EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación y Desarrollo Tecnológico

Tesis de Maestría

Detección y reconocimiento en tiempo real de armas a partir de videos en vías de transporte usando la Red Neuronal Convolucional Yolo V5

presentada por

Ing. Félix Cortés Ramírez

como requisito para la obtención del grado de

Maestro en Ciencias de la Computación

Director de tesis

Dr. Nimrod González Franco

Codirector de tesis

Dr. Dante Mújica Vargas

Cuernavaca, Morelos, México. 17 de febrero de 2025.

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

SEP TecNM CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
RECIBIDO
05 FEB 2025
SUBDIRECCIÓN ACADÉMICA

Centro Nacional de Investigación y Desarrollo tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor., 05/febrero/2025

OFICIO No. DCC/044/2025

Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de **FÉLIX CORTÉS RAMÍREZ** con número de control **M21CE006**, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Detección y reconocimiento en tiempo real de armas a partir de videos en vías de transporte usando la red neuronal convolucional yolo V5"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

Dr. Nimrod González Franco
Director de tesis

Dr. Dante Mújica Vargas
Codirector de tesis

Dr. Juan Gabriel González Serna
Revisor 1

Dr. Raúl Pinto Elías
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
05 FEB 2025
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
SERVICIOS ESCOLARES
cenidet

2025 Año de La Mujer Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,
e-mail: dcc_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

Centro Nacional de Investigación y Desarrollo tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor., 05/FEBRERO2025

Asunto: Liberación de producto académico
Comité tutorial

FÉLIX CORTÉS RAMÍREZ
CANDIDATO AL GRADO DE MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Detección y reconocimiento en tiempo real de armas a partir de videos en vías de transporte usando la red neuronal convolucional yolo V5",** avala que usted tiene el siguiente producto académico, derivado de su investigación.

Artículo: "Detección automática de delitos en sistemas de videovigilancia: una Revisión Sistemática de la Literatura " 7a. Jornada de Ciencia y Tecnología Aplicada, 2021, Vol. 4, Núm. 2, Pag. 133-138. https://jcyta.cenidet.tecnm.mx/2021/

Sin más por el momento, reciba un cordial saludo.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

Dr. Nimrod González Franco
Director de tesis

Dr. Dante Mújica Vargas
Codirector de tesis

Dr. Juan Gabriel González Serna
Revisor 1

Dr. Raúl Pinto Elías
Revisor 2

C.c.p. Depto. Servicios Escolares.

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
05 FEB 2025
cenidet
SERVICIOS ESCOLARES
RECIBIDO

2025 Año de La Mujer Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,
e-mail: dcc_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

100% LIBRE DE PLÁSTICO

Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 05/febrero/2025
Oficio No. SAC/053/2025
Asunto: Autorización de impresión de tesis

FÉLIX CORTÉS RAMÍREZ
CANDIDATO AL GRADO DE MAESTRO
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Detección y reconocimiento en tiempo real de armas a partir de videos en vías de transporte usando la red neuronal convolucional yolo V5"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
SUBDIRECCIÓN ACADÉMICA

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO

c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,
e-mail: acad_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

Dedicatoria

Dedico esta tesis a mi familia por apoyarme en todo momento, impulsarme a seguir con fortaleza y esmero durante esta nueva etapa de vida profesional.

Agradecimientos

Agradezco la Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo económico brindado durante mis estudios. Al Tecnológico Nacional de México / Centro Nacional de Investigación y Desarrollo Tecnológico CENIDET por brindar las instalaciones y permitirme realizar los estudios de maestría.

A mi director de tesis el Dr. Nimrod González Franco y mi codirector el Dr. Dante Mújica Vargas por brindarme su confianza, así como por guiarme durante el desarrollo de esta investigación mediante sus valiosos consejos y observaciones.

A los miembros del comité revisor, el Dr. Juan Gabriel González Serna, y el Dr. Raúl Pinto Elías, por sus acertadas correcciones y comentarios que permitieron enriquecer este trabajo.
¡Gracias!

Resumen

Este trabajo se centra en el desarrollo de un sistema para la detección automática de armas en escenarios de carreteras, abordando los desafíos relacionados con los entornos dinámicos y los cambios visuales drásticos. Para ello, se creó un dataset especializado (CriMex) con imágenes de cuatro clases principales: normal, pre-crimen, crimen y post-crimen, diseñado específicamente para escenarios de carreteras y complementado con imágenes etiquetadas y aumentadas mediante técnicas avanzadas.

Se implementaron modelos de aprendizaje profundo, como YOLO V5 y SSD, reentrenados para detectar armas blancas y de fuego con alta precisión, a su vez separandolos en 3 clases(arma corta, arma larga y arma blanca) alcanzando valores superiores al 90 %. Además, se desarrolló una aplicación móvil que permite la detección en tiempo real, con almacenamiento local de los resultados y facilidad de uso en dispositivos con recursos limitados.

Los experimentos realizados incluyeron pruebas controladas para validar el desempeño de los modelos, mostrando resultados favorables en términos de precisión y eficiencia en diversos escenarios. Este trabajo no solo aporta una solución práctica y replicable para la seguridad en carreteras, sino que también sienta las bases para futuras extensiones, como la detección de placas vehiculares y la integración en sistemas de vigilancia más complejos.

Palabras clave: detección automática de armas, Redes Neuronales Convolucionales, YOLO V5, *dataset* CriMex.

Abstract

This work focuses on the development of a system for the automatic detection of weapons in road scenarios, addressing the challenges posed by dynamic environments and drastic visual changes. To achieve this, a specialized dataset (CriMex) was created with four main classes: normal, pre-crime, crime, and post-crime. The dataset was specifically designed for road scenarios and was complemented with labeled and augmented images using advanced techniques.

Deep learning models, such as YOLO V5 and SSD, were retrained to detect bladed and firearm weapons with high precision,separating them into 3 classes (handgun, long gun and knife) achieving accuracy levels above 90%. Additionally, a mobile application was developed for real-time detection, featuring local storage of results and ease of use on resource-constrained devices.

Experiments included controlled tests to validate the performance of the models, demonstrating favorable results in terms of accuracy and efficiency across various scenarios. This work not only provides a practical and replicable solution for road safety but also lays the foundation for future extensions, such as license plate detection and integration into more complex surveillance systems.

Keywords: automatic weapon detection, Convolutional Neural Networks, YOLO V5, CriMex Dataset.

Índice General

Lista de Figuras xii

Lista de Tablas xiv

1 Introducción 1

1.1 Planteamiento del problema 2
1.1.1 Delimitación del problema 3
1.1.2 Complejidad del problema 3
1.2 Objetivos 4
1.2.1 Objetivo general 4
1.2.2 Objetivos específicos 4
1.3 Alcances y limitaciones 4
1.3.1 Alcances 4
1.3.2 Limitaciones 4
1.4 Organización de la tesis 5

2 Marco Conceptual 6

2.1 Redes Neuronales 6
2.2 Redes Neuronales Convolucionales 7
2.3 Detección de Objetos con CNN 12
2.4 Algoritmo YOLO 13

3 Estado del Arte 16

3.1 Trabajos sobre la detección de delitos usando Redes Neuronales 16
3.2 Discusión del Estado del Arte 27

4 Metodología de Solución 32

4.1 Propuesta de Solución 32
4.2 Implementación por Computadora de la Propuesta de Solución 35

5 Experimentación 37

5.1 Primer Experimento: validación del dataset con modelos CNN 37
5.2 Segundo Experimento: ajustes al dataset 41
5.3 Tercer Experimento: modelos de detección de objetos 45

5.3.1 Modelo SSD 46
5.3.2 Modelo YOLO V5 48
5.4 Informe de Pruebas de Software 52
5.4.1 Casos de prueba 52
5.4.2 Resultados de los Casos de Prueba 63
5.4.3 Anomalías 65

6 Conclusiones **67**
6.1 Objetivos y alcances logrados 67
6.2 Resultados de la investigación 68
6.2.1 Productos 68
6.2.2 Aportaciones 68
6.3 Conclusiones 69
6.4 Trabajo futuro 70

Referencias **71**

Apéndice A Producción **75**

Lista de Figuras

2.1 Una red neuronal simple con cuatro nodos de entrada (Mohamed et al., 2015) 7
2.2 Arquitectura de Red Neuronal Convolucional (Calvo, 2017) 8
2.3 Capa de convolución (Calvo, 2017) 9
2.4 Capa de submuestreo (Calvo, 2017) 10
2.5 Proceso de detección de objetos por medio de cajas delimitadoras del algoritmo YOLO (Xie and Yao, 2023) 12

3.1 Diagrama de flujo para el reconocimiento de actividades anómalas (Maqsood et al., 2021) 25
3.2 Segmentación de video utilizando los momentos obtenidos del método de segmento de comportamiento Pre-Crimen (PCB) (Martínez-Mascorro et al., 2021) 25

4.1 Arquitectura del Sistema de Extracción y Etiquetado 32
4.2 Arquitectura de la Aplicación Móvil 34
4.3 Arquitectura de funcionamiento de la aplicación móvil 35
4.4 GUI de base de datos de detección de armas 36

5.1 Resultados de entrenamiento de modelo EfficientNetB0 1er experimento 38
5.2 Resultados de entrenamiento de modelo VGG16 1er experimento 39
5.3 Resultados de clasificación del modelo EfficientNetB0 1er experimento 39
5.4 Resultados de clasificación del modelo VGG16 1er experimento 40
5.5 Resultados de entrenamiento del modelo Inception V3 2do experimento 41
5.6 Resultados de clasificación del modelo Inception V3 2do experimento 42
5.7 Resultados de entrenamiento de modelo VGG19 2do experimento 42
5.8 Resultados de clasificación del modelo Inception V3 2do experimento 42
5.9 Resultados de entrenamiento de modelo Xception 2do experimento 43
5.10 Resultados de clasificación del modelo Inception V3 2do experimento 43
5.11 Resultados de entrenamiento de modelo ResNet50 2do experimento . 43
5.12 Resultados de entrenamiento de modelo VGG16 2do experimento 44
5.13 Resultados de clasificación del modelo Inception V3 2do experimento 44
5.14 Ejemplo ilustrativo del contenido del dataset de entrenamiento para el sistema de detección de armas 46

5.15 Resultados de pérdida del error del modelo SSD en etapa de entrenamiento 47
5.16 Prueba del modelo SSD con una imagen 47
5.17 Resultado de las curvas P-R correspondiente al entrenamiento de YOLO V5 48
5.18 Resultados de pérdida del error y precisión del modelo YOLO V5 en etapa de entrenamiento 49
5.19 Imagenes ilustrativas de la detección con YOLO en la etapa de validación 50
5.20 Imagen ilustrativa de la detección con YOLO en videos grabados en estacionamiento del CENIDET 1 51
5.21 Imagen ilustrativa de la detección con YOLO en videos grabados en estacionamiento del CENIDET 2 51
5.22 Ejemplos de Funcionalidad de Aplicación Móvil 64
5.23 Ilustración de precisión de detección de armas 65
5.24 Diseño de implementación de cámaras en distintos vehículos 66

A.1 Portada de artículo científico publicado en la 7JCyTA, CENIDET 2021 76
A.2 Portada de poster sobre el dataset Crimex, CENIDET 2022 77
A.3 Certificado de registro de CRIMEX en INDAUTOR, CENIDET 2022 78
A.4 Certificado de registro de software Crimex Image Generator en INDAUTOR, CENIDET 2022 79

Lista de Tablas

3.1 Comparativa entre los artículos relevantes para el Estado del Arte . . 28
3.1 Comparativa entre los artículos relevantes para el Estado del Arte . . 29
3.1 Comparativa entre los artículos relevantes para el Estado del Arte . . 30
3.1 Comparativa entre los artículos relevantes para el Estado del Arte . . 31

5.1 Especificaciones de los equipos utilizados para las pruebas.....37
5.2 Resultados de desempeño de los modelos CNN.....39
5.3 Resultados de prueba de modelos de CNN experimento 2 45
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 53
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 54
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 55
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 56
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 57
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 58
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 59
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 60
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 61
5.4 Casos de prueba de funcionalidad de la Aplicación Móvil 62

6.1 Cumplimiento de objetivos.....67

Capítulo 1

Introducción

La detección y el reconocimiento de objetos en tiempo real a partir de videos es un área de investigación crucial en el campo de la visión por computadora, especialmente en aplicaciones de seguridad pública y monitoreo (Bhatti et al., 2021). En particular, la identificación de armas en entornos como vías de transporte, que incluyen carreteras, estaciones de tren o aeropuertos, es una tarea esencial para la prevención de delitos y la protección de la vida humana. Las armas, debido a su naturaleza diversa y a menudo oculta o disfrazada, representan un desafío significativo para los sistemas de vigilancia tradicionales (Shah et al., 2021).

En los últimos años, el campo de la inteligencia artificial ha avanzado notablemente en la mejora de los sistemas de reconocimiento de objetos mediante el uso de Redes Neuronales Convolucionales (CNN). Estas redes, basadas en la imitación del procesamiento visual del cerebro humano, han demostrado ser altamente eficaces en tareas complejas como la clasificación, segmentación y detección de objetos en imágenes y videos (Alzubaidi et al., 2021). A medida que las redes neuronales han evolucionado, también lo han hecho los modelos específicos para la detección en tiempo real, siendo el modelo You Only Look Once (YOLO) uno de los más destacados debido a su rapidez y precisión (Redmon, 2016). Esta investigación se enfoca en la evaluación de detección con el modelo YOLO V5, el cual ha sido ampliamente adoptado en aplicaciones que requieren detección de objetos en tiempo real, gracias a su alta velocidad de procesamiento y su capacidad para detectar múltiples objetos de interés en una sola pasada. Este modelo ha demostrado ser especialmente efectivo en entornos dinámicos y con variabilidad en las condiciones de la imagen, como en el monitoreo de tráfico en tiempo real (Jocher et al., 2022). En el contexto de la detección de armas, YOLO V5 ofrece una solución viable para la vigilancia automatizada en entornos como las vías de transporte, donde los objetos de interés deben ser identificados y clasificados rápidamente para evitar situaciones de riesgo (Sumi and Dey, 2023).

Este trabajo de investigación tiene como objetivo explorar el uso de YOLO V5 para la detección y reconocimiento en tiempo real de armas en videos capturados en vías de transporte. A través de la implementación de este modelo, se busca no solo mejorar la precisión, sino también adaptar el sistema a las características específicas de estos entornos, como la variabilidad en la iluminación, el movimiento de los objetos y las diferentes perspectivas de la cámara. En este sentido, se pretende contribuir al desarrollo de sistemas de vigilancia más eficientes y seguros que puedan ser implementados en infraestructuras críticas de transporte, contribuyendo de esta manera a la prevención de amenazas en tiempo real. La relevancia de esta investigación radica en su potencial para optimizar los sistemas de seguridad y vigilancia, no solo mejorando la capacidad de detectar armas de manera temprana, sino también ofreciendo soluciones escalables y aplicables a diferentes escenarios de transporte.

1.1 Planteamiento del problema

En la actualidad, los transportistas enfrentan constantemente riesgos asociados a delitos como el robo y el asalto durante sus trayectos, lo que pone en peligro su seguridad y la integridad de las mercancías que transportan. Aunque algunos conductores emplean sistemas de videovigilancia que permiten el monitoreo remoto, estos siguen dependiendo de la intervención humana para activar el llamado de auxilio en caso de una emergencia. Por otro lado, en algunos casos se hace uso de dispositivos móviles para alertar a las autoridades, pero este mecanismo también es limitado, ya que depende de la disponibilidad y la capacidad de los transportistas para realizar la llamada, lo que puede generar retrasos en la respuesta ante un incidente. Estos sistemas tradicionales de seguridad presentan diversas deficiencias que comprometen la eficacia en la detección y respuesta ante situaciones de riesgo, ya que la intervención humana es indispensable y puede resultar ineficaz cuando se requieren acciones rápidas y automáticas.

El problema de esta investigación se centra en automatizar la detección de incidentes relacionados con el robo o asalto a transportistas mediante el uso de técnicas avanzadas de *Deep Learning*, en particular, Redes Neuronales Convolucionales. Este sistema permitirá la identificación de situaciones sospechosas sin intervención humana, con el fin de emitir notificaciones de alerta a las autoridades competentes y proporcionar pruebas visuales de la escena del delito al personal encargado de la videovigilancia.

Para esta investigación se partió con la siguiente pregunta de investigación: ¿Cuáles son las características visuales y comportamentales clave que pueden ser identificadas en tiempo real en videos de carreteras para detectar incidentes de robo o asalto a transportistas mediante técnicas de Deep Learning?

1.1.1 Delimitación del problema

Para esta investigación se trabajó en la automatización de la detección de delitos en carretera utilizando técnicas de aprendizaje profundo, en específico una red neuronal convolucional, tomando como entrada fotogramas de vídeos que serán clasificados en cuatro clases: normal, pre-crimen, crimen, post-crimen, para lo cual se realizó la detección de armas de fuego y de armas blancas. Se utilizaron fotogramas extraídos de vídeos grabados en buenas condiciones climatológicas, descartando aquellos en los que se percibió lluvia, nieve, baja o alta iluminación, accidentes y desenfoque .

1.1.2 Complejidad del problema

Este proyecto presenta una serie de complejidades técnicas y conceptuales que abarcan tanto el diseño del sistema como su implementación y evaluación. Las principales dimensiones de complejidad de este proyecto son las siguientes:

1. **Desafíos en la recolección y preprocesamiento de datos:** La creación de un dataset adecuado y diverso para entrenar el modelo. La detección de armas en contextos de transporte implica capturar una variedad de situaciones, condiciones de iluminación, diferentes tipos de armas y ángulos de visión, lo que puede dificultar la creación de un conjunto de datos representativo.

2. **Complejidad del modelo:** YOLO V5 es un modelo que, aunque es reconocido por su eficiencia y rapidez, también es desafiante de ajustar y entrenar. La afinación de los hiperparámetros es crucial para obtener un buen rendimiento.

3. **Condiciones del mundo real y escenarios dinámicos:** La implementación de un sistema de detección en entornos dinámicos, como las vías de transporte, implica enfrentar varios factores complicados. Las condiciones de iluminación variables, el movimiento rápido de los vehículos, la presencia de múltiples objetos y el cambio constante de ángulos de visión hacen que la tarea de detección sea aún más compleja.

1.2 Objetivos

1.2.1 Objetivo general

Detectar comportamientos sospechosos de robos a transportistas, mediante la aplicación de técnicas de Aprendizaje Profundo basadas en Redes Neuronales Convolucionales.

1.2.2 Objetivos específicos

- Implementar y evaluar un modelo de aprendizaje profundo para el reconocimiento de delitos.
- Crear un dataset de imágenes y videos de situaciones sospechosas y violentas en carreteras, con los descriptores y las etiquetas correspondientes.
- Desarrollar un sistema para la identificación de comportamientos sospechosos y normales para la detección de armas en escenas de videos, reutilizando una red neuronal convolucional que se adapte a esta necesidad.

1.3 Alcances y limitaciones

1.3.1 Alcances

- Integrar videos e imágenes obtenidas de fuentes públicas en un repositorio.
- Aplicar técnicas de preprocesamiento de imágenes.
- Considerar la detección de robos y asaltos a transportistas.
- Implementar un sistema de extracción y etiquetado de imágenes de crimen en carretera.
- Reutilizar o proponer una red neuronal que detecte el crimen en carretera.

1.3.2 Limitaciones

- No se consideró la detección de crímenes en situaciones de baja iluminación y/o mal clima.
- No se trabajó en la detección de accidentes.
- No se desarrolló un sistema de alertas y detección en tiempo real.

1.4 Organización de la tesis

Este documento de tesis se compone por seis capítulos principales, aunado a las secciones de anexos y referencias. En el Capítulo 2 se describen los conceptos teóricos necesarios para comprender el proyecto, como los conceptos de Redes Neuronales Convolucionales y detección de objetos. En el Capítulo 3 se presentan los trabajos más recientes sobre detección de delitos en videos de transportistas y la implementación de Redes Neuronales para la detección de objetos. En el Capítulo 4 se describe la metología de solución que se siguió para realizar la detección y en el Capítulo 5 el diseño de experimentos y resultados obtenidos. Finalmente en el Capítulo 6 se describen las concluciones, productos, aportaciones y trabajo futuro del proyecto.

Capítulo 2

Marco Conceptual

La automatización en la detección de delitos representa un avance significativo en el ámbito de la seguridad, facilitando el trabajo humano y ofreciendo resultados más consistentes, ya que las máquinas no presentan las limitaciones humanas, como la fatiga o la atención dispersa. Este tipo de sistemas de detección pueden basarse en el modelo de Pre-Crime Behavior (PCB), el cual permite identificar patrones de comportamiento previos al delito a partir de muestras de video. El método PCB divide el video en tres segmentos: el segmento de evidencia del crimen, el segmento de comportamiento sospechoso y el segmento de comportamiento previo al crimen. El observador debe analizar el video completo para identificar momentos específicos dentro de estos segmentos (Martínez-Mascorro et al., 2020b). En este contexto, las Redes Neuronales Convolucionales 3D (3DCNN) juegan un papel crucial, ya que son capaces de procesar datos espacio-temporales, como los videos, para extraer características clave que permiten detectar comportamientos sospechosos (Martínez-Mascorro et al., 2020c).

