

CENTRO DE ENSEÑANZA TÉCNICA Y SUPERIOR



Colegio de Ingeniería
Dirección de Posgrado
Campus Mexicali

Tesis

**Navegación Autónoma de Vehículos Aéreos por
Referencia Visual**

para obtener el grado de

Maestría en Ingeniería e Innovación

Presenta

Francisco Armando Castillo Vazquez

Director de Tesis: Dr. Ulises Orozco Rosas

Co-directora de Tesis: Dra. Kenia Picos Espinoza

Mexicali, Baja California. Diciembre de 2023

CENTRO DE ENSEÑANZA TÉCNICA Y SUPERIOR



Colegio de Ingeniería
Dirección de Posgrado
Campus Mexicali

Tesis

**Navegación Autónoma de Vehículos Aéreos por
Referencia Visual**

para obtener el grado de

Maestría en Ingeniería e Innovación

Presenta

Francisco Armando Castillo Vazquez

Comité evaluador:

Dr. Ulises Orozco
Rosas

Director de Tesis

Dra. Kenia Picos
Espinoza

Co-directora de Tesis

Dra. Dania Licea
Verduzco

Miembro del Comité

Dra. Karla Garduño
Palomino

Miembro del Comité

Mexicali, Baja California. Diciembre de 2023

Agradecimiento

Agradezco sinceramente al Dr. Ulises Orozco, mi director de tesis, por haber aceptado dirigir esta investigación. Su orientación constante, apoyo y aliento en cada entrega de avances fueron fundamentales para el éxito de este proyecto. También quiero expresar mi gratitud a la Dra. Kenia Picos por su valioso respaldo y recomendaciones a lo largo de esta investigación, así como a la Dra. Dania Licea, quien siempre me alentó en cada sesión y al cumplimiento de este proyecto.

Dedicatoria

Este logro no habría sido posible sin el apoyo incondicional de mis padres y hermanos, quienes han sido una parte indispensable en mi desarrollo personal y profesional. Les dedico este proyecto con profundo agradecimiento. Asimismo, agradezco a mis amigos, quienes me brindaron su apoyo a través de consejos, ánimos y conversaciones enriquecedoras durante el trayecto de mi maestría. Su presencia ha sido invaluable y ha contribuido significativamente a mi crecimiento académico y personal.

Mexicali B.C, 27 de noviembre de 2023

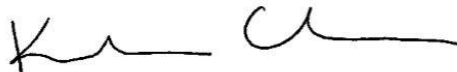
Dra. Dania Licea
CETYS Universidad

Estimada Dra. Dania Licea, anteponiendo un cordial saludo me permito informarle que el alumno **Castillo Vazquez Francisco Armando** perteneciente a la Maestría en Ingeniería e Innovación desarrolló el proyecto titulado “**Navegación autónoma de vehículos aéreos por referencia visual**” que tuvo como objetivo desarrollar y evaluar un sistema de navegación autónoma para vehículos aéreos no tripulados basado en referencia visual, haciendo uso de redes neuronales convolucionales para mejorar la percepción y toma de decisiones de los vehículos aéreos no tripulado en base a los puntos de referencia. Se obtuvieron los siguientes resultados:

- Desarrollo con éxito un sistema de navegación basado en ROS que facilita la comunicación y el control del vehículo aéreo, permitiendo la adquisición de imágenes.
- Desarrollo una red neuronal convolucional altamente eficiente para la clasificación de las imágenes adquiridas por el vehículo aéreo, lo que ha permitido tomar decisiones basadas en esta clasificación de manera efectiva.
- Implementación de la interacción entre el sistema de navegación del vehículo aéreo y la red neuronal convolucional ha sido exitosa, logrando una integración fluida que mejora significativamente la percepción y toma de decisiones del vehículo aéreo en función de los datos visuales recopilados.

Sirva la presente como evidencia para los fines académicos del Posgrado en Ingeniería de CETYS Universidad, campus Mexicali.

Atentamente



Dra. Karla Garduño Palomino
Coordinadora de Posgrado de Ingeniería
CETYS Universidad campus Mexicali

ÍNDICE

	Páginas
RESUMEN	1
1 Introducción	4
1.1 Antecedentes.....	5
1.2 Justificación.....	8
1.3 Planteamiento del problema.....	10
1.4 Preguntas de investigación	11
1.5 Objetivo general	11
1.6 Objetivos específicos.....	11
1.7 Hipótesis.....	12
2 Marco Teórico	13
2.1 Aprendizaje profundo	13
2.2 Mapeo de imágenes a acciones de control	17
2.3 Clasificación de imágenes	20
2.4 Funcionamiento del vehículo aéreo	26
2.5 Casos de uso	29
3 Propuesta Metodológica	37
3.1 Manipulación y obtención de imágenes	38
3.2 Arquitectura de clasificación de imágenes.....	42
4 Resultados	47
4.1 Resultados experimentales para reconocimiento de imágenes	47
4.2 Resultados experimentales para el sistema con la CNN.....	55
5 Conclusiones	59
5.1 Futuros trabajos.....	60
ANEXOS	65

ÍNDICE DE FIGURAS

	Páginas
2.1 Estructura básica de una neurona biológica.....	15
2.2 Arquitectura de una red neuronal artificial.....	16
2.3 Arquitectura LeNet-5.	24
2.4 Arquitectura VGG-19.....	25
3.1 Diagrama del sistema de navegación.....	38
3.2 Diagrama propuesto de la red neuronal convolucional para clasificación de imágenes.	43
4.1 Gráfica de precisión resultante del sistema propuesto de clasificación de imágenes.	49
4.2 Gráfica de perdida resultante del sistema propuesto de clasificación de imágenes.....	51
4.3 Ejemplo de una imagen de señalamiento utilizada en las pruebas de entrenamiento. 52	
4.4 Parámetros identificados en distintas capas para flecha.....	53
4.5 Ejemplo 2 de una imagen de señalamiento utilizada en las pruebas de entrenamiento.....	54
4.6 Parámetros identificados en distintas capas para rombo.	54
4.7 Diagrama del flujo de la misión del UAV.....	57

ÍNDICE DE TABLAS

	Páginas
3.1 Detalles de la arquitectura propuesta de la red neuronal convolucional para clasificación de imágenes	43
4.1 Tabla de imágenes y sus acciones utilizadas en el sistema propuesto	54
4.2 Ejemplo de imágenes de entrada adquiridas por el UAV y su respectivo comando de salida	56

LISTAS DE ABREVIATURAS

UAV: *Unmanned Aerial Vehicle*
CNN: *Convolutional Neural Networks*
ROS: *Robot Operating System*

RESUMEN

Esta tesis se centró en el diseño y desarrollo de un sistema de navegación autónoma para vehículos aéreos no tripulados UAV (*Unmanned Aerial Vehicle*), basado en referencia visual. En la actualidad, la demanda de sistemas autónomos para la navegación de UAV ha experimentado un crecimiento significativo, impulsado por su aplicación en diversas industrias, desde la agricultura hasta la vigilancia.

La investigación se apoya en el uso de redes neuronales convolucionales (CNN, del inglés *Convolutional Neural Networks*) para la clasificación de imágenes en diferentes tipos de señalamientos de tráfico y, en consecuencia, tomar decisiones de navegación en los UAV. Este enfoque representa una respuesta innovadora a los desafíos actuales relacionados con la percepción y toma de decisiones en entornos dinámicos y cambiantes. El sistema se estructura en torno a tres pilares principales: la configuración del entorno ROS (*Robot Operating System*) para la comunicación y control del UAV, la implementación de una CNN especializada para la clasificación de imágenes y la integración de esta red en el sistema de navegación autónoma. Además, se desarrollarán algoritmos para la obtención de imágenes y la manipulación del vehículo aéreo no tripulado basados en las salidas de la red neuronal convolucional.

La metodología se abordó en dos etapas interrelacionadas entre sí, desde el inicio de una configuración del entorno ROS hasta la evaluación exhaustiva del sistema con la CNN y su respuesta a las imágenes obtenidas por el UAV. Cada fase será documentada y analizada meticulosamente para asegurar el cumplimiento de los objetivos específicos y del objetivo general de la investigación.

Esta investigación responde a la demanda de sistemas de navegación eficientes para vehículos aéreos no tripulados (UAV), y aporta de manera destacada a la convergencia entre la visión por computadora y la robótica autónoma, otorgando especial importancia al uso de redes neuronales convolucionales.

Palabras Clave - Vehículos Aéreos No Tripulados, Redes Neuronales Convolucionales, Visión por Computadora, Aprendizaje Profundo, *Robot Operating System*.

ABSTRACT

This thesis focused on the development of an autonomous navigation system for Unmanned Aerial Vehicles (UAVs) based on visual reference. Currently, the demand for autonomous systems for UAV navigation has experienced significant growth, driven by their application in various industries, from agriculture to surveillance.

The study hinges on employing a Convolutional Neural Networks (CNNs) to classify images portraying various types traffic signals, empowering the UAV to make navigation decisions. This approach represents a pioneering solution to address current challenges linked to perception and decision-making in dynamic and evolving environments.

The structured system around three main pillars: the configuration of Robot Operating System (ROS) for UAV communication and control, the implementation of a specialized CNN for image classification, and the integration of this network into the autonomous navigation system. Additionally, algorithms will be developed for image acquisition and UAV manipulation based on the outputs of the convolutional neural network.

The methodology was approached with two interrelated stages, in the initial ROS to environment configuration to the comprehensive evaluation of the system with the CNN and its response to images obtained by the UAV. Each phase will be documented and analyzed meticulously to ensure the fulfillment of specific objectives and the overall goal of the research.

This research addresses the demand for efficient navigation systems for UAVs and contributes significantly to the convergence of computer vision and autonomous robotics, placing special emphasis on the use of convolutional neural networks.

Keywords - Unmanned Aerial Vehicles, Convolutional Neural Networks, Computer Vision, Deep Learning, Robot Operating System.

Capítulo 1

Introducción

En la era actual de la tecnología aeroespacial, la navegación autónoma de vehículos aéreos no tripulados (UAV) se ha convertido en un campo de investigación crucial con aplicaciones potenciales en distintas industrias, desde la cartografía y la vigilancia hasta la entrega de mercancías y la respuesta a emergencias. En este contexto, la utilización de referencias visuales y el empleo de redes neuronales convolucionales (CNN) representan avances significativos en la mejora de la percepción y toma de decisiones de los UAV, esenciales para lograr un vuelo autónomo, preciso y eficiente.

El estudio se enfoca en el desarrollo y evaluación de un sistema de navegación autónoma para UAV basado en referencia visual, aprovechando las capacidades de las CNN. La premisa fundamental es que esta combinación permitirá a los UAV interpretar y utilizar la información visual adquirida por la cámara, permitiendo así una mayor autonomía y capacidad de respuesta en entornos cambiantes y desafiantes.

Para este objetivo, se propone la implementación de un sistema de comunicación y control basado en el *Robot Operating System* (ROS). Este enfoque permitirá la captura eficiente de imágenes y la interacción con la Convolutional Neural Network (CNN). La configuración de este entorno de trabajo se llevará a cabo en el sistema operativo Linux, utilizando Python como lenguaje de programación. Este lenguaje es reconocido por su versatilidad, que será aprovechado gracias a su amplia variedad de librerías y comandos diseñados para optimizar la interacción del UAV en este contexto específico.

