DETECCIÓN DE ARMAS MEDIANTE VISIÓN ARTIFICIAL UTILIZANDO TÉCNICAS DE AUMENTACIÓN DE DATOS

Gilbert F Pérez-García[1]; Alexis de J Flores García[1]; Elías N Escobar-Gómez[1]; Jorge A Sarmiento-Torres[1]; María C Salgado-Gutiérrez[1]

[1] Tecnológico Nacional de México/I. T. de Tuxtla Gutiérrez, Tuxtla Gutiérrez, Chiapas, México. Autor responsable: gilbert.pg@tuxtla.tecnm.mx
Área: Informática

Resumen

En este estudio, se presenta el desarrollo de un sistema de detección de armas basado en visión artificial, diseñado para fortalecer la seguridad en entornos críticos, como aeropuertos, escuelas y áreas públicas. La metodología propuesta se fundamenta en el procesamiento de imágenes y el aprendizaje profundo, orientado a la identificación de armas de fuego. Para este propósito, se implementa el modelo de detección de objetos YOLO, específicamente la versión 8. El proceso de entrenamiento se lleva a cabo empleando un conjunto de datos de dominio público modificado mediante técnicas de aumentación de datos. La ejecución del modelo se realiza en un ordenador de computadora logrando con esto, una detección en tiempo real. Cuando se detecta una posible arma de fuego, el sistema genera una alarma discreta e instantánea, alertando a las autoridades responsables, lo que acelera los tiempos de respuesta y simultáneamente registra el incidente. Es importante destacar que la ética y la privacidad son consideraciones prioritarias en este proyecto, asegurando que el sistema se centre exclusivamente en la identificación de armas de fuego, sin invadir la privacidad de las personas. En resumen, este proyecto representa las bases para el desarrollo de sistemas orientados a la aplicación responsable de la inteligencia artificial para reforzar la seguridad pública, proporcionando una capa adicional de protección respaldada por precisión, rapidez y consideraciones éticas.

Palabras clave: Detección de armas, YOLOv8, visión artificial, Aprendizaje profundo, redes neuronales convolucionales.

Abstract

In this study, the development of a firearm detection system based on computer vision is presented. It is designed to enhance security in critical environments such as airports, schools, and public areas. The proposed methodology relies on image processing and deep learning, with a focus on the identification of firearms. For this purpose, the YOLO object detection model, specifically version 8, is implemented. The training process is conducted using a modified publicly available dataset through data augmentation techniques. The model's execution takes place on a computer, achieving real-time detection. When a potential firearm is detected, the system generates a discreet and instantaneous alarm, alerting the relevant authorities, thereby expediting response times and simultaneously recording the incident. It is essential to emphasize that ethics and privacy are paramount considerations in this project, ensuring that the system is solely centered on the identification of firearms without infringing upon people's privacy. In summary, this project lays the foundation for the development of systems geared toward the responsible application of artificial intelligence to reinforce public safety. It provides an additional layer of protection supported by precision, speed, and ethical considerations.

Keywords: Firearm detection, YOLOv8, computer vision, Deep learning, Convolutional neural networks.

Introducción

La visión artificial es una tecnología revolucionaria que ha transformado la forma en que interactuamos con el mundo que nos rodea. Su capacidad para procesar imágenes y extraer información valiosa de ellas ha encontrado aplicaciones en una amplia gama de campos, desde la medicina hasta la industria automotriz. En particular, la visión artificial ha incursionado en el tema de seguridad [1] como una herramienta de apoyo para el cumplimiento de la ley, desarrollando técnicas de detección para implementación en sistemas integrados. En la actualidad, la detección de armas es una de las preocupaciones más apremiantes en la población, por lo que, contar con sistemas eficientes y confiables para llevar a cabo la detección de armas peligrosas es esencial. Los sistemas tradicionales de detección de armas a menudo dependen de la revisión manual de imágenes o la implementación de escáneres de rayos X, lo que ralentiza el proceso, aumentando la posibilidad de errores y, en muchos casos, infringe la privacidad de las personas. En este sentido, la detección de armas mediante visión artificial ofrece ventajas significativas. Su enfoque no implica la revisión física, lo que preserva la privacidad de las personas. Esto reduce la resistencia a la implementación de estos sistemas, permitiendo una revisión más transparente y sin causar molestias a quienes están siendo evaluados.

