Annals of Computing (2025), 1(6):36-43
https: doi.org /10.5281/zenodo.15729330
Research Article

Visible and Thermal Images Fusion Model for Security Threat Detection: A CNN Study

Richard Ochogwu[1], Ibrahim Babangida Salihu[2], Njideka Nwabuogo Ajumobi[3] and Christopher Okro Uzoigwe[4]

[1]Geemla Technologies, Abuja, Nigeria.
[2,3]Department of Computer Science, Nasarawa State University, Keffi, Nigeria.
[4]Department of Computer Science, Federal University, Wukari, Nigeria.
Corresponding Author's Email: richard.emoche@gmail.com

Received 3 June 20025; Acceptance 16 June 2025; Published 24June 2025.

Abstract

As illegal use of fire arms continues to pose security threat across the globe, the need for novel detection technique is timely. In this study, visible and thermal images fusion model for security threat detection: a CNN study was investigated. The study deployed the fusion of visible and thermal images techniques to design a Convolutional Neural Network (CNN)-based model for security threat detection. Thermal imaging data, preprocessing, supervised learning, detection and classification techniques of concealed objects was utilized in the model design. The study employed performance metrics, such as accuracy, precision, recall, and F1-score, to evaluate the effectiveness of the model. Results reveal that the investigated model demonstrates improved detection rates compared to traditional methods, offering a unique solution for application insecurity threat detection. Based on these findings, the study recommended among others that in the fast-paced landscape of security, it is vital to establish mechanisms for regular model retraining.
Keywords: *Visible, Thermal images, Model, Security threat, Detection, CNN.*

Introduction

Breakthroughs in computer vision and deep learning have created new possibilities for addressing security challenges in recent times [1-3]. Object detection is a key task in computer vision that involves identifying the presence or absence of specific features within image data. Object recognition is a general term to describe a collection of related computer vision tasks that involve identifying objects in digital photographs [4]. Object recognition refers to a collection of related tasks for identifying objects in digital photographs [5]. It is a crucial application of machine learning and deep learning, aiming to teach machines to interpret and recognize the content of an image in a manner similar to human perceptron.

Correspondence to: Richard Ochogwu, e-mail: richard.emoche@gamil.com

Once features are detected, an object can be further classified as belonging to one of a pre-defined set of classes [6]. Object Detection algorithms act as a combination of image classification and object localization. It takes an image as input and outputs one or more bounding boxes each labeled with a class. These algorithms are proficient in handling multi-class classification and localization and detecting objects with multiple occurrences. The main approach to detection in thermal infrared has historically been thresholding, so called *hotspot* detection [7].

Overall, the aim of object detection is to detect all instances of objects from a known class, such as people, cars or faces in an image. Object recognition algorithms rely on matching, learning, or pattern recognition algorithms using appearance-based or feature-based techniques. Generally, only a small number of instances of the object are present in the image, but there is a very large number of possible locations and scales at which they can occur and that need to somehow be explored [5].

Thermal imaging is a non-contact technique where the radiation emitted by an object is converted into a visible image called thermal image or thermogram. Any object with temperature above absolute zero (−273 °C) emits infrared radiation. The infrared radiation in the wavelength from 3 to 14 μm is called thermal infrared region [8]. Thermal imaging is simply the process of converting infrared (IR) radiation (heat) into visible images that depict the spatial distribution of temperature differences in a scene viewed by a thermal camera [9]. Thermal images are visual representations of emitted, reflected and transmitted thermal radiation within a specific area.

Thermal cameras can be classified as either cooled or uncooled depending on their technology and functionality. High-end cooled cameras can deliver hundreds of HD resolution frames per second and have a temperature sensitivity of 20 mK. Images are typically stored as 16 bits per pixel to provide a large dynamic range, such as 0–382.2K with a precision of 10 mK. Uncooled cameras generally use bolometer detectors and operate in LWIR spectrum. Uncooled thermal cameras produce noisier images at a lower framerate, but they are smaller, silent, and more affordable. A thermal camera is considered *thermographic* when it is calibrated to measure temperatures accurately [7].

Thermal cameras are ideal for perimeter or area protection offering a powerful and cost-effective alternative to radio-frequency intruder detection, fences, and floodlights. Since they do not rely on ambient light to produce images, they enable discreet surveillance even in complete darkness. In situations where some light is needed for identification, thermal cameras can minimize the need for excess illumination. They also enhance security in restricted areas, such as, in transportation zones like tunnels, railway tracks, and bridges [10]. Indoor applications of thermal cameras include building security and emergency management. They can detect humans inside a building after business hours or in emergency situations, such as when rooms fill up with smoke. High-security buildings, nuclear power plants, correctional facilities, airports, pipelines, and sensitive sections of railways also benefit from thermal camera surveillance to enhance safety and monitoring [10].

