Development of a Firearms and Target Weapons Recognition and Alerting System Applying Artificial Intelligence

Jhonny Jimenez
Electronic Engineering
Universidad Politécnica Salesiana
Quito - Ecuador
jjimenez@est.ups.edu.ec

Jhonatan Toscano
Electronic Engineering
Universidad Politécnica Salesiana
Quito - Ecuador
jtoscano@ups.edu.ec

William Oñate
Electronic Engineering*)*
Universidad Politécnica Salesiana
Quito - Ecuador
wonate@ups.edu.ec

Gustavo Caiza
Electronic Engineering
Universidad Politécnica Salesiana
Quito - Ecuador
gcaiza@ups.edu,.ec

***Abstract*—Nowadays, surveillance systems that make use of security cameras are indispensable to ensure the protection and security of companies and organizational entities. These systems operate through monitoring by trained individuals. The progress of a system that uses artificial intelligence to identify and recognize firearms and knives is based on the implementation of machine learning techniques and real-time image and video analysis. The main goal is to increase public safety by accurately and quickly detecting the presence of weapons in various environments. The purpose of this research is to improve public safety through the early detection of threats and more effective responses by security forces. To achieve this, convolutional neural networks have been used. During the development of the system, a database has been created using images and videos containing firearms or knives, based on the "You Only Look Once" (YOLO) algorithm, particularly YOLOv5s.**

Keywords—Deep learning, convolutional neural networks, algorithm, YOLO.

I. Introduction

In modern times, surveillance using Closed Circuit Television (CCTV) video surveillance systems is widely used to avoid dangerous situations, such as robberies involving weapons. However, it has several notable disadvantages, such as a high price, the complexity of its maintenance, and the need for operators to constantly monitor the camera recordings [1]; after an uninterrupted 12-minute period of constant monitoring, a study [2] in "Security Oz Magazine" found that an operator is likely to lose up to 45% of screen activity and failure rate after 22 minutes of monitoring. increases to 95% since inspecting images with human intervention would be labor-intensive as well as being a slow process [3]. This can lead to the waste of vulnerable situations due to fatigue, lack of attention, or the omission of crucial information in surveillance recordings [4] Some of these studies found no impact after the installation of CCTV, while others revealed a significant decrease in the incidence of violent crime [5]; Object classification is the first step in object detection, followed by identification of weapons in images or videos [6]. The focus is to provide security to citizens and use this vital tool for crime prevention and citizen protection, proposing to help reduce the occurrence of violent crimes and provide a video surveillance solution, which will go to a new level, reducing statistics in armed crimes.

Artificial intelligence has challenges that have been developed with automatic solutions focused on the detection and recognition of dangerous elements of video surveillance, with the advantage of precision, consistency, and speed [7]. Technology is clinging to every step of our daily lives, from our phones [8]; to the computer in the identification and detection of objects, for which artificial intelligence has been rede-fined to society in the detection of objects in the world and direct time that tries to improve the work capacity of the comparison of correct and incorrect annotations resulting better than human surveillance. On the other hand [9] presented a model that works in different steps, the initial step is the acquisition of images and the detection of motion, and the last one is the detection of weapons, they carried out experiments in three different models, the first is the model based on Convolutional Neural Networks (CNN), the second is the model based on Fast R, the third is the model based on MobileNet, to analyze its behavior and trends (YOLO) these algorithms being the basis for object detection because of the rapid progress in computer vision skills [10][11][12] as well as having two completely connected layers at the end.

YOLO is like CNN in that it uses convolutional layers and max-pooling layers [13]. It is considered the best and fastest algorithm currently available globally, it is preferred by most AI researchers and enthusiasts thanks to its diversity and faster inference time [14]. The algorithm only analyzes images once, requiring forward programming [15]. From the above, this article aims to delve deeper into the uses of artificial intelligence for weapons recognition systems at the same time, with characteristics that present real-time data and implement machine learning algorithms, to contribute to improving the security of society

II. Methodology

To guarantee the safety of citizens, it is essential to conceive a design for the communication structure that we will use in the creation of a detection system. This process requires the development of a large database for model training, as is done by virtually all systems that use machine learning as an analytical approach. The procedure begins by obtaining access to Firebase [16] since it is necessary to have a database hosted in the cloud that stores and synchronizes data in real time.

