

UNIVERSITY OF SALAMANCA



VNiVERSiDAD
D SALAMANCA

INTERNATIONAL DOCTORAL THESIS

**A Platform for Smart Infrastructure
Monitoring with Unmanned Aerial
Vehicles and Deep Learning Techniques**

Author:
Luis Augusto SILVA

Supervisors:
Dr. Juan Francisco DE PAZ
SANTANA
Dr. Gabriel VILLARRUBIA
GONZÁLEZ

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy
in Computer Engineering at the
University of Salamanca*

September, 2023

Declaration of Content

This International Doctoral Thesis consists of a compendium of three scientific papers published in high impact factor international journals that are specified below.

1. An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images

Luis Augusto Silva^{1,2}, Héctor Sanchez San Blas¹, David Peral García¹, André Sales Mendes¹, Gabriel Villarrubia González¹

¹Computer and Automation Department, University of Salamanca, Salamanca, Spain

²Laboratory of Embedded and Distribution Systems, University of Vale do Itajaí, Rua Uruguai 458, C.P. 360, Itajaí 88302-901, Brazil

Sensors, 30 October 2020

DOI: 10.3390/s20216205

2. Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning

Luis Augusto Silva¹, André Sales Mendes¹, Héctor Sanchez San Blas¹, Lia Caetano Bastos², Alexandre Leopoldo Gonçalves², André Fabiano de Moraes^{2,3}

¹Expert Systems and Applications Lab, Faculty of Science, University of Salamanca, Plaza de los Caídos s/n, 37002 Salamanca, Spain

²Department of Knowledge Engineering and Management, Federal University of Santa Catarina, Florianopolis 88040-900, Brazil

³Department Information Technology, IT Institute Federal of Science Technology IFC, Camboriú 88340-055, Brazil

Sensors, 23 December 2022

DOI: 10.3390/s23010138

3. Automated Road Damage Detection Using UAV Images and Deep Learning Techniques

Luis Augusto Silva¹, Valderi Reis Quietinho Leithardt^{3,4}, Vivian F. López Batista², Gabriel Villarrubia González¹, Juan F. de Paz Santana¹

¹Expert Systems and Applications Lab, Faculty of Science, University of Salamanca, Plaza de los Caídos s/n, 37002 Salamanca, Spain

²Department of Computer Science and Automatics - University of Salamanca - Salamanca, Spain

³Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, 1959-007 Lisbon, Portugal

⁴VALORIZA, ESTG, Instituto Politécnico de Portalegre, 7300-555 - Portalegre, Portugal

IEEE Access, vol. 11, pp. 62918-62931, 2023

DOI: 10.1109/ACCESS.2023.3287770

Declaration of Authorship

Dr. Juan Francisco DE PAZ SANTANA, full professor of the Computer Science and Artificial Intelligence area at the University of Salamanca and Dr. Gabriel VILLARRUBIA GONZÁLEZ, professor of the Computer Science and Artificial Intelligence area at the University of Salamanca,

CERTIFY

That the present document, entitled “**A Platform for Smart Infrastructure Monitoring with Unmanned Aerial Vehicles and Deep Learning Techniques**” has been prepared under their supervision at the Computer and Automation Department of the University of Salamanca by Luis Augusto SILVA, and constitutes his thesis in the modality of compendium of articles for the degree of Doctor of Philosophy in Computer Engineering.

Dr. Juan Francisco DE PAZ SANTANA

Dr. Gabriel VILLARRUBIA GONZÁLEZ

Date:

Date:

Luis Augusto SILVA

Date:

Declaration of the Academic Commission

The academic committee of the Ph.D. program in computer engineering authorizes Luis Augusto Silva with NIE Y8119893X to present the doctoral thesis entitled **“A Platform for Smart Infrastructure Monitoring with Unmanned Aerial Vehicles and Deep Learning Techniques”** through the compendium of articles modality.

Coordinator of the Ph.D. program in computer engineering

Date:

«Si tú no trabajas por tu sueños, alguien te contratará para que trabajes por los suyos.»

Steve Jobs

Abstract

A Platform for Smart Infrastructure Monitoring with Unmanned Aerial Vehicles and Deep Learning Techniques

In recent years, with the evolution of artificial intelligence (AI) and machine learning techniques, the capture and processing of aerial images has been boosted, providing innovative solutions to social challenges, including road damage monitoring and urban infrastructures. As a result of some recent economic crises, many countries have reduced the maintenance of their roads, which has increased pavement defects, raising the risk of accidents and traffic complications. Various methods have been proposed to solve this problem, ranging from vibration techniques to 3D reconstruction. However, the predominant use of manual inspections or specific vehicles implies high costs. In urban contexts, the growth can be chaotic. It is essential to monitor the expansion with environmental and urban planning criteria. In the past, this task was done manually or with satellite images that were too complex to process. This research focuses on the use of unmanned aerial vehicles to detect and classify objects, using images and techniques based on artificial intelligence. The results of this work have taken the form of 3 publications in prestigious journals. This research work is meeting the Sustainable Development Goals (SDGs), specifically SDG-9 and SDG-11. The main contributions of this research include a comprehensive review of the state of the art, the development of AI models for detection and classification, an advanced image processing algorithm, and dataset creation. The advances presented here justify the improvement of the new search for technological solutions that pursue greater efficiency in the monitoring of roads and urban areas, with the possibility of being applied in other domains such as rural areas, agriculture, and public health.

Resumen

Plataforma para la Monitorización Inteligente de Infraestructuras con Vehículos Aéreos No Tripulados y Técnicas de Aprendizaje Profundo

En los últimos años y con la evolución de la inteligencia artificial (IA) y las técnicas de aprendizaje automático, se ha potenciado la captación y procesamiento de imágenes aéreas, brindando soluciones innovadoras a desafíos sociales, pudiendo destacar entre otros la monitorización de carreteras e infraestructuras urbanas. A raíz de algunas crisis económicas recientes, muchos países han reducido el mantenimiento de sus carreteras, lo que ha aumentado los defectos en el pavimento, elevando los riesgos de accidentes y complicaciones en el tráfico. Para dar solución a esta problemática, se han propuesto varios métodos, desde técnicas de vibración hasta reconstrucción 3D. Sin embargo, el uso predominante de inspecciones manuales o vehículos específicos implica altos costes. Actualmente, son numerosos los centros de investigación que buscan soluciones utilizando inteligencia artificial y procesamiento avanzado de imágenes. En contextos urbanos, donde el crecimiento puede llegar a ser algo caótico, es fundamental monitorizar su expansión con criterios ambientales y urbanísticos particulares. Antiguamente, esta tarea se realizaba manualmente o con imágenes satelitales que resultaban muy complejas para su procesamiento. Esta investigación se centra en la utilización de vehículos aéreos no tripulados para detectar y clasificar objetos, haciendo uso de imágenes y técnicas basadas en inteligencia artificial. Los resultados de este trabajo se han plasmado en forma de 3 publicaciones en revistas de reconocido prestigio. Durante la elaboración de este trabajo de investigación se ha prestado una especial atención para cumplir con los Objetivos de Desarrollo Sostenible (ODS), específicamente con el ODS-9 y ODS-11. Las principales contribuciones de este trabajo de investigación, incluye una revisión exhaustiva del estado del arte, el desarrollo de modelos de IA para detección y clasificación, un algoritmo avanzado de procesamiento de imágenes y la creación de un conjunto de imágenes para la comunidad científica. Los avances aquí plasmados justifican la mejora la nueva búsqueda de soluciones tecnológicas que persiguen una mayor eficiencia en la monitorización de carreteras y zonas urbanas, con la posibilidad de ser aplicados en otros dominios como áreas rurales, agricultura y salud pública.

Agradecimientos

Desde que empecé mis estudios de doctorado muchas personas me han demostrado su apoyo a lo largo de este viaje. Por lo tanto, quiero expresarles mi gratitud por su apoyo y ayuda.

En primer lugar, quiero agradecer a mis directores Dr. Juan Francisco De Paz Santana y Dr. Gabriel Villarrubia González, su dedicación y apoyo a esta investigación. Gracias por la oportunidad e impulso que habéis dado a mi vida académica y profesional. Esta tesis no sería posible sin vosotros.

En segundo lugar, a los profesores que he tenido a lo largo de mi vida académica, especialmente al Dr. Valderi Leithardt por todo su apoyo y motivación.

Me gustaría reconocer y agradecer a cada uno de los miembros del grupo de investigación ESALab la oportunidad de vivir nuevas y emocionantes aventuras en el mundo de la investigación. En particular, me gustaría mencionar a mis compañeros André, Loza, Héctor, Diego, Dani y María por su colaboración y apoyo continuo, así como a mis compañeros de Béjar, Raúl y Torreblanca.

Un agradecimiento especial a mi novia Beatriz, con cuya paciencia, dedicación y apoyo han sido esenciales. Agradezco tus constantes ayudas y consejos en todos los aspectos de mi vida.

Gracias a mis amigos que siempre me han prestado un gran apoyo moral y humano. En especial a Fátima, que siempre será la hermana española que la vida me regaló.

No puedo dejar de mencionar a mis padres, Paulo y Rosa, así como a mis queridos hermanos, Paulo, Clarissa, Patrick y sobrinas. A pesar de la distancia que nos separa de Brasil, su apoyo ha sido un pilar y una fuente constante de inspiración y motivación. Gracias por creer en mí y por estar siempre a mi lado.

A todos los que han sido parte de este ciclo les agradezco por ser parte de mi historia. Su apoyo y amistad significan el mundo para mí, y estoy emocionado por lo que el futuro nos depara. Este logro no habría sido posible sin cada uno de vosotros.

A todos un *MUITO OBRIGADO!*

Índice general

Declaration of Content	III
Declaration of Authorship	V
Declaration of the Academic Commission	VII
Abstract	XI
Resumen	XIII
Agradecimientos	XV
1. Introducción	1
1.1. Hipótesis y objetivos	3
1.2. Metodología de la investigación	4
1.3. Estructura de la tesis	5
2. Estado del arte	7
2.1. Vehículos aéreos no tripulados (UAV)	7
2.1.1. Clasificación	9
2.1.2. Características	11
2.1.3. Áreas de aplicación de UAVs	13
2.1.4. Aprendizaje automático y aprendizaje profundo aplicados a UAV	16
2.2. Procesamiento de Imágenes y Deep Learning	18
2.2.1. Procesamiento de imágenes	19
2.2.2. Deep Learning	21
2.2.3. Arquitecturas de Redes Neurales Convolucionales	21
2.2.4. Arquitecturas de detección de objetos	24
3. Lista de artículos de revistas	29
3.1. Coherencia y relación entre los artículos	29
3.2. An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images	33
3.2.1. Resumen en Español	55
3.3. Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning	59
3.3.1. Resumen en Español	81
3.4. Automated Road Damage Detection Using UAV Images and Deep Learning Techniques	86
3.4.1. Resumen en Español	99

4. Conclusions	103
4.1. Contributions	103
4.2. Future work	105
A. Factor de impacto de las revistas	107
A.1. Sensors	107
A.2. Sensors	107
A.3. IEEE Access	108
B. Conclusiones	109
B.1. Contribuciones	109
B.2. Trabajos futuros	111
Bibliografía	113

Índice de figuras

2.1. Mercado global de UAV comerciales	8
2.2. UAV de ala fija (MQ-9) con motor turbohélice	9
2.3. DJI Mavic Air	10
2.4. DJI Matrice 300	11
2.5. Dron especializado en aplicaciones agrícolas	11
2.6. Aplicaciones en <i>remote sensing</i>	14
2.7. Diagrama de bloques de un sistema de búsqueda utilizando UAVs con <i>Machine Learning</i>	17
2.8. Relación entre las categorías de UAV según técnicas y aplicaciones . . .	20
2.9. Representación de una red neuronal simple y una red neuronal pro- funda	22
2.10. Mask R-CNN	24
2.11. Single Shot Detector [15]	25
2.12. Flujo del algoritmo YOLO	26
2.13. Comparativa entre versiones del algoritmo YOLO [54]	27

List of Abbreviations

AEC	Asociación Española de la Carretera
AR	Active Research
CNN	Convolutional Neural Network
DL	Deep Learning
DSM	Digital Surface Model
DTM	Digital Terrain Model
FPV	First Person View
GPS	Global Positioning System
HAPs	High Altitude Platforms
IA	Inteligencia Artificial
ILE	Inspection of Lines Electrical
IR	Infrared
LAPs	Low Altitude Platforms
LDA	Linear Discriminant Analysis
LoS	Line of Sight
LSTM	Long Short-Term Memory
ML	Machine Learning
m/s	meters per second
RoI	Region of Interest
RS	Remote Sensing
RTM	Real Time Monitoring
SAR	Search and Rescue
SSD	Single Shot Multibox Detector
SVM	Support Vector Machine
UAS	Unmanned Aerial Systems
UAV	Unmanned Aerial Vehicle
USD	United States Dollars
VTOL	Vertical Take-Off and Landing
YOLO	You Only Look Once

Capítulo 1

Introducción

Esta sección justifica la necesidad de monitorizar infraestructuras, especialmente vías de transporte y estructuras urbanas. Desde el ámbito de la ingeniería informática, se realiza un profundo análisis acerca de los recientes avances en la obtención y procesamiento de imágenes aéreas en combinación con las innovaciones en inteligencia artificial y aprendizaje profundo que se han realizado en los últimos años.

El mantenimiento de las diferentes vías de transporte así como el estado de las estructuras logísticas es clave en la seguridad de los desplazamientos. Si el mantenimiento de una carretera es el adecuado y su estado es óptimo, la posibilidad de accidente es inferior al de una vía de transporte que se encuentra en mal estado [1]. La principal solución para minimizar la posibilidad de un accidente, es decrementar los límites de velocidad, sin embargo esto no suele ser suficiente y es de vital importancia analizar y mejorar las condiciones de las vías. Por estos motivos, y con objeto de poder reducir la siniestralidad de las carreteras, es de vital importancia determinar de una forma ágil y precisa el estado de las mismas.

Estudios elaborados por la Asociación Española de la Carretera (AEC) señalan que el mal estado de las carreteras es la causa principal del 94 % de los accidentes producidos por las condiciones de la vía [2]. Según este organismo, uno de cada trece kilómetros de la red de carreteras española presenta daños relevantes en más del 50 % de la superficie del pavimento, como baches y grietas. Estos desperfectos incurren en la seguridad del vehículo, provocando una conducción menos estable y segura.

El informe de la AEC [2], también constata que España suspende en la conservación del estado de las vías de transporte, además, el citado informe califica el estado del asfalto como muy deficiente. Todo ello se traduce en un aumento del riesgo de sufrir un accidente en el peor de los casos, pero también puede llegar a ocasionar averías en los vehículos. Si se quiere minimizar los riesgos de un posible accidente o disminuir la posibilidad de aparición de una avería, es de vital importancia indagar y economizar las técnicas que posibilitan la mejora de las infraestructuras de las vías de transporte. Además, la difícil situación económica que se vive a nivel mundial repercute en el estado de las carreteras, haciendo que la partida de mantenimiento y mejora de las vías de transporte sea cada vez menor. Para poder determinar qué vías son aquellas que requieren de algún tipo de mejora o mantenimiento urgente, se deben determinar las condiciones del pavimento de dicha vía, por tanto, esta acción puede ser llevada a cabo de una forma manual o automática.

Actualmente, en España estas revisiones se realizan de una forma manual, proceso que conlleva un alto coste debido a la necesidad de un experto humano que identifique el estado de dichas carreteras. Esta labor acaba resultando muy costosa y lenta. Adicionalmente, se debe destacar que la toma de decisiones por parte de un

experto puede ser criticada debido a que la decisión sobre el estado final de un tramo de carretera puede basarse en una opinión muy subjetiva del individuo a partir de percepciones que pueden llegar a fallar generando información errónea.

Por otro lado, la evolución de la ciencia y la tecnología junto al abaratamiento de los procesos de fabricación de componentes electrónicos ha permitido indagar en nuevas técnicas a así como en el desarrollo de herramientas que hagan que la detección pueda ser realizada de una forma objetiva y automática. Debido a lo mencionado anteriormente, ser capaces de obtener una imagen de la carretera de forma clara y detectar cualquier tipo de deterioro producido en una vía de transporte supone un reto en el cual muchas universidades y centros de investigación están trabajando.

Actualmente, se han desarrollado diversas técnicas para detectar el deterioro en el pavimento de las carreteras, tales como la utilización de láser [3], sensores de vibración [4] y técnicas basadas en imágenes [5]. Con el auge de los métodos de procesamiento de imágenes, ha crecido el interés en técnicas de *Machine Learning* destinadas a identificar distintos tipos de deterioro en el pavimento. Sin embargo, la implementación de estas técnicas basadas en aprendizaje profundo enfrenta desafíos significativos. En primer término, su precisión depende en gran medida de que la configuración de la cámara sea dinámica, adaptándose al contexto y a las condiciones lumínicas. En segundo lugar, dichas técnicas conllevan una manipulación extensiva de los píxeles de las imágenes, lo que, computacionalmente hablando, resulta ser un proceso complejo cuando se ejecuta en dispositivos con recursos limitados, especialmente, cuando la eficiencia y la duración de la batería son cruciales [6]. Adicionalmente se pueden encontrar en la literatura diversas investigaciones que buscan solucionar estos retos inherentes [7, 8].

Debido al avance producido en los últimos años en cuanto a la miniaturización de sensores, se ha obtenido una gran cantidad de datos que amplían nuestra comprensión del entorno, pero dificulta el análisis no automático tradicional. El aprendizaje profundo, una rama emergente en el ámbito del aprendizaje automático, utiliza redes neuronales para extraer características de estos datos a diferentes niveles de abstracción [9]. En el procesamiento de imágenes, estas técnicas demuestran ser eficientes para diferentes propósitos, como la clasificación pixel a pixel lo que resulta esencial para la monitorización urbana, donde la necesidad de un análisis rápido y eficiente de grandes conjuntos de datos, como las imágenes satelitales, es crucial para supervisar el crecimiento urbano [10, 11].

Al explorar métodos para automatizar la captación de imágenes geolocalizadas, se pueden encontrar soluciones que utilizan vehículos con cámaras próximas al suelo [12]. No obstante, esta técnica es lenta debido a que el vehículo debe moverse a velocidades reducidas para garantizar imágenes detalladas; el movimiento a mayor velocidad decrementa la calidad de la imagen obtenida. En zonas amplias, se requeriría de múltiples vehículos y conductores, elevando los costes. Las imágenes satelitales podrían constituir una alternativa, pero suelen ser antiguas y el acceso a nivel económico es elevado [13]. Conocidas las limitaciones expuestas anteriormente y aprovechando los avances y el abaratamiento de los vehículos aéreos no tripulados, se propone su integración en uno de los artículos presentados en este trabajo.

Los vehículos aéreos no tripulados (UAV) han demostrado su versatilidad en diversas aplicaciones, incluidas las inspecciones urbanas de objetos y entornos. Cada vez estos vehículos son más utilizados para inspeccionar carreteras, ya que ofrecen varias ventajas sobre los métodos tradicionales [14]. Estos vehículos están equipados con cámaras de alta resolución y otros sensores que permiten captar imágenes

de la superficie de la carretera desde múltiples ángulos, alturas y con una gran resolución, proporcionando una visión completa del estado de la calzada. Además, los UAV pueden cubrir un área extensa con relativa rapidez, lo que reduce la necesidad de inspecciones manuales, que pueden resultar peligrosas para los operadores humanos. Por todo ello, el uso de UAV para inspeccionar carreteras ha suscitado gran interés entre los investigadores e ingenieros que profundizan en estos campos de investigación. La combinación de UAV con técnicas de inteligencia artificial, como el aprendizaje profundo, permite desarrollar enfoques eficientes y rentables para la detección de desperfectos en las carreteras. Estos vehículos son utilizados con mucha frecuencia para inspecciones urbanas en construcciones e infraestructuras, tales como piscinas [15], edificaciones [16], vegetaciones [17], y entornos urbanos [18, 19], añadido a eso hay una gran versatilidad también en la utilización para detección de manchas de aceite y control de epidemias como es el caso del mosquito en países endémicos [20].

Debido a lo mencionado anteriormente, se han realizado diferentes estudios utilizando conjuntos de datos para entrenar modelos, incorporando imágenes adicionales captadas por UAV, cámaras montadas en automóviles e imágenes satélites. Para facilitar el proceso de aprendizaje, estos conjuntos de datos se suelen etiquetar para identificar distintos tipos de desperfectos en la carretera, pudiendo destacar baches y grietas entre otros. La anotación de estas imágenes permite al algoritmo aprender a detectar y clasificar con precisión distintos tipos de daños en la carretera. Utilizando un conjunto de datos amplio y diverso, los investigadores pueden mejorar la precisión y fiabilidad de sus modelos, asegurándose de que pueden identificar y tratar eficazmente los distintos tipos de daños en las infraestructuras.

1.1. Hipótesis y objetivos

En los últimos años, son numerosos los grupos de investigación que han centrado sus esfuerzos en avanzar e indagar el desarrollo de nuevas herramientas para la detección de imágenes. A pesar de que el coste de los dispositivos aéreos no tripulados ha descendido en los últimos años, nos enfrentamos a desafíos tecnológicos como la integración de sistemas de detección de imágenes con sistemas de vuelo o el desarrollo de capacidades tecnológicas que no dependan exclusivamente de un único sistema. Ante este panorama, resulta esencial investigar y diseñar arquitecturas innovadoras que ofrezcan soluciones eficientes para la conservación de infraestructuras urbanas.

Dada esta premisa, la hipótesis que fundamenta este trabajo de investigación es la siguiente:

Hipótesis

La utilización de técnicas basadas en inteligencia artificial en combinación con la obtención y clasificación de imágenes aéreas, permite el desarrollo de soluciones tecnológicas con capacidades de detección e identificación de desperfectos sin intervención humana, mejorando significativamente los sistemas de mantenimiento y detección de estructuras que son empleados en la actualidad.

El objetivo principal de este trabajo es el diseño de una plataforma modular y basada en agentes que permita mejorar el sistema de monitorización del estado de las infraestructuras urbanas empleando para ello diferentes imágenes y técnicas basadas en visión artificial. Para validar la hipótesis de esta investigación, se han planteado inicialmente los siguientes objetivos:

- Realizar una revisión exhaustiva del estado del arte sobre tecnologías de inteligencia artificial y aprendizaje profundo aplicadas al análisis de imágenes aéreas.
- Identificar nuevos desafíos en el campo de la monitorización y el mantenimiento de infraestructuras urbanas.
- Diseñar, desarrollar y validar modelos de aprendizaje para la detección precisa de daños en infraestructuras urbanas a partir de imágenes aéreas.
- Explorar arquitecturas inteligentes que faciliten la integración de tecnologías de inteligencia artificial en sistemas de detección de daños en infraestructuras, y que puedan adaptarse de manera dinámica ante cambios o anomalías.
- Diseñar una red de procesamiento de imágenes aéreas que, en combinación con el Edge Computing, permita una rápida identificación y clasificación de potenciales daños en las infraestructuras.
- Entrenar y evaluar un modelo de aprendizaje profunda, apto para ser integrado en dispositivos con limitaciones computacionales, que ayude en la toma de decisiones durante la monitorización automática de infraestructuras.
- Desarrollar y analizar al menos tres casos de estudio diferentes en los que se empleen técnicas de inteligencia artificial y aprendizaje profundo para la detección, prevención y reparación de daños en infraestructuras urbanas.

1.2. Metodología de la investigación

El proceso de investigación y el método de trabajo en el que se ha fundamentado esta tesis es la metodología *Action Research* (AR) [21]. La metodología AR ha demostrado ser efectiva en numerosos proyectos de software de naturaleza similar.

La AR se orienta hacia la resolución de problemas concretos a partir de una hipótesis claramente definida. Una vez establecido el objetivo de investigación, es esencial recopilar y analizar información acerca del tema de estudio. Esta fase de análisis, facilita la concepción de una propuesta óptima orientada a resolver el desafío identificado. Posteriormente, la solución propuesta se somete a validación a través de diferentes casos de estudio. El planteamiento de los casos de estudios y su correcta resolución permitirán corroborar la validez de la hipótesis y la pertinencia de la solución propuesta en su contexto. Si la solución no respalda adecuadamente la hipótesis, la flexibilidad inherente a la metodología AR permitirá iterar en el proceso, afinando propuestas hasta que la hipótesis se confirme o desestime.

La metodología AR se caracteriza por emplear un método cíclico conformado por las siguientes fases:

- **Definición del problema:** identificación de los problemas y sus variables para obtener la hipótesis de investigación y los objetivos de la misma.

- **Revisión del estado del arte:** un revisión del estado de la literatura y del estudio de los conceptos y técnicas básicas y fundamentales relacionados con la hipótesis de investigación.
- **Análisis y diseño de una propuesta:** un análisis profundo del estado del arte para definir una propuesta que busque comprobar la hipótesis.
- **Definición de estudios de caso:** una definición de uno o varios estudios de caso que buscan probar la propuesta diseñada.
- **Pruebas y extracción de resultados:** se ejecutan los estudios de caso y se obtienen los resultados para su posterior análisis.
- **Análisis y discusión de resultados:** se evalúan y explican en detalle los resultados obtenidos para conocer la razón de los mismos.
- **Definición de conclusiones:** se evalúan las conclusiones con el fin de aceptar y/o refutar la propuesta y/o la hipótesis. Después de este paso, el ciclo puede reiniciarse.

La utilización de esta metodología de investigación permite validar su estudio aplicada a casos reales, facilitando así su transferencia a la sociedad. Además, esta metodología busca comunicar los resultados obtenidos en la investigación a través de publicaciones en revistas o conferencias. Esta última afirmación es de gran importancia, ya que el objetivo principal de toda investigación es transferir el conocimiento adquirido al resto de la comunidad científica.

1.3. Estructura de la tesis

Una vez definida la hipótesis de partida de este trabajo, es esencial profundizar en su estudio. La exploración y la búsqueda de una solución que culmine los objetivos iniciales de este trabajo, conformará los siguientes capítulos de este documento. A continuación, se presenta su estructura organizativa.

- **Capítulo 2. Estado del arte:** Este capítulo aborda tres pilares esenciales. Primero, examina la evolución y aplicabilidad de los UAVs en la monitorización de infraestructuras haciendo foco en vías de transporte y zonas urbanas. Segundo, se adentra en la visión por ordenador y el impacto del aprendizaje profundo en el procesamiento de imágenes. Finalmente, se investigan las arquitecturas y técnicas avanzadas de aprendizaje.
- **Capítulo 3. Lista de artículos de revista:** El tercer capítulo de la tesis doctoral está conformado por los artículos que han dado lugar a que esta tesis sea presentada en la modalidad de compendio de artículos. Primero, se presenta una sección que describe la relación entre los artículos que conforman el trabajo. Posteriormente, se incluyen los artículos originales, seguidos de un resumen en español de cada uno. Estos resúmenes abarcan los objetivos de investigación, la metodología empleada, la explicación de los resultados y las conclusiones derivadas del artículo.
- **Capítulo 4. Conclusiones:** Este capítulo encapsula las principales conclusiones derivadas del trabajo de investigación llevado a cabo durante el periodo doctoral. En el se reflejan los hallazgos más significativos, las implicaciones de la investigación llevada a cabo y las líneas futuras que podrían llevarse a cabo.

Capítulo 2

Estado del arte

Este capítulo detalla el conocimiento fundamental y los trabajos previos relacionados con la hipótesis de investigación y objetivos de esta tesis. El capítulo ha sido segmentado en tres secciones esenciales: UAVs, algoritmos utilizados en el procesamiento de imágenes y los sistemas de monitorización.

La primera sección está dedicada a los UAVs más particularmente aquellos cuyo uso se encuentra orientado a la monitorización de infraestructuras. Esta sección muestra la evolución de los UAVs, desde sus usos iniciales hasta los últimos avances que han permitido resolver retos de la sociedad actual mediante la resolución de tareas complejas. Adicionalmente, en esta parte se enfatiza en los diversos desafíos y ventajas de la utilización de UAVs para el análisis de infraestructura, especialmente en la monitorización de carreteras y espacios urbanos. Se explorará el alcance y las capacidades y mejoras tecnológicas de los UAVs a lo largo de los años, preparando el escenario para su integración con los sistemas inteligentes.

La sección siguiente profundiza en el procesamiento digital de imágenes y la visión por ordenador detallando los algoritmos existentes más importante. Esta sección hará foco en la importancia de la inteligencia artificial en el reconocimiento y procesamiento de imágenes. Esta sección clarifica los desafíos específicos de la visión por ordenador y cómo los modelos de aprendizaje profundo, en particular las Redes Neuronales Convolucionales, se adaptan para resolverlos. No únicamente se describirá los algoritmos, sino que también se reflexionará sobre su eficacia, limitaciones y las áreas potenciales de mejora en el dominio de detección de objetos.

La última sección profundiza en el ámbito de los sistemas multi-agente, resaltando las ventajas de uso, funcionalidad..etc. Se discutirá sobre la aplicabilidad de tales sistemas en la monitorización de infraestructuras, especialmente en combinación con UAVs y aprendizaje profundo. También se hará foco en las implementaciones punteras, estudios de caso novedosos y la evolución prospectiva de sistemas multi-agente en la monitorización de infraestructuras urbanas y carreteras.

2.1. Vehículos aéreos no tripulados (UAV)

En la actualidad, el uso de drones se está extendiendo cada vez más a nuevos campos de aplicación, que van desde la agricultura hasta la seguridad. Una de estas nuevas aplicaciones es la grabación de sonidos en zonas de difícil acceso. El reto que se plantea al utilizar drones con este fin es que hay que separar el sonido de las fuentes grabadas del ruido producido por el propio dron. La intensidad del ruido emitido por el dron depende de varios factores, como la potencia del motor, la velocidad de rotación de la hélice o el tipo de rotor. La reducción del ruido es, por tanto,

uno de los mayores retos para las próximas generaciones de UAV y sistemas aéreos no tripulados (UAS).

Los vehículos aéreos no tripulados, también conocidos como drones, son una tecnología emergente utilizada tanto por civiles como por militares. De hecho, la popularidad de este tipo de vehículos ha aumentado tanto, que ya no se utilizan exclusivamente para el trabajo, sino también para el ocio. La utilización de los mismos que es cada vez mayor, ha sido posible gracias a la fabricación de drones más baratos, precisos y fáciles de usar [22].

El mercado global de UAV comerciales tuvo un valor de 13.46 mil millones de USD en 2020 y se espera que crezca hasta alcanzar una tasa compuesta anual del 57.3% durante el período pronosticado. Los UAV comerciales son aeronaves autónomas o pilotadas remotamente utilizadas en el sector comercial. Esto se debe principalmente a dos razones. El sector de los vehículos y sistemas aéreos no tripulados ha crecido rápidamente en los últimos años, haciendo tareas muy complejas de realizar antes, por ejemplo, el control de los incendios forestales, la localización de personas desaparecidas tras catástrofes naturales, la agricultura [23], o incluso la vigilancia de carreteras. La Figura 2 muestra el crecimiento de los ingresos en los continentes para el mercado de UAV comerciales en diferentes sectores.

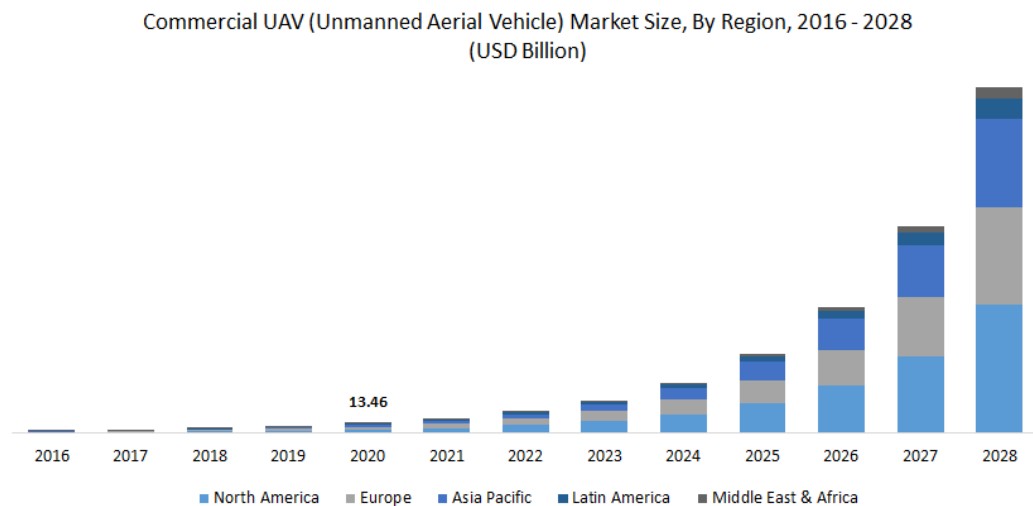


FIGURA 2.1: Mercado global de UAV comerciales

Existen diferentes tipos de UAV comerciales, incluidos los rotativos y de ala fija. Los drones rotativos se usan en sectores como la agricultura o la supervisión de infraestructuras, pero son más lentos y de menor alcance que los de ala fija, que se utilizan para tareas como la fotografía aérea debido a su capacidad de volar más horas a velocidades con mayores exigencias de estabilidad.

El mercado está experimentando un notable crecimiento debido al aumento en las aplicaciones de UAVs en diversas industrias. Esta expansión es potenciada por la creciente demanda de servicios basados en ubicación. A su vez, los avances tecnológicos en cámaras y software refuerzan aún más esta tendencia ascendente. Las mejoras tecnológicas y la miniaturización han incentivado a empresas para invertir en UAV comerciales. Otros impulsores incluyen la demanda creciente de industrias el uso de UAV en tareas de alto riesgo. Por el contrario, las preocupaciones de seguridad y privacidad podrían obstaculizar el crecimiento. La demanda en economías

emergentes y la creciente conciencia representan oportunidades de crecimiento para el mercado de UAV comerciales en el futuro.

El uso de UAV como fines recreativos también ha tenido un gran aumento, en particular, aquellos que son empleados para grabación de fotos y vídeos, carreras de drones, seguridad y tareas industriales. Debido a la gran variedad de usos que pueden tener un dron existen dispositivos con diferentes tamaños, tipos de hélices, geometrías de brazos o potencia de motores.

Sin embargo, aunque son útiles para muchas tareas, los aviones teledirigidos tienen ciertos problemas que pueden restringir su funcionamiento. Uno de los más significativos es el ruido producido por las hélices. Este inconveniente puede hacer que el dron sea inadecuado en situaciones en las que existen regulaciones sobre el ruido o cuando las altas emisiones de ruido pueden afectar negativamente a la tarea realizada [24, 25]. Por ejemplo, muchos aeropuertos de todo el mundo imponen limitaciones estrictas al nivel de ruido permitido durante el día o la noche, añadido a eso, la firma acústica producida por una aeronave militar influye considerablemente en que pueda pasar desapercibida.

2.1.1. Clasificación

Los vehículos aéreos no tripulados pueden clasificarse en términos generales como dispositivos de ala fija y dispositivos de ala rotatoria. Los primeros UAV eran de ala fija, una extensión de las pequeñas aeronaves controladas a distancia, equipados con electrónica adicional para la navegación y reconocimiento. Un gran ejemplo de ala fija pesada UAV militar se ilustra en la Figura 2.2.



FIGURA 2.2: UAV de ala fija (MQ-9) con motor turbohélice

Los UAVs varían en especificaciones, equipamiento, tamaños, alcance y formas. Están disponibles en el mercado con diferentes números de rotores o hélices. Estos vehículos se han desarrollado con distintos motores y estructuras de alas y pueden comunicarse usando tecnologías inalámbricas de corto y largo alcance. Dependiendo de su tamaño, se pueden clasificar como nano, micro o grandes. Gracias a avances

continuos, son excelentes para proporcionar conectividad celular. Están equipados con gafas FPV (Vista en Primera Persona), GPS, sensores, estabilizadores y cámaras. En este estudio, se identifican cuatro categorías de UAVs:

- **Ala fija:** Estos UAVs se componen de alas, cuerpo principal, motor y hélice. Requieren una formación extensa basada en habilidades para operar y pueden permanecer equilibrados verticalmente en el aire durante unas dieciséis horas. Sin embargo, no pueden moverse hacia atrás, mantenerse en un punto en el aire ni girar. No son útiles para ciertas tareas, como la fotografía aérea. Son comúnmente usados para inspecciones de líneas eléctricas y cartografía aérea.
- **Híbrido de ala fija:** Se basan en la automatización y el planeo manual. No son buenos para el vuelo hacia adelante ni para mantenerse en un punto en el aire. Se caracterizan por el despliegue vertical (Vertical takeoff and landing) VTOL.
- **Rotor único:** Son caros y requieren formación para operar. Estos UAVs son mecánicamente complejos y susceptibles a desafíos como vibraciones y ruido.
- **Multi-rotor:** Son los UAVs más económicos y fáciles de fabricar, utilizados comúnmente para imágenes y vigilancia de vídeo. Los multi-rotores pueden ser tricópteros, cuadricópteros, hexacópteros u octocópteros. Los cuadricópteros son los más populares debido a su aterrizaje vertical, alta maniobrabilidad, agilidad, capacidades de despegue, estructura simple, rentabilidad y tamaño reducido.

La versatilidad de los multi-rotores es destacable, permitiendo un desplazamiento en todas las direcciones —tanto horizontal como verticalmente— y la habilidad de permanecer estáticos en una posición determinada. Existen diversas configuraciones populares de multi-rotores. Los cuadricópteros ligeros, dotados de cuatro rotores, son de las variantes más reconocidas, tal como se muestra en la Figura 2.3. Otra configuración innovadora se encuentra en los cuadricópteros con motores orientados hacia abajo, ilustrados en la Figura 2.4. Además, los hexacópteros, equipados con seis rotores, presentan una configuración robusta, representada en la Figura 2.5.



FIGURA 2.3: DJI Mavic Air

El DJI Mavic Air, ejemplificado en la figura anterior, es un dron comercial de introducción, pero que ha ganado una amplia aceptación en el mercado. Pese a su diseño compacto, ofrece una capacidad de vuelo que se extiende alrededor de 25



FIGURA 2.4: DJI Matrice 300



FIGURA 2.5: Dron especializado en aplicaciones agrícolas

minutos, limitado principalmente por la capacidad de su batería. Este problema ya se ha solucionado en las versiones más actuales del dispositivo.

La última figura muestra un dron especializado en la aplicación de insumos agrícolas, destacando la versatilidad y adaptabilidad de estos dispositivos para diversas aplicaciones.

Los cuadricópteros y hexacópteros son más sencillos y comunes en su uso, sobre todo con la configuración de cuadricóptero que se está haciendo especialmente popular para los pequeños vehículos aéreos no tripulados. Tienen un diseño mucho más sencillo, pero son menos redundantes en caso de pérdida del motor. Los octacópteros, sin embargo, son máquinas mucho más potentes que los cuadricópteros o los hexacópteros. Lo que se diferencia es la carga útil que estos vehículos pueden cargar. Determinados servicios de monitorización o inspección necesitan cámaras específicas que no tienen el mismo peso que una cámara común.

2.1.2. Características

Velocidad y tiempo de vuelo

Los UAVs más pequeños pueden volar a una velocidad inferior a 15 m/s, mientras que los UAVs grandes pueden alcanzar velocidades de hasta 100 m/s. Es esencial controlar adecuadamente la velocidad de cualquier UAV en diferentes puntos

de giro de su trayectoria, especialmente cuando sigue una ruta específica para mejorar su eficiencia espectral o energética. Según Azade, Ming y Mahbub [26] hay que analizar el equilibrio entre la agilidad en los giros del UAV y su velocidad.

El tiempo de vuelo de un UAV se refiere al tiempo máximo que puede volar hasta que se agote su batería. Factores como el tamaño, peso y las condiciones climáticas influyen significativamente en la vida útil de la batería de los UAVs. Los UAVs más grandes pueden volar durante horas, mientras que los más pequeños tienen una duración limitada de 20 a 30 minutos. El tiempo de vuelo también puede verse afectado por el sistema de piloto automático y el GPS. El tiempo de vuelo es crucial, al igual que el costo y precio del UAV. Estas informaciones son esenciales para una investigación, en el caso de esta tesis, se necesitan soluciones innovadoras para superar la limitada resistencia de los UAVs, ya que esto es esencial para su despliegue a gran escala y el éxito de sus misiones.

Capacidad de carga

La capacidad de carga (payload) de un UAV se refiere a su habilidad para transportar diferentes cargas. Esta capacidad puede oscilar entre unos pocos gramos hasta cientos de kilogramos. Un mayor peso de carga permite transportar más accesorios, pero suele resultar en un tiempo de vuelo reducido, un mayor consumo de batería y un tamaño más grande del UAV. Los equipamientos más comunes que los UAVs transportan son sensores y cámaras de video destinados a la vigilancia, reconocimiento o aplicaciones comerciales. Además, los UAVs pueden portar equipos de usuario celular, como teléfonos móviles o tabletas, que pesen menos de 1 kg. Se reconoce que cargas pesadas pueden disminuir el tiempo de vuelo de los UAVs. Sin embargo, si un UAV tiene una mayor superficie y más motores, puede almacenar más energía, lo que a la postre podría extender su tiempo de vuelo. Por lo tanto, la calidad del equipo transportado puede ayudar a que el UAV recorra mayores distancias manteniendo la misma precisión y resolución.

Alcance y altitud

El alcance de un UAV indica la distancia desde la cual puede ser controlado de manera remota. Este alcance varía: mientras que los drones pequeños pueden ser controlados a unos pocos metros de distancia, los drones más grandes tienen un alcance de cientos de kilómetros. Por otro lado, la altitud se refiere a la máxima altura que un dron puede alcanzar en vuelo. Típicamente, las plataformas aéreas se dividen en dos categorías según su altitud:

- **Plataformas de baja altitud (LAPs):** Estas plataformas se despliegan comúnmente para respaldar la comunicación celular debido a su rápida implementación y rentabilidad. Además, las LAPs proporcionan una trayectoria visual directa (LoS) que puede mejorar considerablemente el rendimiento de la comunicación.
- **Plataformas de alta altitud (HAPs):** Plataformas como globos se utilizan también para la conectividad celular. Las HAPs brindan una cobertura más amplia en comparación con las LAPs. Aunque el despliegue de HAPs es más complejo, generalmente se ven como medios para respaldar la conectividad a Internet.

2.1.3. Áreas de aplicación de UAVs

Los avances en las capacidades de los sensores desplegados en UAVs han dado lugar a la creación de una nueva generación de servicios y diversas aplicaciones para operaciones no tripuladas. En esta sección, hemos discutido brevemente algunas de las aplicaciones más destacadas de los UAVs.

Vigilancia y monitorización

Los UAVs se han convertido en herramientas esenciales para misiones de vigilancia militar. Diversos países han incorporado estos dispositivos voladores a sus estrategias de defensa. Estos robots aéreos se emplean en la detección de enemigos, combate a la caza furtiva, control fronterizo y supervisión marítima de rutas críticas. Debido a su bajo costo, confiabilidad y versatilidad, los UAVs han ganado relevancia en la vigilancia aérea y el monitorización de áreas específicas para prevenir actividades ilegales. Por ejemplo, ante cualquier amenaza, los drones pueden identificarla y monitorear movimientos en zonas restringidas. Los UAVs facilitan estos servicios al generar alertas automáticas, reduciendo la necesidad de intervención manual [27].

