Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 12 (2024) pp. 1–10
© Cerebration Science Publishing

Article

An Improved Weapons Detection and Classification System

Monday Abutu Idakwo [1],*, Rume Elizabeth Yoro [2], Philip Achimugu [3] and Oluwatolani Achimugu [4]

[1] Department of Computer Engineering, Federal University, Lokoja 260102, Nigeria
[2] Department of Cyber Security, Dennis Osadebay University, Asaba 320001, Nigeria; elizabeth.yoro@dou.edu.ng
[3] Department of Computer Science, Air Force Institute of Technology, Kaduna 800282, Nigeria; check4philo@gmail.com
[4] Department of Information Communication Technology, Air Force Institute of Technology, Kaduna 800282, Nigeria; tolapeace@gmail.com
* Correspondence author: monday.idakwo@fulokoja.edu.ng

Received date: 21 March 2024; Accepted date: 20 May 2024; Published online: 29 May 2024

Abstract: The rapid increase in insecurity orchestrated by the illegal possession and usage of weapons (knives, rifles, handguns, amongst others) has led to a wider deployment of surveillance cameras for real-time video monitoring. However, the small size of these weapons, distance from the surveillance camera and atmospheric conditions have made it impossible for easier identification of the weapon or the crime perpetrators. Interestingly, several research have deployed computer vision through closed-circuit television cameras monitoring and tracking of weapons. Nevertheless, the need to improve detection accuracy, and lowering false alarm rates has remained a bottleneck. This paper addressed some of the inherent issues using a selective tile processing strategy that uses an attention mechanism. The image tiling technique was adopted as weapon images are smaller in size when compared to the entire image and down-sampling the images to a lower resolution will either reduce the features of the small weapon or make it invisible to be detected. Hence, the high-definition images from the surveillance camera in the public Mock Attack dataset were automatically splited into their respective tile images using their respective size ratio to the input of the modified capsule network. The capsule network was adopted for detecting and classification of the system owing to speed in prediction, lower data requirement, ease in pose recognition, texture, and image deformation. An average accuracy, precision, recall, and F1-score of 99.43%, 98.14%, 98.77%, and 98.45% respectively was achieved.

Keywords: camera; capsule network; computer vision; deep learning; weapon

1. Introduction

Insurgency, terrorism, robbery, and diverse crimes have dominated global discussions owing to the disruption of peace and harmony. Nigeria like most countries is not left out as kidnapping, armed robbery, cattle rustling, amongst others is threatening the peace and tranquility of the citizens. These insecurities create panic for the citizens as well as scare investors from investing. Therefore, security plays a critical role in any country's development as well as revenue generation. Hence security is paramount in human life as economic, social, and political achievements are dependent on it [1]. The ravaging insecurities have been perpetuated using weapons. Since the usage of unlicensed weapons [2] is prohibited by several countries just as in Nigeria, surveillance cameras and concealed weapon detection are majorly used for detecting these weapons. The surveillance system [3] approach involved using closed-circuit television (CCTV) surveillancep cameras to monitor strategic locations [4] that will capture any human carrying any weapon [5]. The CCTV systems are special recording systems that are mainly used for security purposes. The concealed weapon detection approach uses image sensor technology like infrared/thermal, x-ray, and millimeter waves to expose any item on the body [6]. Hence, concealed weapon body detectors are body scanners developed based on radar imaging technology. These body scanners like the ones deployed in airports, require the person to take a predefined pose and a stationary position while being scanned. Thus, it is slow and time-consuming with lower output. Therefore, not suitable for crowded places like shopping malls, and railway stations that require fast scanning [7]. It is therefore pivotal for concealed weapon detection to process the images in real time with higher accuracy. Since these crowded locations are predominantly crowded walking people, the hidden weapons are likely to be beneath the legs or arms to reduce detection. Therefore, mitigating these challenges requires fast scanning at a high frame rate while considering the various poses of the walking person. Hence a real-time scan rate of 20 fps minimum will overcome these challenges. Since carrying unlicensed weapons is prohibited, the need for evidence in the judicial system has led to the wider adoption of CCTV surveillance systems owing to their video records of scenes [8]. This evidence is crucial to ease the identification of the offenders, the nature of the crime, arrest, and prosecution. Nonetheless, useful evidence relies on the quality of the video quality as the captured images require high resolution to avoid wrong identification of criminals, and vehicles' licensed plate numbers, amongst others. Thus, CCTV cameras are important security needs to address security issues. While the existing CCTV system can remotely report any intruder or detected objects over the network, the need to further evaluate if the intruder is a security threat becomes necessary.

