of image and extracts histogram with bins for each texture channels using equation (2)

$$S_{texture}(R_i, R_j) = \sum_{k=1}^n \min(t_i^k, t_j^k) \quad (2)$$

where:  $t_i^k, t_j^k = K^{th}$  are the value of texture histogram bin of region  $R_i$  and  $R_j$  respectively

Size Similarity: This make smaller region merge easily by reducing many bounding boxes to fewer ones that contain objects within the given image.

$$S_{size}(R_i, R_j) = 1 - \left( \frac{size(R_i) + size(R_j)}{size(img)} \right) \quad (3)$$

where:  $size(R_i), size(R_j)$  and  $size(img)$  are the sizes of regions  $R_i, R_j$  and image in the pixel respectively.

Fit Similarity: This merges different regions that are fit with each other to reduce the number of bounding boxes to fewer ones for easy identification. This is obtained using equation (4).

$$S_{fit}(R_i, R_j) = 1 - \left( \frac{(size(BB_{ij}) - size(R_i) - size(R_j))}{size(image)} \right) \quad (4)$$

where:  $size(BB_{ij})$  is the size of bounding box around  $i$  and  $j$ .

The equations (1-4) are combined to form equation (5). This will be used to determine the presence of an object(s) in the image. This is obtained in equation (5).

$$S_{R_i R_j} = a_1 * S_{color}(R_i, R_j) + a_2 * S_{texture}(R_i, R_j) + a_3 * S_{size}(R_i, R_j) + a_4 * S_{fit}(R_i, R_j) \quad (5)$$

The suspected regions are cropped out of the image and resized. The CNN classifies the cropped and resized regions into weapons and non-weapons. Performance evaluation was measured using detection rate, accuracy and precision.

Detection rate measures the percentage of true targets that is detected. It is obtained using equation (6).

$$Detection\ Rate(DR) = \frac{TP}{(TP+FN)} \quad (6)$$

Accuracy is the measure of the actual performance of the system with regard to both correctly detecting and rejecting targets. It is calculated by the sum of the true positives and the true negatives relative to the total number of GT objects as obtained in equation (7)

$$Accuracy(A) = \frac{(TP+TN)}{CGT} \quad (7)$$

Precision is the fraction of detected items that are correct. This is obtained in equation (8).

$$Precision = \frac{TP}{(TP+FP)} \quad (8)$$

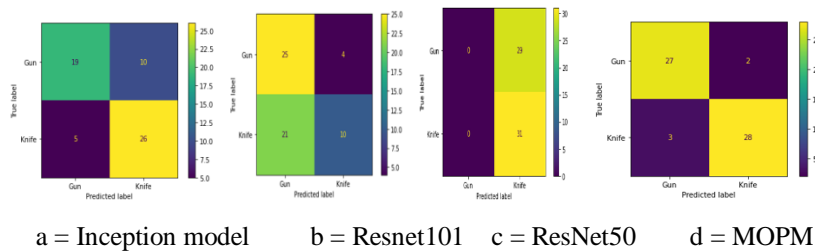
## III. Result and Discussion

### Training the Network for Classification

The datasets were trained for classification. The images were converted to JPG, for the datasets to follow the same format, the datasets were converted to xml file, because the targets were represented in xml files. The target in xml files were converted into strings; "0 to represent gun and 1 to represent knife". These were converted to RGB format size 227 by 227 by 3. The same function was developed for the four models used; ResNet50, ResNet101, Inception model and my own proposed model (MOPM).

**Confusion Matrix for the Models**

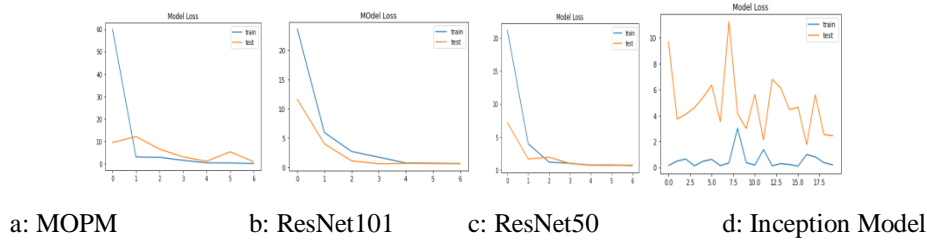
labels = ['Gun', 'Knife']



a = Inception model    b = Resnet101    c = ResNet50    d = MOPM

Fig.1: Confusion matrix for the models.