Por otro lado, la visión artificial tiene la capacidad de analizar imágenes de video en tiempo real y detectar armas de manera automática y precisa. Esto se debe a su capacidad para procesar grandes cantidades de datos en fracciones de segundo, lo que permite a estos sistemas identificar posibles amenazas de manera más rápida. Lo que destaca especialmente es su capacidad para procesar información en tiempo real y enviar alertas a las autoridades casi al instante, lo que representa una ventaja clave de la visión artificial en la detección de armas. Los sistemas tradicionales de detección de armas a menudo generan falsas alarmas o dependen del juicio de los operadores, lo que resulta en falsos positivos y molestias para las personas inocentes. En este sentido, la visión artificial mejora la eficiencia y reduce las interrupciones no deseadas en el proceso de detección de armas. La visión artificial se caracteriza por su alta precisión, y a través de su proceso de entrenamiento, es posible ajustar varios parámetros configurables para adaptar el modelo a las necesidades específicas de una aplicación o del entorno en el que se utilizará el sistema. Esto conlleva a una reducción significativa de los falsos positivos, lo que a su vez resulta en una mejora en la eficacia y una disminución de las interrupciones no deseadas.

En esta investigación, se describe el desarrollo de un sistema inteligente de detección de armas basado en visión artificial y aprendizaje profundo, diseñado para la identificación de armas de fuego en entornos cerrados con iluminación controlada. El propósito principal de este sistema es fortalecer la seguridad en entornos críticos y restringidos, ofreciendo una capa adicional de protección respaldada por la precisión y la velocidad. La implementación se ha realizado utilizando un algoritmo de detección de objetos conocido como YOLO en su versión más reciente, YOLOv8 [2]. Este algoritmo se ha destacado por su capacidad para realizar identificaciones precisas en tiempo real, lo que lo convierte en una herramienta valiosa para la detección de armas. El proceso de entrenamiento se ha llevado a cabo utilizando un conjunto de datos público que se encuentra alojado en los servidores de Roboflow y ha sido compartido por la Universidad de Granada.

Las principales contribuciones de esta investigación son las siguientes:

- Se ha fortalecido el modelo entrenado mediante técnicas de aumentación de datos, lo que ha resultado en un conjunto de datos modificado con 8913 imágenes.

- El desarrollo del modelo entrenado se ha presentado con una precisión de 0.97239, una sensibilidad de 0.96477 y un equilibrio del sistema medido a través de la métrica mAP de 0.98767 con un factor de confianza de 0.5 y 0.86446 con un factor de confianza de 0.95.

El uso de técnicas de aumento de datos en el ámbito de la visión artificial es fundamental para fortalecer los conjuntos de datos utilizados en el entrenamiento de modelos de predicción. Estas técnicas ofrecen diversas formas de enriquecer el conjunto de datos, lo que resulta en un aprendizaje más efectivo. Por ejemplo, en algunos estudios [3,4], se utiliza la minería de datos para generar muestras relacionadas con actividades delictivas. Esto se logra mediante algoritmos que buscan patrones específicos en grandes cantidades de datos. Además, para adaptar los modelos de detección a diferentes contextos, como la elección de detectores de objetos específicos, se utilizan conjuntos de datos variados. Por ejemplo, en [5], los autores utilizan la red Inception v3 como base y fortalecen su conjunto de datos a través de técnicas de aumento de datos. Lo que es interesante es que afirman que su metodología de aumento es dinámica y no se basa en replicar el mismo archivo una y otra vez. Por otro lado, algunos investigadores optan por la generación de datos sintéticos en lugar de simplemente aumentar los datos existentes, como se menciona en [6]. Esta estrategia proporciona información diversa que enriquece aún más el aprendizaje de los modelos de detección.