Interestingly, artificial intelligence technology can be utilized to detect suspicious activities through the surveillance system, like detecting weapons, movement pose, and route tracking amongst others [9]. Tracking a suspect with a weapon and monitoring the behaviour will ease the risk of clamping down on the suspect without causing much havoc to others in the scenario. Therefore, such monitoring systems are useful to the police, immigration, and customs officers among others as anomaly behaviours are reported in real-time for immediate action. Nevertheless, detecting any weapon from CCTV footage is tasking as images may be far away and the weapons are not visible [10]. It becomes more challenging if the weapons are smaller. Hence, a good preprocessing technique is crucial in enhancing the object detection model's efficiency. This paper adopted the mock attack and synthetic dataset [11] as the dataset characteristics conform with the paper's research interest**.** The dataset comprises of digital images [12] knives, short rifles, and handguns of varying sizes with fully or partially concealed weapons owing to the distance of the weapon from the CCTV cameras. The mock dataset was captured using three surveillance cameras (Cam1, Cam5, and Cam7 with 607, 3,511, and 1,031 frames respectively) with a total of 5,149 pictures with 1,920 × 1,080 as dimensions. Since the weapons in the images are small in size, down-sampling the images to a lower resolution which is a common practice in the conventional neural network will either reduce the features of the small weapon or make it invisible to be detected. Hence this paper proposes breaking down the weapon images into smaller tiles [13] that conform to the input of the deep learning network. No doubt that the convolutional neural network and its various modified forms have shown excellent performance in image detection and classification. Nevertheless, the larger data requirement during training, and failure to recognize the pose, deformation, and texture of an image has led to a wider adoption of capsule networks. The capsule network has shown an excellent performance in pose image detection with the added advantage of having the fastest prediction rate compared to other deep learning approaches [14]. Capsule network varies from CNN as it uses vectors instead of scalar, generates dynamic routing algorithms, adopts transformation process, adopts marginal loss in place of cross entropy, uses a squash function, and it embed regularization as extra loss function [15]. Therefore, this paper utilized the equivariance property of the capsule network for the effective detection and classification of knives, handguns, and short riffles. This paper's contributions are as highlighted:

- Improved the small-size weapon image in the surveillance image using a separate tiling processing strategy based on attention and memory mechanism.
- Modified the capsule network model to resolve the pose, occluded, and affine transformation issues inherent in CNN.
- Automated the resizing of surveillance images into their respective tiles-images size to suit the Capsule network input.

2. Literature Review

No doubt that the CCTV surveillance system finds application in diverse areas like pedestrian analysis, traffic systems, transportation services, and intruder detection, amongst others. The CCTV cameras are majorly installed by focusing the camera on the region of interest that needs monitoring. The need to improve the existing security surveillance system using deep learning methods has attracted diverse researchers. The traffic flow of vehicles has been predicted using deep learning in [16]. Furthermore, the

CCTV system has been deployed for human activities [17], pedestrian behavior, and pedestrian detection [18] for risk-threatening population and accident prevention. Weapon detection and the associated risk prevention are crucial to safeguard the lives of pedestrians and others.

Carrobles et al. [19] proposed an efficient weapon detection using the CCTV camera. A Faster Region-based Convolutional Neural Network (Faster R-CNN) technology was developed to detect knives and guns. The system was trained using the public weapon dataset which has varying sizes of weapon object images. Similarly, González et al. [20] implemented a real-time gun detection using the CCTV cameras. A synthetic dataset was adopted to train the model. To solve the issues of obscured weapon images created by the distance between the CCTV camera and the distance of the arm bearer, image tiling was employed. Image tiling is effective in tiny object identification. However, the frequent practice of down-sampling images during training must be avoided once the image is tiled. In image tiling, overlapping tiles are adopted to split the image into smaller images [13]. The lower-quality image is cropped from the source image using the overlapping tiles. Every tile is a representation of a new picture where the ground truth is obtained irrespective of the size of the object.