Gestión de emergencias

En situaciones de desastres, ya sean provocados por el hombre o fenómenos naturales como ataques terroristas, tsunamis o inundaciones, los drones pueden acceder a zonas peligrosas donde la presencia humana podría estar en riesgo. Estos eventos catastróficos pueden causar graves daños a infraestructuras como las telecomunicaciones, transportes y servicios básicos como el agua y la electricidad. Los drones facilitan la recopilación de información, permitiendo respuestas rápidas y la navegación entre escombros. Gracias a radares, sensores y cámaras de alta calidad, estos dispositivos pueden ayudar a los equipos de rescate a determinar el grado de daño y a iniciar rápidamente operaciones de recuperación, enviando recursos como kits médicos y helicópteros tripulados. En situaciones como incendios forestales, un grupo de drones equipados con extintores puede monitorizar y examinar áreas sin poner en peligro vidas humanas. De esta forma, los drones permiten la cobertura en tiempo real de extensas áreas protegiendo al personal involucrado. Además, las alertas tempranas mediante drones pueden ser cruciales para salvar vidas humanas y de la fauna en peligro [28].

Remote sensing

La tecnología de drones, incluso a nivel de amadores, está siendo utilizada para obtener imágenes de alta resolución de áreas remotas, como islas, cimas de montañas y costas. La tecnología de drones sirve como puente entre la detección remota desde el aire, el espacio y el suelo. Sus características económicas y ligeras permiten observaciones de calidad con resoluciones temporales y espaciales elevadas. Las capacidades de detección remota de los drones son útiles en la detección de enfermedades, monitorización de calidad del agua, seguimiento de sequías, industria del petróleo y gas, estimaciones de rendimiento, modelado hidrológico, conservación de la biodiversidad, estudios de desastres geológicos, mapeo de terrenos y bosques, y monitorización de cultivos. Además, esta tecnología se utiliza para el mapeo colaborativo y la creación de mapas ambientales en 3D, convirtiéndose en una herramienta esencial para la arqueología y la cartografía. Los drones económicos ofrecen datos

actualizados que se ajustan a los presupuestos de planificación territorial sin depender de fuentes de mapeo obsoletas. Las diversas aplicaciones de los drones en la detección remota se ilustran en la Figura 2.6



FIGURA 2.6: Aplicaciones en *remote sensing*

Operaciones de búsqueda y rescate

Los drones son esenciales en situaciones críticas, como la gestión de desastres, operaciones de rescate y seguridad pública. Su capacidad para brindar imágenes en tiempo real de ubicaciones específicas permite a los equipos de rescate identificar con precisión dónde se necesita ayuda con urgencia. En situaciones adversas, como la desaparición de personas, aludes, incendios forestales o infiltraciones de gas tóxico, los drones aceleran las operaciones de búsqueda y rescate. Estos dispositivos pueden seguir a montañistas perdidos o salvar vidas en desiertos o bosques remotos. Así, ayudan a localizar a las víctimas en terrenos difíciles o condiciones atmosféricas extremas. Antes de la llegada de ambulancias o equipos médicos, los drones pueden suministrar productos médicos esenciales, equipados con kits médicos, chalecos salvavidas y alimentos, proporcionando recursos esenciales a personas aisladas en áreas inaccesibles.

Inspección de infraestructuras

La cartografía de construcciones, el monitorización de obras y la inspección de sitios se han vuelto más eficientes y rápidos gracias a los drones. Estos permiten supervisar proyectos de construcción desde su inicio hasta su finalización, garantizando un progreso de calidad en las obras. Los drones pueden proporcionar informes con imágenes, videos y mapeo 3D a las partes interesadas. Su aplicación ha revolucionado la inspección de infraestructuras y construcciones. Hay un creciente interés en su uso para la inspección de torres GSM, monitorización de gasoductos, inspección de líneas eléctricas y supervisión de proyectos de construcción [29].

Agricultura de precisión

La utilización de drones en la agricultura de precisión permite recopilar datos de sensores terrestres (como calidad del agua, propiedades del suelo y humedad), rociar pesticidas, detectar enfermedades, programar riegos, detectar malezas y monitorear y gestionar cultivos. Integrar drones en este sector es una técnica que ahorra tiempo y es eficiente en costos, potenciando la rentabilidad, la productividad y el rendimiento de los cultivos. Además, estos dispositivos facilitan la detección de daños causados por plagas, el monitorización de malezas, la fumigación y la gestión agrícola, lo que resulta en un mejor rendimiento de los cultivos según las necesidades de producción. Al combinar drones con sensores remotos, se está revolucionando la agricultura de precisión, proporcionando resoluciones temporales, espaciales y espectrales avanzadas [30].

Monitorización en tiempo real del tráfico vial

En el monitorización en tiempo real (RTM, por sus siglas en inglés), se puede lograr la completa automatización del sector de transporte a través de drones [105]. Esto incluiría la automatización de equipos de rescate, topógrafos de carreteras, policía de tráfico y equipos de apoyo en campo. Los drones confiables e inteligentes pueden ayudar en la automatización de estos elementos. Los drones han surgido como nuevas herramientas prometedoras para recopilar datos sobre las condiciones del tráfico en autopistas. A diferencia de los dispositivos de monitorización convencionales, como sensores de microondas, cámaras de vigilancia y detectores de bucle, los drones rentables pueden monitorear grandes segmentos de carreteras [102]. Los drones pueden ser operados por la policía local para obtener una visión clara de accidentes viales o grandes operativos de seguridad en actividades delictivas en autopistas, como el robo de autos. Otras aplicaciones incluyen la identificación de vehículos, redadas a vehículos sospechosos, persecuciones a ladrones armados y secuestradores, o cualquier persona que viole las normas de tráfico. También pueden detectar vehículos que exceden los límites de velocidad, accidentes y ayudar a evitar atascos y congestiones masivas [31]. En el caso de España, la Dirección General de Tráfico en cooperación con la Guardia Civil ya hace uso de este tipo de dispositivos en las carreteras y autopistas [32].

De manera general, los drones pueden ser utilizados para monitorización de condiciones de las autopistas y carreteras, como indicar cualquier grieta y proporcionar una advertencia temprana para evitar posibles accidentes y reducir daños. Lo que es la principal aportación y discusión de esta tesis. Actualmente, la monitorización e inspección de carreteras se llevan a cabo mediante vehículos tripulados y su nivel de automatización debe mejorarse. La integración de la tecnología de inspección vial y los drones puede reducir significativamente los daños en las superficies. Además, se pueden implementar algoritmos de detección de daños utilizando imágenes tomadas desde drones para una mejora eficiente [33].

Restauración forestal automatizada

En recursos forestales, los drones son utilizados para monitorización y también para la actuación como agentes restauradores. Los drones pueden ser desplegados para asistir en diversas tareas necesarias para la implementación de la restauración forestal, como son el estudio del sitio, estrategia de restauración, infraestructura del sitio, suministro de semillas, gestión del sitio (fertilización, deshierbe, etc.), y el estudio

de biodiversidad después de intervenciones de restauración. Los drones ayudan a llevar a cabo estudios preliminares del sitio previos a la restauración y monitorear varios aspectos de la recuperación de la biodiversidad. Pueden monitorear cambios en el clima, composición del ecosistema y funciones de los bosques, así como apoyar en la inspección de restauraciones forestales [23]. Las cámaras de alta resolución instaladas en los drones pueden proporcionar datos pertinentes sobre los ecosistemas forestales para asistir en proyectos de restauración forestal. La adecuada resolución de las cámaras de los drones puede ayudar en la caracterización y análisis de áreas forestales debido a su facilidad para adquirir datos y su versatilidad. Además, los sensores ópticos en los drones también se utilizan para obtener características geométricas de los bosques, como el diámetro y la altura del dosel. De manera similar, las operaciones de teledetección desde drones también son sustitutos fiables y eficientes para las mediciones forestales tradicionales.

Inspección de líneas eléctricas

La detección y prevención de fallos en las líneas eléctricas es crucial para la disponibilidad y fiabilidad del suministro eléctrico. Las desventajas de las técnicas tradicionales incluyen alto costo, despliegue engorroso y riesgos peligrosos. Por lo tanto, la inspección de líneas de distribución y transmisión eléctrica asistida por drones, ha despertado un interés significativo entre los investigadores. La inspección de líneas eléctricas también se refiere a la seguridad de una red de transmisión eléctrica. El uso de drones equipados con cámaras digitales para tomar imágenes de los corredores de líneas eléctricas es un enfoque conveniente para apoyar estas tareas de inspección [34]. Los drones pueden usarse para rastrear pilones eléctricos en busca de pernos dañados, corrosión o oxidación y golpes de rayos. El cortocircuito de estas líneas eléctricas suele ocurrir debido a condiciones climáticas adversas, incendios forestales y caídas de árboles. En un estudio reciente [35], los autores discutieron la instalación de drones en las líneas eléctricas aéreas para identificar fallos. Tanto los drones como los robots trepadores pueden usarse para localizar fallos. Los drones pueden realizar estas operaciones de inspección a un costo menor que los helicópteros y con un riesgo más bajo en comparación con las patrullas a pie tradicionales. Entre los diferentes drones, los drones de ala fija que pueden volar más rápido y más alto son los más preferidos para inspecciones preliminares y monitorización de vegetación. En contraste, los drones multi-rotor se utilizan para obtener imágenes al permanecer en el aire en una posición más cercana a los objetos. Los drones multi-rotos son adecuados debido a su alta maniobrabilidad en 3D. A pesar de estos potenciales, el uso de drones multi-rotor requiere una planificación y piloto de misión específicos. Se deben investigar algoritmos avanzados de recopilación, compartición y procesamiento de datos para redes cooperativas de drones para apoyar inspecciones más fiables, eficientes y rápidas.

2.1.4. Aprendizaje automático y aprendizaje profundo aplicados a UAV

Las técnicas de *Machine Learning* y *Deep Learning* juegan un papel prometedor en diferentes aplicaciones relacionadas con los drones, como la programación de baterías, la planificación de trayectorias, el seguimiento, la evitación de obstáculos, detección de objetos y la asignación de recursos. El desarrollo de nuevas herramientas de ML y el aumento del poder computacional a bordo ayudan a desarrollar nuevos dispositivos más inteligentes, ligeros y pequeños, capaces de realizar cualquier

operación sin el riesgo de colisión. Usando estas herramientas, los drones pueden modificar autónomamente su movimiento, dirección y ubicación para servir a los usuarios en tierra. Además, la disponibilidad de datos precisos puede apoyar a los drones en tareas inteligentes de control, planificación de trayectorias y visión.

Al instalar diferentes cámaras en un drone, se pueden capturar varios tipos de imágenes para su posterior procesamiento. La planificación de un drone, incluyendo trayectoria, navegación y manipulación, se puede realizar para encontrar un camino deseado. De manera similar, el movimiento de los usuarios en tierra y la comparación de carga se pueden obtener para la planificación de trayectorias de drones. Aunque se han implementado métodos convencionales de aprendizaje automático en drones, los métodos de aprendizaje profundo no se han explorado debido a recursos de energía limitados y capacidades de procesamiento. Por lo tanto, la comunidad investigadora ya investiga métodos de aprendizaje profundo eficientes y de bajo consumo, especialmente para operaciones SAR (Búsqueda y Rescate).

Para ello, se deben emplear técnicas de procesamiento de imágenes para encontrar objetos objetivo de manera rápida y precisa. Los métodos de procesamiento de imágenes pueden utilizarse en sistemas autónomos de drones, tanto individuales como múltiples, para localizar posibles objetivos en apoyo de las operaciones SAR. Además, se puede aumentar la información de ubicación en las imágenes aéreas de objetos objetivo. Los drones pueden integrarse con tecnologías de detección de objetivos, incluyendo cámaras térmicas y de visión. Las cámaras térmicas (por ejemplo, cámaras IR) pueden detectar el perfil térmico para localizar personas desaparecidas. Las cámaras de visión nocturna también pueden ayudar en el proceso de detección de objetos y personas [36]. En la literatura, se han reportado métodos que utilizan una combinación de cámaras térmicas y de visión nocturna, como en [37].

Técnicas de aprendizaje automático pueden aplicarse a imágenes capturadas por drones para ayudar en operaciones de búsqueda. En un estudio, utilizan una Red Neuronal Convolutiva (CNN) pre-entrenada junto con una Máquina de Vectores de Soporte lineal (SVM) entrenada para determinar el fotograma exacto de la imagen/video en el que potencialmente se detecta a una persona perdida. Un ejemplo de como el proceso se encuentra en la Figura 2.7.

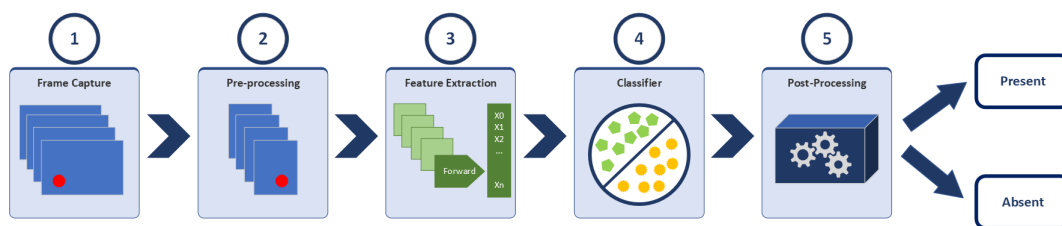


FIGURA 2.7: Diagrama de bloques de un sistema de búsqueda utilizando UAVs con *Machine Learning*

El flujo del proceso es similar para la mayoría de los algoritmos de procesamiento de imágenes captadas a distancia. En el caso de los algoritmos para *Remote Sensing*, primero se utiliza el archivo log del UAV para proporcionar estimaciones iniciales sobre la posición y orientación de cada imagen. Luego, el algoritmo implementa un proceso de triangulación aérea mediante el cual restablece las verdaderas posiciones y orientaciones de las imágenes de la misión aérea. Durante este proceso, el algoritmo genera una gran cantidad de puntos de vinculación automáticos para puntos conjugados identificados en múltiples imágenes.

A continuación, un ajuste de paquete de imágenes utiliza estos puntos para optimizar las posiciones y orientaciones fotográficas, generando una alta cantidad de observaciones redundantes que se emplean para derivar una solución eficiente mediante un ajuste de mínimos cuadrados riguroso. Para ofrecer una comprobación independiente sobre la precisión del ajuste, el algoritmo incluye varios puntos de control. Posteriormente, las imágenes orientadas se utilizan para crear un modelo de superficie digital que ofrece una representación detallada de la superficie del terreno, incluyendo las elevaciones de objetos prominentes como árboles y edificios. Este modelo genera una nube de puntos densa al emparejar características en múltiples pares de imágenes [38].

En esta etapa, se puede generar un modelo digital del terreno, conocido como modelo de terreno desnudo. Este modelo es más útil que el de superficie porque se elimina el ruido de alta frecuencia asociado con la cobertura vegetal. Tras generar el modelo digital del terreno, se puede realizar un proceso de ortorrectificación para eliminar las distorsiones de las imágenes originales. Una vez concluido este proceso, el algoritmo fusiona las imágenes individuales en un mosaico, ofreciendo una imagen continua del área de misión con la resolución deseada.

Tras la revisión exhaustiva de las diversas aplicaciones y características de los UAVs, queda claro que estos dispositivos no se limitan meramente a la captura de imágenes; en realidad, constituyen una herramienta primordial para la adquisición de extensos conjuntos de datos visuales. Ante este escenario, surge una cuestión fundamental: ¿cuál es el siguiente paso tras la captura de tales imágenes? La solución se halla en el dominio del procesamiento de imágenes y del deep learning. Si bien los UAVs actúan como instrumentos de recolección de datos desde una perspectiva aérea, es mediante estas avanzadas técnicas que se procede a la interpretación y análisis detallado de la información captada, permitiendo así la extracción de conocimientos relevantes.

2.2. Procesamiento de Imágenes y Deep Learning

En la era digital actual, la cantidad de imágenes y vídeos capturados diariamente es colosal. Solo con UAVs, obtenemos terabytes de datos visuales que, si se procesan adecuadamente, pueden revelar patrones, anomalías y características esenciales que el ojo humano podría pasar por alto. Aquí es donde el procesamiento de imágenes se vuelve crucial.

El procesamiento de imágenes abarca una variedad de técnicas que permiten mejorar, interpretar y transformar imágenes. Estas técnicas pueden variar desde simples ajustes de contraste hasta complejas operaciones que permiten identificar objetos específicos dentro de una imagen.

Pero el verdadero avance en este campo viene con el *Deep Learning*. Esta subárea del *Machine Learning* utiliza redes neuronales con múltiples capas (conocidas como redes neuronales convolucionales o CNNs, por sus siglas en inglés) para analizar imágenes. Estas redes pueden ser entrenadas para reconocer patrones, caras, objetos y mucho más, con un nivel de precisión que a menudo supera al del ojo humano.

Los beneficios son inmensos: desde la detección temprana de enfermedades en imágenes médicas hasta el reconocimiento de objetos para vehículos autónomos. En el contexto de los UAVs, el *Deep Learning* puede permitir, por ejemplo, identificar

automáticamente áreas afectadas en una imagen post-desastre o detectar signos de estrés en cultivos agrícolas.

En esta sección, nos adentraremos en los algoritmos clave y las aplicaciones del procesamiento de imágenes y deep learning, ilustrando cómo estas tecnologías están redefiniendo lo que es posible en el análisis de datos visuales.

2.2.1. Procesamiento de imágenes

Cada aplicación de UAVs demanda técnicas específicas de procesamiento de imágenes y modelos de aprendizaje automático para interpretar y enriquecer las imágenes capturadas de acuerdo a su área de aplicación. En función de los objetivos perseguidos con estas imágenes, es factible clasificar de manera categórica las aplicaciones de los UAVs. Estas finalidades pueden agruparse en segmentación y análisis, identificación y predicción, y reconstrucción 3D. La Tabla 2.1 presenta una propuesta de clasificación categórica de las aplicaciones de UAVs basadas en estos objetivos, mientras que la Figura 2.8 ilustra las interrelaciones entre dichas categorías.

Categoría	Definición	Aplicación
Segmentación y análisis	Manipulación directa o segmentación de la región de interés (RoI) en imágenes capturadas por drones mediante la aplicación de técnicas de procesamiento de imágenes	-Seguridad en la construcción -Detección de objetos -Cuidado de animales -Detección de vehículos -Detección de personas -Operaciones de rescate
Detección y predicción	Reconocimiento de objetos o escenas segmentadas o no segmentadas y predicción del futuro mediante técnicas de aprendizaje automático	-Reconocimiento de objetos -Identificación de personas -Estimación agrícola -Cuidado de animales
Reconstrucción 3D	Reconstrucción de imágenes 3D usando múltiples o una única escena 2D	-Procesamiento datos geoespaciales -Mapeo ambiental

CUADRO 2.1: Aplicaciones de UAV propuestas según sus propósitos.

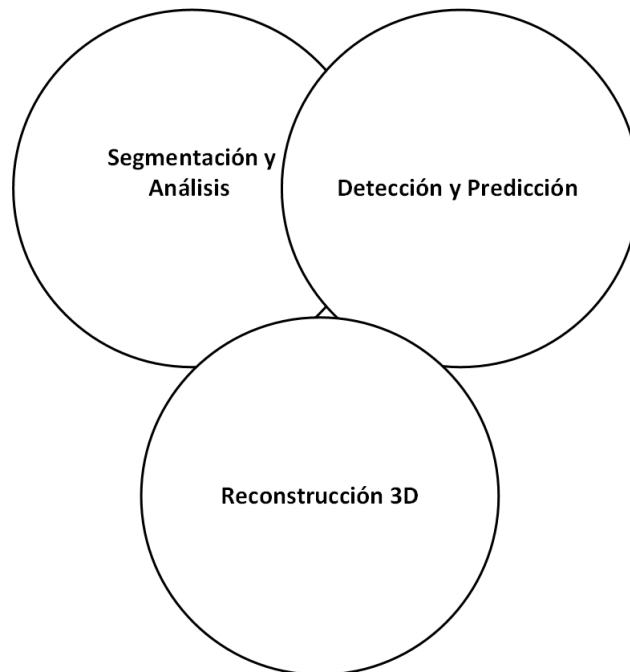


FIGURA 2.8: Relación entre las categorías de UAV según técnicas y aplicaciones

Segmentación y análisis

La fase de análisis y segmentación en el procesamiento de imágenes de UAV suele incluir la binarización de imágenes y operaciones morfológicas para separar la región de interés de la imagen completa o para ofrecer una captura explícita de dicha región.

Para ello se hace uso de procedimientos como la binarización de imágenes para segmentar imágenes, filtros Canny y la transformada de Hough. Para finalizar una segmentación, las regiones segmentadas son identificadas mediante un clasificador de análisis discriminante lineal (LDA). Dependiendo de cada uso se puede tratar la segmentación de formas distintas.

La segmentación juega un papel fundamental en el procesamiento de imágenes al separar un objeto o región de interés de su fondo o entorno. Esto facilita un análisis más detallado del objeto en cuestión y es esencial para aplicaciones como la medicina, donde la segmentación de imágenes médicas permite identificar tumores o anomalías.

- **Técnicas Tradicionales:** Métodos como el umbralizado, la segmentación basada en crecimiento de regiones y la segmentación basada en bordes son técnicas convencionales ampliamente utilizadas.
- **Enfoques Modernos:** El advenimiento del aprendizaje profundo ha llevado al desarrollo de técnicas más avanzadas, como las redes neuronales convolucionales (CNN) para la segmentación semántica y la segmentación de instancias.

Detección y predicción

La detección de objetos es esencial para identificar y localizar objetos específicos dentro de imágenes, mientras que la predicción se centra en inferir comportamientos futuros o estados de esos objetos.

- **Detección:** La detección de objetos ha avanzado enormemente con el uso de técnicas de aprendizaje profundo, como las Redes Neuronales Convolucionales (CNN) y sistemas como YOLO (You Only Look Once) o SSD (Single Shot Multibox Detector).
- **Predicción:** Con el auge de la inteligencia artificial, la predicción basada en imágenes se ha vuelto más sofisticada. Las redes neuronales recurrentes, en particular, las LSTM (Long Short-Term Memory), son esenciales para predecir secuencias temporales basadas en datos visuales.

Reconstrucción 3D

La reconstrucción 3D transforma datos visuales 2D en modelos tridimensionales, siendo fundamental en áreas como la arqueología, la medicina y el diseño industrial.

- **Fotogrametría:** Es una técnica tradicional que utiliza imágenes 2D para reconstruir estructuras 3D, aprovechando la paralaje entre múltiples fotos.
- **Técnicas Modernas:** Con el desarrollo de cámaras de profundidad y la mejora de algoritmos, técnicas como SLAM (Simultaneous Localization and Mapping) y técnicas basadas en aprendizaje profundo están ganando popularidad en la reconstrucción 3D.

El procesamiento de imágenes es un campo en constante evolución que se beneficia enormemente de los avances tecnológicos y algorítmicos, ofreciendo soluciones innovadoras a desafíos en diversos dominios.

2.2.2. Deep Learning

El Aprendizaje Profundo, o Deep Learning, es una subárea del Aprendizaje Automático. Aunque opera de manera similar a una Red Neuronal Artificial, se distingue por tener un mayor número de capas. En este tipo de redes, cada capa tiene la función de extraer características específicas de los datos ingresados. La información fluye de capa en capa, donde la salida de una se convierte en la entrada de la siguiente, creando una cadena de transformaciones y extracciones de características.

Al igual que en las Redes Neuronales Artificiales (RNAs), en una red neural profunda, la primera capa se conoce como camada de entrada y la última como camada de salida. Las camadas que se encuentran entre estas dos se denominan camadas ocultas.

2.2.3. Arquitecturas de Redes Neurales Convolucionales

Las Redes Neurales Convolucionales (CNNs, por sus siglas en inglés) son una categoría especializada de redes neuronales diseñadas para procesar datos con una estructura en forma de rejilla, como una imagen. Estas redes están diseñadas para reconocer patrones visuales directamente desde imágenes en bruto, siendo especialmente efectivas en tareas de procesamiento y clasificación de imágenes. A continuación, describiremos brevemente algunas de las arquitecturas más destacadas de CNNs.

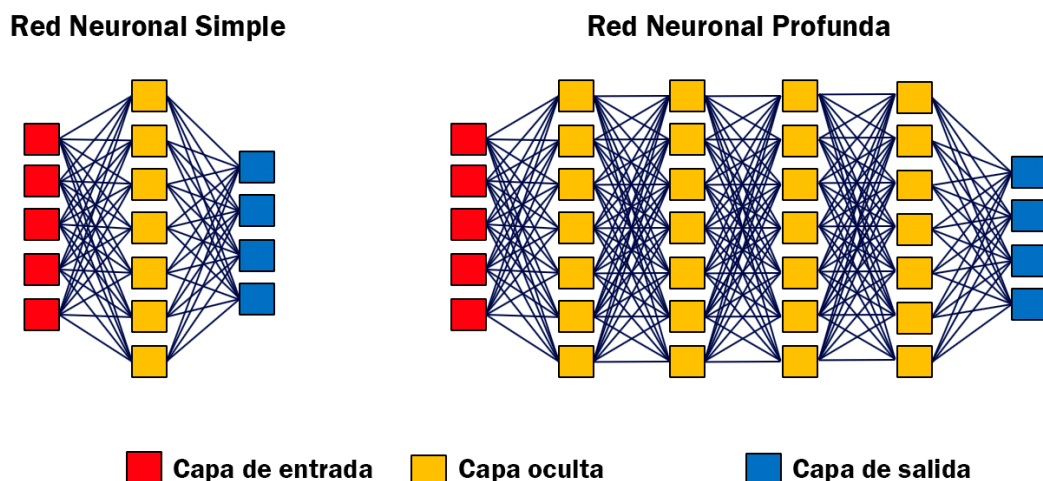


FIGURA 2.9: Representación de una red neuronal simple y una red neuronal profunda

- **LeNet-5:** Desarrollada en 1998 por Yann LeCun [39], esta fue una de las primeras redes neuronales convolucionales y se utilizó principalmente para reconocimiento de dígitos manuscritos. Consiste en dos series de capas convolucionales, seguidas de capas de submuestreo (*pooling*) y finaliza con capas completamente conectadas.

Aunque las redes neuronales ya existían antes de LeNet-5, esta red fue pionera en demostrar que las CNN podían ser efectivas en tareas visuales. A partir de aquí, las redes convolucionales se convirtieron en la arquitectura preferida para el procesamiento de imágenes.

- **AlexNet:** Presentada por Alex Krizhevsky en 2012 [40], esta arquitectura ganó el desafío ImageNet ILSVRC. Es similar en diseño a LeNet pero mucho más grande y fue una de las primeras CNNs en hacerse uso de las unidades de procesamiento gráfico (GPU) para mejorar su rendimiento.
- **VGG (Visual Geometry Group):** Desarrollada en 2014 por el grupo VGG de la Universidad de Oxford, se caracteriza por su uniformidad, utilizando sólo convoluciones 3x3 y pooling 2x2. Se hicieron populares las versiones VGG-16 y VGG-19, que se refieren al número de capas ponderadas en la red.
- **GoogLeNet/Inception:** Desarrollada por Google en 2014, introduce el módulo *Inception*, que permite a la red aprender representaciones a diferentes escalas. GoogLeNet es significativamente más profundo que VGG y utiliza menos parámetros.
- **ResNet (Red Residual):** Propuesta por Kaiming He en 2015, esta arquitectura introdujo conexiones residuales para permitir la formación de redes mucho más profundas, superando el problema de desvanecimiento del gradiente. Con estas conexiones, ResNet puede alcanzar arquitecturas de hasta 152 capas [41]. La red ResNet se alzó como ganadora en el desafío ImageNet en el mismo año de su propuesta. Su innovación radica en la capacidad para entrenar redes de una profundidad extraordinaria, con más de 150 capas. ResNet se caracteriza

por dos aspectos clave: conexiones de atajo (conexiones directas de identidad) y la normalización por lotes (batch normalization).

Estas conexiones de atajo abordan un problema común en redes profundas: el esvanecimiento del gradiente. En contextos donde el gradiente se retropropaga a capas anteriores, la multiplicación repetitiva puede reducir drásticamente su magnitud, haciendo que la red pierda capacidad de aprendizaje a medida que se añaden más capas. Las conexiones de atajo, sin embargo, permiten que ciertas señales se transmitan directamente a través de capas, evitando el proceso normal y potencialmente problemático de propagación. Específicamente, en ResNet, una señal puede ser producida a partir de la suma de la señal de dos capas convolucionales anteriores y la señal que proviene directamente del punto anterior a esas capas. Este método de combinar señales procesadas con señales no procesadas de etapas previas es esencial para el funcionamiento de ResNet.

La estructura de ResNet está formada por bloques residuales, también conocidos como bloques de conexiones de salto o *skip-connection blocks*. Cada uno de estos bloques está compuesto por múltiples capas convolucionales apiladas. Una de las ventajas de la arquitectura ResNet es su eficiente convergencia, debido en gran parte a las conexiones residuales que aceleran la convergencia de las capas más profundas [42].

- **Darknet:** arquitectura de red neuronal convolucional que destaca por su diseño específico orientado a la velocidad y eficiencia. Esta estructura fue concebida por Joseph Redmon y ha ganado notoriedad gracias a su integración con YOLO (You Only Look Once) [43, 44], un método líder en detección de objetos en tiempo real.

Una de las razones detrás de la eficacia de Darknet es su compromiso con la eficiencia y velocidad. Esta arquitectura fue desarrollada para ser ágil, lo que le permite operar eficazmente en dispositivos con recursos limitados, un aspecto esencial para aplicaciones en tiempo real. Además, Darknet se distingue por su diseño modular, lo que facilita su adaptabilidad a distintas necesidades y aplicaciones.

Desde el punto de vista técnico, Darknet fue escrito en C y CUDA, una combinación que le permite beneficiarse de la aceleración por GPU. En particular, su implementación en CUDA, una plataforma de NVIDIA, garantiza un rendimiento óptimo en máquinas equipadas con GPUs de esta marca.

Pero, sin duda, uno de los logros más notables de Darknet es su asociación con YOLO. Esta técnica de detección, que realiza la tarea en una única pasada, contrasta con otras metodologías que dividen el proceso en varias etapas. Esta singularidad otorga a YOLO una velocidad insuperable, consolidándolo como un referente en detección en tiempo real.

Por último, es esencial mencionar que Darknet es un proyecto de código abierto. Esto implica que la comunidad tiene la libertad de usarlo, adaptarlo y distribuirlo según lo vea conveniente, lo que ha fomentado su adopción y evolución continua.

En particular, el procesamiento de imágenes ha logrado integrar múltiples tecnologías de IA no sólo para clasificar imágenes de manera automática, sino también para describir detalladamente los elementos que estas contienen. Los modelos de

aprendizaje profundo, dadas sus estructuras de múltiples niveles, demuestran ser especialmente eficientes al extraer información de gran complejidad de las imágenes procesadas. Adicionalmente, las redes neuronales convolucionales han optimizado notablemente los tiempos computacionales, aprovechando la potencia de las GPU, una característica no siempre explotada por otras redes.

2.2.4. Arquitecturas de detección de objetos

En cuanto a la identificación de objetos en imágenes, dos métodos destacan: los algoritmos de *Region Proposal* y los algoritmos de detección de objetos por regresión [45], el enfoque de *Region Proposal* busca pre-identificar posibles ubicaciones del objeto a detectar en la imagen, garantizando así una alta tasa de recuperación con menos ventanas de selección. Cuando se procesa una imagen, tras una serie de operaciones convolucionales, se obtiene un mapa de características. Cada sector de dicho mapa representa una porción específica de la imagen original. Dentro de estos algoritmos de propuesta regional, se encuentran destacados exponentes como R-CNN, Fast R-CNN [46], Faster R-CNN [47], y MASK-CNN [48].

R-CNN

Entre estos, Mask R-CNN merece una atención especial. Se trata de una evolución del Faster R-CNN, enriquecido con una rama adicional que predice una *object mask* en simultáneo con la determinación del cuadro delimitador. Lo que diferencia a Mask R-CNN es su precisión en la alineación a nivel de píxeles. Este algoritmo opera en dos fases: mientras que la primera fase se asemeja a RPN, la segunda no sólo predice la clase y desplazamiento de la caja, sino que también produce una máscara binaria para cada RoI. Una de las grandes ventajas de Mask R-CNN es su facilidad de implementación y formación, ofreciendo además una operatividad ágil y versátil. Un ejemplo gráfico de la segmentación lograda por este algoritmo puede apreciarse en la Figura 2.10.

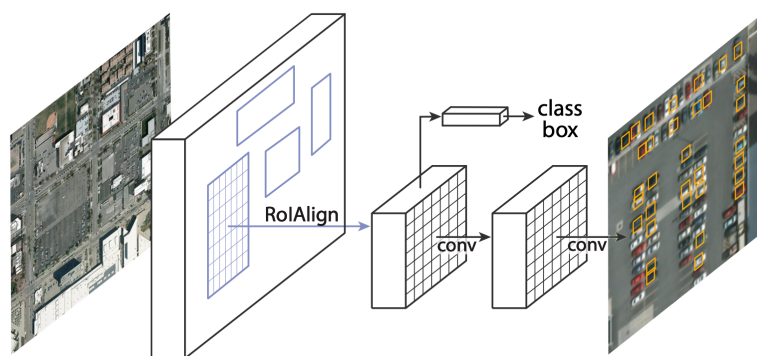


FIGURA 2.10: Mask R-CNN

Detectron

En la misma línea que las anteriores, encontramos Detectron, que es el sistema de software de Facebook AI Research (FAIR) que implementa algoritmos de detección de objetos de última generación, incluyendo Mask R-CNN. El objetivo de Detectron

es proporcionar una base de código de alta calidad y alto rendimiento para la investigación de la detección de objetos. Ha sido diseñado para ser flexible, con el fin de apoyar la rápida implementación y evaluación de nuevas investigaciones. Sin embargo, la versión con más éxito se trata de Detectron2 [49], siendo una mejora de Detectron. La gran diferencia entre versiones se basa en que la más actual presenta un diseño más modular, más flexible y extensible, permitiendo un entrenamiento mucho más rápido en ordenadores que presentan dispositivo GPU.

El Detectron2 incluye implementaciones de alta calidad de los algoritmos más avanzados de detección de objetos, como DensePose, redes de pirámide de características panópticas y numerosas variantes de la pionera familia de modelos Mask R-CNN, también desarrollada por FAIR. Los creadores del algoritmo reproducen las líneas de base de ResNet-50-FPN junto al algoritmo Scale Jitter.

Single Shot Detector (SSD)

Por otro lado, tenemos los métodos de algoritmos de detección de objetos por regresión. Los algoritmos anteriores utilizan la detección como un problema de clasificación, es decir, primero se generan propuestas de objetos y luego estas propuestas se envían a las regiones de clasificación/regresión. Sin embargo, algunos métodos abordan la detección como un problema de regresión basándose en una operación similar. Dentro de este campo destacan los algoritmos de YOLO (You Only Look Once) y SSD (Single Shot Detector).

El algoritmo SSD [50] consigue un buen equilibrio entre velocidad y precisión. SSD ejecuta una red convolucional en la imagen de entrada sólo una vez y calcula un mapa de características. A continuación, se ejecuta un pequeño núcleo convolucional de 3×3 en este mapa de características para predecir los cuadros delimitadores y la probabilidad de clasificación. SSD también utiliza anchor boxes en varias relaciones de aspecto, de forma similar a Faster-RCNN, y aprende el desplazamiento en lugar de aprender la caja. Para manejar la escala, SSD predice los cuadros delimitadores después de múltiples capas convolucionales. Como cada capa convolucional opera a una escala diferente, es capaz de detectar objetos de varias escalas. Un ejemplo del funcionamiento del algoritmo SSD se puede observar en la Figura 2.11.

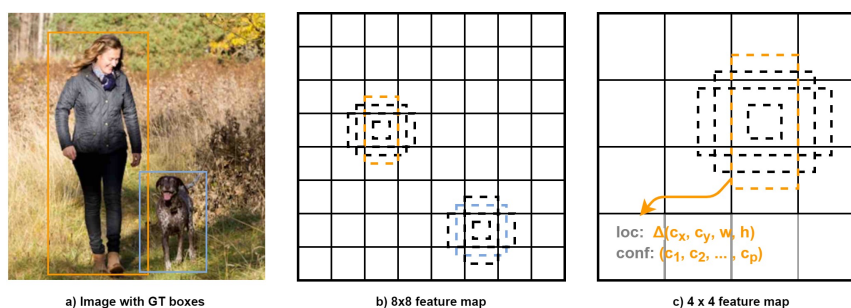


FIGURA 2.11: Single Shot Detector [15]

YOLO (You Only Look Once)

Se trata de una arquitectura de detección de objetos muy rápida y precisa, que fue creada por Joseph Redmond como principal autor [51]. En el algoritmo, la detección es un simple problema de regresión que toma una imagen de entrada y conoce las probabilidades de clase junto a las coordenadas de los bounding boxes. El YOLO divide cada imagen en una cuadrícula de $S \times S$ y cada cuadrícula predice N bounding boxes junto a su confianza. La confianza refleja la precisión del cuadro delimitador y si éste contiene realmente un objeto, independientemente de la clase. YOLO también predice la puntuación de clasificación de cada bounding box para cada clase en el entrenamiento. El algoritmo puede combinar ambas clases para calcular la probabilidad de que cada clase esté presente en un cuadro predicho. Así, se predice un total de $S \times S \times N$ bounding boxes. Sin embargo, la mayoría de estas casillas tienen puntuaciones de bajas confianza, por lo que si establecemos un umbral, por ejemplo del 30 % de confianza, desde eso, se puede eliminar la mayoría de ellas con el ajuste del umbral, como se muestra en la Figura 2.12.

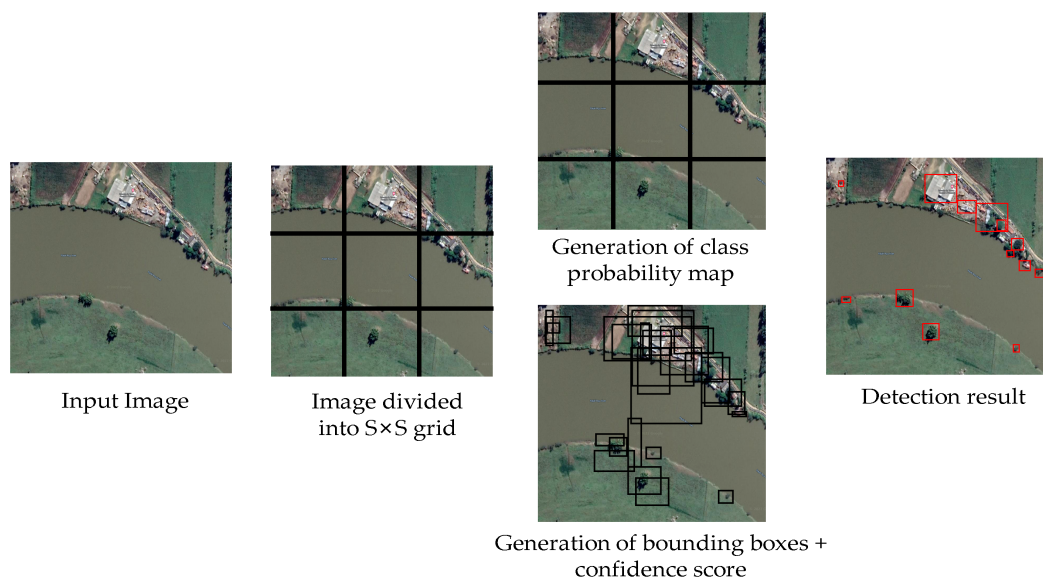


FIGURA 2.12: Flujo del algoritmo YOLO

YOLO es un algoritmo más rápido que el resto de los algoritmos de detección, permitiendo ser ejecutado en tiempo real. Otra diferencia clave es que YOLO ve la imagen completa de una vez, en lugar de mirar sólo las propuestas de una región generada en los métodos anteriores. Así, esta información contextual ayuda a evitar los falsos positivos. Sin embargo, una de las limitaciones de YOLO es que sólo predice un tipo de clase en una cuadrícula, por lo que tiene dificultades con objetos muy pequeños. Hay varias versiones de YOLO como YOLOv2 [51], YOLOv3 [52] y YOLOv4 [53].

La evidencia en la literatura presenta como uno de los algoritmos más utilizados en el campo de detección de objetos. Es un algoritmo popular de detección de objetos, y se han lanzado varias versiones. Al comparar la evolución de toda la serie YOLO, podemos ver una evolución significativa con respecto al tiempo de detección. En la primera versión publicada, se requiere solo una red neuronal convolucional para hacer una predicción, YOLO está hecho para funcionar en dispositivos con

bajo poder de procesamiento. Dado que la versión inicial se basó en AlexNET, este método ha pasado por varias iteraciones adicionales.

En la cronología del algoritmo YOLO, aparecen YOLOv3 y la versión YOLOv4. En resumen, YOLOv3 y YOLOv4 son algoritmos de detección de objetos basados en aprendizaje profundo, pero YOLOv4 es una mejora sobre YOLOv3. YOLOv4 ha sido optimizado para la detección de objetos en tiempo real y entrenado en un gran conjunto de imágenes y videos para mejorar su precisión. YOLOv4 también incluye nuevas técnicas, como la ampliación de datos Mosaic y DropBlock, mejorando su rendimiento.

Hasta 2021, se considera que YOLOv4 es la versión más precisa de YOLO. Está construido sobre una arquitectura de red neuronal diseñada a medida que utiliza una combinación de capas convolucionales y convolucionales transpuestas para detectar objetos en imágenes y videos. Posteriormente, se lanzó la quinta versión del algoritmo, llamada YOLOv5. Este algoritmo resultó ser un modelo perfecto, trayendo más opciones, destacándose la segmentación de imágenes, pero aún necesita acercarse a la quinta actualización principal. Los resultados son muy similares a YOLOv4, donde se realizó una cantidad considerable de trabajo y se tuvieron en cuenta todos los matices. YOLOv5 es una mejora sobre YOLOv4. Se basa en una nueva arquitectura SPADE, que utiliza información semántica y espacial para mejorar la precisión de detección de objetos. YOLOv5 también utiliza un nuevo algoritmo de entrenamiento llamado Mosaic Data Augmentation para mejorar la generalización del modelo.

Más recientemente, se lanzó la séptima versión del algoritmo, YOLOv7, la última iteración en el ciclo de vida de los modelos YOLO. YOLOv7 infiere más rápido y con mayor precisión que sus versiones anteriores (es decir, YOLOv5). YOLOv7 es la última versión de YOLO. Ha sido construido sobre una nueva arquitectura llamada Efficient-YOLO, que utiliza EfficientNet como red principal. YOLOv7 ha sido entrenado en un gran conjunto de datos y ha sido optimizado para la detección de objetos en tiempo real. Es más preciso y rápido que las versiones anteriores de YOLO. En la Figura 2.13 es posible comparar la actual version con los predecesores.

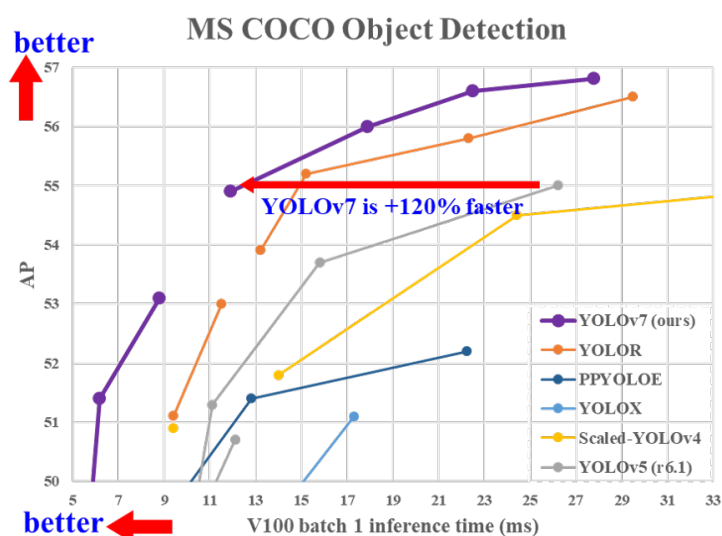


FIGURA 2.13: Comparativa entre versiones del algoritmo YOLO [54]

En conclusión, YOLOv4 se considera la versión más precisa de YOLO hasta 2021 y está optimizado para la detección de objetos en tiempo real. YOLOv7 es la última versión de YOLO y se basa en una nueva arquitectura llamada Efficient-YOLO, que es más precisa y rápida que las versiones anteriores.