En resumen, el uso de técnicas para ampliar los datos desempeña un papel muy importante al ofrecer diversas perspectivas de las imágenes originales presentes en un conjunto de datos. Esta combinación de aumento de datos y visión artificial ha traído una revolución en la detección de armas, proporcionando muchas ventajas. Su capacidad para procesar grandes cantidades de información visual en tiempo real, adaptarse a diferentes entornos y funcionar en diversas condiciones de iluminación la convierte en una herramienta esencial para la seguridad y la aplicación de la ley. Además, su capacidad para integrarse con otros sistemas de seguridad, reducir las alarmas falsas y respetar la privacidad de las personas la convierte en una opción factible para aplicaciones de detección de armas.

Objetivos

El objetivo de este trabajo es realizar un sistema de visión artificial que tenga la capacidad de detectar armas de fuego, específicamente de cañón corto en una escena bajo vigilancia.

Materiales y métodos

En esta sección, se detallan los materiales y la metodología utilizados en este estudio. Nuestro objetivo principal es desarrollar un sistema físico capaz de detectar armas de fuego en entornos cerrados, ya sea a través de imágenes, videos o en tiempo real mediante cámaras de videovigilancia. Para lograr la identificación de las armas, empleamos un modelo previamente entrenado utilizando el conjunto de datos MS COCO (que significa Microsoft Common Objects in Context), que contiene alrededor de 91 tipos de objetos diferentes con un total de 2.5 millones de instancias etiquetadas en 328,000 imágenes. La ventaja de utilizar un modelo preentrenado radica en la capacidad de aprovechar el conocimiento previamente adquirido por la red neuronal.

Además, durante el proceso, utilizamos el algoritmo YOLO (You Only Look Once) (ver Figura *1*) para analizar la escena. Este algoritmo se destaca por su capacidad para emplear capas neuronales convolucionales al final de la red, evitando la necesidad de transformarlas en una red convencional. Gracias a esta característica, el algoritmo es capaz de analizar la escena a una velocidad notable de 60 fotogramas por segundo (FPS), incluso en sistemas informáticos convencionales (CPUs).

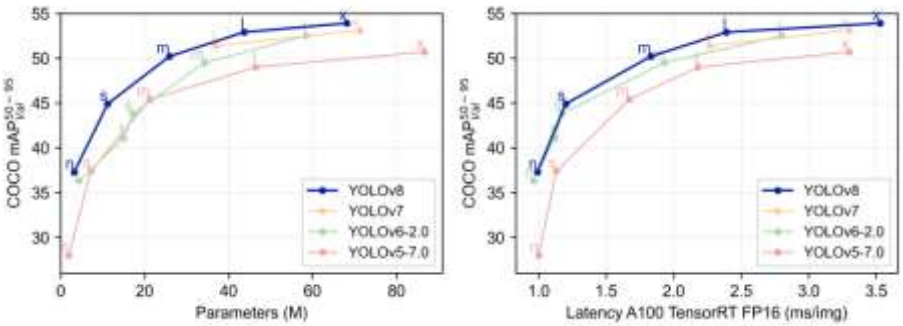

Figura 1. Comparativa de YOLOv8 con otros modelos de YOLO

Para la adquisición de imágenes, se ha empleado una cámara de video IR wyze cam v3, que dispone de una resolución de 1080 píxeles, características de visión nocturna, con un campo de visión de 130°. Para las imágenes se capturan a una resolución de 640 píxeles con una velocidad de grabación de 30 fotogramas por segundo.

Conjunto de datos

Para llevar a cabo la transferencia de conocimiento en la identificación de armas de fuego, utilizamos un conjunto de datos de acceso público que se encuentra alojado en el servidor de Roboflow. Este conjunto de datos consta de 2986 imágenes y 3448 etiquetas que se refieren a una única categoría de anotación: pistolas. Las imágenes muestran una variedad de contextos, como personas sosteniendo pistolas, representaciones de dibujos animados y representaciones de alta calidad en entornos de estudio relacionados con armas de fuego. Este conjunto de datos se originó en un proyecto de la Universidad de Granada y fue sometido a una depuración que incluyó la eliminación de duplicados antes de ser nuevamente alojado en Roboflow por un colaborador.

Preprocesamiento del conjunto de datos

El conjunto de datos proporcionado por Roboflow consta de 2986 imágenes, cada una con una resolución de 415 píxeles. Utilizar técnicas de aumento de datos es esencial para mejorar el entrenamiento del modelo. Estas técnicas generan nuevas versiones de las imágenes, lo que amplía la variedad de patrones presentes en ellas. Esto, a su vez, permite que la red neuronal pueda aprender de manera más efectiva y generalizar mejor. La Figura 2 muestra cómo se ven las imágenes después de aplicar estas transformaciones.

