

# Detección de campamentos de minería ilegal en videos de cámaras FLIR

1<sup>er</sup> Tomas Acosta

*Ingeniería de sistemas y computación)*  
Universidad de los Andes  
Bogotá, Colombia  
t.acosta@uniandes.edu.co

2<sup>do</sup> Juan Andrés Carrasquilla

*Ingeniería de Sistemas y computación)*  
Universidad de los Andes  
Bogotá, Colombia  
j.carrasquillag@uniandes.edu.co

3<sup>ro</sup> Samuel Romero

*Ingeniería de sistemas y computación)*  
Universidad de los Andes  
Bogotá, Colombia  
sj.romero10@uniandes.edu.co

**Abstract**—La minería ilegal en la Amazonía colombiana representa una amenaza crítica para la biodiversidad y los recursos hídricos. Este artículo propone un modelo basado en YOLO que compara el rendimiento de entrenamientos con y sin preprocesamiento de imágenes para detectar campamentos de minería ilegal en videos aéreos. La metodología incluyó: (1) aumentación de datos en el dataset de entrenamiento para mejorar la generalización del modelo, (2) preprocesamiento con filtro mediano y CLAHE, y (3) búsqueda de hiperparámetros con el dataset con y sin procesar. Los resultados mostraron que el preprocesamiento mejora un 5% las métricas clave (recall: 0.837, mAP@0.5: 0.8875) frente a modelos sin procesamiento. Se evidenció que técnicas como CLAHE mejoran la detección de patrones. El trabajo establece un marco replicable para análisis ambiental basado en visión artificial, priorizando sensibilidad sobre falsos positivos para aplicaciones en vigilancia ecológica.

**Index Terms**—componente, formato, estilo, diseño, insertar

## I. INTRODUCCIÓN

La minería ilegal en regiones de alta biodiversidad constituye una de las principales amenazas para la conservación de ecosistemas. En particular, la selva amazónica enfrenta presiones crecientes debido a la expansión no regulada de actividades extractivas, que generan impactos significativos sobre el suelo, los cuerpos de agua y la cobertura forestal. Según Boscó Arias [1], la minería ilegal genera graves consecuencias ambientales, sociales y económicas. Provoca contaminación por mercurio en ríos y deforestación, afectando la biodiversidad y la calidad del agua. Además, altera las dinámicas de las comunidades indígenas, cambiando roles tradicionales y generando conflictos e inseguridad. Económicamente, financia el crimen organizado, con el 85% del oro extraído de forma ilegal [2], lo que resulta en pérdidas significativas para el Estado y fortalece redes de lavado de dinero. Esta situación plantea retos complejos para las autoridades encargadas de la protección ambiental y la seguridad territorial, quienes deben responder con eficacia en contextos geográficos remotos, de difícil acceso y con condiciones operativas adversas.

En este contexto, las tecnologías de observación remota han cobrado relevancia como herramientas para el monitoreo ambiental. La disponibilidad creciente de plataformas aéreas no tripuladas y sensores ópticos ha facilitado la recolección

masiva de datos visuales sobre extensas áreas de difícil vigilancia. Sin embargo, el valor estratégico de esta información depende en gran medida de la capacidad para procesarla de manera oportuna, confiable y automatizada. El análisis visual manual, aunque preciso en ciertos casos, presenta limitaciones inherentes en términos de escalabilidad, eficiencia operativa y susceptibilidad a errores humanos, especialmente cuando se enfrentan condiciones de visibilidad limitadas o camuflaje intencional de las actividades ilícitas.

Frente a estos desafíos, los avances recientes en visión por computador y aprendizaje profundo ofrecen nuevas posibilidades para la detección automatizada de patrones asociados a la minería ilegal. Estos enfoques permiten no solo identificar objetos relevantes en imágenes complejas, sino también adaptarse a contextos dinámicos mediante técnicas de entrenamiento supervisado y adaptación de dominio. No obstante, su aplicación efectiva en entornos amazónicos aún requiere superar obstáculos técnicos significativos, tales como la variabilidad lumínica, la presencia de ruido atmosférico y la diversidad morfológica de los elementos a identificar.

Este trabajo se propone como una contribución al desarrollo de sistemas inteligentes de vigilancia aérea capaces de detectar con precisión y en tiempo real indicios visuales de actividades mineras ilegales en la selva. La investigación se orienta hacia el diseño e implementación de una arquitectura de análisis automatizado que integre procesos de preprocesamiento, modelos de detección y una validación bajo imágenes de operacionales reales. Al abordar simultáneamente aspectos técnicos y operativos, se busca avanzar en la consolidación de soluciones tecnológicas que fortalezcan las capacidades institucionales de monitoreo ambiental y respuesta oportuna.

Concretamente, el enfoque adoptado en este estudio se basa en la aplicación de modelos de detección de objetos de tipo You Only Look Once (YOLO) [3], entrenados sobre datos reales de vigilancia aérea en el amazonas. A lo largo del trabajo, se presentan múltiples configuraciones experimentales orientadas a evaluar el desempeño del modelo bajo distintas condiciones visuales, comparando entrenamientos realizados con imágenes crudas y con imágenes sometidas a técnicas específicas de preprocesamiento. En particular implementamos métodos como el filtro mediano y la ecualización adaptativa del histograma (CLAHE) [4]. Esta última es una técnica

de mejora de contraste que ajusta localmente la iluminación de una imagen, siendo especialmente útil en visión por computador para resaltar detalles en condiciones de iluminación no uniforme.