Adicionalmente, se propone la elaboración de una red neuronal convolucional que se especializa en la clasificación detallada de imágenes que contengan señalamientos de tráfico. Este proyecto incorporará el potencial de las bibliotecas de Python, que se destacan por sus capacidades para gestionar eficientemente grandes volúmenes de datos, manipulaciones avanzadas en imágenes y generar representaciones gráficas claras que

facilitan la comprensión de los resultados.

La utilización de estas bibliotecas en el desarrollo de la red neuronal no solo optimizará el proceso de entrenamiento y evaluación del modelo, sino que también permitirá una exploración más profunda de las características específicas de las imágenes relacionadas con los señalamientos de tráfico. De esta manera, se espera obtener resultados más precisos y comprensibles, contribuyendo así al avance significativo de la tarea de clasificación en el contexto específico de la investigación.

A lo largo de este trabajo, se detallará el proceso de desarrollo, implementación y evaluación de este sistema, así como los resultados obtenidos a partir de pruebas rigurosas en diferentes escenarios. Estos hallazgos no solo respaldarán la hipótesis planteada, sino que también ofrecerán perspectivas para el futuro desarrollo y aplicación de tecnologías similares en la industria de vehículos aéreos no tripulados.

1.1 Antecedentes

En la última década, el campo de la robótica y la aviación ha presenciado un rápido avance en la tecnología de vehículos aéreos no tripulados (UAV) (Beul et al., 2020). Estos dispositivos han encontrado aplicaciones en diversas industrias, como la inspección de infraestructuras, la vigilancia, la agricultura de precisión y la respuesta a desastres. Sin embargo, la navegación autónoma de UAV sigue siendo un desafío crítico, particularmente en entornos complejos donde la precisión y la confiabilidad son esenciales. En este contexto, la navegación autónoma por referencia visual ha surgido como una solución prometedora.

La navegación autónoma por referencia visual aprovecha las capacidades de percepción visual de los UAV para analizar y comprender su entorno (Zhu and Mourikis, 2017). En lugar de depender exclusivamente de sistemas de posicionamiento global (GPS) u otros sensores inerciales, esta metodología utiliza cámaras a bordo para capturar información visual del entorno circundante. Las imágenes obtenidas son luego procesadas para extraer características clave, como puntos de referencia, bordes y objetos, que ayudan a

comprender la topografía y la estructura del entorno.

En esta investigación, se abordará el desafío de la navegación autónoma de UAV mediante el empleo de redes neuronales convolucionales, que son una clase de algoritmos de aprendizaje profundo especialmente diseñados para el procesamiento de imágenes y la detección de patrones. Estas redes han demostrado una capacidad excepcional para realizar tareas como la detección de objetos, el reconocimiento de patrones y la segmentación semántica en imágenes visuales.

En el contexto de los antecedentes para la aplicación de redes neuronales convolucionales en vehículos aéreos no tripulados, se destaca la contribución de Loretta, quien propone una arquitectura de sistema basada en CNN para la detección y segmentación de carreteras en imágenes aéreas capturadas por un UAV (LeCun et al., 2015).

Este enfoque se distingue por su abordaje en dos fases: aprendizaje y operación. Las imágenes aéreas son descompuestas en sus componentes de color y procesadas en Matlab, utilizando un algoritmo de caja deslizante para segmentarlas en pequeñas unidades (LeCun et al., 2015).

La CNN, diseñada con MatConvNet, comprende capas convolucionales, de agrupación, ReLu, totalmente conectada y Softmax. El sistema entrenado con 2,000 cajas exhibe resultados prometedores, destacando por su velocidad de procesamiento y simplicidad. Este trabajo, centrado en la detección de carreteras desde altitudes, aporta a la comprensión de la aplicación de CNN en UAV, especialmente en el procesamiento eficiente de imágenes aéreas para tareas específicas (LeCun et al., 2015).

Por otro lado, Yeon Ji Choi presenta un enfoque innovador que capacita a los UAV para realizar la navegación autónoma en entornos interiores sin señal GPS, centrándose específicamente en la tarea desafiante de ascender escaleras. La propuesta integra una cámara monocromática y un sensor de detección y rango (LiDAR), utilizando un avanzado modelo de red neuronal convolucional (CNN) para el procesamiento de imágenes en tiempo real (Choi et al., 2021).

La singularidad de este enfoque radica en la utilización del flujo de video de la cámara frontal del UAV para que la red neuronal profunda clasifique las imágenes como escaleras o no escaleras, mientras que los datos del sensor LiDAR se emplean para calcular distancias. El sistema, implementado y evaluado en un UAV Parrot Bebop 2 con resultados prometedores, destaca por su capacidad de reconocimiento de escaleras del 92.06% y la ejecución exitosa de ascensos sin colisiones en pruebas reales (Choi et al., 2021).

También existe otro trabajo que contribuye al área del desarrollo de UAV capaces de reconocer y superar obstáculos complejos como escaleras en entornos interiores. La combinación de técnicas de procesamiento de imágenes basadas en CNN y la información de distancia proporcionada por LiDAR ofrece una solución efectiva y novedosa para la navegación autónoma en situaciones desafiantes (Choi et al., 2021).

En la búsqueda de equilibrios ecológicos en parques naturales mediante censos de animales realizados con vehículos aéreos no tripulados, Benjamin Kellenberger propone un enfoque para mejorar la precisión de los detectores de vida silvestre en imágenes. Destacando el uso de redes neuronales convolucionales junto con métodos de detección de objetos, Kellenberger aborda el desafío de la transferencia de dominio entre conjuntos de datos de UAV en diferentes años (Kellenberger et al., 2018).

Su método, basado en transporte óptimo, adapta un detector CNN entrenado en un conjunto de datos de origen a un conjunto de datos de destino, logrando una mejora en la precisión sin requerir etiquetas adicionales del nuevo conjunto de datos. La implementación se evalúa en imágenes de una reserva de juegos en Namibia, demostrando la capacidad de la CNN para adaptarse efectivamente a cambios en condiciones de adquisición de imágenes, lo que sugiere aplicaciones prácticas en la gestión de poblaciones animales en entornos naturales a través de UAV (Kellenberger et al., 2018).

Este otro estudio, se explora la aplicación de vehículos aéreos no tripulados junto con CNN para la detección automática de áreas verdes en imágenes de UAV. La metodología

propuesta utiliza CNNs entrenadas en puntos que representan la presencia o ausencia de árboles, logrando una segmentación precisa a nivel de superpíxeles. La fusión de esta segmentación con un método basado en contorno activo y el análisis de textura permiten distinguir árboles de otras vegetaciones, con resultados experimentales que demuestran la eficacia del enfoque propuesto con errores mínimos (Fikri et al., 2019).

En el ámbito del aprendizaje supervisado con redes neuronales, el artículo destaca las etapas de entrenamiento y evaluación, enfocándose en la propagación hacia adelante y la actualización de pesos y sesgos para optimizar la precisión del modelo. La CNN implementada logra una notable precisión del 90.35% en la detección de la presencia o ausencia de árboles mediante procesos de convolución (Fikri et al., 2019).

Otro ejemplo de avance significativo en este campo es la arquitectura InceptionNet, introducida por Szegedy. Esta red utiliza múltiples tamaños de filtros en paralelo, permitiendo así la detección de características a múltiples escalas. Esta innovación ha demostrado mejoras sustanciales en la precisión de clasificación de imágenes (Szegedy et al., 2016).

1.2 Justificación

La motivación de esta investigación radica en mejorar las limitaciones actuales en la autonomía y operatividad de los vehículos aéreos no tripulados, con un enfoque particular en el modelo Parrot Bebop 2. La necesidad de liberar a estos dispositivos de la dependencia exclusiva de comandos externos a través de controles remotos se hace evidente para mejorar su capacidad de operar de manera autónoma y adaptarse a entornos cambiantes.

La autonomía mejorada no es solo un objetivo tecnológico; es crucial para optimizarla eficiencia operativa de los UAV en diversas situaciones. Capacitar a los UAV para tomar decisiones autónomas en tiempo real, basadas en la interpretación directa de su entorno a través de técnicas de procesamiento visual avanzado, tiene el potencial de transformar misiones críticas como búsqueda y rescate, vigilancia y reconocimiento, así como la entrega autónoma en lugares de difícil acceso.

Además, abordar las limitaciones de la comunicación inalámbrica entre los controles remotos y los UAV es esencial. Superar las restricciones geográficas y las interferencias electromagnéticas que pueden afectar el alcance de la comunicación busca proporcionar a los UAV una mayor libertad operativa y capacidad de maniobra eficiente en diversos escenarios.

La contribución al desarrollo tecnológico es un aspecto destacado de esta investigación. La aplicación práctica de referencia visual y redes neuronales convolucionales en el Parrot Bebop 2 no solo responde a desafíos específicos de la autonomía UAV, sino que también abre perspectivas para futuras innovaciones en la aplicación de técnicas avanzadas de visión por computadora y aprendizaje profundo en sistemas aéreos no tripulados.

Los resultados de esta investigación no solo beneficiarán la eficiencia operativa en la industria, sino que también pueden tener impactos positivos en la sociedad. Desde operaciones de respuesta rápida en desastres naturales hasta la optimización de la entrega de suministros en áreas remotas, los avances tecnológicos pueden traducirse en mejoras tangibles en la seguridad, la eficiencia y la capacidad de respuesta en diversas aplicaciones prácticas.

1.3 Planteamiento del problema

La navegación autónoma de vehículos aéreos no tripulados representa un campo en constante evolución, enfrentando desafíos que van más allá de la mera capacidad de percepción visual. En este contexto, el Parrot Bebop 2, un UAV ampliamente utilizado en

diversas aplicaciones, presenta un paradigma de control que depende tradicionalmente de la intervención humana mediante controles remotos. Esta dependencia, aunque común, conlleva limitaciones críticas en términos de autonomía y eficiencia operativa.

El control del Parrot Bebop 2 se realiza típicamente a través de dispositivos remotos, estableciendo una conexión inalámbrica para la transmisión de comandos y la recepción de datos. Sin embargo, esta comunicación presenta restricciones que afectan la autonomía y la velocidad de respuesta del UAV. El alcance de distancia entre el control remoto y el Parrot Bebop 2, aunque significativo, puede ser limitado en entornos con interferencias electromagnéticas o topografías complejas. Este factor impone restricciones geográficas y operativas, comprometiendo la utilidad del UAV en escenarios donde la cobertura extensa es esencial.

Además, la latencia en la comunicación entre el control remoto y el UAV introduce un intervalo de tiempo entre la emisión de un comando y la ejecución correspondiente por parte del vehículo aéreo. Este retraso, aunque aparentemente breve, puede ser crítico en situaciones que demandan respuestas rápidas y precisas. Por ejemplo, en operaciones de búsqueda y rescate o en misiones de vigilancia, la capacidad de maniobra instantánea puede marcar la diferencia entre el éxito y el fracaso.

El Parrot Bebop 2, al depender en gran medida de comandos externos para su navegación, se ve afectado por la incapacidad de tomar decisiones autónomas en tiempo real basadas

en la interpretación directa de su entorno. La necesidad de una conexión constante con un control remoto limita la autonomía y dificulta la adaptación a entornos impredecibles o en evolución.

Este problema se agrava por la dependencia de sensores tradicionales, como el GPS, cuya precisión puede verse afectada por diversas condiciones ambientales. En consecuencia, la autonomía de los UAV, como el Parrot Bebop 2, se ve comprometida, creando un escenario en el cual las capacidades operativas se ven restringidas, especialmente en entornos urbanos densos o áreas geográficas de difícil acceso.