Para llevar a cabo este proceso, se realizaron cuatro tipos diferentes de transformaciones en las imágenes del conjunto de datos.

Donde:

x : denota el eje horizontal

y : denota el eje vertical

θ : denota el ángulo de rotación

La primera transformación es la reflexión horizontal (Ec. 1), que implica una inversión de la imagen alrededor de un eje vertical imaginario, generando una imagen reflejada horizontalmente.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(Ec. 1)}$$

La segunda transformación es la reflexión vertical (Ec. 2), en la que utilizamos un eje horizontal imaginario y aplicamos una transformación geométrica que invierte la imagen alrededor de este eje, generando una imagen reflejada verticalmente.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(Ec. 2)}$$

La tercera transformación es una reflexión combinada o reflexión en el origen (Ec. 3), en la que aplicamos la transformación geométrica para invertir la imagen alrededor de ambos ejes imaginarios, es decir, el eje horizontal y el eje vertical. Esto crea una imagen que está reflejada tanto horizontal como verticalmente de manera simultánea.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{(Ec. 3)}$$

En la cuarta transformación se realiza la rotación de la imagen sobre sus coordenadas centrales (Ec. 4).

$$I(\theta) = \begin{bmatrix} \cos\theta & -sin(\theta) \\ sin(\theta) & \cos(\theta) \end{bmatrix} \quad \text{(4)}$$

La combinación de estas transformaciones aumentó significativamente el tamaño del conjunto de datos, expandiéndolo hasta tres veces su tamaño original, lo que resultó en un total de 8958 imágenes. La inclusión de datos adicionales en el proceso de entrenamiento contribuyó a evitar el sobreajuste que puede ocurrir cuando se entrena un modelo con una cantidad limitada de datos. De estas imágenes adicionales, el 80 %, es decir, 7166 imágenes, se utilizaron como datos de entrenamiento para el algoritmo de detección, mientras que el 20 % restante, es decir, 1792 imágenes, se reservaron como datos de validación.

Entrenamiento

En el proceso de entrenamiento, se establecieron varios parámetros cruciales. Se fijaron 100 épocas de entrenamiento (epochs = 100), se definió un tamaño de lote que especifica cuántas imágenes se procesan simultáneamente (batch = 10), y se determinaron las dimensiones de reescalamiento de las imágenes (image_size = 640). El entrenamiento se llevó a cabo en Google Colab, aprovechando el GPU de entrenamiento y utilizando Ultralytics YOLOv8.0.20. Se configuró un entorno con Python-3.10.12, torch-2.1.0+cu118 y CUDA:0 (Tesla T4, 15102MiB). El sistema de cómputo estaba equipado con 2 CPU con 12.7 GB de RAM y disponía de un espacio en disco de 27.1/78.2 GB.

