Detección de armas de fuego en imágenes digitales mediante redes neuronales convolucionales

Tomas López Pérez,1Jose Edgar Lara Ramírez ,2 María de Jesús Estudillo Ayala3,\*

1Facultad de Sistemas Biológicos e Innovación tecnológica, Universidad Autónoma “Benito Juárez” de Oaxaca (FASBIT-UABJO), Av. Universidad S/N, Ex-Hacienda 5 Señores, C.P. 68120, Oaxaca de Juárez, México

2Departamento de Electrónica, División de Ingenierías, Universidad de Guanajuato, Carretera Salamanca-Valle de Santiago km 3.5 + 1.8, Comunidad de Palo Blanco, Salamanca, Gto., C.P. 36885, México

[\*104231.fasbit @uabjo.com](mailto:*104231.fasbit%20@uabjo.com) , [\* [lramirez.cat@uabjo.mx](mailto:lramirez.cat@uabjo.mx)](mailto:*auto.dos@gmail.com), [\* mestudillo.cat@uabjo.mx](mailto:*autor.tres@gmail.com)

Recibido XX mesXXXX; revisado XXmes, XXXX; aceptado XX mes XXXX; publicadoXX mes XXXX

**RESUMEN**

**La inseguridad en México es una preocupación nacional e internacional, especialmente por los frecuentes asaltos en la calle y transporte público, a menudo con armas de fuego. Este trabajo propone una metodología innovadora para detectar armas de fuego mediante un modelo neuronal de aprendizaje profundo, utilizando una Red Neuronal Convolucional. La RNC destaca por su rapidez y eficiencia, para el desarrollo de esta metodología se usaron recursos en la nube para construir la base de conocimiento para la implementación y entrenar el modelo. Los resultados experimentales muestran buen desempeño en la detección de armas en imágenes digitales, siendo esta propuesta una vía para mejorar la seguridad pública.**

**ABSTRACT**

Insecurity in Mexico is a national and international concern, especially due to frequent assaults on the street and public transport, often with firearms. This work proposes an innovative methodology to detect firearms through a deep learning neural model, using a Convolutional Neural Network. The RNC stands out for its speed and efficiency. For the development of this methodology, cloud resources were used to build the knowledge base for the implementation and train the model. The experimental results show good performance in detecting weapons in digital images, this proposal being a way to improve public safety.

Palabras clave: Deep Learning, Redes Neuronales Convolucionales, Tecnologías para la seguridad.

1. **INTRODUCCIÓN**

La detección de armas de fuego en imágenes digitales es un campo de estudio en el ámbito de la visión por computadora y el procesamiento de imágenes. A lo largo de los años, se han desarrollado diferentes técnicas y enfoques para abordar este problema, tal como se hace en el modelo de redes neuronales convolucionales, que es un enfoque popular y efectivo para la detección de objetos en imágenes. Este enfoque se caracteriza por ser capaz de realizar la detección y clasificación de objetos en tiempo real, ya que realiza la detección y clasificación de objetos en una sola pasada de la red neuronal convolucional. Es importante tener en cuenta que la detección de armas de fuego en imágenes digitales es un desafío complejo y puede haber limitaciones en la precisión y confiabilidad de los métodos actuales. Además, es necesario contar con un conjunto de datos de entrenamiento diverso y representativo que incluya una variedad de imágenes de armas de fuego en diferentes escenarios para obtener resultados más precisos y generalizables, sin dejar de lado las consideraciones éticas y legales asociadas con la detección de armas de fuego en imágenes digitales, como la privacidad y el uso adecuado de los datos recopilados.

1. **Desarrollo**
2. *Fundamento Teórico*
3. *Redes neuronales convolucionales*

Las CNN son un tipo especial de redes neuronales inspiradas en los sistemas biológicos. El trabajo desarrollado por Huble y Wiessel [1], en el que al trabajar con la corteza visual de los gatos descubrieron que una porción específica del campo visual es la responsable de excitar ciertas neuronas en particular, estableció el principio usado para diseñar una arquitectura esparcida usando redes neuronales artificiales.

Las CNN se especializan en procesar información que está estructurada en forma de matriz o grids, razón por la cual son ampliamente usadas para trabajar en imágenes [2]. La característica principal de las CNN es que como su nombre lo indica, utilizan la operación matemática llamada convolución, junto con una arquitectura general de una ANN [3].