1.4 Preguntas de investigación

- ¿Cuáles son los componentes esenciales para desarrollar un sistema de navegación basado en ROS que permita la comunicación y control efectivo del vehículo aéreo, incluyendo la adquisición de imágenes?
- ¿Qué arquitectura de red neuronal convolucional es la más adecuada para la clasificación de las imágenes adquiridas por el vehículo aéreo y la toma de decisiones basadas en esta clasificación?
- ¿Cómo realizar una integración de la red neuronal convolucional con el sistema de navegación del vehículo aéreo de forma apropiada?

1.5 Objetivo general

- Desarrollar y evaluar un sistema de navegación autónoma para vehículos aéreos no tripulados basado en referencia visual, haciendo uso de redes neuronales convolucionales para mejorar la percepción y toma de decisiones de los UAV en base a los puntos de referencia.

1.6 Objetivos específicos

- Desarrollar un sistema de navegación basado en ROS para la comunicación y el control del vehículo aéreo, permitiendo la adquisición de imágenes.
- Desarrollar una red neuronal convolucional para la clasificación de las imágenes adquiridas por el vehículo aéreo y en base a esto tomar decisiones.
- Implementar la interacción entre el sistema de navegación del vehículo aéreo y la red neuronal convolucional.

1.7 Hipótesis

La implementación de un sistema de navegación autónoma para vehículos aéreos no tripulados, fundamentado en referencias visuales mediante el uso de redes neuronales convolucionales, se traducirá en una mejora sustancial de la percepción y la capacidad de toma de decisiones de los UAV en comparación con la visión asistida proporcionada por la CNN. Este avance, a su vez, se reflejará en una navegación más precisa y eficaz.

Capítulo 2

Marco Teórico

Este marco se centra en el papel del aprendizaje profundo en la navegación autónoma de vehículos aéreos no tripulados. Profundizando en cómo un sistema puede utilizar las redes neuronales convencionales para aprender a mapear directamente imágenes que provienen de la cámara de un UAV y realizar distintas acciones de control para el uso de los actuadores de UAV.

2.1 Aprendizaje profundo

El aprendizaje profundo se refiere a una clase de modelos de aprendizaje automático que utilizan múltiples capas de representaciones para aprender y realizar tareas de clasificación, regresión y generación de datos (Goodfellow et al., 2016a). Para comprender el concepto de aprendizaje profundo, es útil familiarizarse con una serie de conceptos clave en el ámbito de la inteligencia artificial y el aprendizaje automático.

Las redes neuronales artificiales (ANN, del inglés *Artificial Neural Network*) tienen un linaje intelectual que se remonta a los primeros intentos de emular el funcionamiento del cerebro humano en una máquina. La idea de estas redes surgió en la década de 1940, cuando Warren McCulloch y Walter Pitts propusieron un modelo matemático de una neurona, sentando así las bases teóricas para el desarrollo de las ANN (McCulloch y Pitts, 1943). Sin embargo, fue en la década de 1950 cuando Frank Rosenblatt propuso el Perceptrón, un modelo de neurona artificial capaz de aprender y clasificar patrones binarios (McCulloch y Pitts, 1943; Rosenblatt, 1958).

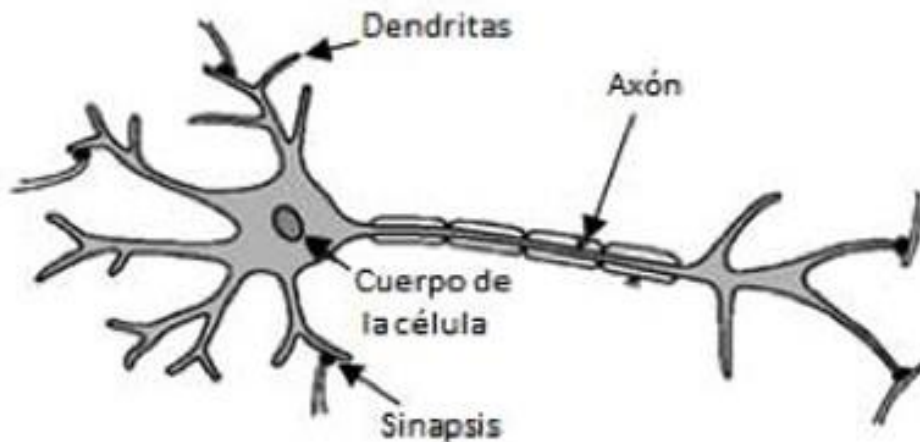
A pesar de este prometedor inicio, el Perceptrón tenía limitaciones fundamentales en su capacidad para abordar problemas más complejos. Este revés condujo a un período conocido como “la temporada de invierno de la inteligencia artificial”, donde las expectativas

sobre las capacidades de las ANN se vieron frustradas (Nilsson, 2009).

Sin embargo, en la década de 1980, el campo experimentó un resurgimiento con el desarrollo del algoritmo de *backpropagation*, un método para entrenar redes neuronales profundas mediante la propagación eficiente del error a través de la red (Rumelhart et al., 1988). Esto marcó un hito crucial y permitió que las ANN empezaran a resolver problemas mucho más complejos que antes parecían inabordables.

Las neuronas artificiales están basadas en las neuronas biológicas, que son los bloques fundamentales del sistema nervioso en organismos vivos. Su estructura intrincada comprende un cuerpo celular, dendritas receptoras de señales y un axón que transmite estas señales a otras células como se puede observar en la Figura 2.1. La comunicación entre neuronas ocurre a través de sinapsis, donde se transmiten señales eléctricas y químicas, permitiendo así la codificación y procesamiento de información en el cerebro humano (Tavanaei et al., 2019).

Figura 2.1: Estructura básica de una neurona biológica.



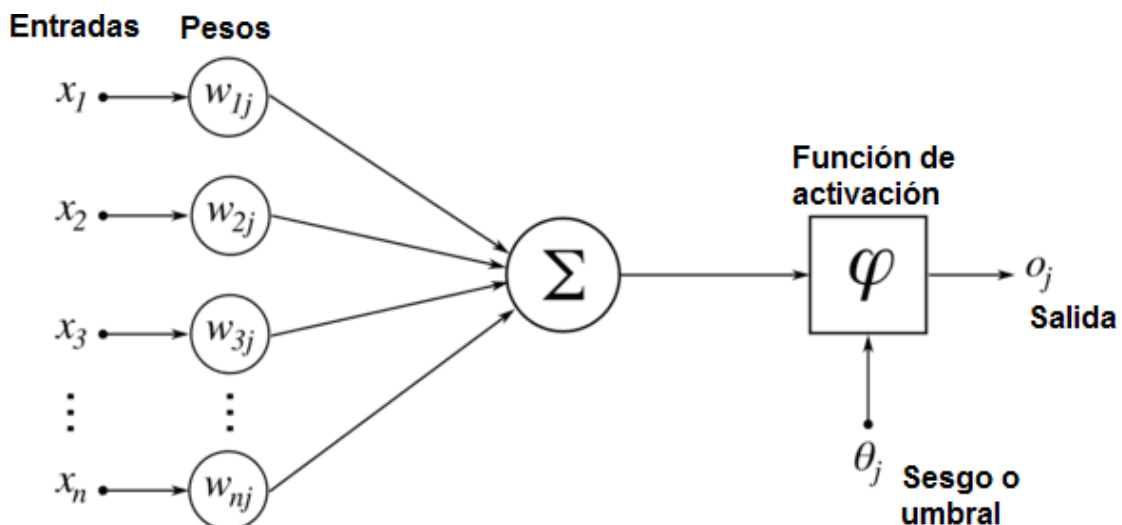
Fuente: Rivas-Asanza et al.(2018).

Esta maquinaria biológica ha servido como inspiración para las Redes Neuronales Artificiales, logrando distintos diseños modelos matemáticos para emular el funcionamiento de las neuronas biológicas. Aunque las neuronas artificiales siguen una ruta más simplificada, cada neurona artificial recibe múltiples entradas, ponderadas por valores conocidos como pesos, y luego estas son canalizadas a través de una función de activación. Esta función, esencialmente, introduce no linealidades al proceso, emulando de manera simplificada el complejo proceso de integración y transmisión de señales que se observa en las neuronas biológicas (Guclu and van Gerven, 2017).

A pesar de la diversidad de modelos matemáticos en las Redes Neuronales Artificiales como Perceptrón, Multilayer Perceptron (MLP), Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN), Long Short-Term Memory (LSTM), Redes Neuronales Generativas (GAN) y Transformers, estas comparten varios componentes en sus arquitecturas.

En la Figura 2.2 muestra la arquitectura de una red neuronal artificial y los componentes en común, como lo son las entradas, nodos, conexiones, pesos, sesgo, funciones de activación y salidas.

Figura 2.2: Arquitectura de una red neuronal artificial.



Fuente: Farfán (2005).

Los nodos o neuronas artificiales reciben múltiples entradas, que representan diversas características o información, y se encuentran conectadas por vías de comunicación a lo largo de las cuales la información fluye a través del sistema. Las neuronas aplican una operación matemática a sus entradas, ponderándolas por medio de valores llamados pesos (LeCun et al., 2015). Luego, estos valores ponderados se suman, y el resultado se pasa a través de una función de activación no lineal. Esta función introduce complejidad y no linealidades al proceso, lo que permite a la red aprender relaciones y patrones complejos en los datos (LeCun et al., 2015).

Las neuronas están organizadas en capas de entrada, ocultas y de salida. La capa de entrada recibe directamente los datos. Luego, las capas ocultas procesan la información, extrayendo características relevantes a medida que se profundiza en la red. Finalmente, la capa de salida produce la predicción o clasificación final. Las conexiones entre las neuronas están asociadas con pesos sinápticos que determinan la fuerza de la influencia de una neurona sobre otra. Estos pesos se ajustan durante el proceso de entrenamiento de la red para mejorar la precisión de las predicciones.

Las funciones de activación en una red neuronal artificial son esenciales para la capacidad de la red de aprender y representar relaciones no lineales entre las entradas y las salidas. Estas funciones determinan si una neurona se activa o no en respuesta a un cierto estímulo. Una de las funciones de activación más comunes es la Rectified Linear Unit (ReLU), que asigna cero a todos los valores negativos y mantiene los valores positivos sin cambios. Esta función ha demostrado ser especialmente efectiva en el entrenamiento de redes neuronales profundas (Goodfellow et al., 2016b).

Existe un fenómeno en el que el modelo se adapta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos, esto se llama sobre ajuste. Para prevenir el sobre ajuste, se aplican técnicas como la regularización y el *dropout* (Goodfellow et al., 2016b). Estas estrategias ayudan a la red a generalizar mejor a datos no vistos, lo que mejora su desempeño en situaciones del mundo real.

2.2 Mapeo de imágenes a acciones de control

El mapeo de imágenes es un paso crítico en la comprensión de datos visuales, especialmente en áreas como la visión por computadora y el aprendizaje profundo. Implica transformar una imagen, que es esencialmente un conjunto de píxeles con valores de intensidad, en una forma que una computadora pueda entender y analizar. Dado que las computadoras operan en el dominio numérico, esta conversión es esencial para que puedan procesar y extraer información relevante de las imágenes.