Capítulo 3

Lista de artículos de revistas

En este apartado se describe la coherencia de cada uno de los artículos que conforman esta tesis doctoral presentada en formato de “*Compendio de artículos*”. A continuación, se detallan los tres artículos ya publicados en revistas internacionales, todas ellas indexadas en Journal Citation Reports (JCR). Se debe mencionar que los artículos han sido sometidos a un proceso de revisión ciega por pares de autores. Este capítulo se encuentra estructurado de la siguiente forma: en primer lugar se explica la coherencia y la relación que tienen los artículos con la hipótesis de este trabajo. A continuación, se presenta cada uno de los artículos en su formato original, incluyendo para cada uno de ellos un breve resumen en castellano conformado por los objetivos de la investigación, la metodología utilizada, los resultados obtenidos y conclusiones en cada uno de ellos.

3.1. Coherencia y relación entre los artículos

En esta sección se describe la coherencia y relación que existe entre los artículos y en relación con la hipótesis de investigación planteada en la sección 1.1. Como ya se ha mencionado anteriormente se trata de un total de tres artículos de investigación científica publicados en revistas internacionales con factor de impacto, es decir, indexadas en el JCR. Cada uno de estos artículos se alinea directamente con la temática de la hipótesis propuesta. A continuación, se enumeran las publicaciones acompañadas por un resumen muy breve:

- ***An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images***: Dada la disminución en el mantenimiento de rutas de transporte público en diversos países por situaciones económicas desfavorables, este artículo propone una plataforma distribuida que utiliza drones para detectar desperfectos en las rutas de transporte. En particular, en este artículo se hace uso de un sistema multi-agente basado en PANGEA haciendo uso de técnicas de computación ubicua. Como uno de los resultados principales, se presenta un caso de estudio que utiliza el clasificador You Only Look Once (YOLO) v4, logrando una precisión del 95 %. Las imágenes utilizadas que conforman el dataset han sido compartidas de forma abierta para un uso gratuito y abierto por parte de la comunidad científica.
- ***Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning***: Debido al crecimiento urbanístico de la ciudad, la investigación aborda la necesidad de mapear zonas urbanas y rurales. Utiliza clasificadores basados en objetos junto con GeoDMA y algoritmos como OneR, NaiveBayes, J48, IBk, y Hoeffding

Tree, además de una primera aproximación con R-CNN y el algoritmo YOLO. El propósito es caracterizar zonas de vegetación, suelo expuesto, asfalto y edificaciones. El estudio presenta una metodología para identificar activamente objetos geoespaciales y seleccionar el mejor clasificador, logrando una precisión del 85 % y un índice kappa del 0.76. Finalmente, ofrece una estrategia práctica para extraer y recomendar información con relevancia científica.

- ***Automated Road Damage Detection Using UAV Images and Deep Learning Techniques:*** Este artículo aborda la problemática de la recolección manual de datos sobre daños en carreteras, identificando su ineficiencia y riesgos para humanos. Propone una solución mediante el uso de UAVs y tecnologías de Inteligencia Artificial para poder detectar daños en carreteras con mayor eficiencia y precisión. Se explora el potencial de los algoritmos YOLOv4, YOLOv5, y YOLOv7, utilizando imágenes tomadas por UAVs. Los resultados experimentales, basados en conjuntos de datos de China y España, demuestran la efectividad del enfoque, con un alto grado de precisión en la detección automática de daños en carreteras.

En la era actual de crecimiento acelerado de las ciudades y la evolución constante de la tecnología, es imperativo abordar los desafíos que surgen en la gestión y el mantenimiento de las infraestructuras urbanas. Estos desafíos no sólo están relacionados con la expansión física de las ciudades, sino también con la garantía de seguridad, calidad y sostenibilidad en los espacios que habitamos. La infraestructura vial, en particular, se encuentra en el centro de esta discusión, dada su importancia crítica en el transporte, la economía y la seguridad pública. Es en este contexto, donde tres investigaciones distintas, pero intrínsecamente relacionadas, emergen para ofrecer soluciones innovadoras mediante la combinación de vehículos aéreos no tripulados y técnicas de aprendizaje profundo.

Los tres artículos identifican problemas relacionados con la monitorización urbana. Mientras el primer y tercer artículo se centran en la detección de daños en rutas de transporte y carreteras, el segundo aborda el mapeo de zonas debido al desarrollo urbano. Es evidente que existe una necesidad creciente de soluciones tecnológicas en la monitorización y el mantenimiento de infraestructuras urbanas y rurales.

La utilización de UAVs y la detección de características en imágenes es una constante que se encuentra presente en los tres artículos. Los drones proporcionan de forma eficiente y segura de recopilar datos desde una perspectiva aérea, ofreciendo una vista amplia y detallada de las áreas bajo estudio. Además, todos los artículos aprovechan el poder del machine learning, en particular, los algoritmos y sus variantes, demostrando su eficacia en la detección y clasificación de daños y objetos urbanos.

En la primera investigación se propone un sistema multi-agente para coordinar diferentes partes de la arquitectura, demostrando la potencia de los sistemas distribuidos en la gestión de grandes volúmenes de datos. En la segunda investigación, por otro lado, explora diversos algoritmos y técnicas para identificar y clasificar zonas urbanas, poniendo énfasis en la importancia de la precisión y la relevancia científica de la información extraída. El tercer artículo destaca la seguridad y eficiencia de usar UAVs en comparación con los métodos tradicionales de recolección de datos, a la vez que profundiza en la eficacia de diferentes versiones del algoritmo una vez probado en un conjunto de datos más grande y con características de países distintos.

Otro punto en común que presentan los artículos que componen este compendio de artículos es su contribución a la comunidad científica. El conjunto de datos de imágenes propias utilizado en el primer artículo ha sido puesto a disposición de la comunidad científica, contribuyendo con la ciencia abierta. Siguiendo con el enfoque colaborativo y global en el tercer artículo se ha empleado un conjunto de datos tanto de China como de España. En la segunda contribución, se pretende contrastar y analizar imágenes de satélite e imágenes de UAVs, con el fin de evaluar la concordancia y validez de los datos. Estas contribuciones facilitan la replicabilidad y extensión de los estudios, promoviendo avances futuros en el campo.

La confluencia de UAVs y machine learning está redefiniendo el campo de la monitorización urbana. Los tres artículos examinados en este informe destacan diferentes facetas de esta revolución, pero comparten una visión común sobre la importancia de la innovación tecnológica en la solución de problemas urbanos críticos. Al observar estos estudios conjuntamente, no sólo se identifican áreas específicas de aplicación y avance, sino que también vislumbramos el potencial futuro de estas tecnologías en la construcción de ciudades más seguras, eficientes y bien mantenidas.

Para finalizar, aunque cada artículo aborda un problema específico y particular, todos convergen en la utilización de UAV que hacen uso de técnicas avanzadas de machine learning para mejorar y optimizar la gestión y el mantenimiento de infraestructuras urbanas. La recurrencia del uso del clasificador YOLO en los tres estudios también indica la relevancia y eficiencia de este algoritmo en tales aplicaciones. Estos trabajos no únicamente subrayan la relevancia de la tecnología en la solución de problemas urbanos actuales, sino que también sugieren un camino claro para futuras investigaciones en este ámbito dinámico y esencial.

3.2. An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images



Article

An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images

Luís Augusto Silva ^{1,2,*}, Héctor Sanchez San Blas ¹, David Peral García ¹,
André Sales Mendes ¹ and Gabriel Villarubia González ^{1,*}

¹ Expert Systems and Applications Lab—ESALAB, Faculty of Science, University of Salamanca, Plaza de los Caídos s/n, 37008 Salamanca, Spain; hectorsanchezsanblas@usal.es (H.S.S.B.); daverall@usal.es (D.P.G.); andremendes@usal.es (A.S.M.)

² Laboratory of Embedded and Distribution Systems, University of Vale do Itajaí, Rua Uruguai 458, C.P. 360, Itajaí 88302-901, Brazil

* Correspondence: luisaugustos@usal.es (L.A.S.); gvg@usal.es (G.V.G.)

Received: 9 September 2020; Accepted: 27 October 2020; Published: 30 October 2020



Abstract: In recent years, maintenance work on public transport routes has drastically decreased in many countries due to difficult economic situations. The various studies that have been conducted by groups of drivers and groups related to road safety concluded that accidents are increasing due to the poor conditions of road surfaces, even affecting the condition of vehicles through costly breakdowns. Currently, the processes of detecting any type of damage to a road are carried out manually or are based on the use of a road vehicle, which incurs a high labor cost. To solve this problem, many research centers are investigating image processing techniques to identify poor-condition road areas using deep learning algorithms. The main objective of this work is to design of a distributed platform that allows the detection of damage to transport routes using drones and to provide the results of the most important classifiers. A case study is presented using a multi-agent system based on PANGEA that coordinates the different parts of the architecture using techniques based on ubiquitous computing. The results obtained by means of the customization of the You Only Look Once (YOLO) v4 classifier are promising, reaching an accuracy of more than 95%. The images used have been published in a dataset for use by the scientific community.

Keywords: smart applications; drones; YOLOv4; crack detection; virtual organizations of agents

1. Introduction

Road safety is key to reducing the number of road accidents. The main solutions often focus on reducing speed limits, but road conditions are also important in the increase in road accidents. A recent study carried out by the AEC stated that the poor condition of roads is the main cause of 94% of accidents [1]. According to this organization, 1 out of every 13 km of the Spanish road network shows significant deterioration in more than 50% of the surface of the pavement as a result of accumulating potholes, ruts, and cracks. This damage affects vehicle safety, creating a less stable and safe driving environment. The report [1] also shows that Spain is failing to maintain its roads, and describes the condition of the asphalt as very poor. This translates into an increased risk of an accident in the worst-case scenario, but can also lead to vehicle breakdowns. To minimize the risks of a possible accident or decrease the possibility of a breakdown, it is vitally important to investigate and develop techniques that can be used to improve the road infrastructure. The difficult worldwide economic situation has repercussions for the state of roads, making maintenance and improvement of transport routes lower priorities.

filtering, and coordinating the information handled by the platform. The multi-agent system, which is used in the case study, is based on the use of the open-source platform PANGAEA [16], allowing the different elements of the system to behave dynamically according to the requirements of the platform at any given moment. As a fundamental pillar in the identification of road damage, we customized the YOLOv4 algorithm. Notably, the machine learning models depend on the quality of the data used for their training; therefore, to obtain a robust model for this task, a new dataset was created for use by deep learning algorithms. Using a UAV is proposed to obtain the images automatically using a tracking road [17–19], thus minimizing the effort required for the sample collection process in terms of both cost and time. Through the proposed multi-agent architecture, coordination and communication between the different elements of the system are achieved so that the different classifiers that were tested are transparent to the end users.

The article is organized as follows: Section 2 provides an in-depth review of the current literature on crack and drone detection algorithms. Section 3 describes the architecture of the proposed system. The experiments and results are outlined and explained in Section 4. Finally, the conclusions are presented in Section 5. In addition, images of the pothole detection are provided in Appendix A.

2. Background

In this section, the different works in the different areas involved in this study are analyzed. This section is divided into different sub-sections to provide a more detailed analysis of each of the research areas. First, we begin by explaining the state-of-the-art technology used to capture the images from our dataset (drones). Then, we analyze the different methods and algorithms of image detection. We then focus on one of these image detection systems, YOLO, which is used in our framework, as well as some of the most used datasets. Finally, we refer to the framework in which we include the proposed system, the multi-agent system, highlighting its major advances and utilities.

2.1. Drones

The first step in determining the status of a pathway is to take images of it. To obtain images with good quality and perspective, we chose to use a drone. UAVs are becoming an important tool for surveyors, engineers, and scientists, as the number of cost-effective and easy-to-use systems is increasing rapidly. These platforms offer an attractive alternative for mapping small areas with centimeter resolution [20]. Drones are widely used today for tasks such as surveillance, monitoring, and data collection in buildings, infrastructure, and hazardous environments [12], detection of excessive vegetation growth [13], Internet of Things (IoT) systems [21], and in combination with virtual reality devices for control [22]. The main reason for why we chose a drone to recognize the terrain and obtain the data of the dataset was to acquire a large number of images in a short period of time with an acceptable resolution. Using these images, it is possible to associate the cracks in the recorded images with their real positions in the real world and to calculate their real sizes.

2.2. Pattern Detection in Images

Once we determined how to obtain the road image data, we focused on the area of object detection—in our case, cracks and potholes—in the road. In recent years, demand has been growing for analysis and recovery of large amounts of information from images on a network. The main purpose of image processing is to separate relevant information (foreground) from non-relevant information (background) [23].

Cracks are a common defect in structures, and they can cause serious structural failures. Surface crack detection using image processing has emerged as an important topic in the last 20 years [24]. Most studies have targeted crack detection for concrete surfaces [25,26], glass fiber composite laminates [27], solar cells [28], and different types of surfaces [29]. The use of the image detection techniques, which involve the detection of cracks and potholes in the captured image, allowed

us to extract relevant information without pre-processing the image in a reasonable time with acceptable results.

2.3. YOLO

One of these systems used for image detection is YOLO. YOLO is an open-source application that allows the classification of images in real time.

We selected YOLO because it has been tested and used for the detection of cracks in different surfaces, such as defects in turbine propellers [30], cracks in cement surfaces [31,32], and in composite panels [33]. YOLO has previously been used for the detection of cracks in roads [34–38].

To select the appropriate version of YOLO, we compare the main versions of YOLO in Table 1, focusing on different aspects like the number of convolutional layers, activation function, calculation function, and the specific characteristics of each version.

Table 1. You Only Look Once (YOLO) versions.

Version	Article	Features
YOLOv1	[4,30,31,34,35,39,40]	26 layers, SoftMax activation function, sum-squared risk calculation function, may not detect objects too close together, non-maximal suppression to eliminate duplicates.
YOLOv2	[35,41].	30 layers, function of calculation of the risk of mean square error, requires fewer training times, batch normalization in the layers (an increase in the mean average precision (mAP)).
YOLOv3	[37,42–45].	106 layers, multi-label sorting, SoftMax activation function with independent logistic sorters, binary cross-entropy loss calculation function for each label, small object enhancement.
YOLOv4	[10].	53 layers, FPN for greater accuracy, SAM allows focus on a specific part of the image, SPP detects object deformations.

Table 2 compares the mean average precision (mAP) and speed (FPS) for the PASCAL VOC2007 dataset made in [46]. The YOLOv2 version has the highest FPS, and YOLOv3 + SPP obtains the highest mAP.

Table 2. mAP and speed comparison.

Method	mAP (%)	Speed (fps)
Faster RCNN	76.4	5.0
SSD	74.3	46.0
DSSD 321	78.6	9.5
STDN	78.1	40.3
YOLO	63.4	45.0
YOLOv2	76.8	67.0
YOLOv3	79.3	37.9
DC-SPP-YOLO	78.4	55.7
YOLOv3(DC)+SPP	79.7	37.0

To improve YOLO’s efficiency, it is used in combination with different algorithms and techniques, as well as different variants of the original YOLO model.

The variants include the YSODA model for the detection of small objects [30], the DC-YOLO model based on YOLOv3 [42], the Tiny-YOLO-V2 model, which has a lower computational cost,

and pre-trained detection models, such as DarkNet [42]. The DarkNet model has pre-trained networks that are capable of recognizing and detecting cracks in pavement, which is why we chose to use it in our system.

For crack detection, there are two-phase and single-phase methods. Both have their disadvantages: The first is too slow (low FPS) and the second is less accurate than the two-phase method. Xia et al. [36] used the single-phase method, focusing on improving the accuracy of the method by incorporating a new detection pipe (cumulative feature module and a smoothed loss function). This update of the original algorithm was selected and added to our system because we needed high FPS and significantly higher accuracy.

2.4. Multi-Agent

We placed the image detection algorithm in a multi-agent system platform framework to control the drone movements and coordinate all the functions of the system. We selected the multi-agent system because it is an ideal platform for improving communication between agents due to their maneuverability and their ability to be deployed in a multitude of environments [47] due to the use of the RL and Q-learning algorithms. They are useful for tasks in different areas, both in commercial and industrial environments, such as search and rescue of people, sensorization, monitoring of the environment, surveillance, and recognition, as well as in other scientific fields [47,48].

For our purpose, we selected the MARLsystem, which performs general-purpose tasks in a coordinated and efficient manner [48]. After a detailed review of the literature, no reference or knowledge was found of a crack detection application that makes use of distributed technological capabilities adapted to the recognition of Spanish roads.

3. Proposed System

3.1. Proposed Architecture

The proposed architecture aims to provide a solution to the problem and can be adapted to the introduction of new functionality in different environments. An architecture to deal with the proposed problem must contain a series of well-defined characteristics to correctly operate. To this end, this research project was designed to focus on a multi-agent architecture that allows for the automation and detection of irregularities on roads by employing the detection of images captured by a drone. The main characteristics that led to the decision to use this type of architecture are its distributed architecture, the high level of communications, and the existence of message queues for information processing. A multi-agent architecture also allows a functionality used by one system to be replaced by another without having to modify the whole system, such as the image processing system or the type of database. A multi-agent system, in addition to the characteristics described above, is mainly characterized by each agent having a well-defined task and coordinating among themselves to achieve a common goal.

For the construction of multi-agent systems, several systems have already been developed that allow the process to be faster, including libraries such as SPADE, Python's library [49], or more complex systems, such as JADE [50], PANGEA [15,51], and osBrain [52]. In this architecture, we chose to use the PANGEA multi-agent system as a starting point. This system allows the elements of the system to dynamically enter and leave the platform, thus allowing the specific demands of the system to be met. The MAS architectures have to perform services on demand, which means that each of the agents has to report on the services they have available and can offer to other entities within the architecture. PANGEA is based on organizational technology, which allows for visual representation and can be applied to any type of system, allowing for human interaction within the system. Finally, PANGEA can be distinguished from other systems by its rules engine that allows for the distribution of the computational load. Figure 1 shows the proposed architecture using MAS PANGEA, the different virtual organizations, and the main agents that form the architecture designed for this system.

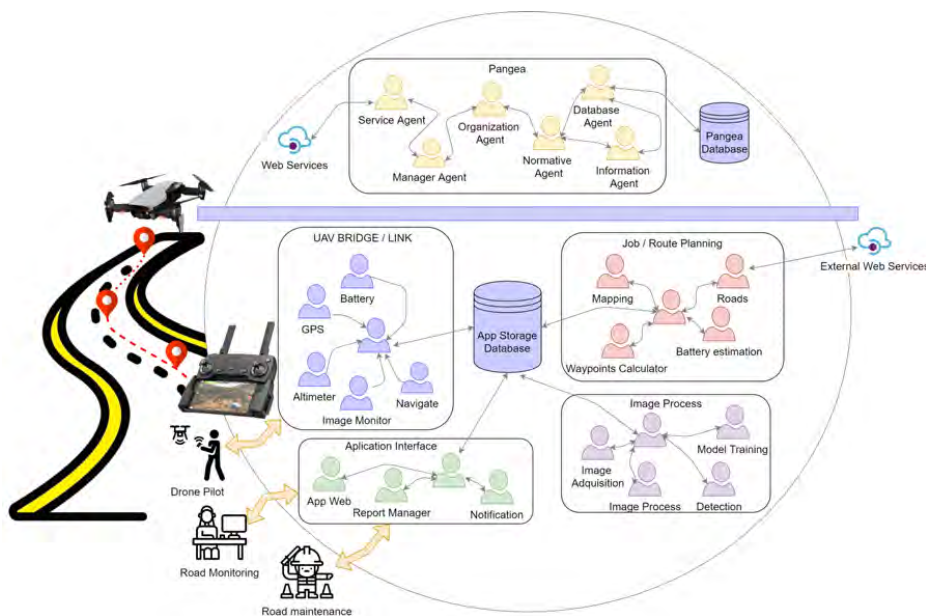


Figure 1. Proposed architecture using multi-agent system (MAS) PANGEA.

The designed architecture is divided into different parts. There are two well-defined parts: The upper part of the image displays the minimum agents necessary for the functioning of the MAS PANGEA system; the lower part shows the virtual organizations upon which the case study was based.

The organizations in the system are detailed below.

Job/Route Planning Organization: This organization refers to the generation of the routes to be followed by the drones to capture the video of the roads. Its main functionality is based on transforming data between two coordinates into a safe route that the drone can follow to capture information. Communications between the agents of this organization and the central agent are bi-directional. This agent is responsible for requesting services from each of the agents, and they respond with data for each of the queries. This central agent is responsible for communicating with the other organizations for the exchange of information. The agents found in this organization are listed in Table 3.

Table 3. Agents of the job/route planning organization.

Agent	Description
Mapping	Allows iteration with maps to identify the existence of high-voltage cables in the flying area, distances to airports, etc.
Waypoint Calculator	Calculates optimal waypoints from route data.
Roads	Obtains a route from a road indicating two coordinates. This agent connects to external service APIs. One can use the services of Google or Open-Street Maps to obtain this information.
Battery Estimation	Estimates the battery life for the realization of a route.

Image Process Organization: This organization is responsible for processing images by applying deep-learning techniques. To achieve the objective of this organization, it has several agents that can

be replaced by agents of the same characteristics that apply different techniques for labeling, training models, or detection. This architecture, which allows for exchanges of agents, makes it easy to conduct tests or apply new techniques without making changes to the whole system. The main agents that can be found in this organization are as follows:

- Image acquisition: Obtains the images from the source to perform pre-processing, training, or detection tasks.
- Image pre-process: Performs pre-processing of the images, such as color adjustment, scaling, and adaptation of each of the images to the inputs of the different models.
- Model training: Takes the tagged images as input and performs training with different deep learning algorithms.
- Detection: Uses the deep learning models trained by the training model agent to detect cracks or anomalies in the new images inserted into the system.

UAV Bridge/Link Organization: This organization refers to the application that is used as an interface for communication and information exchange between the drone and the platform. It contains several agents that obtain data from the drone sensors, such as the battery agent, the altimeter agent, or the image monitoring agent. In this organization, we also have a navigation agent that is in charge of the navigation of the drone. This agent uses the information from the other agents in the organization and the coordinates of the waypoints from the job/route planning organization to make the drone reach the indicated coordinates. The global positioning system (GPS) agent, apart from being useful for navigation, is crucial for the whole system, since it allows tagging of the locations of the photos and, thereby, locating where crack repairs are necessary.

Application Interface Organization: This organization can adapt the generated information to the application layer. This organization is used as an interface; the applications that interact with this interface can interact directly with the system. That is, the organization is responsible for converting the system's raw data to human-readable data. In this particular case, it is used for the user in charge of monitoring and deciding which roads are to be inspected, the applications of the workers responsible for carrying out repairs on these roads, the generation of reports, and sending notifications.

PANGEA Multi-Agent System Organization: This organization is composed of the minimum agents necessary for PANGEA's operation. The objective of this organization is to manage virtual organizations and the agents in each of them. The agents of this organization include the following:

- Database Agent: The only agent with database access privileges, it stores the information present within the organization. This agent is also in charge of performing backups and ensuring consistency of information.
- Information Agent: Manages the services within the virtual organization by making the services of each of the agents available to the other agents. When an agent joins the system, it indicates to this agent the services it provides so that when another agent needs to use that service, it makes a query to this agent.
- Normative Agent: Responsible for imposing and ensuring that the rules comply with the communications they establish between agents.
- Service Agent: Aims to distribute functionality through web services. It is used as a gateway that allows the communication of external services with the organization's agents. These web services allow the easy construction of external agents in different programming languages.
- Manager Agent: Responsible for checking the status of the system periodically. It is in charge of detecting if there is any overloaded functionality and any possible failures in the agents of the different organizations.
- Organization Agent: Responsible for the verification of all the operations of the virtual organizations, checking the security and the load balancing, and offering encryption of the most important agents.

The system uses two different databases for its operation: The PANGAEA Database is used by the PANGAEA organization to store the available agents and the services available in each of them. The APP Storage Database contains information about the specific system. In this case, this includes the images obtained by the drone, their locations, the next inspections to be performed, routes, etc.

In the proposed architecture, three agents external to the system were designed. These represent the users who interact with the system. Firstly, Road Monitoring is the main agent in charge of defining the inspections of the roads to be carried out, labeling the images, and supervising of the results. The second agent of the system is the Drone Pilot, which is in charge of conducting road inspections. Its main task is to supervise the flying of the drone in case there is any incident and the drone needs to be controlled manually. Finally, the Road Maintenance agent represents the person or company responsible for completing road maintenance and resolving incidents.

The architecture is based on different modules and each one is specialized in a different task or objective. The challenge of this type of architecture is focused on simplifying communication between nodes or agents and allowing them to be decoupled from the system so that they can be replaced by others with similar characteristics or services. To this end, when the system is initiated, we have a set of agents from the PANGAEA organization that are responsible for displaying the available services and coordinating agents.

The protocol used in this architecture to achieve the objectives is the contract-net protocol, where an external agent can search, find, and execute a required service. To do this, the agent that requires a service sends a message to the Agent Manager indicating the type of service required with the name of the service and the parameters for that service. This agent, with the help of the Organization Agent, Information Agent, and Database Agent, responds with a list of available agents that have the services and resources to carry out the required service. The agent accepts the proposal of the Agent Manager by choosing an agent from the list to carry out the service. An example of requiring a detection service is provided in Figure 2.

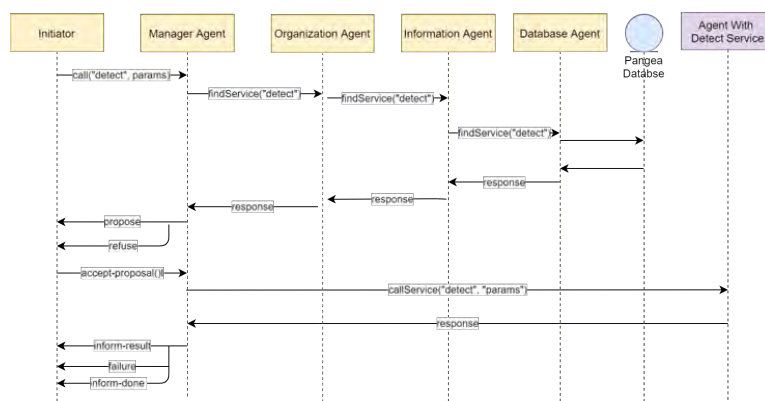


Figure 2. Image call service.

The architecture in the case study based on PANGAEA allows the dynamic integration of new functionality without affecting the other parts of the system. The second reason why PANGAEA was chosen is its user license, as it is open source can be used for commercial purposes.

3.2. Proposed Method

We evaluated the different forms of malfunctions and cracks in roads using deep learning for detection. The aerial system consisted of a DJI Mavic Air 2 quadcopter (DJI, Shenzhen, China), with a maximum flight time of 34 min, combining a 4K digital camera and location information, which were used for aerial imaging. The camera mounted on the UAV had a 24 mm lens with a field of view (FOV)

of 84 degrees and a resolution of 4000×3000 , and it was capable of shooting 48-megapixel photos; the camera was three-axis stabilized by its gimbal drone <https://www.dji.com>.

The image data used for model training and testing of deep neural networks were collected using the conventional UAV camera, since its resolution capacity was 3840×2160 pixels at a distance of 60 m from the ground. According to some authors [53], during this training process, images can be used from third parties with different environments but following the same process. The general idea of this proposed solution is illustrated in Figure 3.

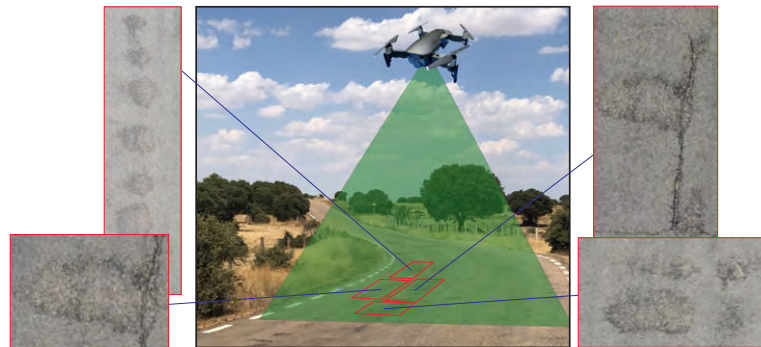


Figure 3. Image data acquisition.

During the capture process, videos were created. From them, frames were made for use in training. A total of 568 high-resolution images were generated, which were then classified and labeled as potholes and cracks. Then, the data from the labeled images were used to train and evaluate the convolutional neural net; the dataset used in this work is detailed in this section.

We first prepared of research data, followed by annotation and labeling. The prepared dataset was divided into training and testing sets. The labeled training data were used to build a model using the YOLOv4 architecture. The result of the modeling phase was a model, also called weight. To evaluate the model's performance, we detected the cracks found in the roads that were saved in the test data. The proposed method is shown in Figure 4.

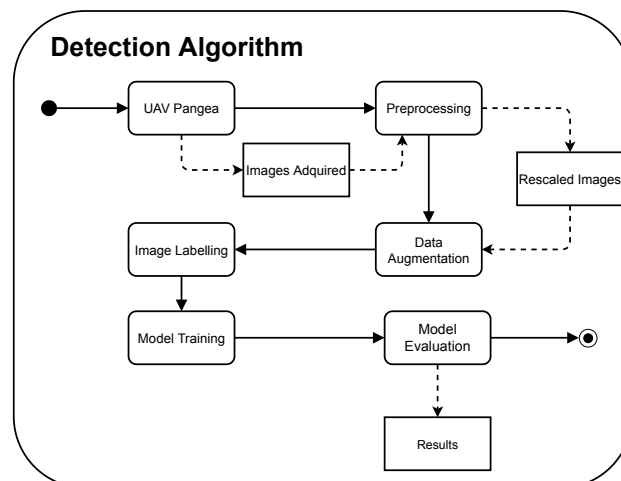


Figure 4. Block diagram of the drone detection solution.

3.2.1. Annotation and Labeling

The classification process is a manual process and must be performed by a person who knows how to differentiate whether the object in question is in the image and where the object is located. For this process, we used LabelImg software [54]. This graphic tool allows the annotation of images and delimitation of boxes of labeled objects in images. During this process, more than one object can be found, and each one must be delimited.

In this process of marking, we completed the annotations on all 568 images acquired. The information was stored in text, starting with the identification or annotation, differentiating potholes and cracks. In this case, a Class ID, the central position of the bounding box (x and y), and width and height of the bounding box (w and h) were stored. The Class ID is an integer value, starting at 0, and the bounding box information is a decimal format on a scale of 0–1. Each image in .jpg format has a file in .txt format with information about the holes, as shown in Equation (1).

$$\begin{aligned} &0 \ 0.365625 \ 0.745370 \ 0.020313 \ 0.055556 \\ &0 \ 0.665755 \ 0.531481 \ 0.048698 \ 0.053704 \end{aligned} \tag{1}$$

3.2.2. Model Training

This work was based on the YOLOv4-tiny model, which follows the principle of prediction of coordinates like YOLOv2 and YOLOv3. It is possible to multi-classify in YOLOv4 instead of classifying only one class, as in the older versions. YOLOv4 adopts a cross-loss entropy function instead of the multi-class loss function.

The network was configured to detect two classes. The filter number must be configured directly in the layer of the convolutional network. Thus, the formula used to apply the filter number is represented by Equation (2).

$$Filters = (Classes + 5) \times 3 \tag{2}$$

This number of filters must be replaced in the three convolutional layers before each YOLO layer, which only has to be the last convolution before each of the YOLO layers. In the model used in this work, the two last layers before each YOLO instance have 21 filters each. The networks used and configured are presented in Figure 5, highlighting the YOLO instances in blue.

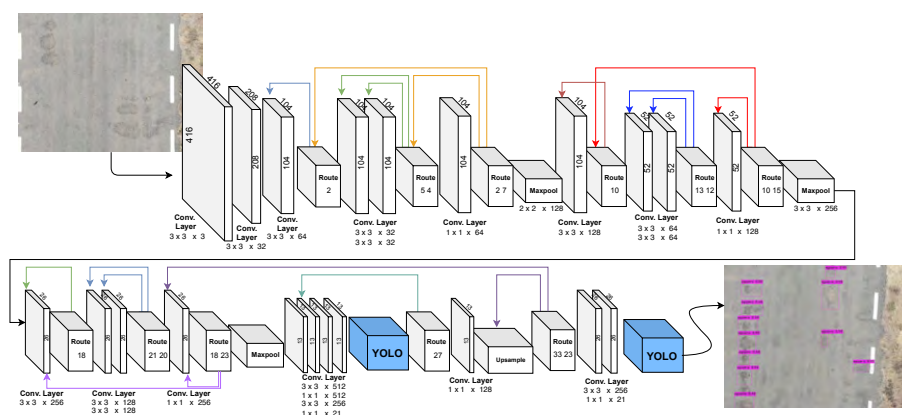


Figure 5. The network structure of the YOLOv4 pothole detection model.

By using this model in a graphics processing unit (GPU), YOLOv4 can be used with real-time images [55]. However, in devices with low processing capacity, such as the Nvidia Jetson, the conventional YOLOv4 algorithm works slowly. As an alternative, the YOLOv4-tiny network can be used, which satisfies the requirements in real time based on limited hardware resources. Therefore, we chose to use the YOLOv4-tiny algorithm. The structure of the YOLOv4-tiny network used is shown in Table 4.

Table 4. Network structure of YOLOv4-tiny.

Layer	Type	Filters	Size/Stride	Input	Output
0	Convolutional	32	$3 \times 3/2$	$416 \times 416 \times 3$	$208 \times 208 \times 32$
1	Convolutional	64	$3 \times 3/2$	$208 \times 208 \times 32$	$104 \times 104 \times 64$
2	Convolutional	64	$3 \times 3/1$	$104 \times 104 \times 64$	$104 \times 104 \times 64$
3	Route 2				
4	Convolutional	32	$3 \times 3/1$	$104 \times 104 \times 32$	$104 \times 104 \times 32$
5	Convolutional	32	$3 \times 3/1$	$104 \times 104 \times 32$	$104 \times 104 \times 32$
6	Route 5 4				
7	Convolutional	64	$1 \times 1/1$	$104 \times 104 \times 64$	$104 \times 104 \times 64$
8	Route 2 7				
9	Maxpool		$2 \times 2/2$	$104 \times 104 \times 128$	$52 \times 52 \times 128$
10	Convolutional	128	$3 \times 3/1$	$52 \times 52 \times 128$	$52 \times 52 \times 128$
11	Route 10				
12	Convolutional	64	$3 \times 3/1$	$52 \times 52 \times 64$	$52 \times 52 \times 64$
13	Convolutional	64	$3 \times 3/1$	$52 \times 52 \times 64$	$52 \times 52 \times 64$
14	Route 13 12				
15	Convolutional	128	$1 \times 1/1$	$52 \times 52 \times 128$	$52 \times 52 \times 128$
16	Route 10 15				
17	Maxpool		$2 \times 2/2$	$52 \times 52 \times 256$	$26 \times 26 \times 256$
18	Convolutional	256	$3 \times 3/1$	$26 \times 26 \times 256$	$26 \times 26 \times 256$
19	Route 18				
20	Convolutional	128	$3 \times 3/1$	$26 \times 26 \times 128$	$26 \times 26 \times 128$
21	Convolutional	128	$3 \times 3/1$	$26 \times 26 \times 128$	$26 \times 26 \times 128$
22	Route 21 20				
23	Convolutional	256	$1 \times 1/1$	$26 \times 26 \times 256$	$26 \times 26 \times 256$
24	Route 18 23				
25	Maxpool		$2 \times 2/2$	$26 \times 26 \times 512$	$13 \times 13 \times 512$
26	Convolutional	512	$3 \times 3/1$	$13 \times 13 \times 512$	$13 \times 13 \times 512$
27	Convolutional	256	$1 \times 1/1$	$13 \times 13 \times 512$	$13 \times 13 \times 256$
28	Convolutional	512	$3 \times 3/1$	$13 \times 13 \times 256$	$13 \times 13 \times 512$
29	Convolutional	21	$1 \times 1/1$	$13 \times 13 \times 512$	$13 \times 13 \times 21$
30	YOLO				
31	Route 27				
32	Convolutional	128	$1 \times 1/1$	$13 \times 13 \times 256$	$13 \times 13 \times 128$
33	Upsample		2x	$13 \times 13 \times 128$	$26 \times 26 \times 128$
34	Route 33 23				
35	Convolutional	256	$3 \times 3/1$	$26 \times 26 \times 384$	$26 \times 26 \times 256$
36	Convolutional	21	$1 \times 1/1$	$26 \times 26 \times 256$	$26 \times 26 \times 21$
37	YOLO				

For the training, we used the 1362 images present in the dataset. An image could be reproduced several times, increasing the training of our model. During the pre-processing stage, we manually detected and recognized the holes in the road, as explained in Section 3.2.1. The model training environment is composed of the DarkNet network. This neural network structure is written in C and CUDA language and can be executed directly on the GPU. It is installed depending on the GPU; we used it with Google Colab, which allowed us to perform calculations on a Tesla K80 GPU with 12 GB of memory.

4. Results

To evaluate our system, we first considered quantitative aspects, comparing the labeled dataset images with the final images classified by the algorithm. Next, for the qualitative evaluation, a series of applications were constructed to allow the user to carry out, in a centralized and unified way, all the tasks of our system, from the definition and follow-up of routes to the visualization and verification of the results.

4.1. Dataset

According to [34], the main objective in creating better roads is to avoid holes and cracks in the track, which requires precise diagnosis and differentiation between the types of problems that can be found in the pavement.

In the bibliographical research, we identified several datasets of holes and cracks in asphalt, but the datasets were not adequate for the proposed method of using a drone in a multi-agent system to capture pictures at a safe distance from the road. Therefore, we created a new set of data to represent the situation of Spanish roads. A total of 600 photos were taken at a resolution of 3840×2160 pixels. The images were resized to 1200×900 pixels. Some images of the dataset are demonstrated in Figure 6.

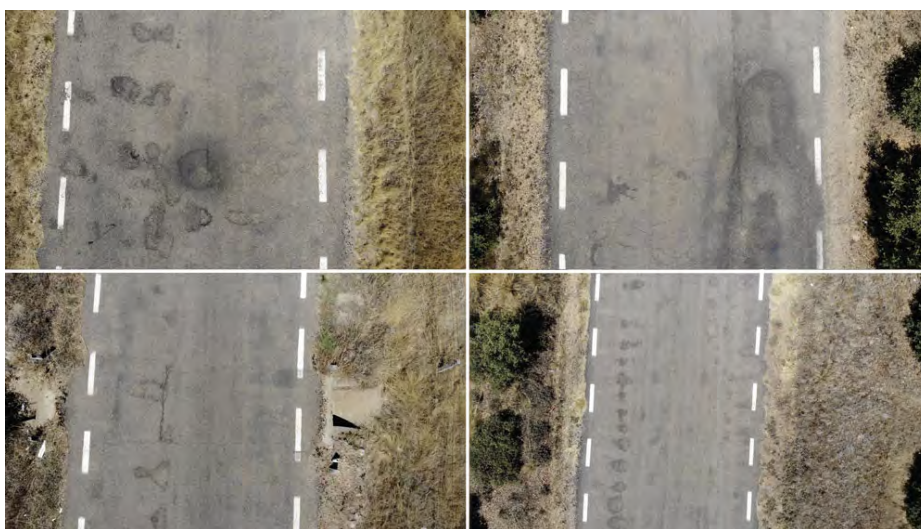


Figure 6. Dataset.

After labeling, the dataset contained a total of 568 labeled images. During the pre-processing stage, adjustments were made to the image orientation and resizing, as mentioned above. For each image in the set, different versions of the image were created using magnification techniques. The images were zoomed in and out, ranging from 0% to 15%. This process was repeated to increase the data set size from N to $2N$, repeating only the images that contained defects in the track.

The total number of images in the dataset was 1362 images. Of these images, 70% were for training, 20% for validation, and 10% for testing the effectiveness of the trained model. The dataset is composed of the images and their respective labels. An application (<https://roboflow.ai>) was used to generate the dataset, according to Figure 7.

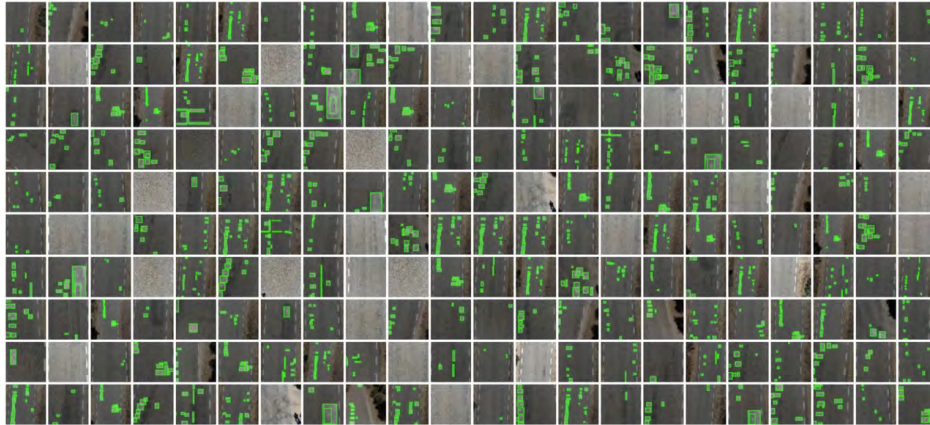


Figure 7. Dataset.

The dataset used for the validation of the results of this scientific article has been published at <https://github.com/luisaugustos/Pothole-Recognition> for verifying the results or for testing new algorithms by the community.

For the quantitative evaluation, we used the following metrics: accuracy, i.e., the relationship between true positives (TPs) and true positives (TPs) along with false positives (FPs) (Equation (3)); recall, which is the probability that an image is classified as positive and the relationship between the TPs and the TPs together with the false negatives (FNs) (Equation (4)); and F1, which is combination of the two previous metrics (Equation (5)). We classified the speed measured in frames per second (FPS); the mean average precision (mAP), calculated by the precision and recall curve; and the intersection over union (IoU), which is the overlapping area between the area found the image and the detected area. Lastly, we used the kappa metric (Equation (6)), which is the relationship between the relative observed agreement among raters $Pr(a)$, and the hypothetical probability of chance agreement $Pr(e)$, which is used to measure inter-rater reliability for qualitative items.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

$$Kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (6)$$

The training process of the model was evaluated in stages, alternating between iterations and image resolution. The first stage was used for YOLOv4-tiny using a pre-trained weight model, altering the convolution layers as needed. Every 1000 iterations during the training process, the model was stored; the result of the success of each of the values is shown in Figure 8. In the image, the value reached 6000 iterations and stabilized with an average success value of 94.6%.