A complete set of data has been prepared for training the weapons detection model, which was obtained from different sources on the web, to provide training information. All this was carried out following the MakeSense protocol, as it offers an extremely intuitive graphical interface [17]. A representation of the image selection process with appropriate quality must be provided. After analyzing the images, it was determined that their quality is ideal for training the network, taking into consideration aspects such as focus, resolution, and luminosity, which allows a more detailed visualization of all the characteristics present in the image.

A. Database

This choice was chosen to provide training examples to the neural network that resemble the situations that can be found during its use, in which the firearm or knife is displayed in complex environments with the presence of numerous nearby objects. The original database consists of 10,354 images and is divided into two classes (firearms and edged weapons), each with approximately 50% of the samples.

First Class – Firearms. For the first class, images are given, which represent illustrations of people participating in shooting exercises, showing common postures of individuals carrying a firearm. These postures turned out to be similar to those found in real robbery situations. Including these images in an environment different from that of a real robbery allows the system to obtain more generalized data, avoiding over-training with the same data.

Second Class – Edged Weapons. The second category of the database comprises images of individuals carrying knives, collected from different web places where these weapons were involved in robbery or were simply found in various positions and contexts. Using the suggested approach, a set of 10,354 images was generated, and from these, the corresponding "tags" were extracted. In these tags, the detection target is a small region of the image in a complex environment, specifically when the firearm or knife is adjacent to people and not in other areas of the image.

Through the suggested alterations, it was possible to reduce images of large proportions to smaller images, preserving the crucial information, which in this context is people carrying firearms. A total of 5,179 images were then supplied to the firearms detection model, while 5,176 images were provided to the knife detection model, depending on the classes that needed to be identified. In both instances, the training process encompassed 70% of the data, while the testing comprised the remaining 30%.

B. Training.

Encryption, replay, and training of the element and image detection model were performed in the Google Colab interactive environment. The programming also included the YOLO v5s algorithm. With the help of these tools, it was easy to tune the model parameters to previously labeled images and store a substantial data set. The YOLO v5s network was then trained for convergence through Images, Obtaining Label iterations of the additional Class A: Firearms and Class B: White Weapons classes using 70% of the images. The model was then evaluated using the remaining 30% of the images using the Python text editor, the programming we relied on for training is shown in Unified Modeling Language (UML) diagrams in Figure 1. On the other hand, retraining of the neural model was used to achieve convergence if the accuracy was between 80% and 81%. Ultimately, if the result was positive, the algorithm was corroborated by the analysis carried out by the video surveillance system.

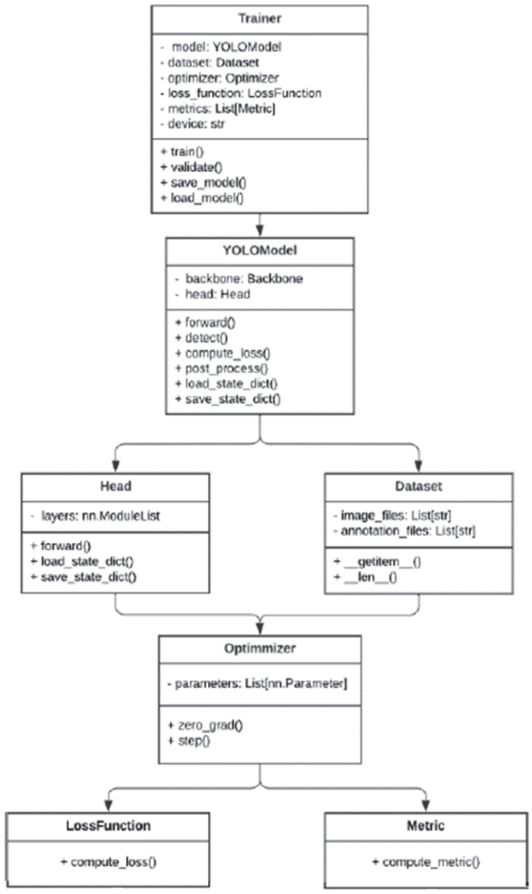

Fig. 1. Basic structure for YOLO training on UML diagrams.