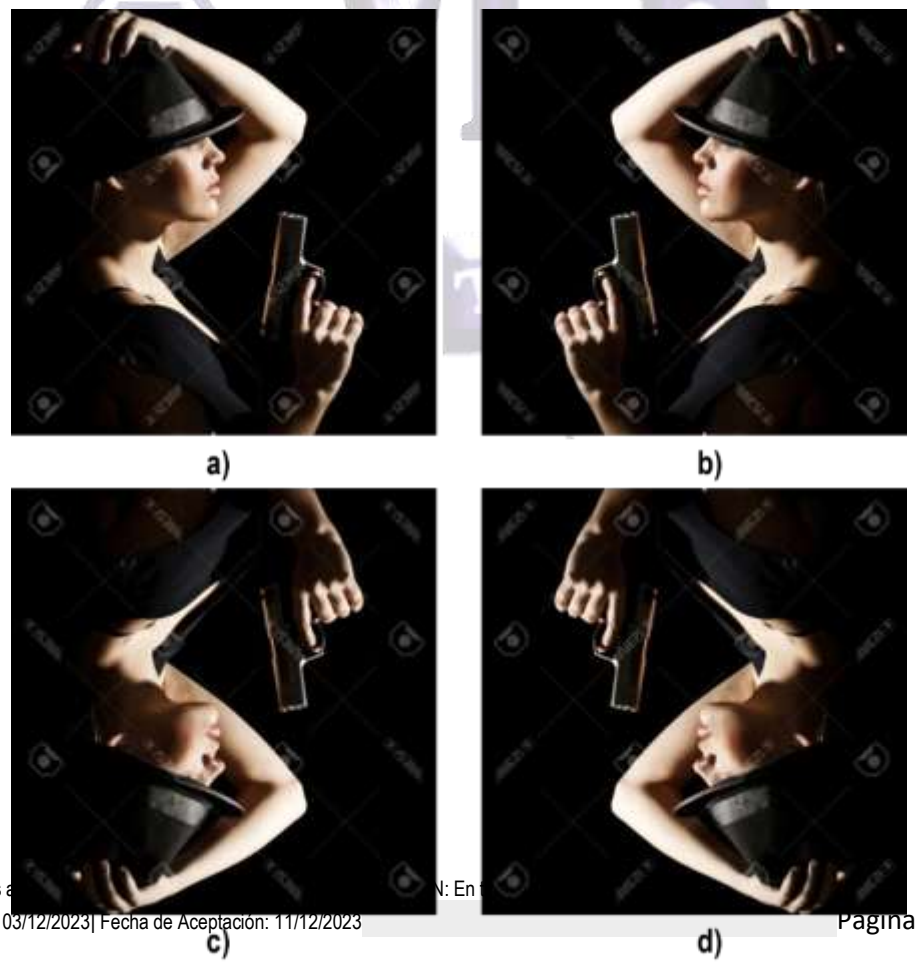

Figura 2. Ejemplo de imagen del conjunto de datos. a) Imagen original, b) Imagen con inversión horizontal, c) Imagen con inversión vertical, d) Imagen con inversión horizontal y vertical.

Resultados y discusión.

En este estudio, se ha desarrollado un modelo de detección de armas basado en visión artificial. A continuación, se exponen los resultados obtenidos después de 100 épocas de entrenamiento y se presentan los resultados de la validación del sistema mediante diferentes escenarios de prueba.

Modelo Entrenado

El sistema fue entrenado utilizando un conjunto de datos previamente procesado mediante técnicas de aumento de datos, generando un conjunto de 8958 imágenes en total. Para evaluar el rendimiento del modelo resultante, se presentan en la Tabla 1 las métricas empleadas en su evaluación.

Una métrica ampliamente reconocida para evaluar la calidad de las predicciones en modelos de detección de objetos es la puntuación media de precisión (mAP). Los resultados del entrenamiento presentan dos variantes de esta métrica: la primera, que considera un factor de confianza igual o superior a 0.5 (mAP_50), alcanza un valor de 0.98767, mientras que la segunda, aplicada con un factor de confianza igual o superior a 0.95 (mAP_95), muestra un valor de 0.86446.

Tabla 1. Métricas resultantes para la evaluación del modelo re-entrenado

Métrica	Valor
Factor de confianza	0.478
Factor F1	0.97
Precisión	0.97239
Sensibilidad	0.96477
mAP @ 0.5	0.98767
mAP @ 0.95	0.86446

Cada predicción se ajusta utilizando un umbral de operación, conocido como el factor de confianza. El éxito en la predicción de resultados positivos está estrechamente relacionado con la elección de este umbral. La métrica F1 se emplea para determinar este valor, ya que combina tanto la precisión como la sensibilidad del sistema en un solo indicador, proporcionando así una evaluación completa y equilibrada de su desempeño. Aunque esta métrica se utiliza principalmente en situaciones de desequilibrio en los datos, suele recomendarse como punto de partida para establecer el nivel de detección en la mayoría de los procesos de entrenamiento.

En el contexto de nuestro estudio, alcanzamos un valor de F1 de 0.97 al utilizar un factor de confianza de 0.478. Este valor representa el punto de equilibrio óptimo, garantizando la máxima cantidad de predicciones positivas con precisión sin sacrificar la sensibilidad del sistema (consulte la Fig. 3).