En general, los modelos de *Deep Learning* (DL), como las Redes Neuronales Convolucionales (CNN), se utilizan para detectar a personas cuyo comportamiento indica una alta probabilidad de cometer un delito. Esto se logra analizando la evolución del comportamiento de las personas en los videos antes de que el delito ocurra, lo cual resulta crucial para predecir y prevenir crímenes (Martínez-Mascorro et al., 2020a). A continuación, se presentan algunos conceptos clave que permiten comprender los fundamentos de esta investigación.

2.1 Redes Neuronales

Una Red Neuronal se puede definir como un sistema que modela la relación entre entradas y salidas, inspirado en el funcionamiento del sistema nervioso humano. A diferencia de los sistemas de computación tradicionales, las redes neuronales no emplean una lógica secuencial; en su lugar, procesan información de manera paralela, lo que les permite aprender y generalizar patrones no presentes explícitamente en los

datos de entrenamiento (Serna et al., 2018).

Una Red Neuronal Artificial consta de una estructura de unidades interconectadas que simulan las neuronas en el cerebro. Cada neurona recibe señales de entrada, las procesa y transmite una señal de salida. Este proceso involucra la sumatoria de los datos de entrada ponderados por la fuerza sináptica respectiva, y una función de activación que limita la amplitud de la salida (Ahmadi et al., 2021).

La Figura 2.1 ilustra un ejemplo simple de una red neuronal con cuatro nodos de entrada, dos nodos intermedios y un nodo de salida, mostrando los pesos de conexión y el resultado del proceso de entrenamiento.

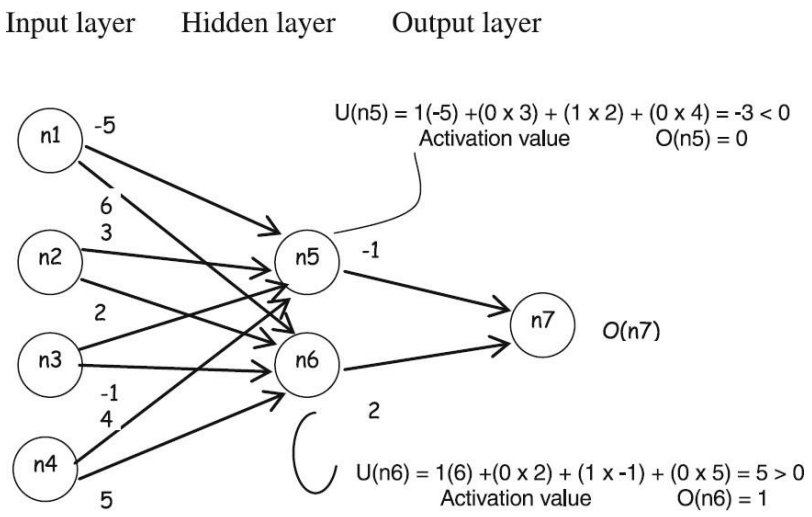


Figura 2.1 Una red neuronal simple con cuatro nodos de entrada (Mohamed et al., 2015)

2.2 Redes Neuronales Convolucionales

El aprendizaje profundo ha logrado avances notables en el reconocimiento de imágenes y videos, especialmente mediante el uso de Redes Neuronales Convolucionales (CNN). Estas redes se destacan por su alto rendimiento en tareas de visión por computadora y reconocimiento de patrones (Martínez-Mascorro et al., 2020a). El principal objetivo de una CNN es extraer características relevantes de una imagen y utilizarlas para detectar o clasificar objetos dentro de ella. En la Figura 2.2 se ilustra la arquitectura básica de una Red Neuronal Convolucional, donde las capas de convolución juegan un papel fundamental en la extracción de características.

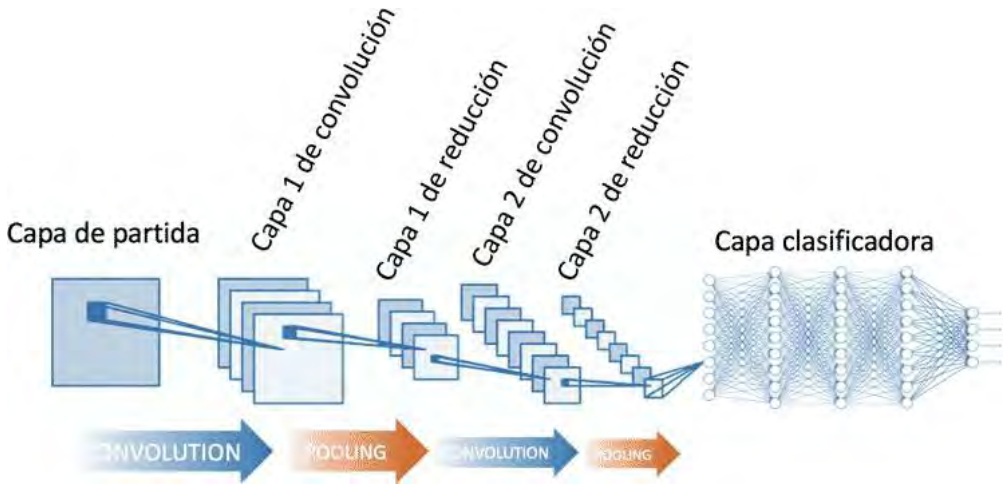


Figura 2.2 Arquitectura de Red Neuronal Convolucional (Calvo, 2017)

La convolución es una operación matemática que combina dos funciones, en este caso, una imagen y un filtro (kernel), para extraer características. En el contexto de las CNNs, la convolución en una imagen discreta se realiza mediante la siguiente definición:

$$I_{out}(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \tag{2.1}$$

donde $I(i, j)$ es la imagen de entrada y $K(m, n)$ es el filtro que se aplica a la imagen.

La convolución se realiza deslizando el filtro sobre la imagen. Para cada posición del filtro, se realiza una multiplicación elemento por elemento entre la subregión de la imagen y el filtro, seguida de una suma. Este proceso se repite en todas las posiciones del filtro sobre la imagen, como se muestra en la Figura 2.3. Estas características corresponden a cada posible ubicación del filtro en la imagen original (Taye, 2023).

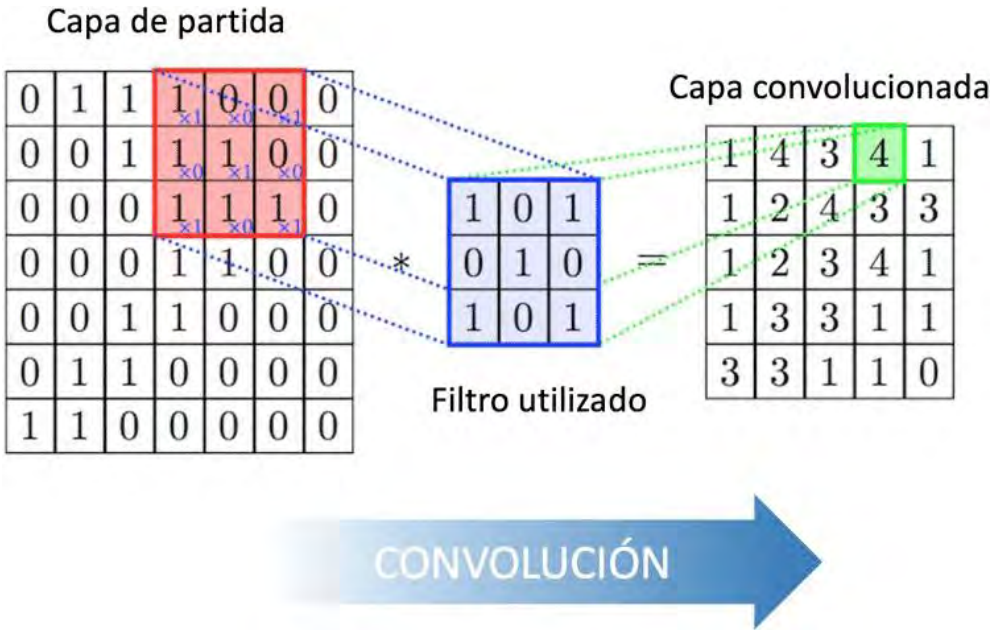


Figura 2.3 Capa de convolución (Calvo, 2017)

El tamaño de la salida de la operación de convolución depende de las dimensiones de la imagen y del filtro. Este se calcula utilizando las siguientes operaciones:

$$H_{out} = \frac{H - F}{S} + 1 \tag{2.2}$$

$$W_{out} = \frac{W - F}{S} + 1 \tag{2.3}$$

donde H y W son las dimensiones de la imagen, F es el tamaño del filtro, y S es el paso ($stride$) con el que se mueve el filtro.

Ventajas de la Convolución

La convolución tiene las siguientes características importantes:

- **Localidad:** La operación es local, es decir, cada valor de salida depende de una región local de la imagen.
- **Parámetros Compartidos**: El mismo filtro se aplica en todas las posiciones de la imagen, reduciendo el número de parámetros a entrenar.
- **Invarianza a la Translación:** El filtro puede detectar patrones sin importar su posición en la imagen.

Tras la convolución, los mapas de características pasan por una función de activación, donde se suelen usar funciones rectificadoras como ReLU. Además, existen otras funciones como Leaky ReLU (Xu et al., 2020) o Maxout (Goodfellow et al., 2013), pero se recomienda evitar el uso de la función sigmoide logística debido a su tendencia a saturarse en ciertas condiciones.

Reducción de características

En el proceso de reducción, se disminuye la cantidad de parámetros al seleccionar las características más relevantes de cada mapa. Las capas de *pooling* se utilizan para reducir las dimensiones de los mapas de características, lo que ayuda a disminuir la carga computacional y prevenir el sobreajuste. Existen dos tipos comunes de *pooling*:

Max Pooling

El *max pooling* toma el valor máximo de una subregión de la imagen (ventana). Matemáticamente, para una ventana de tamaño $k \times k$ (Murray and Perronnin, 2014) se tiene la siguiente ecuación:

$$P_{\text{out}}(i, j) = \max\ (I(i + m, j + n)) \tag{2.4}$$

donde se selecciona el valor máximo de la subregión $k \times k$, como se muestra en la Figura 2.4.

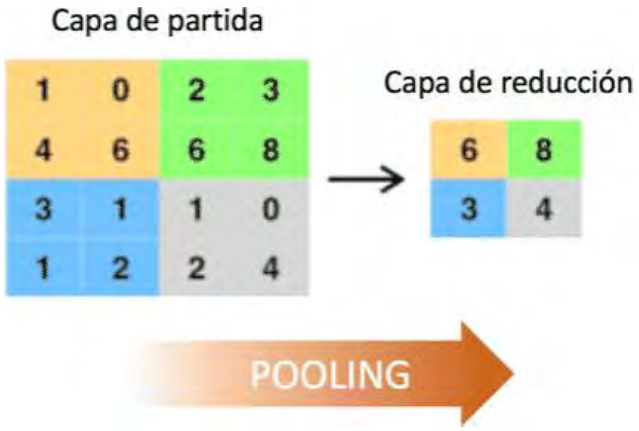


Figura 2.4 Capa de submuestreo (Calvo, 2017)

Average Pooling

El *average pooling* (Yu et al., 2014) calcula el promedio de los valores dentro de una subregión:

$$P_{out}(i,j) = \frac{1}{k^2} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} I(i+m, j+n) \tag{2.5}$$

Max pooling y *average pooling* son técnicas esenciales en redes neuronales convolucionales para reducir dimensiones y extraer patrones relevantes, cada una con enfoques complementarios. *Max pooling* prioriza las características más destacadas al seleccionar los valores máximos, lo que es ideal para resaltar activaciones fuertes como bordes o texturas. En contraste, *average pooling* proporciona una representación más equilibrada al considerar todos los valores de una región, manteniendo información contextual. La elección entre ambas depende del objetivo específico y las características del problema, permitiendo ajustar el modelo según las necesidades de análisis y precisión (Bieder et al., 2021). Sin embargo, para ofrecer una clasificación, un modelo neuronal depende de una capa clasificadora.

Clasificación

La capa final de la red es una capa clasificadora, que tiene tantas neuronas como el número de clases que se desean predecir. La salida de una neurona en la capa densa se puede calcular utilizando:

$$z_j = \sum_{i=1}^{n} w_{ij} \cdot x_i + b_j \tag{2.6}$$

donde: z_j es la salida de la neurona j de la capa densa, w_{ij} es el peso de la conexión entre la neurona i de la capa anterior y la neurona j de la capa actua, x_i es la entrada de la neurona i. b_j es el sesgo o *bias* de la neurona j y n es el número de entradas de la capa, es decir, el número de neuronas en la capa anterior.

Una vez que se calcula el valor z_j, la salida de la neurona se pasa a través de una función de activación, como la función sigmoide, ReLU, softmax, etc., dependiendo del tipo de tarea que se realice, ya sea clasificación binaria, clase continua o multiclase.

2.3 Detección de Objetos con CNN

You Only Look Once (YOLO) es un algoritmo de visión artificial basado en Aprendizaje Profundo que ha revolucionado la detección de objetos en imágenes y videos debido a su eficiencia y velocidad. A diferencia de otros métodos tradicionales de detección de objetos, que suelen dividir el proceso en múltiples etapas, como la propuesta de regiones y la clasificación de objetos, YOLO aborda la detección de manera única y *end-to-end*, es decir, en un solo paso (Redmon, 2016). Este enfoque integral le permite predecir de forma simultánea las clases de los objetos y sus localizaciones en una imagen o un video.

La arquitectura de YOLO se basa en una red neuronal convolucional (CNN) que divide una imagen en una rejilla de celdas. Cada celda es responsable de predecir un conjunto de cajas delimitadoras junto con las probabilidades de clase correspondientes a los objetos que contiene, este proceso es mostrado visualmente en la Figura 2.5. A través de un único pase de la red, YOLO produce una predicción simultánea para todas las clases y ubicaciones de objetos, lo que hace que sea extremadamente rápido, ya que no requiere pasar por diferentes etapas de procesamiento (Redmon, 2016).

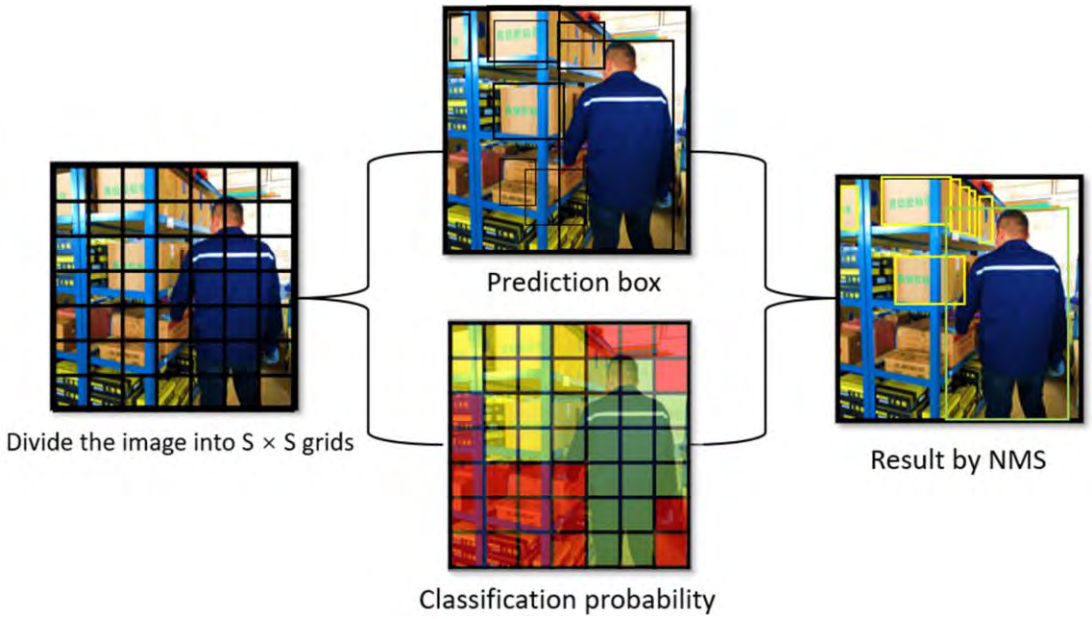


Figura 2.5 Proceso de detección de objetos por medio de cajas delimitadoras del algoritmo YOLO (Xie and Yao, 2023)

2.4 Algoritmo YOLO

División de la Imagen en una Rejilla

Cuando YOLO procesa una imagen, primero la divide en una rejilla de $S \times S$ celdas. Cada celda de la rejilla tiene la responsabilidad de predecir objetos cuyas cajas delimitadoras (*bounding boxes*) tienen el centro dentro de esa celda. La imagen de entrada tiene un tamaño de $W \times H$, y se divide en una rejilla de $S \times S$, por lo que el tamaño de cada celda es $\frac{W}{S} \times \frac{H}{S}$ (Jiang et al., 2022).

Predicciones de cada Celda

Cada celda de la rejilla debe predecir varias cantidades, que incluyen (Thuan, 2021):

- B cajas delimitadoras por celda, cada una representada por:
  - x, y: las coordenadas del centro de la caja en relación con la celda de la rejilla (estas coordenadas son relativas a la celda de la rejilla).
  - w, h: el ancho y la altura de la caja, que se predicen con respecto al tamaño de la imagen original.
  - p_{obj}: la probabilidad de que haya un objeto dentro de esa caja.
- Para cada caja, YOLO también predice la probabilidad de que esa caja pertenezca a una clase de objeto específica. Esto se calcula como $P(\text{class}_i)$, donde i representa las clases de objetos posibles.

En resumen, para cada celda, YOLO realiza las siguientes predicciones:

- B cajas delimitadoras con:
  - 4 valores: (x, y, w, h) para la ubicación y tamaño de la caja.
  - 1 valor: p_{obj}, la probabilidad de que haya un objeto dentro de la caja.
- C valores de probabilidad para cada clase C, es decir, $P(\text{class}_1), P(\text{class}_2), \ldots, P(\text{class}_C)$.

Cálculo de las Predicciones

La salida de YOLO se organiza como un tensor de dimensiones $S \times S \times (B \times 5 + C)$, donde $S \times S$ es el tamaño de la rejilla, además, considerando que $B \times 5$ cada caja tiene 5 parámetros que representan su localización y la probabilidad de que contenga un objeto, para B cajas, se tienen $B \times 5$ y C número de clases.

Fórmulas de Predicción

Para cada celda de la rejilla, el algoritmo predice lo siguiente:

Las coordenadas (x, y) se calculan considerando la *probabilidad* de que el centro de la caja esté dentro de esa celda de la rejilla. Estas coordenadas se normalizan entre 0 y 1 con respecto al tamaño de la celda de la rejilla.

El Tamaño de la Caja w y h, que se predicen como proporciones relativas a las dimensiones de la imagen. YOLO usa una representación *logarítmica* para calcular las dimensiones relativas:

$$w = \text{sigmoid}(t_w) \cdot W \tag{2.7}$$

$$h = \text{sigmoid}(t_h) \cdot H \tag{2.8}$$

donde t_w y t_h son las predicciones de la red para el ancho y la altura, y W y H son las dimensiones de la imagen original.

La probabilidad de que un objeto esté presente dentro de la caja es predicha por la red mediante la fórmula:

$$p_{obj} = \text{sigmoid}(t_{obj}) \tag{2.9}$$

donde t_{obj} es la salida de la red neuronal para esa celda.

La probabilidad de la Clase, es de que la caja pertenezca a una clase específica $P(\text{class}_i)$, se utiliza una *función sigmoide* para cada clase:

$$P(\text{class}_i) = \text{sigmoid}(t_{\text{class}_i}) \tag{2.10}$$

donde t_{class_i} es la salida de la red para la clase i.

Cálculo de la Confianza (Confidence Score)

La *confianza* de cada predicción de la caja C_{box} es el producto de dos probabilidades:

- p_{obj}, la probabilidad de que haya un objeto dentro de la caja.
- $P(\text{class}_i)$, la probabilidad de que esa caja pertenezca a la clase i.

La fórmula general para la confianza de una caja es:

$$C_{\text{box}} = p_{obj} \times P(\text{class}_i) \tag{2.11}$$

Esta puntuación es multiplicada por el valor de la caja para cada clase y por cada celda de la rejilla.