La metodología seguida comprende tres fases principales: la expansión del conjunto de datos mediante técnicas de aumentación visual (con movimientos de traslación y aplicación de filtros), la preparación de dos flujos paralelos de datos con y sin procesamiento previo, y la aplicación de la búsqueda de hiperparámetros para encontrar el modelo de mejor desempeño. Si bien se consideran múltiples indicadores de rendimiento, este trabajo da especial importancia al recall como métrica principal, con el objetivo de minimizar los falsos negativos. Esta elección responde a una necesidad operativa crítica: en el contexto de vigilancia ambiental, omitir la detección de una actividad ilícita puede tener consecuencias irreversibles para el ecosistema y dificultar las acciones correctivas oportunas.

El artículo se estructura en cinco secciones principales. La Sección 2 presenta una revisión del estado del arte, en la que se discuten enfoques previos para la detección de actividades ilícitas mediante visión por computador, haciendo énfasis en aplicaciones recientes de redes neuronales convolucionales y en las limitaciones de los sistemas actuales. En la Sección 3 se describe detalladamente la metodología propuesta, incluyendo la construcción del dataset, los procesos de aumentación y preprocesamiento, las arquitecturas empleadas y los criterios de validación. La Sección 4 reporta los resultados obtenidos, comparando el desempeño de los modelos con distintas configuraciones y analizando su efectividad bajo condiciones visuales adversas. La Sección 5 discute las implicaciones de los hallazgos en contextos operativos reales, identifica posibles limitaciones del enfoque propuesto y sugiere direcciones para investigaciones futuras. Finalmente, se incluye una sección bibliográfica con las referencias citadas a lo largo del documento.

## II. ESTADO DEL ARTE

El desarrollo de sistemas automatizados para la detección de actividades ilícitas en entornos naturales ha sido objeto de creciente atención en la literatura científica. Diversos estudios han aplicado técnicas de visión por computador y aprendizaje profundo para tareas como la identificación de inundaciones, deforestación, cuerpos de agua y minería aurífera en regiones de difícil acceso. Esta sección presenta una revisión de trabajos representativos que, por su enfoque metodológico o dominio de aplicación, resultan relevantes para el diseño del sistema propuesto en este estudio.

Gomes et al. [5] exploran el uso de YOLOv8 junto con imágenes de radar (SAR) para detectar pistas de aterrizaje ilegales en la selva amazónica, un método adaptable también a la identificación de minería ilegal. La integración de YOLOv8 permite detectar estructuras anómalas en entornos selváticos de manera rápida, avanzando en el análisis satelital con redes neuronales, mientras que el uso de radar supera las limitaciones de las imágenes ópticas al no verse afectado

por la nubosidad. Sin embargo, la precisión del modelo está ligada a la calidad y cantidad de los datos de entrenamiento, que suelen ser escasos y de baja resolución. Además, la detección automatizada puede producir falsos positivos si no se complementa con información geoespacial o validaciones de campo, y el entrenamiento de YOLOv8 demanda recursos computacionales significativos. En la región del Chocó colombiano, Rico et al. [6] analizan el uso de imágenes satelitales de radar (SAR) para monitorear la minería aurífera aluvial, aprovechando la capacidad del radar para penetrar la vegetación y detectar cambios superficiales en áreas de difícil acceso. Este enfoque facilita un seguimiento continuo, independientemente de las condiciones climáticas, y la identificación de alteraciones en la morfología de ríos y suelos permite trazar la expansión minera a lo largo del tiempo. A pesar de ello, diferenciar entre actividades mineras legales e ilegales requiere datos adicionales, ya que el radar por sí solo no ofrece esa distinción, y la resolución de las imágenes limita la detección de operaciones de menor escala.

Los trabajos relacionados con la identificación de cuerpos de agua y detección de inundaciones son especialmente relevantes en el contexto de la minería, ya que muchas de sus operaciones utilizan o afectan recursos hídricos. Reconocer zonas con presencia de agua permite monitorear impactos ambientales, prevenir riesgos por contaminación y mejorar la vigilancia de actividades ilegales que dependen del uso intensivo del agua. En este sentido, Özdemir et al. [7] desarrollan un marco para extraer cuerpos de agua de imágenes aéreas y satelitales de alta resolución mediante modelos de fundación visual, como SAM y CLIP. El proceso combina segmentación automática con SAM y clasificación de segmentos con CLIP, logrando una identificación precisa de cuerpos de agua con métricas como F1-Score, OA e IoU elevadas, y demostrando versatilidad al aplicarse a distintos tipos de imágenes. Sin embargo, el método enfrenta dificultades para segmentar cuerpos de agua pequeños o en áreas de alto contraste, lo que afecta la definición de bordes, y su eficacia depende de ajustes manuales en parámetros específicos del dataset, como el stability score offset, lo que restringe la automatización completa. Ansari et al. [8] presentan una metodología basada en el algoritmo de k-means y técnicas de umbralización para cuantificar variaciones en áreas de flujo superficial y crecimiento de vegetación en el río Kerian. El estudio reporta altos niveles de precisión en la detección de estas características —con coeficientes de Dice y Jaccard de hasta 97.86% y 94.36%, respectivamente— lo que demuestra el potencial de estas técnicas para el monitoreo de ríos, la planificación agrícola y el análisis del entorno urbano en zonas fluviales. En el ámbito de la detección de inundaciones con UAV y aprendizaje profundo, Yang et al. [9] presentan un método que emplea UAV y el algoritmo YOLOv3 para identificar edificios y vegetación afectados por inundaciones. Este enfoque procesa imágenes aéreas mediante el análisis de parámetros de vuelo y características visuales, lo que permite detectar objetos de interés y estimar el área inundada. Los resultados muestran una precisión del 88% en la detección de