La Fig. 1 tomada de [4], muestra la arquitectura general de una CNN, en la que se muestran las etapas por las que pasa la información para lograr la creación del modelo. La arquitectura general de la CNN está dividida en dos etapas principales: extracción de y clasificación

Diagrama, Dibujo de ingeniería

Descripción generada automáticamente

**Fig. 1.** Arquitectura general de una CNN

1. *Metodología*

La metodología propuesta en este trabajo constituye una serie de procesos y elementos que nos permiten procesar imágenes digitales a color o en escala de grises para detectar armas de fuego en ellas. El proceso de detección de armas de fuego se realizará utilizando la arquitectura neuronal. La metodología utilizada para este trabajo se resume en el diagrama que se muestra en la Fig. 2.

Icono

Descripción generada automáticamente

**Fig. 2.** Metodología para la detección de armas de fuego.

1. *Software*

Para implementar un modelo de red neuronal que permita realizar la detección de armas de fuego en imágenes digitales, es necesario disponer de un entorno de trabajo apropiado, que permita instalar componentes con las características apropiadas y el uso de librerías de aprendizaje profundo compatibles con Python, así como librerías de TensorFlow y OpenCV, y algunas más de utilidad como: Numpy, Keras, Sklearn y Matplotlib. La plataforma Google colab Pro fue el entorno elegido para escribir y probar códigos en el lenguaje Python para la detección de armas de fuego.

1. *Hardware*

La visión por computadora utiliza modelos de aprendizaje profundo, lo que conlleva procesos de entrenamiento que requieren alto poder de cómputo. En lo que respecta al Hardware, en primer lugar, se utilizó una computadora portátil, y posteriormente se optó por usar el un servidor en línea para acelerar el proceso de entrenamiento de los modelos neuronales, ya que la computadora portátil no contaba con los recursos de hardware apropiados, situación que provocaba esperar un tiempo excesivo para el entrenamiento.

1. *Adquisición de la información*

La materia prima de todo trabajo de en el área de IA son los datos, la calidad y cantidad de datos usados para entrenar un modelo de IA repercutirá directamente en el desempeño del sistema implementado. Debido al gran interés que existe desde hace algunos años en el desarrollo de sistemas inteligentes, se han desarrollado proyectos encaminados a la recopilación de información, lo que se ve reflejado en la existencia de bancos de datos de acceso libre o bajo suscripción, los que permiten a los interesados en el desarrollo de sistemas de IA, contar con la información para entrenar modelos para desempeñar alguna tarea. Los bancos de datos existentes evitan el costoso proceso de construir una base del conocimiento propia. Para este trabajo las imágenes usadas fueron descargadas de la plataforma Kaggle.

1. *Preparación de la información*

En esta etapa se utilizó la plataforma de Roboflow para etiquetar las imágenes que se usarán para entrenar el modelo. El etiquetado de las imágenes consiste en el proceso de marcar el área de la imagen en la que se visualiza un arma de fuego. La plataforma Roboflow nos proporciona las herramientas apropiadas para hacer la adecuada gestión de la información etiquetada Fig. 3.



**Fig. 3.** Interfaz para el etiquetado de imágenes

1. *Aprendizaje*

Para la detección de armas en imágenes, se utilizó el modelo fasterrcnn\_resnet50\_fpn de torchvision, es modelo preentrenado que combina el detector de objetos Faster R-CNN con una red neuronal convolucional avanzada. Este modelo se seleccionó por su capacidad para detectar objetos de diferentes tamaños y su robustez en aplicaciones de detección de objetos.

1. *Evaluación*

Medir el rendimiento del desempeño del modelo entrenado es una etapa muy importante en el proceso del aprendizaje. El usar métricas adecuadas para evaluar el modelo de aprendizaje permite tener un panorama de que tan bien el modelo generaliza la información del problema, y proporciona una pauta para hacer mejoras, cambios en la información o cambios el modelo, tales como cantidad de datos, distribución de los datos, parámetros del modelo, estructura del modelo, etc.