Post-Procesamiento: NMS (Non-Maximum Suppression)

Una vez que YOLO ha generado las predicciones de cajas delimitadoras, se aplica un *algoritmo de supresión de no máximos (NMS)* para eliminar las predicciones redundantes. NMS es un proceso que elimina las cajas que se solapan excesivamente y que tienen una baja confianza. Esto se realiza en dos pasos:

1. Se seleccionan las cajas con mayor puntuación de confianza.
2. Se descartan las cajas que se solapan con la caja seleccionada si el índice de Intersección Sobre la Unión (IoU) entre ellas es superior a un umbral predefinido.

La salida final de YOLO es un conjunto de cajas delimitadoras y sus correspondientes clases, que indican la ubicación y la categoría de los objetos detectados en la imagen.

En resumen, el algoritmo YOLO se destaca por su capacidad de realizar detección de objetos en tiempo real con alta precisión y eficiencia, lo que lo convierte en una herramienta ideal para aplicaciones en entornos dinámicos y de alta exigencia, como las vías de transporte. Su enfoque unificado permite procesar imágenes en una sola pasada a través de la red, optimizando recursos computacionales y reduciendo el tiempo de inferencia.

Este marco conceptual sienta las bases para explorar el estado del arte en la detección de armas mediante YOLO y otras arquitecturas similares, evaluando sus avances, limitaciones y las soluciones más recientes desarrolladas para abordar los desafíos específicos de este campo.

Capítulo 3

Estado del Arte

En esta sección se presenta brevemente los trabajos más relevantes relacionados con la implementación de Redes Neuronales para la detección de objetos y detección de escenas de crimen. Esta revisión es un compilado de las aportaciones más relevantes para la investigación. De la misma manera se presentan los trabajos que anteceden a este, dentro del mismo centro de investigación.

3.1 Trabajos sobre la detección de delitos usando Redes Neuronales

Edge artificial intelligence and super-resolution for enhanced weapon detection in video surveillance (Berardini et al., 2025)

El artículo propuso YOLOSR, un modelo que combina YOLOv8-small con una red de superresolución (EDSR) para mejorar la detección de armas en tiempo real en entornos con recursos limitados. Durante el entrenamiento, la superresolución ayudó a extraer mejores características, pero en la inferencia fue eliminada para mantener la eficiencia.

Probado en el dataset WeaponSense y un NVIDIA Jetson Nano, YOLOSR mejoró la precisión promedio en 10.2 puntos porcentuales respecto a YOLOv8-small, sin aumentar la complejidad computacional, ofreciendo una solución eficiente y precisa para dispositivos edge.

WeaponVision AI: a software for strengthening surveillance through deep learning in real-time automated weapon detection (Yadav et al., 2025)

El artículo presentó WeaponVision AI, un sistema avanzado capaz de detectar armas de fuego de manera autónoma en transmisiones en vivo, videos grabados e imágenes, incluso en condiciones de poca iluminación.

El sistema se basó en una arquitectura de aprendizaje profundo derivada de YOLOv7, modificada y entrenada con un amplio conjunto de datos de 79,558 imágenes de armas. Tras el entrenamiento, el modelo alcanzó resultados satisfactorios, con una precisión del 91.75% y una precisión promedio (mAP) del 92.15%.

Robust Weapon Detection in Dark Environments using Yolov7-DarkVision (Yadav et al., 2024)

En este artículo se propuso un enfoque novedoso para la detección de armas en condiciones de poca luz o escenarios nocturnos, un área que hasta ahora ha recibido poca atención en comparación con la detección de armas en entornos iluminados.

Los autores adaptaron el modelo de aprendizaje profundo YOLOv7, desarrollando una variante llamada YOLOv7-DarkVision, que combina un algoritmo de mejora de brillo y técnicas avanzadas de procesamiento de imágenes integradas en la arquitectura de YOLOv7.

Para entrenar el modelo, se utilizó un dataset compuesto por 15,367 imágenes y cinco videos oscuros de diversas fuentes para evaluar su rendimiento. Los resultados mostraron que este modelo tiene una precisión del 95.50% y un puntaje F1 del 93.41%, demostrando ser robusto y preciso para detectar armas en condiciones nocturnas desafiantes.

Effective Strategies for Enhancing Real-Time Weapons Detection in Industry (Torregrosa-Domínguez et al., 2024)

El artículo propone un sistema para mejorar la detección de armas mediante cámaras de vigilancia, abordando problemas como la detección en tiempo real, la precisión en objetos pequeños y la reducción de falsos positivos.

El esquema incluye dos módulos que optimizan el rendimiento de un detector reconocido sin afectar significativamente el tiempo de inferencia. Además, utiliza una técnica de coincidencia de escala para mejorar la detección de armas con proporciones pequeñas, logrando un aumento del 13.23% en precisión promedio al detectar objetos pequeños.

Los resultados experimentales mostraron que el sistema reduce los falsos positivos en un 71% respecto al modelo base, manteniendo un tiempo de inferencia bajo de 34

cuadros por segundo en una NVIDIA GeForce RTX-3060 con resolución de 720p, comparado con los 47 cuadros por segundo del modelo base.

Improving Armed People Detection on Video Surveillance Through Heuristics and Machine Learning Models (Amado-Garfias et al., 2024)

Este artículo abordó la identificación de personas armadas mediante cámaras de vigilancia en tiempo real, un área menos estudiada en comparación con la detección general de armas. La solución propuesta utilizó el modelo YOLOv4 para detectar personas, armas de fuego (pistolas y revólveres) y rostros. A partir de los videos en tiempo real, se extrajo información como coordenadas de las cajas delimitadoras, distancias y áreas de intersección entre las armas y las personas en cada fotograma para identificar a los individuos armados.

Se enfrentaron desafíos como la oclusión, armas ocultas y la proximidad entre personas. Para abordar estos problemas, los autores desarrollaron tres heurísticas (método de centros, intersecciones y distancias) y siete modelos de aprendizaje automático: Random Forest, Multilayer Perceptron, k-Nearest Neighbors, Support Vector Machine, Regresión Logística, Naive Bayes y Gradient Boosting.

El modelo Random Forest obtuvo el mejor desempeño, con una precisión del 85.44%, un valor F1 de 87.87%, una sensibilidad del 88.68% y una precisión promedio del 87.07%.

A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras (Berardini et al., 2024)

El artículo abordó la detección temprana de armas de fuego y cuchillos en videos de vigilancia, enfocándose en superar dos desafíos principales: el tamaño reducido de las armas en el campo de visión y la necesidad de retroalimentación en tiempo real en dispositivos edge de bajo costo.

Para ello, se desarrolló un enfoque de doble paso con redes neuronales convolucionales (CNN). La primera CNN detectaba personas, mientras que la segunda identificaba armas. Este sistema fue implementado en un dispositivo NVIDIA Jetson Nano conectado a una cámara IP, logrando un rendimiento cercano al tiempo real sin requerir hardware costoso.

El sistema alcanzó una precisión promedio de 79.30% y una velocidad de procesamiento de 5.10 cuadros por segundo, destacándose frente a otros métodos de vanguardia y promoviendo el uso de sistemas de videovigilancia automatizados y accesibles.

Weapon Violence Dataset 2.0: A synthetic dataset for violence detection (Nadeem et al., 2024)

Este artículo abordó el desafío de obtener datos auténticos para entrenar modelos en áreas sensibles como la detección de violencia, donde la escasez de datos reales y las restricciones éticas complicaron la investigación. Para superar estas limitaciones, los autores crearon el Weapon Violence Dataset (WVD), el primer conjunto de datos sintético para la detección de violencia, generado dentro del videojuego fotorrealista Grand Theft Auto V (GTA-V).

El WVD incluyó clips cuidadosamente seleccionados de peleas entre personas desde una vista frontal, con armas de fuego (violencia "caliente"), armas blancas (violencia "fría") y escenarios sin violencia (control). Además, los videos incluyeron tanto imágenes RGB normales como flujo óptico, permitiendo a la comunidad de investigación entrenar modelos profundos en datos sintéticos con la opción de expandir el corpus si fuera necesario.

El dataset fue diseñado para ser accesible a la comunidad científica y se puso a disposición pública en Kaggle.

A Comprehensive Study towards High-level Approaches for Weapon Detection using Classical Machine Learning and Deep Learning Methods (Yadav et al., 2023)

En el artículo se abordó la necesidad de sistemas automáticos para detectar actividades delictivas, como robos a mano armada, en imágenes de cámaras de circuito cerrado (CCTV), eliminando la dependencia de intervención humana. Aunque se han desarrollado algoritmos de alto rendimiento, estos mostraron limitaciones en condiciones específicas.

Los autores identificaron brechas en las tecnologías actuales para la detección de armas y exploraron un área emergente: la detección intraclase, que consiste en identificar tipos específicos de armas de fuego utilizadas en un ataque. Se analizaron y clasificaron fortalezas y debilidades de diversos algoritmos de aprendizaje clásico y

profundo, evaluando su desempeño en diferentes conjuntos de datos.

Los resultados indicaron que las técnicas de aprendizaje profundo superaron a las de aprendizaje clásico en términos de velocidad y precisión, destacando su potencial para futuras aplicaciones en la investigación de escenas del crimen y vigilancia automática.

Improving video surveillance systems in banks using deep learning techniques (Zahrawi and Shaalan, 2023)

Se presentó un marco para la detección temprana de armas utilizando sistemas avanzados de detección de objetos en tiempo real, como YOLO y SSD (Single Shot Multi-Box Detector). Además, se consideró la reducción de falsas alarmas como un aspecto clave para que el modelo fuera aplicable en situaciones de la vida real.

El sistema propuesto fue diseñado para cámaras de vigilancia en interiores de bancos, supermercados, centros comerciales y estaciones de servicio, entre otros. También se sugirió su uso en cámaras exteriores como un sistema preventivo para evitar robos armados.

Improving handgun detection through a combination of visual features and body pose-based data (Ruiz-Santaquiteria et al., 2023)

En este artículo se destacó la importancia de la detección temprana de objetos peligrosos, como armas de fuego, en imágenes de cámaras de circuito cerrado (CCTV) para reducir posibles daños. Se propuso un método novedoso basado en una arquitectura combinada que utilizó la estimación de poses corporales junto con características visuales de armas, logrando una detección más robusta en entornos de vigilancia.

La combinación de características de armas y poses corporales fue aplicada para superar limitaciones en casos donde las características visuales eran insuficientes, como cuando las armas eran pequeñas o poco visibles. Para la extracción de características visuales, se utilizaron tanto redes neuronales convolucionales (CNN) como arquitecturas basadas en transformadores.

El método fue evaluado en múltiples conjuntos de datos, donde se demostró que mejoró el desempeño de detectores basados en poses. Asimismo, se realizó un estudio

de ablación para analizar la contribución de la rama de procesamiento de poses y del filtro de falsos positivos.

Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications (Huszar et al., 2023)

Se abordó el desafío de analizar en tiempo real los grandes volúmenes de datos generados por cámaras de vigilancia. Para ello, se propuso un método automático de detección de violencia que utilizó redes con convoluciones 3D para capturar relaciones espaciales y temporales, apoyándose en un modelo preentrenado de reconocimiento de acciones.

El enfoque fue probado en conjuntos de datos públicos diversos, logrando una mejora del 2% en precisión frente a métodos avanzados, con menos parámetros y robustez ante artefactos de compresión comunes en sistemas de procesamiento remoto.

Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos (Ahmed et al., 2022)

En este trabajo la detección de armas en videos de vigilancia de cámaras CCTV representó un desafío importante debido al fácil acceso y posible mal uso de estas. Para enfrentar este problema, se desarrolló un sistema mejorado de detección de armas en tiempo real, aprovechando los avances en visión por computadora y detección de objetos, con el objetivo de tomar decisiones inteligentes que protejan a las personas en situaciones de peligro.

El sistema empleó el modelo Scaled-YOLOv4 con un dataset personalizado, alcanzando una precisión promedio (mAP) de 92.1 y una velocidad de procesamiento de 85.7 cuadros por segundo (FPS) en una GPU de alto rendimiento (RTX 2080TI). Además, el modelo fue optimizado para dispositivos edge de bajo costo, como el Jetson Nano, utilizando el optimizador TensorRT para reducir la latencia, aumentar el rendimiento y mejorar la privacidad.

Human pose estimation for mitigating false negatives in weapon detection in video-surveillance (Lamas et al., 2022)

Este trabajo presentó una metodología *top-down*, denominada WeDePE (*Weapon Detection over Pose Estimation*), que primero identificó las regiones de las manos guiándose por la estimación de poses humanas y luego analizó dichas regiones con

un modelo de detección de armas. Para optimizar la localización de cada mano, se introdujo un nuevo factor denominado Factor de Pose Adaptativo, que consideró la distancia del cuerpo respecto a la cámara.

Los experimentos demostraron que esta metodología fue más robusta que los enfoques *bottom-up* y los modelos avanzados existentes, tanto en escenarios de vigilancia en interiores como en exteriores.

ACF: An Armed CCTV Footage Dataset for Enhancing Weapon Detection (Hnoohom et al., 2022)

Para esta investigación se creó el dataset Armed CCTV Footage (ACF), compuesto por grabaciones simuladas de peatones armados con pistolas y cuchillos en diferentes escenarios. El estudio propuso una metodología basada en image tiling (división de imágenes) para mejorar la detección de armas pequeñas mediante aprendizaje profundo.

Los experimentos, realizados en un conjunto de datos público (Mock Attack), mostraron que el enfoque de tiling mejoró significativamente la precisión promedio (mAP), alcanzando valores hasta 10.22 veces mayores. En pruebas con el modelo SSD MobileNet V2, el dataset ACF procesado con tiling logró un mAP de 0.758 en la detección de pistolas y cuchillos.

Integrating Deep Learning-Based IoT and Fog Computing with Software-Defined Networking for Detecting Weapons in Video Surveillance Systems (Fathy and Saleh, 2022)

El trabajo abordó la necesidad de desarrollar técnicas inteligentes y adaptativas para mejorar la calidad del servicio (QoS) en aplicaciones de Internet de las Cosas (IoT), especialmente debido al aumento del tráfico multimedia durante la pandemia de COVID-19. Se investigó la integración de técnicas de aprendizaje profundo con arquitecturas de Redes Definidas por Software (SDN) para soportar aplicaciones sensibles al retraso en entornos IoT.

Se entrenaron y evaluaron múltiples modelos basados en aprendizaje profundo, eligiendo el de mayor rendimiento para integrarlo en un modelo de inteligencia artificial en el edge. Este sistema extrajo los primeros fotogramas detectados con armas para enviarlos rápidamente a las autoridades, permitiendo una detección temprana de crímenes mientras se redujo el tráfico en la red y el consumo de ancho

de banda.

La evaluación del modelo, realizada con el emulador Mininet, mostró mejoras significativas: un aumento del 75% en el rendimiento promedio, una reducción del 14.7% en el jitter medio y una disminución del 32.5% en la pérdida de paquetes, optimizando la QoS al programar dinámicamente la red según el tráfico y su destino.

Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities (Ingle and Kim, 2022)

El artículo abordó la dificultad de monitorear múltiples cámaras de vigilancia para detectar y clasificar armas (como pistolas y cuchillos) en tiempo real, especialmente en dispositivos con recursos computacionales limitados. Para solucionar esto, se propuso un método ligero basado en redes neuronales convolucionales, diseñado para clasificar, localizar y detectar armas de forma eficiente en entornos de tiempo real.

El modelo incluyó un clasificador multicategoría que diferenciaba entre objetos normales y anormales en los fotogramas de video. Los experimentos mostraron que el método alcanzó una precisión promedio (mAP) de 97.50% en los datasets ImageNet e IMFDB, 90.50% en el dataset Open Images, 93% en el dataset Olmos, y 90.7% en cámaras multivista. En dispositivos con recursos limitados, logró una precisión del 85.5% para la detección en entornos multivista, demostrando un rendimiento satisfactorio y eficiente para escenarios de videovigilancia.

A neural network aided attuned scheme for gun detection in video surveillance images (Manikandan and Rahamathunnisa, 2022)

El artículo abordó el uso de cámaras de circuito cerrado (CCTV) para la detección de objetos peligrosos en entornos de seguridad residencial y comercial. Se presentó un esquema denominado Attuned Object Detection Scheme (AODS), basado en redes neuronales convolucionales (CNN), para detectar y clasificar objetos peligrosos a partir de grabaciones de video.

El esquema utilizó las características de los objetos, extraídas y analizadas mediante CNN, para realizar la clasificación. Durante este proceso, se implementaron análisis basados en restricciones de características y una atenuación de las representaciones dimensionales para evitar errores en la detección de múltiples objetos. Además, se identificaron desviaciones en las representaciones dimensionales

para mejorar la precisión del modelo.

El rendimiento del esquema fue validado mediante métricas de precisión, exactitud y puntaje F1, logrando una mejora del 8.08% en la precisión, así como reducciones en el error y la complejidad de 7.47 y 8.23 puntos porcentuales, respectivamente, gracias al entrenamiento con conjuntos de datos externos. Se espera que en el futuro se implemente la clasificación de objetos basada en etiquetas.

An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos (Ullah et al., 2021)

Este artículo presentó un marco de reconocimiento basado en una Red Neuronal Convolucional (CNN) ligera y eficiente para videovigilancia con baja complejidad temporal, complementada con una red LSTM de atención residual para extraer características espaciales.

El modelo fue entrenado y validado usando los datasets UCF-Crime, UMN y Avenue, los cuales incluyen miles de fotogramas de eventos normales y anormales. Los resultados mostraron que el modelo superó a los métodos de vanguardia, con mejoras de precisión del 1.77%, 0.76% y 8.62% en cada uno de los datasets, respectivamente.

Anomaly Recognition from Surveillance Videos using 3D Convolution Neural Network (Maqsood et al., 2021)

En este artículo, los autores propusieron un enfoque para aprender características espacio-temporales mediante redes neuronales convolucionales tridimensionales (3D ConvNets). Utilizaron el dataset UCF-Crime, compuesto por 1900 videos, que incluían 950 normales y 810 anómalos, con clases como robo, arrestos con disparos, accidentes de tráfico y asaltos.

El procesamiento de imágenes consistió en separar los videos en fotogramas, redimensionarlos a 170 x 170 píxeles, normalizarlos, aplicarles aumentos espaciales y formar cubos 3D con una longitud fija. Estos cubos fueron empleados para capturar características espaciales y temporales en una arquitectura profunda optimizada, como se ilustra en la Figura 3.1.

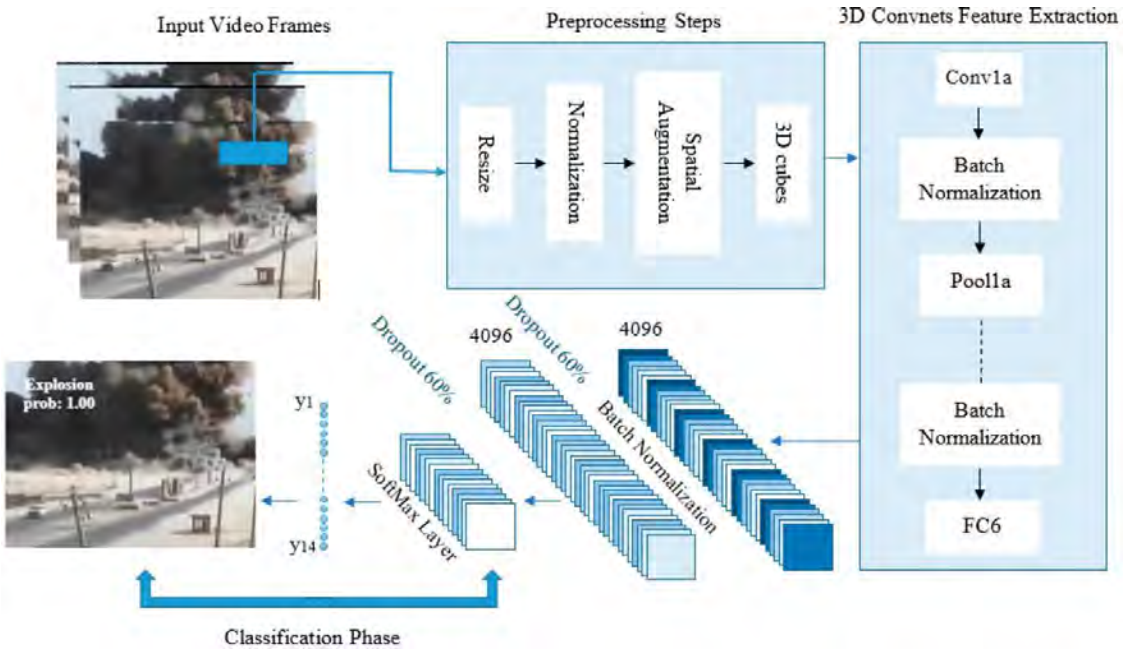


Figura 3.1 Diagrama de flujo para el reconocimiento de actividades anómalas (Maqsood et al., 2021)

Criminal Intention Detection at Early Stages of Shoplifting Cases by Using 3D Convolutional Neural Networks (Martínez-Mascorro et al., 2021)

Este artículo se centró en la prevención del delito de hurto en tiendas, para lo cual se llevó a cabo un modelado de comportamientos típicos que conducían a la comisión de un delito. Se implementó un modelo de Red Neuronal Convolucional 3D (3DCNN) para extraer características de video, logrando un 85.71% de precisión en la detección de comportamientos sospechosos en los distintos escenarios evaluados. Se describió el proceso de segmentación de un video basado en los momentos relacionados con comportamientos sospechosos, como se ilustró en la Figura 3.2.