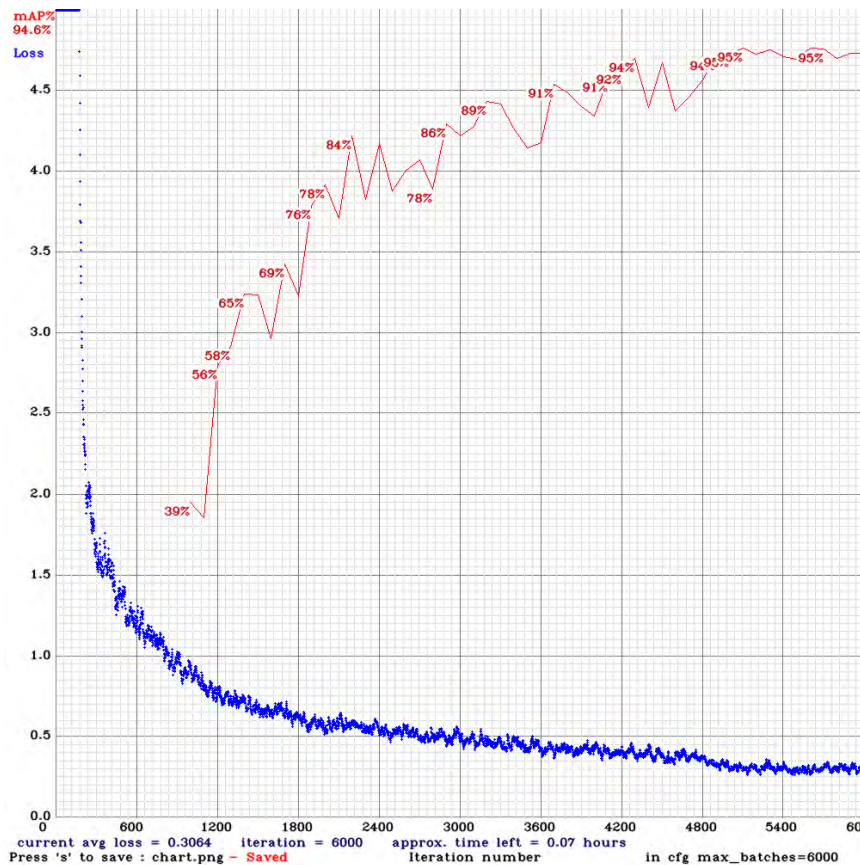


Figure 8. The result using 608 × 608 images with 6000 batches.

4.2. Detection

YOLOv4 has four data augmentation parameters provided by the author that were also used to generate more training samples by rotating the angle and adjusting the saturation, exposure, and hue of the images. YOLOv4 uses a new method of data enhancement, which was explained earlier in Section 2. The speed of image detection with YOLO initially averages 45 FPS. Its biggest failure is inaccuracy with small objects in the image.

The bounding box (bbox) method, for each grid cell, predicts B (bbox) and C probabilities of being one of the trained classes. A bbox prediction has five components: (x, y, w, h, and confidence). The coordinates (x, y) represent the center of the bbox relative to the location of the cell. If the center of a bbox is not in a cell, it will not be responsible for it and will not represent it. Cells only have a reference to objects whose center falls inside them. These coordinates are normalized to [0, 1]. The dimensions of the bbox (w, h) are also normalized to [0, 1] relative to the image size. Figures 9–11 depicts the result in the detection of road defects. In addition, results in Appendix A and a video demonstration added in Supplementary Materials.



Figure 9. Detection results with zoom. (a) Single detection. (b) Multiple detection.



Figure 10. Detection results in a large number of potholes. (a) Multiple detection. (b) Single pothole.

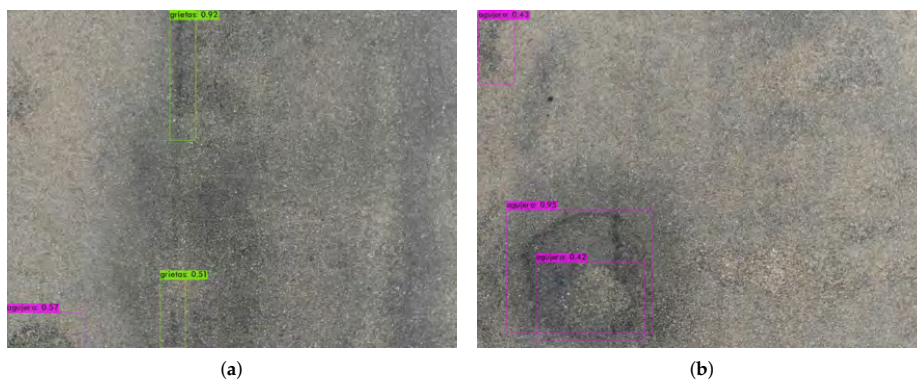


Figure 11. Crack detection and a detection of a pothole inside other. (a) Detection of a crack in the track. (b) Pothole inside another.

In experiment 1 (Table 5), concerning the label that classifies the potholes, we observed an accuracy of 96.25% and an AP of 98.46%. For the classification of the cracks, we obtained slightly worse accuracy and AP than for potholes, at 90.38% and 90.89%, respectively. This decrease occurred due to the lower number of images of cracks in the dataset. For all the images, the precision was 95.70%, the AP was at 94.67%, and the prediction time was 5.53 ms. We obtained a kappa metric of 0.73, which represents a high concordance.

Table 5. Obtained results.

Label	TP	FP	FN	Precision (%)	AP (%)	Kappa	Prediction Time (ms)
Pothole	977	38	-	96.25	98.46	-	-
Crack	94	10	-	90.38	90.89	-	-
Total	1070	48	62	95.70	94.67	0.73	5.53

In experiment 2 (Table 6), we obtained similar results to those in experiment 1. In the classification of the potholes, we obtained a precision of 96.54% and an AP of 98.45%; however, in the classification of the cracks, the precision was higher than in experiment 1, obtaining a similar AP of 92.00% and 90.68%, respectively. In all the images, the precision was 96.13%, the AP was 94.56%, and the prediction time was 5.52 ms. We obtained a kappa metric of 0.72, which represents high concordance, as in the first experiment.

Table 6. Obtained results.

Label	TP	FP	FN	Precision (%)	AP (%)	Kappa	Prediction Time (ms)
Pothole	977	35	-	96.54	98.45	-	-
Crack	92	8	-	92.00	90.68	-	-
Total	1069	43	64	96.13	94.56	0.72	5.52

To achieve complete functioning of the system, three different applications were developed, each with a well-defined objective, which can be used by users with different roles within the system.

4.3. Application

The first application in Figure 12a refers to the application used by the Road Monitor. In this application, the user is responsible for defining the routes where inspections should be performed, preparing the data for training, and validating the data of the detection output for the generation of reports and worksheets for the operators in charge of performing road maintenance. Although the system allows automatic operation, we decided to use expert supervision to avoid possible failures, identify the cracks or holes that need intervention, and assign these tasks to the operators. This reclassification will allow feedback on the system to improve the algorithm in the future.

The second application we developed is the Road Maintenance users' application, which is shown in Figure 12b. In this application, the user has a list of incidents, for which details, the photo, and the location of the incident are provided to enable its repair.

The last application is that used by the drone pilot. This application specifically uses the DJI SDK, since the drone used for the case study is of this brand. In this application, a list of tasks to be performed is provided and used by the pilot to know where the inspections are required and to perform real-time monitoring of the route followed by the drone during the inspection. This application takes the route coordinates calculated in the Road Monitor application to automatically create a route.

The main advantage offered by the development of the set of applications compared to the use of applications that currently exist on the market is a complete system from the definition of routes, to inspection, to the monitoring of results.



Figure 12. Application details.

5. Conclusions

In this work, we designed a platform capable of detecting damage on transport routes using drones and a multi-agent architecture. At the level of detection of anomalies, accurate results were obtained, since the proposed system enabled the platform to achieve precision higher than 95%. Despite the different crack detection techniques that were tested and analyzed during the literature review phase, none of them produced a result higher than 47%. Notably, one of the key factors producing this low performance was that the datasets found in different articles corresponded to non-European roads or that cracks were too large, so they were not similar to the small cracks and potholes that were detected in this work. The dataset for the verification of results in this work was developed on a regional road in the province of Salamanca. The imaging was performed according to the current legislation for the flight of drones in public locations. No variation was observed in the results obtained by the predictive algorithm between images taken at 70 m and at 90 m. The results were invariant regardless of the speed of the drone, which was 15–25 km/h during the tests.

The findings demonstrated how the procedures currently carried out by companies and regional governments responsible for road maintenance can be improved. Our method saves time, costs, and labor by establishing a more objective method for determining the road areas in need of repair as soon as possible.

The design of the solution proposed in the case study and based on virtual organizations of agents allows the testing of different techniques and the definition of parameters that are suitable for the user's final application, that is, without the final user detecting any kind of change during the reprogramming. The PANGAEA-based multi-agent architecture allows for parallelization of work according to the requirements of the platform at any given time, thus adapting to the needs of the context. The advantages of using an MAS include facilitating the development of case studies and ensuring compatibility between the different entities that compose the platform. The communication of the different agents implemented by means of the RFC 1459 Internet Relay Chat Protocol allows the optimization of the energy consumption necessary for communication, thus optimizing the battery of the drone as much as possible during the flight. To allow the scientific community to carry out research, both the pre-trained model and the datasets have been made public in a repository.

As future lines of this work, we plan to work on the coordination and distribution of works using drones with the main goal of distributing or subdividing a sampling area among several UAVs with different characteristics, such as weight, speed, and flight time. Due to this, we are currently studying different techniques to optimize tasks and routes based on battery life, altitude, and distance from the mapping zone to the initial point of takeoff using ardupilot as a flight controller. We also plan to continue our research by incorporating new 3D cameras or a LiDAR sensor in this drone or in a custom-made drone that allows knowledge of the distance from the drone to the road to calculate the size of the cracks or holes. Although we have started tests using techniques based on drone examples, no important results are yet available for dissemination.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-8220/20/21/6205/s1>.

Author Contributions: Conceptualization, L.A.S., D.P.G., A.S.M. and G.V.G.; Data curation, G.V.G.; Investigation, L.A.S., A.S.M., H.S.S.B. and D.P.G.; Methodology, H.S.S.B., D.P.G. and G.V.G.; Project administration, L.A.S. and G.V.G.; Resources, G.V.G.; Software, L.A.S., D.P.G. and G.V.G.; Supervision, A.S.M. and G.V.G.; Visualization, D.P.G.; Writing—original draft, L.A.S., H.S.S.B., D.P.G., A.S.M. and G.V.G.; Writing—review & editing, L.A.S., H.S.S.B., D.P.G., A.S.M. and G.V.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Junta De Castilla y León—Consejería De Economía Y Empleo: System for simulation and training in advanced techniques for occupational risk prevention through the design of hybrid-reality environments with ref J118. The research of André Filipe Sales Mendes was co-financed by the European Social Fund and Junta de Castilla y León (Operational Programme 2014–2020 for Castilla y León, EDU/556/2019 BOCYL).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AEC	Spanish Road Association
AP	Average Precision
FPS	Frames Per Second
FPN	Feature Pyramid Network
GPS	Global Positioning System
IoT	Internet of Things
JADE	Java Agent Development Framework
mAP	Mean Average Precision
MAS	Multi-Agent System
MARL	Multi-Agent Reinforcement Learning
YOLO	You Only Look Once
PANGAEA	Platform for Automatic coNstruction of orGanizations of intElligent Agents
UAV	Unmanned Aerial Vehicle
FPN	Feature Pyramid Network
SAM	Spatial Attention Module
SPP	Spatial Pyramid Pooling
RL	Reinforcement Learning
PANet	Prototype Alignment Network
SPADE	Smart Python Agent Development Environment

Appendix A



Figure A1. Detection results.

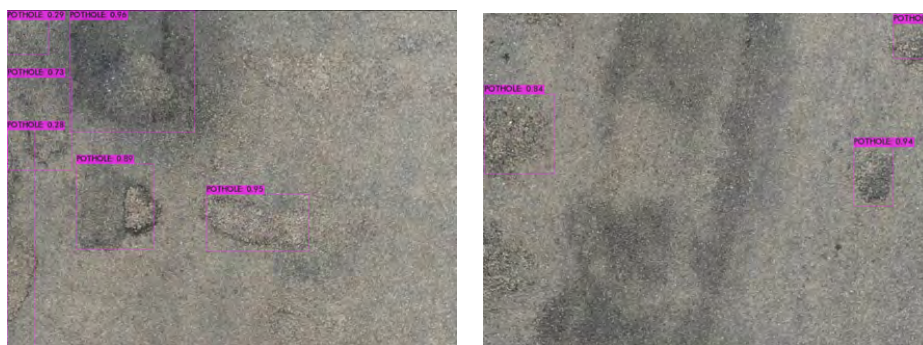


Figure A2. Detection results.

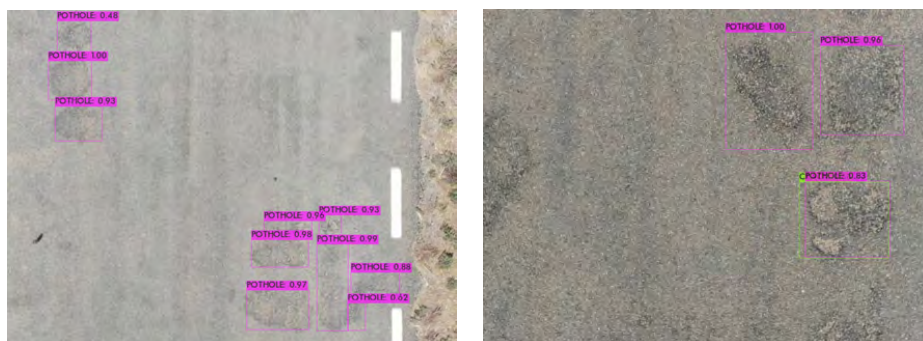


Figure A3. Detection results.



Figure A4. Detection results.

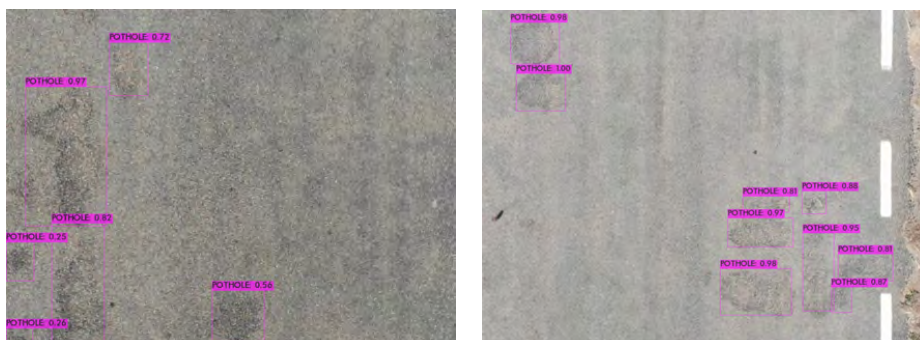


Figure A5. Detection results.

References

1. Harvey, J.; Al-Qadi, I.L.; Ozer, H.; Flintsch, G. *Pavement, Roadway, and Bridge Life Cycle Assessment 2020: Proceedings of the International Symposium on Pavement, Roadway, and Bridge Life Cycle Assessment 2020 (LCA 2020, Sacramento, CA, 3–6 June 2020)*; CRC Press: Boca Raton, FL, USA, 2020.
2. She, X.; Zhang, H.; Wang, Z.; Yan, J. Feasibility study of asphalt pavement pothole properties measurement using 3D line laser technology. *Int. J. Transp. Sci. Technol.* **2020**. [\[CrossRef\]](#)
3. Wang, H.; Chen, C.H.; Cheng, D.; Lin, C.H.; Lo, C.C. A Real-Time Pothole Detection Approach for Intelligent Transportation System. *Math. Probl. Eng.* **2015**. [\[CrossRef\]](#)
4. Li, W.; Shen, Z.; Li, P. Crack Detection of Track Plate Based on YOLO. In Proceedings of the 2019 12th International Symposium on Computational Intelligence and Design, ISCID 2019, Hangzhou, China, 14–15 December 2019; pp. 15–18. [\[CrossRef\]](#)
5. Kim, J.; Kim, S.; Jeong, J.; Kim, H.; Park, J.S.; Kim, T. CBDN: Cloud-Based Drone Navigation for Efficient Battery Charging in Drone Networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 4174–4191. [\[CrossRef\]](#)
6. Celtek, S.A.; Durdu, A.; Kurnaz, E. Design and Simulation of the Hierarchical Tree Topology Based Wireless Drone Networks. In Proceedings of the 2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018, Malatya, Turkey, 28–30 September 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019. [\[CrossRef\]](#)
7. Wang, D.; Hu, P.; Du, J.; Zhou, P.; Deng, T.; Hu, M. Routing and Scheduling for Hybrid Truck-Drone Collaborative Parcel Delivery With Independent and Truck-Carried Drones. *IEEE Internet Things J.* **2019**, *6*, 10483–10495. [\[CrossRef\]](#)
8. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)

9. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
10. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
11. Kim, N.V.; Mikhailov, N.A.; Mokrova, M.I. Drone Searches in Challenging Conditions. *Russ. Eng. Res.* **2020**, *40*, 583–585. [[CrossRef](#)]
12. Hodge, V.J.; Hawkins, R.; Alexander, R. Deep reinforcement learning for drone navigation using sensor data. *Neural Comput. Appl.* **2020**, 1–19. [[CrossRef](#)]
13. Shukla, M.; Chen, Z.; Lu, C.T. DIMPL: A distributed in-memory drone flight path builder system. *J. Big Data* **2018**, *5*, 1–29. [[CrossRef](#)]
14. Abou, U.; Belkaid, B. Dynamic Reconfiguration of LPWANs Pervasive System Using Multi-agent Approach. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 300–305.
15. Sales Mendes, A.; Jiménez-Bravo, D.M.; Navarro-Cáceres, M.; Reis, V.; Leithardt, Q.; Villarrubia González, G. Multi-Agent Approach Using LoRaWAN Devices: An Airport Case Study. *Electronics* **2020**, *9*, 1430. [[CrossRef](#)]
16. Villarrubia, G.; De Paz, J.F.; Bajo, J.; Corchado, J.M. Ambient agents: Embedded agents for remote control and monitoring using the PANGEA platform. *Sensors* **2014**, *14*, 13955–13979. [[CrossRef](#)] [[PubMed](#)]
17. Das, L.B.; Mohan, V.; George, G. Human Target Search and Detection using Autonomous UAV and Deep learning. In Proceedings of the 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 7–8 July 2020.
18. Yang, Q.; Shi, L.; Han, J.; Yu, J.; Huang, K. A near real-time deep learning approach for detecting rice phenology based on UAV images. *Agric. For. Meteorol.* **2020**, *287*. [[CrossRef](#)]
19. Chew, R.; Rineer, J.; Beach, R.; O’Neil, M.; Ujeneza, N.; Lapidus, D.; Miano, T.; Hegarty-Craver, M.; Polly, J.; Temple, D.S. Deep Neural Networks and Transfer Learning for Food Crop Identification in UAV Images. *Drones* **2020**, *4*, 7. [[CrossRef](#)]
20. Cledat, E.; Jospin, L.V.; Cucci, D.A.; Skaloud, J. Mapping quality prediction for RTK/PPK-equipped micro-drones operating in complex natural environment. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 24–38. [[CrossRef](#)]
21. Alwateer, M.; Loke, S.W.; Fernando, N. Enabling Drone Services: Drone Crowdsourcing and Drone Scripting. *IEEE Access* **2019**, *7*, 110035–110049. [[CrossRef](#)]
22. Watanabe, K.; Takahashi, M. Head-synced Drone Control for Reducing Virtual Reality Sickness. *J. Intell. Robot. Syst. Theory Appl.* **2020**, *97*, 733–744. [[CrossRef](#)]
23. Deng, L.L. Pre-detection Technology of Clothing Image Segmentation Based on GrabCut Algorithm. *Wirel. Pers. Commun.* **2018**, *102*, 599–610. [[CrossRef](#)]
24. Miao, Y.; Jeon, J.Y.; Park, G. An image processing-based crack detection technique for pressed panel products. *J. Manuf. Syst.* **2020**, *57*, 287–297. [[CrossRef](#)]
25. Ren, Y.; Huang, J.; Hong, Z.; Lu, W.; Yin, J.; Zou, L.; Shen, X. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr. Build. Mater.* **2020**, *234*, 117367. [[CrossRef](#)]
26. Wang, Y.; Zhang, J.Y.; Liu, J.X.; Zhang, Y.; Chen, Z.P.; Li, C.G.; He, K.; Yan, R.B. Research on Crack Detection Algorithm of the Concrete Bridge Based on Image Processing. *Procedia Comput. Sci.* **2018**, *154*, 610–616. [[CrossRef](#)]
27. Miskdjian, I.; Hajikazemi, M.; Van Paeppegem, W. Automatic edge detection of ply cracks in glass fiber composite laminates under quasi-static and fatigue loading using multi-scale Digital Image Correlation. *Compos. Sci. Technol.* **2020**, *200*, 108401. [[CrossRef](#)]
28. Dhimish, M.; Holmes, V. Solar cells micro crack detection technique using state-of-the-art electroluminescence imaging. *J. Sci. Adv. Mater. Devices* **2019**, *4*, 499–508. [[CrossRef](#)]
29. Gehri, N.; Mata-Falcón, J.; Kaufmann, W. Automated crack detection and measurement based on digital image correlation. *Constr. Build. Mater.* **2020**, *256*, 119383. [[CrossRef](#)]
30. Qiu, Z.; Wang, S.; Zeng, Z.; Yu, D. Automatic visual defects inspection of wind turbine blades via YOLO-based small object detection approach. *J. Electron. Imaging* **2019**, *28*, 1. [[CrossRef](#)]
31. Park, S.E.; Eem, S.H.; Jeon, H. Concrete crack detection and quantification using deep learning and structured light. *Constr. Build. Mater.* **2020**, *252*, 119096. [[CrossRef](#)]

32. Fang, F.; Li, L.; Gu, Y.; Zhu, H.; Lim, J.H. A novel hybrid approach for crack detection. *Pattern Recognit.* **2020**, *107*, 107474. [[CrossRef](#)]
33. Hwang, H.; Oh, J.; Lee, K.H.; Cha, J.H.; Choi, E.; Yoon, Y.; Hwang, J.H. Synergistic approach to quantifying information on a crack-based network in loess/water material composites using deep learning and network science. *Comput. Mater. Sci.* **2019**, *166*, 240–250. [[CrossRef](#)]
34. Du, Y.; Pan, N.; Xu, Z.; Deng, F.; Shen, Y.; Kang, H. Pavement distress detection and classification based on YOLO network. *Int. J. Pavement Eng.* **2020**. [[CrossRef](#)]
35. Majidifard, H.; Adu-Gyamfi, Y.; Buttlar, W.G. Deep machine learning approach to develop a new asphalt pavement condition index. *Constr. Build. Mater.* **2020**, *247*, 118513. [[CrossRef](#)]
36. Xia, B.; Cao, J.; Zhang, X.; Peng, Y. Automatic concrete sleeper crack detection using a one-stage detector. *Int. J. Intell. Robot. Appl.* **2020**, *4*, 319–327. [[CrossRef](#)]
37. Nie, M.; Wang, C. Pavement Crack Detection based on yolo v3. In Proceedings of the 2019 2nd International Conference on Safety Produce Informatization (IICSPI 2019), Chongqing, China, 28–30 November 2019; pp. 327–330. [[CrossRef](#)]
38. Peng, C.; Yang, M.; Zheng, Q.; Zhang, J.; Wang, D.; Yan, R.; Wang, J.; Li, B. A triple-thresholds pavement crack detection method leveraging random structured forest. *Constr. Build. Mater.* **2020**, *263*, 120080. [[CrossRef](#)]
39. Zhuang, Y.; Ke, R.; Wang, Y. Edge-Based Traffic Flow Data Collection Method Using Onboard Monocular Camera. *J. Transp. Eng. Part A Syst.* **2020**, *146*, 04020096. [[CrossRef](#)]
40. Banerjee, S.; Singh, S.K.; Chakraborty, A.; Das, A.; Bag, R. Melanoma Diagnosis Using Deep Learning and Fuzzy Logic. *Diagnostics* **2020**, *10*, 577. [[CrossRef](#)] [[PubMed](#)]
41. Mandal, V.; Uong, L.; Adu-Gyamfi, Y. Automated Road Crack Detection Using Deep Convolutional Neural Networks. In Proceedings of the 2018 IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, 10–13 December 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 5212–5215. [[CrossRef](#)]
42. Zhang, J.; Yang, X.; Li, W.; Zhang, S.; Jia, Y. Automatic detection of moisture damages in asphalt pavements from GPR data with deep CNN and IRS method. *Autom. Constr.* **2020**, *113*, 103119. [[CrossRef](#)]
43. Luo, Z.; Yu, H.; Zhang, Y. Pine cone detection using boundary equilibrium generative adversarial networks and improved YOLOv3 model. *Sensors* **2020**, *20*, 4430. [[CrossRef](#)]
44. Morera, Á.; Sánchez, Á.; Moreno, A.B.; Sappa, Á.D.; Vélez, J.F. Ssd vs. Yolo for detection of outdoor urban advertising panels under multiple variabilities. *Sensors* **2020**, *20*, 4587. [[CrossRef](#)]
45. Cao, C.; Wu, J.; Zeng, X.; Feng, Z.; Wang, T.; Yan, X.; Wu, Z.; Wu, Q.; Huang, Z. Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network. *Sensors* **2020**, *20*, 4696. [[CrossRef](#)]
46. Huang, Z.; Wang, J.; Fu, X.; Yu, T.; Guo, Y.; Wang, R. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection. *Inf. Sci.* **2020**, *522*, 241–258, [[1903.08589](#)]. [[CrossRef](#)]
47. Wu, G.; Gao, X.; Wan, K. Mobility control of unmanned aerial vehicle as communication relay to optimize ground-to-air uplinks. *Sensors* **2020**, *20*, 2322. [[CrossRef](#)]
48. Walker, O.; Vanegas, F.; Gonzalez, F. A framework for multi-agent UAV exploration and target-finding in GPS-denied and partially observable environments. *Sensors* **2020**, *20*, 4739. [[CrossRef](#)] [[PubMed](#)]
49. Welcome to SPADE's Documentation!—SPADE 3.1.4 Documentation. Available online: <https://spade-mas.readthedocs.io/en/latest/index.html> (accessed on 8 September 2020).
50. Anandan, R. Machine Condition Monitoring Software Agent Using JADE and Data Mining. *J. Inst. Eng. Ser. B* **2015**, *96*, 61–67. [[CrossRef](#)]
51. Zato, C.; Villarrubia, G.; Sanchez, A.; Barri, I.; Rubión, E.; Fernández, A.; Sánchez, C.; Cabo, J.; Álamos, T.; Sanz, J.; et al. PANGEA -Platform for Automatic coNstruction of orGanizations of intElligent Agents. In *Advances in Intelligent and Soft Computing*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 151. [[CrossRef](#)]
52. osBrain—0.6.5—osBrain 0.6.5 Documentation. Available online: <https://osbrain.readthedocs.io/en/stable/index.html> (accessed on 8 September 2020).
53. Dhiman, A.; Klette, R. Pothole Detection Using Computer Vision and Learning. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3536–3550. [[CrossRef](#)]

Sensors **2020**, *20*, 6205

23 of 23

54. Tzutalin/LabelImg. Free Software: MIT License. Available online: <https://github.com/tzutalin/labelImg> (accessed on 20 August 2020).
55. Fang, W.; Wang, L.; Ren, P. Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments. *IEEE Access* **2020**, *8*, 1935–1944. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.2.1. Resumen en Español

Esta subsección muestra un resumen en castellano de los aspectos más relevantes del artículo científico titulado “*An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images*” que en castellano se podría traducir a «Una arquitectura multiagente para monitorización de pavimentos con reconocimiento de baches en imágenes de vehículo aéreo no tripulado».

Objetivos de la investigación

En los últimos años, las tareas de mantenimiento en las vías de transporte de uso público se han reducido drásticamente debido a la difícil situación económica que presentan muchos países. Los diferentes estudios que han sido realizados por agrupaciones de conductores y colectivos relacionados con la seguridad vial, concluyen que se está experimentando un aumento de accidentes debido al mal estado del pavimento de la carretera, afectando incluso, al estado de los vehículos mediante la aparición de costosas averías. Actualmente, los procesos de detección de cualquier tipo de desperfecto en la vía, son llevados a cabo manualmente o se basan en la utilización de un vehículo de carretera, lo que conlleva un alto coste de mano de obra. Para poder dar solución a este problema, son muchos los centros de investigación que se esfuerzan en indagar sobre técnicas de procesamiento de imágenes con el objetivo de identificar el mal estado de la calzada mediante la utilización de algoritmos de aprendizaje profundo.

El objetivo de este artículo ha sido el desarrollo de una plataforma capaz de monitorizar el estado de las carreteras siendo capaz de detectar diferentes desperfectos en la misma, de forma que todas las personas y organizaciones que forman parte del proceso de identificación se encuentren coordinadas y comunicadas. Con el objetivo de coordinar y comunicar todas las entidades partícipes durante la realización de esta tarea, se propone la utilización de un sistema multi-agente (MAS) que permite una reconfiguración dinámica del sistema. Además, el sistema multi-agente lleva a cabo una distribución de recursos y capacidades para evitar problemas que acontecen a los sistemas centralizados tales como cuello de botella o el acceso recurrente a recursos críticos.

Cabe destacar que, los modelos de machine learning dependen de la calidad de los datos usados para su entrenamiento, por tanto, para poder obtener un modelo robusto para esta tarea, se ha creado un novedoso dataset preparado para su uso por algoritmos de deep learning. Además, se ha propuesto el uso de un UAV para la obtención de las imágenes de forma automática mediante el seguimiento de una carretera, de forma que se reduzca el esfuerzo requerido para el proceso de captación de muestras, tanto a nivel económico como de tiempo. Mediante la arquitectura multi-agente propuesta, se ha propuesto coordinación y comunicación entre los distintos elementos del sistema de forma que los diferentes clasificadores que sean probados sean transparente a los usuarios finales.

El objetivo principal de este trabajo es el diseño de una plataforma distribuida que permita detectar desperfectos en las vías de transporte utilizando drones. Para la consecución exitosa de este trabajo, se ha conformado un caso de estudio haciendo uso de un sistema multi-agente que coordine las diferentes partes de la arquitectura mediante la utilización de técnicas basada en la computación ubicua.

El presente artículo establece varios objetivos de investigación que buscan innovar en la detección de desperfectos en las vías de transporte. Estos objetivos son:

- Diseño de una plataforma distribuida que utilice drones para identificar imperfecciones en carreteras.
- Explorar algoritmos especializados en detección de patrones en imágenes, así como en técnicas de aprendizaje profundo.
- Compilar un repositorio de imágenes de carreteras españolas que muestren desperfectos específicos, diferenciando entre agujeros y grietas, todas obtenidas mediante drones.
- Entrenar y evaluar un modelo de inteligencia artificial basado en el repositorio de imágenes creado.
- Implementar un sistema multi-agente que optimice y controle rutas de inspección para inspecciones automatizadas.

Metodología

La metodología utilizada durante la elaboración de este artículo científico es la misma que se ha explicado en la sección 1.2.

- **Revisión del estado del arte:** Para contextualizar y fundamentar la investigación presentada en este artículo, se realizó una exhaustiva revisión de trabajos previos en el campo de estudio. Esta revisión buscó comprender, ampliar, mejorar y, en algunos casos, reutilizar investigaciones existentes, con el fin de garantizar que los aportes de este trabajo sean relevantes y de calidad para la comunidad científica. Tras una revisión detallada del estado de la literatura no se ha encontrado referencia o conocimiento de que exista una aplicación de detección de grietas que haga uso de capacidades tecnológicas distribuidas adaptadas para el reconocimiento de las vías españolas.
- **Análisis y diseño de la propuesta:** analizando lo aprendido en el estado del arte, se ha propuesto el diseño de subsistemas que interaccionan entre sí. Dentro de la propuesta diseñada, se desarrollaron tres aplicaciones interconectadas. La primera, Aplicación Road Monitor, permite al usuario definir rutas de inspección, preparar datos y validar detecciones, optando por una supervisión experta para garantizar precisión. La segunda, Aplicación Road Maintenance, está destinada a operarios de mantenimiento, ofreciendo detalles y ubicaciones de desperfectos. La última es para el piloto de drones, integrada con el SDK de DJI, permitiendo supervisión en tiempo real y navegación automática basada en las rutas definidas previamente.
- **Definición del caso de estudio:** Aunque se identificaron diversos conjuntos de datos en investigaciones bibliográficas sobre estos problemas en el asfalto, ninguno se adaptó al enfoque innovador de este estudio: usar un drone en un sistema multi-agente para tomar fotografías desde una distancia segura. Los conjuntos de datos tampoco se referían a carreteras y vías europeas. Por ello, se creó un conjunto de datos específico para reflejar la situación de las carreteras en España, compuesto por 600 fotos de 3840x2160 píxeles, que luego se redimensionaron a 1200x900 píxeles. Estas imágenes, capturadas por la cámara

del UAV a 50 metros del suelo, sirvieron para entrenar y probar redes neuronales profundas,

- **Desarrollo:** A lo largo del proceso de recopilación de datos, se capturaron vídeos y se extrajeron fotogramas de estos para ser usados en el entrenamiento. Este proceso dio lugar a un total de 600 imágenes de alta resolución, las cuales se categorizaron en dos clases: grietas y agujeros. Posteriormente, estas imágenes fueron etiquetadas según su clasificación. El estudio siguió distintas fases. Se inició con la preparación de los datos, seguida de la anotación y etiquetado. A continuación, el conjunto de datos se dividió en datos para entrenamiento y datos para pruebas. Utilizando la arquitectura YoloV4 y también una versión adaptada del YoloV4-tiny, se emplearon los datos etiquetados de entrenamiento para construir un modelo. El resultado de esta fase de modelado se refiere comúnmente como "ponderación". Para valorar la eficacia del modelo, se pusieron a prueba las detecciones de grietas en las imágenes de prueba.

Resultados

Con el objetivo de evaluar el sistema, se ha centrado en primer lugar a aspectos cuantitativos, comparando las imágenes del dataset etiquetadas con las imágenes finales clasificadas por el algoritmo. En segundo lugar para la evaluación del aspecto cualitativo se realizan una serie de aplicaciones que permiten al usuario realizar de forma centralizada y unificada todas las tareas de nuestro sistema, desde la definición y el seguimiento de rutas hasta la visualización y verificación de los resultados.

En la evaluación cuantitativa, empleamos métricas como la precisión, definida por la relación entre verdaderos positivos (TP) y la suma de TP y falsos positivos (FP). También se ha considerado la velocidad de clasificación en FPS, el mean average precision (mAP) derivado de la curva de precisión y recall, y el intersection over union (IoU), que mide el solapamiento entre áreas detectadas y reales.

Los resultados conseguidos mediante la adaptación del clasificador YoloV4 han alcanzado una precisión media de 95 %. En los experimentos se observaron tasas de acierto para baches de 96.25 % y 96.54 % respectivamente. En grietas, la precisión fue de 90.38 % en el primer experimento y aumentó al 92.00 % en el segundo. Estas variaciones están influenciadas por la cantidad de imágenes de grietas en el conjunto de datos. En general, las precisiones globales fueron del 95.70 % y 96.13 %, y los valores AP de 94.67 % y 94.56 %, respectivamente.

Conclusión

Diseñamos una plataforma con drones y una arquitectura multi-agente para detectar imperfecciones en vías, logrando una precisión del 95 %. Aunque se analizaron técnicas existentes, ninguna superó el 47 % en precisión, en parte debido a datasets no apropiados para carreteras europeas. El dataset creado se basó en carreteras de Salamanca, respetando la legislación de drones. Los resultados fueron consistentes para imágenes tomadas desde diferentes altitudes y velocidades de dron e entre 15-25km/h. Por fin, como contribución, las imágenes utilizadas han sido publicadas en una dataset para que la comunidad científica pueda hacer uso de las mismas.

3.3. Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning



Article

Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning

Luis Augusto Silva ^{1,*}, André Sales Mendes ¹, Héctor Sánchez San Blas ¹, Lia Caetano Bastos ², Alexandre Leopoldo Gonçalves ² and André Fabiano de Moraes ^{2,3,*}

- ¹ Expert Systems and Applications Lab (ESALAB), Faculty of Science, University of Salamanca, 37008 Salamanca, Spain
- ² Department of Knowledge Engineering and Management, Federal University of Santa Catarina, Florianópolis 88040-900, Brazil
- ³ Department Information Technology, IT Institute Federal of Science Technology IFC, Camboriú 88340-055, Brazil
- * Correspondence: luisaugustos@usal.es (L.A.S.); andre.moraes@ifc.edu.br (A.F.d.M.)

Abstract: Due to the increasing urban development, it has become important for municipalities to permanently understand land use and ecological processes, and make cities smart and sustainable by implementing technological tools for land monitoring. An important problem is the absence of technologies that certify the quality of information for the creation of strategies. In this context, expressive volumes of data are used, requiring great effort to understand their structures, and then access information with the desired quality. This study are designed to provide an initial response to the need for mapping zones in the city of Itajaí (SC), Brazil. The solution proposes to aid object recognition employing object-based classifiers OneR, NaiveBayes, J48, IBk, and Hoeffding Tree algorithms used together with GeoDMA, and a first approach in the use of Region-based Convolutional Neural Network (R-CNN) and the YOLO algorithm. All this is to characterize vegetation zones, exposed soil zones, asphalt, and buildings within an urban and rural area. Through the implemented model for active identification of geospatial objects with similarity levels, it was possible to apply the data crossover after detecting the best classifier with accuracy (85%) and the kappa agreement coefficient (76%). The case study presents the dynamics of urban and rural expansion, where expressive volumes of data are obtained and submitted to different methods of cataloging and preparation to subsidize rapid control actions. Finally, the research describes a practical and systematic approach, evaluating the extraction of information to the recommendation of knowledge with greater scientific relevance. Allowing the methods presented to apply the calibration of values for each object, to achieve results with greater accuracy, which is proposed to help improve conservation and management decisions related to the zones within the city, leaving as a legacy the construction of a minimum technological infrastructure to support the decision.

Keywords: machine learning; information extraction; object spatial; smart cities; gis detection



Citation: Silva, L.A.; Sales Mendes, A.; Sánchez San Blas, H.; Caetano Bastos, L.; Leopoldo Gonçalves, A.; Fabiano de Moraes, A. Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning. *Sensors* **2023**, *23*, 138. <https://doi.org/10.3390/s23010138>

Academic Editor: Natividad Duro Carralero

Received: 30 October 2022
Revised: 16 December 2022
Accepted: 20 December 2022
Published: 23 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Constant changes occur due to technological evolution in various areas of knowledge, generating new evolutionary cycles, bringing demands, and presenting a large number of still fragmented solutions. For example, the laws of fiscal responsibility, access to information, the civil mark of the internet, and the general law of data protection, guarantee the right to be well informed about what is produced in the public sector. They regulate the implementation of general concepts on data protection, rules for active transparency, and operational procedures. However, the provision of public data in an open format

To be able to make all the contributions, a methodology of standardization of the steps is applied with the DSR (Design Science Research) proposed by [21], to add new practices and build a literature review in parallel to the technological steps that help in the implementation of the proposal. To this end, the processes of (a) classification in the context of remote sensing (RS) are also reviewed; (b) text mining techniques with the software Weka (workbench for machine learning) performing text analysis, quantifying the words and extracting statistics with the TF-IDF method (Frequency-Inverse Document Frequency). Through a case study carried out in the municipality of Itajaí in the state of Santa Catarina, Brazil. The results collected for the urban and regional planning evaluation period are presented, considering the crossing of environmental and social data, referring to territorial occupation.

For a better understanding of the research and results of this work, the paper is organized as follows: Section 2 presents the development; Section 3 describes the methodology, in Section 4 the materials and methods used to apply the case study are detailed; in Section 5 we present the discussion of the results, and finally, in Section 6 are the final conclusions and future work.

2. Development

To develop the project, actions were initially planned for the acquisition of spatial data and satellite images. In the second stage, cataloging and standardization processes are carried out for data interoperability. The third stage presents the technological solutions adopted from GeoDMA and TerraView for remote sensing (RS). In the fourth step occurs the Weka implementation applied to the spatial context. To orchestrate the processes, the main methodology of the work aims from the experiments to define the best practices for the classification process focused on the SR, enabling the integration of all processes towards the active learning of objects. At the end, the case study is presented with specific results and discussions.

2.1. Spatial Data Acquisition and Satellite Imagery

In particular for matrix data, [22] defines spatial data acquisition from images from a few approaches, those being: input data definition, segmentation, detection cycles, cell space (matrix) creation, and preliminary extraction features. From the proposals [23–25] implemented through the GeoDMA framework (GEOBIA), in synthesis provides the realization of the steps of segmentation of satellite images, extraction of attributes, creation of classification rules, hierarchical classification and visualization of results. Additionally, the works [19,26–29] describe in detail the precautions to be taken in image acquisition and processing. In particular, according to [27,30], the monitoring of the interactions with the terrestrial surfaces is very important, where each intensity of the solar radiation must be observed. That is, the intensity of solar radiation received by the surface depends on the solar zenith angle that is calculated, considering [27,31] the incident solar radiation, the solar radiation intensity and the solar zenith angle. After acquiring the digital data from the sources provided according to [32], new processes for cataloging emerge. However, the extraction characteristics, initially present the need for specific module of resources extraction and storage in a database. From this, it is possible to extract several features, depending on the size of the scanning process and the amount of objects this task can be very time consuming to complete. Therefore, cataloging guarantees that the features will be extracted without losing original characteristics. Experiments conducted using the images collected by the CBERS-4 (China-Brazil Earth Resources Satellite) and the CBERS-4A satellite, located in <http://www.dgi.inpe.br/documentacao/dgi/documentacao/satelites/cbers/capa-cbers> (accessed on 1 June 2021), considering different periods.

2.2. Machine Learning and Processing

The availability of images from satellites and aerial platforms over the Earth's surface in the most diverse resolutions has been enabling an unprecedented approach between

technology and society, as [28] the processing of large volumes of data and geolocation for the use of mobile devices in most different devices makes the insertion of various technologies flexible. However, large volumes of data are generated, and for analysis, new challenges arise involving interoperability, from those related to data collection and storage, through ethics and privacy [33–35], to the development of efficient and robust algorithms to extract the most unimaginable information.

However, processing large volumes of data requires technical expertise in remote sensing, raw data processing, information extraction, a transformation of orthogonal models and spectral mixing, calculation of physical indices, arithmetic operations, frequencies, and statistical classification of data. However, all of these resources seek to assist in classifying image pixels associated with the presence of observed spatial object characteristics. To enable this, only with the use of consolidated technologies as [36,37], various classification algorithms have been developed, as there is a growth in the data obtained by images.

Each classifier has its strengths and weaknesses. Hybridizing classifiers with each other have the potential to combine the strengths and overcome the weaknesses by analyzing level by level as per Figure 1.

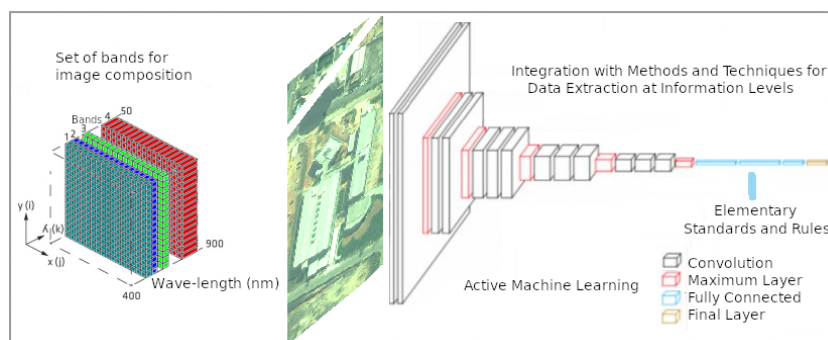


Figure 1. Architecture of the ecosystem adopted for active detection, extraction, and learning of geospatial objects.