En este trabajo se utilizan las métricas tradicionales del machine learning precisión, recall y matriz de confusión para evaluar el desempeño del modelo entrenado.

La precisión proporciona la calidad del modelo en tareas de clasificación. La métrica calcula la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión. Esta dada por la Ec. (1).

(1)

El Recall es la capacidad de modelo para discriminar los casos positivos, de los negativos. Para calcularla usamos la Ec. (2).

(2)

El puntaje F1 es una forma de combinar la precisión y la sensibilidad del modelo, la definimos en la Ec. (3).

(3)

La matriz de confusión es una tabla que describe el rendimiento de un modelo. Se llama matriz de confusión porque muestra de forma fácil en dónde el modelo está clasificando erróneamente las clases.

Tiene 4 métricas a considerar, los verdaderos positivos (VP), Verdaderos negativos (VN), Falsos negativos(FN), Falsos positivos(FP) como se muestra en la Fig. 4.

Diagrama

Descripción generada automáticamente

**Fig. 4.** Matriz de confusión binaria.

1. **Resultados**

Para evaluar la metodología desarrollada se diseñaron 2 experimentos en los que se cambia la cantidad de imágenes usadas para el entrenamiento lo que da como resultado contar con dos modelos a evaluar.

La siguiente **Tabla I**muestra la información de las cantidades de imágenes usadas para entrenar las dos versiones de modelos (M1 y M2). La distribución de los datos en los conjuntos fue aleatoria usando el 70% para entrenamiento, el 20% para validación y 10% para pruebas.

**Tabla I.** Distribución de los datos por modelo.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Total** | **Entrenamiento** | **Validación** | **Prueba** |
| M1 | **302** | 211 | 60 | 31 |
| M2 | **2986** | 2091 | 596 | 299 |

Adicional a estos conjuntos se hicieron pruebas con otro conjunto de 100 imágenes, de las cuales, el 50 contenía armas y el restante 50 no contenía armas. Para que los modelos determinen si en una imagen hay armas o no.

1. *Resultados de entrenamiento y pruebas*

Para evaluar el desempeño de los modelos entrenados, se utilizaron métricas del aprendizaje precisión y sensibilidad o recall. La **Tabla II** muestra los resultados obtenidos por ambos modelos.

**Tabla II.** Tabla de métricas.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precisión** | **Recall** | **F1** |
| M1 | **1** | 0.99 | 0.96 |
| M2 | **1** | 0.97 | 0.96 |

Estos resultados muestran que, al evaluar el modelo con los datos de prueba, se tiene un excelente desempeño ya que la métrica precisión nos indica que todas las imágenes que tienen armas las detecta correctamente, lo que se ve corroborado con los datos de la matriz de confusión mostrados en la **Tabla III**, donde se proporciona una visión general del rendimiento del modelo al resumir las predicciones correctas e incorrectas.

**Tabla III.** Tabla de matriz de confusión.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **TP** | **TN** | **FP** | **FN** |
| M1 | **0.95** | 1 | 0.05 | 0 |
| M2 | **0.95** | 1 | 0.05 | 0 |

En lo que respecta a la precisión y el en función de las épocas, de forma similar al modelo M1, en el modelo M2 se alcanza una estabilidad alrededor de la época 100, lo que se puede observar en las Fig. 3 y Fig. 4. Esta estabilidad indica que el modelo ha logrado una convergencia en su capacidad para detectar correctamente las armas de fuego y minimizar los falsos negativos.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente **Fig. 5.** Precisión y recall M1.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**Fig. 6.** Precisión y recall M2.

1. *Clasificación*

Previamente se mencionó que se realizó una prueba de clasificación usando un conjunto de 100 imágenes, donde 50 contienen armas y 50 no contienen. Para evaluar el desempeño de los dos modelos, se construyeron matrices de confusión que reflejan los resultados de la clasificación.

El M1fue evaluado utilizando la matriz de confusión que se muestra en la Fig. 5.

Para el M2, también se realizó la evaluación y se obtuvo la matriz de confusión correspondiente, que se presenta en la Fig. 6.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente**Fig. 7.** Matriz de confusión M1

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente**Fig. 8.** Matriz de confusión M2

*Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente*

**Fig. 9.** Detección de armas

1. **Conclusiones**

El modelo basado en redes neuronales convolucionales demostró ser una buena alternativa para la detección de armas de fuego en las imágenes utilizadas en las pruebas. Aunque presenta algunas limitaciones en la precisión de la detección de clases, su desempeño es generalmente aceptable.