Figura 3.2 Segmentación de video utilizando los momentos obtenidos del método de segmento de comportamiento Pre-Crimen (PCB) (Martínez-Mascorro et al., 2021)

Se identificó el instante en el que el delincuente aparecía por primera vez en el video, definido como el Momento de la Primera Aparición (FAM). El análisis de conductas sospechosas comenzaba a partir de este punto. Se detectó el momento en que el delincuente cometía el delito, conocido como el Momento del Crimen Estricto (SCM). Este momento contenía las pruebas necesarias para sustentar el delito. Entre el FAM y el SCM, se identificaba el momento en el que el delincuente comenzaba a actuar de manera sospechosa, denominado Momento del Crimen Integral (CCM), el cual iniciaba en cuanto se detectaba un comportamiento sospechoso en el video. Después del SCM, se ubicaba el instante en el que concluía el crimen, momento en el que todo parecía volver a la normalidad. Si el video comenzaba a partir de este punto, no se dispondría de evidencia de ningún crimen ocurrido previamente. Este momento se denominó el Momento de Regreso a la Normalidad (B2NM).

3.2 Discusión del Estado del Arte

Durante la revisión y análisis de los trabajos presentados, se identificó que la mayoría propone métodos para la prevención del crimen mediante la detección de movimientos sospechosos. Estos enfoques hacen uso de Redes Neuronales Convolucionales 3D y técnicas de aprendizaje profundo, complementados en su mayoría con técnicas de preprocesamiento de imágenes, como redimensionamiento y conversión a escala de grises.

En la literatura revisada, se encontraron investigaciones centradas principalmente en la detección automática de crímenes en centros comerciales, domicilios privados y pequeñas empresas.En su mayoría reutilizan modelos de Redes Neuronales. El dataset más utilizado es el UCF-Crime, que incluye 128 horas de video e imágenes clasificadas en categorías como abuso, robo, asalto, arresto y accidentes, entre otras. Las clases empleadas suelen ser comportamiento sospechoso y normal, o bien pre-crimen, crimen y post-crimen.

Sin embargo, no se encontró evidencia de datasets enfocados específicamente en la detección automática de delitos o armas en escenarios de carreteras. Esta carencia motivó la creación de un dataset especializado para este ámbito, con cuatro clases: normal, pre-crimen, crimen y post-crimen.

En la Tabla 3.1 se presentan el resultado de la revisión y comparación de los veinte artículos más relevantes para este tema de tesis; se agrupan los artículos más relevantes de acuerdo con las técnicas utilizadas para la detección de crímenes.

Tabla 3.1 Comparativa entre los artículos relevantes para el Estado del Arte

Referencia	**Datos utilizados**	**Temática**	**Resultados obtenidos**
Berardini et al. (2025)	Dataset WeaponSense, NVIDIA Jetson Nano	Detección de armas en video vigilancia mediante superresolución	Mejora del mAP en 10.2 puntos con YOLOSR respecto a YOLOv8-small, manteniendo eficiencia.
Yadav et al. (2025)	Dataset con 79,558 imágenes de armas	Detección de armas en tiempo real en condiciones de poca iluminación	Precisión del 91.75% y mAP del 92.15%.
Yadav et al. (2024)	Dataset de 15,367 imágenes y cinco videos oscuros	Detección de armas en condiciones de poca luz	Precisión del 95.50% y puntaje F1 del 93.41%.
Torregrosa-Domínguez et al. (2024)	NVIDIA GeForce RTX-3060, resolución de 720p	Detección de armas en tiempo real con objetos pequeños	Aumento del 13.23% en mAP, reducción del 71% en falsos positivos.
Amado-Garfias et al. (2024)	Videos de vigilancia, algoritmos heurísticos y modelos clásicos	Detección de personas armadas en tiempo real	Random Forest alcanzó precisión del 85.44% y F1 del 87.87%.
Berardini et al. (2024)	Cámara IP, NVIDIA Jetson Nano	Detección de armas pequeñas en dispositivos edge	mAP de 79.30% con velocidad de procesamiento de 5.10 FPS.

Tabla 3.1 Comparativa entre los artículos relevantes para el Estado del Arte

Referencia	**Datos utilizados**	**Temática**	**Resultados obtenidos**
Nadeem et al. (2024)	Dataset sintético WVD generado en GTA V	Detección de violencia usando datos sintéticos	Disponible en Kaggle, diseñado para expandir el corpus de investigación.
Yadav et al. (2023)	Diversos datasets de aprendizaje clásico y profundo	Estudio comparativo de métodos clásicos y profundos para detección	Las técnicas de aprendizaje profundo superaron a las clásicas en velocidad y precisión.
Zahrawi and Shaalan (2023)	Datos de cámaras de bancos, supermercados y estaciones de servicio	Detección de armas en interiores	Uso de YOLO y SSD, reducción de falsas alarmas clave para la aplicabilidad real.
Ruiz-Santaquiteria et al. (2023)	Múltiples datasets, estimación de poses y características visuales	Detección combinada de armas y poses corporales	Mejora significativa en la detección con evaluaciones en varios datasets.
Huszar et al. (2023)	Conjuntos de datos públicos	Detección de violencia en tiempo real con redes 3D convolucionales	Mejora del 2% en precisión frente a métodos avanzados con menos parámetros.
Ahmed et al. (2022)	Dataset personalizado, GPU RTX 2080TI	Detección de armas en tiempo real optimizada para dispositivos edge	mAP de 92.1 y velocidad de 85.7 FPS, optimizado con TensorRT.

Tabla 3.1 Comparativa entre los artículos relevantes para el Estado del Arte

Referencia	**Datos utilizados**	**Temática**	**Resultados obtenidos**
Lamas et al. (2022)	Estimación de poses humanas	Detección de armas guiada por estimación de manos	Metodología robusta para interiores y exteriores con poses humanas.
Hnoohom et al. (2022)	Dataset Armed CCTV Footage (ACF), Mock Attack	Detección de armas pequeñas mediante image tiling	Mejoras de hasta 10.22 veces en mAP en pruebas con SSD MobileNet V2.
Fathy and Saleh (2022)	Emulador Mininet, múltiples modelos de aprendizaje profundo	Integración de IoT, fog computing y SDN para detección de armas	Reducción del 14.7% en jitter y 32.5% en pérdida de paquetes, optimizando la QoS.
Ingle and Kim (2022)	Datasets ImageNet, IMFDB, Open Images, Olmos	Detección de objetos anómalos en tiempo real	mAP del 97.50% en ImageNet y 85.5% en dispositivos limitados.
Manikandan and Rahamathunnisa (2022)	Cámaras CCTV, análisis basado en CNN	Detección de objetos peligrosos en entornos residenciales y comerciales	Mejora del 8.08% en precisión con reducción de errores y complejidad.
Ullah et al. (2021)	Datasets UCF-Crime, UMN, Avenue	Reconocimiento de anomalías en videovigilancia con LSTM residual	Mejora de precisión del 1.77% al 8.62% dependiendo del dataset.

Tabla 3.1 Comparativa entre los artículos relevantes para el Estado del Arte

Referencia	**Datos utilizados**	**Temática**	**Resultados obtenidos**
Maqsood et al. (2021)	Dataset UCF-Crime	Reconocimiento de anomalías espacio-temporales	Captura de características 3D, mejora en precisión y robustez ante artefactos de compresión.
Martínez-Mascorro et al. (2021)	Modelado de comportamientos en videovigilancia	Detección temprana de intención delictiva en hurto	Precisión del 85.71% en detección de comportamientos sospechosos.

Capítulo 4

Metodología de Solución

En este trabajo, se desarrolló una metodología diseñada específicamente para abordar los desafíos inherentes a la detección automática de armas en escenarios dinámicos, como las carreteras. La propuesta combina técnicas avanzadas de aprendizaje profundo con estrategias de preprocesamiento de imágenes para mejorar la precisión y robustez del modelo en condiciones visuales adversas. Además, se implementó un enfoque basado en Redes Neuronales Convolucionales, específicamente una variante adaptada de YOLOv5, que fue reentrenada con un conjunto de datos especializado creado para este estudio. Esta metodología no solo permite la detección de objetos como armas fuego en tiempo real, sino que también integra herramientas para el etiquetado eficiente de datos y la implementación en dispositivos móviles, asegurando su aplicabilidad en escenarios reales de vigilancia y seguridad.

4.1 Propuesta de Solución

La propuesta de solución tiene dos elementos, un sistema de extracción y etiquetado para crear un repositorio de imágenes y una aplicación móvil orientada a la detección automática de armas en entornos de carreteras. Además, la creación de un *dataset* con 4 clases balanceadas de nombre CRIMEX. Como se muestra en la Figura 4.1, el sistema de extracción y etiquetado consta de los siguientes módulos:

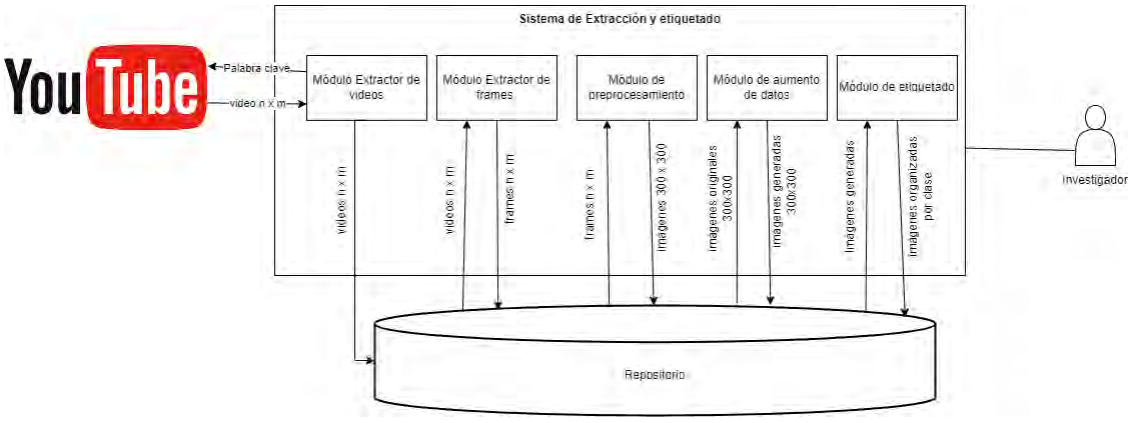


Figura 4.1 Arquitectura del Sistema de Extracción y Etiquetado

- **Módulo de Extracción de Videos:** Se utilizó la plataforma YouTube como fuente de videos relevantes mediante la búsqueda por palabras clave. Los 15 videos seleccionados por el usuario se almacenaron en un repositorio.
- **Módulo de Extracción de Frames:** Este módulo extrajo frames a una tasa de 30 frames por segundo, que luego fueron almacenados en el mismo repositorio.
- **Módulo de Preprocesamiento:** Se aplicaron técnicas de redimensión y recorte a las imágenes originales, obteniendo imágenes con un tamaño uniforme de 300×300 píxeles.
- **Módulo de Aumento de Datos:** Se implementaron filtros como espejo, rotación ($\pm 10^{\circ}$) y ajustes de contraste para aumentar el volumen y la diversidad del dataset.
- **Módulo de Etiquetado:** Las imágenes procesadas fueron organizadas en carpetas según su clase (arma corta, arma larga, arma blanca).

Por otro lado, la aplicación móvil cuenta con los siguientes componentes:

- **Módulo de Monitoreo y Visualización:** Captura frames desde la cámara del dispositivo móvil y muestra las imágenes con el cuadro delimitador del objeto detectado.
- **Módulo Clasificador:** Realiza la detección y registra información relevante (precisión, fecha, hora e imagen del objeto detectado) en una base de datos local.

Estos módulos son detallados gráficamente en la Figura 4.2.

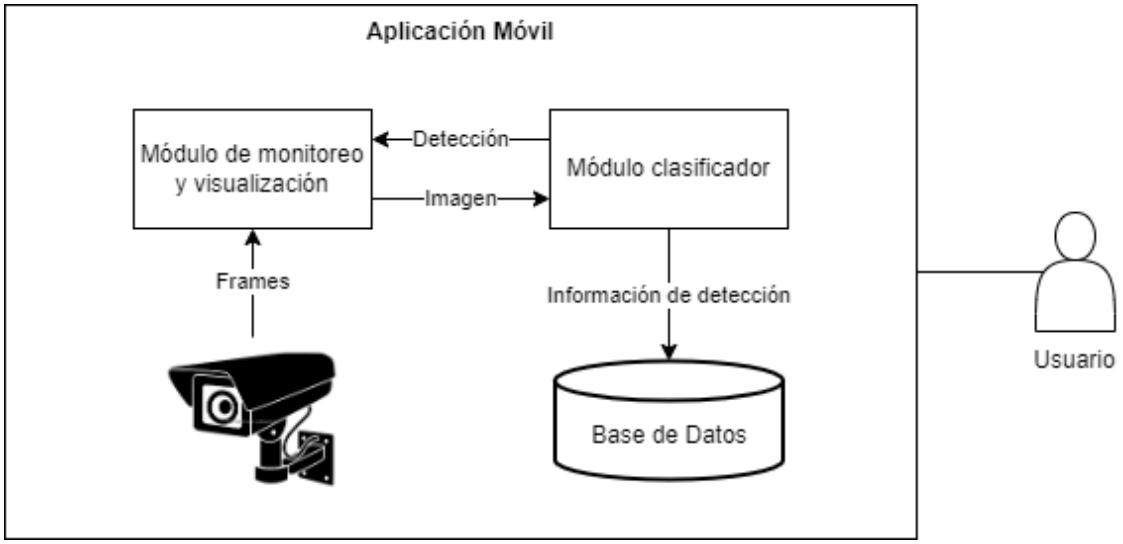


Figura 4.2 Arquitectura de la Aplicación Móvil

Para la creación del dataset utilizado en el reentrenamiento de los modelos YOLO y SSD, se descargó un subconjunto de 5,078 imágenes para la detección de armas del repositorio de *Data Science and Computational Intelligence* (Andaluz, 2020), el cual es de acceso libre. Estas imágenes contienen armas en exhibición, es decir, sin la presencia de personas portando las armas. Como parte del proceso, se extrajeron las imágenes que mostraban las armas en exhibición. Durante la etapa de etiquetado de las imágenes, se utilizó el programa LabelImg, desarrollado por la comunidad de Python, que permite delimitar el objeto a detectar mediante un cuadro. Para complementar este conjunto de datos, se incluyeron 379 imágenes adicionales de CriMex, seleccionando solo aquellas con la mejor calidad visual. Aunado a esto, la cantidad de imágenes no era lo suficiente para el proceso de reentranamiento, para esto, se utilizó el sistema de extracción y etiquetado, aplicando filtros de rotación, traslación, color y espejo a las imágenes, logrando aumentar de manera considerable el número de imágenes. El conjunto de imágenes etiquetado está disponible en el enlace proporcionado en la nota al pie de página [1].

La arquitectura de funcionamiento de la aplicación móvil se basa en varios procesos interconectados. La entrada son frames obtenidos a través de la cámara nativa del dispositivo móvil, con un tamaño de 320 x 320 píxeles. Primero, se realiza un preprocesamiento para obtener la región de interés. El siguiente paso consiste en la extracción de características mediante una CNN llamada Darknet, la cual recibe un frame como entrada y devuelve un valor dentro del intervalo de 0-1. A partir de este resultado, se clasifica si existe o no una detección. En caso de no detectarse,

[1] https://drive.google.com/drive/folders/1vGt4wFIoZ8Xqd-zVUoW6IWN78aksLSkh?usp=drive_link

se procede a procesar el siguiente frame. Si se detecta un objeto, se delimita y se obtienen los datos de interés, como la precisión de confianza, la fecha, la hora y la imagen con el marco delimitador del objeto. Estos datos se registran en una base de datos local. Los procesos descritos se ilustran en la Figura 4.3.

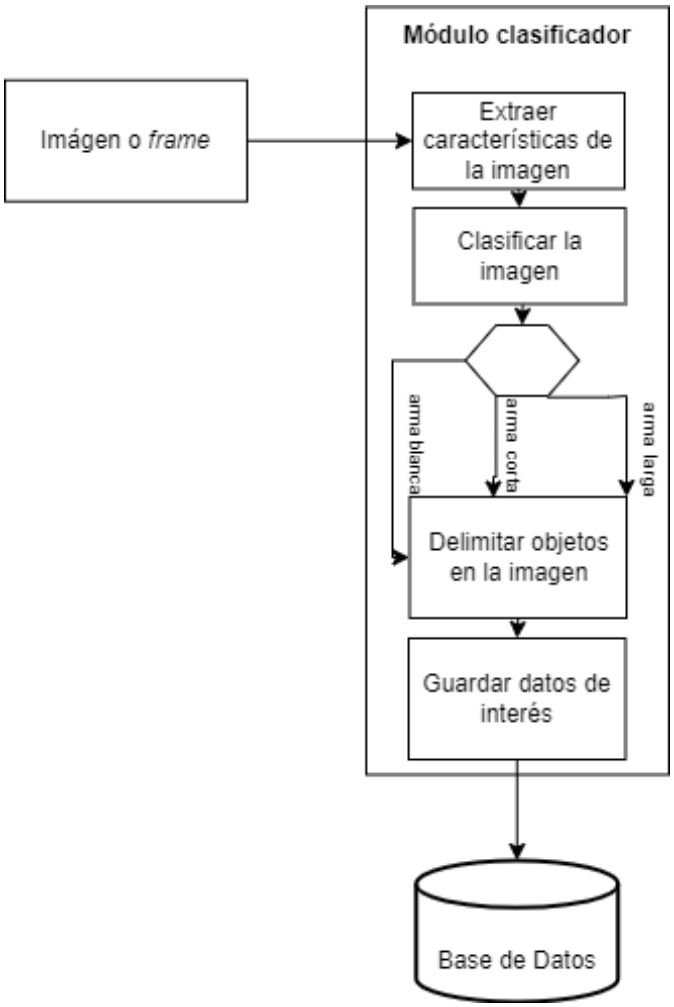


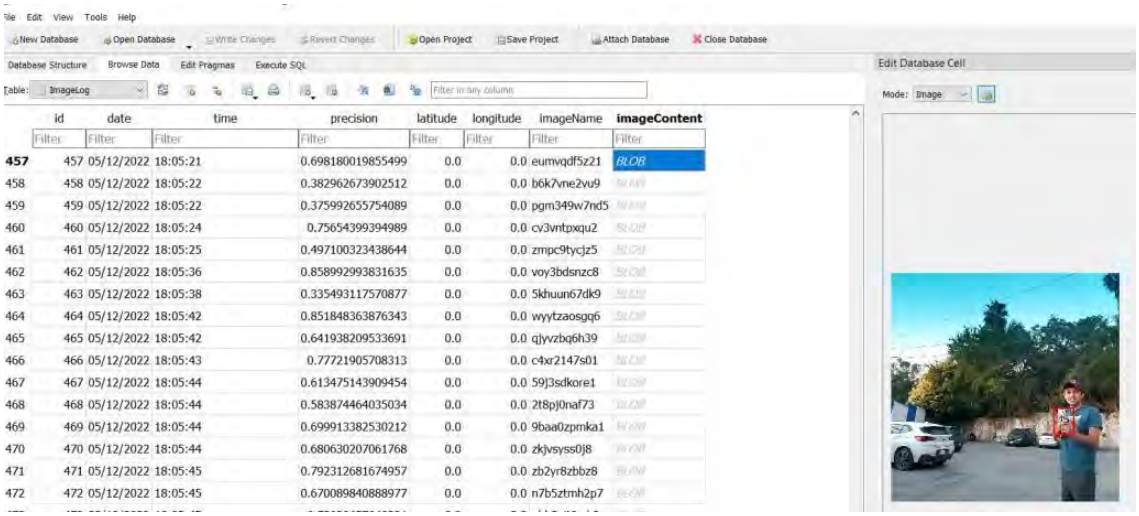
Figura 4.3 Arquitectura de funcionamiento de la aplicación móvil

4.2 Implementación por Computadora de la Propuesta de Solución

La aplicación móvil se desarrolló en el entorno Android Studio utilizando Java como lenguaje de programación y SQLite para la gestión de la base de datos. La aplicación fue optimizada para procesar frames en tiempo real y almacenar detecciones localmente. Adicionalmente, se implementó el modelo reentrenado en Yolo V5 en TensorFlow Lite, lo que permitió su adaptación a dispositivos móviles, las etiquetas

utilizadas durante la detección son acorta = arma_corta, alarga = arma_larga y navaja = arma_blanca.