edificios y del 85% en vegetación, además de proporcionar una estimación cuantitativa del área afectada, útil para la gestión de emergencias. Sin embargo, la calidad de las imágenes condiciona su eficacia, ya que condiciones de baja iluminación o alta oclusión disminuyen el rendimiento. Asimismo, el método exige una base de datos etiquetada y un entrenamiento que, según se reporta, consume 60 horas de procesamiento. Por otro lado, Barlow et al. [10] proponen una metodología basada en la arquitectura U-Net para segmentar límites de cauces en regiones polares hiperáridas, con un estudio centrado en Taylor Valley, Antártida. A partir de rasters generados con datos lidar, como elevación, pendiente, intensidad y acumulación de flujo, el modelo extrae características topográficas que facilitan la delineación precisa de cauces. Este enfoque alcanza índices de precisión, recall y F1 cercanos a 0.94, especialmente en cauces meandrinicos con discontinuidades marcadas, y reduce la dependencia de procesos manuales gracias a la automatización mediante aprendizaje profundo. No obstante, depende de datos etiquetados manualmente, lo que puede introducir errores si no reflejan la complejidad del terreno, y su desempeño decae en cauces de geometría simple, como canales rectos con variaciones sutiles en la pendiente.

Finalmente, el uso de vehículos aéreos no tripulados (UAVs) ha cobrado relevancia en aplicaciones de monitoreo ambiental y vigilancia, gracias a su capacidad para capturar imágenes de grandes extensiones de terreno. En este contexto, mencionamos algunos los trabajos que se enfocan en el procesamiento de imágenes obtenidas desde UAVs. Diversos autores han propuesto variaciones a los modelos de YOLO con el objetivo de optimizar su desempeño en tareas de detección de objetos desde (UAV), atendiendo a los desafíos particulares que implican estos escenarios, como la detección de objetos pequeños, condiciones ambientales cambiantes y recursos computacionales limitados. Por ejemplo, Cui et al. [11] propusieron una mejora sobre YOLOv5 para incrementar la precisión en imágenes de alta resolución capturadas por UAVs, introduciendo una estrategia de segmentación de imágenes y adaptaciones en la arquitectura del modelo, lo cual permitió aumentar significativamente el desempeño en la detección de drones pequeños. De manera complementaria, Chen et al. [12] presentaron una revisión integral de la tecnología basada en YOLO para UAVs (YBUT), resaltando su impacto en sectores como agricultura, transporte y automatización, y destacando cómo la fusión de tecnologías ha ampliado el potencial de aplicación del algoritmo. Asimismo, Qiu et al. [13] desarrollaron el modelo YOLO-GNS, enfocado en la detección de vehículos especiales desde drones. Este modelo incorpora estructuras como SSH y GhostNet para mejorar la capacidad de detección en objetos parcialmente ocultos o pequeños, al mismo tiempo que reduce los costos computacionales, logrando mejoras notables en precisión y velocidad. Estos trabajos demuestran cómo la adaptación estratégica de modelos YOLO puede responder eficazmente a los retos de la visión computacional desde UAVs en contextos reales.

### III. METODOLOGÍA

1) *Exploración y aumentación de datos*: El conjunto de datos utilizado en este estudio fue inicialmente dividido en tres carpetas correspondientes a los subconjuntos de entrenamiento (train), validación (val) y prueba (test). La distribución original fue: 178 imágenes en train, 107 en val y 372 en test. Estos subconjuntos incluyen imágenes etiquetadas y no etiquetadas, siendo algunas de las no etiquetadas únicamente de fondo, sin presencia explícita de objetos de interés.

Se revisó la distribución de clases y anotaciones por conjunto. En la Tabla I se resume la cantidad total de imágenes y anotaciones por subconjunto, diferenciando aquellas que contienen objetos (anotadas) de las que sólo presentan fondo.

TABLE I  
DISTRIBUCIÓN POR CONJUNTO

Conjunto	Total	Anotadas	Fondo
Train	1178	948	230
Valid	107	85	22
Test	372	281	91

Se definieron seis clases dentro del conjunto de datos, correspondientes a categorías visuales relevantes para la detección de actividades mineras ilegales:

- **Clase -1: Background**, imágenes sin objetos de interés.
- **Clase 0:** Vehículos.
- **Clase 1:** Bodegas.
- **Clase 2:** Caminos.
- **Clase 3:** Ríos.
- **Clase 4:** Zonas de minería ilegal.

Con el fin de realizar una validación visual cualitativa, se seleccionaron cinco imágenes representativas por cada clase, incluyendo el fondo, como subconjunto de referencia. En las Figuras 1 a 6, se muestra un ejemplo de cada clase considerada en el conjunto de datos.



Fig. 1. Ejemplo de la **Clase -1 (Background)**: imagen sin objetos de interés.

Para robustecer el proceso de entrenamiento y mejorar la generalización del modelo ante condiciones operacionales diversas, se implementó un esquema de *aumentación de datos* diferenciado según el tipo de contenido. Esta técnica permite incrementar la variabilidad del conjunto de entrenamiento sin introducir nuevas imágenes, preservando la información



Fig. 2. Ejemplo de la **Clase 0 (Vehículos)**: anotación de un vehículo en vista aérea.



Fig. 3. Ejemplo de la **Clase 1 (Bodegas)**: estructura identificada como bodega.