En este contexto, ROS (*Robot Operating System*) juega un papel importante al proporcionar un marco de trabajo robusto para el desarrollo de sistemas robóticos, incluidos los que están involucrados en el procesamiento de imágenes. ROS ofrece herramientas y bibliotecas que facilitan la adquisición y procesamiento de imágenes a través de una variedad de sensores, como cámaras y LiDAR. Esto permite a los investigadores y desarrolladores implementar algoritmos de mapeo de imágenes con mayor facilidad y eficacia en entornos robóticos (Quigley et al., 2009).

Algunos de los paquetes de ROS, como “imageTransport” y “cvBridge”, son particularmente útiles para el manejo de imágenes en tiempo a su entorno. Además, ROS proporciona un entorno de simulación que permite probar y validar algoritmos de mapeo de imágenes en un entorno controlado antes de implementarlos en un robot físico (Bradski and Kaehler, 2008a).

ROS, consiste en un *framework* especializado para el desarrollo de software enfocado al área de la robótica, en dicho *framework* se cuenta con diversas herramientas de código abierto para su aplicación directa a ambientes y sistemas robóticos (ROS, Fecha de acceso). Ofrece una arquitectura modular que permite la fácil interconexión de diferentes nodos y componentes. Esto es crucial en el contexto de mapeo de imágenes, donde se requiere la colaboración de múltiples sistemas, como la adquisición de imágenes, procesamiento de visión, y algoritmos de mapeo. La arquitectura de ROS facilita la comunicación entre estos componentes, lo que resulta en una operación más eficiente y coordinada del UAV (Quigley et al., 2009).

Existe una librería llamada Bebop Autonomy, que es una biblioteca de código abierto diseñada para facilitar la integración y control del vehículo aéreo Parrot Bebop dentro del entorno de desarrollo ROS (Parrot, n.d.). Esta biblioteca ofrece una amplia gama de funcionalidades, desde el control de velocidad y orientación del vehículo aéreo hasta el acceso a datos de la cámara integrada. Permite a los desarrolladores aprovechar las capacidades del vehículo aéreo Bebop en aplicaciones que abarcan desde la visión por computadora hasta la navegación autónoma. La integración de Bebop Autonomy con ROS proporciona un marco de trabajo robusto y escalable para el desarrollo de soluciones robóticas basadas en vehículos aéreos.

El desarrollo de *scripts* y nodos para el control de vehículos aéreos es una parte esencial en la programación de sistemas robóticos aéreos (Quigley et al., 2009). Estos *scripts* son piezas de software especialmente diseñadas para comunicarse con el vehículo aéreo, dictando su comportamiento y permitiendo una interacción precisa y automatizada. Por ejemplo, en el contexto de ROS, los *scripts* pueden ser escritos en lenguajes de programación populares como Python o C++ (Bradski and Kaehler, 2008a). Estos *scripts* no solo son capaces de enviar comandos al vehículo aéreo, sino también de recibir y procesar una variedad de datos provenientes del vehículo aéreo. Esto incluye información vital como la posición GPS, datos de sensores de a bordo y las imágenes capturadas por las cámaras del vehículo aéreo. Esta capacidad bidireccional de intercambio de información permite a los vehículos aéreos tomar decisiones más informadas y responder en tiempo a su entorno.

Python es un lenguaje de programación de alto nivel y propósito general, su fortaleza radica en su sintaxis intuitiva y fácil de leer, lo que lo hace especialmente accesible para programar. Además, Python es un lenguaje interpretado, lo que significa que no requiere un proceso de compilación previo y puede ejecutarse directamente desde su código fuente. Esto agiliza el proceso de desarrollo y facilita las pruebas y la interacción (Downey, 2012).

Cuenta con una amplia colección de bibliotecas y módulos que cubren una amplia gama de funcionalidades. La biblioteca estándar de Python proporciona un conjunto robusto de herramientas, mientras que la comunidad ha desarrollado bibliotecas adicionales como *OpenCV*, *NumPy*, *Pandas*, *TensorFlow*, entre otras, que extienden sus capacidades en áreas como la ciencia de datos, el procesamiento de imágenes, y la inteligencia artificial.

OpenCV es una biblioteca de código abierto ampliamente utilizada para el procesamiento de imágenes y visión por computadora. Ofrece una amplia gama de funciones que van

desde la manipulación básica de imágenes hasta técnicas avanzadas de visión por computadora (Bradski and Kaehler, 2008b).

Pandas es otra biblioteca esencial para el procesamiento y análisis de datos. Proporciona estructuras de datos flexibles y eficientes, como los *DataFrames*, que permiten manipular grandes conjuntos de datos de manera sencilla y rápida (McKinney, 2012).

NumPy es una base fundamental para el cómputo científico en Python. Proporciona un poderoso conjunto de herramientas para la manipulación de matrices y operaciones matemáticas de alto nivel (Harris et al., 2020).

TensorFlow cuenta con una capacidad para construir y entrenar modelos de aprendizaje profundo con facilidad. Ofrece una amplia variedad de herramientas y recursos, como *APIs* de alto nivel para construir y entrenar modelos rápidamente, así como *APIs* de bajo nivel para una mayor flexibilidad y personalización en la construcción de modelos (Abadi et al., 2016).

2.3 Clasificación de imágenes

A diferencia de las redes neuronales tradicionales, las CNN están especialmente

diseñadas para reconocer patrones en datos con una estructura espacial, como las imágenes. Esta especialización proviene de la observación de que, en imágenes, ciertas características locales son más importantes que las relaciones globales entre píxeles. El desarrollo de las redes neuronales convolucionales ha revolucionado la percepción visual en tareas de reconocimiento de objetos, detección de características y segmentación de imágenes.

Las CNN son capaces de extraer automáticamente características relevantes de las imágenes, permitiendo a los vehículos aéreos identificar elementos del entorno, como puntos de referencia, obstáculos y características geográficas. Esto demuestra una sorprendente capacidad para aprender representaciones jerárquicas de datos visuales, lo que resulta en una mejora significativa en la precisión del reconocimiento de objetos y la interpretación del entorno (LeCun et al., 2015).

En las redes neuronales convolucionales, las capas son los componentes básicos que realizan operaciones de convolución en las imágenes de entrada. Las capas convolucionales aplican filtros o *kernels* a las regiones de la imagen para extraer características locales, como bordes, texturas y patrones relevantes (Bengio, 2016). Cada capa convolucional contiene múltiples filtros que generan mapas de características. Estos mapas de características se utilizan como entradas para capas posteriores, lo que permite aprender representaciones cada vez más abstractas de la imagen.

Los *kernels* son matrices de pesos que se utilizan en las capas convolucionales para realizar la operación de convolución. Cada *kernel* se desliza sobre la imagen de entrada y realiza una multiplicación y suma ponderada en las regiones locales de la imagen para generar un valor en el mapa de características. Los *kernels* son entrenados durante el proceso de aprendizaje de la CNN y determinan qué características se detectan en la imagen (Bengio, 2016).

En cuanto a la aplicabilidad, las CNN han revolucionado la detección de objetos, como lo destacan Ren et al., con la introducción de la Red Neuronal Convolucional Regional, que combina la detección de objetos basada en regiones y el aprendizaje profundo para

lograr un rendimiento sobresaliente en tareas de detección (Ren et al., 2015). Para el procesamiento de imágenes médicas, Litjens et al., subrayan cómo las CNN han demostrado ser esenciales en la segmentación de órganos y lesiones, lo que facilita diagnósticos más precisos y una atención médica de mayor calidad (Litjens et al., 2017).

Además, las CNN han evolucionado para abordar desafíos más específicos. Las redes residuales, propuestas por He, K. et al., introdujeron conexiones residuales que permiten entrenar redes extremadamente profundas sin sufrir degradación en el rendimiento, lo que impulsó aún más el avance en la profundidad de las CNN (He et al., 2016).

La clasificación de objetos a partir de imágenes se basa en la utilización de un conjunto de características específicas y un algoritmo de clasificación. Sin embargo, en la práctica, alcanzar una clasificación perfecta es difícil debido a las variaciones dentro de las clases, las similitudes entre ellas y la posible interferencia de ruido en las imágenes.

Como mencionan González et al. “la elección del clasificador y la regla de decisión desempeñan un papel crucial en la precisión de la clasificación” (González et al., 2018) . Es esencial elegir y extraer las características de clasificación con meticulosidad, lo cual requiere un análisis detenido del contexto del problema. Las características se obtienen de las imágenes de entrenamiento y se asignan a un algoritmo de clasificación para entrenar el modelo predictivo.

Luego, este mismo conjunto de características se extrae de la imagen de prueba y se suministra al modelo ya entrenado. El modelo devuelve la etiqueta de clase de la muestra de prueba, basándose en la probabilidad de concordancia con una categoría específica. Para el modelo de aprendizaje se debe tener en cuenta los siguientes puntos.

- **Conjunto de datos etiquetados:**

Se requiere un conjunto de datos extenso y etiquetado que contenga imágenes representativas de las categorías que se desean clasificar. Cada imagen debe estar

etiquetada con la categoría a la que pertenece, lo que permite a la CNN aprender las asociaciones entre características visuales y etiquetas correspondientes (Krizhevsky et al., 2012).

- **Arquitectura de la CNN:**

La elección de la arquitectura de la CNN es crucial. Esto incluye el diseño de capas convolucionales, de *pooling* y totalmente conectadas, así como la configuración específica de cada capa. La arquitectura debe ser seleccionada considerando la complejidad de la tarea y la cantidad de datos disponibles (Szegedy et al., 2015).

- **Funciones de activación y función de pérdida:**

Las funciones de activación, como ReLU (*Rectified Linear Unit*), y la función de pérdida, como la entropía cruzada, son esenciales para la capacidad de aprendizaje de la red. Las funciones de activación introducen no linealidades en el modelo, permitiendo a la CNN aprender relaciones no lineales entre las características de entrada (Nair y Hinton, 2010).

Una de las arquitecturas de CNN es LeNet-5, que se caracteriza por constar de siete capas, excluyendo la capa de entrada. Cada una de estas capas posee parámetros entrenables, que se despliegan de manera detallada en la Figura 2.3. La entrada inicial es una imagen de 32x32 píxeles, un tamaño significativamente mayor al carácter más grande de la base de datos, permitiendo que los detectores de características de alto nivel identifiquen rasgos distintivos cruciales como los puntos finales de trazos o las esquinas.

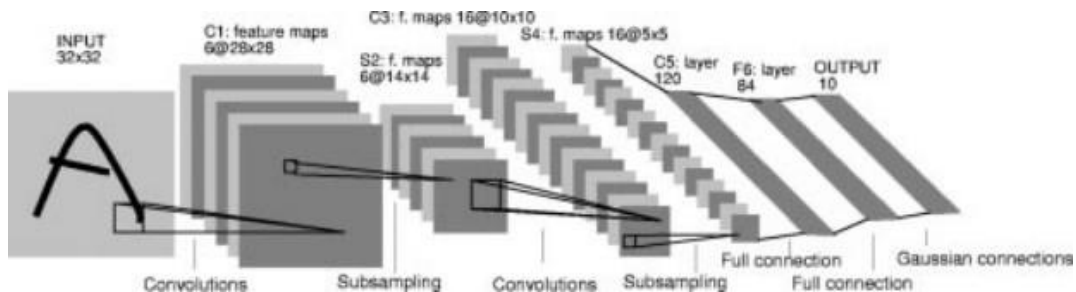
La primera capa, conocida como C1, es una capa convolucional con seis mapas de características, donde cada unidad en cada mapa se conecta a un vecindario de 5x5 en la entrada. Esto genera mapas de características de 28x28, evitando que las conexiones caigan fuera de los límites. C1 engloba 156 parámetros entrenables y 122,304 conexiones. A esta le sigue la capa S2, que es una capa de submuestreo con seis mapas de características de tamaño 14x14. S2 reduce a la mitad el número de filas y columnas

en comparación a C1.