It is necessary to advance with new studies on the behavior of different algorithms for hybrid classification as KNN (K-Nearest Neighbors) and SVM (Support Vector Machine) addressed by [38], with genetic programming, decision trees with an artificial neural network, Naive Bayes with decision trees and decision trees with K-means. Additionally, in hybrid image processing, specifically for the segmentation process, basic procedures are established according to the works of [22,39], and there are still procedures to be evaluated.

For the processing of images according to [29], together with the extraction of attributes from the regions and their spectral characteristics, they must be previously identified, validated, and calculated. The attributes can be determinant in interpretations and classification processes that involve many classes and some with little separability.

From this, the challenge arises to increase the amount of coherent information to facilitate the discrimination of spectrally similar classes. To this end, determining procedures that help increase identification with greater accuracy provides a set of rules that can be instrumental in identifying distributed objects.

For image processing according to [29], besides the extraction of attributes from the regions and their spectral characteristics, they must be previously identified, validated and calculated. The attributes can be determinant in interpretations and classification processes that involve many classes and some with little separability. From this, the challenge arises to increase the amount of coherent information to facilitate the discrimination of spectrally similar classes.

To this end, determining procedures that help increase identification with greater accuracy provides a set of rules that can be instrumental in identifying distributed objects.

However, new attributes can be associated with the spectral attributes from new dimensions attached to the original image. The spectral attributes refer to the color of the pixels and are calculated separately for each band of the input image based on the pixels belonging to the segment.

As per categorical data analysis [40], three standards of texture verification are still evaluated: structural, statistical, and spectral approaches. A statistical approach is the most widely used and considers the texture of an image as a quantitative measure of the arrangement of intensities in a given region. In this context, the concurrency matrix contemplates the numerical characteristics of the texture using similar shades of gray between a pixel and its adjacencies determined by the $N \times N$ pixel quadrant. The main formulations adopted to quantify the texture in digital images, in this case, obtained by (the CBERS-4 satellite), are mean, variance and entropy. The mean, according to (1), corresponds to the value of the arithmetic mean of the gray levels of a region in each band of the image. Where, $R(i)$ equals each element (i) segment R and N the total number of pixels.

$$Average = \left(\sum_{i=1}^N .Ri \right) N \quad (1)$$

$$Variance = \left(\sum_{i=1}^N .i - M^2 \right) \quad (2)$$

$$Entropy = \left(\sum_{i=1}^N .Pi.InPI \right) \quad (3)$$

The variance (2) is a measure of the dispersion of the gray level values of the pixels in the region around the mean, and M is the mean of the gray levels of the segment. Entropy (3) is calculated based on the distribution of pixel values in the region and is a measure equivalent to the “distortion” of the values in the region. Where $P(i)$ contains a normalized histogram of the segment elements. The geometric attributes [29] are calculated based on the polygon that defines the segment boundary, being: area, perimeter, compactness, convexity, and elongation. In other words, the geometric metrics of a segment are defined in advance for the behavior of the processing algorithm. The main considerations about image segmentation refer to the selection of the optimal parameters for each application. However, new active and adaptive processes have presented important results with machine learning, such as GeoDMA.

2.3. GeoDMA and TerraVIEW for Remote Sensing

To analyze altered patterns [22,23], in large remote sensing datasets, GeoDMA was created. Implemented in TerraVIEW software, a tool to integrate the most essential image analysis algorithms, ecology metrics, a scheme for multitemporal analysis [41] and data mining techniques to automate the analysis of large databases. Addressing only implementation aspects of active extraction features, it seeks to provide new perspectives for generating automaton functions for data collection, management, analysis, and representation, both for basic functionalities, and the integration extraction, and transformation of geospatial objects [19].

2.4. Weka Applied to the Spatial and Geographic Context

Through the work environment for machine learning Weka, it is possible to perform various analyses on a specific data set, or on several sets, provided that these sets have the format in which it can perform the analyses. In this regard, several discussions about GDPM (Geographic Data Preprocessing Module) [23,42,43] arise, extending the Weka Data Mining Toolkit to support geographic data. Additionally, [22] presents discussions of geographic data integration techniques, for example, the ID3, C4.5, and C5.0 algorithms for rule generation. Improvements with Weka-3.9.3 (2019), operating through MOA (Massive

Online Analysis) different types of datasets available at <http://moa.cms.waikato.ac.nz/downloads> (accessed on 20 January 2022).

That is, types such as JSON, XML, SHP, DAT, TXT, CSV, PostgreSQL/PostGIS tables, MySQL/MyGIS, ARFF and XRRF, among others, must necessarily be evaluated and formatted for machine learning processes. The discretization process of Weka is another interesting way to enhance the processes and information extraction, considering the geospatial representation. However, it is necessary to improve experiments for the automation of many of the data transformation tasks for the generation of information with higher quality [44,45].

2.5. Deep Learning and Object Detection

The main problem that arises in the processes of acquiring knowledge from images is that of relating the images collected by satellites or drones to object detection systems and the corresponding verification of the same within the databases of local systems. One approach may be to follow the advances in the machine learning algorithm literature, with a focus on using Deep Learning (DL), which is a class of Machine Learning algorithms. This type of algorithm uses multiple layers to progressively extract features from the input images [46].

DL-based approaches are efficient when large datasets are available. The word deep specifies more layers and deep neural networks. DL uses nonlinear functions. Thanks to deep learning, Intelligent Document Processing (IDP) is able to combine various AI technologies not only to automatically classify photographs, but also to describe the different elements of images. Deep learning models, with their multi-level structures, are very useful for extracting complicated information from input images. Convolutional Neural Networks (CNN) are also able to dramatically reduce computational time by leveraging the GPU for computation, something that many networks do not utilize. In the field of object identification in images, two methods stand out: regional proposal algorithms and regression object detection algorithms.

The first method is to discover in advance the possible target locations to be detected in the picture. This can ensure that the highest retrieval rate is maintained when fewer windows are selected. Suppose an image is input and, after a series of convolutions and backbone clustering, a feature map of size $M \times N$ is obtained, which corresponds to the division of the original image into areas $M \times N$. The center of each area of the original image is represented by the coordinates of a pixel in this feature map.

Region Proposition Algorithms are used to determine whether the k anchor boxes corresponding to each pixel contain a target. The network must learn to classify the anchor boxes as background or foreground. From this, it must calculate regression coefficients to modify the position, width and height of the foreground anchor box. Within these classifiers, we find algorithms such as R-CNN [47], Fast R-CNN [48], Faster R-CNN [49] and MASK-CNN [50]. Of the algorithms mentioned above, Mask R-CNN stands out. This algorithm is an extension of Faster R-CNN and works by adding a branch to predict an object mask in parallel with the existing branch for bounding box recognition. The key element of R-CNN Mask is pixel-to-pixel alignment, which is the main missing piece in Fast/Faster R-CNN. The R-CNN mask adopts the same two-phase procedure with an identical first phase (which is RPN). In the second phase, in parallel with class prediction and box clearing, Mask R-CNN also produces a binary mask for each RoI. This is in contrast to more recent systems, where classification depends on mask predictions. Furthermore, Mask R-CNN is simple to implement and train thanks to the faster R-CNN framework, which facilitates a wide range of flexible architecture designs. Furthermore, the mask branch only adds a small computational overhead, allowing for a fast system and rapid experimentation. Figure 2 shows a visual example of the segmentation performed by the algorithm.

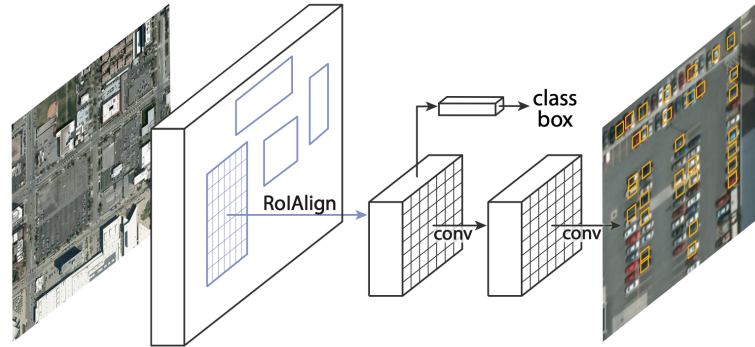


Figure 2. Mask R-CNN framework.

The above algorithms use detection as a classification problem, i.e., object proposals are first generated and then these proposals are sent to the classification/regression regions. However, some methods approach detection as a regression problem based on a similar operation. Within this field, the YOLO (You Only Look Once) and SSD (Single Shot Detector) algorithms stand out. The SSD algorithm [51] strikes a good balance between speed and accuracy. SSD runs a convolutional network on the input image only once and computes a feature map. It then runs a small 3×3 convolutional kernel on this feature map to predict bounding boxes and classification probability. SSD also uses anchor boxes in various aspect ratios, similar to Faster-RCNN, and learns the offset instead of learning the box. To handle scale, SSD predicts bounding boxes after multiple convolutional layers. Since each convolutional layer operates at a different scale, it is able to detect objects of various scales. An example of how the SSD algorithm works can be seen in Figure 3.

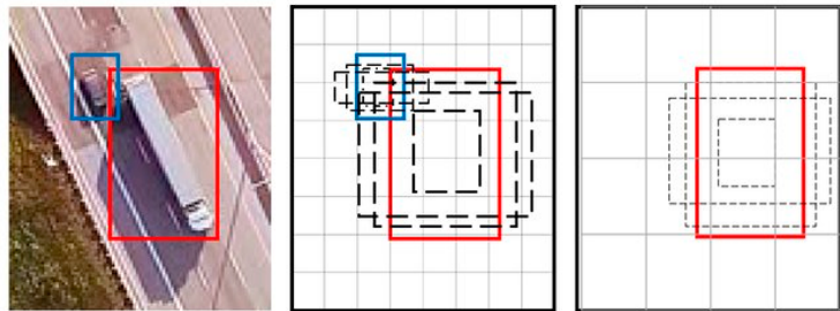


Figure 3. SSD framework.

For YOLO [52], detection is a simple regression problem that takes an input image and learns the class probabilities along with the coordinates of the bounding box. YOLO divides each image into an $S \times S$ grid, and each grid predicts N bounding boxes along with their confidence. The confidence reflects the accuracy of the bounding box and whether the bounding box actually contains an object, regardless of the class. YOLO also predicts the classification score of each bounding box for each class in the training. It can combine both classes to calculate the probability that each class is present in a predicted box. Thus, a total of $S \times S \times N$ bounding boxes are predicted. However, most of these boxes have low confidence scores, so if we set a threshold, for example of 30% confidence, we can eliminate most of them, as shown in Figure 4.

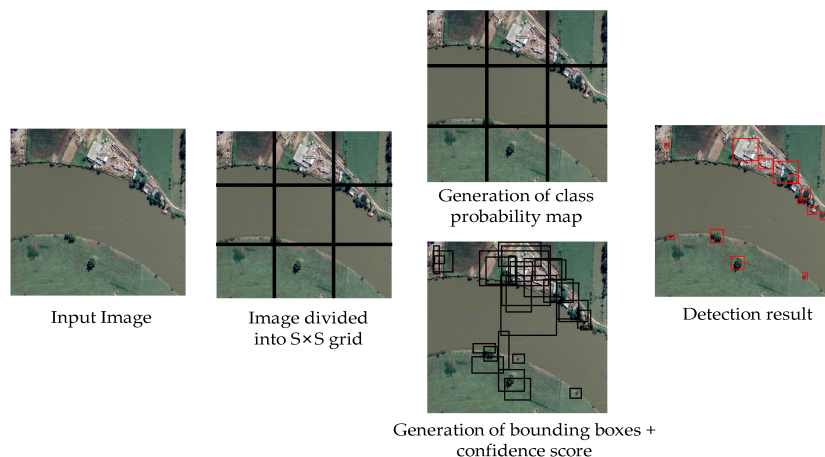


Figure 4. Yolo framework example.

YOLO is a faster algorithm than all other detection algorithms, allowing it to be run in real time. Another key difference is that YOLO sees the entire image at once, rather than looking only at the proposals of a region generated in previous methods. Thus, this contextual information helps to avoid false positives. However, one of the limitations of YOLO is that it only predicts one type of class in a grid, so it has difficulties with very small objects. There are several versions of YOLO such as YOLOv2 [53], YOLOv3 [54], YOLOv4 [55], YOLOv4-tiny [56,57], YOLO-Fine [58] and recently YOLOv7 [59]. There are also available versions of YOLO applied to Satellite Imagery, such as YOLT [60] and MRFF-YOLO [61].

3. Methodology

From the concepts of DSR (Design Science Research) proposed by [21], whose study considers it essential to also deepen the area of management. In this context, according to [62], hierarchies are applied for knowledge-intensive tasks on each identified problem. Added to the discussions of [24,63–66] allied to the classification methods being divided according to the processing, into visual or digital, known as supervised, unsupervised and hybrid as per [36].

Additionally, observing the metrics, in parametric or non-parametric and according to the approach by pixel or by regions (objects), the methodology proposed in the work aims, from the survey of satellite images and/or images obtained by RPAS and also by crossing previously shared textual data, to identify through active learning the recognition of geospatial objects with the generation of elementary rules. For this, an architecture for systematic detection and extraction supported by machine learning is proposed, see Figure 5.

In stage 1 meetings, interviews, surveys for questionnaires implementation, documentation for support, and a survey of the prerequisites of the required project are planned. The installation, testing, and homologation processes are also planned. This is where different work platforms are made available for individual or collaborative use (groupware) for integration and standardization. In stage 2, the requirements engineering processes are carried out with the construction of artifacts using UML (Unified Modeling Language). The important delimitation of the coverage area also takes place with the objective of project execution. Data acquisition processes, images, and related documents. The cataloging of data with centralized and shared storage. The processes of treatment, qualification, and homologation of the collected data with due certification. In step 3, acquisitions are made, such as contracting satellite image collection services with specific parameters. Scene

processing for example (CBERS) and (LANDSAT-8). Definition of scenes imaged by the satellites through date parameters, bands, and other relevant details, and also the survey and integration of demands with the definition of the goals.

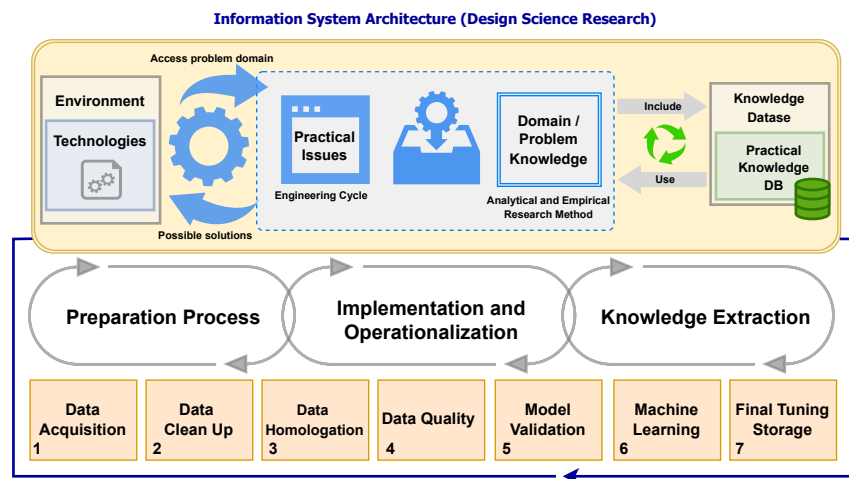


Figure 5. Architecture of the ecosystem adopted for active detection, extraction and learning of geospatial objects.

Step 4 is homologation and image processing, with the choice of the contrast method chosen for the visualization of the objects to be evaluated. Optionally, the methods can vary; for example, the linear method, histogram equalization, square contrast, square root contrast, log contrast, mean and standard deviation, decorrelation enhancement, cumulative 2% enhancement, composition and decomposition method, arithmetic operations of image bands with NDBI (Normalized Difference Built-in Index), fusion method, and image segmentation method.

In step 5, the homologation of each processing generated by the choice and application of the methods is subsequently performed, the indexes are prepared, and the methods for detecting and extracting spatial or textual objects are made specifically available. The data structures generated in the previous step are necessarily reused, according to cataloging by date, time, function, data sources, and coverage regions, enabling the import and centralized integration for sharing, through specific infrastructure for networks and sensors.

In step 6, specifically different algorithms for adaptive rule generation are evaluated. Adaptive rules are statistical patterns detected for representing the analyzed dataset. This made it possible to combine them with the intersection of new attributes already stored in the repository. In step 7, the optimization of the data structures for the repository and subsequent reverse engineering is a priority. From known rules, it is possible to actively generate learning about the experiences with the availability of large volumes of data to support the other decision-support processes. To synchronize each step and process, an infrastructure [67–69] computer network for remote communication between various devices, data collectors, cameras, and sensors is implemented.

Furthermore, through the fruit of several research discussions comes the development of the model for active knowledge extraction, presented by [45], intensifying the interoperability of the data and the advancement of the implementation of the concepts in this work. It also aims at improving the functions through the prototype as presented in Figure 6. From the application architecture idealized by [22], the same also provides the opportunity for the derivation of new experiments for machine learning, since the collection of data, cataloging, discretization of data and application of algorithms is of great importance for the detailing of each process and recording of operations for possible control.

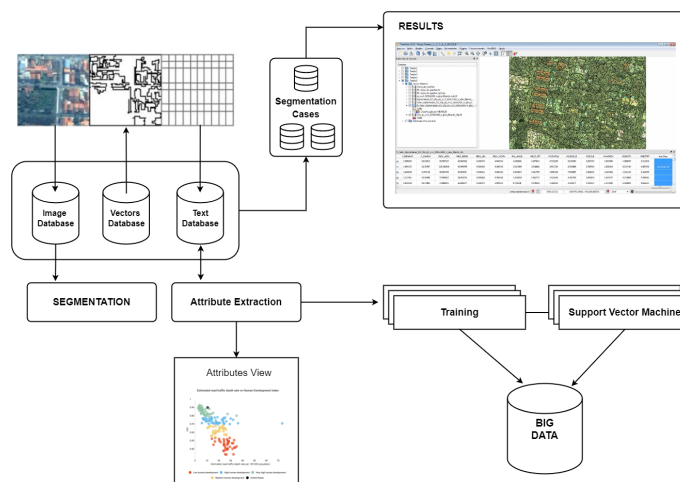


Figure 6. Adapted architecture for systematization and generation of recommendation cases.

This defines a process for detecting geographic patterns, urban and rural devices through segmentation and other records for automatic observation and evaluation of territorial expansion from the extraction of shared knowledge.

One of the main reasons is the difficulty of constantly processing a large volume of data, due to satellite images collected periodically and which can present files with expressive sizes (megabytes or gigabytes), fundamentally important for pattern recognition.

3.1. Remote Sensing Classification

In applications that require image classification, the availability of labeled samples (training data) is closely associated with the choice the analyst will make for extracting information from the images. Two families of techniques are distinguished, called supervised and unsupervised, according to the presence or absence of labeled samples, respectively.

In the context of SR, classification is the process that seeks to assign a label to certain data described by a set of attributes. In digital terrestrial remote sensing imagery, this process is equivalent to determining, for each pixel, which category is present on the surface, such as water, soil, and forest, which is usually done by spectral attributes, such as the gray level in each band.

They are commonly used over the radiometric indices as [70], arithmetic contrast operations with NDBI, NDVI (Normalized Difference Vegetation Index), and NDWI (Normalized Difference Water Index). Considering the processes of unsupervised classification in (SR) and supervised classification, from data collected by (satellites) or unmanned aircraft, these images can be analyzed in different scenarios, whether knowledgeable or not about the observed area. The discussions of [29] describe in detail the implementation of each process that can be adapted to different experiments.

3.2. Active Training and Machine Learning

We also consider using parametric classifiers that model the decision boundaries between training classes with a fixed number of parameters, regardless of the number of samples [71,72].

It is the simplest classifier in existence and therefore ends up having a more didactic than operational role. The decision boundaries are positioned on lines equidistant between midpoints of the various classes present.

The classification process by Euclidean minimum distance is performed by Figure 7A the distribution of the sample elements of each class in two bands of a generic image and

Figure 7B the averages calculated for each sample and respective distances to a pixel to be classified.

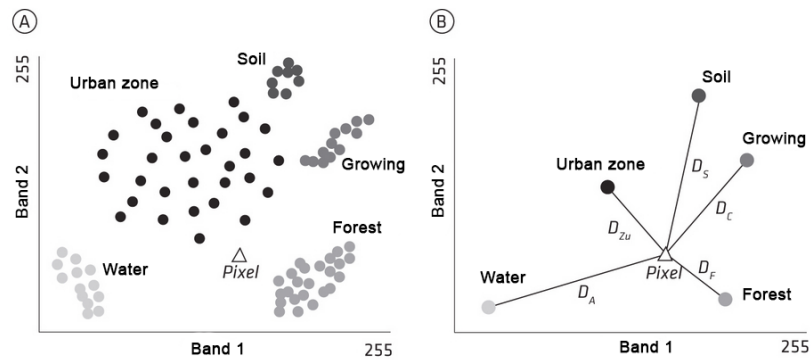


Figure 7. Representation of Euclidean distances.

Especially with the evolution of text mining techniques according to [36], the development of the StringToWordVector function can optionally assist in scanning large amounts of text using TF-IDF concepts in Weka treated IDFTransform and TFTransform.

Additionally, when searching with textual data, a statistical measure is adopted that is intended to indicate the importance of a word in a document relative to a collection of documents or in a linguistic corpus. According to [73], it is possible to distinguish the importance between different word features, and it is necessary to calculate the weights of the prominent words. For this, the TF-IDF method is implemented and used to calculate the weight according to (4).

$$TF - IDF = tf \times idf = \left(\frac{a}{t}\right) \times \log\left(\frac{b}{c + 1}\right) \quad (4)$$

In the formula, a is the frequency of the resource in the document set, this is the total number of times of all resources in the document set, b is the document number in the document set, and t is the number of documents that contain the resource. Then, with the use of the TF-IDF method, it is possible to select n resources with the maximum value of TF-IDF as per the candidate resource set. Using IDFTransform and TFTransform, scans and learning are performed on the datasets prepared for the textual data matching, as detailed in the case study of Figure 8.

This enabled the cataloging of each geographic object properly identified from the satellite image with the items found with IDF-TFTransform. In parallel, from the discussions of [74], another viable strategy to relate and spatially represent different information can be through a geographical matrix of spatial queries, being a two-dimensional representation of intrinsic relationships between locations. To exemplify, implement the forest code and limits of permanent preservation areas in each municipality. Many impasses arise, and although this is not a complex task, it requires great human effort and skilled labor for permanent monitoring. In this sense, the municipalities that make up the basin of the Itajaí-Açu River were mapped to generate the geographical matrix. Through the implementation, the integration of data was carried out, resulting in a large volume of distinct information, providing important relationships for the evolution of monitoring through the recognition of spatial objects as explored in the case study.

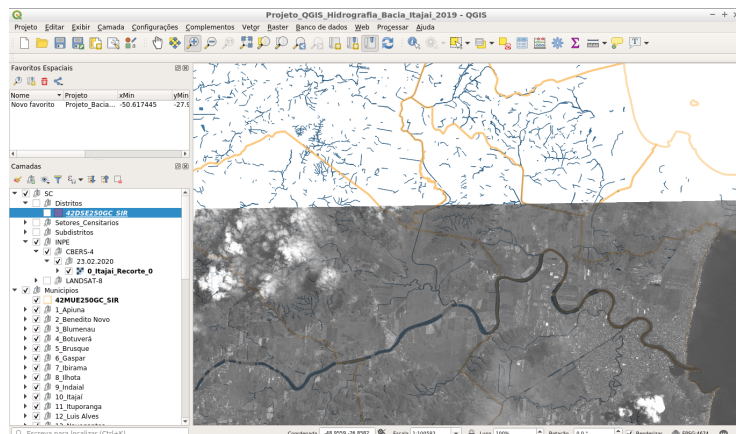


Figure 8. Integration in QGIS with CBERS4 satellite images and CTM data (Itajaí-SC) for monitoring the expansion on riverbanks.

4. Materials and Methods

4.1. Study Area

The study sites were located in the south of Brazil, in the Itajaí Municipality along the Itajaí-Açu basin river, in Santa Catarina (SC) state, Brazil (Figure 9).

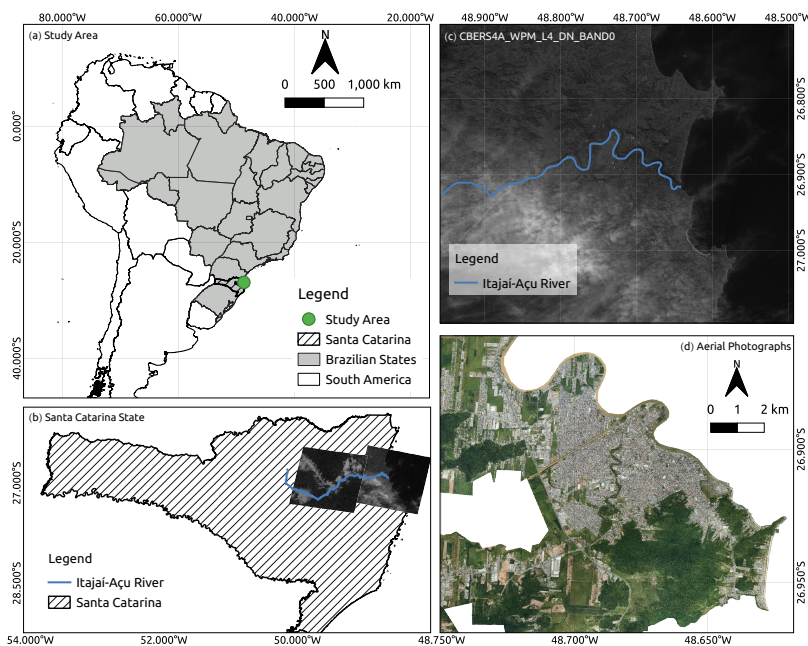


Figure 9. Location of study sites: (a) in South America, Brazil, Santa Catarina state. (b) State of Santa Catarina with the CBERS4A Panchromatic 2 m Image, including the Itajaí-Açu River layer. (c) Satellite image from the CBERS4A of the mouth of the Itajaí-Açu River (d) Airborne from the Itajaí—SC Municipality.

4.2. Materials

New challenges arise with updating a municipality's land registry. According to [73], a good cadastre contributes to the equitable distribution of tax resources, promotes property security and creates bases for urban and regional planning. For this, bases for urban and regional planning are created, using the Multifinality Technical Cadastre (*Cadastro Técnico Multifinalitário*—CTM), available at <https://geoitajai.github.io/geo/plantacadastral.html> (accessed on 20 October 2022). And from the collection of images through the CBERS-4 and CBERS-4A satellite with spatial resolution of 5 and 2 m, respectively <http://www.dgi.inpe.br/catalogo> (accessed on 20 October 2022). Additionally, the aerial photogrammetry are obtained from the portal <https://arctis.itajai.sc.gov.br/geoitajai/plantacadastral/plantacadastral.html> (accessed on 20 October 2022), they are used to identify spatial objects in the municipality of Itajaí, state of Santa Catarina—Brazil. The map layouts were generated using QGIS software [75].

4.3. Methods

With the advancement of technologies and the emergence of large volumes of data, it is possible to cross data through automated processes and solutions with greater reliability and precision. The policies are evaluated with the crossing of environmental and social data, referring to territorial occupation, and considering traditional economic, physical, and legal aspects, among others.

However, one of the limiting factors for updating base registers is still the high cost of developing the entire cartographic framework. However, with the advancement of collaborative technologies, more affordable alternatives should be considered, especially for small and medium-sized cities. In this sense, we highlight the possibility of using photogrammetric techniques, since the logistics involved in the operation of these systems are more flexible and economical when compared to cartography by topographic or geodetic techniques, or even conventional photogrammetry with UAV [76].

In the case study, image acquisition processes were carried out, and subsequently the prior selection of the delimited perimeter for application of the present study. Figure 10 presents the result of this process and the application of the linear contrast method. During the investigation, the pixel transformation functions were evaluated from the image contrast, and in the application of active object learning, the contrast presented a more suitable visual result. From the transformation function (T) for a single pixel (r = original pixel value), the resulting pixel value (S) is generated through different techniques to obtain better processing and visualization of objects. Where, $(s) = T(r)$. Table 1 shows the results with the contrast method.



Figure 10. Image of the municipality of Itajaí SC—Brazil, after the contrast method in the area chosen for object detection and active learning.

Table 1. Preparation for active learning experiments.

Satellite		Segmentation		Generated Records
CBERS Sensor Banda ^a	Contrast Method ^b	CPU Time i7 6.5 GB GNU/Linux	Similarity	Object Detection
PAN5M—Band Espec = 1	Linear	327 s 70	0.050	1,066,788
		390 s 46	0.045	1,345,492
		435 s 84	0.040	1,639,031
		464 s 94	0.035	2,158,357
		516 s 02	0.030	2,770,752
		572 s 11	0.025	3,760,484

^a Image acquisition criteria, e.g., (date, location, quality and others). ^b Software used for image processing, such as TerraLib/TerraView (INPE) and Qgis.

The lower the degree of similarity threshold, the higher the generation of objects for analysis processing and active learning. In other words, higher demand and availability of hardware resources to support higher processing volumes are essential.

Additionally, the availability of a relational database management system, such as PostgreSQL/PostGIS, is another important implementation due to the need of storage for manipulation. The case study was designed to integrate public data with the main objective of evaluating the existence of relationships, direct or indirect crossover. Different data sources were considered for experiments of the algorithms.

After two comparative implementations, over the data set (49,325 occurrences) obtained from the CTM. Where, initially in the first implementation was prioritized application of the classification method to elaborate the decision tree with ID3 algorithm, using the Weka tool (version 2). It generated 78.83% (38,885) correctly classified instances and 21.15% (10,436) incorrectly classified instances.

In the second implementation on the same dataset, a new tree structure was generated with J48 algorithm using Weka tool (version 3.9.3). After visualizing the tree, it was possible to detect the levels and the class (Conservation), with more information gain. For correctly classified instances, it was obtained 85.2002% (42025) of success and 14.79% (7300) not classified. The final kappa statistic of 76.11% also determined the classifier that obtained the best learning. After the implementations with the software Weka, for the same area investigated, another study was performed using the software TerraView/GeoDMA, where the methods of segmentation, image vectorization and later extraction of the set of attributes were applied, see Figure 10.

Subsequently, with GeoDMA for object classification, 212 attributes were generated. From the identification of the characteristics of each pixel was performed the conference of the generation of each polygon and definition of relations for characterization of basic rules of each object. For a better understanding of the objects detected in the study, it was important to evaluate the different types of scales.

5. Discussion

From the expected results are presented some reflections and discussions about the development of the work, initially idealized and later obtained. That is, the results achieved by the research and their implementation are presented.

Reflection 1: Through the studies presented, was it possible to detect the applicability of knowledge extraction in CTM in conjunction with other areas? Clearly, and as is proven through the case study presented in Figures 5 and 6, along with a comparison of the Weka software in version 3.2 (2001) and version 3.9.3 (2019), both versions allowed for obtaining models, enabling machine learning, and expansion of analysis with new processes of data integration and information extraction.

Reflection 2: Specifically in this paper, new performance tests were presented with the Weka classifiers: OneR, IBk, NaiveBayes, and J48. All classifiers use the same CTM data and under the same conditions. That is, at this point it is worth mentioning the inclusion of analysis with the Weka “Hoeffding Tree” classifier, allowing the generation of a tree with less criticality.

Reflection 3: Was it possible to use new satellite images and run GeoDMA for automatic learning of new objects? Yes, in this regard, it is worth highlighting the specifications adopted according to Table 1, enabling the correct acquisition of images through standards for the next stage of segmentation, then the vectorization of objects. In particular, several procedures are performed for image segmentation, which not part of the scope of this work, but will be detailed with new experiments.

Reflection 4: Did the development and discussion of the experiments occur with other CTM databases integrated for the recommendation? Partially, some experiments use static datasets successfully, but have not been evaluated by mining continuous stream data using Weka-MOA.

Reflection 5: Were performance evaluations of the Weka software performed? Yes, exhaustive performance experiments were conducted, as shown in Table 2 and made available for online access. In all classifiers, the CTM dataset was analyzed with the same parametrizations and specifications, such as the 10-fold cross-validation over the “property conservation” class. The Weka classifiers performed well on a dataset containing 18 attributes and 49,325 instances. The OneR, NaiveBayes, J48, IBk, and Hoeffding Tree classifiers showed satisfactory results. The results of the classifiers are explained one by one below:

For Weka → OneR: OneR processing took 0.09 s to build and run the model. Additionally, with the application of OneR, a correct classification of the instances of (38,404) records was obtained totaling an accuracy of 77.8591% by the algorithm. It also presented satisfactory learning with 63.75% evaluated by the kappa statistic.

For Weka → NaiveBayes: Initially, the processing took 0.13 s to build and run the model. Additionally, with the application of NaiveBayes, a correct classification of the instances of (37,445) records was obtained, totaling an accuracy of 75.9149% by the algorithm. It also presented satisfactory learning with 62.48% evaluated by the kappa statistic.

For Weka → J48: The processing took 2.45 s to build and run the model. Additionally, with an application of J48, a correct classification of the instances of (42,025) records was obtained, totaling an accuracy of 85.2002% by the algorithm. It also showed satisfactory learning with 76.11% evaluated by the kappa statistic.

For Weka → IBk: Processing took 3 min 34 s to build and run the model. Additionally, with an application of IBk, a correct classification of the instances of (40,042) records was obtained, totaling an accuracy of 81.1799% by the algorithm. It also showed satisfactory learning with 69.97% evaluated by the kappa statistic.

For Weka → Hoeffding Tree: Processing took 0.69 s to build and run the model. Additionally, with an application of the Hoeffding Tree, a correct classification of the instances of (39,203) records was obtained, totaling an accuracy of 79.48% by the algorithm. It also presented satisfactory learning with 66.54% evaluated by the kappa statistic.

Experiments with Weka → IBk from the vector with 25 attributes and containing (1,000,000) instances, generated after image segmentation Weka with the IBk classifier built the classification model quickly, but presented a very large slowness (9 h) to measure the distances of all instances. However, the alternative found to speed up the processing was to retain in memory only a “window” of instances, instead of the complete dataset. In Weka, the default parameter “window size = 0” allows you to set the maximum number of instances allowed in the training pool, and adding additional instances simply removes the old ones, freeing up memory to improve performance.

In addition, for a better understanding, analysis was performed on the datasets below, being separated into three different sets to initially compare the training with 100 (objects), 1000 (objects), and 15,000 respective training results, it was possible to decide which

algorithm to use first to be prioritized and adopted for further processing, as shown in Table 2, containing the results obtained from training the different datasets and compared with the performance of processing the entire set of objects. The final test was evaluated with a total of 49,325 instances, and these instances are isolated from the training dataset.

Table 2. Training datasets (objects).

Algorithm	Training (100 Objects) %	Training (1000 Objects) %	Training (15,000 Objects) %	Performance Evaluation (Final) %
J48	52	77.9	80.4067	85.2002
IBk	78	78.2	80.4467	81.1799
Hoeffding Tree	76	77.1	80.3733	79.4800
OneR	75	77.9	80.3600	77.8591
NaiveBayes	76	77.1	79.4133	75.9149

Training datasets and evaluation/tests are available in: <http://sadpreaigeo.org/ufsc-egc/mtec2022/> (accessed on 15 December 2022).

With Weka → Hoeffding Tree, a Hoeffding tree (VFDT—Very Fast Decision Trees) is a very fast decision tree algorithm for incremental decision tree induction at any time, capable of learning from massive data streams, assuming the distribution generation instances do not change over time. Hoeffding trees exploit the fact that a small sample can be sufficient to choose an optimal splitting attribute. This idea is supported mathematically by the Hoeffding limit, which quantifies the number of observations (examples) needed to estimate some statistics within a prescribed precision (according to the goodness of an attribute). A theoretically attractive feature of Hoeffding Trees not shared by other additional decision tree learners is that it has good performance guarantees. Using the Hoeffding boundary, one can show that its output is asymptotically nearly identical to that of a non-incremental study using infinite examples proposed by [41]. This classifier is a successful reference in dealing with large spatial representation datasets, for example, the evaluated dataset (Weka → ConvtypNom), regarding spatial coverage of forests with quadrants defined in 30 × 30 m, 581,012 instances, and 54 attributes, elaborated the model in 41 s and completed the evaluation processing with cross-validation 10 times in 6 min and 49 s. All the results of Weka 3.9.3 performance are presented in Table 3.

Table 3. Performance comparison of Weka 3.9.3.

Nr	Weka 3.9.3 Classifiers	Time	Correct Instances	% Hits	% Kappa
1°	J48	0.30 s	42.0250	85.2002%	76.11%
2°	IBk	3 m 34 s	40.0420	81.1799%	69.97%
3°	Hoeffding Tree	0.69 s	39.2030	79.4800%	66.54%
4°	OneR	0.03 s	38.4040	77.8591%	63.75%
5°	NaiveBayes	0.04 s	37.4450	75.9149%	62.48%

Results available in: <http://sadpreaigeo.org/ufsc-egc/mtec2019/> (accessed on 20 January 2022).

In all classifiers, the CTM dataset was analyzed with equal parameterizations and with the same specifications, such as the “k = 10 cross-validation” on the “property conservation” class. For property conservation in the “good condition” category, from the confusion matrix generated by the first classifier established, the true positives (TP) with 22,279 units, and the true negatives (VN) with 19,746, totaling 42,025 units to be certified, were first identified. Afterward, the false positives (FP) with 4002 units enabled the separation for re-evaluation of each occurrence. However, unlike the IBk classifier, the other evalu-

ated classifiers had good performance regarding processing time using 18 attributes and 49,325 instances, over the same computational infrastructure provided.

With the satisfactory results obtained with OneR, NaiveBayes, J48, IBk, and Hoeffding Tree, allows the use of GeoDMA for automatic learning of new objects to be positively proven by Figure 10. Additionally, the fusion between textual classifiers and geospatial classifiers made possible through this work, the verification of an innovative form of knowledge extraction engineering.

From the use of GeoGMA [22], to perform the extraction of attributes, after exhaustive performance tests on the hardware used Table 1, adjustments were applied opting for the selection of all statistical methods, except “Percent of each class by area”, because this method increases the consumption of processing and memory. Still in the process of extraction of attributes, it was possible from this procedure to obtain a better response time with a duration of up to 15 min of processing load.

Specifically, in Figure 11, it is shown how this made it possible to start the elaboration of queries through the filter on the attribute “B0Mean” > 0.4, especially to obtain the selection of objects with the highest “vegetation index” in the image.

In the second query, it was possible through the filter on the attribute “B7Mean” > 300.0, to obtain the selection of objects with the “shadow” characterization on the image. In this particular case, the query changed the return color for the objects. However, it was not rendered after processing, changing the color parameter set as “Yellow” to “Green”, but remaining the best identification of the color “yellow” for the recognition of the object “shadow” in the image.

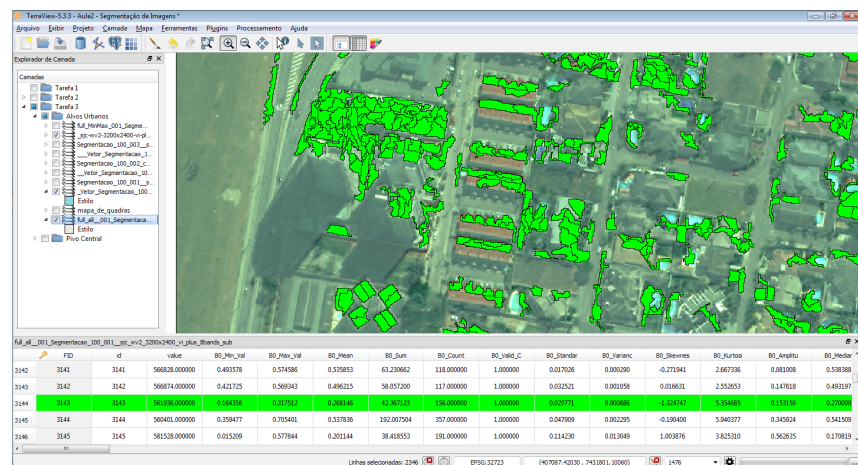


Figure 11. Results after calibration of values and recognition of objects.

Still, for the definition and characterization of the objects, some queries were performed, allowing, through the filter on the attribute “Band 5-Mean” using the GeoDMA [22], us to obtain different values for the selection of objects and characterization, such as “ceramic roofs”, as shown in Figure 12. Thus, for the tested objects, the following values were obtained and are available in Table 4.

Table 4. Obtained values.

ID	Min	Max	Mean
2128	386.000000	511.000000	426.769231
2135	241.000000	724.000000	425.036585
2240	235.000000	631.000000	532.342541
2302	326.000000	692.000000	531.554622
2398	27.000000	556.000000	355.121339
2999	310.000000	664.000000	551.215962
3056	385.000000	645.000000	553.728000
3075	460.000000	650.000000	596.877160
3116	260.000000	526.000000	426.401042
3144	372.000000	529.000000	436.227848
3732	218.000000	638.000000	425.622951
3767	270.000000	521.000000	391.483974
3768	416.000000	584.000000	529.207207
3867	229.000000	555.000000	398.095023

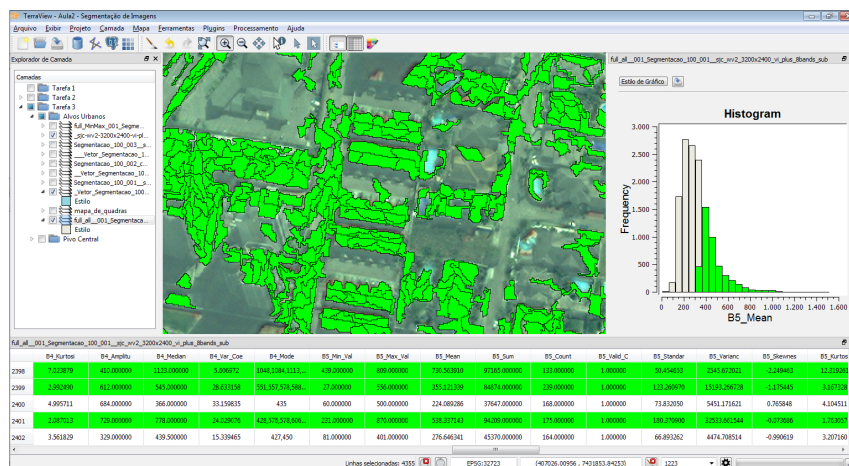


Figure 12. Results after calibration of values and recognition of objects with the results in the histogram.

The choice of samples was made randomly within the dataset images used for training, aiming initially to understand how the learning was performed, especially with the use of GeoDMA, as in Figures 13 and 14. This provided an important experience in the choice of classes. Specifically, the classes of investigation were: (a) asphalt, (b) roofs—light, dark or ceramic, (c) swimming pools, (d) shadows, (e) exposed soil, and (f) vegetation. Figure 13 describes the calibration process to obtain the final result, detailed in Figure 14.

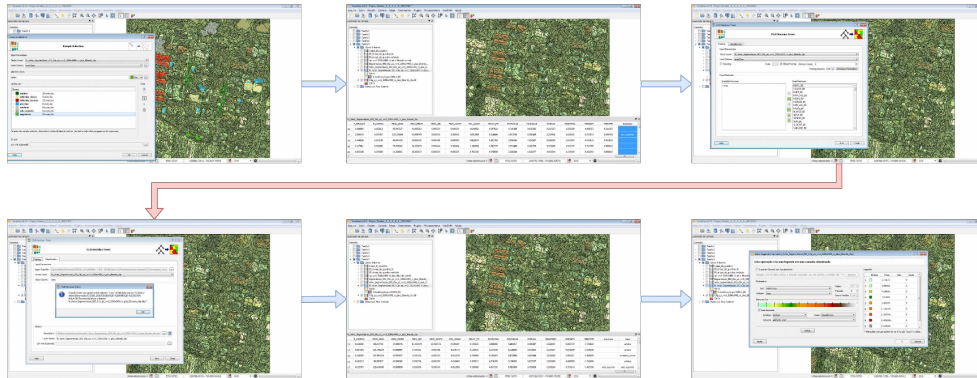


Figure 13. Calibration for segmentation process.