Fig. 4. Ejemplo de la **Clase 2 (Caminos)**: vía no pavimentada detectada en el entorno selvático.

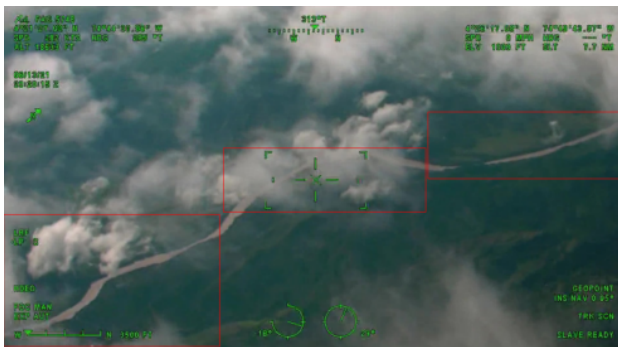


Fig. 5. Ejemplo de la **Clase 3 (Ríos)**: cuerpo de agua delimitado manualmente.



Fig. 6. Ejemplo de la **Clase 4 (Zonas de minería ilegal)**: área con intervención evidente por maquinaria.

espacial relevante mediante el ajuste automático de las cajas delimitadoras (bounding boxes), gracias al uso de la biblioteca *Albumentations* para objetos detectables y *OpenCV* para imágenes de fondo. Las transformaciones aplicadas incluyen:

- **HorizontalFlip**: simula capturas desde distintos ángulos horizontales con probabilidad del 50
- **RandomBrightnessContrast**: modifica aleatoriamente tanto el brillo como el contraste para reflejar variaciones en las condiciones de iluminación.
- **MotionBlur**: introduce desenfoque direccional controlado para emular movimiento del dron o de la cámara durante la captura.
- **RandomGamma**: ajusta el contraste no lineal, especialmente útil en condiciones ambientales difíciles como nubosidad o sombra densa.
- **Affine**: aplica transformaciones geométricas incluyendo escalado (0.8-1.2x), traslación ( $\pm 10$ )
- **GaussianBlur**: añade desenfoque gaussiano sutil para emular condiciones atmosféricas o imperfecciones ópticas.
- **Ruido gaussiano**: incorpora variabilidad estocástica para simular interferencias del sensor y condiciones de captura no ideales.

Para imágenes de fondo sin objetos detectables, se desarrolló un pipeline específico basado en *OpenCV* que incluye las mismas transformaciones adaptadas para preservar las características ambientales relevantes sin requerir ajuste de coordenadas de detección. En las Figuras 7 y 8 se muestran unos ejemplos de las transformaciones realizadas sobre una muestra de los datos.

Estas transformaciones se aplicaron tanto a imágenes anotadas como a imágenes de fondo. El objetivo fue balancear el número de muestras por clase hasta alcanzar superar las 2000 imágenes por categoría. Finalmente, se calculó la distribución de objetos (labels) por clase, después de aplicar la aumentación de los datos. En la tabla II se muestra la distribución final por clases.





Fig. 7. Ejemplo de transformaciones aplicadas sobre las muestras de datos en particular desenfoque y movimiento



Fig. 8. Ejemplo de transformaciones aplicadas sobre las muestras de datos en particular movimiento y ajuste de contraste

TABLE II  
DISTRIBUCIÓN DE OBJETOS E IMÁGENES POR CLASE

Clase	Descripción	Cantidad	Porcentaje (%)
<b>Imágenes de Fondo</b>			
-1	Fondo (sin objetos)	2530	-
<b>Objetos Detectados</b>			
0	Vehículos	2396	10.80
1	Bodegas/construcciones	4654	20.98
2	Caminos	5353	24.14
3	Ríos	4832	21.79
4	Minería ilegal	4944	22.29
<b>Total Objetos</b>		<b>22179</b>	<b>100.00</b>

#### A. Preprocesamiento de las imágenes

El preprocesamiento de los datos consistió en dos etapas principales: la reducción de ruido y la mejora del contraste en imágenes con nubes. Estas operaciones se implementaron con el objetivo de facilitar la detección de objetos de interés y robustecer el desempeño del modelo durante el entrenamiento. Para reducir el ruido de tipo sal y pimienta, se utilizó un filtro mediano. Este se aplicó únicamente a imágenes en las que se detectó una probabilidad de ruido superior a un umbral de 0.01, de modo que no se afectaran imágenes limpias innecesariamente. Por otro lado, la mejora del contraste se aplicó a todo el conjunto de datos, sobre todo para mitigar el efecto que tienen las nubes sobre algunas imágenes. Para ello, se utilizó la técnica CLAHE, que permite redistribuir el histograma de intensidades de la imagen localmente,

mejorando el contraste sin sobresaturar regiones homogéneas. Ambas técnicas fueron aplicadas primero a un conjunto reducido de imágenes para evaluar visualmente su impacto. Se generaron visualizaciones comparativas que incluyen la imagen original, la imagen con reducción de ruido, y la imagen con mejora de contraste, acompañadas por sus respectivos histogramas. En la Figura 9, se muestra un ejemplo del procesamiento de imágenes que se aplicó.

El preprocesamiento se aplicó sobre el conjunto completo de datos y se almacenaron las imágenes procesadas en un nuevo directorio que mantiene la misma estructura del dataset original, tanto para la versión filtrada como para la versión completa. Una vez finalizado el preprocesamiento, se generaron los archivos YAML correspondientes, necesarios para que los modelos pudieran identificar la estructura de las rutas, nombres de clases y organización de los datos.