La capa C3, una capa convolucional subsiguiente, contiene 16 mapas de características. Cada unidad se conecta a varios vecindarios de 5x5 en ubicaciones idénticas en un subconjunto de mapas de características de S2. Por otro lado, la capa S4 es una capa de submuestreo que tiene 16 mapas de características de tamaño 5x5, conectándose a vecindarios de 2x2 en el mapa de características correspondiente en C3. La capa C5, a pesar de estar etiquetada como convolucional, posee 120 mapas de características y conecta unidades a vecindarios de 5x5 en los 16 mapas de características de S4.

A continuación, la capa F6 contiene 84 unidades y está completamente conectada a C5, con un total de 10,164 parámetros entrenables. Finalmente, la capa de salida consta de unidades con funciones de base radial euclidiana (RBF), una por cada clase, cada una con 84 entradas [29]. Esta estructura única ha demostrado ser efectiva en tareas de clasificación de dígitos escritos a mano.

Figura 2.3: Arquitectura LeNet-5.



Fuente: Lecun et al. (1998).

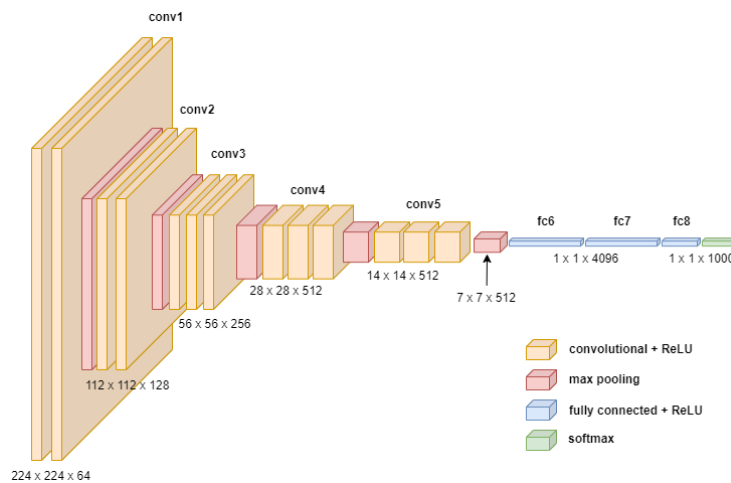
Habiendo explorado la arquitectura LeNet-5 y su eficacia en la clasificación, nos enfocaremos en el terreno más profundo de la visión por computadora con la arquitectura VGG-19. Esta última se distingue por su impresionante profundidad, compuesta por un total de 19 capas.

A diferencia de LeNet-5, VGG-19 conlleva un análisis más detallado de su estructura compleja y las estrategias innovadoras que emplea para comprender la información visual de manera más profunda.

La arquitectura VGG-19 se caracteriza por su profundidad, compuesta por un total de 19 capas. A la entrada de la ConvNet, se proporciona una imagen RGB de tamaño fijo de 224x224 píxeles; esta arquitectura se muestra en la Figura 2.4. El único preprocesamiento realizado es la substracción del valor RGB medio calculado en el conjunto de entrenamiento de cada píxel.

La imagen atraviesa una serie de capas convolucionales, donde se utilizan filtros de 3x3, el tamaño más pequeño que puede capturar relaciones espaciales elementales como izquierda/derecha, arriba/abajo y centro. En sus configuraciones, se incorporan filtros de convolución 1x1, que actúan como una transformación lineal de los canales de entrada seguida de una operación no lineal. La zancada de convolución se mantiene constante en 1 píxel, y el relleno espacial se ajusta de modo que la resolución espacial se conserve después de la convolución. Esto se logra con un relleno de 1 píxel para las capas de convolución de 3x3.

Figura 2.4: Arquitectura VGG-19.



Fuente: Ahad et al. (2023).

A lo largo de la arquitectura, se incluyen cinco capas de agrupación máxima que siguen a algunas de las capas de convolución. La agrupación máxima se lleva a cabo en ventanas de 2x2 píxeles, con un paso de 2. La etapa final de la red consta de tres capas completamente conectadas. Las dos primeras tienen 4,096 canales cada una, mientras que la tercera se encarga de realizar una clasificación ILSVRC de 1,000 vías, lo que implica 1,000 canales, uno para cada clase. La última capa es la capa softmax, que realiza la asignación de clases. Es importante destacar que la configuración de las capas completamente conectadas es la misma en todas las variantes de la red VGG-19.

2.4 Funcionamiento del vehículo aéreo

Parrot Bebop 2, fabricado por la empresa francesa Parrot, representa una notable evolución en el mundo de los drones recreativos. Este dron, que es la versión mejorada de su predecesor, el Parrot Bebop, ha sido diseñado con una combinación de características avanzadas que lo distinguen en el mercado (Parrot, n.d.).

El diseño aerodinámico y ligero del Parrot Bebop 2 no solo le confiere una estética moderna, sino que también contribuye a su agilidad y maniobrabilidad durante el vuelo (Parrot, n.d.).

Cuenta con un acelerómetro de 3 ejes, que configura para medir la aceleración en tres dimensiones espaciales distintas, es decir, a lo largo de los ejes x, y, y z (Silva and Santos, 2018). Este instrumento desglosa la aceleración, representando el cambio en la velocidad de un objeto respecto al tiempo (Beeharee y Clare, 2006). Para el vehículo aéreo no tripulado (UAV), estas mediciones son fundamentales para comprender y gestionar el movimiento tridimensional de la aeronave.

En el eje x, mide la aceleración de izquierda a derecha o viceversa, el acelerómetro proporciona información crucial para ajustar la orientación lateral del UAV. De manera similar, en el eje y, que mide la aceleración de arriba a abajo o viceversa, se obtiene información valiosa sobre los cambios en la orientación vertical de la aeronave. El eje

z, que mide la aceleración hacia adelante o hacia atrás, permite ajustes precisos para el movimiento frontal o posterior del UAV.

Un giroscopio de tres ejes es un componente esencial en la instrumentación de vehículos aéreos no tripulados, ya que mide la velocidad angular en tres direcciones distintas: eje x, eje y y eje z (Brown y Wang, 2018; Li et al., 2019). En el contexto de los UAV, este tipo de giroscopio proporciona datos cruciales sobre las tasas de cambio en la orientación de la aeronave en cada uno de sus ejes, permitiendo un seguimiento preciso y reactivo del movimiento angular en tres dimensiones.

En el eje x, se capturan las variaciones de orientación lateral, mientras que el eje y mide los cambios en la orientación vertical. Por último, el eje z registra las rotaciones alrededor del eje frontal o posterior del UAV. Esta información detallada es fundamental para el control de vuelo y la estabilidad de la aeronave, asegurando respuestas precisas a los comandos del usuario y contribuyendo a un rendimiento óptimo del UAV.

También utiliza un magnetómetro, que en el contexto de vehículos aéreos no tripulados, es un sensor crucial que mide la intensidad y, en algunos casos, la dirección del campo magnético en su entorno (Hagen y Ruotsalainen, 2018). Esta tecnología se utiliza para determinar la orientación del UAV con respecto al campo magnético terrestre, siendo especialmente útil para establecer la dirección del norte magnético. La información proporcionada por el magnetómetro contribuye a la navegación precisa del UAV, mejorando su capacidad para mantener una trayectoria específica y responder eficientemente a cambios en la dirección.

Contiene un sensor de presión barométrica, que es un componente esencial diseñado para medir la presión atmosférica en la ubicación del UAV (Elitaş y Zengin, 2019). Este tipo de sensor aprovecha el principio de que la presión atmosférica varía con la altitud, permitiendo así estimar la altitud a la que se encuentra el UAV.

La evaluación de la presión barométrica se ha destacado en la literatura especializada como una herramienta valiosa para la estimación precisa de la altitud en UAV (Kulesza et al., 2018). Al utilizar la información de la presión atmosférica, estos vehículos pueden

mejorar significativamente su capacidad de navegación y mantener altitudes específicas durante el vuelo, lo que es esencial para diversas aplicaciones, desde la cartografía hasta la monitorización ambiental.

Utiliza motores *brushless*, que son componentes esenciales en la propulsión de vehículos aéreos no tripulados. En contraste con los motores convencionales con escobillas, los motores *brushless* eliminan el contacto físico directo entre las escobillas y el rotor, lo que resulta en una reducción significativa de la fricción y el desgaste, y al mismo tiempo mejora la eficiencia general del motor (Yang et al., 2017). Este diseño sin escobillas también proporciona una mayor durabilidad y vida útil, factores críticos para el rendimiento y la confiabilidad a largo plazo de los motores en aplicaciones aeroespaciales.

Estos motores *brushless* no solo benefician a los UAV al proporcionar un rendimiento más eficiente, sino que también contribuyen a la autonomía y capacidad de respuesta mejorada de los UAV durante el vuelo (Rahman et al., 2016).

2.5 Casos de uso

En esta sección, nos sumergiremos en un análisis exhaustivo de casos de uso cruciales que ocupan un lugar central en la investigación en curso. La conjunción estratégica entre Redes Neuronales Convolucionales, Vehículos Aéreos No Tripulados y el Marco de Trabajo de Robótica de Código Abierto ROS que configura un ámbito de investigación.

Al explorar estos casos de uso, descubrimos las complejidades y las soluciones ingeniosas que se aplican para enfrentar desafíos específicos en la intersección de estas tecnologías avanzadas. Desde mejorar la capacidad de procesamiento visual con las CNN, hasta la autonomía y agilidad únicas de los vehículos aéreos no tripulados, y la coordinación eficiente proporcionada por ROS, cada componente desempeña un papel especial en la creación de soluciones integradas.

Este segmento nos ofrece una visión práctica de las aplicaciones y los avances logrados en la convergencia de estas tecnologías, estableciendo así una base sólida para discusiones y análisis más detallados en las secciones siguientes.

- **Caso de uso 1:** *Enfoque de Red Neuronal Convolutiva Multiescala Basado en Superpíxeles para la Segmentación de Objetos de Múltiples Clases a partir de Imágenes Aéreas Capturadas por UAV*

Tanmay ha contribuido significativamente al estado del arte en el campo de las redes neuronales convolucionales aplicadas a la visión de vehículos aéreos no tripulados. Su trabajo, como se describe, destaca por su enfoque innovador en la utilización de la arquitectura CNN para abordar los desafíos de la segmentación semántica en imágenes aéreas de UAV (Behera et al., 2023).

Proporcionando una solución robusta para la identificación de clases de objetos como carreteras, vehículos y vegetación. Este enfoque tiene el potencial de ser integrado en soluciones de inteligencia artificial basadas en robótica para aplicaciones como la extracción inteligente de carreteras y la detección de vegetación a través de imágenes aéreas de UAV, contribuyendo así al avance en la visión por computadora aplicada a vehículos aéreos no tripulados (Behera et al., 2023).