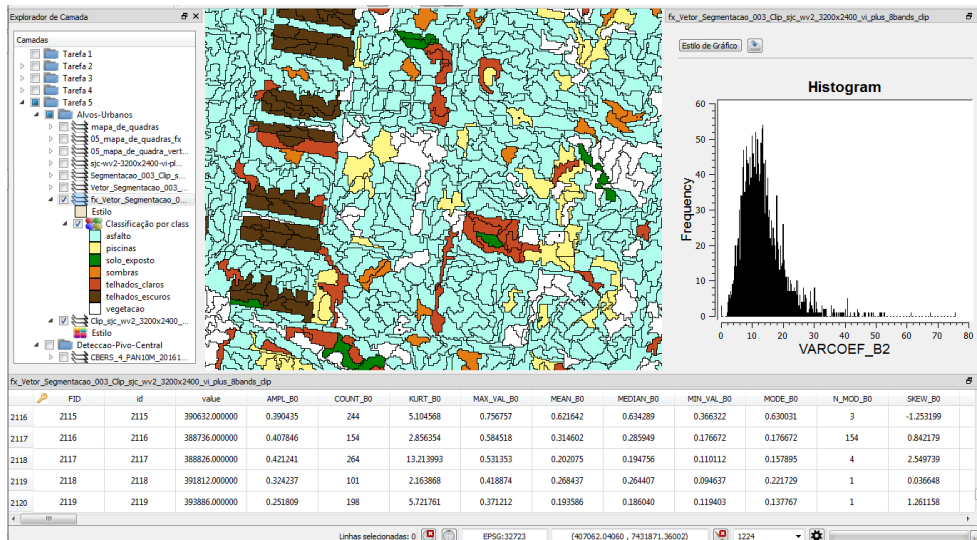


Figure 14. Segmentation results of the recognized objects.

From this study, it was possible to carry out new experiments to acquire a larger number of samples and improve the rule generation. For example, with respect to the class “asphalt”, the rule was very broad, in which a greater number of representations were obtained, as illustrated in Figure 14. On the other hand, for “dark ceramic roofs”, the rule was generated with better results. However, for future work, new implementations will be developed, aiming to train new classes.

Certainly, much still has to evolve computationally, especially with regard to constant active machine learning. In particular, urban object recognition and monitoring is noted. All documents related to the Weka training, datasets and other complementary documents are available at <http://sadpreaigeo.org/ufsc-egc/mtec2019> and <http://sadpreaigeo.org/ufsc-egc/mtec2022> (accessed on 15 December 2022).

6. Conclusions

Finally, according to the initial objectives of this work, through extensive research, it was possible to prove the applicability of the extraction of knowledge with the integration of data collected from the Cadastro Técnico Multifinalitário (CTM). Additionally, as

prioritized, the research contemplated through investigation of the publications made in the last 5 years. During the study, it was noted that there is a great involvement of the academic and scientific community in the development of technologies that understand the geospatial and earth phenomena.

Growth and strong trends were observed in the use of the SVM (Support Vector Machine) method for evaluating large volumes of textual and geospatial data, as was the use of data discretization to enhance the execution (performance) of classifiers (algorithms). In this sense, the possibility of implementation and integration of the software Weka with TerraView/GeoDMA was proven, and they were compatible and operationalized because both complement each other from the collection to the structuring of textual and geospatial data for the evaluation of datasets, as presented in the Section 5.

An important characteristic to highlight with the Weka software was experimentation with vectors that presented only numeric values, in which linear regression (Weka 3.9.3) proved to be faster to deal with large volumes of data. Regarding linear regression, the J48 decision tree showed the best results with the best classifier with accuracy (85%) and the kappa agreement coefficient (76%) in an average time of 0.30 s.

For future work, further studies to advance experiments from evolving data streams—those generated by mechanisms that change or fluctuate over time, by implementing the Weka/MOA package, designed specifically for data stream mining including new adaptations with Deep Learning algorithms. Furthermore, we intend to advance the development of a module for a pre-processing face as proposed [77], prioritizing data collection, transformation, and preparation of datasets and images. This is essential for the crossing of data and construction of rules to ensure the quality of the information to the user and decision maker. Another need concerns the improvement of models, being more or less robust and that can be reusable through vectors for TF-IDF x TerraViewGeoDMA application.

Although great demands of work are generated, one must prioritize care for the quality of data and information both for technical and operational issues, as well as for strategic issues that require constant validations, allowing certifications to occur for each step performed during all the extraction processes, generating a reliability indicator for the quality of information. The continuity of actions to intensify the implementation of processes for information quality, in this work, is indispensable so that all stages of knowledge extraction are guaranteed and certified.

Importantly, our results provide evidence for an implementation of an innovative practical, and systematic approach. The extraction of information and recommendation of knowledge shows a greater scientific relevance. Allowing the methods presented to apply calibration parameters for each object and achieve results with greater accuracy.

Author Contributions: Conceptualization, L.A.S. and A.F.d.M.; methodology, L.A.S. and A.F.d.M.; software, L.A.S., A.S.M. and H.S.S.B.; validation, L.A.S. and A.F.d.M.; formal analysis, L.A.S.; investigation, L.A.S. and A.F.d.M.; resources, L.A.S. and A.F.d.M.; data curation, L.A.S. and A.F.d.M.; writing—original draft preparation, L.A.S. and A.F.d.M.; writing—review and editing, L.A.S., A.S.M., H.S.S.B., A.L.G., L.C.B. and A.F.d.M.; visualization, L.A.S. and A.F.d.M.; supervision, A.F.d.M.; project administration, A.F.d.M.; funding acquisition, L.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research of Luis Augusto Silva has been funded by the call for predoctoral contracts USAL 2021, co-financed by Banco Santander. Héctor Sánchez San Blas's research was supported by the Spanish Ministry of Universities (FPU Fellowship under Grant FPU20/03014).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shankaranarayanan, G.; Blake, R. From Content to Context: The Evolution and Growth of Data Quality Research. *J. Data Inf. Qual.* **2017**, *8*, 9. [[CrossRef](#)]
2. Li, D.; Gao, X.; Deng, Y. A Generalized Expression for Information Quality of Basic Probability Assignment. *IEEE Access* **2019**, *7*, 174734–174739. [[CrossRef](#)]

3. Lin, G.; Jing, G.; Fang-Fang, D. Evaluation method of enterprise information quality based on QFD. In Proceedings of the 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), Xianning, China, 16–18 April 2011; pp. 325–328.
4. Tseng, Y.H.; Lin, C.J.; Lin, Y.I. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [[CrossRef](#)]
5. Madnick, S.E.; Wang, R.Y.; Lee, Y.W.; Zhu, H. Overview and Framework for Data and Information Quality Research. *J. Data Inf. Qual.* **2009**, *1*, 2. [[CrossRef](#)]
6. Naumann, F.; Rolker, C. Assessment Methods for Information Quality Criteria. Mathematisch-Naturwissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Berlin, Germany, 2005. . [[CrossRef](#)]
7. Wang, R.Y.; Strong, D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. *J. Manage. Inf. Syst.* **1996**, *12*, 5–33. [[CrossRef](#)]
8. Mahdianpari, M.; Granger, J.E.; Mohammadimanesh, F.; Warren, S.; Puestow, T.; Salehi, B.; Brisco, B. Smart solutions for smart cities: Urban wetland mapping using very-high resolution satellite imagery and airborne LiDAR data in the City of St. John's, NL, Canada. *J. Environ. Manag.* **2021**, *280*, 111676. . [[CrossRef](#)] [[PubMed](#)]
9. Hu, Q.; Wu, W.; Xia, T.; Yu, Q.; Yang, P.; Li, Z.; Song, Q. Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping. *Remote Sens.* **2013**, *5*, 6026–6042. [[CrossRef](#)]
10. Fawcett, D.; Bennie, J.; Anderson, K. Monitoring spring phenology of individual tree crowns using drone-acquired NDVI data. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 227–244. [[CrossRef](#)]
11. Mahdavi Saeidi, A.; Babaie Kafaky, S.; Mataji, A. Detecting the development stages of natural forests in northern Iran with different algorithms and high-resolution data from GeoEye-1. *Environ. Monit. Assess.* **2020**, *192*, 1–15. [[CrossRef](#)]
12. Alphan, H.; Çelik, N. Monitoring changes in landscape pattern: Use of Ikonos and Quickbird images. *Environ. Monit. Assess.* **2016**, *188*, 1–13. [[CrossRef](#)]
13. Albuquerque, R.; Vieira, D.; Ferreira, M.; Soares, L.; Olsen, S.; Araujo, L.; Vicente, L.; Tymus, J.; Balieiro, C.; Matsumoto, M.; et al. Mapping Key Indicators of Forest Restoration in the Amazon Using a Low-Cost Drone and Artificial Intelligence. *Remote Sens.* **2022**, *14*, 830. [[CrossRef](#)]
14. Habibie, M.I.; Ahamed, T.; Noguchi, R.; Matsushita, S. Deep Learning Algorithms to determine Drought prone Areas Using Remote Sensing and GIS. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS), Jakarta, Indonesia, 7–8 December 2020; pp. 69–73. [[CrossRef](#)]
15. Hosseiny, B.; Mahdianpari, M.; Brisco, B.; Mohammadimanesh, F.; Salehi, B. WetNet: A Spatial–Temporal Ensemble Deep Learning Model for Wetland Classification Using Sentinel-1 and Sentinel-2. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
16. Klein, V.B.; Todesco, J.L. Um Modelo Conceitual para uso de Big Data e Open Data nas Smart Cities. *Intern. J. Knowl. Eng. Manag.* **2017**, *6*, 28–50. [[CrossRef](#)]
17. Wagstaff, K.L. Machine Learning that Matters. In Proceedings of the International Conference on Machine Learning, Edinburgh, UK, 26 June–1 July 2012; pp. 529–536.
18. Souza, J.; Francisco, A.; Piekarski, C.; Prado, G.; Oliveira, L. Data Mining and Machine Learning in the Context of Sustainable Evaluation: A Literature Review. *IEEE Lat. Am. Trans.* **2019**, *17*, 372–382. [[CrossRef](#)]
19. Mei, J.; Jiang, X.; Cai, J. Learning local feature representation from matching, clustering and spatial transform. *J. Vis. Commun. Image Represent.* **2019**, *63*, 102601. [[CrossRef](#)]
20. Gonçalves, V.P.; Ribeiro, E.A.W.; Imai, N.N. Mapping Areas Invaded by Pinus sp. from Geographic Object-Based Image Analysis (GEOBIA) Applied on RPAS (Drone) Color Images. *Remote Sens.* **2020**, *12*, 2805. [[CrossRef](#)]
21. Bayazit, N. Investigating Design: A Review of Forty Years of Design Research. *Des. Issues* **2004**, *20*, 16–29. [[CrossRef](#)]
22. Korting, T.S. GEODMA: A Toolbox Integrating Data Mining with Object-Based and Multi-Temporal Analysis of Satellite Remotely Sensed Imagery. Ph.D. Thesis, INPE, Sao Jose dos Campos, Brazil, 2012.
23. Korting, T.S.; Fonseca, L.M.G.; Camara, G. GeoDMA-Geographic Data Mining Analyst. *J. Comput. Geosci.* **2013**, *57*, 133–145. [[CrossRef](#)]
24. Maretto, R.V.; Korting, T.S.; Fonseca, L.M.G. An Extensible and Easy-to-use Toolbox for Deep Learning Based Analysis of Remote Sensing Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 17–22 July 2019; pp. 9815–9818.
25. Braz, A.M.; Korting, T.S.; Martins, A.P.; Braz, A.M. Geobias E Mineração de Dados para Classificação de Imagens de Altíssima Resolução Espacial. *Estud. Geográficos Rev. Eletrônica Geogr.* **2021**, *19*, 209–224. [[CrossRef](#)]
26. Kanjir, U.; Veljanovski, T.; Oštir, K. Multilevel Segmentation of Soil Sealing in Urban Areas. In Proceedings of the Asian Conference on Remote Sensing (ACRS), Taipei, Taiwan, 3–7 October 2011; pp. 1–7.
27. Liu, W.T.H. Aplicacoes de Sensoriamento Remoto. In *Plastics*; Oficina de Textos: Rio de Janeiro, Brazil, 2019; pp. 10–100.
28. Rajbhandari, S.; Aryal, J.; Osborn, J.; Lucieer, A.; Musk, R. Leveraging Machine Learning to Extend Ontology-Driven Geographic Object-Based Image Analysis (O-GEOBIA): A Case Study in Forest-Type Mapping. *Remote Sens.* **2019**, *11*, 503. [[CrossRef](#)]
29. Zanotta, D.C.; Ferreira, M.P.; Zortea, M. Processamento de Imagens de Satélite. In *Plastics*; Blucher: São Paulo, Brasil, 2019; pp. 107–301.
30. Barbosa, A.L.; Loureiro, G.; Manea, S.; Duarte, J.M.L.; Garbi, G.P. Ranking of Fault Mitigation Techniques for Spatial Radiation in Commercial Off-the-Shelf Field Programmable Gate Array. *IEEE Lat. Am. Trans.* **2020**, *18*, 736–743. [[CrossRef](#)]

31. Obando, E.D.; Carvajal, S.X.; Agudelo, J.P. Solar Radiation Prediction Using Machine Learning Techniques: A Review. *IEEE Lat. Am. Trans.* **2019**, *17*, 684–697. [[CrossRef](#)]
32. Pletsch, M.A.; Körting, T. Information mining for automatic search in remote sensing image catalogs. *Rev. Bras. Cartogr.* **2019**, *70*, 1860–1884. [[CrossRef](#)]
33. Cesconetto, J.; Augusto Silva, L.; Bortoluzzi, F.; Navarro-Cáceres, M.; Zeferino, C.A.; Leithardt, V.R.Q. PRIPRO—Privacy Profiles: User Profiling Management for Smart Environments. *Electronics* **2020**, *9*, 1519. [[CrossRef](#)]
34. de Moraes Rossetto, A.G.; Segal, C.; Leithardt, V.R.Q. An Architecture for Managing Data Privacy in Healthcare with Blockchain. *Sensors* **2022**, *22*, 8292. [[CrossRef](#)]
35. Lima, R.; Filippetto, A.S.; Heckler, W.; Barbosa, J.L.V.; Leithardt, V.R.Q. Towards ubiquitous requirements engineering through recommendations based on context histories. *PeerJ Comput. Sci.* **2022**, *8*, 794. [[CrossRef](#)]
36. Mehta, R.; Kumar, A. An Analysis of Hybrid Layered Classification Algorithms for Object Recognition. *IOSR J. Comput. Eng.* **2018**, *20*, 57–64. [[CrossRef](#)]
37. Yahav, I.; Shehory, O.; Schwartz, D. Comments Mining with TF-IDF: The Inherent Bias and Its Removal. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 437–450. [[CrossRef](#)]
38. Khorshid, M.M.H.; Abou-El-Enien, T.H.M.; Soliman, G.M.A. A Comparison among Support Vector Machine and other Machine Learning Classification Algorithms. *Int. J. Comput. Sci.* **2015**, *3*, 25–35.
39. Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; Valen, D.V. Deep learning for cellular image analysis. *Nat. Methods* **2019**, *16*, 1233. [[CrossRef](#)]
40. Giolo, S.R. Introdução à análise de dados categóricos com aplicações. In *Plastics*; Blucher: São Paulo, Brasil, 2017; pp. 75–190.
41. Hulten, G.; Spencer, L.; Domingos, P. Mining time-changing data streams. In Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining SIGKDD, San Francisco, CA, USA, 26–29 August 2001; pp. 97–106.
42. Bogorny, V.; Avancini, H.; de Paula, B.C.; Rocha Kuplich, C.; Alvares, L.O. Weka-STPM: A Software Architecture and Prototype for Semantic Trajectory Data Mining. *Trans. GIS* **2011**, *15*, 227–248. [[CrossRef](#)]
43. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Data Mining: Practical Machine Learning Tools and Techniques. In *Plastics*; Kaufmann Morg.: Cambridge, MA, USA, 2016; pp. 515–621.
44. Damacena, A.P.C. Avaliação da qualidade de dados geoespaciais. In *Plastics*, 2nd ed.; IBGE: Rio de Janeiro, Brasil, 2019; pp. 41–52.
45. Moraes, A.F.; Bastos, L.C.; Koehler, F.S.; Kracik, M.S.; Moraes, N.I.G. Open integration for knowledge quality distributed: Approach to geospatial data collect in sustainable cities. *Braz. J. Dev.* **2019**, *5*, 31219–31236. [[CrossRef](#)]
46. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
47. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [[CrossRef](#)]
48. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [[CrossRef](#)]
49. Rampersad, H. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Total Performance Scorecard*; Routledge: London, UK, 2020; pp. 159–183. [[CrossRef](#)]
50. Doll, P.; Girshick, R.; Ai, F. Mask R-CNN. *arXiv* **2017**, arXiv:1703.06870.
51. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. *SSD: Single Shot MultiBox Detector*; Springer: Berlin/Heidelberg, Germany, 2015. [[CrossRef](#)]
52. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
53. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [[CrossRef](#)]
54. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
55. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
56. Tang, Y.; Zhou, H.; Wang, H.; Zhang, Y. Fruit detection and positioning technology for a *Camellia oleifera* C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [[CrossRef](#)]
57. Silva, L.A.; Sanchez San Blas, H.; Peral García, D.; Sales Mendes, A.; Villarubia González, G. An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images. *Sensors* **2020**, *20*, 6205. [[CrossRef](#)]
58. Pham, M.T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-Stage Detector of Small Objects Under Various Backgrounds in Remote Sensing Images. *Remote Sens.* **2020**, *12*, 2501. [[CrossRef](#)]
59. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
60. Li, W.; Li, W.; Yang, F.; Wang, P. Multi-Scale Object Detection in Satellite Imagery Based On YOLT. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 162–165. [[CrossRef](#)]

61. Xu, D.; Wu, Y. MRFF-YOLO: A Multi-Receptive Fields Fusion Network for Remote Sensing Target Detection. *Remote Sens.* **2020**, *12*, 3118. [CrossRef]
62. Schreiber, G.; Akkermans, H.; Anjewierden, A.; Hoog, R.; Shadbolt, N.; Velde, W.V.; Wielinga, B. Knowledge Engineering and Management—The CommonKADS Methodology. In *Plastics*; MIT Press: Cambridge, MA, USA, 2000; pp. 69–121.
63. Dresch, A.; Lacerda, D.P.; Júnior, J.A.A. Design Science Research: Método de pesquisa para a engenharia de produção. In *Plastics*; Bookman: Porto Alegre, Brazil, 2015; pp. 22–102.
64. Rajyalakshmi, D.; Raju, K.K.; Varma, G.P.S. Taxonomy of Satellite Image and Validation Using Statistical Inference. In Proceedings of the IEEE 6th Intern Conference on Advanced Computing, Toulouse, France, 18–21 July 2016; pp. 352–361.
65. Li, Y.; Fang, Y.; Cheng, R.; Zhang, W. Spatial Pattern Matching: A New Direction for Finding Spatial Objects. *ACM SIGSPATIAL Spec.* **2019**, *11*, 3–12. [CrossRef]
66. Arganda-Carreras, I.; Kaynig, V.; Rueden, C. Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics* **2017**, *33*, 2424–2426. [CrossRef] [PubMed]
67. Kotikot, S.M.; Kar, B.; Omitaomu, O.A. A Geospatial Framework Using Multicriteria Decision Analysis for Strategic Placement of Reserve Generators in Puerto Rico. *IEEE Trans. Eng. Manag.* **2020**, *67*, 659–669. [CrossRef]
68. Gimenez, F.; Zerbini, C.; Riva, G. Extending SMS Service Coverage in Rural Areas by using LoRa Communication Technology. *IEEE Lat. Am. Trans.* **2020**, *18*, 214–222. [CrossRef]
69. Huai, M.; Miao, C.; Li, Y.; Suo, Q.; Zhang, A. Learning Distance Metrics from Probabilistic Information. *ACM Trans. Knowl. Discov. Data* **2020**, *14*, 53–86. [CrossRef]
70. Rodrigues, M.L.; Körting, T.S.; de Queiroz, G.R.; Sales, C.P.; Silva, L.A.R.d. Detecting Center Pivots in Matopiba Using Hough Transform and Web Time Series Service. In Proceedings of the 2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 22–26 March 2020; pp. 189–194. [CrossRef]
71. Russell, S.; Norvig, P. Artificial Intelligence: A moen approach. In *Plastics*; Prentice Hall: Hoboken, NJ, USA, 2009; pp. 149–297.
72. Zhang, Q.; Zhou, J.; He, J.; Xiaodong, C.; Zhang, Z.S.B. A shell dataset, for shell features extraction and recognition. *Sci. Data* **2019**, *6*, 226. [CrossRef]
73. Erba, D.A.; Loch, C. Cadastro técnico multifinalitário rural e urbano. In *Plastics*; Lincoln: Cambridge, MA, USA, 2007; pp. 50–108.
74. Ferreira, M.C. *Iniciação à Análise Geoespacial: Teoria, técnicas e Exemplos para Geoprocessamento*; Unesp: Sao Paulo, Brazil, 2014; pp. 297–333.
75. QGIS Development Team. QGIS Geographic Information System, 2022. Open Source Geospatial Foundation Project. Available online: <http://qgis.osgeo.org> (accessed on 15 December 2022).
76. Awrangjeb, M.; Ravanbakhsh, M.; Fraser, C.S. Automatic detection of residential buildings using LIDAR data and multispectral imagery. *J. Photogramm. Remote Sens.* **2010**, *65*, 457–467. [CrossRef]
77. Chenthamarakshan, V.; Desphande, P.; Krishnapuram, R.; Varadarajan, R.; Stolze, K. *WYSIWYE: An Algebra for Expressing Spatial and Textual Rules for Visual Information Extraction*; Cornell University Computer Science: Ithaca, NY, USA, 2016.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

3.3.1. Resumen en Español

Esta subsección muestra un resumen en castellano de los aspectos más relevantes del artículo científico titulado “*Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning*” que en castellano se podría traducir a «Acciones activas en la extracción de objetos urbanos para la calidad de la información y la recomendación de conocimientos con aprendizaje automático».

Objetivos de la investigación

Con el creciente uso del aprendizaje automático impulsando la evolución tecnológica en varias áreas del conocimiento, surgen nuevas demandas para mapear procesos específicos y soluciones. Esto es especialmente relevante para las ciudades inteligentes y sostenibles, donde el monitorización territorial es una herramienta esencial. Un desafío existente es la falta de tecnologías que garanticen la calidad de la información necesaria para crear estrategias. En este escenario, se manejan grandes volúmenes de datos que requieren un esfuerzo considerable para comprender sus estructuras y, posteriormente, acceder a la información con la calidad deseada.

Por lo tanto, se propone una solución para ayudar en el reconocimiento de objetos, identificar patrones estadísticos y combinar el uso de métodos y técnicas para extraer conocimiento. A través del modelo desarrollado para la identificación activa de objetos geoespaciales con niveles de similitud, se pudo cruzar datos tras detectar el mejor clasificador con una tasa de acierto del 85 % y un coeficiente de acuerdo kappa del 76 %. En el estudio de caso, se muestran las dinámicas de expansión urbana y rural, donde se obtienen grandes cantidades de datos, que son sometidos a diferentes métodos de catalogación y preparación para respaldar acciones de control rápidas.

La investigación ofrece un enfoque práctico y sistemático, evaluando la extracción de información para recomendar conocimientos. Los resultados se discuten en relación con la construcción de una infraestructura tecnológica mínima que respalde la toma de decisiones.

El presente artículo tiene como propósito fundamental abordar, estudiar y profundizar en temas esenciales relacionados con el planeamiento urbano y regional, y su intersección con las tecnologías modernas de procesamiento y adquisición de datos. Específicamente, los objetivos se desglosan de la siguiente manera:

- **Adquisición de Datos Espaciales y Análisis de Imágenes de Satélites:** Estudiar y comprender las técnicas y herramientas actuales para la adquisición de datos espaciales y su significado en la toma de decisiones urbanas y regionales.
- **Procesamiento de Imágenes y Aprendizaje Automático:** Analizar y evaluación de metodologías vigentes en el procesamiento de imágenes y cómo las técnicas de aprendizaje automático pueden mejorar la precisión y eficiencia de este proceso.
- **Aplicación de Tecnologías GeoDMA y TerraView para el Sensoriamento Remoto:** Evaluar las funcionalidades y beneficios del uso de GeoDMA y TerraView, y su impacto en el sensoriamento remoto.
- **Uso del Software Weka en Contextos Espaciales y Geográficos:** Investigar las aplicaciones del software Weka en el análisis y tratamiento de datos espaciales y geográficos, identificando sus ventajas y limitaciones.

- **Clasificación en el Contexto del Sensoriamento Remoto (SR):** Delimitar y describir las técnicas actuales de clasificación, particularmente en relación con el sensoriamento remoto.
- **Técnicas de Minería de Texto con Weka:** Profundizar en el uso de Weka para la minería de texto, analizando en detalle técnicas como el método TF-IDF y su aplicabilidad en el análisis cuantitativo de textos.
- **Estudio de Caso en Itajaí, Santa Catarina, Brasil:** Presentar y analizar un estudio de caso concreto en Itajaí, identificando patrones, resultados y conclusiones en relación con el planeamiento urbano y regional.
- **Análisis de Datos Ambientales y Sociales:** Considerar la intersección y el impacto de datos ambientales y sociales en la ocupación territorial, aportando una visión holística del desarrollo urbano.

Además de los objetivos mencionados, el artículo estructura su contenido en secciones específicas, desde el desarrollo y metodología hasta la discusión de resultados y conclusiones finales, con el fin de ofrecer una comprensión completa y detallada del tema en estudio.

Metodología

Esta sección proporciona un esquema detallado de la metodología utilizada para desarrollar la investigación presentada en este trabajo. La metodología se divide en tres subsecciones principales: Revisión del estado del arte, Análisis y diseño de la propuesta, y Definición del caso de estudio.

Revisión del estado del arte: El estado del arte se basa en el marco de DSR (Design Science Research). A través de DSR, se establecen jerarquías para tareas intensivas en conocimiento específicas para cada problema identificado. Los métodos de clasificación se dividen según su modo de procesamiento, ya sea visual o digital. Estos métodos se categorizan como supervisados, no supervisados e híbridos. En adición, las métricas se observan en dos tipos: paramétricas o no paramétricas y según el enfoque por píxel o por regiones (objetos). Se presta atención especial a la dificultad de procesar volúmenes masivos de datos, especialmente con imágenes satelitales que pueden ser extensas en tamaño.

Análisis y diseño de la propuesta: El enfoque propuesto en esta investigación se basa en el levantamiento de imágenes satelitales y/o obtenidas por RPAS. También se aprovechan datos textuales compartidos previamente para identificar objetos geoespaciales a través del aprendizaje activo y la generación de reglas elementales. La propuesta incluye una arquitectura de detección y extracción sistemática con soporte en aprendizaje automático.

La implementación de esta propuesta se desarrolla en varias etapas:

1. **Planificación:** Se organizan reuniones, entrevistas y encuestas. Se implementan cuestionarios y se documentan los prerrequisitos del proyecto.
2. **Ingeniería de Requisitos:** Se construyen artefactos utilizando UML (Unified Modeling Language).
3. **Adquisiciones:** Se contratan servicios de recolección de imágenes satelitales, como (CBERS) y (LANDSAT-8), con parámetros específicos.

4. **Procesamiento de Imágenes:** Se realiza la homologación y procesamiento de las imágenes, eligiendo métodos de contraste para visualizar los objetos de interés.
5. **Homologación y Extracción:** Posteriormente se elaboran índices y se aplican métodos específicos de detección y extracción de objetos espaciales o textuales.
6. **Generación de Reglas Adaptativas:** Se evalúan algoritmos para crear reglas adaptativas, patrones estadísticos para representar el conjunto de datos.
7. **Optimización y Ingeniería Inversa:** Se prioriza la optimización de estructuras de datos para el repositorio y la ingeniería inversa posterior.

Definición del caso de estudio: El caso de estudio se enfoca en la detección de patrones geográficos en zonas urbanas y rurales. La segmentación y otros registros son vitales para observar y evaluar automáticamente la expansión territorial a partir del conocimiento compartido extraído. Se utiliza la teledetección para clasificación, comúnmente basada en índices radiométricos, operaciones de contraste aritmético con NDBI, NDVI (Índice de Vegetación de Diferencia Normalizada) y NDWI (Índice de Agua de Diferencia Normalizada). Esto se integra con la planificación urbana y regional, utilizando el Catastro Técnico de Multifinalidad (Catastro Técnico Multifinalitario—CTM) similar al Catastro de España.

La metodología presentada busca combinar múltiples técnicas y herramientas para obtener un reconocimiento preciso de patrones, esencial para la planificación y gestión urbana y territorial.

Resultados

■ Implementación con Algoritmo ID3 (Weka):

Método: Clasificación para elaborar la árbol de decisión. Resultados de la clasificación: Instancias clasificadas correctamente: (38.885) resultando en 78,83 % de tasa de acierto. Instancias clasificadas incorrectamente: 21,15 % (10.436).

■ Implementación con Algoritmo J48 (Weka):

Se generó una nueva estructura de árbol. Se detectaron niveles y la clase "Conservação". Resultados de clasificación: - Instancias clasificadas correctamente: 85,2002 % (42.025). - Instancias no clasificadas: 14,79 % (7.300). - Estadística Kappa: 76,11 % (indicando el mejor aprendizaje del classificador).

■ Uso de Software TerraView/GeoDMA: Se aplicaron técnicas de segmentación, vectorización de imagen y extracción de atributos (ilustrado en la Figura 7). Con GeoDMA, se generaron 212 atributos. Se identificaron características de cada pixel y se generaron y definieron polígonos y reglas básicas para cada objeto. Se evaluaron diferentes tipos de escalas para comprender mejor los objetos detectados.

■ Aplicabilidad de la Extracción de Conocimiento: Integración de imágenes de satélite y otros datos como textos del CTM para el periodo 2001-2019. Software Weka en sus versiones 3.2 (2001) y 3.9.3 (2019) permitió la construcción de modelos y el aprendizaje de máquina.

- **Pruebas de Rendimiento con Weka:** Clasificadores evaluados: OneR, IBk, NaiveBayes, J48. Todos los clasificadores fueron probados en las mismas condiciones y con los mismos conjuntos de datos. Se destaca el uso del clasificador "Hoeffding Tree" de Weka para una generación de árbol menos crítica.
- **Uso de Imágenes de Satélite y GeoDMA:** Se amplió el espectro de información extraída, especialmente útil para proyectos que requieren monitoreo frecuente. El satélite CBERS-4 facilitó la rápida adquisición de nuevas imágenes.

Conclusiones

A partir de la sexta etapa de la metodología, se identificó una marcada inclinación hacia el uso del método SVM (utilizando la biblioteca LibSVM-Weka) para evaluar grandes conjuntos de datos, tanto textuales como geoespaciales. La discretización de estos datos ha demostrado ser una estrategia valiosa para optimizar la performance de los clasificadores.

Asimismo, se confirmó de manera efectiva la integración del software Weka con TerraView/GeoDMA. Ambos programas se complementan en todas las fases, desde la preparación hasta la estructuración de datos textuales y geoespaciales.

Para investigaciones futuras en esta línea, se planea perfeccionar el módulo de preprocesamiento. Esta mejora se centrará en aspectos esenciales, como la recolección de datos de sensores, transformación, preparación de datasets y la extracción de información. Estos pasos son cruciales para el cruce de datos, construcción de reglas y aseguramiento de la calidad de la información obtenida.

3.4. Automated Road Damage Detection Using UAV Images and Deep Learning Techniques

IEEE Access

Multidisciplinary | Rapid Review | Open Access Journal

Received 27 April 2023, accepted 11 June 2023, date of publication 19 June 2023, date of current version 27 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3287770



Automated Road Damage Detection Using UAV Images and Deep Learning Techniques

LUÍS AUGUSTO SILVA¹, (Member, IEEE),
VALDERI REIS QUIETINHO LEITHARDT^{2,3}, (Senior Member, IEEE),
VIVIAN FÉLIX LÓPEZ BATISTA⁴, GABRIEL VILLARRUBIA GONZÁLEZ¹,
AND JUAN FRANCISCO DE PAZ SANTANA¹

¹Expert Systems and Applications Laboratory (ESALAB), Faculty of Science, University of Salamanca, 37008 Salamanca, Spain

²VALORIZA, Polytechnic Institute of Portalegre, 7300-555 Portalegre, Portugal

³Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, 1959-007 Lisbon, Portugal

⁴Department of Computer Science and Automatics, University of Salamanca, 37008 Salamanca, Spain

Corresponding author: Luís Augusto Silva (luisaugustos@usal.es)

The work of Luís Augusto Silva was supported by the Call for Predoctoral Contracts USAL 2021, co-financed by Banco Santander.

This work was supported in part by the Spanish Agencia Estatal de Investigación, and in part by the Project Monitoring and Tracking

Systems for the Improvement of Intelligent Mobility and Behavior Analysis (SiMoMIAC) under Grant PID2019-108883RB-C21/AEI/

10.13039/501100011033, and supported by the Junta Castilla y León through the Project Investigación y desarrollo en tecnología y

algoritmos inteligentes para plataforma unificada, distribuida y de bajo coste de monitorización de inmuebles e infraestructuras públicas o

privadas a través de imágenes satelitales o de drones under Grant 07/18/SA/0022.

ABSTRACT This paper presents a novel automated road damage detection approach using Unmanned Aerial Vehicle (UAV) images and deep learning techniques. Maintaining road infrastructure is critical for ensuring a safe and sustainable transportation system. However, the manual collection of road damage data can be labor-intensive and unsafe for humans. Therefore, we propose using UAVs and Artificial Intelligence (AI) technologies to improve road damage detection's efficiency and accuracy significantly. Our proposed approach utilizes three algorithms, YOLOv4, YOLOv5, and YOLOv7, for object detection and localization in UAV images. We trained and tested these algorithms using a combination of the RDD2022 dataset from China and a Spanish road dataset. The experimental results demonstrate that our approach is efficient and achieves 59.9% mean average precision mAP@.5 for the YOLOv5 version, 65.70% mAP@.5 for a YOLOv5 model with a Transformer Prediction Head, and 73.20% mAP@.5 for the YOLOv7 version. These results demonstrate the potential of using UAVs and deep learning for automated road damage detection and pave the way for future research in this field.

INDEX TERMS UAV, road damage detection, deep learning, object-detection.

I. INTRODUCTION

Managing the maintenance of all the roads in a country is essential to its economic development. A periodic assessment of the condition of roads is necessary to ensure their longevity and safety. Traditionally, state or private agencies have carried out this process manually, who use vehicles equipped with various sensors to detect road damage. However, this method can be time-consuming, expensive, and dangerous for human operators. To address these challenges, researchers and engineers have turned to Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence (AI) technologies to automate the pro-

cess of road damage detection. In recent years, there has been a surge of interest in using UAVs and deep learning-based methods to develop efficient and cost-effective approaches for road damage detection.

Unmanned aerial vehicles have proven to be versatile in various applications, including urban inspections of objects and environments. They have been increasingly used for road inspections, offering several advantages over traditional methods. These vehicles are equipped with high-resolution cameras and other sensors that can capture images of the road surface from multiple angles and heights, providing a comprehensive view of the condition of the road. Additionally, UAVs can cover a large area relatively quickly, reducing the need for manual inspections, which can be dangerous

The associate editor coordinating the review of this manuscript and approving it for publication was Halil Ersin Soken.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

62918

VOLUME 11, 2023

real-time object detection and trained on a large dataset of images and videos to improve its accuracy.

Subsequently, the fifth version of the algorithm, called YOLOv5 [17], was released. This algorithm turned out to be a perfect model, bringing more options as we can highlight the image segmentation, but it still needs to be closer to the 5th major update. The results are very similar to YOLOv4 which a considerable amount of work was done, and all the nuances were taken into account. YOLOv5 is an improvement over YOLOv4. It is based on a new SPADE architecture, which uses semantic and spatial information to improve object detection accuracy. YOLOv5 also uses a new training algorithm called Mosaic Data Augmentation to enhance the model's generalization.

Later and more recently, the seventh version of the algorithm was released [18], the latest iteration in the life cycle of YOLO models. YOLOv7 infers faster and more accurately than its previous versions (i.e., YOLOv5). YOLOv7 is the latest version of YOLO. It has been built on a new architecture called Efficient-YOLO, which uses EfficientNet as the backbone network. YOLOv7 has been trained on a large dataset, and it has been optimized for real-time object detection. It is more accurate and faster than previous versions of YOLO.

In conclusion, YOLOv4 is considered the most accurate version of YOLO until 2021 and is optimized for real-time object detection. YOLOv7 is the latest version of YOLO, and it is based on a new architecture called Efficient-YOLO, which is more accurate and faster than previous versions.

C. OBJECTIVES AND STRUCTURE

This paper builds upon a previous project proposing an architecture for a pavement monitoring system with pothole recognition in UAV images [7]. In this new research, we expand upon the previous solution by comparing it with new algorithms and datasets, introducing new classes of damage, and adopting data augmentation during training, which promotes adapting to dramatic size changes of objects in images. Finally, in this work, the YOLOv5 and YOLOv7 are compared, and an improvement was made in the YOLOv5 model using the Transformer Prediction Head for the UAV use case.

We have used a merged dataset from previous work and Crowdsensing-based Road Damage in this work. Detection Challenge, including new damage classes for a more comprehensive understanding of pavement damage. Experimental results demonstrate the effectiveness and efficiency of our proposed solution, achieving more accuracy on the test dataset.

The main objective of this project is to improve the autonomous monitoring system for the state of roads using images captured by drones and advanced artificial vision and intelligence techniques. The proposed system will notify the maintenance company about detected road damage, including the ability to send messages with the geographical coordinates of the damages found.

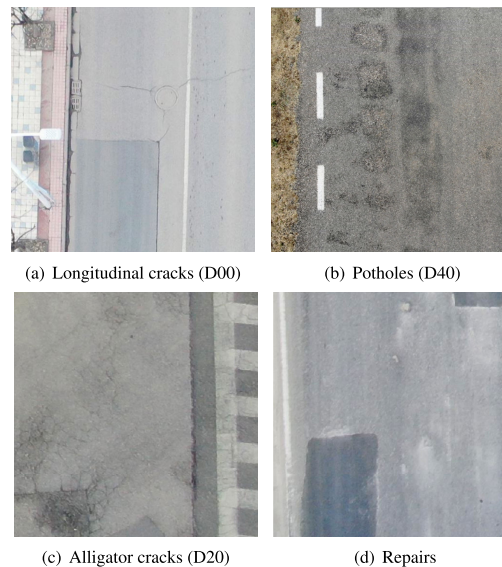


FIGURE 1. Road damage classes in the dataset.

Our team has made several contributions, including:

- Adding an extra prediction head to address the issue of large variations in object scales.
- Incorporating Transformer Prediction Heads (TPH) into the YOLOv5 model, resulting in improved object localization in high-density scenes.
- Providing a range of useful techniques and filtering out ineffective approaches for object detection in drone-captured scenarios.
- Enhancing the classification accuracy of certain ambiguous categories by utilizing a self-trained classifier.
- The project has introduced several new classes of pavement damage, as depicted in Figure 1. These include longitudinal cracks, alligator cracks, potholes, bumps, and repairs. The project offers a more comprehensive understanding of pavement damage by including these additional classes, enabling more precise and efficient road infrastructure monitoring.

In the overall project, convolutional neural networks detect asphalt defects, allowing for operator overrides or suggestions for improved accuracy over time. Additionally, we will implement a feature for automatically planning routes to cover the entire road, eliminating the need for manual operation by the pilot and utilizing PIX4D to automate the route planning.

The structure of this paper is as follows: Section II thoroughly analyzes existing literature on damage detection methods and UAVs. Section III delves into the proposed system's architectural design, the dataset used, and its implementation. The experiments carried out, and their results are discussed in Section IV. Finally, Section V concludes the

paper by presenting a summary of the findings and outlining potential future work.

II. RELATED WORKS

Imagery capture plays a crucial role in the initial assessment of a road or highway's condition. A UAV, specifically a drone, is an efficient and cost-effective way to capture high-quality and detailed photographs of the road surface from various perspectives. In this study, we have used the DJI Mavic Air 2S drone, a more recent drone version budgeted for this project. This drone has advanced features such as a high-resolution camera, GPS, and obstacle avoidance sensors, enabling it to capture high-quality road surface images with minimal distortion. Additionally, using a UAV allows for more comprehensive coverage of the road surface, especially in hard-to-reach areas, and can be done safely and quickly.

Related articles focus on improving existing algorithms in deep learning and unmanned aerial vehicles (UAVs). For example, autonomous UAVs have been used for structural health monitoring and real-time damage mapping using deep learning methods and ultrasonic beacons with geo-tagging [19], [20]. Deep learning techniques, such as CNNs, have shown promising results in various domains, including vehicle traffic monitoring [21], large population monitoring [22], animal identification [23], wind generator inspection [24], and electric component detection [25]. These techniques can also be used to analyze images or video from cameras mounted on vehicles to detect road potholes, making them an effective approach for automated road damage detection.

The transportation industry is no exception, and the task of road damage identification is ready to profit from the rapid advancement and diffusion of deep learning technologies. Using convolutional neural networks (CNNs) or other deep learning techniques to analyze images or video from cameras mounted on vehicles to detect road potholes is possible. One of the fundamental approaches for automated road damage detection is using deep learning algorithms. These algorithms effectively detect a range of objects, including damage.

Standard deep learning methods in this area include the implementation of Convolutional Neural Networks (CNNs). In the paper [26], the authors proposed a deep convolutional neural network (CNN) for road damage detection from UAV images. The proposed CNN was trained and tested on a dataset of UAV images, and the results showed that it could detect road damage accurately. In [27] proposes a novel approach for detecting concrete cracks using a deep architecture of CNN without the need for image processing techniques (IPTs) to extract defect features. The CNN is trained on a large dataset of 40,000 images and achieves an accuracy of about 98%. The proposed method is tested on a different structure under various conditions and performs better than traditional Canny and Sobel edge detection methods.

In another recent work [28], the authors proposed a deep learning-based object detection method for automated road damage detection using UAV images. They used the Faster R-CNN algorithm as the object detector. Results reflected

that the proposed method is superior to other methods of road damage detection.

Also, with Regions with Convolutional Neural Network (R-CNN) and their improvements called Faster R-CNN, the authors in [29] and [30] proposes for structural visual inspection, which can detect multiple types of damages, including concrete cracks, steel corrosion, bolt corrosion, and steel delamination. The proposed method achieves an average precision rating of 87.8%. The proposed method provides a remarkably fast test speed of 0.03 seconds per image and can potentially be used for quasi-real-time damage detection on video using the trained networks.

Finally, [31] proposes a crack detection and quantification method using Faster R-CNN and modified TuFF and DTM algorithms. The proposed method achieved high accuracy with 95% average precision, 83% intersection over union, and 93% accuracy for crack length.

In [32], the authors developed a new sensor technology for road damage detection using a deep learning-based image processing algorithm with super-resolution and semi-supervised learning methods based on GAN. Tested on 400 road images, the proposed method showed an average recognition performance of 81.54% and 79.228% in terms of mean intersection over union and F1-score, respectively. The paper suggests that the proposed method can be used for efficient road management in the future.