To ensure lower computation, neural networks usually adopt smaller picture sizes. Thereby affecting the microscopic object recognition. Hence, tiling enhances the performance of the CNN detection performance when compared to CNNs that scale and analyze a single image. Huang et al. [21] unveiled that stride convolutions and zero paddings result in forecasting heterogeneity within the tile border. Ozge Unel et al. [22] enhanced smaller object detection using an overlapping tile to achieve a higher precision. The tiling technique is most suitable for smaller items as well as medium-sized items. Daye [23] discussed feeding individual patches created from the non-overlapping and overlapping of decomposed tiled images into the network instead of reducing the image size. Nonetheless, inference time as well as post-processing overhead are increased due to merging and refined prediction.

Sankaranarayanan et al. [24] discussed how objects can be tracked using multiple surveillance cameras. Furthermore, mean fluctuation an aspect of segmentation used in object tracking was explored extensively. A Gaussian blend with a Bayesian Kalman filter was deployed in tracking the detected object which is pivotal in tracking a detected unlicensed weapon in a CCTV for immediate action. Jain et al. [25] equally proposed automatic weapon detection using Faster R-CNN and SSD. An accuracy of 84.60% and 73.80% were obtained with the Faster R-CNN and SSD respectively. The SDD offered real-time detection owing to a higher speed at 0.73 fps while the Faster R-CNN achieved the best accuracy with 1.606 fps speed. Salido et al. [26] performed a comparative analysis using three CNN models to automatically detect pistols in video surveillance. Furthermore, the effect of posture information on false alarms was evaluated. The results revealed that RetinaNet tuned using the unfrozen ResNet-50 backbone achieved 96.36% and 97.23% as average precision and recall respectively. The YOLOv3 achieved the highest accuracy and F1-score at 96.23% and 93.36% respectively. Nevertheless, the 21 and 8 obtained as the false negatives and false positives respectively were too high for the small dataset with poor image resolution.

Singh et al. [27] similarly adopted YOLOv4 to train an image dataset consisting of swords, knives, machine guns, shotguns, pistols, and other weapons. The various classes of weapons were classified into a class and achieved 77.75% and 1.314 as mean average precision and average loss respectively. Nonetheless, mean average precision and average loss alone cannot significantly measure real-time weapon detection. Sliding window and region proposal/object detection were two methodologies used by Bhatti et al. [5]. Explored the combination of sliding window and region object detection to effectively detect firearms. A comparative evaluation of Faster R-CNN, VGG16, SDD, Inception-ResnetV2, SSD, Inception-V3, MobileNetV1, YOLOv3, Inception-ResnetV2, and YOLOv4 algorithms were performed. The 8,327 images in the dataset consist of pistols and non-pistols. The training and test dataset consists of 7,328 images and 999 images. The results obtained showed that YOLOv4 with 91% and 91.73% as average F1 score and accuracy respectively outperforming other algorithms. However, the system had a higher false positive and false negative of 54 and 52 respectively. Lamas et al. [28] designed a traceable and reproducible top-down weapon detection using pose estimation with the ability to exploit the human holding the weapon. The system was trained to detect guns and knives using four different CNN architectures (Faster R-CNN, ResNet50, EfficientDet, and CenterNet). The Sohas weapon dataset was adopted for the training and EfficientDet outperformed the other deep learning models. The system was effective in detecting human-handled weapons. Khalid et al. [29] adopted data augmentation to resolve the rotation, affine, size, and occlusion inherent problem. The YOLOv5 model was adopted to train the augmented dataset to achieve an accuracy of 95.43. The need to curb or reduce illegal ammunition in society has remained a bottleneck to research. Although the CCTV monitoring system has shown a good prospect in detecting and tracking down crime perpetrators, real-time detection of weapons amidst large crowds has remained an issue. The reason for these hinges on the small size of this weapon, distance from the CCTV camera (occlusion), atmospheric condition, and other look-alike smaller devices like phones, and touch lights amongst others. The large data requirement of the convolutional neural network, the inability to detect image pose, and extra computational overhead created by data augmentation have made the existing approaches computationally expensive and lowered the accuracy. Hence, the false alarm rate and accuracy become a paramount issue to be addressed.

2.1. Image Evaluation Metrics

It is paramount to evaluate the performances of any newly developed model using the standard metrics to evaluate its performances. The evaluation metrics are as discussed.

2.1.1. Accuracy

Accuracy measures the number of weapons images correctly classified using Equation (1) [13];

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where:

TP is the True Positive and represents weapon images accurately predicted.
FN is False Negative and represents incorrectly predicted weapons images as normal instances.
FP is False Positive and represents tomato images inaccurately classified.
TN is the True Negative that represents instances accurately classified as normal instances.