C. Configuration

The system was expanded using Raspberry Pi 4 [18], specific parameters serve as a solid foundation for system operation. The 5 Mpx camera (Raspberry Pi 4) was used for trial and error and in the implementation the 12 Mpx webcam was chosen and at the same time the Raspbian OS system because it works with the hardware requirements of the Raspberry Pi 4 model. Real-time imaging system camera integrated with, Firebase Cloud, Android App and YOLOv5s with its two class types, Class A Class B, is a crucial part, and the detection model processes the information of that image to determine if a weapon may or may not be present.

D. Firebase Integration

To manage and store data in applications, Firebase offers several functions and services when used as a cloud solution. These logs contain information about detection events and their associated details, are stored in real-time, and are kept accessible for later analysis and consultation. The architectural representation in Figure 2 demonstrates how its use in the system allowed these functionalities to be leveraged and ensure an effective approach to the management of data generated by weapons detection.

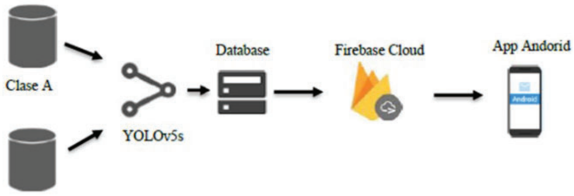

Fig. 2. Integration with Firebase.

To facilitate smooth and effective interaction between the system and a connected mobile application such as Android Studio, Firebase also takes care of communication with the mobile application, offering a collection of real-time communication tools. By ensuring that only authorized users can access the data generated by the system, the authentication and security mechanisms of your platform add an extra layer of protection to the information we store.

III. RESULTS

The checks carried out during the training process of the neural network using a dimension or batch size of 940 images with a validation accuracy of 0.45% for knives and 0.771% for firearms, the image size is set to 640 having the precision of a total of classes of 0.61% in mAP (Mean Average Precision) with 0.5% being 98%, for YOLOv5s it can incredibly detect objects with a speed of 140 fps (frames per seconds), which is 180% faster for obtaining training loss and accuracy. You get the Google Colab configuration with "batch" 16 and "epoch" 75.

For the execution of tests, the YOLOv5s architecture mentioned in section 2 was used, it was verified through the train.py file that the model correctly selected the weapons to be detected shown in Figure 3, from these we proceed with the analysis and optimal system performance.

Fig. 3. Weapon selection using Google Colab training.

The results obtained in the precision-recovery curve graphs observed in Figure 4 are detailed. This identifies the precision values of firearms with 0.681%, and knives with 0.699%. With these values, we can evaluate the performance of the model in object detection which is an accepted value for the performance of the implemented system.

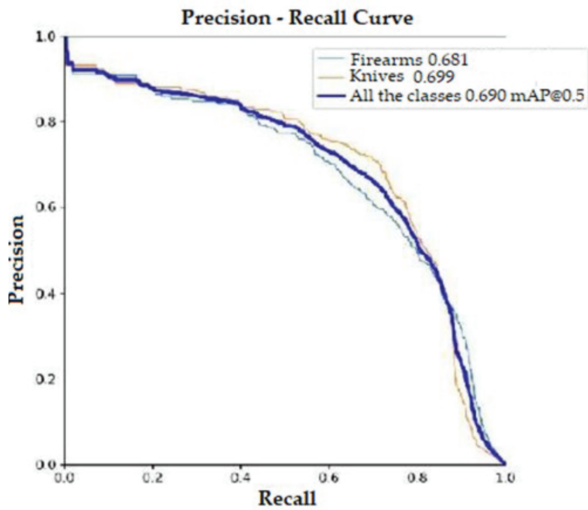

Fig. 4. Precision-recall curve.

The performance of the YOLOv5s model is measured in terms of Mean Average Precision (mAP). The higher the mAP value, the more accurate the model is. The model gives 0.95% as the Average Precision. See Figure 5.