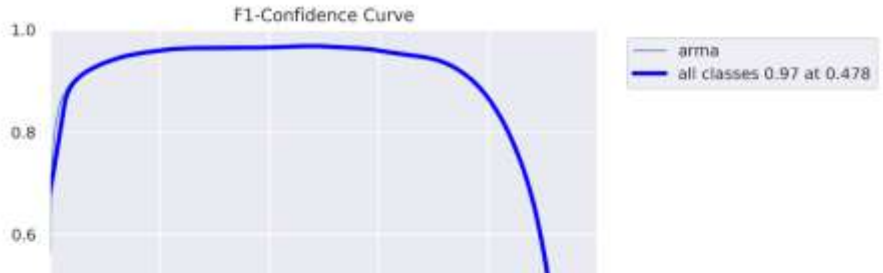

Figura 3. Curva característica de la métrica F1

En la Fig. 4 se muestran las curvas características de precisión y sensibilidad, dos métricas fundamentales que se complementan mutuamente. La precisión evalúa la proporción de predicciones correctas (verdaderos positivos) en relación con la suma de predicciones correctas y predicciones incorrectas (falsos positivos). Mientras que, la sensibilidad, también conocida como Recall, se enfoca en la identificación de los falsos negativos en lugar de los falsos positivos. Estas curvas permiten un análisis independiente de ambas métricas y ofrecen una comprensión más completa del rendimiento del sistema.

En la Fig. 4-b, se aprecia que a medida que aumentamos el nivel de confianza, la sensibilidad del modelo disminuye. Aunque esta disminución conlleva a una mayor precisión del sistema, también implica que el modelo se vuelve más cauteloso y menos capaz de realizar predicciones en situaciones inciertas. Esto se traduce en que el modelo se vuelve más estricto, ignorando predicciones que no cumplen con un alto estándar de confianza, tal como se ilustra en la Fig. 4-a.