As innovations and breakthroughs in technology continue, the integration of different image modalities and innovative tools for security threat detection such as concealed weapons, will play a pivotal role in improving public safety measures and fortifying security protocols in various settings. Convolutional Neural Networks (CNNs) have displayed unprecedented performance in image analysis, facilitating the development of robust object detection and recognition systems. By leveraging the power of CNNs, it becomes possible to build a model capable of identifying armed individuals by analyzing visual data, such as images or video footage.

In essence, integrating thermal imaging into the model's architecture provides an additional layer of information, enhancing its overall effectiveness in recognizing security threats. This multi-modal approach can offer superior accuracy and reliability, especially in low-light or challenging environmental conditions. It is on this premise that visible and thermal images fusion model for security threat detection: a CNN study is presented.

Research Method

The primary data for this study was captured using a thermal camera, providing valuable insights into temperature variations and heat signatures. The decision to use a thermal camera was driven by its capability to offer real-time and accurate data in circumstances whether in industrial settings, medical applications or in field of security. It consists of 1062 JPEG thermal images of dimension 2976 x 3968 with different temperature ranges. When collecting the experimental data, the temperature was set to a range of 28.4 °C to 42.3 °C; the imaging speed was set to 10 frames per second (FPS); emissivity of 0.95, and the imaging distance was about 0.6~1.0 m. This method was chosen for its direct and in-depth insights, aligning perfectly with the project's focus on detection and recognition of armed persons. In preparation for training and evaluation, the thermal images went through a series of preprocessing steps. These steps were essential for refining the data captured by the thermal camera, ensuring it is optimized for subsequent model training and analysis in the context of designing a model for detection and recognition of armed persons using CNN. The dataset was annotated with bounding boxes to mark the concealed weapon's location within each image. To account for real-world variations, we applied data augmentation techniques, such as random rotations and translations, to create a robust and representative dataset.

Model Evaluation

For evaluating the model, the dataset is divided into 2 parts. 80 percent was used in training the model, while the remaining 20 percent was used in testing the model.

Evaluation Metrics

Scikit learn python library version 1.1.1 will be used for analysis to determine the performance of the developed model. When making classification predictions, there are four possible outcomes: True Negative

(TN), True Positive (TP), False Negative (FN), and False Positive (FP). These four standard indicators will be used to evaluate the performance accuracy of the machine learning model.

i. True Positive (TP): is when the image is predicted to contain to a weapon and it actually does belong to that class.
ii. False Positive (FP) is when the image is predicted to contain to a weapon and it actually does not belong to that class.
iii. True Negative (TN)is when the image is predicted not to contain to a weapon and it actually does not belong to that class.
iv. False Negative (FN) is when the image is predicted not to contain to a weapon and it actually does belong to that class.

These four indicators help in the understanding of what types of mistakes the learned model made in the classification of hidden weapons. They are plotted on confusion matrix which gives a glance view of each measurement indicators.

Model Performance Measurement

From the four standard indicators, three different measurement criteria can be used to evaluate the rate of the correctness of the prediction; they are: Accuracy, Recall, and Precision using Scikit learn version 1.1.1.

Accuracy

Accuracy measures the percentage of correct predictions of the model over the entire dataset. It shows the total number of predictions that are correct from the entire labeled dataset. The equation 5.1 is used to calculate the accuracy of the model.

$$Accuracy = \frac{\text{correct predictions}}{\text{all predictions}} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (1)$$

But accuracy is not good enough to determine the correctness of the performance of the prediction model because if there is a large class skew in the labeled dataset, there will be a differential misclassification costs. For our model, we have got 0.96 which means our model is approx. 96% accurate.

Receiver Operating Characteristic (ROC) curve

A better way of determining the accuracy of the model is by plotting a Receiver Operating Characteristic (ROC) curve. Our ROC value is 94.30%. The graph is obtained by plotting the True Positive (TP) rate against the False Positive (FP) rate while varying the threshold of the positive rate.