#### B. Entrenamiento

Para la tarea de detección de zonas de minería ilegal, se optó por utilizar YOLO, uno de los modelos más empleados en detección de objetos debido a su equilibrio entre velocidad y precisión. Su arquitectura permite procesar imágenes en un solo paso, lo que lo hace eficiente para aplicaciones en tiempo real. Además, su amplia adopción en la comunidad científica y su soporte para transfer learning facilitan la implementación y el ajuste a casos específicos, como la identificación de áreas afectadas por minería en imágenes aéreas o satelitales. Esta combinación de eficiencia, precisión y flexibilidad lo convierte en una opción adecuada para el proyecto.

La Tabla III resume las principales características de estas variantes. Modelos pequeños como yolov11n y yolov11s son ideales para tareas rápidas o dispositivos con recursos limitados, mientras que versiones más grandes como yolov11l o yolov11x están pensadas para contextos donde se dispone de suficiente potencia de cómputo y un conjunto de datos diverso. En nuestro caso, debido a la complejidad visual de las imágenes (presencia de nubes, vegetación densa, múltiples clases pequeñas) y al contar con GPUs con al menos 12 GB de VRAM, optamos por modelos de tamaño medio a grande (yolov11m y yolov11l) que permiten capturar mejor estas características sin comprometer la eficiencia.

TABLE III  
COMPARACIÓN DE VARIANTES DE YOLOv11

Modelo	Tamaño	Uso	VRAM
yolov11n	Pequeño	Móviles/tiempo real	4 GB
yolov11s	Compacto	Velocidad alta	6 GB
yolov11m	Mediano	Balanceado	8 GB
yolov11l	Grande	Alta precisión	12 GB
yolov11x	Extra grande	Máximo rendimiento	16 GB

Para evaluar sistemáticamente el modelo, analizamos múltiples configuraciones experimentales centrándonos en dos aspectos fundamentales: (1) el efecto del preprocesamiento de imágenes y (2) el impacto de diferentes parámetros de

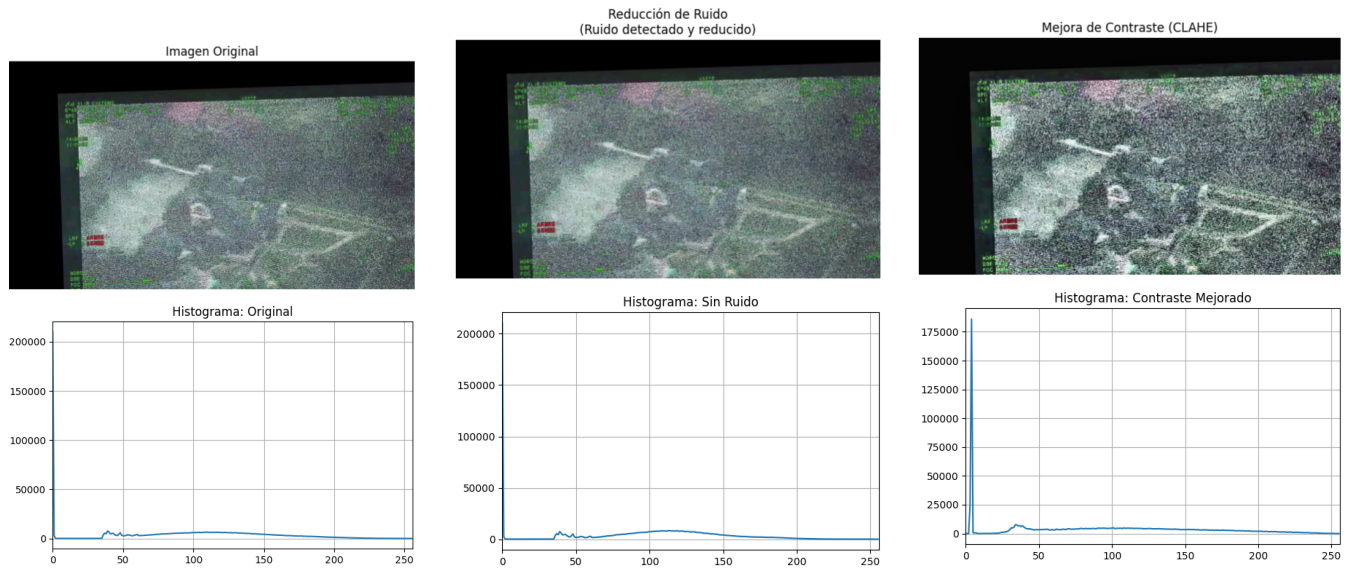


Fig. 9. Proceso de preprocesamiento aplicado a una imagen: izquierda, imagen original; centro, imagen con reducción de ruido mediante filtro mediano; derecha, imagen con mejora de contraste mediante CLAHE.

entrenamiento. Este enfoque nos permitió determinar el mejor modelo para finalmente aplicar fine-tuning sobre la mejor configuración. La metodología buscaba encontrar el mejor modelo para resolver la tarea.