En la Figura 4.4 se muestra la GUI de los registros que se almacena en la Base de Datos, así como la imagen que fue capturada durante la detección, los campos de latitud y longitud fueron considerados por si en un futuro llega a implementarse un sistema de alerta.



	id	date	time	precision	latitude	longitude	imageName	imageContent
457	457	05/12/2022	18:05:21	0.698180019855499	0.0	0.0	eumvqdf5z21	BLOB
458	458	05/12/2022	18:05:22	0.382962673902512	0.0	0.0	b6k7vne2vu9	
459	459	05/12/2022	18:05:22	0.375992655754089	0.0	0.0	pgm349w7nd5	
460	460	05/12/2022	18:05:24	0.75654399394989	0.0	0.0	cv3vntpxqu2	
461	461	05/12/2022	18:05:25	0.497100323438644	0.0	0.0	zmpc9tycjz5	
462	462	05/12/2022	18:05:36	0.858992993831635	0.0	0.0	voy3bdsnzc8	
463	463	05/12/2022	18:05:38	0.335493117570877	0.0	0.0	5khuun67dk9	
464	464	05/12/2022	18:05:42	0.851848363876343	0.0	0.0	wyytzaosgq6	
465	465	05/12/2022	18:05:42	0.641938209533691	0.0	0.0	qjyvzbq6h39	
466	466	05/12/2022	18:05:43	0.77721905708313	0.0	0.0	c4xr2147s01	
467	467	05/12/2022	18:05:44	0.613475143909454	0.0	0.0	59j3sdkore1	
468	468	05/12/2022	18:05:44	0.583874464035034	0.0	0.0	2t8pj0naf73	
469	469	05/12/2022	18:05:44	0.699913382530212	0.0	0.0	9baa0zpmka1	
470	470	05/12/2022	18:05:44	0.680630207061768	0.0	0.0	zkjvsyss0j8	
471	471	05/12/2022	18:05:45	0.792312681674957	0.0	0.0	zb2yr8zbbz8	
472	472	05/12/2022	18:05:45	0.670089840888977	0.0	0.0	n7b5ztmh2p7	

Figura 4.4 GUI de base de datos de detección de armas

Capítulo 5

Experimentación

En esta sección se presentan los experimentos realizados, siguiendo una combinación de la Guía de Tesis e Informes de la Universidad de Oxford y la estructura propuesta por Jedlitschka y Pfahl para reportes de experimentación. Además de describir el proceso experimental, se detallarán los inconvenientes y observaciones que surgieron durante el mismo. También se incluirán gráficas que ilustran los resultados obtenidos.

5.1 Primer Experimento: validación del dataset con modelos CNN

En esta etapa de experimentación, se utilizaron modelos de CNN existentes para evaluar la viabilidad de realizar una clasificación binaria entre *crimen* y *normal* en el conjunto de datos de crímenes en carretera. Para ello, se emplearon los modelos EfficientNet_B0, MobileNet V2, VGG16 y AlexNet. Se crearon tres subconjuntos de imágenes a partir del dataset CriMex: un conjunto de entrenamiento con 46,080 imágenes, un conjunto de validación con 11,520 imágenes y un conjunto de prueba con 6,398 imágenes. Todos los subconjuntos estaban balanceados entre las clases Normal y Crimen. Cabe destacar que se aseguró que en cada subconjunto estuvieran representadas imágenes de los 15 diferentes videos extraídos previamente de YouTube. Las pruebas se realizaron empleando los equipos de cómputo mostrados en la Tabla 5.1

Tabla 5.1 Especificaciones de los equipos utilizados para las pruebas.

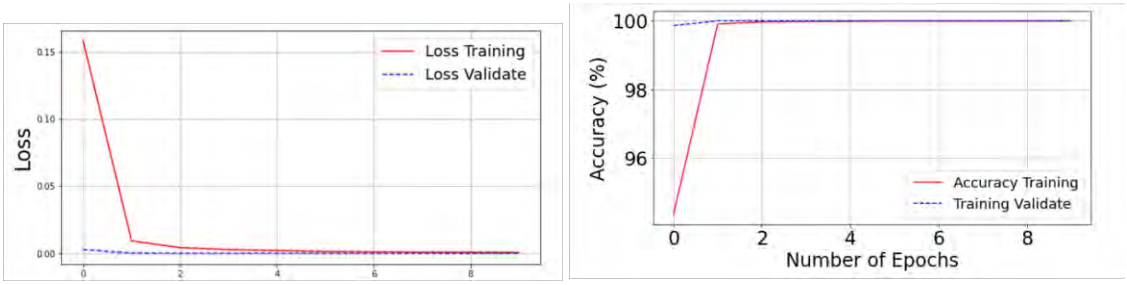
Marca	Procesador	Tarjeta Gráfica	Memoria RAM
Asus	Intel Core i7 de 11va generación	4GB de video NVIDIA GeForce RTX 3050 Ti	8GB
Lenovo Legion Y530	Intel Core i5 de 8va generación	2GB de video NVIDIA GeForce GTX 1050	No especificada
Google Colab Pro	No aplicable (entorno cloud)	GPU Tesla P100-PCIE	16GB

Con el objetivo de realizar una comparativa entre el rendimiento de las CNN previamente mencionadas, todas ellas fueron implementadas con la siguiente configuración:

- Tasa de aprendizaje: 0.001
- Técnica de optimización: ADAM
- Número total de épocas para el entrenamiento: 10
- Número de lotes: de 5 a 10, dependiendo de las capacidades del equipo de cómputo utilizado en cada prueba y de la cantidad de capas en cada red.

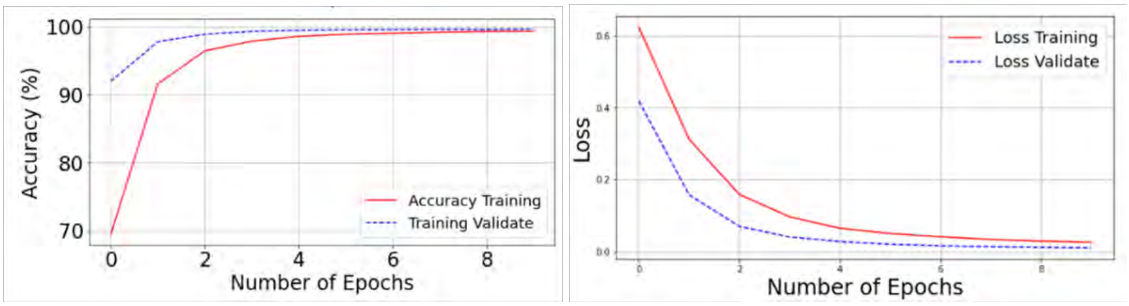
Se eligió esta configuración para todos los modelos con el objetivo de evaluar y seleccionar aquellos que presentaran buenos resultados, para posteriormente mejorar su rendimiento en la clasificación. Se aplicó la técnica de transferencia de aprendizaje, utilizando los pesos de la base de datos ImageNet, que contiene 1.2 millones de imágenes distribuidas en 1000 clases diferentes. Esto permitió utilizar una red neuronal convolucional preentrenada capaz de detectar patrones básicos. La principal ventaja de este enfoque es la posibilidad de realizar configuraciones mínimas, especialmente cuando se dispone de una base de datos pequeña para el entrenamiento.

En la gráfica de la Figura 5.1 se muestran los resultados obtenidos durante la fase de entrenamiento del modelo EfficientNetB0, ilustrando el porcentaje de pérdida y la precisión de clasificación. De la misma manera, la gráfica de la Figura 5.2 presenta los resultados del modelo VGG16 durante el entrenamiento.



(a) Pérdida (b) Precisión

Figura 5.1 Resultados de entrenamiento de modelo EfficientNetB0 1er experimento



(a) Error (b) Precisión

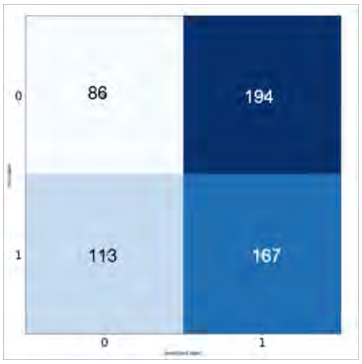
Figura 5.2 Resultados de entrenamiento de modelo VGG16 1er experimento

Los resultados obtenidos para este primer experimento durante la etapa de prueba se ilustran en la Tabla 5.2, los cuáles en primera instancia se ven muy buenos con tres de los modelos, a diferencia del modelo VGG16.

Tabla 5.2 Resultados de desempeño de los modelos CNN.

Modelo CNN	Exactitud (%)	Precisión (%)	F1-Score (%)
AlexNet	98.00%	96.00%	97.96%
EfficientNet_B0	99.34%	99.75%	99.35%
MobileNet_V2	97.34%	96.81%	97.33%
VGG16	69.46%	100%	76.61%

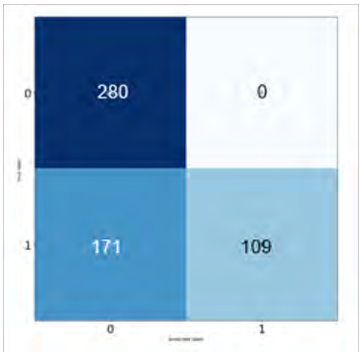
El análisis de los resultados iniciales mostró un comportamiento prometedor, por lo que se decidió realizar pruebas con otro subconjunto de imágenes para evaluar su desempeño. En la Figura 5.3 se presenta la matriz de confusión correspondiente a los resultados obtenidos al validar el modelo EfficientNetB0 con un subconjunto de imágenes diferente. Este subconjunto contiene imágenes del mismo tipo que las utilizadas en la primera prueba, pero se observa que los resultados fueron menos favorables.



Métrica	Crimen	Normal
Precisión	0.4322	0.4626
Recall	0.3071	0.5964
F1-Score	0.3591	0.5211

Figura 5.3 Resultados de clasificación del modelo EfficientNetB0 1er experimento

A pesar de que los resultados de la prueba anterior no fueron favorables, se decidió probar con el modelo VGG16 para observar y analizar su comportamiento. Los resultados obtenidos se ilustran en la Figura 5.4, donde se puede apreciar que los resultados tampoco fueron los esperados. Como resultado, se realizaron nuevas configuraciones y se exploraron otros modelos.



Métrica	Crimen	Normal
Precisión	0.6467	1.0000
Recall	1.0000	0.4536
F1-Score	0.7854	0.6241

Figura 5.4 Resultados de clasificación del modelo VGG16 1er experimento

Como primera observación, al utilizar equipos de cómputo local, se presentó un inconveniente relacionado con el traslado de los dispositivos, lo que interrumpió los entrenamientos debido al apagado o suspensión de los equipos. Además, para preservar la vida útil de la batería, era necesario mantener los dispositivos siempre conectados a la corriente eléctrica. Otro desafío fue que, al estar el equipo ocupado realizando el entrenamiento, se hacía necesario cancelar otras tareas secundarias, lo que provocaba retrasos en el avance de otras actividades. Debido a estas limitaciones, se tomó la decisión de utilizar el entorno de Google Colab para realizar todos los entrenamientos.

En cuanto a Google Colab, una de las principales limitaciones fue la necesidad de contar con acceso a Internet constante durante el proceso de entrenamiento. La pérdida de la conexión provocaba la interrupción del entrenamiento, lo que obligaba a reiniciarlo. Este problema se vio agravado por la duración de cada entrenamiento, que tenía al menos 10 horas de duración.

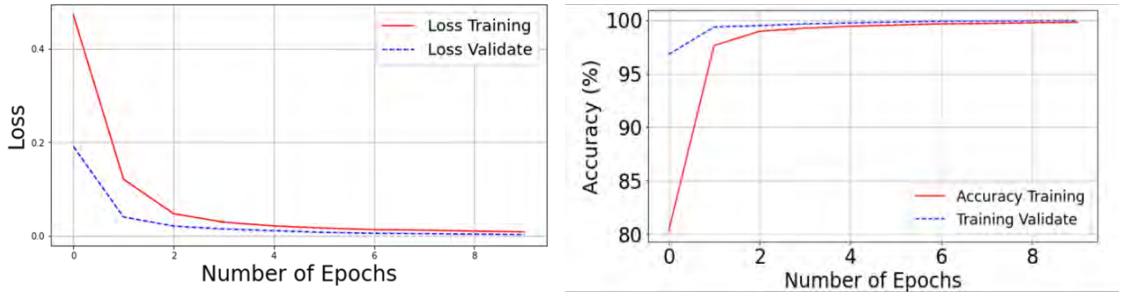
Por último, los resultados obtenidos con el conjunto de datos de entrenamiento, que contenía imágenes similares a las del dataset, fueron positivos. Sin embargo, al validar el modelo con un segundo subconjunto, se observó que la precisión disminuía. Esto sugirió la necesidad de explorar otras configuraciones para el entrenamiento, analizar el conjunto de datos más a fondo o incluso considerar la posibilidad de probar otros modelos para mejorar los resultados.

5.2 Segundo Experimento: ajustes al dataset

Al revisar las imágenes del repositorio, se observó que muchas de ellas eran muy similares, lo que se debió al tener una alta tasa de extracción de frames, al momento de procesar los videos obtenidos con el sistema de Extracción y etiquetado. Debido a esto, se realizó una sustracción de imágenes casi idénticas, lo que permitió obtener una mayor diferencia entre las imágenes y mejorar la capacidad de la CNN para discriminar durante el entrenamiento. Como resultado de este proceso, se obtuvo lo siguiente: un 70% del dataset para el entrenamiento equivalente a 5,592 imágenes, un 20% de 1,598 imágenes para la validación y un subconjunto de prueba de 800 imágenes, todos con un balance entre las clases Normal y Crimen.

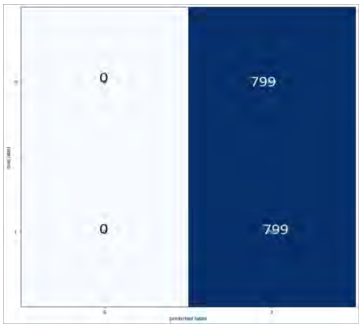
A raíz de los resultados obtenidos en el primer experimento, se decidió aplicar distintas configuraciones a cada modelo, variando el número de épocas, la tasa de aprendizaje y la técnica de optimización. Cabe destacar que se ajustó el número de épocas para cada modelo en función de la mejora en el entrenamiento. Para todos los modelos se utilizó el optimizador ADADELTA y una tasa de aprendizaje de 0.01, trabajando en el entorno Colab Pro.

El modelo Inception V3 mostró buenos resultados durante la fase de entrenamiento. En la gráfica mostrada en la Figura 5.5 se ilustran las primeras 8 épocas, las cuales evidencian un claro sobreajuste de la CNN. Este fenómeno indicó que no era posible mejorar más los resultados, ya que el modelo alcanzó un 100% de precisión. Se utilizaron 50 épocas como valor inicial y se probó el modelo, obteniendo un rendimiento deficiente en la clasificación. La Figura 5.5 muestra la matriz de confusión correspondiente.



(a) Error (b) Precisión

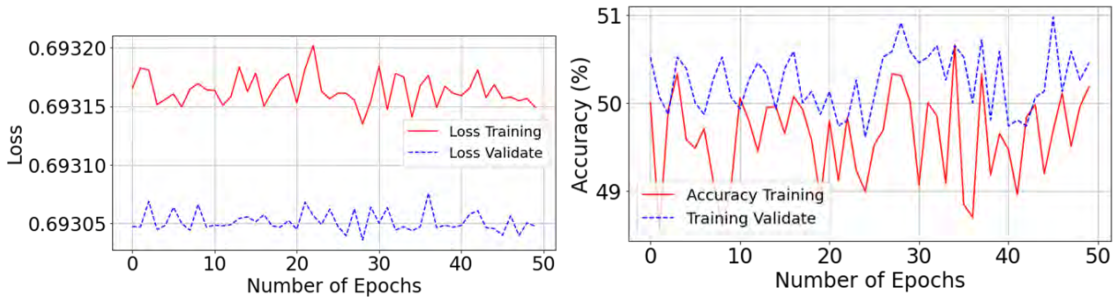
Figura 5.5 Resultados de entrenamiento del modelo Inception V3 2do experimento



Métrica	Crimen	Normal
Precisión	0.0000	0.5000
Recall	0.0000	1.0000
F1-Score	0.0000	0.6667
Support	799	799

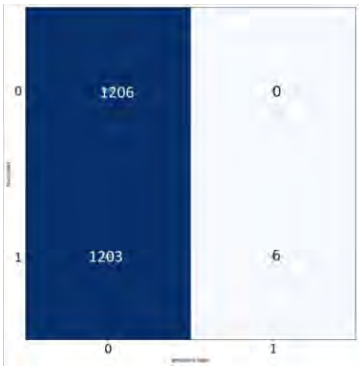
Figura 5.6 Resultados de clasificación del modelo Inception V3 2do experimento

En el modelo VGG19 se utilizaron 50 épocas de entrenamiento. A lo largo de este proceso, la precisión no superó el 60%, y la tasa de pérdida no mostró mejoras significativas. Los resultados de la prueba de clasificación no fueron los esperados, ya que se observó una inclinación hacia una sola clase en las predicciones. Esta tendencia se ilustra en la gráfica de la Figura 5.7, y los resultados de la prueba se muestran en la Figura 5.8, los cuales evidencian un desbalance hacia una única clase.



(a) Error (b) Precisión

Figura 5.7 Resultados de entrenamiento de modelo VGG19 2do experimento

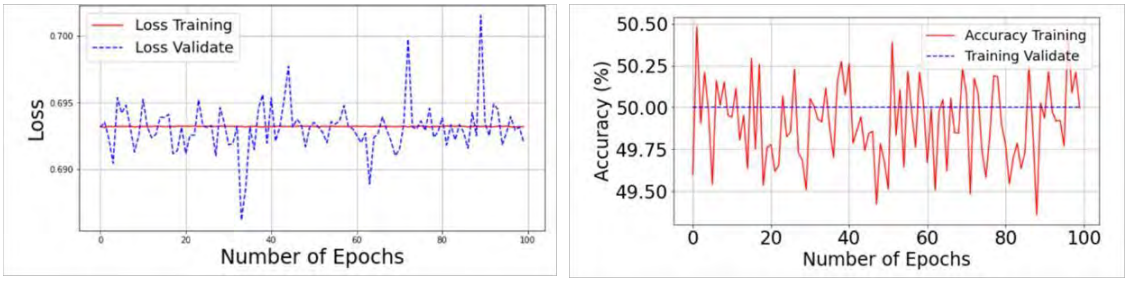


Métrica	Crimen	Normal
Precisión	0.5006	1.0000
Recall	1.0000	0.0050
F1-Score	0.6672	0.0099
Support	1206	1206

Figura 5.8 Resultados de clasificación del modelo Inception V3 2do experimento

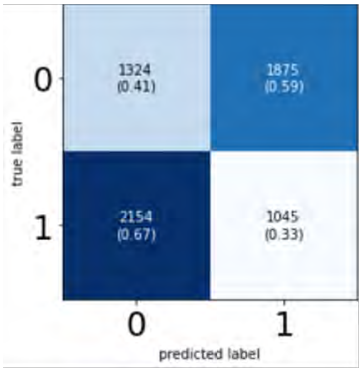
El modelo Xception fue entrenado con 100 épocas. Se decidió no aumentar el número de épocas debido a que la precisión no mostró mejoría, como se observa en

la gráfica de la Figura 5.9. En este experimento, los resultados de la clasificación no fueron favorables, ya que las predicciones se comportaron de manera aleatoria, similar al lanzamiento de una moneda al aire. Esta tendencia se ilustra en la Figura 5.10.