True Positive rate is obtained by the equation:

$$\text{True Positive Rate (recall)} = \frac{\text{TP}}{\text{Actual Positive}} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

While False Positive rate is given by:

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{Actual Negative}} = \frac{\text{FP}}{\text{TN+FP}} \quad (3)$$

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. We have a 0.94 precision which is pretty good. It is given by the formula in equation (4):

$$Precision = \frac{\text{true positives}}{\text{true positives + false positives}} = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

Recall (Sensitivity)

Recall is the ratio of correctly predicted positive observations to the all observations in actual class. We got a recall of 0.92 which is good for this model as it's above 0.5, it is given by the equation (5):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (5)$$

F1-Score

The last performance metric is F1-Score which combines the precision and recall metrics to give a holistic view of the overall performance of the model. Our F1 score is 0.931. It is denoted by the formula in equation (6):

$$\text{F}_\beta = (1 + \beta^2 \frac{\text{precision.recall}}{(\beta^2.precision)+recall}) \quad (6)$$

Results and Discussion

Performance comparison of the proposed system (DRAP) and existing system (CCTV and X-ray scanner).

The performance analysis of the models is evaluated based on accuracy, precision, recall, and F1-score. The rate of properly detecting the affected photographs from all images is referred to as recall, also called sensitivity. Precision is the opposite of recall. The F1-score is a combined measure of precision and recall, which shows how often the predicted value is accurate.

The summarized results of the training is given as follows:

Accuracy: 0.96

Precision: 0.94

Recall: 0.92

F1 Score: 0.93

Performance comparison of the proposed system (DRAP) and existing system (CCTV and X-ray scanner)

Table 1 presents the comparison between the existing systems and the new system.

Table 1. Comparison between the existing systems and the new system.

Classifiers	Approaches	Accuracy	Precision	Recall	F1 Score	ROC
VGC-16 [11]	VGC-16 framework	90.6%	84.21%	100%	91.43%	93.53%
YOLOv3-13 [12]	YOLO framework	72.7%	80.50%	90.30%	83.00%	78.05%
DRAP	**CNN + Tensorflow**	**96.00%**	**94.00%**	**92.00%**	**93.10%**	**94.30%**