2.1.2. Precision

Precision measures the ratio of accurately predicted images to all the samples predicted. It is mathematically given by Equation (2) [17].

$$Pre = \frac{TP}{TP + FP} \tag{2}$$

2.1.3. Recall

A recall is equally known as the detection rate. It evaluates all accurately classified weapons images ratio to the samples. As represented by Equation (3) [18]

$$Rec = \frac{TP}{TP + FN} \tag{3}$$

2.1.4. F1-Score

It measures the harmonic mean of the recall and precision as given by Equation (4) [13];

$$F1 - score = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{4}$$

3. Materials and Methods

3.1. Weapon Dataset

The performance of any deep learning model is dependent on the subjected dataset and the model's ability to learn the image features effectively. Hence, it becomes necessary to develop a high-fidelity weapon image dataset. Therefore, a robust dataset with high-quality weapon images from the public dataset Mock attack dataset was obtained. The dataset consists of 5,149 full HD images with 1,920 × 1,080 dimensions. The dataset has 2,722 weapon labels of types short rifles, handguns, and knives. The dataset was split into 80% and 20% as training and testing respectively. Therefore 2,178 and 544 labels were used for training and testing datasets. Given the high image dimension, high memory demands of processing the full HD images, and the smaller sizes of the weapons (knife, handgun, and knife) in the images, the images were separated into tiles. To avoid reprogramming the tiles to suit any surveillance-captured images a selective tile processing (STP) scheme was designed. The tile separates the input weapon images into smaller parts. The smaller parts serve as input to the capsule network thereby avoiding resizing the captured image and maintaining the weapon resolution. Therefore, for any input image obtained from the CCTV cameras, the tile numbers to be generated must be calculated. The calculation is performed using the CCTV captured images' width and height ratio size against the capsule network input size as shown in Equation (5).

$$Ratio = \frac{Imagesize}{CapsuleNetworkinputsize} \quad (5)$$

The value obtained as the ratio is rounded up and multiplied together to obtain the number of the total tile. Furthermore, the evaluated tile numbers in the horizontal and vertical direction are uniformly distributed across the input image to achieve full coverage of the image while ensuring a constant overlap between the tiles. The processing of the tile by the capsule network is based on the STP scheme that uses an attention mechanism [30]. The attention mechanism uses the statistical information obtained over time to process only a few tiles per frame while using a memory mechanism for keeping track of all non-processed tiles' activities. For instance, an image obtained from Cam5 with dimension 1920 × 1080 when subjected to the Capsule network input size of 512 × 512 will be split into 2 × 2 image tiles as shown in Figure 1.

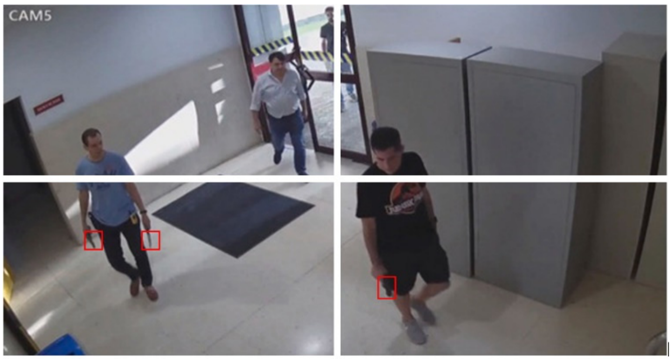


Figure 1. A 2 × 2 Tiled Images from the original Image.

3.2. Capsule Network

A special feature of the Capsule network is the ability to learn the needed image features through the computation of the desired weights. Since the proposed system does not use a defiled tile image but automatically computes the number of tiles based on the input image to capsule network input ratio, the weights are evaluated using Equation (6)

$$Weight = L \times Tw \times Th \times 3 \quad (6)$$

where:

L is the learning rate.
T_w is the tiled image width.
T_h is referred to as the tiled image height.

The proposed system hyperparameters used for the evaluation are summarized in Table 1.

Table 1. Proposed System Hyperparameters.