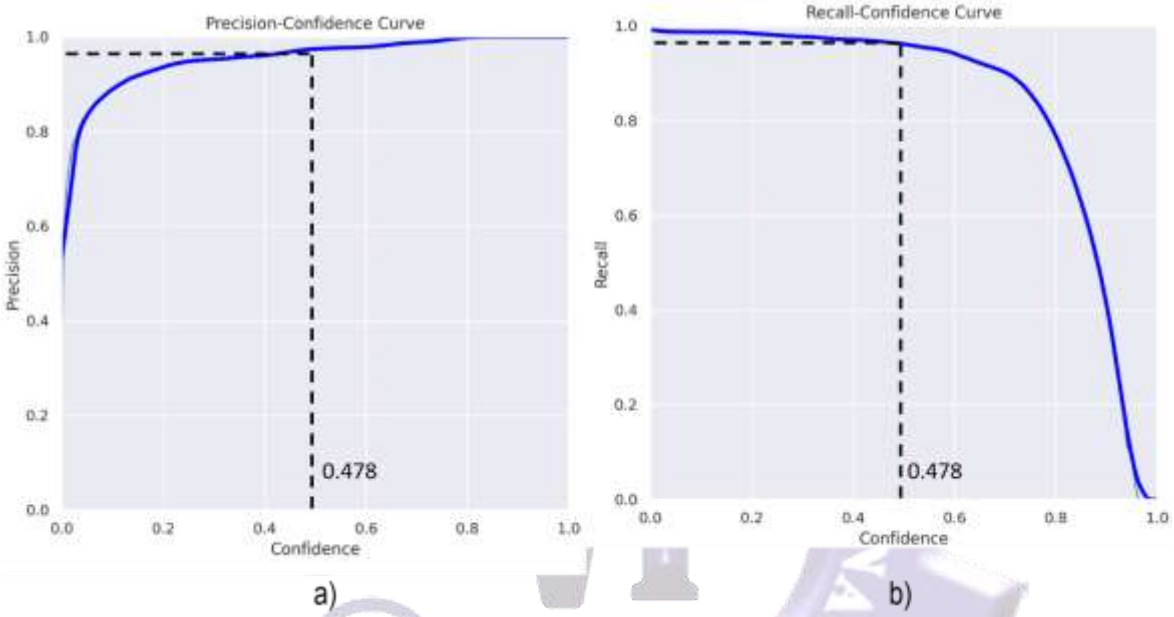

Figura 4. Curva características de a) Precisión b) Sensibilidad

De acuerdo con la métrica F1, se eligió un factor de confianza de 0.478. Al observar este umbral en las gráficas de precisión (Fig. 4-a) y sensibilidad (Fig. 4-b), se nota que, en ambos casos, tanto la precisión como la sensibilidad coinciden, logrando un equilibrio que se considera una probabilidad aceptable. Por lo tanto, este umbral puede considerarse óptimo. Sin embargo, es importante destacar que, según el contexto de implementación, este umbral puede ser ajustado para adaptarse a las condiciones específicas del entorno en el que se realizará la detección.

Validación

En esta sección, se procedió a evaluar el rendimiento del modelo propuesto, el cual ha sido implementado en la plataforma Python. El conjunto de datos utilizado previamente fue sometido a técnicas de aumento de datos como parte del proceso. En la Fig. 5, se presentan ejemplos de imágenes de entrada junto con las detecciones correspondientes. Estas imágenes fueron adquiridas de fuentes en línea y representan situaciones delictivas reales. A pesar de que el modelo fue originalmente concebido para operar en entornos cerrados como bancos, escuelas y empresas, donde la iluminación tiende a ser uniforme, las variaciones en la intensidad de la luz en las escenas de la Fig. 5 resaltan la robustez del modelo. Esta robustez se logró gracias a la selección del conjunto de datos, al preprocesamiento de imágenes y a la configuración de los parámetros durante el proceso de entrenamiento.

Figura 5. Escenas delictivas a) Entrada de la imagen b) Salida de la imagen con caja de predicción

Conclusiones.

En este estudio, se planteó la creación de un modelo de visión artificial diseñado para identificar armas de fuego en situaciones cotidianas. Para lograr esto, se utilizó un conjunto de datos que se sometió a un proceso de mejora mediante técnicas de procesamiento de imágenes, incluyendo cambios horizontales, verticales y con respecto al origen. El modelo se entrenó utilizando aprendizaje profundo con este conjunto de datos mejorado, con un total de 8958 imágenes. Las métricas clave que se utilizaron para evaluar el rendimiento del modelo incluyen un factor de confianza de 0.478 y un factor F1 de 0.97. Estos valores destacan la capacidad del modelo para equilibrar eficazmente la precisión de sus predicciones y su sensibilidad. Además, las métricas de puntuación media de precisión (mAP) respaldan sólidamente la eficacia del modelo, lo que confirma su rendimiento. La flexibilidad para ajustar el factor de confianza según las condiciones específicas del entorno de detección subraya la versatilidad del sistema en diferentes situaciones prácticas.

En términos de futuras investigaciones, este estudio establece una base sólida para explorar más a fondo la detección de objetos. Se abre la puerta a la adaptación del sistema a aplicaciones específicas y al desarrollo de prototipos en tiempo real en entornos cambiantes. Esta evolución tiene el potencial de hacer contribuciones valiosas en términos de seguridad.

Referencias bibliográficas

[1] Oñate Miranda, F. P. (2020). *Diseño y construcción de nodos inteligentes para detección de armas dentro de una red de video-vigilancia utilizando visión artificial* (Bachelor's thesis, Escuela Superior Politécnica de Chimborazo).

[2] Terven, J., & Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv preprint arXiv:2304.00501*.

[3] Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., & García-Martínez, R. (2008). Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. In JIISIC (pp. 31-40).

[4] Perversi, I., Valenga, F., Fernández, E., Britos, P. V., & García Martínez, R. (2007). Identificación y detección de patrones delictivos basada en minería de datos. In *IX Workshop de Investigadores en Ciencias de la Computación*.

[5] Sánchez, J., & Campos, M. A. (2021). Red neuronal artificial para detección de armas de fuego y armas blancas en video vigilancia. Revista de Iniciación Científica, 7(2), 83-88.

[6] Vallez, N., Velasco Mata, A., Cotorro, J. J., & Deniz, O. (2019). ¿Es posible entrenar modelos de aprendizaje profundo con datos sintéticos? In *XL Jornadas de Automática* (pp. 859-865). Universidade da Coruña, Servizo de Publicacións.