(a) Error (b) Precisión

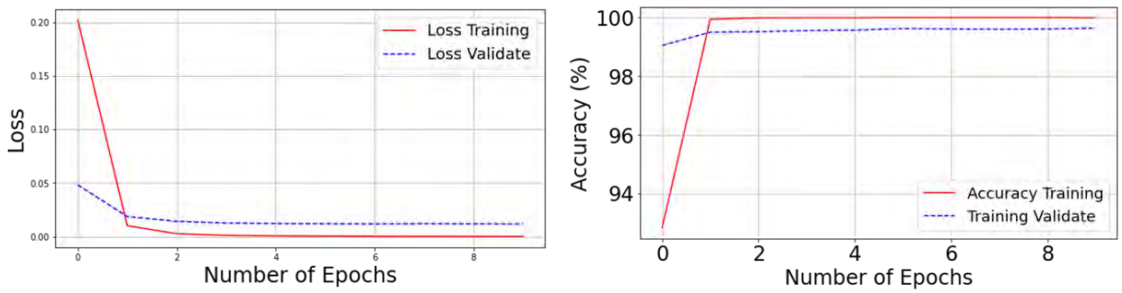
Figura 5.9 Resultados de entrenamiento de modelo Xception 2do experimento



Métrica	Crimen	Normal
Precisión	0.38	0.36
Recall	0.41	0.33
F1-Score	0.40	0.34
Support	3199	3199

Figura 5.10 Resultados de clasificación del modelo Inception V3 2do experimento

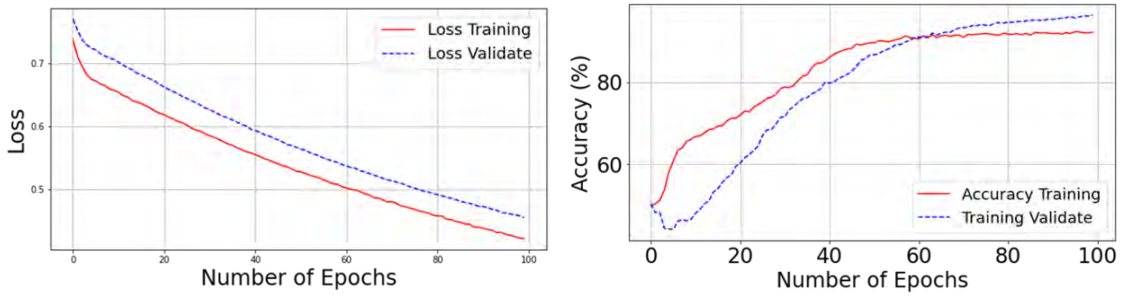
En el modelo ResNet50, durante la décima época se alcanzó un 100% de precisión y casi 0 de error en la pérdida, lo que indica un claro sobreajuste de la CNN. Esto sugirió que no se obtendrán mejoras adicionales, por lo que se decidió continuar con otro modelo. Este resultado se ilustra en la gráfica de la Figura 5.11.



(a) Error (b) Precisión

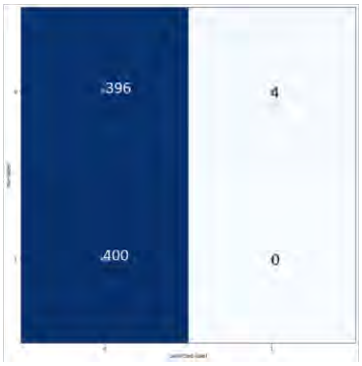
Figura 5.11 Resultados de entrenamiento de modelo ResNet50 2do experimento

El modelo VGG16 mostró un buen desempeño en cuanto a precisión y error de pérdida, como se ilustra en la Figura 5.12. Sin embargo, durante la etapa de clasificación, no logró una clasificación adecuada para ambas clases, tal como se muestra en la matriz de confusión de la Figura 5.13.



(a) Error (b) Precisión

Figura 5.12 Resultados de entrenamiento de modelo VGG16 2do experimento



Métrica	Crimen	Normal
Precisión	0.4575	0.0000
Recall	0.9900	0.0000
F1-Score	0.6622	0.0000
Support	400	400

Figura 5.13 Resultados de clasificación del modelo Inception V3 2do experimento

En la Tabla 3 se presenta un resumen de los resultados obtenidos para cada CNN. En las gráficas mostradas previamente, se puede observar que algunos modelos alcanzaron valores cercanos al 100% de precisión durante la etapa de entrenamiento. Sin embargo, estos valores cayeron por debajo del 70% en la etapa de prueba, lo que resulta demasiado bajo para cumplir con el objetivo de lograr una precisión igual o superior al 80%.

Tabla 5.3 Resultados de prueba de modelos de CNN experimento 2

Modelo CNN	Cant. de épocas	Exactitud (%)	Precisión (%)	F1-Score (%)
Inception v3	50	100	100	100
VGG19	50	50.16	66.72	58.14
Xception	100	37.03	41.39	39.66
ResNet50	50	100	100	100
VGG16	100	49.59	66.22	57.76

Los resultados obtenidos con los diferentes modelos de CNN no fueron favorables, lo que lleva a la conclusión de que no es viable utilizar estos modelos para una tarea de clasificación simple en escenas de carretera. Esto puede deberse a la naturaleza del dataset, que presenta una gran diversidad de escenarios, lo que introduce una serie de factores que interfieren con los resultados, como puentes, árboles, nubes, autos, entre otros. Además, este tipo de dataset carece de un fondo definido, a diferencia de escenarios cerrados, como el de un cajero automático, donde el ambiente siempre es constante. En ese tipo de entorno, lo único que varía son los elementos en movimiento o, en algunos casos, el cambio de color de las paredes, lo que facilita el aprendizaje de patrones. Debido a estas dificultades, se optó por recurrir a un enfoque de detección de objetos.

5.3 Tercer Experimento: modelos de detección de objetos

Este experimento surge debido a los resultados insatisfactorios obtenidos en una tarea de clasificación simple en escenas de delitos en carretera. En esta nueva prueba, se utilizaron modelos de detección de objetos para identificar escenas de crimen y escenas normales, considerando como "escena de crimen" aquellas en las que aparecieran personas portando armas blancas o de fuego. Las armas blancas incluyen elementos como navajas, cuchillos y machetes, mientras que las armas de fuego comprenden pistolas, metralletas, rifles, entre otras. Se eligieron los modelos SSD (Mehta, 2021) y YOLOv5 (Jocher et al., 2022) debido a que son de los más utilizados y con mejores resultados. Aunque YOLOv5 es algo más lento en comparación con SSD, se destaca por una mejor discriminación en la detección de objetos.

El dataset utilizado, descrito previamente, contiene al menos 5,000 imágenes que presentan objetos de tres clases: arma_blanca, arma_corta y arma_larga. Esta

clasificación se hizo para reflejar la diferencia de patrones entre los objetos y facilitar un mejor aprendizaje por parte de la CNN. La diferencia entre las tres clases se ilustra en la Figura 5.14.

Figura 5.14 Ejemplo ilustrativo del contenido del dataset de entrenamiento para el sistema de detección de armas

Se empleó la herramienta LabelImg, que se describe en la sección 3 y se puede acceder a través del enlace proporcionado en (Sell, 2022), para realizar el proceso de etiquetado del conjunto de imágenes. Este proceso se llevó a cabo en dos etapas, ya que cada modelo de detección de objetos requiere un formato específico de archivo para su reentrenamiento. Es decir, cada modelo necesita una estructura de etiquetas y anotaciones distinta, lo que obligó a realizar el etiquetado de las imágenes dos veces, ajustando los archivos según los requisitos de cada modelo. Este procedimiento garantizó que los datos fueran compatibles con los diferentes modelos y optimizó el proceso de entrenamiento, permitiendo que se pudieran comparar los resultados obtenidos con cada uno de ellos.

5.3.1 Modelo SSD

Se implementó el modelo SSD utilizando la Red Neuronal Convolucional MobileNet V2, con un total de 60,000 pasos durante el proceso de entrenamiento y aplicando transferencia de aprendizaje sobre el dataset COCO17. Las imágenes de entrada utilizadas para el entrenamiento tuvieron una resolución de 320 × 320 píxeles. El entrenamiento y la prueba se llevaron a cabo en el entorno de Colab, donde se observaron resultados positivos en cuanto a la reducción del error de pérdida. Estos resultados se ilustran en la Figura 5.15.

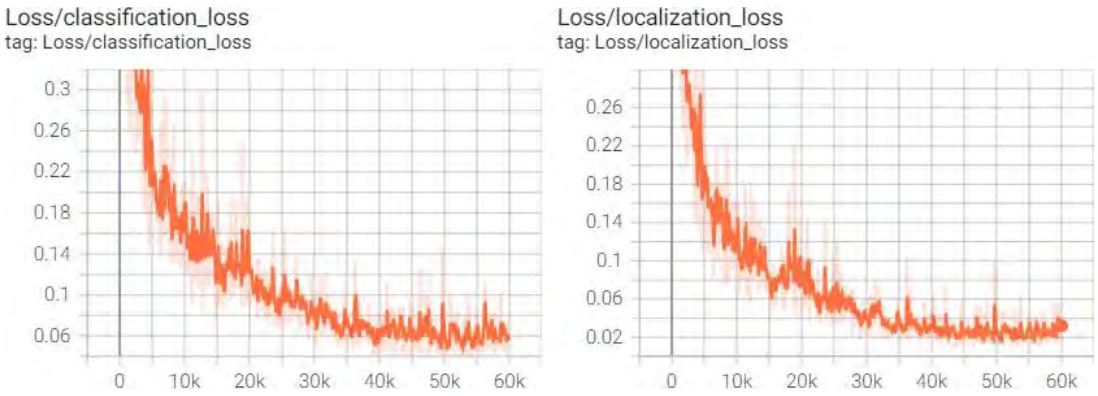


Figura 5.15 Resultados de pérdida del error del modelo SSD en etapa de entrenamiento

Durante la etapa de prueba del modelo, se logró una precisión del 94% en la detección del objeto. En la Figura 5.16 se presenta un ejemplo de los resultados obtenidos, donde se muestra una imagen en buenas condiciones con la presencia de un arma blanca.

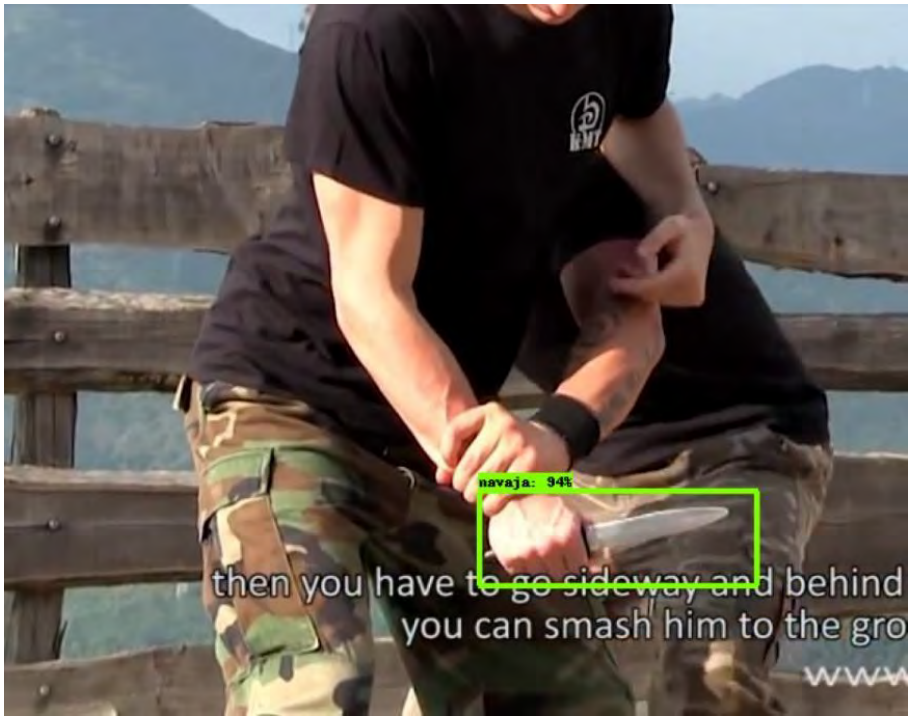


Figura 5.16 Prueba del modelo SSD con una imagen

En resumen, el modelo SSD con MobileNet V2 demostró un rendimiento notable, alcanzando una precisión del 94% en la detección de objetos durante la etapa de prueba. Los resultados obtenidos fueron satisfactorios, especialmente al detectar escenas con armas blancas en condiciones favorables, lo que valida la efectividad del modelo para esta tarea.

5.3.2 Modelo YOLO V5

Se implementó el modelo YOLOv5 utilizando la red neuronal convolucional entrenada durante 500 épocas con la configuración preestablecida, que incluye el optimizador SGD y una tasa de aprendizaje de 0.01. Además, se aplicó transferencia de aprendizaje usando los pesos de COCO128 (Lin et al., 2014). Los resultados de las curvas de precisión y recall durante el entrenamiento se muestran en la Figura 5.17, donde se observa un buen desempeño en la detección de objetos de las tres clases.

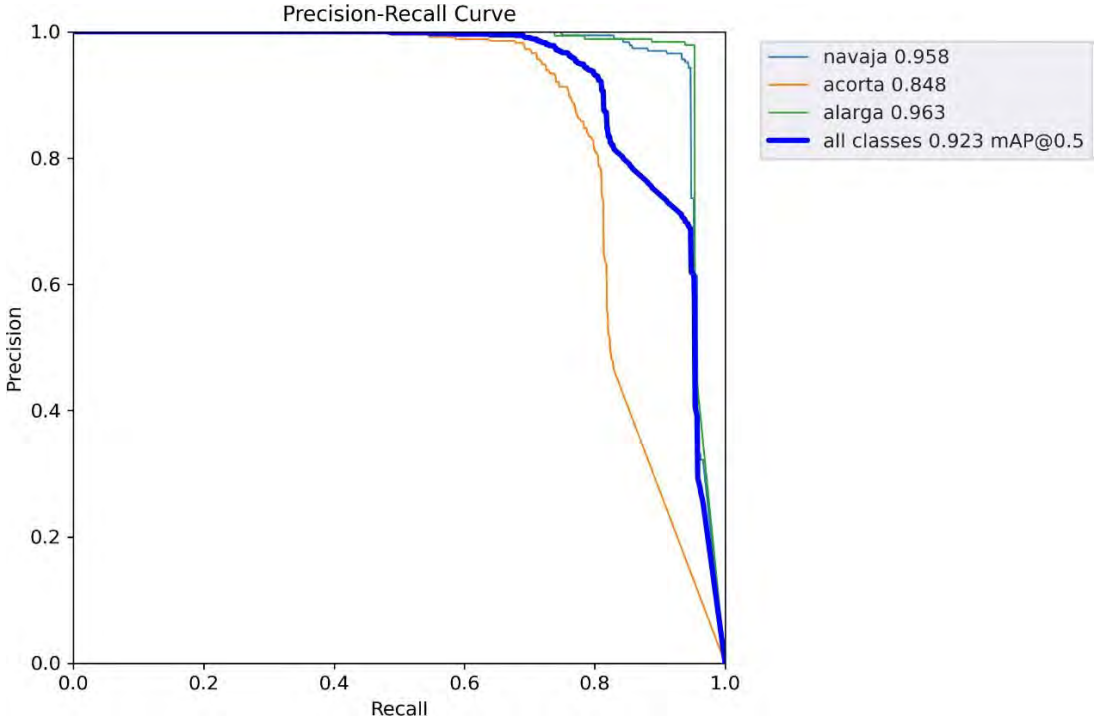


Figura 5.17 Resultado de las curvas P-R correspondiente al entrenamiento de YOLO V5

En la Figura 5.18 se presenta el comportamiento de la pérdida de error y la precisión durante el entrenamiento. Cabe aclarar que los datos mostrados no corresponden a las 500 épocas completas, ya que el proceso de entrenamiento se vio interrumpido por la pérdida de conexión a Internet. Después de cada interrupción, los pesos se actualizaron desde los últimos valores calculados, reiniciando el entrenamiento desde ese punto.

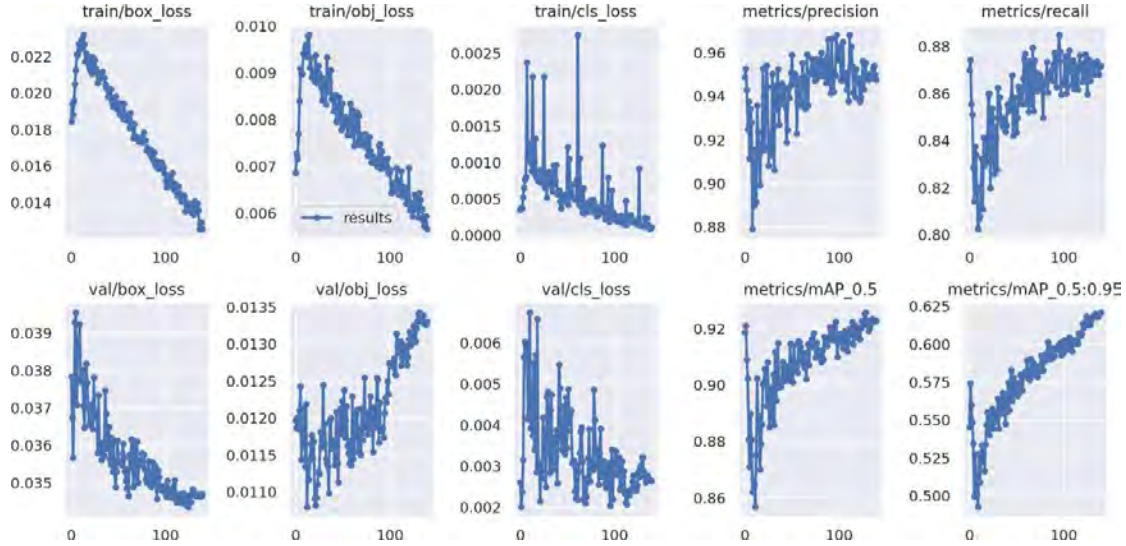


Figura 5.18 Resultados de pérdida del error y precisión del modelo YOLO V5 en etapa de entrenamiento

En las Figura 5.19 se ilustran los resultados de las detecciones durante la etapa de validación del modelo. Se observan resultados con una precisión superior al 90% y una correcta clasificación de los objetos. Cabe destacar que la precisión mostrada representa el porcentaje de la Intersección sobre la Unión (IoU) del objeto. A mayor precisión, mejor es el resultado del enmarcado.

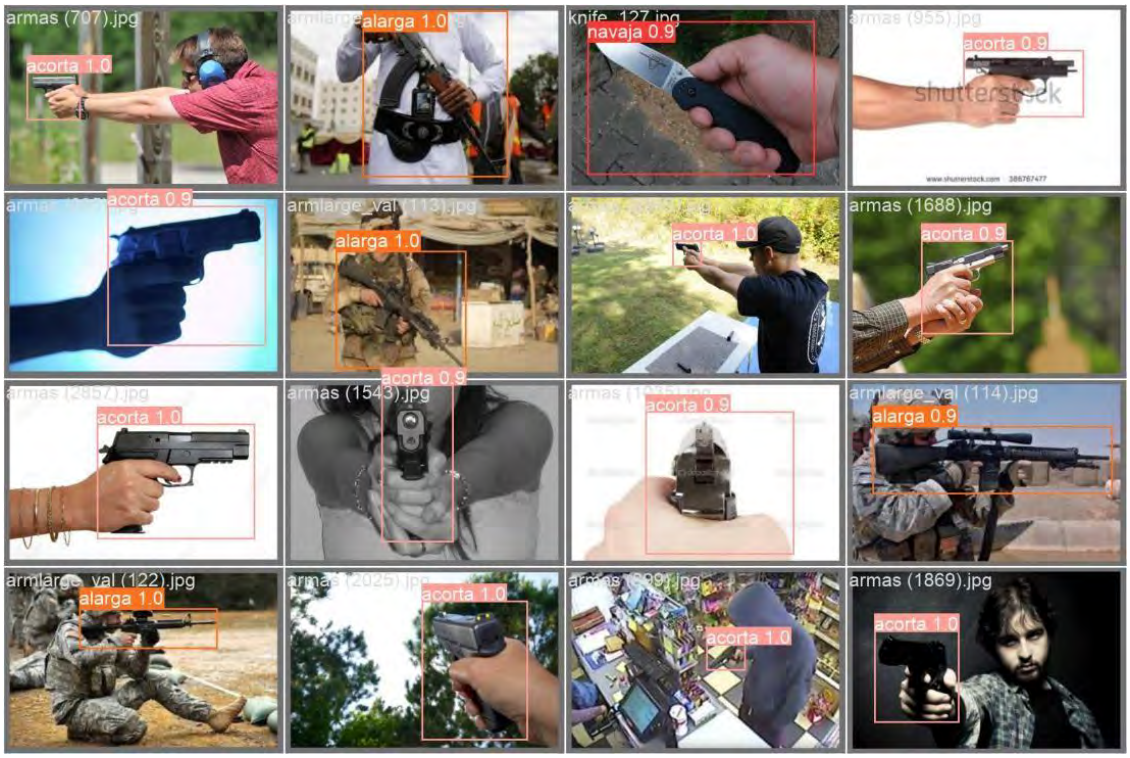


Figura 5.19 Imagenes ilustrativas de la detección con YOLO en la etapa de validación

En la Figura 5.20 y 5.21 se muestran ejemplos de detección de armas a partir de videos grabados desde el interior de un vehículo en un estacionamiento con buenas condiciones ambientales. Para las pruebas se utilizó una pistola de juguete, y los resultados obtenidos variaron entre un 80% y un 98% de precisión en el enmarcado de las armas detectadas. Los videos utilizados en las pruebas están disponibles en (Cortes Ramirez, 2022b) YoloV5Custom/runs/detect/exp9 y exp10.