Routing	Learning Rate	Batch Size	Optimizer	Epoch
1,3,5	0.0001	64	Adam	50

The Matplotlib, NumPy, and Keras libraries were imported to implement the system using Python programming language. Each library installation was performed using the pip install “module name”. The tiled images from the CCTV image are the input to the system. The capsule network input size was set to 512 × 512. Therefore, all input images were tiled to fit into the input. The number tile in an image is evaluated using Equation (5). This paper adopted the squashing activation layer to develop a five-convolution layer. The first, second, and third layers have 16, 32, and 64 kernels with 5 × 5 kernel sizes using 1 stride. Maxpooling with stride and size of 2 was applied to the layers exit. While the fourth layer contains 128 kernels with a dimension of 9 × 9 and a stride of 1. Finally, the primary layer which is the fifth layer has 32 distinct capsules using 9 × 9 convolutional kernels having 1 stride. Finally, the layer responsible for labelling the weapon images uses 16 dimensions. The developed system architecture is shown in Figure 2.

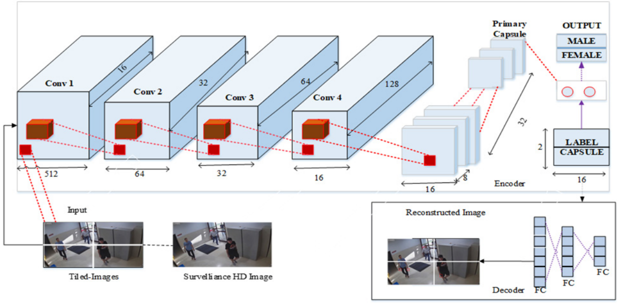


Figure 2. Developed System Architecture.

From Figure 2, the proposed capsule network architecture has three encoders (convolution layer, primary layer, and label layer) and the decoder uses the three fully connected layers for the reconstruction of the input image. The decoder reconstructs any of the input tiled images using the mean square error differences between the input tiled image and the output image. A lower mean square error is an indication that the rebuilt image is similar to the input image. The developed system uses 10-fold cross-validation to evaluate the system's performance. The dataset was split into 10 parts where 9 parts and the remaining 1 part were used for training and testing respectively. The system performance was calculated using the total average. The developed system simulation was carried out on the Google Collab Jupiter notebook [31].

4. Results and Discussions

To obtain the accurate routing parameter suitable for the proposed system. Different simulations were performed varying the routing parameters (1, 3, and 5) parameters presented in Table 2. The routing parameter r = 5 showed the best results in detecting the weapons in the CCTV images. Hence, the r = 5 was used to implement the developed system. The system was evaluated under three input conditions, tiled-images, tiled-images + full-image, and full-image. The results obtained from the evaluation of the three scenarios using the accuracy, recall, and F1-score represented by Equations (1), (2), and (3) respectively are presented in Figure 3.

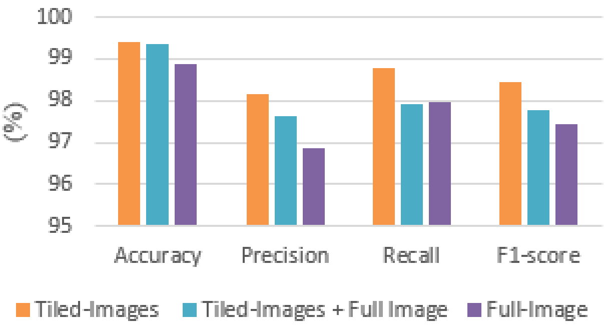
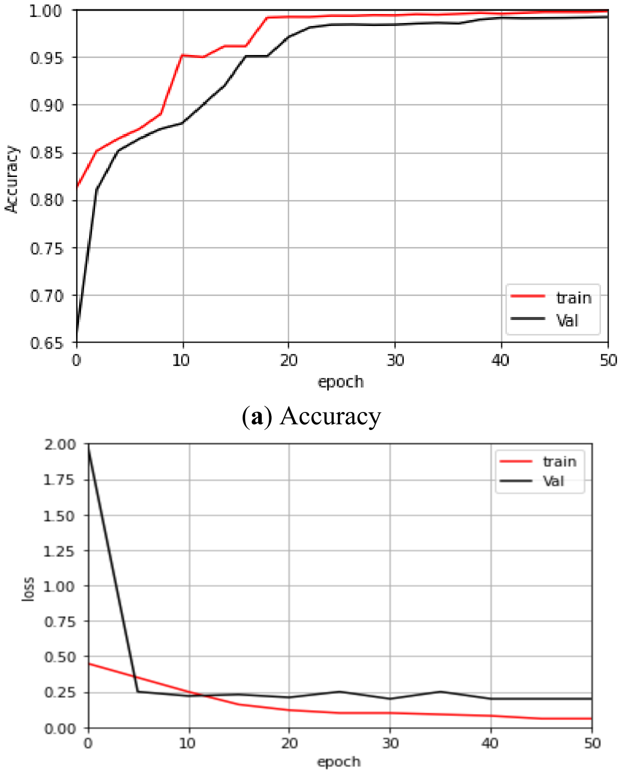


Figure 3. Performance of the Developed System.