Nowadays, simply detecting damage in structural images is not enough. To fully understand and assess the extent of the damage, it is necessary to quantify it by measuring the size of the detected defects. This requires a more advanced technique known as pixel-level segmentation, which can accurately delineate the boundaries of the damaged areas in the image.

Kang [33] proposes a novel semantic transformer representation network (STRNet) for crack segmentation in complex scenes, achieving high performance and fast processing speed. The network was evaluated and compared with other advanced networks, showing superior performance and processing speed compared with other networks, including attention on [34], which proposes a high-performance deep-learning network for real-time pixel-level segmentation of internal damages in concrete members using active thermography. The attention-based IDSNet outperforms state-of-the-art networks with a mean intersection over the union of 0.900, a positive predictive value of 0.952, an F1-score of 0.941, and a sensitivity of 0.942.

Single-Shot Detection (SSD) is another point-of-view specifically for road or concrete damage detection. The work [35] presents the SDDNet, a deep learning model for real-time segmentation of concrete cracks in images, achieving high accuracy on a manually created dataset. The model is compared with recent models and outperforms them while processing images at 36 FPS, which is significantly faster than previous works.

In another work, Arya et al. [36] reported a set of state-of-the-art solutions in global road damage detection and classification tasks. For example, Pham et al. [37] experimented

with a study with Detectron2, implementing Faster R-CNN. Generally, these reviewed studies show that the Faster R-CNN model provides better accuracy with the trade-off of prediction time (8 frames per second) than the YOLO model (40 frames per second). In contrast, SSD balances the two concerning prediction accuracy and time.

A. YOLO IMPLEMENTATIONS

In [38], presents a deep learning approach to identifying potholes on Indian roads using the YOLO algorithm to improve road maintenance and reduce accidents. A new dataset of 1500 images of Indian roads is created and trained using YOLOv3, YOLOv2, and YOLOv3-tiny, and the results are compared in terms of accuracy. In contrast, the authors in [39] present the M-YOLO, which uses a light network architecture based on MobileNet V3 and YOLOv5S to improve the detection efficiency of pavement oil repair using UAV images. The results of experiments showed that the M-YOLO algorithm has an accuracy of 98.3%, an average accuracy of 95.5%, and a detection speed of 96.6fps, which is significantly better than YOLOv3 in terms of accuracy, speed, and number of parameters.

In addition, the authors in [12] present a novel automated pavement distress detection framework that combines stereo vision and deep learning. The proposed method is tested on asphalt roads under various conditions. The results show that it can achieve millimeter-level accuracy in the crack and pothole segmentation and that the enhanced 3D crack segmentation model is superior to other models in terms of accuracy and inference speed. It also uses the high-resolution pothole segmentation map to measure the pothole volume accurately.

Another solution related to the literature is using multispectral images to detect road damages [40]. Multispectral imaging using UAVs is a powerful tool for detecting and analyzing road damage. Another approach is using hyperspectral images to detect road pavement cracks. In the study [41], an asphalt crack index is introduced and found to be effective for crack detection, with an average 21.37% increase in F1-score compared to the existing metric in literature.

Convolutional Neural Networks (CNNs) and Transformer can also be used for hyperspectral image classification [42]. CNNs can extract features from hyperspectral data by learning spatial patterns in the spectral domain, while Transformers can capture global contextual information by modeling long-range dependencies. Both approaches have shown promising results in hyperspectral image classification tasks.

In the literature, many works are connected to this field, which is developing very fast, and more experiments should be performed to find a better approach in this specific case. In this work, we approached YOLO because it is the most efficient technique. When we wrote the first article, the most recent version was YOLOv4. YOLOv7 is the current version of this algorithm, which has been tried extensively in this work. Currently, the work of Wang et al. [18] is the official

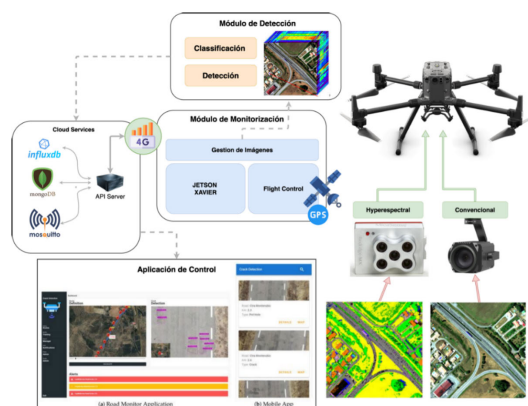


FIGURE 2. Design of the proposed method.

implementation, which some of the authors are the same as the YOLOv4 version [16].

The speed and precision of YOLOv7 are between five and 160 frames per second. This project tested various hyperparameters (using the freebies) and models (adding modules and custom configuration files) to train models for road damage detection and classification tasks.

III. DESIGN AND IMPLEMENTATION

A. IMPLEMENTATION

The proposal's main objective is the detection of deformities on street surfaces, roads, highways, and other vehicle traffic surfaces. The initial proposal of this project, as seen in Figure 2, uses a commercial drone integrated with a high-resolution camera and, in its case, also the use of a multispectral camera. The multispectral camera, as its name suggests, is a camera that is capable of capturing several light spectra. In the case of the dataset of this article, the use of a multispectral camera is not involved and uses only images from high-resolution cameras.

B. UAV IMAGE DATASET

First, we conducted a literature search to find a dataset of potholes and cracks in asphalt at the outset of this study. However, the databases were different from the current suggestion of utilizing an unmanned aerial aircraft to take photographs at a safe distance from the road. Therefore, a new dataset was required to depict the Spanish road situation accurately. In total, 600 pictures with a resolution of 3840×2160 pixels were taken. The images were taken from a DJI Air 2S drone 50 meters from the ground on roads in Spain and only had two classes, potholes (D40) and cracks (D00).

Upon dataset creation and labeling of all photographs, 568 tagged photos were recovered. The photos' orientation was adjusted during the pre-processing stage and got a new size (640×640). Different iterations of each image in the collection were created utilizing augmentation techniques.

TABLE 1. Spain roads dataset.

Class	Spain Annotations
D00 (Crack)	327
D40 (Pothole)	3480

The zoom levels of the photographs ranged from 0% to 15%. In total, 1362 images are included in the collection. 70% of these photos were used for training, 20% for validation, and 10% for testing the trained model's efficacy. This dataset was used in previous work [7], and its repository is available.¹ Table 1 shows how the classification is organized.

Continuing to compose the dataset, we approached the previous datasets (Spain) as a reference for training deep learning models to detect road damage from the collected videos automatically. Added to this, we joined the dataset provided in the CRDDC2022. This dataset is a dataset of road damage in multiple countries [43].

This benchmark dataset is used for training and testing machine learning models for automated pavement distress detection. The dataset contains 47,420 road photos from five countries (China, Japan, the Czech Republic, Norway, the US, and India). We use these photos to train and test models to identify four types of pavement damage: alligator cracks (D20), transverse cracks (D10), longitudinal cracks (D00), and pothole cracks (D40).

The training set of this dataset is used to train machine learning models to recognize the four types of pavement damage. The models learn to identify the characteristics of each type of damage from the photos in the training set. The testing set of this dataset is used to evaluate the performance of the trained models. The models are applied to the photos in the testing set, and their predictions are compared to the actual labels to evaluate the model's accuracy.

This dataset is useful for researchers and engineers working on automated pavement distress detection because it provides a large and diverse set of images that can be used to train and test models. Including images from different countries ensures that the models trained using this dataset can generalize well to different road conditions and environments.

These images were obtained from smartphones, high-resolution cameras, and satellite images. All are obtained by employing cars, motorcycles, and drones. The distribution of damage types (of the four relevant damage types) by countries is displayed explicitly in Table 2. For China, two datasets were made available: Ch_M, which refers to images taken by mobile phones, and Ch_UAV, which refers to images taken by drones.

To compose the dataset for this article, we used the first dataset of roads in Spain and a small part of the images taken by drones in China **Ch_UAV**, as mentioned in the table above. This dataset also includes two complementary classes **Repair** that refers to some repair done on the road and **Block Crack**.

¹<https://github.com/luisaugustos/Pothole-Recognition>

TABLE 2. Damage category-based data statistics for RDD2022.

	JPN	India	CZ	NW	US	Ch_M	Ch_UAV
D00	4049	1555	988	8570	6750	2678	1426
D10	3979	68	399	1730	3295	1096	1263
D20	6199	2021	161	468	834	641	293
D40	2243	3187	197	461	135	235	86



(a) Air 2S is DJI used to adquiry (b) The drone and camera set-up used to capture China_Drone data included in RDD2022.

FIGURE 3. UAV used to obtain images for the dataset.**TABLE 3.** Damage category-based data statistics for the merged dataset.

Class	China_Drone	Spain	Total
D00	1426	327	1753
D10	1263	0	1263
D20	293	0	293
D40	86	3480	3566
Repair	769	0	769
Block Crack	3	0	3

TABLE 4. Dataset split.

	Number of Images	Percentage
Training Set	4000	82%
Validation Set	584	12%
Testing Set	289	6%
TOTAL	4873	100%

We noticed that including the China_Drone data in the proposed training set increased the dataset's heterogeneity. They were aligned with RDD2020 and this work, which focuses on low-cost and affordable automatic road damage detection considering feasible methods for the public.

The final dataset has 2893 images and comprises images from both countries obtained by UAV detailed in Fig. 3(a) and Fig. 3(b).

Table 3 provides an overview of the damage category-based data statistics for the merged dataset, including the distribution of classes and annotations for the China_Drone and Spain datasets.

This dataset is augmented and preprocessed using auto-orientation and also resized to 640×640 . The augmentation was performed on the images to increase the size and diversity of the dataset artificially. Each training example has been augmented to produce two outputs in this case. The rotation has been applied to randomly rotate the images between -15° and $+15^\circ$ to make the model more robust to different orientations of the objects being detected. The final dataset description was presented in Table 4.

C. DATA PREPARATION

According to the previous table, the two data sets were combined and divided into three versions formatted for YOLOv4, YOLOv5, and YOLOv7. One directory is needed for training, while the other is for validation. Additionally, these two folders must have the label and image directories. The labels would include a text file holding the image annotation for each labeled image, while the images would contain the actual photos. The text file's name must match that of the associated image. After producing new YOLO annotations, the folder follows the YOLO dataset structure. Information about the data set, including names and the number of classes, is contained in the file "data.yaml". All this was done thanks to the platform Roboflow,² on which all datasets are stored.

D. MODEL TRAINING

The YOLOv4-tiny model served as the initial basis for this work, which adheres to the coordinated prediction concept just like YOLOv2 and YOLOv3 did. Multi-class classification is possible instead of single-class classification, as in the older versions. This initial network was set up to detect two classes within the 568 images. Later with the use of YOLOv5, YOLOv5-Transformer, and YOLOv7, the number of classes increased to 6. Afterward, new training was performed with the six classes and the 4000 images.

To train our YOLO models, we prepared our data and fed them with the necessary data set. The model trained is capable of detecting the following sorts of cracks: longitudinal cracks (D00), transverse cracks (D10), alligator cracks (D20), pothole cracks (D40), and repair and block cracks. We used 4873 photos from the dataset generated by the roboflow platform to train the model. The training and validation of the models described in this research were carried out using an Intel(R) Core(TM) i9-10940X CPU @ 3.30GHz computer with 128GB of RAM and an RTX3090 GPU with 24GB of integrated memory due to the availability of reasonably priced GPUs.

E. IMAGE AUGMENTATIONS

Image augmentation is a technique to expand the training dataset by applying various transformations to the existing images. Image augmentation aims to introduce variability and diversity in the training dataset, which helps improve the model's generalization ability. YOLOv7 and YOLOv5 can use various image augmentation techniques, such as:

- **Random horizontal flipping:** This technique randomly flips the image horizontally, giving the model more examples of the same object in different orientations.
- **Random cropping:** This technique randomly crops a portion of the image, giving the model more examples of the object in different scales and positions.
- **Random rotation:** This technique randomly rotates the image, providing the model with more examples of objects in different orientations.

²<https://roboflow.ai>

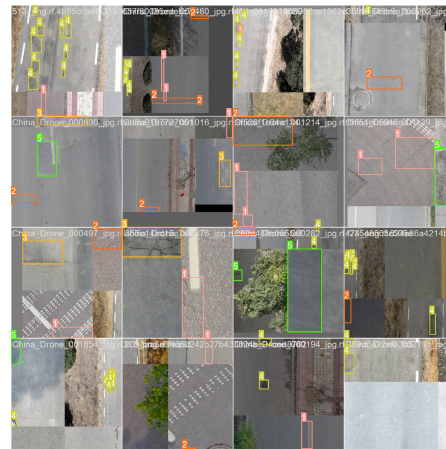


FIGURE 4. YOLOv5's default parameters on augmented images, showcasing the potential for image distortion and inconsistencies.

- **Random brightness and contrast:** This technique randomly adjusts the brightness and contrast of the image, providing the model with more examples of the object under different lighting conditions.
- **Random color jitter:** This technique randomly changes the image's color, giving the model more examples of the object under different color variations.

YOLOv5 and YOLOv7 are state-of-the-art algorithms that enhance and upscale images to improve the robustness and accuracy of the model. However, as seen in Figures 4 and 5, several issues with the default parameters negatively impact the results. Using these techniques, YOLOv7 and YOLOv5 can increase the size and diversity of the training dataset. This can help prevent overfitting and improve the model's generalization ability.

Three models were trained: the first using a YOLOv4 architecture, which replicated the experiment from the prior study; the second with a YOLOv5 design; and the third with a YOLOv7 architecture. We will now go over each model's outcomes and guess how and why we came to those conclusions to choose the ultimate design that is most appropriate for the task at hand.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Intending to test our system, we compare the labeled dataset photos with the final images identified by the algorithm, paying attention to quantitative characteristics. In this case, different experiments result in different models, and in the case of using three different models, we must pay attention to the evaluation process.

Therefore, we need a robust metric to select the best models among all experiments. There are two standard evaluation metrics used in this area. The first is the Mean Accuracy (mAP) calculated in IoU (Intersection overlapping) limit of 0.5 (mAP@0.5). The second is the F1 score. The MAP is

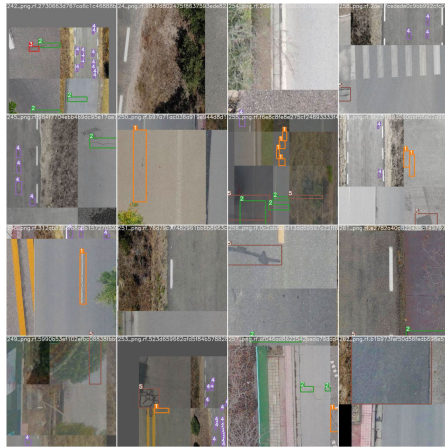


FIGURE 5. YOLOv7's default parameters on augmented images, showcasing the potential for image distortion and inconsistencies.

a good measure when we must ensure the model is stable at different confidence limits (robust) while the F1 score is computed for a specific confidence limit. The common practice is to use mAP@0.5 in the validation set to select the best model and use the F1 score to report the model performance in the test data set. This project also follows this common practice (using mAP@0.5 to select the best models and report the F1 scores in the test sets).

In terms of the quantitative assessment, we'll employ the following metrics: the precision, or the ratio of true positives (TP) to true positives (TP) plus false positives (FP) Equation 1. Equation 2 combines the recall, the likelihood that a picture will be categorized as positive, and the ratio of true positives (TP) to true positives plus false negatives (FN). The third and final metric is the F1 metric, which combines the first two abovementioned metrics.

On the other hand, we also have the classification speed, which is expressed in frames per second (FPS), the mean average precision (mAP), which is determined by using the precision and recall curve, and the IoU, or the overlap area between the detected and imaged areas.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

The model training process was evaluated in 3 steps, switching between iterations and image resolution. The measures mAP@0.5 are used to select the best models during training based on the validation data.

Additionally, with the main evaluation process, we introduced four ways to evaluate the comparative analysis and the

TABLE 5. Performance metric for YOLOv4.

Label	TP	FP	AP (%)
Block crack	0	0	0.00%
D00	87	106	27.14%
D10	149	232	35.17%
D20	4	14	10.39%
D40	162	42	44.25%
Repair	63	69	44.24%

implementation. Using hyperparameter tuning (evolve), error analysis, transfer learning, and ensemble methods.

- **Hyperparameter tuning:** This involves adjusting the settings and parameters of a model or algorithm to find the best configuration for a given task. By systematically varying the values of different parameters, researchers can determine which settings lead to the best performance and gain insight into which factors are most important.
- **Error analysis:** This involves examining the errors made by a model or algorithm and identifying patterns or trends. By analyzing the specific cases where a system fails, researchers can better understand its limitations and identify areas for improvement.
- **Transfer learning:** This involves using a pre-trained model or algorithm as a starting point and fine-tuning it for a specific task. By leveraging the knowledge and experience encoded in a pre-trained model, researchers can often achieve better results with less data and training time.
- **Ensemble methods:** This involves combining the predictions of multiple models or algorithms to improve overall performance. By leveraging the strengths of different models and compensating for their weaknesses, ensemble methods can often achieve better results than any individual model.

A. YOLOv4 EXPERIMENTS

With YOLOv4, the first processing step was performed. Convolutional layers were adjusted as necessary using a pre-trained weight model. Compared to the earlier work, the results of this training could have been better, implying that the prior work involved overtraining or overfitting. In this instance, we achieved a precision of 0.50, with a recall of 0.32 and an F1 score of 0.39. They chose a mean average precision (mAP@0.50) of 0.268638, or 26.86%, for this training, in just three seconds of detecting time. The performance is broken down per class in Table 5.

B. YOLOv5 EXPERIMENTS

Employing YOLOv5 v7.0 for road damage detection showed significant improvements compared to YOLOv4, as indicated in Table 6. The mean average precision (mAP) at an IoU threshold of 0.5 (mAP@.5) increased from 26.86% to 59.90%, indicating a substantial increase in the model YOLOv5x ability to detect road damage accurately.

TABLE 6. Performance metric for YOLOv5.

Class	Imgs	Labels	P	R	mAP@.5	mAP@.95
all	584	1447	0.787	0.561	0.599	0.344
Block	584	1	1	0	0.0203	0.0142
D00	584	367	0.68	0.557	0.594	0.294
D10	584	248	0.778	0.742	0.829	0.46
D20	584	65	0.702	0.507	0.566	0.299
D40	584	623	0.814	0.785	0.817	0.419
Repair	584	143	0.745	0.776	0.766	0.578

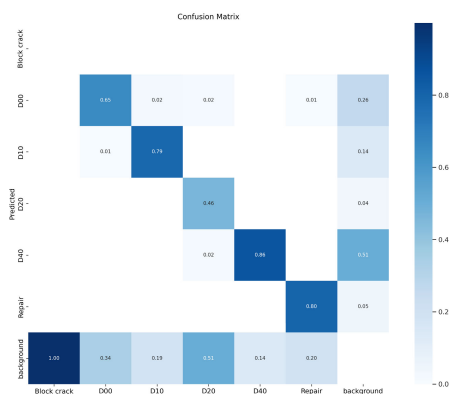


FIGURE 6. Confusion matrix for the YOLOv5 model.

Additionally, the mAP at an IoU threshold of 0.5 and a recall threshold of 0.95 (mAP@.5:.95) also showed a significant improvement, with a boost of around 27%.

Furthermore, the recall percentage, which measures the proportion of actual positive instances that are correctly detected, also increased from 32% to 56.10%. These results demonstrate the effectiveness of YOLOv5 in detecting road damage with high accuracy and recall. Additionally, the inference time of YOLOv5 is 17.2 milliseconds, with a pre-processing time of 0.9 milliseconds, an inference time of 17.2 milliseconds, and a Non-maximal Suppression (NMS) of 6.8 milliseconds per image at shape (1, 3, 640, 640). These results show that YOLOv5 is accurate and efficient in processing time. As shown in Figure 6, the confusion matrix for classifying six groups using test data and YOLOv5 revealed that the model correctly classified most classes.

The horizontal axis represents the ground truth, and the vertical axis represents the predicted classes. The diagonal elements, which represent the correctly classified classes, are the highest among all elements in the matrix, indicating a high level of accuracy. However, it can also be observed that classes D10 and D40 have some misclassifications. These classes are overrepresented in the dataset, which could have made the model more sensitive to detecting these specific classes. However, overall, the results demonstrate the effectiveness of YOLOv5.

Additionally, we calculated the F1 score for each class, as shown in Figure 7. Compared to accuracy, this is a more

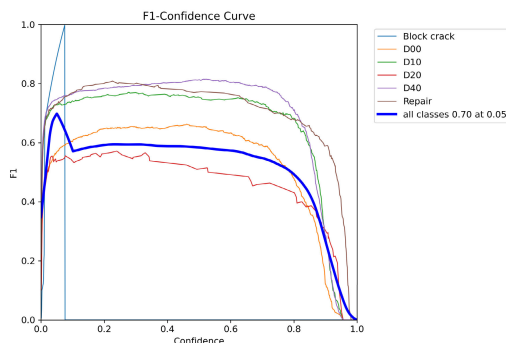


FIGURE 7. F1-Confidence curve.

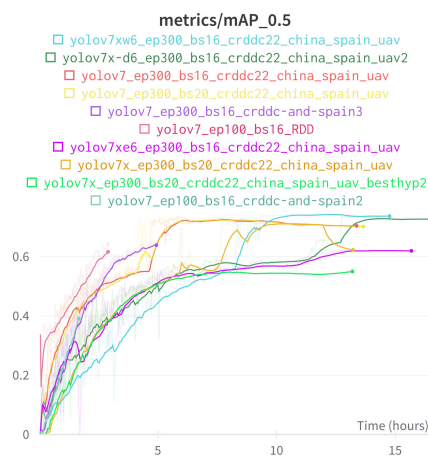


FIGURE 8. mAP@0.5 with the YOLOv7 versions implementing hyperparameters finetuning.

effective metric. The six groups differ in look and structure, making the issue more challenging. The system successfully got a 70% total F1 score. However, the percentage of the F1 score allocated to each class varies from 50% to 80%. The D20 class has the lowest F1 score due to weaker recall, whereas the D40 class has a higher score. A conclusion that the earlier-presented Confusion Matrix supports.

C. YOLOv7 EXPERIMENTS

The best experiment using the YOLOv7 model was trained using the YOLOv7x-W6 model with 300 epochs. A batch size of 16 concluded after the grid research proving the accuracy of prediction and classification try the other YOLO models, YOLOv7-W6, YOLOv7-E6, YOLOv7-D6, YOLOv7-E6E. Figure 8 shows the ten performed trains with these models, changing the batch size and hyperparameters.

Overall, YOLOv7 achieved an mAP of 0.737, indicating that it performed reasonably well in detecting the different

TABLE 7. YOLOv7 performance on pavement damage dataset.

Class	Img	Labels	P	R	mAP@.5	mAP@.95
all	584	1463	0.788	0.714	0.737	0.453
D00	584	349	0.735	0.573	0.590	0.259
D10	584	247	0.751	0.708	0.731	0.348
D20	584	57	0.633	0.439	0.475	0.203
D40	584	627	0.814	0.769	0.802	0.395
Repair	584	182	0.797	0.799	0.827	0.615

classes of pavement damage. The highest mAP was obtained for the **Repair** class, which had an mAP of 0.827. The **Block crack** class achieved perfect precision and recall, indicating that the model could detect this type of damage accurately, but with the low presence of this class in all the datasets, this class was excluded.

However, some classes, such as **D20**, had lower precision and recall values, suggesting that the model struggled to detect this damage accurately. Nonetheless, the model achieved an overall precision of 0.788 and recall of 0.714, indicating that it could detect pavement damage in general accurately. The YOLOv7 is speedier, with an inference time of only 11.4 ms, demonstrating all of this. Speed per 640 × 640 image: 11.4/6.7/18.1 ms inference/NMS/total at batch-size 1

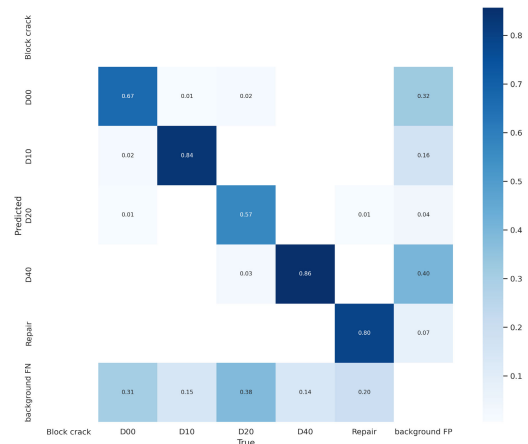
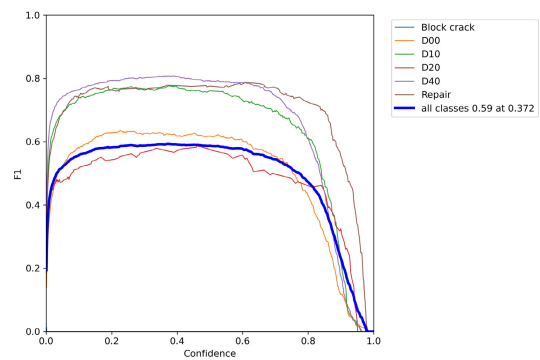
Table 7 shows the Precision (P), Recall (R), and mean Average Precision (mAP) scores for different classes of pavement damage.

In the Table, the detection effect of D00 and D20 is lower than that of other classes. This could be because both D00 (potholes) and D20 (alligator cracks) share some similarities in their visual appearance, which makes it more difficult for the model to distinguish between them and other classes.

Some specific strategies could be employed to improve the detection performance of YOLOv7 on D00 and D20. For example, data augmentation techniques could be used to increase the diversity of the training data, including variations in lighting, angle, and distance. Additionally, transfer learning could be applied to pre-train the model on a large dataset of similar objects, such as road surface images, before fine-tuning the specific classes.

Furthermore, the model architecture was adjusted in the YOLOv5 with Transformer Head to better handle the complexities and variations of potholes and alligator cracks. This improvement will be explained in the next subsections. As shown in Figure 9, the confusion matrix for classifying six groups using test data and YOLOv7 shows that the model accurately categorizes most classes. The horizontal axis represents the ground truth, and the vertical axis represents the predicted classes. The diagonal elements, which represent the correctly classified classes, are the highest among all elements in the matrix, indicating a high level of accuracy.

Compared to the YOLOv5 matrix, it can be observed that there is an increase in the D10 and D40 classes, which were overrepresented in the dataset. This could be due to the improved capability of YOLOv7 to detect these specific classes, even though they are overrepresented, which leads to

**FIGURE 9. Confusion matrix for the YOLOv7 model.****FIGURE 10. F1-Confidence curve.**

improved overall identification. Overall, the results demonstrate the effectiveness of YOLOv7.

The performance of the YOLOv7 algorithm is further analyzed by calculating the F1 score for each class, as shown in Figure 10. The system achieved a total F1 score of 59%. The individual classes have F1 scores ranging from 50% to 80%. The D10, D40, and Repair classes have higher scores than the others. Interestingly, the D00 class has a lower F1 score in YOLOv7 than in YOLOv5, as highlighted by the Confusion Matrix previously presented.

D. YOLOv5 + TRANSFORMER PREDICTION HEAD EXPERIMENTS

In our YOLOv5 + Transformer Prediction Head experiments, the transformer architecture has been used as the prediction head of the YOLOv5 model, replacing the conventional CNN-based prediction head. The transformer head takes in the features extracted by the YOLOv5 backbone network and generates predictions for object locations and class probabili-

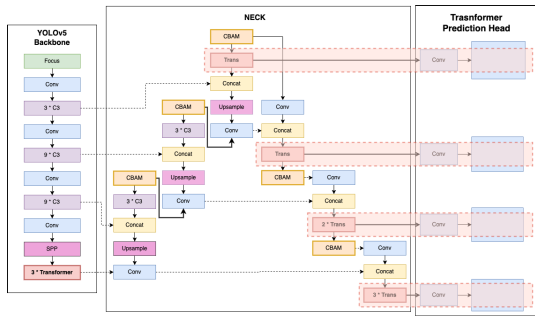


FIGURE 11. YOLOv5 + transformer prediction head model.

TABLE 8. YOLOv5 with Transformer Prediction Head performance on pavement damage dataset.

Class	Img	Labels	P	R	mAP@.5	mAP@.95
all	584	1462	0.71	0.672	0.657	0.35
D00	584	349	0.694	0.576	0.551	0.244
D10	584	247	0.734	0.737	0.735	0.34
D20	584	57	0.53	0.439	0.375	0.177
D40	584	627	0.775	0.793	0.791	0.396
Repair	584	182	0.818	0.816	0.832	0.594

ties. Figure 11 shows the YOLOv5 + Transformer Prediction Head model.

The transformer head consists of multiple layers similar to those used in NLP tasks. Each transformer layer contains a self-attention mechanism, which allows the model to attend to different parts of the input features to make predictions. The transformer head also includes fully connected layers and sigmoid activation functions to produce the final output.

The Experiments have shown that the YOLOv5 + Transformer model can achieve higher accuracy than the standard YOLOv5 model while reducing the computational cost of object detection. In the training process, the model takes 212 epochs completed in 6.121 hours, 10 hours less when compared with YOLOv7x-W6. Table 8 shows the performance metrics for object detection using a YOLOv5 model with a transformer head prediction.

The use of transformer architecture allows the model to capture more complex spatial relationships between objects, resulting in better localization and classification performance. Additionally, the transformer-based prediction head can be trained more efficiently, requiring fewer iterations to converge to a good solution. Figures 12 and 13 shows the performance of the model.

The YOLOv5 + Transformer model achieved a mAP@.5 of 0.657 on all objects, with a precision of 0.71 and a recall of 0.672. The model also achieved high mAP values for individual classes. Overall, this experiment demonstrates the high performance of the YOLOv5 + transformer model with the dataset.

E. VISUAL ANALYSIS

As shown in Figure 14, YOLOv5 could accurately recognize and identify road damage structures from UAV images. The predicted model outcome closely reflects the initial test

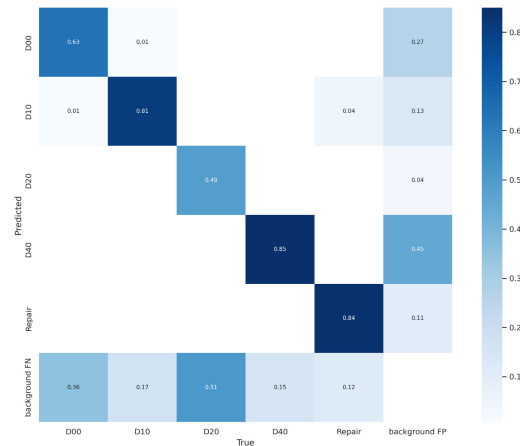


FIGURE 12. Confusion matrix for the YOLOv5 + transformer prediction head model.

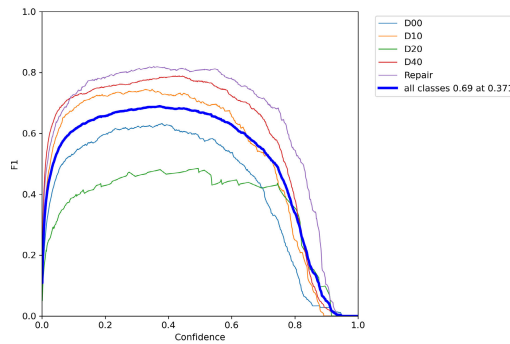


FIGURE 13. F1-Confidence curve.

photographs, indicating high accuracy and performance. Similarly, YOLOv7’s ability to identify road damage structures can be observed in Figures 15, 16 and 17. The model correctly detects the damages, even in cases where the images contain minor flaws or unclear classes.

Visual analysis shows that both YOLOv5 and YOLOv7 can accurately identify and locate road damage structures in UAV images. Specifically, when comparing Figures 14 and 15, it can be seen that YOLOv5 has a slightly more accurate and precise bounding box placement around the damage, whereas YOLOv7 tends to have slightly larger bounding boxes that encompass more of the surrounding area. It can also be observed that YOLOv7 tends to split the damage class into smaller sub-regions, which can be seen in the increased number of bounding boxes around the same area of damage compared to YOLOv5. This is particularly noticeable for the D40 category, which is overrepresented in the data.

In Figure 16, the bounding box around the recheck at the center of the image may be inadequate, possibly due to occlusion or a high box loss value. This highlights the need

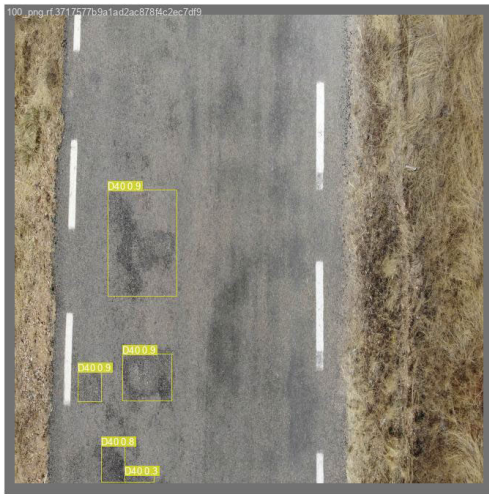


FIGURE 14. Validation YOLOv5.

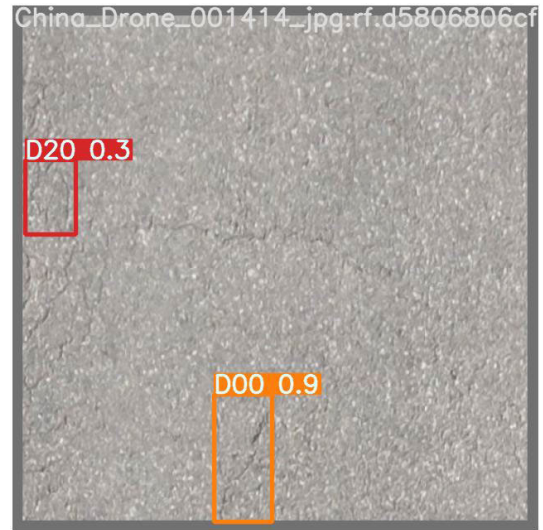


FIGURE 16. Validation YOLOv7.



FIGURE 15. Validation YOLOv7.



FIGURE 17. Validation YOLOv7X-W6.

TABLE 9. Performance comparison for YOLOv4, YOLOv5 and YOLOv7.

	Imgs	P	R	@.5	.5:.95	Time
YOLOv4	584	0.50	0.32	0.26	-	3s
YOLOv5x	584	0.78	0.56	0.59	0.895	17.2ms
YOLOv5 + TPH	584	0.71	0.67	0.65	0.35	6.1ms
YOLOv7	584	0.65	0.78	0.73	0.289	11.4ms

for further improvement in the model’s ability to correctly identify and draw bounding boxes around such instances of damage.

F. COMPARISON

Table 9 compares the performance of the YOLOv4, YOLOv5, and YOLOv7 architectures for road damage identification.

The results show that YOLOv4 performed poorly in precision and detection speed, while YOLOv5 and YOLOv7 performed significantly better. Specifically, YOLOv7 stood out for its high accuracy and fast detection time. The table also includes information on the number of images used for testing, precision, recall, mean average precision (mAP), and inference time for each model.

V. CONCLUSION AND FUTURE WORKS

In conclusion, this study compares the YOLOv4 from past work, the YOLOv5 and YOLOv7 architectures, and includes

an implementation of the YOLOv5 with Transformer for road damage identification using UAV images. The research successfully achieved its goal of creating an architecture capable of detecting road damage and demonstrated that new architecture versions, such as YOLOv5 and YOLOv7, can improve upon previous work.

A significant contribution of this study was the development of a UAV image database tailored explicitly for training the YOLO versions, which was further enhanced by merging with the RDD2022 dataset. This improved detection of road damage samples, particularly for Spanish and Chinese roads, and helped reduce class imbalance for specific forms of road damage, such as potholes and alligator cracks. The findings of this study provide a valuable contribution to the field and pave the way for future research in this area. As presented in the results section, our implementation achieved a mAP.5 of 26.8% with YOLOv4, 59.9% with YOLOv5, and 73.20% with YOLOv7, finally the implemented Transformer achieved 65.7%. There is still scope for improvement in our work.

Future research can explore the different types of images, such as multispectral images and LIDAR sensors, to further enhance the performance. The fusion of such information is potentially possible to yield better results using embedded computer. Moreover, another approach to this work is the use of fixed-wing UAV.

REFERENCES

- H. S. S. Blas, A. C. Balea, A. S. Mendes, L. A. Silva, and G. V. González, "A platform for swimming pool detection and legal verification using a multi-agent system and remote image sensing," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 2023, pp. 1–13, Jan. 2023.
- V. J. Hodge, R. Hawkins, and R. Alexander, "Deep reinforcement learning for drone navigation using sensor data," *Neural Comput. Appl.*, vol. 33, no. 6, pp. 2015–2033, Jun. 2020, doi: [10.1007/s00521-020-05097-x](https://doi.org/10.1007/s00521-020-05097-x).
- A. Safonova, Y. Hamad, A. Alekhina, and D. Kaplun, "Detection of Norway spruce trees (*Picea abies*) infested by bark beetle in UAV images using YOLOs architectures," *IEEE Access*, vol. 10, pp. 10384–10392, 2022.
- D. Gallacher, "Drones to manage the urban environment: Risks, rewards, alternatives," *J. Unmanned Vehicle Syst.*, vol. 4, no. 2, pp. 115–124, Jun. 2016.
- L. A. Silva, A. S. Mendes, H. S. S. Blas, L. C. Bastos, A. L. Gonçalves, and A. F. de Moraes, "Active actions in the extraction of urban objects for information quality and knowledge recommendation with machine learning," *Sensors*, vol. 23, no. 1, p. 138, Dec. 2022, doi: [10.3390/s23010138](https://doi.org/10.3390/s23010138).
- L. Melendy, S. C. Hagen, F. B. Sullivan, T. R. H. Pearson, S. M. Walker, P. Ellis, A. K. Sambodo, O. Roswintarti, M. A. Hanson, A. W. Klassen, M. W. Palace, B. H. Braswell, and G. M. Delgado, "Automated method for measuring the extent of selective logging damage with airborne LiDAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 139, pp. 228–240, May 2018, doi: [10.1016/j.isprsjprs.2018.02.022](https://doi.org/10.1016/j.isprsjprs.2018.02.022).
- L. A. Silva, H. S. S. Blas, D. P. García, A. S. Mendes, and G. V. González, "An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images," *Sensors*, vol. 20, no. 21, p. 6205, Oct. 2020, doi: [10.3390/s20216205](https://doi.org/10.3390/s20216205).
- M. Guerrieri and G. Parla, "Flexible and stone pavements distress detection and measurement by deep learning and low-cost detection devices," *Eng. Failure Anal.*, vol. 141, Nov. 2022, Art. no. 106714, doi: [10.1016/j.engfailanal.2022.106714](https://doi.org/10.1016/j.engfailanal.2022.106714).
- D. Jeong, "Road damage detection using YOLO with smartphone images," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5559–5562, doi: [10.1109/BIGDATA50022.2020.9377847](https://doi.org/10.1109/BIGDATA50022.2020.9377847).
- M. Izadi, A. Mohammadzadeh, and A. Haghghattalab, "A new neuro-fuzzy approach for post-earthquake road damage assessment using GA and SVM classification from QuickBird satellite images," *J. Indian Soc. Remote Sens.*, vol. 45, no. 6, pp. 965–977, Mar. 2017.
- Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, "Convolutional neural networks based potholes detection using thermal imaging," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 3, pp. 578–588, Mar. 2022, doi: [10.1016/j.jksuci.2019.02.004](https://doi.org/10.1016/j.jksuci.2019.02.004).
- J. Guan, X. Yang, L. Ding, X. Cheng, V. C. Lee, and C. Jin, "Automated pixel-level pavement distress detection based on stereo vision and deep learning," *Automat. Constr.*, vol. 129, p. 103788, Sep. 2021, doi: [10.1016/j.autcon.2021.103788](https://doi.org/10.1016/j.autcon.2021.103788).
- D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "RDD2022: A multi-national image dataset for automatic road damage detection," 2022, *arXiv:2209.08538*.
- J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 6517–6525, doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).
- J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*. [Online]. Available: <https://pjreddie.com/yolo/>
- A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, C. Wong, D. Montes, Z. Wang, C. Fati, J. Nadar, V. Sonck, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, "Ultralytics/YOLOv5: v7.0—YOLOv5 SOTA realtime instance segmentation," Zenodo, Tech. Rep., Nov. 2022. [Online]. Available: <https://zenodo.org/record/7347926>
- C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
- R. Ali, D. Kang, G. Suh, and Y.-J. Cha, "Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures," *Autom. Construct.*, vol. 130, Oct. 2021, Art. no. 103831. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S092658052100282X>
- D. Kang and Y.-J. Cha, "Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging: Autonomous UAVs for SHM," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 10, pp. 885–902, Oct. 2018.
- Z. Xu, H. Shi, N. Li, C. Xiang, and H. Zhou, "Vehicle detection under UAV based on optimal dense YOLO method," in *Proc. 5th Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2018, pp. 407–411, doi: [10.1109/ICSAI.2018.8599403](https://doi.org/10.1109/ICSAI.2018.8599403).
- P. Kannadaguli, "YOLO v4 based human detection system using aerial thermal imaging for UAV based surveillance applications," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Nov. 2020, pp. 1213–1219, doi: [10.1109/DASA51403.2020.9317198](https://doi.org/10.1109/DASA51403.2020.9317198).
- T. Petso, R. S. Jamsola, D. Mpoeleng, and W. Mmereki, "Individual animal and herd identification using custom YOLO v3 and v4 with images taken from a UAV camera at different altitudes," in *Proc. IEEE 6th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2021, pp. 33–39, doi: [10.1109/ICSIP52628.2021.9688827](https://doi.org/10.1109/ICSIP52628.2021.9688827).
- L. Wang and Z. Zhang, "Automatic detection of wind turbine blade surface cracks based on UAV-taken images," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7293–7303, Sep. 2017, doi: [10.1109/TIE.2017.2682037](https://doi.org/10.1109/TIE.2017.2682037).
- D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1599–1601, Jun. 2020, doi: [10.1109/TPWRD.2019.2944741](https://doi.org/10.1109/TPWRD.2019.2944741).
- M. A. A. Khan, M. Alsawwaf, B. Arab, M. AlHashim, F. Almarsharawi, O. Hakami, S. O. Olatunji, and M. Farooqui, "Road damages detection and classification using deep learning and UAVs," in *Proc. 2nd Asian Conf. Innov. Technol. (ASIANCON)*, Aug. 2022, pp. 1–6, doi: [10.1109/ASIANCON55314.2022.9909043](https://doi.org/10.1109/ASIANCON55314.2022.9909043).
- Y.-J. Cha, W. Choi, and O. Büyükköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017.
- M. Böyük, R. Duvar, and O. Urhan, "Deep learning based vehicle detection with images taken from unmanned air vehicle," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–4, doi: [10.1109/ASYU50717.2020.9259868](https://doi.org/10.1109/ASYU50717.2020.9259868).