En cuanto al objetivo general del proyecto, se concluye que el modelo cumple con su propósito a pesar de los desafíos inherentes al trabajo con imágenes. Este resultado es motivador, ya que sugiere que el modelo tiene potencial para mejorar aún más con ajustes adicionales.

1. **Bibliografía**

[1] D. H. Hubel y T. N. Wiesel, “Receptive fields of single neurones in the cat's striate cortex,” *The Journal of Physiology*, vol. 148, no. 3, p. 574, 1959. Consultado en: Mes, Día, Año, [Online].

[2] C. Aggarwal, “Neural networks and deep learning,” Springer, vol. 10, no. 978, p. 3, 2018. Consultado en: Mes, Día, Año, [Online].

[3] Y. Bengio, I. Goodfellow y A. Courville, Deep learning, vol. 1, 2017. Consultado en: Mes, Día, Año, [Online].

[4] S. Saha, “A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,” 2018. Consultado en: Mes, Día, Año, [Online]. Disponible: <https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/>.

[5] M. L. Guevara, J. D. Echeverry y W. A. Urueña, “Detección de rostros en imágenes digitales usando clasificadores en cascada,” *Scientia et Technica*, vol. 1, no. 38, p. 1, jun. 2008. Consultado en: Mes, Día, Año, DOI: 10.22517/23447214.3679, [Online].

[6] C. M. Padilla, “Detección de objetos en imágenes digitales,” *Universidad de Camagüey Ignacio Agramonte y Loinaz*, 2016. [Online].

[7] J. A. Pacco Enriquez, “Desarrollo e implementación de un sistema CCTV antirrobo inteligente capaz de detectar armas de manera eficiente y bajo costo al interior de la joyería Chavelis de Arequipa usando un Raspberry PI 4,” 2022.

1. **Currículo y Fotografía**

Un hombre con lentes y camiseta verde

Descripción generada automáticamente**Tomas López-Perez** nació el 19 de febrero de 1998 en Oaxaca de Juárez. Obtuvo la licenciatura en Computación en la Universidad Autónoma “Benito Juárez” de Oaxaca en junio de 2023. Actualmente, cursa el primer año de la maestría en Ingeniería en la misma universidad. Además, trabaja como desarrollador web en la empresa Metrics ubicada en la ciudad de México. Sus áreas de interés incluyen inteligencia artificial, procesamiento de imágenes y visión por computadora.

Un hombre con una camiseta roja

Descripción generada automáticamente**José Edgar Lara-Ramírez** obtuvo la Licenciatura en Ingeniería en Sistemas Computacionales en el Instituto Tecnológico de Querétaro. Obtuvo la Maestría en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), y el grado de Doctor en Ingeniería Eléctrica por la División de Ingenierías de la Universidad de Guanajuato. Sus áreas de interés son las aplicaciones de Machine Learning y Deep Learning, Procesamiento de Imágenes, Procesamiento del Lenguaje Natural y Arquitecturas de Cómputo de paralelo.

Mujer con lentes

Descripción generada automáticamente

**María de Jesús Estudillo-Ayala** nació el 30 de diciembre de 1971 en Izúcar de Matamoros, estado de Puebla, México. Obtuvo la licenciatura en Computación por la Benemérita Universidad Autónoma de Puebla en diciembre de 2002. En 2003, obtuvo una especialidad en Ciencias Computacionales del Instituto Nacional de Astrofísica, Óptica y Electrónica. En septiembre de 2007, obtuvo el grado de Maestría en Ingeniería Eléctrica por la Universidad de Guanajuato y actualmente es candidata a doctora en Ingeniería Eléctrica por la misma universidad.

Trabajó en la Universidad del Mar desde septiembre de 2007 hasta febrero de 2010. En marzo de 2010, comenzó a trabajar en la Universidad Autónoma "Benito Juárez" de Oaxaca. Sus áreas de interés incluyen Inteligencia Artificial, Procesamiento de Imágenes y Procesamiento del Lenguaje.