Figura 5.20 Imagen ilustrativa de la detección con YOLO en videos grabados en estacionamiento del CENIDET 1

Figura 5.21 Imagen ilustrativa de la detección con YOLO en videos grabados en estacionamiento del CENIDET 2

5.4 Informe de Pruebas de Software

Los transportistas recorren diferentes tipos de rutas, que van desde los centros logísticos hasta los puntos de entrega de pedidos. Según el Instituto Mexicano del Transporte (del Transporte, 2016), se pueden identificar dos ámbitos distintos en el autotransporte de carga: el urbano y el interurbano.

En el primer caso, los trayectos suelen transitar por vialidades urbanas, que en la Norma Oficial Mexicana PROY-NOM-034-SCT2-2003 (de la Federación, 2004) se clasifican como Carreteras Tipo B, Tipo C o Tipo D. Estas conforman las redes primaria, secundaria y alimentadora para el servicio de comunicación interestatal y municipal, con longitudes relativamente cortas, estableciendo conexiones con la red secundaria. Además, se pueden recorrer caminos de terracería, definidos como "vías secundarias abiertas a la circulación vehicular y sin pavimento". En el caso interurbano, se trata de trayectos por autopistas. Según la norma mencionada, las carreteras Tipo ET o Tipo A corresponden a las autopistas, que forman parte de los ejes de transporte establecidos por la Secretaría de Comunicaciones y Transportes. Este ámbito también puede incluir tramos de carreteras Tipo B, especialmente aquellos habilitados como vías rápidas.

Considerando lo anterior, se decidió enfocar los casos de prueba en trayectos por autopistas, carreteras, caminos de terracería y estacionamientos que simulen áreas de salida y llegada de centros logísticos. Además, para garantizar la seguridad, en los trayectos por autopistas, solo se evaluaría el funcionamiento de la aplicación en condiciones normales.

5.4.1 Casos de prueba

El propósito de este plan de pruebas es presentar los resultados finales de las pruebas funcionales del modelo implementado a través de una aplicación móvil denominada CriMex. Esta aplicación está diseñada para detectar la presencia de armas en un ambiente controlado. Las pruebas de la aplicación se basó en el estándar ISO/IEC/IEEE 29119 (Ali and Yue, 2015). En la Tabla 5.4 se detallan los casos de prueba realizados.

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 01	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma blanca en terracería.	• El usuario se encuentra dentro de un vehículo estacionado en una terracería. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una navaja o machete. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una navaja o machete hacia el usuario. 4. La aplicación detecta un arma blanca y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar una arma blanca y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 02	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma corta en terracería.	• El usuario se encuentra dentro de un vehículo estacionado en una terracería. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una pistola de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una pistola hacia el usuario. 4. La aplicación detecta un arma corta y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar una arma corta y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 03	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma larga en terracería.	• El usuario se encuentra dentro de un vehículo estacionado en una terracería. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una metralleta de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una metralleta de fantasia hacia el usuario. 4. La aplicación detecta el arma larga y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar una arma larga y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 04	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma blanca en un estacionamiento.	• El usuario se encuentra dentro de un vehículo en un estacionamiento. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una navaja o machete. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una navaja o machete hacia el usuario. 4. La aplicación detecta un arma blanca. 5. El usuario detiene la aplicación móvil.	Detectar un arma blanca y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 05	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma corta en un estacionamiento.	• El usuario se encuentra dentro de un vehículo en un estacionamiento. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una pistola de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una pistola hacia el usuario. 4. La aplicación detecta un arma corta y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar un arma corta y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 06	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma larga en un estacionamiento.	• El usuario se encuentra dentro de un vehículo en un estacionamiento. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una metralleta de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una metralleta hacia el usuario. 4. La aplicación detecta un arma larga y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar un arma larga y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 07	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma blanca en una vía urbana.	• El usuario se encuentra dentro de un vehículo estacionado en una vía urbana. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una navaja o machete. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una navaja o machete hacia el usuario. 4. La aplicación detecta un arma blanca y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar un arma blanca y almacenar los datos correspondientes en la base de datos.	Pendiente

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 08	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma corta en una vía urbana.	• El usuario se encuentra dentro de un vehículo estacionado en una vía urbana. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una pistola de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección. 3. El usuario comienza a transitar la vía urbana. 4. El voluntario intercepta el vehículo acercándose por la derecha y apuntando una pistola hacia el usuario. 5. La aplicación detecta un arma corta y registra la detección. 6. El usuario detiene la aplicación móvil.	Detectar un arma corta y almacenar los datos correspondientes en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 09	Prueba funcional de caja negra diseñada para comprobar que el sistema puede detectar una situación de delito con un arma larga en una vía urbana.	• El usuario se encuentra dentro de un vehículo estacionado en una vía urbana. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El voluntario porta una metralleta de fantasía. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El voluntario se acerca apuntando con una metralleta hacia el usuario. 4. La aplicación detecta un arma larga y registra la detección. 5. El usuario detiene la aplicación móvil.	Detectar un arma larga y almacenar los datos correspondientes en la base de datos.	Éxito
CP 10	Prueba funcional de caja negra diseñada para comprobar si el sistema no detecta armas en una situación normal en una terracería.	• El usuario se encuentra dentro de un vehículo estacionado en una terracería. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El usuario conduce al menos 1km en terracería. 4. El usuario detiene la aplicación móvil.	No detectar ningún arma y no almacenar ningún dato de detección en la base de datos.	Éxito

Tabla 5.4 Casos de prueba de funcionalidad de la Aplicación Móvil

ID	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 11	Prueba funcional de caja negra diseñada para comprobar si el sistema no detecta armas en una situación normal en un estacionamiento.	• El usuario se encuentra dentro de un vehículo en un estacionamiento. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El usuario detiene la aplicación móvil.	No detectar ningún arma y no almacenar ningún dato de detección en la base de datos.	Éxito
CP 12	Prueba funcional de caja negra diseñada para comprobar si el sistema no detecta armas en una situación normal en vía urbana.	• El usuario se encuentra dentro de un vehículo conduciendo en una vía urbana. • El dispositivo móvil está montado en un soporte para celular orientado hacia el frente del vehículo. • El ambiente presenta buenas condiciones de iluminación y clima.	1. El usuario inicia la aplicación. 2. La aplicación móvil comienza el proceso de detección de objetos. 3. El usuario conduce al menos 1km en vía urbana. 4. El usuario detiene la aplicación móvil.	No detectar ningún arma y no almacenar ningún dato de detección en la base de datos.	Éxito

5.4.2 Resultados de los Casos de Prueba

Se lograron completar 11 de los 12 casos de prueba, con éxito en la detección o no detección de un arma, mientras que 1 pruebas quedaron pendientes. La localización seleccionada para la realización de los casos de prueba fueron definidas de la siguiente forma:

Ubicaciónes

1. Interior internado palmira s/n, col. Palmira, C.P. 62490, Cuernavaca, Morelos
2. P.º de La Reforma 285, Lomas de Cuernavaca, 62584 Tres de Mayo, Mor.

Dispositivos móviles

1. Samsung a72, 6GB RAM, 128GB almacenamiento, cámara de 64mpx

Los resultados de las pruebas funcionales reflejan un buen indicador para el problema de detección de casos de robo a transportistas en carretera. En las escenas clasificadas como "Normales", las pruebas realizadas en zonas controladas fueron exitosas, ya que, en ausencia de armas, no se detectó ningún objeto. Estos datos no se almacenaron en la base de datos para evitar el almacenamiento innecesario de información. En la Figura 5.22 se ilustran los resultados obtenidos durante la etapa de funcionalidad de la aplicación móvil. En esta primera etapa, se logró detectar el objeto correctamente, y los datos de detección, junto con la imagen, fueron insertados en la base de datos de manera adecuada.

(a) Caso de Prueba 05 (b) Caso de Prueba 06

(c) Caso de Prueba 08 (d) Caso de Prueba 09

Figura 5.22 Ejemplos de Funcionalidad de Aplicación Móvil

Las pruebas realizadas están disponibles en (Cortes Ramirez, 2022b), donde se incluyen fragmentos de videos que muestran claramente la escena de crimen. Además, los registros generados por la aplicación durante la etapa de prueba están almacenados en (Cortes Ramirez, 2022a).

Durante la ejecución, inicialmente se utilizó un nivel de confianza de 0.8, lo que afectó negativamente la detección durante la fase de prueba. Este indicador refleja la precisión con la que se enmarca el objeto y también influye en el descarte de objetos que no fueron enmarcados correctamente. El nivel de confianza por defecto es 0.25, pero para este proyecto se ajustó a un valor final de 0.3. Esta decisión se tomó porque el objetivo no es necesariamente enmarcar el objeto de manera óptima, sino detectar la presencia del arma. Esta configuración se ilustra en la Figura 5.23.

Figura 5.23 Ilustración de precisión de detección de armas

La conclusión que se puede extraer de lo anterior es que, al ajustar el nivel de confianza del modelo de detección, se optimizó el balance entre precisión y la capacidad de identificar la presencia del objeto (en este caso, un arma). El valor inicial de 0.8 resultó demasiado estricto y afectó la detección durante la prueba, por lo que se decidió reducirlo a 0.3. Este ajuste permitió que el modelo no solo enmarcara los objetos de manera más flexible, sino que también se centrara en la detección de la presencia del arma, en lugar de buscar una precisión perfecta en el enmarcado. Esto demuestra que, en contextos de detección, a veces es más importante asegurar la presencia de un objeto relevante que la perfección en su localización exacta.

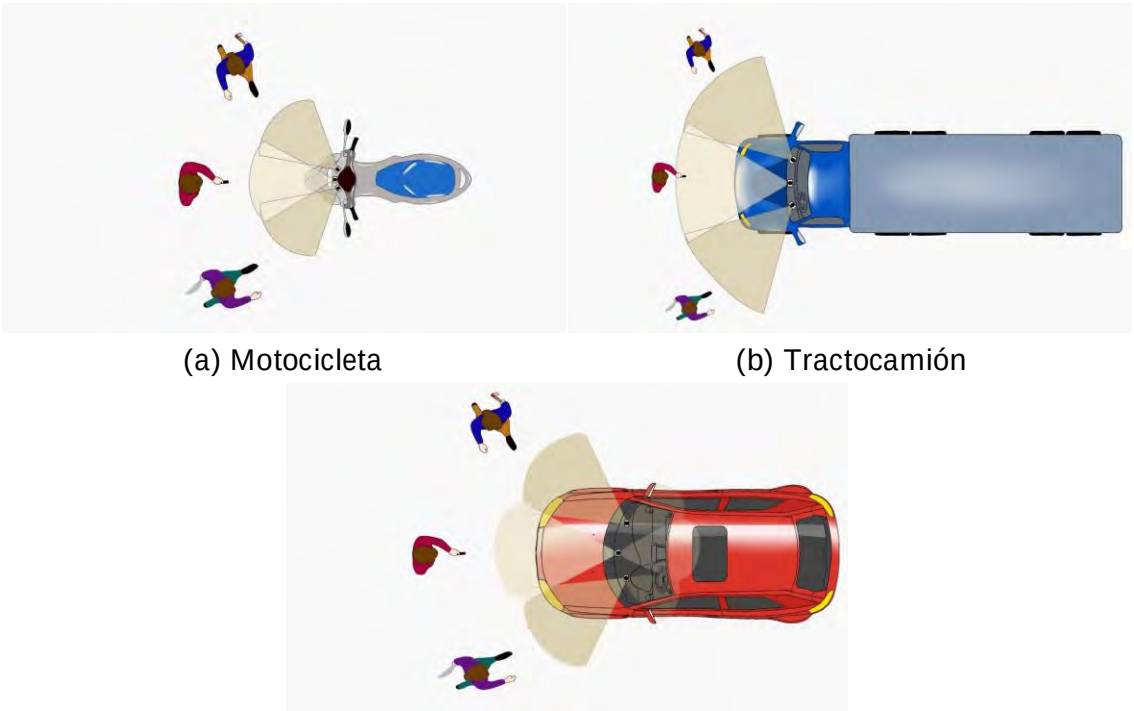
5.4.3 Anomalías

Durante la ejecución de los casos de prueba en zonas controladas, como estacionamientos, la aplicación mostró un buen desempeño al identificar correctamente la presencia de armas. Sin embargo, en las pruebas realizadas en zonas urbanas, que son escenarios no controlados, la aplicación detectó vehículos y otros objetos como si fueran armas. Esto se debe a la gran variedad de objetos que presentan patrones similares a los de un arma, lo que ocurre porque las CNNs se basan en el conocimiento empírico adquirido durante el entrenamiento.

Las cámaras de los dispositivos móviles suelen tener un ángulo de visión de 60°, lo que significa que en una situación de delito, el sospechoso podría no ser captado por la aplicación móvil. Para mitigar este problema, se propone el esquema presentado en la Figura 5.24, que combina hasta tres dispositivos móviles con la aplicación en

funcionamiento. Este enfoque busca reducir los puntos ciegos, con cada dispositivo orientado en una de las siguientes direcciones:

- F-LI: Frente-Lateral Izquierdo del vehículo
- F: Frente del vehículo
- F-LD: Frente-Lateral Derecho del vehículo



(a) Motocicleta (b) Tractocamión

(c) Automóvil

Figura 5.24 Diseño de implementación de cámaras en distintos vehículos

Aunque la aplicación mostró un buen desempeño en zonas controladas, presentó dificultades en escenarios urbanos no controlados debido a la similitud de patrones entre objetos y armas. Además, las limitaciones del ángulo de visión de las cámaras móviles pueden provocar que el sospechoso no sea captado. Para solucionar esto, se propone el uso de tres dispositivos móviles orientados en diferentes direcciones, lo que reduciría los puntos ciegos y mejoraría la detección.

Capítulo 6

Conclusiones

6.1 Objetivos y alcances logrados

Detalles sobre el cumplimiento de los objetivos planteados para el desarrollo de esta investigación se presentan en la Tabla 6.1.

Tabla 6.1 Cumplimiento de objetivos

Objetivo	Estado	Evidencia del cumplimiento del objetivo
Implementar y evaluar un modelo de aprendizaje profundo para el reconocimiento de crímenes	Cumplido	Se tiene como producto un modelo de DL reentrenado de nombre DarkNet, basado en el algoritmo de detección de objetos YOLOv5. Disponible en (Cortes Ramirez, 2022c)
Crear un dataset de imágenes y videos de situaciones sospechosas y violentas en carreteras, con los descriptores y las etiquetas correspondientes	Cumplido	Se cuenta con un repositorio de imágenes dividido en clases, un repositorio etiquetado de armas y un conjunto de videos. Disponible en (Cortes Ramirez, 2022a)
Desarrollar un sistema para la identificación de comportamientos sospechosos y normales para la detección de armas en escenas de videos, reutilizando una red neuronal convolucional que se adapte a esta necesidad	Cumplido	Se tiene como producto una aplicación móvil para la detección de armas en tiempo real, que almacena las detecciones en una base de datos local. Disponible en (Cortes Ramirez, 2023)

6.2 Resultados de la investigación

6.2.1 Productos

Durante el desarrollo de esta investigación se obtuvieron los siguientes productos:

1. **Repositorio CRIMEX**
   Repositorio de imágenes con 4 clases: normal, pre-crimen, crimen y post-crimen, con un total de 53,881 imágenes balanceadas. Además, en las imágenes se encuentran objetos etiquetados pertenecientes a las clases arma_corta, arma_larga y arma_blanca Acceso al repositorio

2. **Sistema para la extracción de frames**
   Herramienta para la extracción de frames, con capacidad para ajustar la tasa de extracción según las necesidades. Acceso al sistema

3. **CriMex Image Generator**
   Sistema que permite aplicar filtros de recorte, redimensión, espejo y rotación para aumentar los datos del dataset. Acceso al generador

4. **Artículo científico**
   Publicación titulada *Detección automática de delitos en sistemas de videovigilancia: una Revisión Sistemática de la Literatura*, presentada en la 7ª Jornada de Ciencia y Tecnología del Cenidet. Acceso al artículo

5. **Registro ante Indautor del repositorio CRIMEX**
   Certificación como coautor del repositorio CRIMEX, dividido en las clases normal, crimen, pre-crimen y post-crimen.

6.2.2 Aportaciones

Las aportaciones obtenidas con esta investigación son:

- **Creación de un dataset especializado**
  Se creó un conjunto de datos único enfocado en crímenes en carreteras, con clases específicas y balanceadas, llenando un vacío en la literatura existente.

- **Metodología para escenarios específicos**
  Se realizaron experimentos controlados excluyendo condiciones adversas y se desarrolló un sistema de etiquetado replicable.

- **Desarrollo de una solución práctica mediante una aplicación móvil** Se desarrolló una aplicación móvil que facilita la detección de armas en tiempo real, con almacenamiento local de resultados.

6.3 Conclusiones

La revisión de los trabajos relacionados permitió identificar que la mayoría se enfocaron en la detección de humanos en espacios cerrados, como cajeros automáticos, centros comerciales y viviendas. Estos entornos se caracterizan porque los frames solo presentan cambios significativos cuando aparece un humano o un animal. En estos estudios se implementaron técnicas de aprendizaje profundo, principalmente redes neuronales convolucionales y LSTM, logrando precisiones entre el 80 % y el 99 % en la detección de casos sospechosos de crimen.

Como principal hallazgo, se verificó que no fue viable aplicar clasificación simple con Redes Neuronales Convolucionales en escenas de delitos en carreteras, debido a los drásticos cambios visuales presentes en todos los frames analizados, lo que dificulta la identificación de patrones consistentes para la clasificación.

En este trabajo se desarrolló un modelo para la detección de objetos como armas blancas y armas de fuego, acompañado de un repositorio de imágenes etiquetadas previamente. Además, se propuso una arquitectura específica para la detección de casos sospechosos de crimen en carreteras. El modelo entrenado se implementó en una aplicación móvil, logrando cumplir con los objetivos planteados en esta tesis.

Los experimentos realizados fueron diseñados considerando las limitaciones identificadas, creando escenarios con los descriptores necesarios para la aparición de casos sospechosos y excluyendo escenas nocturnas, lluviosas o con condiciones climáticas adversas. La aplicación móvil permitió identificar aspectos clave para trabajar con modelos de redes neuronales convolucionales en la detección de objetos en tiempo real, demostrando su ventaja en términos de facilidad de uso, ya que solo se requiere acceso a un dispositivo móvil. El modelo reentrenado destacó por su capacidad de reutilización en diferentes contextos, a diferencia de las tareas de clasificación simple, que tienden a obtener buenos resultados solo en los entornos específicos para los que fueron entrenados.

En conclusión, este trabajo demostró que es posible detectar eventos asociados a casos de robo a transportistas mediante la identificación de armas con un rango de

precisión del 80% al 98%. Además, el modelo desarrollado puede ser adaptado para otros escenarios.

6.4 Trabajo futuro

Como línea de trabajo futuro, se propone expandir las capacidades del modelo desarrollado para incluir la detección automática de placas vehiculares en escenas de crimen, utilizando técnicas avanzadas de reconocimiento óptico de caracteres (OCR) integradas con redes neuronales convolucionales. Esta funcionalidad permitiría identificar y almacenar las placas en una base de datos en tiempo real, facilitando su posterior análisis por parte de las autoridades. Además, se sugiere explorar la adaptación del modelo para condiciones adversas, como escenas nocturnas o bajo condiciones climáticas desfavorables, mediante el uso de técnicas de mejora de imágenes o redes neuronales entrenadas con datos específicos para estos entornos. Estas mejoras podrían fortalecer aún más la aplicabilidad del sistema en escenarios de vigilancia complejos y mejorar la precisión en contextos diversos.

Referencias

Ahmadi, S., Beyhaghi, H., Blum, A., and Naggita, K. (2021). The strategic perceptron. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 6–25.

Ahmed, S., Bhatti, M. T., Khan, M. G., Lövström, B., and Shahid, M. (2022). Development and optimization of deep learning models for weapon detection in surveillance videos. *Applied Sciences*, 12(12):5772.

Ali, S. and Yue, T. (2015). Formalizing the iso/iec/ieee 29119 software testing standard. In *2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS)*, pages 396–405. IEEE.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.

Amado-Garfias, A. J., Conant-Pablos, S. E., Ortiz-Bayliss, J. C., and Terashima-Marín, H. (2024). Improving armed people detection on video surveillance through heuristics and machine learning models. *IEEE Access*.

Andaluz, I. (2020). Detección de armas open data. https://dasci.es/es/transferencia/open-data/deteccion-de-armas/. Accessed: 2025-01-18.