- **Caso de uso 2:** *Modelo de Aprendizaje Profundo Basado en Optimización de Enjambre de Partículas Multiobjetivo para la Clasificación de Escenas en Vehículos Aéreos no Tripulados*

En la investigación reciente realizada por Aghila Rajagopal, se aborda la creciente aplicación de vehículos aéreos no tripulados y sensores de imágenes livianos en la teledetección (Rajagopal et al., 2020). Rajagopal destaca la complejidad de procesos como la recolección, calibración y registro de datos provenientes de UAV, así como la interpretación semántica de estos datos. En este contexto, el artículo

propone un novedoso modelo de optimización multiobjetivo para evolucionar redes neuronales convolucionales de última generación, mejorando la eficiencia en la clasificación de escenas en imágenes aéreas de UAV (Rajagopal et al., 2020).

La investigación se centra en superar limitaciones anteriores asociadas con métodos tradicionales de codificación de características y aprendizaje no supervisado, utilizando CNNs para extraer características invariantes a la escala. Dada la restricción de recursos en UAV, el modelo propuesto automatiza la generación de soluciones no dominantes en el frente de Pareto, equilibrando resultados de clasificación y complejidad computacional (Rajagopal et al., 2020).

- **Caso de uso 3:** *Clasificación de Calidad de Imágenes Basada en CNN Considerando la Degradación de la Calidad en la Inspección de Puentes Utilizando un Vehículo Aéreo No Tripulado.*

En su investigación, Gi-Hun Gwon aborda la problemática de la evaluación de estructuras, especialmente puentes, mediante el procesamiento de imágenes adquiridas por vehículos aéreos no tripulados (Gwon et al., 2023). Gwon propone un método innovador de evaluación de calidad de imagen que utiliza una red neuronal convolucional para considerar varios factores de degradación presentes en las imágenes de inspección estructural (Gwon et al., 2023).

El enfoque se divide en dos etapas esenciales. En la primera, se desarrolla un método para obtener imágenes de calidad consistente al ajustar los parámetros de la cámara y distinguir entre imágenes de baja y alta calidad. La segunda etapa implica la clasificación del conjunto de datos de inspección utilizando un modelo de clasificación de calidad de imagen basado en CNN, entrenado con datos clasificados según su calidad (Gwon et al., 2023).

- **Caso de uso 4:** *Desarrollo y Verificación de Sistema de Pruebas de Vuelo Multi-DOF Basado en ROS para Vehículos Aéreos No Tripulados.*

En el marco de la investigación liderada por Shinhyung Kim, se ha implementado con éxito un sistema de prueba de vuelo para vehículos aéreos no tripulados, destacando la adopción de ROS como piedra angular de esta innovadora propuesta (S. Kim et al., 2023).

Este enfoque se erige con el propósito fundamental de garantizar la seguridad y eficiencia en el desarrollo de UAV, otorgando prioridad a la robustez y la integridad del sistema. El diseño concebido por Kim comprende un banco de pruebas con 4 grados de libertad, una unidad de control electrónica para la recolección de datos y la ejecución del software de operación, una unidad de suministro de energía, herramientas de software en ROS para la integración armoniosa de los sistemas de software y hardware, y un túnel de viento para simular el vuelo (S. Kim et al., 2023).

Los experimentos llevados a cabo con un UAV propulsado por cohete validan la efectividad del sistema, resaltando la capacidad del controlador de actitud para guiar el UAV hacia los objetivos con estabilidad. A pesar de que algunos aspectos del sistema aún están pendientes de desarrollo, se reconoce su utilidad para diversas pruebas de vuelo y estudios en áreas cruciales como control, diseño de navegación, análisis de cinemática y dinámica, estabilidad de vuelo, gestión de energía y control autónomo.

Este enfoque integral, respaldado por la implementación de ROS, proporciona una plataforma versátil para la investigación y desarrollo continuo de UAV. Logros tangibles ya se han alcanzado, incluyendo la finalización del dispositivo y la realización de pruebas de vuelo adicionales, consolidando la construcción exitosa de un sistema integrado de prueba de vuelo mediante la aplicación efectiva de ROS (S. Kim et al., 2023).

- **Caso de uso 5:** *EmergencyNet: Clasificación Eficiente de Imágenes Aéreas para*

la Monitorización de Emergencias con Drones Utilizando Fusión de Características Convolutivas Atrous

En el trabajo liderado por Christos Kyrkou, se aborda la mejora de las capacidades de teledetección de vehículos aéreos no tripulados mediante algoritmos basados en aprendizaje profundo, específicamente redes neuronales convolucionales (Kyrkou y Theocharides, 2020). Se destaca la importancia de esta tecnología en aplicaciones críticas de respuesta a emergencias y gestión de desastres, donde los UAV equipados con cámaras pueden desempeñar un papel crucial (Kyrkou y Theocharides, 2020).

El artículo presenta una contribución significativa al introducir una arquitectura de red neuronal convolucional eficiente llamada EmergencyNet. Esta red, basada en convoluciones atrous, se diseñó para ejecutarse de manera eficaz en plataformas con recursos limitados, como las que suelen encontrarse en UAV. Se demostró que EmergencyNet supera a modelos existentes al lograr hasta 20 veces mayor rendimiento con requisitos mínimos de memoria, manteniendo una pérdida de precisión inferior al 1% en comparación con los modelos de vanguardia (Kyrkou y Theocharides, 2020).

- **Caso de uso 6:** *Red Neuronal con Atención Profunda para la Clasificación Multietiqueta en Imágenes de Vehículos Aéreos no Tripulados.*

El estudio reciente en el ámbito de la clasificación multietiqueta de imágenes de vehículos aéreos no tripulados ha introducido una solución vanguardista que destaca por su enfoque en el uso de redes neuronales convolucionales y redes neuronales recurrentes.

La investigación presenta una arquitectura de red neuronal codificador-decodificador que incorpora ingeniosos mecanismos de atención de canal y espacial. El módulo codificador, basado en una CNN, transforma eficientemente las imágenes de UAV, mejorando la representación de características mediante una capa de excitación por compresión. Simultáneamente, el módulo decodificador, respaldado por una red

LSTM, realiza predicciones secuenciales de las clases presentes en la imagen mediante un innovador mecanismo de atención espacial (Alshehri et al., 2019).

Este enfoque integral ha demostrado ser altamente efectivo al proporcionar soluciones *end-to-end*, mejorando significativamente la representación de características y logrando resultados de clasificación más precisos. Validado en conjuntos de datos de UAV con una resolución espacial de 2 cm, el método propuesto supera a varios enfoques del estado de la técnica, allanando el camino para avances futuros en la clasificación multietiqueta de imágenes de UAV (Alshehri et al., 2019).

- **Caso de uso 7:** *Navegación de Drones Utilizando Redes Neuronales Convolucionales Profundas Basadas en la Explotación de Regiones y Bordes.*

En este trabajo llevado a cabo por Muhammad, se propone una estrategia innovadora basada en una CNN para abordar los desafíos de la navegación autónoma de UAV en entornos complejos y dinámicos. La CNN, denominada *Drone Split-Transform-and-Merge Region-and-Edge*, utiliza bloques convolucionales para realizar de manera sistemática operaciones de región y borde, preservando propiedades específicas en varios niveles, especialmente en entornos congestionados (Arshad et al., 2022).

La fusión de estos bloques convolucionales a varios niveles permite aprender la variación de texturas, facilitando la discriminación eficiente del objetivo del fondo y ayudando en la evasión de obstáculos. Los resultados de validación en conjuntos de datos urbanos demuestran un rendimiento considerable en términos de varianza explicada, *recall*, precisión y *F-score*, respaldando la aplicabilidad del modelo propuesto para la navegación autónoma en tiempo real y vuelos del mundo real (Arshad et al., 2022).

- **Caso de uso 8:** *CRowNet: Red Neuronal Profunda para la Detección de Hileras de Cultivos en Imágenes de Vehículos Aéreos no Tripulados (UAV).*

En el contexto de la investigación sobre la automatización de la detección de hileras

de cultivos mediante vehículos aéreos no tripulados, destaco el trabajo de Mamadou, quien propuso una innovadora metodología denominada CRowNet (Bah et al., 2020). Este enfoque, basado en una combinación de una red neuronal convolucional (CNN) y la transformada de Hough, emplea un modelo conformado por SegNet y una CNN específica basada en la transformada de Hough. La evaluación cuantitativa de CRowNet revela un rendimiento sobresaliente, logrando una tasa de detección de hileras de cultivos del 93.58%, con un puntaje de Intersección sobre Unión superior al 70% (Bah et al., 2020).

Este trabajo de Mamadou se presenta como un hito significativo en el ámbito de la agricultura de precisión y la automatización de la detección de cultivos, especialmente cuando se utilizan UAV para la captura de imágenes. La combinación de técnicas avanzadas de CNN y la transformada de Hough demuestra ser eficaz en la detección precisa de hileras, abriendo nuevas posibilidades para la optimización de tareas agrícolas. Esta contribución se alinea estrechamente con la temática de mi tesis, que se centra en la aplicación de CNN en vehículos aéreos, brindando valiosas perspectivas para futuras investigaciones en este campo (Bah et al., 2020).

- **Caso de uso 9:** *Detección y Seguimiento de Objetos en Tiempo Real basado en Redes Neuronales Convolucionales para el Parrot AR Drone 2*

En su trabajo innovador, Ali Rohan investigó el empleo de una red neuronal convolucional para la detección y seguimiento de objetos en UAV, con un enfoque específico en el Parrot AR Drone 2 (Rohan et al., 2019).

Destacando una precisión del 98% en la detección de objetos y una eficiencia del 96.5% en el seguimiento en tiempo real, este estudio proporciona valiosas aportaciones para mejorar la autonomía y eficacia de vehículos aéreos no tripulados. Este enfoque, relevante para mi tesis centrada en la aplicación de CNN en UAV, destaca la capacidad de la CNN para abordar desafíos cruciales en la vigilancia y

seguimiento de objetivos en entornos dinámicos, ofreciendo perspectivas significativas para el desarrollo de estrategias similares en el ámbito de la autonomía de UAV (Rohan et al., 2019).

- **Caso de uso 10:** *Un Acelerador CNN de Bajo Costo Totalmente Basado en Números Enteros en FPGA para el Reconocimiento en Tiempo Real de Señales de Tráfico.*

En la investigación liderada por Jaemyung Kim, se ha logrado un avance significativo al desarrollar un acelerador de hardware en tiempo real para el reconocimiento de señales de tráfico mediante el empleo de una CNN. La tecnología de reconocimiento de señales de tráfico se erige como esencial para los sistemas avanzados de asistencia al conductor, y la propuesta de Kim se destaca por abordar los desafíos inherentes a la implementación de CNN en entornos automotrices con recursos limitados (J. Kim et al., 2022).

El método de cuantización introducido no solo optimiza la complejidad computacional, sino que también permite la reconstrucción de la CNN de manera que todas las operaciones, incluso las asociadas con la ruta de conexión de salto de bloques residuales utilicen aritmética entera exclusivamente. Esto se traduce en un tamaño de modelo reducido drásticamente, un rendimiento de cómputo mejorado y una precisión excepcional de 99.07% en la clasificación de señales de tráfico, marcando un hito en comparación con implementaciones previas de señales de tráfico basadas en CNN en plataformas integradas (J. Kim et al., 2022).