From Figure 3, it was observed that using the CCTV-captured full image as input to the capsule network showed a lower accuracy, precision, recall, and F1-score compared to the other conditions. The reason for this border on the small size of the weapon in the original image. The combination of the tiled images and the full image as input to the developed system showed an improved performance when compared to the previous results. This approach enables both the detection of small weapons in the tiled images as well as the weapons in the full image. Hence, a larger weapon image can easily be detected from the full image. The tiled images as input to the capsule network showed the best results in terms of accuracy, precision, recall, and F1-score. The performances of the full-image and tiled-images + full-image scenarios revealed that down-sampling the original image to fit in the capsule network input size can lead to the loss of some weapon image features, thereby lowering the performance of the approach.

From the evaluated results, the tiled-images input approach was adopted for the proposed system. Thus, the proposed system preprocessed the original images into their respective tiles as input to the system. Therefore, the developed system divides every input image into an M × N tiled image. The bounding boxes and probability for the weapon images with the centre in the grid cell are simultaneously generated. The accuracy and loss graph for the system is shown in Figure 4.



(**a**) Accuracy

(**b**) Loss

***Figure* 4.** Accuracy and Loss Graph.

From Figure 4, the developed system after 50 epochs showed higher accuracy and lower losses for the training and validation data. Furthermore, the recall and false alarm rates were evaluated. The obtained result is presented in Figure 5.

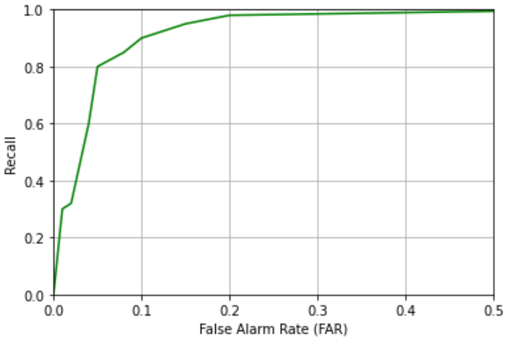


Figure 5. Recall and False Alarm Rate (FAR).

From the graph of recall against false alarm rate (FAR) in Figure 5, it is seen that the developed system showed a lower false alarm rate. To further evaluate the developed system performance, results were compared with the state-of-art methods. The comparative results are presented in Table 2.

Table 2. Comparative Analysis of the Developed System Against State of the Art.

Models	Weapons	Algorithm	Dataset	Acc.
Khalid et al. [29]	Gun	YOLOv5	University of Granada	95.43%
Rao et al. [30]	Gun, Knives	NSGCU-CNN's	Public	97.85%
Pullakandam et al. [31]	Gun, Knives	YOLOv8	Custom	90.1%
Vijayakumar et al. [32]	Axe, Knife,Pistol, Rifle and Sword	Faster R-CNN and YOLOv4	Custom	96.04
Proposed System	Knives, Rifles, Shotguns	Tiled-images and Modified Capsule Network	Mock Attack	99.43%

From Table 2, it is seen that the developed system has an average higher accuracy of 99.43 when compared to the state-of-the-art. Thus, the image-tiling approach adopted magnified the small weapon-size image for easier detection. Furthermore, the capsule network eases the detection of weapons image features effectively.

5. Conclusions

This paper has presented an improved weapon detection and classification system. The developed system resolved the inherent problem of the weapon's small size and lower detection rate by enhancing the weapon image through tiling. This process magnifies the weapons images in the entire image for easier feature detection and classification. The 99.43% average accuracy of the system shows that the system can effectively be deployed in a surveillance system. The classification of the weapons into their respective classes will guide the security agencies on the best approach to track and apprehend the culprit.

Author Contributions

M.A.I. modeled the system. R.E.Y. simulated the model. P.A. determined the performance of the model and O.A. analyzed the results. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive any funding.