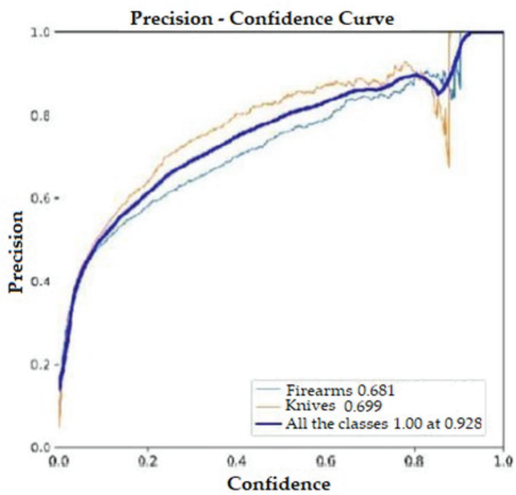

Fig. 5. Precision-confidence curve.

The value of precision and recall varies as the number of epochs increases. In epoch 75, the precision reached 0.98% and the recall was 0.87%, showing a good sign of performance. See Figure 6.

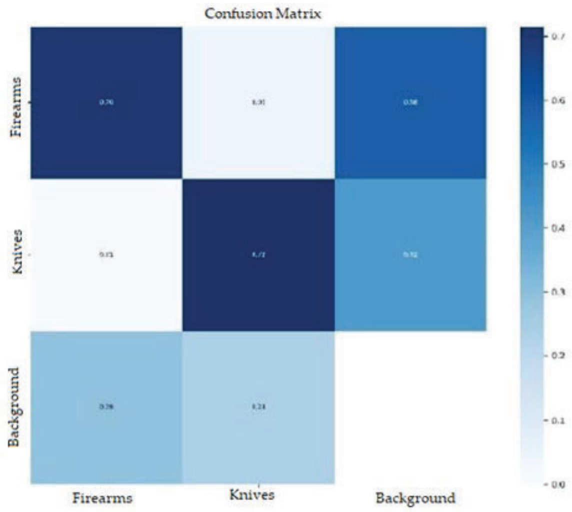

Fig. 6. Confusion matrix.

According to the results obtained from the training of the model, the levels of pressure and loss can be distinguished in the analysis of the images since these are on par in the precision curves, and do not present much difference since they remain with stable values in the confusion matrix; The improved susceptibility of the model which was obtained from the training is noted. With the values obtained in the precision curves, the values are optimal for the correct functioning of the detection system since it does not present drawbacks in its use. At the time of detecting the object, the image is saved and subsequently sent to Firebase, the database that was used in this stores the image in Storage and turns it into the real-time database, in the latter it is saved in a link with the date, and time of the entry of the image, which is displayed in the application.

When considering the times of saving, sending, and displaying the image of the detected object, we can consider that they are optimal for a detection system since these occur almost instantly and do not exceed 4 seconds in detecting and sending the information to the user. With the data obtained, the results in the detection of weapons and knives will be displayed in the application, which is indicated in Figure 7, and in turn, the interface is indicated, which consists of two buttons, the first shows the detections and the second button calls emergencies. By clicking the "Show detection" button, the detections achieved are indicated and the date and time of these in their different classes are detailed.

Fig. 7. Recognition of edged weapons and firearms.

IV. CONCLUSIONS

The multi-class object detection model was even less accurate than the single-class object detection model. Despite the infrastructure environments demonstrating the need for a technological application of great relevance in the field of security, even though some small elements can still confuse the system and cause erroneous identifications, the operation is efficient. The camera detects weapons held by certain individuals, weapons that could go unnoticed by a human being who sometimes has multiple monitors around or does not have images of sufficient quality for adequate perception. In addition to having superior observation capacity, the system also re-acts more quickly; In less than a tenth of a second, it could warn of the presence of weapons and save human lives.

The weapon detection period is 400 ms, that is, immediately, and the latency metric for sending data from detected weapons to the mobile application is 48 seconds, indicating that this is the communication time between the cloud and the mobile application. The use of "YOLO" in the identification system made it possible to provide the developed model with the parts of the image that are relevant to detecting firearms and edged weapons. This allows the system to focus exclusively on areas where weapons are found and ignore other regions of the image where firearms are not expected to be present. and knives, this helps to make the criminal activities that occur in the environment irrelevant to current common scenarios.

REFERENCES

[1] J A Hidalgo, E. (2012). Sistema cctv (circuito cerrado de televisión) entre edificios, para la seguridad y vigilancia en el aeropuerto internacional cotopaxi. Universidad Técnica de Ambato.

[2] R. K. Tiwari and G. K. Verma, "A Computer Vision based Framework for Visual Gun Detection Using Harris Interest Point Detector," Procedia., vol. 54, pp. 703– 712, 2015.