La Tabla IV resume las combinaciones de configuraciones evaluadas:

TABLE IV  
CONFIGURACIONES EXPERIMENTALES YOLO

Config.	Parámetros	Objetivo
1	<b>Ultra Recall</b> Batch=40, LR=0.02 Mosaic=1.0, Mixup=0.15	Máximo recall
2	<b>High Resolution</b> Batch=40, Img=832px Modelo=yolo11l.pt	Detalles finos
3	<b>Balanced</b> Batch=32, LR=0.01 Mosaic=0.85, Mixup=0.05	Balance recall/precision
4	<b>Conservative</b> Batch=40, LR=0.005 Sin mixup, Dropout=0.1	Estable y confiable <b>Mejor resultado</b>

**Total: 8 experimentos**

*4 configuraciones × 2 datasets (Con procesamiento y sin procesamiento)*

Para consultar en detalle los resultados obtenidos por cada uno de los modelos evaluados, así como los parámetros de entrenamiento y configuración de cada experimento realizado, se puede acceder al repositorio del proyecto disponible en el siguiente enlace:

[https://github.com/Tomas1307/FLIR\\_GPF2](https://github.com/Tomas1307/FLIR_GPF2)

Este repositorio incluye scripts, configuraciones, conjuntos de datos anotados y documentación técnica adicional que respalda el desarrollo y análisis presentados en este trabajo.

## IV. RESULTADOS

A continuación se presentan los resultados obtenidos para cada una de las configuraciones de entrenamiento descritas previamente. La evaluación de los modelos se realizó utilizando métricas estándar en detección de objetos: *Recall*, *mAP@50* y *mAP@50:95*. Estas métricas se calcularon sobre un conjunto de validación común para todos los modelos, permitiendo comparaciones consistentes. Dada la naturaleza crítica del problema —la detección de zonas de minería ilegal en imágenes satelitales— se priorizó el **recall** por encima de otras métricas, ya que el objetivo principal es minimizar los falsos negativos, es decir, evitar que regiones con actividad ilegal pasen desapercibidas para el modelo. Aunque otras métricas como la precisión y el F1-score son relevantes, en este contexto es preferible tolerar algunos falsos positivos si con ello se asegura una mayor cobertura de los casos verdaderamente positivos.

En la Tabla V se resumen los indicadores principales obtenidos para los mejores 5 experimentos. A partir de este análisis, se seleccionó el modelo con mejor desempeño.

TABLE V  
RESULTADOS DE EXPERIMENTOS YOLO

ID	Config/Dataset	Recall Min.	Precisión	mAP@50 Gral.
1	Conservative/Prep.	<b>0.863</b>	0.758	<b>0.895</b>
2	Ultra Recall/Prep.	0.838	0.788	0.888
3	Ultra Recall LR/Prep.	0.838	0.788	0.888
4	Balanced/Orig.	0.831	0.681	0.787
5	Balanced/Prep.	0.825	<b>0.805</b>	0.886

A partir de los resultados presentados en la Tabla V, se seleccionó la configuración con mejor desempeño. El modelo con mejor Recall para la clase de minería fue el Conservative con dataset preprocesado. Dado que la prioridad

del sistema es minimizar los falsos negativos —es decir, evitar que se omitan zonas con actividad minera ilegal—, se consideró fundamental privilegiar aquellas configuraciones que maximicen esta métrica. Posteriormente, se aplicó una estrategia de fine-tuning sobre este modelo utilizando un learning rate reducido durante 45 épocas adicionales. El fine-tuning permite aprovechar el conocimiento previamente adquirido por el modelo pre-entrenado YOLOv11 en el dataset COCO, adaptando gradualmente las representaciones internas para optimizar la detección de patrones específicos presentes en imágenes aéreas de zonas mineras. A continuación, se presentan los resultados obtenidos.

TABLE VI  
MÉTRICAS POR CLASE PARA EL MODELO CON CONFIGURACIÓN  
CONSERVATIVE TRAS FINE-TUNING

Clase	Imágenes	Instancias	Precision	Recall	mAP@50	mAP@50-95
Todas	509	1569	0.914	0.908	0.934	0.815
Vehículos	22	23	0.931	1.000	0.995	0.986
Bodegas	249	897	0.935	0.871	0.949	0.703
Caminos	245	400	0.947	0.955	0.958	0.848
Ríos	114	169	0.892	0.911	0.926	0.798
Zonas de minería ilegal	65	80	<b>0.865</b>	<b>0.803</b>	<b>0.844</b>	<b>0.740</b>

Finalmente, presentamos en las figuras 10 y 11 ejemplos de las predicciones generadas por los modelos seleccionados. Estas visualizaciones permiten observar el comportamiento del modelo.

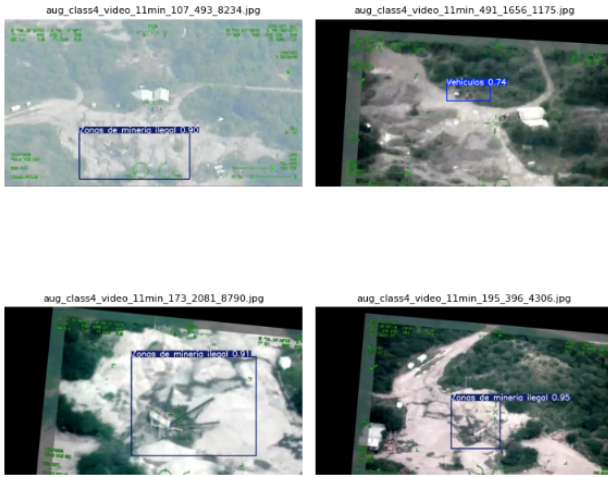


Fig. 10. Ejemplo de las predicciones del modelo con mejor Recall promedio por clases