**Plot\_result for the Model Loss**



a: MOPM    b: ResNet101    c: ResNet50    d: Inception Model

Fig. 2: Training loss rate for the models

**Plot\_result Accuracy for the 4 Models**

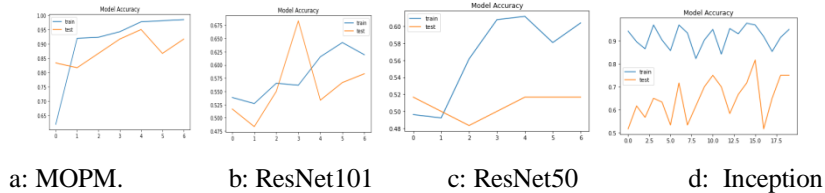


Fig.3: Training accuracy rate for the models

Table1: Training reports for the ResNet50, ResNet101, Inception and the MOPM

	RsesNet50	ResNet101	Inception	MOPM
Sensitivity	0	86	65	93
Specificity	1	71	84	90
Accuracy	52	58	75	92
Loss Rate	2	3	3	2

From table1 above and from figures 2 and 3, MOPM, has a sensitivity rate of 93%, specificity rate of 90% and Accuracy rate of 92% which was better than inception model, ResNet50 and ResNet101. This implies that the new model performed better than the three other models in classifying weapon objects when trained with the same datasets using the same functions. This is represented in the graph in figure 4.

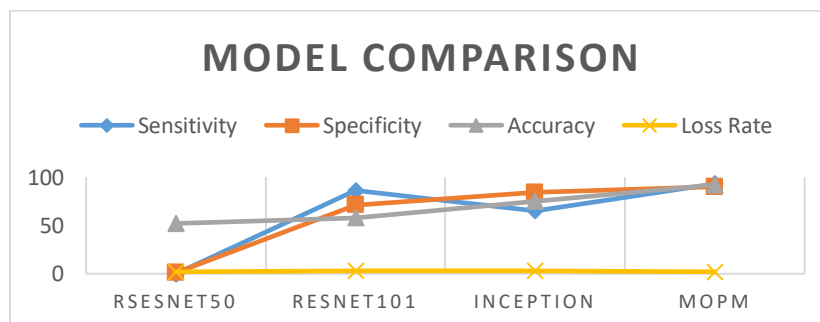


Fig. 4: Models comparison graph

### Performance Evaluation

The performance of the MOPM model was evaluated by prepared test data from images of guns and knives captured from human body which were downloaded from the internet. In order to test the accuracy of the model network, Region-based Convolutional Neural Network deep learning approach from selective search technique was adopted. This is shown in figure 5.

From the figure5: (a) is an input image, (b) extracts series of region proposals from the image, (c) computes features for each proposal using a convolutional neural network (CNN), and (d) classified each region (Kaya *et al.*, 2021).

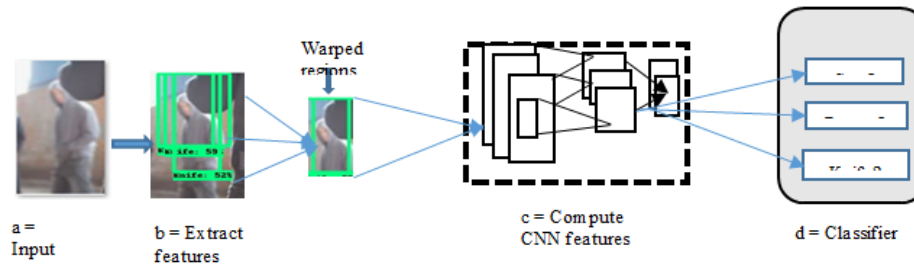


Fig. 5: Region proposal generation (Modified by the author)