Conflict of Interest Statement

Authors declares no conflict of interest.

Data Availability Statement

The dataset used is a real-time gun detection CCTV open problem dataset found at https://github.com/Deepknowledge-US/US-Real-time-gun-detection-in-CCTV-An-open-problem-dataset for University of Seville.

References

1. Zubairu, N. (2020). Rising insecurity in Nigeria: Causes and solution. *Journal of Studies in Social Sciences*, *19*.
2. Woolridge, A. W. (2023). Using federal firearm statutes to prosecute domestic terrorists. Dep't of Just. J. Fed. L. & Prac., 71, 131.
3. Tapas Badal Suneeth Kumar Guptha P Srinivasa Rao, Sushma Rani N. Object detection in infrared images using convolutional neural networks. Journal of Information Assurance and Security, 15:136–143, 2020.
4. Idakwo, M. A., Umoh, I. J., Tekanyi, A. M. S., & Adedokun, E. A. Real Time Universal Scalable Wireless Sensor Network for Environmental Monitoring Application.
5. Bhatti, M. T., Khan, M. G., Aslam, M., & Fiaz, M. J. (2021). Weapon detection in real-time cctv videos using deep learning. IEEE Access, 9, 34366-34382.
6. Gosain, S., Sonare, A., & Wakodkar, S. (2021). Concealed weapon detection using image processing and machine learning. International Journal for Research in Applied Science and Engineering Technology, 9(12), 1374-1384.
7. Hegde, S. M., Neh, N., & Shivaprasad, K. (2015). Imaging for concealed weapon detection. International Journal of Computer Applications, 975, 8887.
8. Butorac, K., & Filipović, H. (2022). Evidential Validity of Video Surveillance Footage in Criminal Investigation and Court Proceedings. European law enforcement research bulletin, (22).