Capítulo 3

Propuesta Metodológica

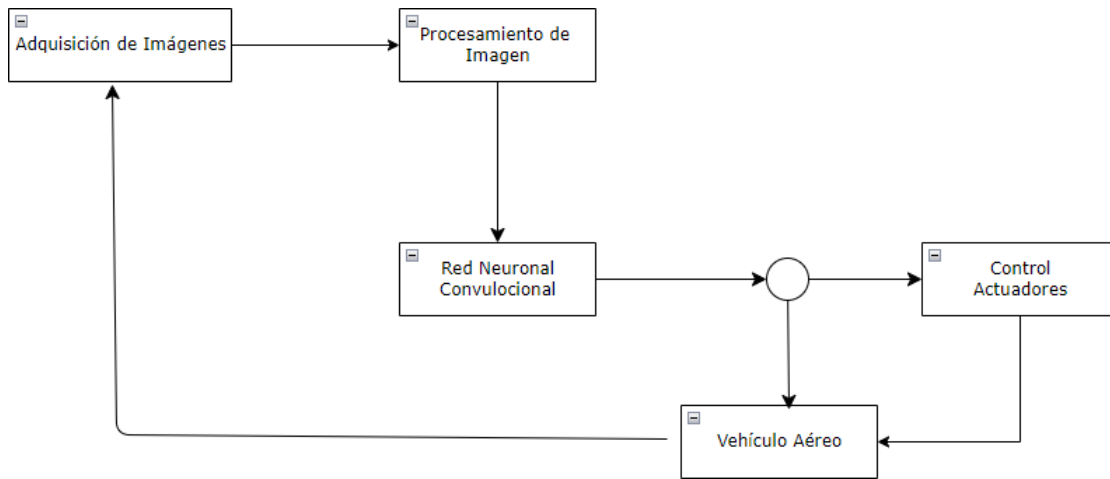
La metodología propuesta se enfocó en la implementación y evaluación de un sistema de navegación autónoma para vehículos aéreos no tripulados basado en referencia visual. Este enfoque integrador se compone de etapas clave, cada una diseñada para alcanzar los objetivos específicos del proyecto.

En la Figura 3.1 se muestra el diagrama de bloques que representa el flujo para el sistema de navegación, donde el proceso comienza con la entrada de imágenes, que consiste en cámaras que capturan el entorno del vehículo. Estas imágenes luego pasan por el procesamiento de imágenes, donde se aplican operaciones como la normalización de la iluminación y la reducción de ruido para prepararlas para el siguiente paso.

Posteriormente, las imágenes procesadas son enviadas a la red neuronal convolucional, con una arquitectura especializada en el procesamiento de datos de tipo imagen. Esta CNN utiliza capas de convolución y *pooling* para extraer características clave como bordes, texturas y formas.

La salida de la CNN se somete a una validación de la imagen para determinar si ha identificado correctamente referencias visuales relevantes. Si la validación es exitosa, los resultados se utilizan en el control de actuadores para generar comandos que ajusten la altitud, orientación y velocidad del vehículo aéreo.

Figura 3.1: Diagrama del sistema de navegación.



Fuente: Elaboración propia.

En caso de que la validación indique que no se ha reconocido ninguna referencia visual, el sistema regresa al bloque de entrada de imágenes para adquirir nuevas imágenes y continuar el proceso de navegación.

Cada etapa desempeña un papel esencial en el procesamiento y utilización de la información visual para guiar el vuelo de manera autónoma.

3.1 Manipulación y obtención de imágenes

En esta sección, se presenta una metodología detallada para manipular el UAV comercial Parrot Bebop 2 utilizando el sistema operativo Linux y el entorno de desarrollo ROS. Se describen los pasos necesarios para establecer la comunicación entre la computadora y el UAV, así como para controlar sus actuadores.

- **PASO 1:** *Configuración del entorno de trabajo*

Es esencial contar con un entorno de desarrollo adecuado en la distribución de Ubuntu 16.04 del sistema operativo Linux, como lo es la instalación de ROS y la distribución de Kinetic con los paquetes para proporciona una interfaz esencial para

interactuar con el vehículo aéreo que en esta investigación será un Parrot Bebop 2.

Para que ROS funcione correctamente, es necesario configurar el entorno en cada sesión de terminal. Esto se logra configurando el archivo tipo *bashrc*, el cual es ejecutado en una terminal. También se inicializa *rosdep*, para que se encargue de identificar y proporcionar todas las bibliotecas, herramientas y recursos adicionales que un paquete de ROS pueda necesitar para funcionar correctamente en el entorno de desarrollo.

- **PASO 2:** *Desarrollo de scripts*

El desarrollo de nodos en ROS implica la creación de unidades de procesamiento autónomas y modularizadas, como publicadores, suscriptores y servicio que se comunican a través del sistema de comunicación de ROS. Con los publicadores se puede enviar mensajes a un tópico específico, los suscriptores pueden recibir mensajes de un tópico y los servicios pueden comunicarse con otros nodos de manera síncrona. Estos elementos facilitan la interacción y colaboración entre los diferentes componentes de la red ROS. Se desarrollaron nodos para despegar, navegar por los ejes *x*, *y* y *z*, protección de interrupciones y aterrizaje.

- **Paso 3:** *Obtención de imágenes.*

El desarrollo de este script implicó programar una serie de instrucciones que permitieron a la cámara del vehículo aéreo, en este caso el Parrot Bebop 2, capturar imágenes a una resolución alta de 4,096 x 2,160 píxeles a intervalos regulares, inicialmente cada segundo. Estas imágenes se sometieron a un proceso de análisis y se dividieron en una matriz de imágenes más pequeñas. Este paso fue crucial para identificar señales o detalles de interés que pueden estar distribuidos en diferentes partes de la imagen completa.

La división de imágenes permitió mejorar la precisión en la identificación, especialmente cuando las señales se encuentran a una distancia considerable o

cuando la calidad de los píxeles es menor. Además, este enfoque proporcionó al vehículo aéreo la capacidad de determinar la dirección en la que es más probable encontrar una señal de interés.

Para el desarrollo de los *scripts*, fue necesario diseñar distintos algoritmos. El Algoritmo 1 está diseñado para guiar la operación de un UAV, centrándose en la captura y procesamiento continuo de imágenes. En su ejecución continua, el UAV intenta capturar imágenes utilizando sus propios sensores o cámaras integradas. La falta de entradas específicas sugiere que el procedimiento confía en los sensores incorporados para realizar la tarea.

Dentro del bucle infinito, el algoritmo establece las dimensiones del objeto de interés, posiblemente una señal visual, con anchos y altos predefinidos de 100 píxeles cada uno. A continuación, se procede a centrar el objeto en la imagen capturada, utilizando las dimensiones previamente establecidas. Este proceso implica ajustar la posición del objeto en la imagen para garantizar que esté en el centro, lo que puede ser crucial para una correcta interpretación y toma de decisiones basada en la señal visual.

Después de este paso de procesamiento, el algoritmo visualiza tanto la imagen original como la imagen resultante con el objeto centrado. Esta visualización podría ser útil para la supervisión humana o el análisis posterior, permitiendo a los operadores o sistemas externos entender la información capturada por el UAV.

Finalmente, el bucle continúa ejecutándose indefinidamente, indicando que el UAV sigue capturando y procesando imágenes de manera continua. Este tipo de algoritmo podría ser relevante en escenarios de navegación autónoma, donde el UAV utiliza señales visuales para la toma de decisiones autónoma, como la identificación de puntos de referencia o la detección de obstáculos.

En el Algoritmo 2 se diseñó el procedimiento de navegación para el UAV, llamado Parrot Bebop 2. Inicia con la tentativa de despegue del UAV, marcando el comienzo del proceso de navegación. Un bucle infinito se establece después del despegue, indicando

que la navegación será continua.

Algoritmo 1: Captura de imagen y centrado de señal en la imagen

Data: None

Result: Ejecutar el procedimiento principal

Procedimiento Imagen():

while *True* **do**

Intentar:

 # Capturar la imagen

 imagen_original = CapturarImagen()

 # Dimensiones de la imagen de la señal

 ancho_objeto = 100

 alto_objeto = 100

 # Centrar la señal en la imagen

 imagen_centrada = CentrarObjeto(imagen_original, ancho_objeto,
alto_objeto)

 # Mostrar la imagen original y la imagen centrada

 MostrarImagen("Imagen Original", imagen_original)

 MostrarImagen("Objeto Centrado", imagen_centrada)

end

Dentro de este bucle, el UAV captura y centra una imagen utilizando el Algoritmo 1. Luego, la imagen centrada se somete a procesamiento mediante una red neuronal convolucional previamente entrenada. La CNN clasifica la imagen, proporcionando una etiqueta de clase predicha que describe el objeto presente en la imagen.

La interacción del UAV se realiza según la clasificación obtenida. Si la clase predicha coincide con el objetivo deseado, se ejecutan instrucciones específicas para desplazar al UAV de acuerdo con la situación. En caso de que la clase sea otro objeto, el UAV gira hacia la izquierda. Si ninguna clase es reconocida, el UAV avanza hacia adelante.

Este bucle continúa ejecutándose indefinidamente, lo que refleja la naturaleza continua del proceso de navegación basado en la clasificación de imágenes. En resumen, el algoritmo establece un marco para que el UAV Parrot Bebop 2 tome decisiones de navegación en tiempo real, utilizando la clasificación de imágenes como base para determinar sus acciones.

Algoritmo 2: Procedimiento de navegación con el UAV

Data: None

Result: Realizar los comandos para los actuadores para el UAV

Procedimiento **NavegacionConBebop():**

Intentar:

 Despegar el UAV

while True do

 CapturarCentradoImagen

Procesar la imagen:

 Clasificar la imagen utilizando una red neuronal convolucional
 entrenada

 Obtener la etiqueta de la clase predicha

Interactuar según la clasificación obtenida:

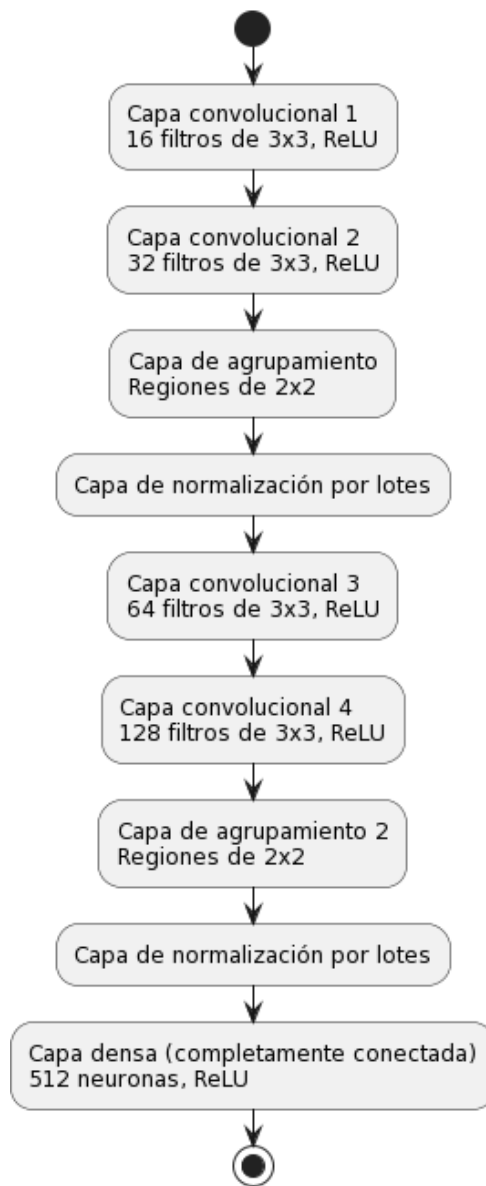
 Si la clase es correcta, enviar los comandos al UAV

end

3.2 Arquitectura de clasificación de imágenes

La arquitectura de la red neuronal convolucional que se muestra en un diagrama en la Figura 3.2, se compone de varias capas, cada una desempeñando un papel crucial en la extracción y comprensión progresiva de características en las imágenes de señales de tráfico. Comienza con capas convolucionales que detectan patrones simples como bordes y texturas, y se profundiza en capas adicionales para reconocer detalles más complejos y específicos. El pseudocódigo proporciona una descripción detallada de la estructura de la red, desde las capas convolucionales iniciales hasta las capas densas completamente conectadas que permiten tomar decisiones finales. Cada capa se describe con el número de filtros, el tamaño de los *kernels* y las funciones de activación utilizadas.