The weights of the individual neurons was adjusted to extract the right features from the images. The process of adjusting these weights is called training the neural network. The CNN starts the training with random weights, the researcher provided the neural network with datasets of images annotated with their corresponding classes. The ConvNet processes each image with its random values and then compares its output with the image's correct label. If the network's output does not match the label, which is likely the case at the beginning of the training process it makes a small adjustment to the weights of its neurons so that the next time it sees the same image, its output will be a bit closer to the correct answer. These are done by back propagation which optimizes the tuning process and makes it easier for the network to decide which parts to adjust instead of making random corrections. Every run of the entire training dataset is called an epoch. During training, the ConvNet passes through a number of epochs and modifies its weights. The neural network gets a little bit more adapt at categorizing the training images after each epoch. The CNN makes fewer and smaller changes to the weights as it becomes better until the network converges. Sufficient datasets are necessary to train sophisticated models, but in this situation, obtaining large datasets for the study was difficult. Datasets were immediately loaded from Google Drive utilizing the mount drive approach. The runtime instance received a full import of the drive's data. In order to access the datasets, the Google Drive was mounted. Image labeler was used to construct the object detection system in a more organized manner. The predicted bounding box coordinates were evaluated and contrast with the actual ground truth bounding box coordinates in order to evaluate an item detector. In addition to classifying the object in the image, their location within the image were also determined. Figure 6 shown some of the objects detected with the developed model with their interval rate. This implied that the developed model can be used to detect an object hiding by human being.



Fig. 6: Objects detected by the model.

## IV. Conclusion and Recommendation

### Conclusion

Criminal activities are increasing daily in our societies, it is very important to classify and detect any form of weapons on a person based on images taken from security cameras, without requiring human intervention. Therefore, Weapon detection is important to prevent criminal activities before they occur and to take appropriate measures to prevent their nefarious actions. A detection model for detecting weapon objects has been designed and implemented in this study. The benefit of this is to help an individual to avoid being exposed to the danger of weapons like improvised explosives, guns and knives, particularly while in a crowd.

### Recommendation

This research work is limited to gun and knife, which I regarded as simple implements it is recommended that further research could be carried out on the detection of complex implements such as explosive materials and others.



## References

1. Adewumi, M. G., Adewale, O. S., Akinwumi, A. O., & Ajisola, K. T. (2022). Security Intelligence Framework for Suicide Bombers Identification in a Crowd. *International Journal of Academic Research in Business and Social Sciences*, 12(4), 697–706. <https://doi.org/10.6007/ijarbss/v12-i4/13124>
2. Akcay, S., Kundegorski, M. E., Willcocks, C. G., & Breckon, T. P. (2020). On Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-ray Baggage Security Imagery. 1–13.
3. Fox, W., Vesecky, J., & Laws, K. (2017). Mathematical Modeling and Simulation for Detection of Suicide bomber. 993943, 22–30.
4. Jong Hyun Kim, H. G. H. and K. R. P. (2017). Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light Camera Sensors. MPDI.
5. Kaya, V., Tuncer, S., & Baran, A. (2021). Detection and classification of different weapon types using deep learning. *Applied Sciences (Switzerland)*, 11(16). <https://doi.org/10.3390/app11167535>
6. Kingdom, U., Nnamdi, A. C., Assembly, N., Sebastine, A. I., Junior, E. O., Kingdom, U., & Anyanwu, K. (2015). Boko-haram Crisis and Implications for Development in the Northern Nigeria. III(4), 1–12.
7. Lim, S. (2017). Classification in Few Class Problem. *Elsevier Journal*. 21(1); 144–150.
8. Qiang, B., Chen, R., Zhou, M., Pang, Y., & Zhai, Y. (2020). Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic. MPDI.
9. Rafi, U. (2016). An Efficient Convolutional Network for Human Pose Estimation. *IEEE Journal of Computer Society*, 10(6); 1002-1034.
10. Sultana, F., Sufian, A., & Dutta, P. (2019). A Review of Object Detection Models based on Convolutional Neural Network.
11. Vanimireddy, A., & Kumari, D. A. (2012). Detection of Explosives Using Wireless Sensor Networks. 3, 277–280.
12. Wei, S., Ramarkrishna, V., Kanade, T., and Sheikah, Y., (2019). Convolutional Pose Machines. *CVF. IEEE Xplore*. 4724-4731.