9. Thomas, A. L., Piza, E. L., Welsh, B. C., & Farrington, D. P. (2022). The internationalisation of cctv surveillance: Effects on crime and implications for emerging technologies. International journal of comparative and applied criminal justice, 46(1), 81-102.
10. Ruiz-Santaquiteria, J., Velasco-Mata, A., Vallez, N., Bueno, G., Alvarez-Garcia, J. A., & Deniz, O. (2021). Handgun detection using combined human pose and weapon appearance. IEEE Access, 9, 123815-123826.
11. Hnoohom, N., Chotivatunyu, P., & Jitpattanakul, A. (2022). ACF: an armed CCTV footage dataset for enhancing weapon detection. Sensors, 22(19), 7158.
12. Idakwo, M. A., Mu'azu, M. B., Adedokun, E. A., & Sadiq, B. O. (2022). Development of a Non-Separable Integer-to-Integer Wavelet Transform-Based Digital Image Steganography System. Applications of Modelling and Simulation, 6, 20-27.
13. Reina, G. A., Panchumarthy, R., Thakur, S. P., Bastidas, A., & Bakas, S. (2020). Systematic evaluation of image tiling adverse effects on deep learning semantic segmentation. Frontiers in neuroscience, 14, 494590.
14. Patrick, M. K., Adekoya, A. F., Mighty, A. A., & Edward, B. Y. (2022). Capsule networks–a survey. Journal of King Saud University-computer and information sciences, 34(1), 1295-1310.
15. Alaoui-Elfels, E., & Gadi, T. (2021). TG-CapsNet: Two Gates Capsule Network for Complex Features Extraction. Journal of Information Assurance & Security, 16(5).
16. Essien, A., Petrounias, I., Sampaio, P., & Sampaio, S. (2021). A deep-learning model for urban traffic flow prediction with traffic events mined from twitter. World Wide Web, 24(4), 1345-1368.
17. Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recognition, 108, 107561.
18. Kim, D., Kim, H., Mok, Y., & Paik, J. (2021). Real-time surveillance system for analyzing abnormal behavior of pedestrians. Applied Sciences, 11(13), 6153.
19. Fernandez-Carrobles, M. M., Deniz, O., & Maroto, F. (2019, July). Gun and knife detection based on faster R-CNN for video surveillance. In Iberian conference on pattern recognition and image analysis (pp. 441-452). Cham: Springer International Publishing.
20. González, J. L. S., Zaccaro, C., Álvarez-García, J. A., Morillo, L. M. S., & Caparrini, F. S. (2020). Real-time gun detection in CCTV: An open problem. Neural networks, 132, 297-308.
21. Huang, B., Reichman, D., Collins, L. M., Bradbury, K., and Malof, J. M. (2018). Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations. arXiv:1805.12219. Available online at: http://arxiv.org/abs/1805.12219
22. Ozge Unel, F., Ozkalayci, B. O., & Cigla, C. (2019). The power of tiling for small object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 0-0).
23. Daye, M.A. A Comprehensive Survey on Small Object Detection. Master's Thesis, İstanbul Medipol Üniversitesi Fen Bilimleri Enstitüsü, Istanbul, Turkey, 2021.
24. Sankaranarayanan, A. C., Veeraraghavan, A., & Chellappa, R. (2008). Object detection, tracking and recognition for multiple smart cameras. Proceedings of the IEEE, 96(10), 1606-1624.
25. Jain, H., Vikram, A., Kashyap, A., & Jain, A. (2020, July). Weapon detection using artificial intelligence and deep learning for security applications. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 193-198). IEEE.
26. Salido, J., Lomas, V., Ruiz-Santaquiteria, J., & Deniz, O. (2021). Automatic handgun detection with deep learning in video surveillance images. Applied Sciences, 11(13), 6085.
27. Singh, A., Anand, T., Sharma, S., & Singh, P. (2021, July). IoT based weapons detection system for surveillance and security using YOLOV4. In 2021 6th international conference on communication and electronics systems (ICCES) (pp. 488-493). IEEE.
28. Lamas, A., Tabik, S., Montes, A. C., Pérez-Hernández, F., García, J., Olmos, R., & Herrera, F. (2022). Human pose estimation for mitigating false negatives in weapon detection in video-surveillance. Neurocomputing, 489, 488-503.
29. Khalid, S., Waqar, A., Tahir, H. U. A., Edo, O. C., & Tenebe, I. T. (2023, March). Weapon detection system for surveillance and security. In 2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD) (pp. 1-7). IEEE.
30. Rao, A., Kainth, S., & Bhattacharya, A. An Efficient Weapon Detection System Using Nsgcu-Dcnn Classifer in Surveillance. Shivam and Bhattacharya, Ansuman, An Efficient Weapon Detection System Using Nsgcu-Dcnn Classifer in Surveillance.
31. Pullakandam, M., Loya, K., Salota, P., Yanamala, R. M. R., & Javvaji, P. K. (2023, June). Weapon Object Detection Using Quantized YOLOv8. In 2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE) (pp. 1-5). IEEE.
32. Vijayakumar, K. P., Pradeep, K., Balasundaram, A., & Dhande, A. (2023). R-CNN and YOLOV4 based Deep Learning Model for intelligent detection of weaponries in real time video. Mathematical Biosciences and Engineering, 20(12), 21611-21625.

Author Biographies

Monday Abutu Idakwo obtained his Bachelors degree in Computer Engineering from Caritas University Enugu Nigeria in 2012, Masters of Science and Doctor of Philosophy in Computer Engineering from Ahmadu Bello University Zaria, Nigeria in 2017 and 2021 respectively. His strong interest is in Machine learning, Wireless Sensor network, Image processing and embedded Systems.

Dr. Rume Elizabeth Yoro is from Dennis Osadebay University Asaba. She is currently the HoD Cyber Security and Sub-Dean, Faculty of Computing. She had her BSc and MSc degrees in Computer Science from the University of Benin in 2000 and 2009 respectively and a PhD degree in Computer Science from Babcock University Ilishan-Remo, Nigeria. Her area of specialty are Intrusion detection and digital Forensic.

Philip Achimugu is a Professor of Computer Science at Air Force Institute of Technology Kaduna. He had his first and second degrees from Kogi State University and Obafemi Awolowo University in Nigeria and third degree from University of Technology Malaysia, all in Computer Science. He has held many administrative positions and review papers for many academic platforms. His area of sepcialty is in Software Engineering and Intelligent Systems.

Oluwatolani Achimugu is a Professor of Computer Science at Air Force Institute of Technology Kaduna. She had her first and second degrees from Ambrose Alli University and Obafemi Awolowo University in Nigeria. She then proceeded for her third degree at University of Technology Malaysia. All her degrees are in Computer Science. She has held many administrative positions. Her area of sepcialty is in Software Engineering.