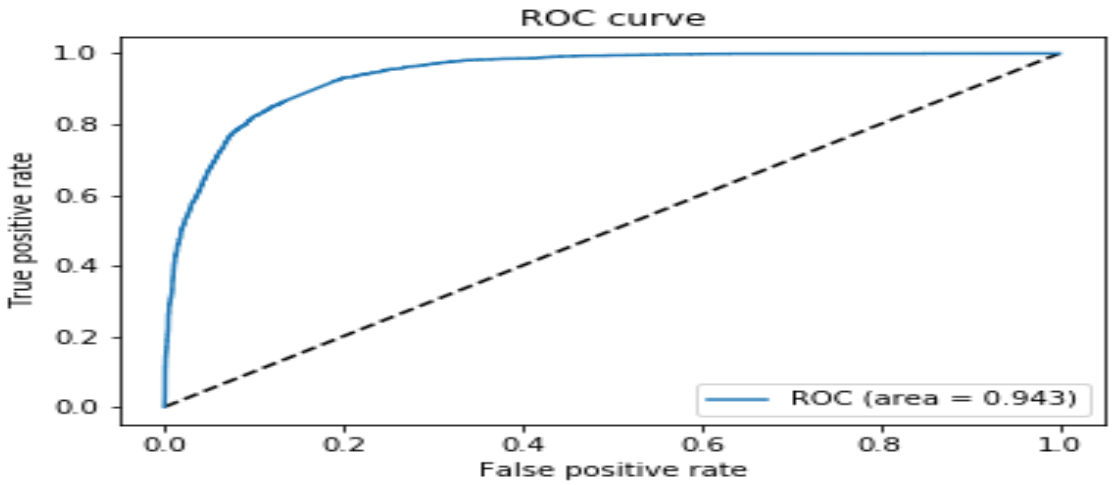

Figure 1. ROC and AUC of the CNN model

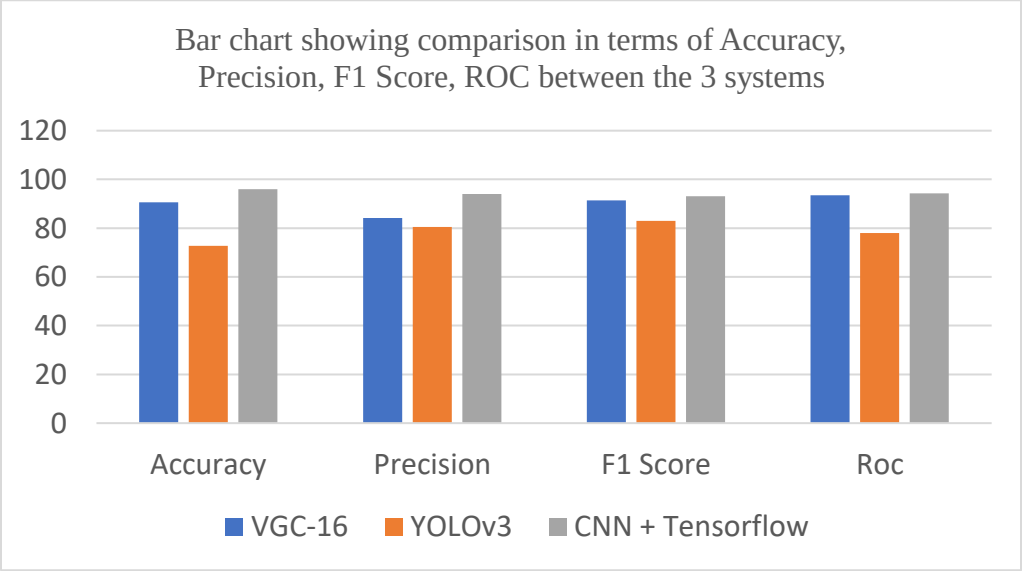

Figure 2. Bar chart showing comparison in metrics for the 3 systems

Discussion of Results

From the results obtained, which are presented in Table 1 as well as graphically in Figure 2, it can be inferred that the CNN algorithm (DRAP) predicts the best result with an accuracy of 96.00%. This shows an improvement beyond those of [11-12]. Additionally, it is known that the higher the precision and the lower the error rate, the better is the model, evidently, CNN, has the highest precision and lowest error rate of 0.94 and 0.021 respectively. Therefore, CNN + Tensorflow algorithm was used in developing the detection and recognition of security threat system because of its highest accuracy and precision score, lowest error rate value and highest area under curve (auc) value of 94.30 as shown in Figure 1.

Conclusion

The study concludes that the CNN-based model significantly improved the accuracy and efficiency of detecting and recognizing security threat in complex and challenging settings. In comparison with traditional models, findings show that the CNN model in this study demonstrates an accuracy of 96%. Thereby, making the present model a suitable alternative to traditional models, especially, when accuracy, speed and practical applicability count. Furthermore, the real-world applications of the present model have shown its potential for deployment in scenarios that demand the utmost in public safety and security. Based on these findings, it was therefore recommended among others that in the fast-paced landscape of security, it is vital to establish mechanisms for regular model retraining. This ensures the model's adaptability to evolving threats, changes in environmental conditions, and improvements in technology.

References

1. Huang, Y. (2021). *Open Area Concealed Weapon Detection (CWD) Sensor System Development* (Doctoral dissertation, University of Huddersfield).
2. Tiwari, R. K., & Verma, G. K. (2015). A computer vision based framework for visual gun detection using harris interest point detector. *Procedia Computer Science*, *54*, 703-712.
3. Tiwari, R. K., & Verma, G. K. (2015, January). A computer vision based framework for visual gun detection using SURF. In *2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)* (pp. 1-5). IEEE.
4. Brownlee, J. (2021, January 27). *object-recognition-with-deep-learning*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/object-recognition-with-deep-learning/
5. Sai, R. S., Sindhusha, R., & Sainagesh, V. (2019). OBJECT DETECTION AND IDENTIFICATION. A Project Report, 11, 22.
6. Boudjit, K., & Ramzan, N. (2022). Human detection based on deep learning YOLO-v2 for real-time UAV applications. *Journal of Experimental & Theoretical Artificial Intelligence*, *34*(3), 527-544.
7. Berg, A. (2016). *Detection and tracking in thermal infrared imagery* (Doctoral dissertation, Linköping University Electronic Press).
8. Teena, M., & Manickavasagan, A. (2014). Thermal infrared imaging. In *Imaging with Electromagnetic Spectrum: Applications in Food and Agriculture* (pp. 147-173). Berlin, Heidelberg: Springer Berlin Heidelberg.

9. Rahman, S. S. (2023). *Development and Evaluation of Low-Cost Infrared Imaging Based System for Assessing Inflammatory Activity Related to Rheumatoid Arthritis* (Doctoral dissertation, New York University Tandon School of Engineering).
10. Nilsson, D., & Fahy, R. (2016). Selecting scenarios for deterministic fire safety engineering analysis: life safety for occupants. *SFPE Handbook of Fire Protection Engineering*, 2047-2069.
11. Olmos, R., Tabik, S., & Herrera, F. (2018). Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, *275*, 66-72.
12. Velasco-Mata, A., Ruiz-Santaquiteria, J., Vallez, N., & Deniz, O. (2021). Using human pose information for handgun detection. *Neural Computing and Applications*, *33*(24), 17273-17286.