[3] Jain, Harsh y Vikram, Aditya y Mohana, Mohana y Kashyap, Ankit y Jain, Ayush. (2020). Detección de Armas usando Inteligencia Artificial y Aprendizaje Profundo para Aplica- ciones de Seguridad. 193-198. 10.1109/ICESC48915.2020.9155832.

[4] GF Shidik, E. Noersasongko, A. Nugraha, PN Andono, J. Jumanto y EJ y Kusuma, "Una revisión sistemática de la videovigilancia de

inteligencia: tendencias, técnicas, marcos y con- juntos de datos". Acceso IEEE, vol. 7,págs. 457–473, 2019.

[5] Lai, Y. L., Sheu, C. J., & Lu, Y. F. (2019). El esquema de circuito cerrado de televisión monitoreado por la policía realmente importa en la reducción del crimen, 63(1), 101-134. doi:10.1177/0306624X18780101.

[6] Redmon et al., Solo se mira una vez: Detección unificada de objetos en tiempo real. IEEE CVPR, 2016.

[7] D. Schuller y BW Schuller, "La era de la inteligencia emocional artificial", Computadora IEEE, vol.51, número 9, pp. 38-46, septiembre de 2018.

[8] S. Narejo, B. Pandey, C. Rodríguez, MR Anjum et al., "Detección de armas usando yolo v3 para sistema de vigilancia inteligente,"Problemas Matemáticos en Ingeniería, vol. 2021, 2021.

[9] Elmir, Yourssef, Sid Ahmed Laouar y Larbi Hamdaoui. "Aprendizaje profundo para la de- tección automática de pistolas en secuencias de video".JERI.2019.

[10] ST Ratnaparkhi, A. Tandasi y S. Saraswat, "Detección y reconocimiento de rostros para el sistema de identificación de delincuentes", 202111.ª (Confluence), 2021, págs. 773-777, doi :10.1109/Confluence51648.2021.9377205.

[11] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. 2012. Clasificación de ImageNet con redes neuronales convolucionales profundas. En Actas de la 25.ª Conferencia Internacional sobre Sistemas de Procesamiento de Información Neural - Volumen 1 (NIPS'12). Curran Associates Inc., Red Hook, Nueva York, EE. UU., 1097–1105.

[12] GF Shidik, E. Noersasongko, A. Nugraha, PN Andono, J. Jumanto y EJ Kusuma, "Una re- visión sistemática de la videovigilancia de inteligencia: tendencias, técnicas, marcos, (CCECE) pp. 1-6, abril de 2012 y conjuntos de datos", en IEEE Access, vol. 7, págs. 170457-170473, 2019, doi: 10.1109/ACCESS.2019.2955387.

[13] R. Girshick, J. Donahue, T. Darrell y J. Malik, "Jerarquías de características ricas para la detección precisa de objetos y la segmentación semántica", IEEE, págs. 580–587, 2014.PyTorch, "YOLOv5"

[14] U. Nepal y H. Eslamiat, "Comparación de yolov3, yolov4 y yolov5 para la detección autó- noma de puntos de aterrizaje en vehículos aéreos no tripulados defectuosos" Sensores, vol. 22, núm. 2, pág. 464, 2022.

[15] Rothe, Rasmus y Guillaumin, Matthieu y Van Gool, Luc. (2015). Supresión no máxima para la detección de objetos al pasar mensajes entre ventanas. LNC. 9003.10.1007/978-3-319- 16865-4_19.

[16] Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019). Detección de objetos con aprendizaje pro- fundo: una revisión. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212-3232. doi: 10.1109/TNNLS.2018.2876865.

[17] P. Kapica, "Implementing YOLO v3 in Tensorflow (TF-Slim) – ITNEXT," 2018. [Online]. Available: https://itnext.io/implementing-yolo-v3-in- tensorflow-tf-slim-c3c55ff59dbe. [Accessed: 13-Dec-2018]

[18] G. J. Caiza and M. V. García, "Implementación de sistemas distribuidos de bajo costo bajo norma IEC-61499, en la estación de clasificación y manipulación del MPS 500," Ingenius, vol. 18, pp. 40–46, 2017, doi: 10.17163/ings.n18.2017.