Berardini, D., Migliorelli, L., Galdelli, A., Frontoni, E., Mancini, A., and Moccia, S. (2024). A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras. *Multimedia Tools and Applications*, 83(7):19109–19127.

Berardini, D., Migliorelli, L., Galdelli, A., and Marín-Jiménez, M. J. (2025). Edge artificial intelligence and super-resolution for enhanced weapon detection in video surveillance. *Engineering Applications of Artificial Intelligence*, 140:109684.

Bhatti, M. T., Khan, M. G., Aslam, M., and Fiaz, M. J. (2021). Weapon detection in real-time cctv videos using deep learning. *Ieee Access*, 9:34366–34382.

Bieder, F., Sandkühler, R., and Cattin, P. C. (2021). Comparison of methods generalizing max-and average-pooling. *arXiv preprint arXiv:2103.01746*.

Calvo, D. (2017). Clasificación de redes neuronales artificiales. *Diego Calvo*, 13.

Cortes Ramirez, F. (2022a). Base de datos. https://drive.google.com/drive/folders/1VXnHuJtrc9SyRp1A-zVzjLUTvoM10lB8?usp=share_link. Accessed: 2025-01-18.

Cortes Ramirez, F. (2022b). Detecciones de armas. https://drive.google.com/drive/folders/1CwDXuPbBNhfXse6Xh0xNVXvCrj18Geaa?usp=share_link. Accessed: 2025-01-18.

Cortes Ramirez, F. (2022c). Yolov5custom. https://drive.google.com/drive/folders/1CE39Vas5k2miSoblwaztzrAcTYqhyrUt?usp=share_link. Accessed: 2025-01-18.

Cortes Ramirez, F. (2023). Sis app para la detección de crimen. https://drive.google.com/drive/folders/1UkyOt9Ako3uTR8h9sHI9jwBDhOowzasq?usp=share_link. Accessed: 2025-01-18.

de la Federación, D. O. (2004). Norma oficial mexicana proy-nom-034-sct2-2003. https://www.dof.gob.mx/nota_detalle.php?codigo=668546&fecha=04/06/2004#gsc.tab=0. Accessed: 2025-01-18.

del Transporte, I. M. (2016). LogÍstica del autotransporte de carga: Estrategias de gestiÓn. https://imt.mx/archivos/Publicaciones/PublicacionTecnica/pt483.pdf. Accessed: 2025-01-18.

Fathy, C. and Saleh, S. N. (2022). Integrating deep learning-based iot and fog computing with software-defined networking for detecting weapons in video surveillance systems. *Sensors*, 22(14):5075.

Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. (2013). Maxout networks. In *International conference on machine learning*, pages 1319–1327. PMLR.

Hnoohom, N., Chotivatunyu, P., and Jitpattanakul, A. (2022). Acf: an armed cctv footage dataset for enhancing weapon detection. *Sensors*, 22(19):7158.

Huszar, V. D., Adhikarla, V. K., Négyesi, I., and Krasznay, C. (2023). Toward fast and accurate violence detection for automated video surveillance applications. *IEEE Access*, 11:18772–18793.

Ingle, P. Y. and Kim, Y.-G. (2022). Real-time abnormal object detection for video surveillance in smart cities. *Sensors*, 22(10):3862.

Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia computer science*, 199:1066–1073.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D., et al. (2022). ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. *Zenodo*.

Lamas, A., Tabik, S., Montes, A. C., Pérez-Hernández, F., García, J., Olmos, R., and Herrera, F. (2022). Human pose estimation for mitigating false negatives in weapon detection in video-surveillance. *Neurocomputing*, 489:488–503.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer.

Manikandan, V. and Rahamathunnisa, U. (2022). A neural network aided attuned scheme for gun detection in video surveillance images. *Image and Vision Computing*, 120:104406.

Maqsood, R., Bajwa, U. I., Saleem, G., Raza, R. H., and Anwar, M. W. (2021). Anomaly recognition from surveillance videos using 3d convolution neural network. *Multimedia Tools and Applications*, 80(12):18693–18716.

Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., Garcia-Collantes, A., and Terashima-Marín, H. (2021). Criminal intention detection at early stages of shoplifting cases by using 3d convolutional neural networks. *Computation*, 9(2):24.

Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., and Terashima-Marín, H. (2020a). Suspicious behavior detection on shoplifting cases for crime prevention by using 3d convolutional neural networks. *arXiv preprint arXiv:2005.02142*.

Martínez-Mascorro, G. A., Ortiz-Bayliss, J. C., and Terashima-Marín, H. (2020b). Detecting suspicious behavior: How to deal with visual similarity through neural networks. *arXiv preprint arXiv:2007.15235*.

Martínez-Mascorro, G. A., Ortiz-Bayliss, J. C., and Terashima-Marín, H. (2020c). Detecting suspicious behavior on surveillance videos: Dealing with visual behavior similarity between bystanders and offenders. In *2020 IEEE ANDESCON*, pages 1–7. IEEE.

Mehta, V. (2021). Object detection using ssd mobilenet v2. https://vidishmehta204.medium.com/object-detection-using-ssd-mobilenet-v2-7ff3543d738d. Retrieved December 5, 2022.

Mohamed, H., Negm, A., Zahran, M., and Saavedra, O. C. (2015). Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: Case study el burullus lake. In *International water technology conference*, pages 12–14.

Murray, N. and Perronnin, F. (2014). Generalized max pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2473–2480.

Nadeem, M. S., Kurugollu, F., Atlam, H. F., and Franqueira, V. N. (2024). Weapon violence dataset 2.0: A synthetic dataset for violence detection. *Data in brief*, 54:110448.

Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Ruiz-Santaquiteria, J., Velasco-Mata, A., Vallez, N., Deniz, O., and Bueno, G. (2023). Improving handgun detection through a combination of visual features and body pose-based data. *Pattern Recognition*, 136:109252.

Sell, L. (2022). Labelimg. https://github.com/heartexlabs/labelImg.

Serna, E. et al. (2018). Desarrollo e innovación en ingeniería. *ANTIOQUIA: INSTITUTO ANTIOQUEÑO DE INVESTIGACION. Recuperado el*, 12.

Shah, S. A. A., Al-Khasawneh, M. A., and Uddin, M. I. (2021). Review of weapon detection techniques within the scope of street-crimes. In *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, pages 26–37. IEEE.

Sumi, L. and Dey, S. (2023). Yolov5-based weapon detection systems with data augmentation. *International Journal of Computers and Applications*, 45(4):288–296.

Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3):52.

Thuan, D. (2021). Evolution of yolo algorithm and yolov5: The state-of-the-art object detention algorithm.

Torregrosa-Domínguez, Á., Álvarez-García, J. A., Salazar-González, J. L., and Soria-Morillo, L. M. (2024). Effective strategies for enhancing real-time weapons detection in industry. *Applied Sciences*, 14(18):8198.

Ullah, W., Ullah, A., Hussain, T., Khan, Z. A., and Baik, S. W. (2021). An efficient anomaly recognition framework using an attention residual lstm in surveillance videos. *Sensors*, 21(8):2811.

Xie, T. and Yao, X. (2023). Smart logistics warehouse moving-object tracking based on yolov5 and deepsort. *Applied Sciences*, 13(17).

Xu, J., Li, Z., Du, B., Zhang, M., and Liu, J. (2020). Reluplex made more practical: Leaky relu. In *2020 IEEE Symposium on Computers and communications (ISCC)*, pages 1–7. IEEE.

Yadav, P., Gupta, N., and Sharma, P. K. (2023). A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. *Expert Systems with Applications*, 212:118698.

Yadav, P., Gupta, N., and Sharma, P. K. (2024). Robust weapon detection in dark environments using yolov7-darkvision. *Digital Signal Processing*, 145:104342.

Yadav, P., Gupta, N., and Sharma, P. K. (2025). Weaponvision ai: a software for strengthening surveillance through deep learning in real-time automated weapon detection. *International Journal of Information Technology*, pages 1–11.

Yu, D., Wang, H., Chen, P., and Wei, Z. (2014). Mixed pooling for convolutional neural networks. In *Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9*, pages 364–375. Springer.

Zahrawi, M. and Shaalan, K. (2023). Improving video surveillance systems in banks using deep learning techniques. *Scientific Reports*, 13(1):7911.

Apéndice A

Producción

Detección automática de delitos en sistemas de videovigilancia: una Revisión Sistemática de la Literatura

Félix Cortés Ramírez, Nimrod González Franco, Dante Mújica Vargas, Juan Gabriel González Serna, Raúl Pinto Elías

Departamento de Ciencias de la Computación, TecNM/CENIDET, Morelos 62490, México.
m21ce3003@cenidet.tecnm.mx (F.C.R), nimrod.gf@cenidet.tecnm.mx (N.G.F), dmujica@cenidet.tecnm.mx (D.M.V), Gabriel.gs@cenidet.tecnm.mx (G.G.S), raul.pe@cenidet.tecnm.mx (R.P.E)

Resumen: Una Revisión Sistemática de la Literatura (RSL), proporciona información cuantitativa y cualitativa sobre la investigación de algún tema en específico, detallando sobre conceptos utilizados durante la investigación como son: bases de datos, palabras clave utilizadas, resultados, etc. En este artículo, se presentara una revisión sistemática sobre el tema de Detección Automática de delitos en sistemas de videovigilancia mediante técnicas de aprendizaje profundo.

Palabras clave: Videovigilancia, detección de crimen, aprendizaje profundo, redes neuronales, robos.

1. INTRODUCCIÓN

En la actualidad, la detección de crimen [1 - 4] es uno de los problemas que sigue afectando a muchos países en diferentes aspectos a pesar de que existen sistemas [5, 6] de videovigilancia [7–10] que tienen como propósito la detección en tiempo real [11] y lograr manejar de manera eficiente un lugar de crimen [12], estos no han logrado obtener buenos resultados, dado que dichos sistemas son monitoreados [13] por un humano [14], y esto a su vez genera ineficiencia en algunas ocasiones, por lo tanto, no se cumple el objetivo que es la detección de crimen.

Por otro lado, la detección [15] de crimen resulta difícil [9], dado que no hay un patrón [16] o algún método [17] que haga eficiente este proceso, además, los movimientos [18] antes del crimen [19] cada vez son cambiantes y esta razón hace que sea un verdadero reto, además, por otro lado se encuentran los problemas de alta, baja iluminación y el desenfoque de las cámaras de seguridad [6 -7], [11], los cuales generan problemas para la visualización correcta de escenas de crimen.

Las Revisiones Sistemáticas de la Literatura (RSL) son importantes para lectores que desean realizar una búsqueda rápida sobre algún tema en específico, por lo que, ayudan a conocer sobre fuentes de consulta, técnicas de búsqueda y selección de la información, además hacen mención sobre la cantidad de información que existe sobre el tema.

En esta Revisión Sistemática, se mencionará sobre una amplia investigación de los trabajos relacionados con la detección de crimen en sistemas de videovigilancia usando aprendizaje profundo, así como las técnicas utilizadas para la obtención de información y los resultados finales.

El objetivo que se siguió en esta investigación fue conocer el estado del arte sobre el diseño, la implementación y la evaluación de un modelo de aprendizaje profundo para el reconocimiento de crímenes, además, se creó como pregunta principal de investigación la siguiente:

- PI: ¿Cuáles son las técnicas utilizadas actualmente en sistemas de videovigilancia para la detección de crimen?

Los materiales y métodos usados en la RSL se describen en la sección 2 del artículo, mientras que los resultados obtenidos se presentan en la sección 3. Finalmente, la sección 4 presenta las conclusiones derivadas de la revisión.

2.MATERIALES Y METODOS

En toda investigación científica es indispensable la utilización de algún método y/o material de estudio, y para el presente trabajo se utilizaron técnicas de cribado para reducir el sesgo de información, criterios de inclusión y exclusión de artículos, los cuales se dan a conocer en esta sección.

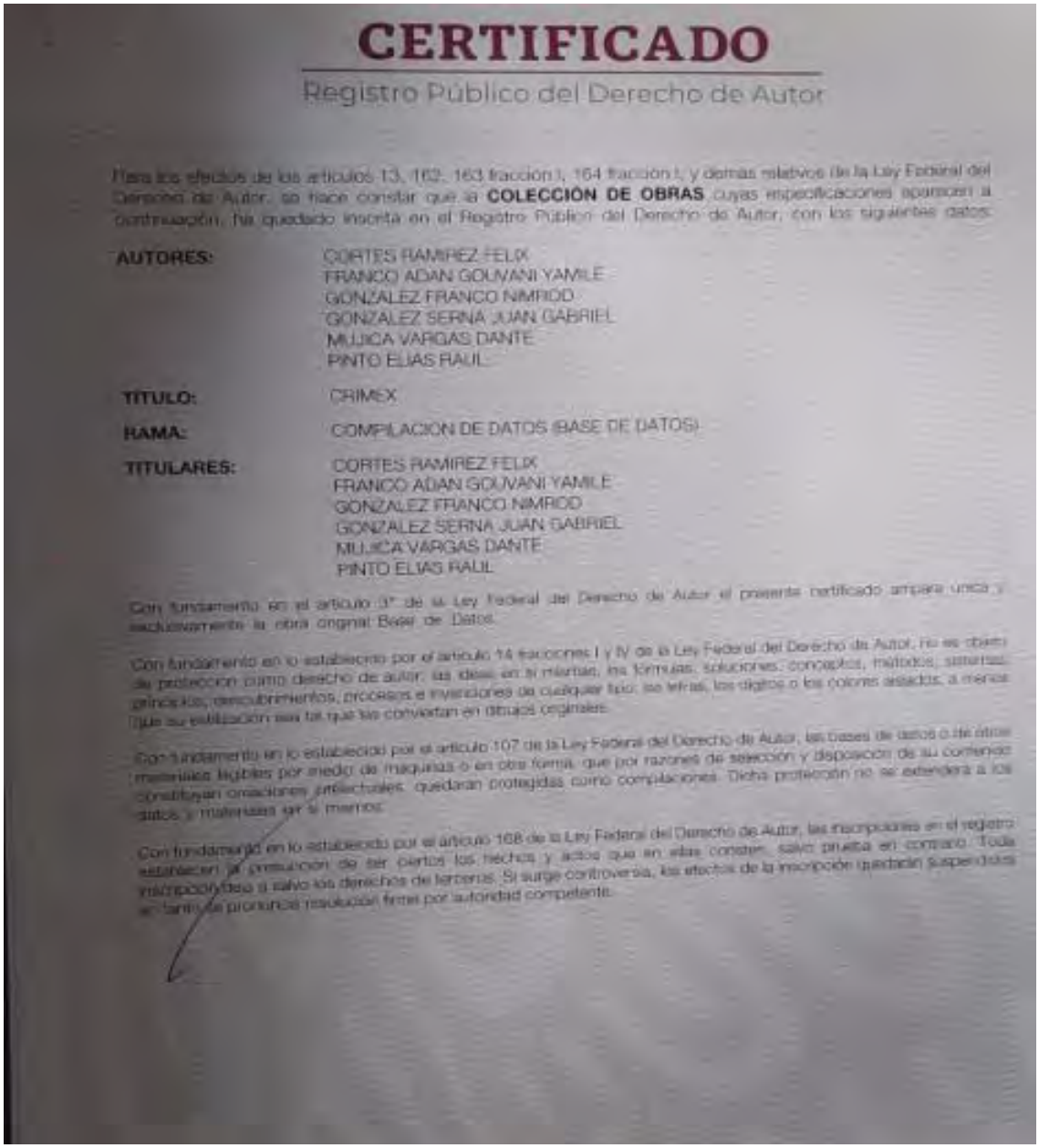
2.1 Criterios para la selección de información

Para la selección de información, se utilizaron criterios de inclusión y exclusión, los cuales fueron de gran importancia por el hecho de que el volumen de información relacionada fue impresionante, y con la aplicación de estos criterios, hubo una reducción considerable del mismo, lo cual beneficia a obtener información más relacionada al tema de estudio. Los criterios para la inclusión de trabajos fueron cinco:

- Trabajos con menos de 5 años de antigüedad.
- Trabajos redactados en idioma español e inglés.
- Considerar solo artículos científicos y libros.

Figura A.1 Portada de artículo científico publicado en la 7JCyTA, CENIDET 2021

Figura A.2 Portada de poster sobre el dataset Crimex, CENIDET 2022



CERTIFICADO

Registro Público del Derecho de Autor

Para los efectos de los artículos 13, 162, 163 fracción I, 164 fracción I, y demás relativos de la Ley Federal del Derecho de Autor, se hace constar que la **COLECCIÓN DE OBRAS** cuyas especificaciones aparecen a continuación, ha quedado inscrita en el Registro Público del Derecho de Autor, con los siguientes datos:

AUTORES: CORTES RAMIREZ FELIX
FRANCO ADAN GOUVANI YAMILE
GONZALEZ FRANCO NIMROD
GONZALEZ SERNA JUAN GABRIEL
MUJICA VARGAS DANTE
PINTO ELIAS RAUL

TITULO: CRIMEX

RAMA: COMPILACION DE DATOS (BASE DE DATOS)

TITULARES: CORTES RAMIREZ FELIX
FRANCO ADAN GOUVANI YAMILE
GONZALEZ FRANCO NIMROD
GONZALEZ SERNA JUAN GABRIEL
MUJICA VARGAS DANTE
PINTO ELIAS RAUL

Con fundamento en el artículo 3° de la Ley Federal del Derecho de Autor el presente certificado ampara única y exclusivamente la obra original Base de Datos.

Con fundamento en lo establecido por el artículo 14 fracciones I y IV de la Ley Federal del Derecho de Autor, no es objeto de protección como derecho de autor, las ideas en sí mismas, las fórmulas, soluciones, conceptos, métodos, sistemas, principios, descubrimientos, procesos e invenciones de cualquier tipo; las letras, los dígitos o los colores aislados, a menos que su estilización sea tal que las convierta en dibujos originales.

Con fundamento en lo establecido por el artículo 107 de la Ley Federal del Derecho de Autor, las bases de datos o de otros materiales legibles por medio de máquinas o en otra forma, que por razones de selección y disposición de su contenido constituyan creaciones intelectuales, quedarán protegidas como compilaciones. Dicha protección no se extenderá a los datos o materiales en sí mismos.

Con fundamento en lo establecido por el artículo 168 de la Ley Federal del Derecho de Autor, las inscripciones en el registro establecen la presunción de ser ciertos los hechos y actos que en ellas consten, salvo prueba en contrario. Toda inscripción deja a salvo los derechos de terceros. Si surge controversia, los efectos de la inscripción quedarán suspendidos en tanto se pronuncie resolución firme por autoridad competente.

Figura A.3 Certificado de registro de CRIMEX en INDAUTOR, CENIDET 2022

CERTIFICADO

Registro Público del Derecho de Autor

Para los efectos de los artículos 13, 162, 163 fracción I, 164 fracción I, y demás relativos de la Ley Federal del Derecho de Autor, se hace constar que la **OBRA** cuyas especificaciones aparecen a continuación, ha quedado inscrita en el Registro Público del Derecho de Autor, con los siguientes datos:

AUTORES: CORTES RAMIREZ FELIX
FRANCO ADAN GOUVANI YAMILE
GONZALEZ FRANCO NIMROD
MUJICA VARGAS DANTE

TÍTULO: CRIMEX IMAGE GENERATOR

RAMA: PROGRAMAS DE COMPUTACION

TITULARES: CORTES RAMIREZ FELIX
FRANCO ADAN GOUVANI YAMILE
GONZALEZ FRANCO NIMROD
MUJICA VARGAS DANTE

Con fundamento en lo establecido por el artículo 3° de la Ley Federal del Derecho de Autor, el presente certificado ampara única y exclusivamente la obra original solo de programa de computo.

Con fundamento en lo establecido por el artículo 168 de la Ley Federal del Derecho de Autor, las inscripciones en el registro establecen la presunción de ser ciertos los hechos y actos que en ellas consten, salvo prueba en contrario. Toda inscripción deja a salvo los derechos de terceros. Si surge controversia, los efectos de la inscripción quedarán suspendidos en tanto se pronuncie resolución firme por autoridad competente.

Con fundamento en los artículos 2, 208, 209 fracción III y 211 de la Ley Federal del Derecho de Autor; artículos 64, 103 fracción IV y 104 del Reglamento de la Ley Federal del Derecho de Autor; y artículos 1, 3 fracción I, 4, 8 fracción I y 9 del Reglamento Interior de Instituto Nacional del Derecho de Autor, se expide el presente certificado.

Número de Registro: 03-2022-062909280100-01

Ciudad de México, a 01 de julio de 2022

EL DIRECTOR DEL REGISTRO PÚBLICO DEL DERECHO DE AUTOR

JESÚS PARETS GÓMEZ

SECRETARÍA DE CULTURA
INSTITUTO NACIONAL DEL DERECHO DE AUTOR

Figura A.4 Certificado de registro de software Crimex Image Generator en INDAUTOR, CENIDET 2022