- [29] R. Li, J. Yu, F. Li, R. Yang, Y. Wang, and Z. Peng, "Automatic bridge crack detection using unmanned aerial vehicle and faster R-CNN," *Construct. Building Mater.*, vol. 362, Jan. 2023, Art. no. 129659, doi: [10.1016/j.conbuildmat.2022.129659](https://doi.org/10.1016/j.conbuildmat.2022.129659).
- [30] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyükoztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 731–747, Sep. 2018.
- [31] D. Kang, S. S. Benipal, D. L. Gopal, and Y.-J. Cha, "Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning," *Autom. Construct.*, vol. 118, Oct. 2020, Art. no. 103291.
- [32] S. Shim, J. Kim, S.-W. Lee, and G.-C. Cho, "Road damage detection using super-resolution and semi-supervised learning with generative adversarial network," *Autom. Construct.*, vol. 135, Mar. 2022, Art. no. 104139, doi: [10.1016/j.autcon.2022.104139](https://doi.org/10.1016/j.autcon.2022.104139).
- [33] D. H. Kang and Y.-J. Cha, "Efficient attention-based deep encoder and decoder for automatic crack segmentation," *Struct. Health Monitor.*, vol. 21, no. 5, pp. 2190–2205, Sep. 2022.
- [34] R. Ali and Y.-J. Cha, "Attention-based generative adversarial network with internal damage segmentation using thermography," *Autom. Construct.*, vol. 141, Sep. 2022, Art. no. 104412.
- [35] W. Choi and Y. Cha, "SDDNet: Real-time crack segmentation," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 8016–8025, Sep. 2020.
- [36] D. Arya, H. Maeda, S. Kumar Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Global road damage detection: State-of-the-art solutions," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5533–5539, doi: [10.1109/BIGDATA50022.2020.9377790](https://doi.org/10.1109/BIGDATA50022.2020.9377790).
- [37] V. Pham, C. Pham, and T. Dang, "Road damage detection and classification with Detectron2 and faster R-CNN," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5592–5601, doi: [10.1109/BIGDATA50022.2020.9378027](https://doi.org/10.1109/BIGDATA50022.2020.9378027).
- [38] L. Parameswaran, "Deep learning based detection of potholes in Indian roads using YOLO," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Feb. 2020, pp. 381–385, doi: [10.1109/ICICT48043.2020.9112424](https://doi.org/10.1109/ICICT48043.2020.9112424).
- [39] Y. Liu, G. Shi, Y. Li, and Z. Zhao, "M-YOLO based detection and recognition of highway surface oil filling with unmanned aerial vehicle," in *Proc. 7th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Apr. 2022, pp. 1884–1887, doi: [10.1109/ICSP54964.2022.9778782](https://doi.org/10.1109/ICSP54964.2022.9778782).
- [40] Y. O. Ouma and M. Hahn, "Pothole detection on asphalt pavements from 2D-colour pothole images using fuzzy c-means clustering and morphological reconstruction," *Autom. Construct.*, vol. 83, pp. 196–211, Nov. 2017, doi: [10.1016/j.autcon.2017.08.017](https://doi.org/10.1016/j.autcon.2017.08.017).
- [41] M. Abdellatif, H. Peel, A. G. Cohn, and R. Fuentes, "Pavement crack detection from hyperspectral images using a novel asphalt crack index," *Remote Sens.*, vol. 12, no. 18, pp. 1–10, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/18/3084>
- [42] F. Viel, R. C. Maciel, L. O. Seman, C. A. Zeferino, E. A. Bezerra, and V. R. Q. Leithardt, "Hyperspectral image classification: An analysis employing CNN, LSTM, transformer, and attention mechanism," *IEEE Access*, vol. 11, pp. 24835–24850, 2023.
- [43] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Crowdsensing-based road damage detection challenge (CRDDC-2022)," 2022, *arXiv:2211.11362*.



LUÍS AUGUSTO SILVA (Member, IEEE) is currently pursuing the Ph.D. degree in computer engineering with the University of Salamanca, Spain. He is a Researcher with the Expert Systems and Applications Laboratory (ESALab). His research interests include object detection, image segmentation, deep learning applied to aerial/satellite images, the Internet of Things, and embedded UAV systems.



VALDERI REIS QUIETINHO LEITHARDT (Senior Member, IEEE) received the Ph.D. degree in computer science from INF-UFRGS, Brazil, in 2015. He is currently a Professor with the Polytechnic Institute of Portalegre and a Researcher integrated with the VALORIZA Research Center for Endogenous Resource Valorization. He is also a Collaborating Researcher with the Expert Systems and Applications Laboratory (ESALab), University of Salamanca, Spain. His main research interests include distributed systems, focusing on data privacy, communication, and programming protocols, involving scenarios and applications for the Internet of Things, smart cities, big data, cloud computing, and blockchain.



VIVIAN FÉLIX LÓPEZ BATISTA received the Ph.D. degree in computer science from the University of Valladolid, in 1996. Since 1998, she has been a Full Professor of computer science with the University of Salamanca, Spain. She performed a research stay with the Big Data Analytics and Data Mining Program, Center for Computational Science, University of Miami, in 2012. She was the Director of the Master's in Intelligent System and Program of Ph.D. in computer science with the University of Salamanca, from 2009 to 2012. She is currently a member of the Data Mining Group (MIDA). She has done research on natural language processing, machine learning, and neural networks. Furthermore, she has also papers published in recognized journals, workshops and symposiums, books, and book chapters on these topics. She has two six-year of research granted the first one, in June 2013, and the second one covering the evaluation period, from 2012 to 2018, was granted in 2019. She has a total of 25 JCR. Her average number of citations/year during the last five years is 18.4. According to Mendeley, her H-index: 17. She has supervised Ph.D. theses within the Doctoral Program, University of Salamanca. She has been a member of the organizing and scientific committee of several international symposiums.



GABRIEL VILLARRUBIA GONZÁLEZ received the master's degree in intelligent systems from the University of Salamanca, in 2012, the master's degree in internet security, in 2014, the master's degree in information systems management, in 2015, and the Ph.D. degree from the Department of Computer Science and Automation, University of Salamanca. He was a Computer Engineer with the Pontifical University of Salamanca, in 2011. He is currently an Associate Professor with the University of Salamanca and a Researcher with the Expert Systems and Applications Laboratory (ESALab). Throughout his training, he has followed a well-defined line of research focused on applying multi-agent systems to ambient intelligence environments, particularly concerning the definition of intelligent architectures and information fusion.



JUAN FRANCISCO DE PAZ SANTANA received the degree in technical engineering in systems computer sciences, the degree in engineering in computer sciences, the degree in statistics, and the Ph.D. degree in computer science from the University of Salamanca, Spain, in 2003, 2005, 2007, and 2010, respectively. He is currently a Full Professor with the University of Salamanca and a Researcher with the Expert Systems and Applications Laboratory (ESALab). He is the coauthor of published papers in several journals, workshops, and symposiums.

3.4.1. Resumen en Español

Esta subsección muestra un resumen en castellano de los aspectos más relevantes del artículo científico titulado “*Automated Road Damage Detection Using UAV Images and Deep Learning Techniques*” que en castellano se podría traducir a «Comparación de algoritmos para la detección automática de daños en carreteras mediante imágenes de vehículos aéreos no tripulados».

Objetivos de la investigación

Este artículo introduce un vanguardista método automatizado para detectar daños en carreteras, empleando imágenes capturadas por vehículos aéreos no tripulados (UAV) combinado con avanzadas técnicas de aprendizaje profundo. La preservación de la infraestructura vial es esencial para asegurar un transporte seguro y sostenible. No obstante, el registro manual de daños en carreteras es tedioso y potencialmente peligroso.

De ahí que proponemos una innovadora alternativa que integra UAVs y tecnologías de *Machine Learning* para optimizar la eficiencia y precisión en la identificación de deterioros viales. Aunque tradicionalmente este proceso se ha hecho manualmente mediante vehículos con sensores, esto puede ser lento, costoso y peligroso. Frente a ello, la tecnología de UAV, que emergen como soluciones automatizadas.

En apoyo a esto, el este trabajo propone el uso de un dataset proveniente del Crowdsensing-based Road Damage Detection Challenge (CRDDC), un reto paradetección de daños en carreteras basado en ayuda de participantes de todo el mundo, se organizó dentro de la IEEE BigData Cup 2022. Este concurso global ofreció un dataset con 47,420 imágenes viales de seis países, etiquetadas con 55,000 casos de daños, como grietas y baches. Muchas organizaciones trabajaron con el conjunto de datos RDD2022, utilizando algoritmos como los de la serie YOLO y Faster RCNN.

El principal objetivo de este proyecto es mejorar el sistema autónomo de detección del estado de las carreteras mediante imágenes captadas por drones. El sistema propuesto notificará a la empresa de mantenimiento sobre los daños detectados en la carretera, incluyendo la capacidad de enviar mensajes con las coordenadas geográficas de los daños encontrados.

Como objetivos específicos se plantean:

- Incorporación de Transformer Prediction Heads (TPH) en el modelo YOLOv5, lo que da como resultado una mejor localización de objetos en escenas de alta densidad (imágenes aéreas)
- Ofrecer técnicas efectivas y descartar métodos ineficientes para la detección de objetos en imágenes captadas por drones
- Introducir clases nuevas de daños al pavimento. Estos incluyen grietas longitudinales, grietas de cocodrilo, baches, golpes y reparaciones
- Eliminar la necesidad de operación manual por parte del piloto y utilizando PIX4D para automatizar la planificación de rutas.

Metodología

El diseño inicial de este proyecto se basa en un dron comercial equipado con una cámara de alta resolución, y potencialmente, una cámara multiespectral, la cual, como sugiere su nombre, captura varios espectros de luz. Sin embargo, en este estudio solo se utilizaron imágenes de cámaras de alta resolución.

Iniciamos con una búsqueda bibliográfica intentando encontrar conjuntos de datos sobre baches y grietas en el asfalto, sin resultados positivos. Por ello, creamos un conjunto personalizado de 600 imágenes de 3840x2160 píxeles cada una. Estas se obtuvieron desde un dron DJI Air 2S, a una altura de 50 metros sobre carreteras en España, clasificando los daños en dos categorías: baches (D40) y grietas (D00).

Posteriormente, tras etiquetar las imágenes se ha seleccionado 568 de ellas, en la etapa de preprocesamiento, ajustamos su orientación y las redimensionamos a 640x640 píxeles. Se generaron múltiples versiones de cada imagen aplicando técnicas de aumento. El zoom de estas imágenes varió entre 0 % y 15 %, totalizando 1362 imágenes en la colección. Del total, el 70 % se destinó a entrenamiento, 20 % a validación y 10 % a pruebas.

Adicionalmente, incorporamos el conjunto de datos del CRDDC2022, con 47.420 imágenes sobre daños en carreteras de diversos países. Estas imágenes sirvieron para entrenar y evaluar modelos capaces de identificar cuatro tipos de daños en el asfalto. Decidimos incluir datos del conjunto China_Drone para enriquecer la heterogeneidad del dataset. Este conjunto de datos, alineado con RDD2020, se enfoca en técnicas económicas de detección automática de daños viales. El conjunto de datos consolidado contiene 2893 imágenes capturadas por UAVs de ambos países. El procesamiento de datos se llevó a cabo en Roboflow, combinando ambos conjuntos y adaptándolos para YOLOv4, YOLOv5 y YOLOv7.

El entrenamiento del modelo se efectuó en una computadora Intel(R) Core(TM) i9-10940X @ 3.30GHz con 128GB de RAM y una GPU RTX3090 de 24GB. Para enriquecer el conjunto de entrenamiento, aplicamos transformaciones como: volteado, recorte, rotación y ajustes aleatorios de brillo, contraste y color. Mediante estas técnicas, YOLOv7 y YOLOv5 pueden ampliar y diversificar el conjunto de entrenamiento, evitando el sobreajuste y potenciando la generalización del modelo.

Resultados

Para discernir entre los modelos más óptimos en todos los experimentos, es imprescindible una métrica robusta. En este ámbito, se utilizan dos métricas de evaluación estándar. La primera es la Precisión Media (mAP) calculada con un límite de IoU (Intersección sobre Unión) de 0.5 (mAP@0.5). La segunda es la puntuación F1 score.

CUADRO 3.1: Comparación de rendimiento para YOLOv4, YOLOv5 y YOLOv7.

	Imgs	P	R	@.5	.5:.95	Time
YOLOv4	584	0.50	0.32	0.26	-	3s
YOLOv5x	584	0.78	0.56	0.59	0.895	17.2ms
YOLOv5 + TPH	584	0.71	0.67	0.65	0.35	6.1ms
YOLOv7	584	0.65	0.78	0.73	0.289	11.4ms

En los experimentos se ha conseguido resultados prometedores: 59,9 % mAP@.5 para YOLOv5, 65,70 % mAP@.5 con un modelo YOLOv5 mejorado con el Transformer Prediction Head y 73,20 % mAP@.5 con YOLOv7.

Estos indicadores subrayan el potente potencial de los UAVs y el aprendizaje profundo en la identificación automática de daños carreteros, sentando las bases para futuras exploraciones en esta área. Las nuevas clases resultaron en una comprensión más completa de los daños al pavimento al incluir clases adicionales, lo que permite una monitorización más precisa y eficiente de la infraestructura vial.

Conclusions

En conclusión, el presente estudio contrasta las arquitecturas YOLOv4, YOLOv5 y YOLOv7, incorporando además una variante del YOLOv5 con Transformer para detectar daños en carreteras a través de imágenes obtenidas por UAVs. Esta investigación destaca no solo por desarrollar una arquitectura robusta para la detección de daños, sino también por evidenciar cómo las versiones más recientes, como YOLOv5 y YOLOv7, superan a sus predecesoras.

Un logro destacable es la creación de una base de datos específica de imágenes UAV para entrenar las versiones YOLO, enriquecida con la adición del conjunto RDD2022. Esta integración potenció la detección de daños en carreteras, con un enfoque particular en vías españolas y chinas, y contribuyó a equilibrar las clases de daños específicos como baches y grietas de cocodrilo. Según nuestra sección de resultados, logramos un mAP@0.5 de 26.8 % con YOLOv4, 59.9 % con YOLOv5, 73.20 % con YOLOv7, y con la implementación de Transformer, alcanzamos un 65.7 %, indicando oportunidades de optimización.

Las futuras líneas de investigación podrían considerar la integración de imágenes multispectrales y el uso de sensores LIDAR para enriquecer el desempeño del modelo. La combinación de estas fuentes de información podría arrojar resultados más precisos con el apoyo de computadoras avanzadas. Adicionalmente, se sugiere explorar el uso de drones de ala fija para abordar esta temática desde otra perspectiva.

Capítulo 4

Conclusions

This last chapter of the doctoral thesis summarises the conclusions drawn from the research carried out over the last few years and presented in the previous chapters. The chapter reviews the results and contributions of the research and describes the future lines of research for the coming years concerning the topic of the thesis.

4.1. Contributions

This section describes the main contributions made in this work and the most noteworthy results of the research carried out in this investigation.

Originally, a comprehensive state-of-the-art review of monitoring applications and protocols using aerial imagery, artificial intelligence, and deep learning was conducted. This analysis has facilitated a deeper understanding of the merits and challenges inherent in intelligent infrastructure monitoring and clarified the relationship between UAVs and deep learning methodologies. During the study, a set of images was created and subsequently made available to the scientific community.

During this research, an in-depth study of the existing protocols was carried out to determine which of them were the most favorable to be employed to satisfy the requirements of the case studies in the research articles. Additionally, multiple algorithms were evaluated to identify the most effective techniques for each problem. These identified algorithms have been tested with different datasets to evaluate their performance in aerial image identification. In particular, the different versions of the YOLOv5 and YOLOv7 algorithms were examined, and the focus was on recommendation algorithms and control systems to determine the best candidate for each case study.

In the development of the three case studies considered in this work, different image processing techniques have been employed for the improvement and effective implementation of the algorithms, with a notable focus on data augmentation and enriching the dataset. It can also be highlighted, in one of our publications, that a collective detection dataset has been used, validating the effectiveness of the platform.

Based on the comprehensive reviews carried out in this research, a system was developed that achieved the following milestones: (I) A comprehensive review of the current state of the art of artificial intelligence and deep learning technologies relevant to aerial image analysis. This review highlighted the main current opportunities and challenges within the field of urban infrastructure monitoring and maintenance. (II) The design, development, and validation of deep learning models were demonstrated to be effective in detecting damage to urban infrastructure using aerial imagery. (III) Exploration and establishment of intelligent architectures were

carried out. These architectures, adapted for infrastructure damage detection, reflected adaptability in responding to identified changes or anomalies. (IV) An advanced network for aerial image processing was established. Together with Edge Computing, this network demonstrated the ability to quickly identify and classify potential infrastructure damage. (V) A deep learning model was trained and evaluated. The model became an essential resource for decision-making in automated infrastructure monitoring. (VI) The research examined three different case studies covering road damage datasets from Spain and China, as well as urban imagery from Brazil. This analysis highlighted the practicality and potential of artificial intelligence and deep learning techniques in the detection, prevention, and repair of urban infrastructure.

Estos caso de estudio ejemplifican las aplicaciones prácticas y el potencial de las metodologías y técnicas propuestas en esta investigación. Las explicaciones detalladas de estos caso de estudio son expuestas en cada artículo científico que conforma esta tesis doctoral.

These case studies exemplify the practical applications and potential of the methodologies and techniques proposed in this research. Detailed explanations of these case studies are presented in each scientific article that makes up this doctoral thesis.

During the development of this work, three different studies have been carried out. The first article proposes a distributed platform using drones, which has been designed to identify different damages in transport routes using a multi-agent system with PANGEA and ubiquitous computing. This work has employed the YOLOv4 classifier, achieving 95% accuracy. The dataset from this study has been made openly and freely available to the scientific community. The second paper focused on classifying different urban and rural terrains using object-based classifiers, GeoDMA, and multiple algorithms, resulting in the identification of vegetation, soil, asphalt, and buildings. The methodology presented in this second paper achieved 85% accuracy and a kappa coefficient of 0.76, providing an effective strategy for extracting vital geospatial data. In the third case, addressing inefficiencies in the manual collection of road damage data, in more detail, a solution has been proposed using UAVs combined with AI technologies, including the YOLO suite of algorithms. Results from the Chinese and Spanish datasets validated the high degree of accuracy of the approach for auto-detection of road damage.

To complement my knowledge and disseminate my results, I have attended numerous scientific events.

4.2. Future work

This section reviews the future lines of research that could be pursued in the future and constitutes the continuation of the studies developed.

First, all the initial objectives have been successfully achieved. However, it would be interesting to test the proposed system in different areas with a larger number of buildings.

In particular, the results of the first and third articles could easily be improved by increasing the number of image sets in the training model. Additionally, the use of a computationally more complex algorithm to identify patterns in the data could increase the performance of the system and thus the user experience. Not only does the result depend on the number of images alone, but also on the structural characteristics of the country of origin.

The second study could be extended to other cities, focusing on the detection of a different object, depending on the circumstances.

Furthermore, as this study is based on real data from two countries, it would be very interesting to collaborate with a local public authority. To improve our work, the possibility of testing other types of cameras, including multispectral cameras and LiDAR-based cameras, has been raised.

The results of artificial intelligence models could also be validated on datasets from underdeveloped countries, where the platform could have a crucial influence on the country's development. This technology could be applied not only to infrastructure control and maintenance but also to public health policies such as reservoir control or environmental measures for agricultural control and deforestation.

As discussed throughout the thesis, the social and real-life focus of the research application, together with the internationalization of the research, shows the potential of the contributions made.

Apéndice A

Factor de impacto de las revistas

A.1. Sensors

- **Nombre de la revista:** Sensors
- **Url:** <https://www.mdpi.com/journal/sensors>
- **Editorial:** MDPI
- **ISSN:** 1424-8220
- **Categorías:** ENGINEERING, ELECTRICAL & ELECTRONIC, INSTRUMENTS & INSTRUMENTATION, CHEMISTRY, ANALYTICAL
- **Factor de impacto 2020:** 3,576
- **Quartil 2020:**
 - ELECTRONIC, INSTRUMENTS & INSTRUMENTATION: Q1
 - ENGINEERING, ELECTRICAL: Q2
 - CHEMISTRY, ANALYTICAL: Q2
- **Ranking 2020:**
 - ELECTRONIC, INSTRUMENTS & INSTRUMENTATION: 14/64
 - ENGINEERING, ELECTRICAL: 82/273
 - CHEMISTRY, ANALYTICAL: 26/87

A.2. Sensors

- **Nombre de la revista:** Sensors
- **Url:** <https://www.mdpi.com/journal/sensors>
- **Editorial:** MDPI
- **ISSN:** 1424-8220
- **Categorías:** ENGINEERING, ELECTRICAL & ELECTRONIC, INSTRUMENTS & INSTRUMENTATION, CHEMISTRY, ANALYTICAL
- **Factor de impacto 2022:** 3,9
- **Quartil 2022:**

- ENGINEERING, ELECTRICAL: Q2
- ELECTRONIC, INSTRUMENTS & INSTRUMENTATION: Q2
- CHEMISTRY, ANALYTICAL: Q2

■ **Ranking 2022:**

- ENGINEERING, ELECTRICAL: 100/275
- ELECTRONIC, INSTRUMENTS & INSTRUMENTATION: 19/63
- CHEMISTRY, ANALYTICAL: 27/86

A.3. IEEE Access

■ **Nombre de la revista:** IEEE Access

■ **Url:** <https://ieeaccess.ieee.org/>

■ **Editorial:** IEEE

■ **ISSN:** 2169-3536

■ **Categorías:** COMPUTER SCIENCE, INFORMATION SYSTEMS & ENGINEERING, ELECTRICAL & ELECTRONIC & TELECOMMUNICATIONS

■ **Factor de impacto 2022:** 3,9

■ **Quartil 2022:**

- COMPUTER SCIENCE, INFORMATION SYSTEMS: Q2
- ENGINEERING, ELECTRICAL & ELECTRONIC: Q2
- TELECOMMUNICATIONS: Q2

■ **Ranking 2022:**

- COMPUTER SCIENCE, INFORMATION SYSTEMS: 79/250
- ENGINEERING, ELECTRICAL & ELECTRONIC: 109/349
- TELECOMMUNICATIONS: 42/116

Apéndice B

Conclusiones

Este último capítulo de la tesis doctoral resume las conclusiones extraídas de la investigación desarrollada durante los últimos años y presentada en los capítulos anteriores. El capítulo repasa los resultados y aportaciones de la investigación y describe las futuras líneas de investigación para los próximos años en relación con el tema de la tesis.

B.1. Contribuciones

Esta sección describe las principales contribuciones realizadas en este trabajo y los resultados más destacables de la investigación llevada a cabo en esta investigación.

Originalmente, se llevó a cabo una revisión exhaustiva del estado del arte sobre aplicaciones y protocolos de monitorización empleando imágenes aéreas, inteligencia artificial y aprendizaje profundo. Este análisis ha facilitado una comprensión más profunda de los méritos y desafíos inherentes a la monitorización inteligente de infraestructuras y ha aclarado la relación entre los vehículos aéreos no tripulados y las metodologías de aprendizaje profundo. Durante el estudio llevado a cabo, se ha creado un conjunto de datos específico que posteriormente se ha distribuido de una forma abierta a disposición de la comunidad científica.

Durante esta investigación se ha realizado un estudio profundo acerca de los protocolos existentes para determinar cuales de ellos eran los más favorables para ser empleados y poder satisfacer los requerimientos de los casos de estudio de los artículos de investigación. Adicionalmente, se evaluaron múltiples algoritmos con el objetivo de identificar las técnicas más efectivas para cada problemática. Estos algoritmos ya identificados han sido testados con diferentes conjuntos de datos con objeto de poder evaluar su comportamiento en la identificación de imágenes aéreas. Especialmente, se examinaron las distintas versiones de los algoritmos YOLOv5 y YOLOv7, y se hizo foco en los algoritmos de recomendación y sistemas de control con objeto de poder determinar el mejor candidato para cada caso de estudio.

En el desarrollo de los tres casos de estudio contemplados en este trabajo, se han empleado diferentes técnicas de procesamiento de imágenes para la mejora y la aplicación efectiva de los algoritmos, con un enfoque notable en la ampliación de datos (*data augmentation*) y con objeto de enriquecer el conjunto de datos. Se puede destacar también, en una de nuestras publicaciones, que se ha utilizado un conjunto de datos de detección colectiva, validando la eficacia de la plataforma.

Basándonos en las revisiones exhaustivas llevadas a cabo en esta investigación, se desarrolló un sistema que alcanzó los siguientes hitos: (I) Un examen exhaustivo

del estado actual de las tecnologías de inteligencia artificial y aprendizaje profundo pertinentes al análisis de imágenes aéreas. Esta revisión destacó las principales oportunidades y desafíos actuales dentro del ámbito de la monitorización y mantenimiento de infraestructuras urbanas. (II) El diseño, desarrollo y validación de modelos de aprendizaje profundo demostraron su eficacia en la detección de daños en infraestructuras urbanas mediante imágenes aéreas. (III) Se llevó a cabo la exploración y establecimiento de arquitecturas inteligentes. Estas arquitecturas, adaptadas para la detección de daños en infraestructuras, reflejaron adaptabilidad al responder a cambios o anomalías identificados. (IV) Se estableció una red avanzada para el procesamiento de imágenes aéreas. Junto con *Edge Computing*, esta red demostró habilidad para identificar y clasificar rápidamente posibles daños en infraestructuras. (V) Se entrenó y evaluó un modelo de aprendizaje profundo. El modelo se convirtió en un recurso esencial para la toma de decisiones en la monitorización automatizada de infraestructuras. (VI) La investigación llevó a cabo el examen de tres casos de estudio distintos que abarcan conjuntos de datos de daños en carreteras de España y China, así como imágenes urbanas de Brasil. Este análisis subrayó la practicidad y potencialidad de las técnicas de inteligencia artificial y aprendizaje profundo en la detección, prevención y reparación de infraestructuras urbanas.

Estos casos de estudio ejemplifican las aplicaciones prácticas y el potencial de las metodologías y técnicas propuestas en esta investigación. Las explicaciones detalladas de estos caso de estudio son expuestas en cada artículo científico que conforma esta tesis doctoral.

Durante el desarrollo de este trabajo, se han llevado a cabo tres estudios diferentes. El primer artículo, propone una plataforma distribuida utilizando drones que ha sido diseñada para identificar diferentes desperfectos en rutas de transporte utilizando un sistema multiagente con PANGEA y computación ubicua. Este trabajo ha empleado el clasificador YOLOv4, consiguiendo un 95 % de precisión. El conjunto de datos de este estudio ha sido puesto a disposición de la comunidad científica de una forma abierta y gratuita. El segundo trabajo, se centró en clasificar diferentes terrenos urbanos y rurales utilizando clasificadores basados en objetos, GeoDMA y múltiples algoritmos y cuyo resultado permitió identificar vegetación, suelo, asfalto y edificios. La metodología presentada en este segundo artículo logró un 85 % de precisión y un coeficiente kappa del 76 %, ofreciendo una estrategia efectiva para extraer datos geospaciales vitales. El tercer caso, aborda las ineficiencias en la recopilación manual de datos de daños en carreteras, más en detalle, se ha propuesto una solución utilizando UAVs combinados con tecnologías de IA, incluyendo la serie de algoritmos YOLO. Los resultados de los conjuntos de datos de China y España permitieron validar el alto grado de precisión del enfoque para la auto-detección de daños en carreteras.

Finalmente y con objeto de cumplimentar mis conocimientos y divulgar mis resultados he asistido a numerosos eventos científicos.

B.2. Trabajos futuros

Esta sección repasa las líneas futuras de investigación que podría seguir en un futuro y que constituye la continuación de los estudios desarrollados.

En primer lugar, se han culminado de forma exitosa todos los objetivos planteados inicialmente. Sin embargo, sería realmente interesante probar el sistema propuesto en diferentes áreas con un mayor número de edificios.

En concreto, los resultados del primer y tercer artículo podrían mejorarse fácilmente aumentando el número de conjuntos de imágenes en el modelo de entrenamiento. De forma adicional, el uso de un algoritmo computacionalmente más complejo para identificar los patrones en los datos podría aumentar los resultados del sistema y, por lo tanto, la experiencia de los usuarios. El resultado esperado no depende únicamente del número de imágenes, sino también de las características estructurales del país de origen.

El segundo estudio podría ampliarse extendiéndolo a otras ciudades centrándose en la detección de un objeto diferente, según la circunstancia.

Además, dado que este estudio se basa en datos reales de varios países, sería muy interesante colaborar con una autoridad local pública. Para mejorar nuestro trabajo, se ha planteado la posibilidad de probar otro tipo de cámaras, incluyendo cámaras multiespectrales y cámaras basadas en LiDAR.

También se podría validar los resultados de los modelos de inteligencia artificial en conjuntos de datos de países subdesarrollados, donde la plataforma podría tener una influencia crucial en el desarrollo del país. Esta tecnología podría aplicarse no sólo al control y mantenimiento de infraestructuras, sino también a políticas de salud pública como el control de embalses o medidas medioambientales para el control agrícola y la deforestación.

Como se ha expuesto a lo largo de la tesis, el enfoque social y real de la aplicación de la investigación, junto con la internacionalización de la misma, muestra el potencial de las contribuciones realizadas.

Bibliografía

- [1] Giacomo Chiesa et al. «A fuzzy-logic IoT lighting and shading control system for smart buildings». En: *Automation in Construction* 120 (dic. de 2020). ISSN: 09265805. DOI: [10.1016/j.autcon.2020.103397](https://doi.org/10.1016/j.autcon.2020.103397).
- [2] John Harvey et al. *Pavement, Roadway, and Bridge Life Cycle Assessment 2020: Proceedings of the International Symposium on Pavement, Roadway, and Bridge Life Cycle Assessment 2020 (LCA 2020, Sacramento, CA, 3-6 June 2020)*. CRC Press, 2020.
- [3] Xin She et al. «Feasibility study of asphalt pavement pothole properties measurement using 3D line laser technology». En: *International Journal of Transportation Science and Technology* (2020). ISSN: 20460449. DOI: [10.1016/j.ijstst.2020.07.004](https://doi.org/10.1016/j.ijstst.2020.07.004).
- [4] Hsiu Wen Wang et al. «A Real-Time Pothole Detection Approach for Intelligent Transportation System». English. En: *Mathematical Problems in Engineering* (2015). ISSN: 1024-123X. DOI: [10.1155/2015/869627](https://doi.org/10.1155/2015/869627).
- [5] Wenju Li, Zihao Shen y Peigang Li. «Crack Detection of Track Plate Based on YOLO». En: *Proceedings - 2019 12th International Symposium on Computational Intelligence and Design, ISCID 2019* (2019), págs. 15-18. DOI: [10.1109/ISCID.2019.10086](https://doi.org/10.1109/ISCID.2019.10086).
- [6] Jinyong Kim et al. «CBDN: Cloud-Based Drone Navigation for Efficient Battery Charging in Drone Networks». En: *IEEE Transactions on Intelligent Transportation Systems* 20.11 (2019), págs. 4174-4191. ISSN: 15580016. DOI: [10.1109/TITS.2018.2883058](https://doi.org/10.1109/TITS.2018.2883058).
- [7] Seyit Alperen Celtek, Akif Durdu y Ender Kurnaz. «Design and Simulation of the Hierarchical Tree Topology Based Wireless Drone Networks». En: *2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018*. Institute of Electrical y Electronics Engineers Inc., 2019. ISBN: 9781538668788. DOI: [10.1109/IDAP.2018.8620755](https://doi.org/10.1109/IDAP.2018.8620755).
- [8] Desheng Wang et al. «Routing and Scheduling for Hybrid Truck-Drone Collaborative Parcel Delivery With Independent and Truck-Carried Drones». En: *IEEE Internet of Things Journal* 6.6 (2019), págs. 10483-10495. ISSN: 23274662. DOI: [10.1109/JIOT.2019.2939397](https://doi.org/10.1109/JIOT.2019.2939397).
- [9] Menghua Zhai et al. «Predicting Ground-Level Scene Layout from Aerial Imagery». En: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 2017, págs. 4132-4140. ISBN: 978-1-5386-0457-1. DOI: [10.1109/CVPR.2017.440](https://doi.org/10.1109/CVPR.2017.440). URL: <http://doi.ieeecomputersociety.org/10.1109/CVPR.2017.440>.
- [10] Gang Fu et al. «Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network». En: *Remote Sensing* 9.5 (2017). ISSN: 2072-4292. DOI: [10.3390/rs9050498](https://doi.org/10.3390/rs9050498). URL: <https://www.mdpi.com/2072-4292/9/5/498>.

- [11] Lei Ma et al. «Deep learning in remote sensing applications: A meta-analysis and review». En: *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (2019), págs. 166-177. ISSN: 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2019.04.015>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271619301108>.
- [12] D. Deepa y A. Sivasangari. «An effective detection and classification of road damages using hybrid deep learning framework». En: *Multimed Tools Appl* 82 (2023), págs. 18151-18184. DOI: <https://doi.org/10.1007/s11042-022-14001-9>.
- [13] Marco Sozzi et al. «Economic Comparison of Satellite, Plane and UAV-Acquired NDVI Images for Site-Specific Nitrogen Application: Observations from Italy». En: *Agronomy* 11.11 (2021). ISSN: 2073-4395. DOI: [10.3390/agronomy11112098](https://doi.org/10.3390/agronomy11112098). URL: <https://www.mdpi.com/2073-4395/11/11/2098>.
- [14] Eugen Valentin Butilă y Răzvan Gabriel Boboc. «Urban Traffic Monitoring and Analysis Using Unmanned Aerial Vehicles (UAVs): A Systematic Literature Review». En: *Remote Sensing* 14.3 (2022). ISSN: 2072-4292. DOI: [10.3390/rs14030620](https://doi.org/10.3390/rs14030620). URL: <https://www.mdpi.com/2072-4292/14/3/620>.
- [15] Héctor Sánchez San Blas et al. «A Platform for Swimming Pool Detection and Legal Verification Using a Multi-Agent System and Remote Image Sensing». En: *International Journal of Interactive Multimedia and Artificial Intelligence In Press*. In Press (2023), págs. 1-13. ISSN: 1989-1660. DOI: [10.9781/ijimai.2023.01.002](https://doi.org/10.9781/ijimai.2023.01.002).
- [16] Victoria J. Hodge, Richard Hawkins y Rob Alexander. «Deep reinforcement learning for drone navigation using sensor data». En: *Neural Computing and Applications* (2020), págs. 1-19. ISSN: 14333058. DOI: [10.1007/s00521-020-05097-x](https://doi.org/10.1007/s00521-020-05097-x). URL: <https://doi.org/10.1007/s00521-020-05097-x>.
- [17] Anastasiia Safonova et al. «Detection of Norway Spruce Trees (Picea Abies) Infested by Bark Beetle in UAV Images Using YOLOs Architectures». En: *IEEE Access* 10 (2022), págs. 10384-10392. DOI: [10.1109/ACCESS.2022.3144433](https://doi.org/10.1109/ACCESS.2022.3144433).
- [18] David Gallacher. «Drones to manage the urban environment: Risks, rewards, alternatives». En: *Journal of Unmanned Vehicle Systems* 4.2 (jun. de 2016), págs. 115-124. ISSN: 2291-3467. DOI: [10.1139/juvs-2015-0040](https://doi.org/10.1139/juvs-2015-0040).
- [19] Luis Augusto Silva et al. «Active Actions in the Extraction of Urban Objects for Information Quality and Knowledge Recommendation with Machine Learning». En: *Sensors-basel*. 23.1 (dic. de 2022), pág. 138. ISSN: 1424-8220. DOI: [10.3390/s23010138](https://doi.org/10.3390/s23010138). URL: <https://doi.org/10.3390/s23010138>.
- [20] Wesley L. Passos et al. «Toward improved surveillance of Aedes aegypti breeding grounds through artificially augmented data». En: *Engineering Applications of Artificial Intelligence* 123 (2023), pág. 106488. ISSN: 0952-1976. DOI: <https://doi.org/10.1016/j.engappai.2023.106488>. URL: <https://www.sciencedirect.com/science/article/pii/S0952197623006723>.
- [21] Peter Reason e Hilary Bradbury. *The SAGE Handbook of Action Research*. SAGE Publications Ltd, mayo de 2012. DOI: [10.4135/9781848607934](https://doi.org/10.4135/9781848607934).
- [22] Hazim Shakhatreh et al. «Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges». En: *IEEE Access* 7 (2019), págs. 48572-48634. DOI: [10.1109/ACCESS.2019.2909530](https://doi.org/10.1109/ACCESS.2019.2909530).

- [23] Barry Allred et al. «Overall results and key findings on the use of UAV visible-color, multispectral, and thermal infrared imagery to map agricultural drainage pipes». En: *Agricultural Water Management* 232 (2020), pág. 106036. ISSN: 0378-3774. DOI: <https://doi.org/10.1016/j.agwat.2020.106036>. URL: <https://www.sciencedirect.com/science/article/pii/S0378377419317597>.
- [24] Anwar. M. Fanan, Meftah. A. Mehdawi y J. C. Murray. «Detecting and filtering multiple drone controller signals from background noise using bearing amplitude data». En: *2021 29th Telecommunications Forum (TELFOR)*. 2021, págs. 1-4. DOI: [10.1109/TELFOR52709.2021.9653345](https://doi.org/10.1109/TELFOR52709.2021.9653345).
- [25] Francisco García Encinas et al. «Singular Spectrum Analysis for Source Separation in Drone-Based Audio Recording». En: *IEEE Access* 9 (2021), págs. 43444-43457. DOI: [10.1109/ACCESS.2021.3065775](https://doi.org/10.1109/ACCESS.2021.3065775).
- [26] Azade Fotouhi, Ming Ding y Mahbub Hassan. «Understanding autonomous drone maneuverability for Internet of Things applications». En: *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2017, págs. 1-6. DOI: [10.1109/WoWMoM.2017.7974336](https://doi.org/10.1109/WoWMoM.2017.7974336).
- [27] Riccardo Bassoli et al. «A Virtualized Border Control System based on UAVs: Design and Energy Efficiency Considerations». En: *2019 IEEE Aerospace Conference*. 2019, págs. 1-11. DOI: [10.1109/AERO.2019.8742142](https://doi.org/10.1109/AERO.2019.8742142).
- [28] Yuting Wan et al. «An accurate UAV 3-D path planning method for disaster emergency response based on an improved multiobjective swarm intelligence algorithm». En: *IEEE Transactions on Cybernetics* 53.4 (2022), págs. 2658-2671.
- [29] Luis Felipe Aguirre Siachoque. «Modelamiento térmico 3D aplicado a la inspección de infraestructuras usando vehículos aéreos no tripulados». En: *Memorias del Congreso Internacional I+ D+ i-Sostenibilidad Energética*. Vol. 1. 1. 2019, págs. 1-20.
- [30] E.K. Raptis, M. Krestenitis, K. Egglezos et al. «End-to-end Precision Agriculture UAV-Based Functionalities Tailored to Field Characteristics». En: *J Intell Robot Syst* 107 (2023), pág. 23. DOI: [10.1007/s10846-022-01761-7](https://doi.org/10.1007/s10846-022-01761-7).
- [31] Mouna Elloumi et al. «Monitoring road traffic with a UAV-based system». En: *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. 2018, págs. 1-6. DOI: [10.1109/WCNC.2018.8377077](https://doi.org/10.1109/WCNC.2018.8377077).
- [32] María del Carmen Girón Tomás. «El uso de la tecnología en los vehículos a motor y en los RPAS para la mejora de la seguridad vial y su posible uso al servicio de la seguridad pública. Las nuevas versiones de la ética pública: carética y dronética». En: *Revista de Derecho de la UNED (RDUNED)* 26 (2020), 195-226. DOI: [10.5944/rduned.26.2020.29171](https://doi.org/10.5944/rduned.26.2020.29171). URL: <https://revistas.uned.es/index.php/RDUNED/article/view/29171>.
- [33] Brenden Herkenhoff et al. «Integration of Drones with 5G Connectivity to Airfields for Enhancing Mission Readiness and Structural Health Monitoring». En: *AIAA AVIATION 2023 Forum*. 2023, pág. 3784.
- [34] Stefano Frizzo Stefenon et al. «Classification of distribution power grid structures using inception v3 deep neural network». En: *Electrical Engineering* 104.6 (2022), págs. 4557-4569.
- [35] Nicolai Iversen et al. «Design, Integration and Implementation of an Intelligent and Self-recharging Drone System for Autonomous Power line Inspection». En: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2021, págs. 4168-4175. DOI: [10.1109/IROS51168.2021.9635924](https://doi.org/10.1109/IROS51168.2021.9635924).

- [36] Yan Wang et al. «MAda-Net: Model-Adaptive Deep Learning Imaging for SAR Tomography». En: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), págs. 1-13. DOI: [10.1109/TGRS.2023.3239405](https://doi.org/10.1109/TGRS.2023.3239405).
- [37] Shubo Wang et al. «A Deep-Learning-Based Sea Search and Rescue Algorithm by UAV Remote Sensing». En: *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. 2018, págs. 1-5. DOI: [10.1109/GNCC42960.2018.9019134](https://doi.org/10.1109/GNCC42960.2018.9019134).
- [38] Shih-Hong Chio y Cheng-Chu Chiang. «Feasibility Study Using UAV Aerial Photogrammetry for a Boundary Verification Survey of a Digitalized Cadastral Area in an Urban City of Taiwan». En: *Remote Sensing* 12.10 (2020). ISSN: 2072-4292. URL: <https://www.mdpi.com/2072-4292/12/10/1682>.
- [39] Y. Lecun et al. «Gradient-based learning applied to document recognition». En: *Proceedings of the IEEE* 86.11 (1998), págs. 2278-2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [40] Alex Krizhevsky, Ilya Sutskever y Geoffrey E Hinton. «Imagenet classification with deep convolutional neural networks». En: *Advances in neural information processing systems*. 2012, págs. 1097-1105.
- [41] Kaiming He et al. «Deep Residual Learning for Image Recognition». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [42] Jiuxiang Gu et al. «Recent advances in convolutional neural networks». En: *Pattern Recognition* 77 (2018), págs. 354-377. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2017.10.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0031320317304120>.
- [43] Joseph Redmon et al. «You only look once: Unified, real-time object detection». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 779-788.
- [44] Joseph Redmon y Ali Farhadi. *YOLO9000: Better, Faster, Stronger*. Inf. téc. URL: <http://pjreddie.com/yolo9000/>.
- [45] Ross Girshick et al. «Rich feature hierarchies for accurate object detection and semantic segmentation». En: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (2014), págs. 580-587. ISSN: 10636919. DOI: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81). URL: <http://arxiv..>
- [46] Ross Girshick. «Fast R-CNN». En: *Proceedings of the IEEE International Conference on Computer Vision* 2015 Inter (2015), págs. 1440-1448. ISSN: 15505499. DOI: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
- [47] Hubert Ramsperger. «Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks». En: *Total Performance Scorecard* (2020), págs. 159-183. ISSN: 0013-1954. DOI: [10.4324/9780080519340-12](https://doi.org/10.4324/9780080519340-12).
- [48] Piotr Doll, Ross Girshick y Facebook Ai. «Mask R-CNN». En: ().
- [49] Yuxin Wu et al. *Detectron2*. 2019. URL: <https://github.com/facebookresearch/detectron2>.
- [50] Wei Liu et al. «SSD: Single Shot MultiBox Detector». En: (dic. de 2015). DOI: [10.1007/978-3-319-46448-0_{_}2](https://doi.org/10.1007/978-3-319-46448-0_{_}2). URL: <http://arxiv.org/abs/1512.02325>https://dx.doi.org/10.1007/978-3-319-46448-0_2.
- [51] Joseph Redmon y Ali Farhadi. «YOLO9000: Better, faster, stronger». En: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017), págs. 6517-6525. DOI: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).

-
- [52] Joseph Redmon y Ali Farhadi. «YOLOv3: An Incremental Improvement». En: (2018). URL: <http://arxiv.org/abs/1804.02767>.
- [53] Alexey Bochkovskiy, Chien-Yao Wang y Hong-Yuan Mark Liao. «YOLOv4: Optimal Speed and Accuracy of Object Detection». En: (abr. de 2020). URL: <http://arxiv.org/abs/2004.10934>.
- [54] Chien-Yao Wang, Alexey Bochkovskiy y Hong-Yuan Mark Liao. *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. 2022. arXiv: [2207.02696](https://arxiv.org/abs/2207.02696) [cs.CV].