## V. DISCUSIÓN

Los experimentos abarcaron diversas combinaciones de técnicas de preprocesamiento y configuraciones de entrenamiento, con el propósito de optimizar métricas clave como el recall, el mAP@50 y el mAP@50:95. Dado el contexto crítico de la aplicación —donde omitir zonas afectadas por minería ilegal podría tener serias implicaciones ambientales y sociales—, se priorizó el recall como métrica principal para minimizar los falsos negativos. A continuación, se examinan los resultados, se justifican las elecciones

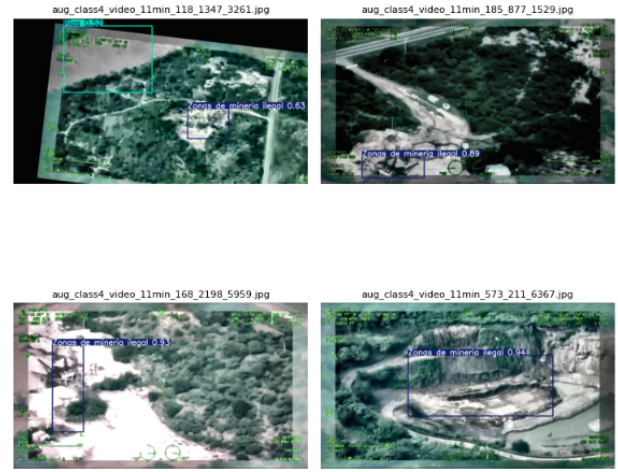


Fig. 11. Ejemplo de las predicciones del modelo con mejor Recall para la clase Zona de Minería Ilegal

realizadas en las configuraciones, se detalla el desempeño del modelo seleccionado y se exploran posibles mejoras. Los experimentos, cuyos resultados se presentan en la Tabla V, incluyeron 8 configuraciones experimentales que evaluaron sistemáticamente el impacto de dos factores fundamentales: (1) el efecto del preprocesamiento de imágenes (dataset procesado vs. original) y (2) diferentes estrategias de entrenamiento mediante cuatro configuraciones principales: Ultra Recall, High Resolution, Balanced y Conservative. Esta aproximación más focalizada permitió evaluar de manera controlada cómo la preparación de los datos y los parámetros de entrenamiento influían específicamente en la detección de "zonas de minería ilegal", que constituía el foco principal del problema. El preprocesamiento aplicado, que incluyó reducción de ruido mediante filtros medianos y mejora de contraste usando CLAHE, demostró ser crucial para el rendimiento del modelo. Los resultados evidenciaron consistentemente que las configuraciones entrenadas con el dataset preprocesado superaron a sus contrapartes entrenadas con datos originales, con una mejora promedio en recall de 0.835 vs 0.791. Esto sugiere que las técnicas de preprocesamiento efectivamente facilitan la extracción de características relevantes en imágenes aéreas, particularmente en presencia de nubes y condiciones de iluminación variable. La elección de las métricas evaluadas —recall, mAP@50 y mAP@50:95— se fundamentó en las necesidades específicas del problema. El recall fue priorizado porque en este contexto es crítico identificar la mayor cantidad posible de zonas de minería ilegal, incluso si ello implica aceptar algunos falsos positivos, antes que pasar por alto áreas críticas. Las métricas mAP complementaron esta evaluación al medir la precisión en la localización de objetos, garantizando que las detecciones fueran confiables para aplicaciones prácticas. El modelo con configuración Conservative aplicado al dataset preprocesado emergió como la mejor solución, alcanzando el recall más alto para minería ilegal (0.863) y

el mejor mAP@50 general (0.895). Su configuración —que incluía batch=40, learning rate=0.005, ausencia de mixup y dropout=0.1— proporcionó el equilibrio óptimo entre estabilidad de entrenamiento y capacidad de generalización. El enfoque conservador en los parámetros de entrenamiento evitó el sobreajuste, mientras que el preprocesamiento robusteció la capacidad del modelo para manejar variabilidad en las condiciones de captura. La aplicación subsecuente de fine-tuning durante 45 épocas adicionales con learning rate reducido permitió una adaptación más refinada del modelo pre-entrenado YOLOv11. Los resultados finales —mostrados en la Tabla VI— demuestran el éxito de esta estrategia: recall general de 0.908, precisión de 0.914, y específicamente para zonas de minería ilegal, recall de 0.803 con precisión de 0.865. Estos valores reflejan un modelo robusto capaz de detectar eficazmente las instancias críticas mientras mantiene un nivel aceptable de precisión. El análisis comparativo revela que configuraciones como Ultra Recall, a pesar de su nombre, no lograron superar al enfoque Conservative, sugiriendo que la estabilidad en el entrenamiento y la calidad de los datos preprocesados son más determinantes que la agresividad en los parámetros de augmentación. La configuración Balanced mostró la mayor precisión (0.805) pero a costa de un recall menor, confirmando el trade-off esperado entre estas métricas. A pesar de los resultados prometedores, se identificaron áreas de mejora. La clase "bodegas" mostró el recall más bajo (0.871), posiblemente debido a su mayor variabilidad visual o confusión con estructuras naturales. Para la clase objetivo "zonas de minería ilegal", aunque el recall de 0.803 es satisfactorio, existe margen para mejora mediante la recolección de datos adicionales, técnicas de balanceamiento más sofisticadas, o la exploración de arquitecturas de ensemble que combinen las fortalezas de diferentes configuraciones. La visualización de predicciones en las Figuras 10 y 11 proporciona evidencia cualitativa del desempeño del modelo y sugiere direcciones para refinamiento futuro.

## VI. CONCLUSIONES

En este trabajo se abordó el problema de la detección automática de zonas de minería ilegal en imágenes de cámaras FLIR mediante técnicas de aprendizaje profundo basadas en YOLOv11. Se evaluaron sistemáticamente 8 configuraciones experimentales que exploraron el impacto del preprocesamiento y cuatro estrategias de entrenamiento, priorizando el recall para minimizar falsos negativos críticos. Los resultados demostraron la superioridad del dataset preprocesado sobre el original (recall promedio 0.835 vs 0.791), confirmando que las técnicas de reducción de ruido y mejora de contraste son fundamentales para imágenes aéreas con condiciones adversas. El modelo Conservative aplicado al dataset preprocesado emergió como la solución óptima, alcanzando el mejor recall para minería ilegal (0.863) y mAP@50 general (0.895). La estrategia de fine-tuning durante 15 épocas adicionales refinó el modelo seleccionado, logrando resultados finales excepcionales: recall general de

0.908, precisión de 0.914, y para zonas de minería ilegal, recall de 0.803 con precisión de 0.865. El análisis reveló que la estabilidad en el entrenamiento es más determinante que estrategias agresivas de augmentación.

Se identificaron limitaciones en la detección de "bodegas" (recall 0.871) debido a su variabilidad visual, pero el sistema demostró viabilidad operacional para monitoreo automatizado de actividades mineras ilegales.

Para trabajos futuros se plantean tres direcciones principales de investigación. Primero, la implementación de aprendizaje activo para optimizar la recolección de datos mediante la identificación automática de muestras más informativas, especialmente para clases subrepresentadas como "bodegas" que mostraron menor rendimiento. Segundo, el desarrollo de sistemas de detección temporal que integren información secuencial de múltiples capturas para monitoreo continuo, permitiendo no solo detectar zonas de minería sino también analizar su evolución y expansión temporal. Finalmente, la evaluación de arquitecturas más avanzadas como YOLOv8 para mejorar la precisión general, y la integración de modelos de segmentación complementaria (SAM) que proporcionen máscaras de píxel precisas para delimitar con mayor exactitud las áreas afectadas, facilitando así el cálculo de superficies impactadas y el análisis espaciotemporal cuantitativo de la actividad minera ilegal.

## REFERENCES

- [1] J. A. Boscó Arias, "Minería ilegal y el impacto a la seguridad ambiental en la región amazónica en Colombia," in *El crimen organizado en la Amazonía: Escenario de desafíos para la seguridad regional*, A. Cerón Rincón, Ed. Bogotá, Colombia: Sello Editorial ESDEG, 2023, pp. 111–140, doi: 10.25062/9786287602649.05.
- [2] Foro Nacional Ambiental. *Informe Nacional: Minería ilegal y contaminación por mercurio en Colombia*. Diciembre 2024. Disponible en: <https://foronacionalambiental.org.co/wp-content/uploads/2024/12/Informe-Nacional-Mineria-Ilegal-y-Contaminacion-por-Mercurio-en-Colombia-2.pdf>
- [3] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4), 1680–1716. MDPI. <https://www.mdpi.com/2504-4990/5/4/86>
- [4] Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, 38, 35–44. Springer.
- [5] L. da Silva Gomes, E. H. Shiguemori, T. N. Kuck, and D. I. Alves, "Evaluation of a YOLOv8-based method for detecting unauthorized airstrips in the Amazon Rainforest using SAR imagery," *ResearchGate*, 2024, doi: 10.14209/sbrt.2024.1571036260.
- [6] L. J. Rico Trujillo, "Identification of alluvial gold mining in the Colombian Chocó region using radar satellite images," *elib.dlr.de*, 2022. [Online]. Available: <https://elib.dlr.de>
- [7] S. Ozdemir, Z. Akbulut, F. Karsli, and T. Kavzoglu, "Extraction of water bodies from high-resolution aerial and satellite images using visual foundation models," *Sustainability*, vol. 16, no. 7, p. 2995, Apr. 2024, doi: 10.3390/su16072995.
- [8] E. Ansari, M. N. Akhtar, M. N. Abdullah, W. A. F. W. Othman, E. A. Bakar, A. F. Hawary, and S. S. N. Alhady, "Image processing of UAV imagery for river feature recognition of Kerian River, Malaysia," *Sustainability*, vol. 13, no. 17, p. 9568, Jun. 2021, doi: 10.3390/su13179568.
- [9] K. Yang, S. Zhang, X. Yang, N. Wu, and C. Liu, "Flood detection based on unmanned aerial vehicle system and deep learning," *Complexity*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/6155300.



- [10] M. C. Barlow, X. Zhu, and C. L. Glennie, "Stream boundary detection of a hyper-arid, polar region using a U-Net architecture: Taylor Valley, Antarctica," *Remote Sens.*, vol. 14, no. 1, p. 234, 2022, doi: 10.3390/rs14010234.
- [11] Cui, B., Zhang, Y., Li, J., & Liu, Z. (2022). Unmanned Aerial Vehicle (UAV) Object Detection in High-Resolution Image Based on Improved YOLO v5. In *Proceedings of the 3rd International Conference on Computer Information and Big Data Applications (CIBDA)*, Wuhan, China, pp. 1–4.
- [12] Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., & Lan, Y. (2023). Yolo-based UAV technology: A review of the research and its applications. *Drones*, 7(3), 190. MDPI.
- [13] Qiu, Z., Bai, H., & Chen, T. (2023). Special vehicle detection from UAV perspective via YOLO-GNS based deep learning network. *Drones*, 7(2), 117. MDPI.