Figura 3.2: Diagrama propuesto de la red neuronal convolucional para clasificación de imágenes.



Fuente: Elaboración propia.

Tabla 3.1: Detalles de la arquitectura propuesta de la red neuronal convolucional para clasificación de imágenes.

Capa	Tipo	Filtros/Neuronas	Kernel	Activación	Operaciones adicionales
Conv1	Convolutacional	16	3x3	ReLU	2x2 MaxPooling
Conv2	Convolutacional	32	3x3	ReLU	2x2 MaxPooling
Pool1	Agrupamiento	-	2x2	-	2x2 MaxPooling
Norm1	Normalización	-	-	-	Batch Normalization
Conv3	Convolutacional	64	3x3	ReLU	2x2 MaxPooling
Conv4	Convolutacional	128	3x3	ReLU	2x2 MaxPooling
Pool2	Agrupamiento	-	2x2	-	2x2 MaxPooling
Norm2	Normalización	-	-	-	Batch Normalization
Dense	Densa	512	-	ReLU	-

Fuente: Elaboración propia.

La Tabla 3.1 presenta una especificación detallada de las capas que componen la arquitectura de una red neuronal convolucional. Cada fila representa una capa diferente, con la columna Tipo indicando el tipo de capa (Convolutacional, Agrupamiento o Densa). La columna Filtros/Neuronas proporciona la cantidad de filtros para las capas convolucionales o neuronas para la capa densa. En la columna *Kernel* se especifica el tamaño del *kernel* utilizado en las capas convolucionales. La columna Activación indica la función de activación aplicada, siendo ReLU en las capas convolucionales y densa. La última columna, Operaciones adicionales, detalla las operaciones específicas realizadas por cada capa, como *MaxPooling* y *Batch Normalization*. Estos elementos describen de manera concisa la configuración y operaciones de cada capa en la CNN.

Esta arquitectura de red neuronal convolucional se ha diseñado con una capa convolutacional de entrada que tiene 16 filtros de 3x3 y una función de activación ReLU para detectar patrones de características simples en las imágenes de entrada como los bordes, esquinas, texturas simples de los tipos de señalamientos que son distinguidos con un conjunto de filtros, y cada filtro escanea la imagen en busca de patrones específicos, como un filtro podría estar diseñado para detectar bordes horizontales, mientras que otro podría buscar esquinas.

El tamaño del *kernel* de 3x3 se utiliza para capturar patrones locales, enfocándose en regiones o áreas pequeñas para comprender las señales que se encuentran en imagen.

La segunda capa se agregan 32 filtros de 3x3 y ReLU para profundizar la extracción e identificación de características y detalles más sofisticados y de alto nivel en las imágenes. A medida que la información se propaga a través de las capas convolucionales, las neuronas en capas posteriores comienzan a reconocer patrones cada vez más complejos y específicos que pueden ser cruciales para distinguir entre diferentes tipos de señales de tráfico. Puede aprender a reconocer características como letras, números, símbolos y combinaciones de formas que son representativas de señales específicas.

Después de las capas convolucionales, se aplica una capa de agrupamiento (*pooling*) dividiendo la entrada en regiones de 2x2 píxeles para reducir la resolución espacial de las características obtenidas en las capas convolucionales. Esto reduce la cantidad de cálculos necesarios en capas posteriores, hace que la red sea invariante ante pequeñas traslaciones en caso de que una característica importante se desplace ligeramente en la imagen, la red aún puede reconocerla y sea menos propensa al sobreajuste.

La capa de normalización es una técnica importante cuando entrenamos una red neuronal, las activaciones y gradientes en cada capa pueden variar significativamente durante el proceso de entrenamiento. Esto puede hacer que el aprendizaje sea más difícil y lento, ya que las capas posteriores tienen que adaptarse constantemente a los cambios en las distribuciones de activaciones de las capas anteriores. Para ayudar a mitigar este problema se agrega una capa de normalización que transforma las activaciones de cada capa para que tengan una media cercana a cero y una desviación estándar cercana a uno.

La arquitectura continúa con dos capas convolucionales adicionales. La primera tiene 64 filtros y la segunda 128, ambas con funciones de activación ReLU. Cada filtro actúa como un detector de características y se desplace por toda la imagen para reconocer patrones específicos.

En el caso de la primera capa, hay 64 de estos detectores de características trabajando simultáneamente para identificar distintos patrones y aprender una amplia variedad de características diferentes, desde bordes y texturas hasta formas más complejas como esquinas, círculos o patrones geométricos. Los 128 filtros de la segunda capa trabajan de

manera similar a los de la capa anterior, pero ahora están sintonizados para reconocer características más específicas y sofisticadas. Estos pueden incluir detalles más finos, como características de texturas más detalladas o componentes de formas más elaboradas.

Después de aplicar las capas convolucionales, se aplica otra capa de *pooling* de tamaño 2x2 para seguir reduciendo la resolución espacial en las salidas de las capas convolucionales. Nuevamente, se aplica una capa de normalización por lotes para estabilizar y transformar las activaciones 2D en un vector unidimensional que puede alimentar una capa completamente conectada.

Se continua con agregar una capa densa completamente conectada con 512 neuronas que cada una de las 512 neuronas en esta capa está conectada a todas las salidas de las capas anteriores, lo que significa que cada neurona puede influir en el resultado final basándose en todas las características extraídas hasta este punto. También tiene una función de activación ReLU (*Rectified Linear Unit*) que se aplica a la salida de cada neurona en la capa densa. Su fórmula es

$$f(x) = \max(0, x)$$

lo que significa que, si la entrada es positiva, la función devuelve el mismo valor, y si es negativa, devuelve cero.

Capítulo 4

Resultados

En esta sección, se presentan y discuten los resultados obtenidos que fueron divididos en dos etapas. En primer lugar, examinamos los resultados experimentales del reconocimiento de imágenes mediante Redes Neuronales Convolucionales, profundizando en el funcionamiento interno de la CNN, la precisión lograda y los datos utilizados para su entrenamiento. La segunda subsección destaca los resultados experimentales de un sistema con CNN y comunicación con AUV, detallando el proceso del vehículo aéreo al capturar y procesar imágenes, su interacción con la CNN y la generación de instrucciones para controlar sus actuadores.

4.1 Resultados experimentales para reconocimiento de imágenes

La implementación de la red neuronal convolucional diseñada para la clasificación de 43 tipos de señales de tráfico ha arrojado resultados altamente prometedores. La elección de una arquitectura cuidadosamente diseñada, combinada con la extensa cantidad de datos de entrenamiento, ha demostrado ser efectiva en la resolución de esta tarea crítica en la conducción autónoma y asistencia de conducción.

La elección de la cantidad de datos para los conjuntos de entrenamiento y validación desempeña un papel esencial en el rendimiento y la generalización de un modelo de aprendizaje profundo, como la CNN implementada en este estudio.

En este caso, se utilizó un conjunto de entrenamiento, que consta de 27,446 imágenes de señales de tráfico. Esta dimensión proporciona una cantidad sustancial de ejemplos, permitiendo que la red neuronal capture una variedad significativa de patrones y características presentes en las señales de tráfico. La elección de imágenes de 30x30 píxeles con 3 canales de color también asegura una representación adecuada de la información visual asociada con las señales, manteniendo un equilibrio entre la complejidad

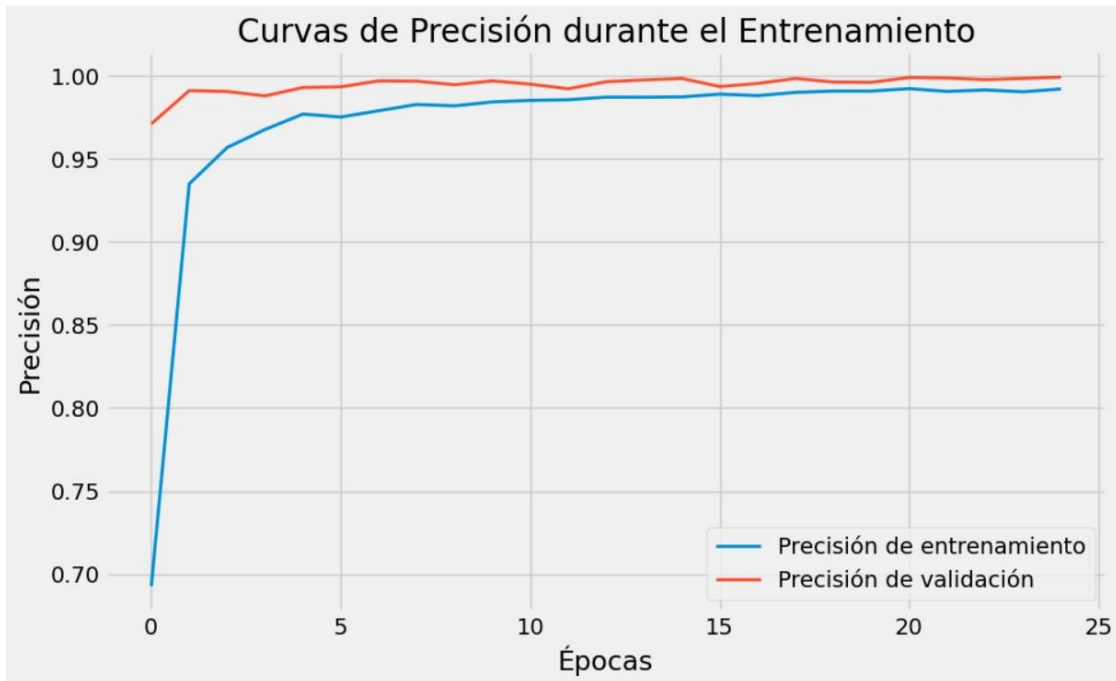
del modelo y la eficiencia computacional.

Por otro lado, el conjunto de validación, compuesto por 11,763 imágenes, se seleccionó para evaluar la capacidad de generalización del modelo en datos no vistos durante el entrenamiento. La dimensión de este conjunto se determinó con el propósito de proporcionar una evaluación robusta sin comprometer la representación de la diversidad presente en las imágenes de señales de tráfico.

Para visualizar de cómo mejora el rendimiento del modelo con el tiempo, se generó una gráfica que se muestra en la Figura 4.1. En el eje vertical, la precisión muestra la capacidad del modelo para clasificar correctamente, mientras que el eje horizontal representa las épocas, o iteraciones a través del conjunto de datos. Esta representación visual permite observar la evolución de la precisión a medida que el modelo se expone repetidamente a los datos de entrenamiento, identificando patrones de aprendizaje, posibles problemas como el sobreajuste, y proporcionando información crucial para la optimización del modelo.

El modelo de red neuronal convolucional entrenado para la clasificación de señales de tráfico ha demostrado un rendimiento notablemente sólido, evidenciado por la gráfica de precisión a lo largo de las épocas. Durante el proceso de entrenamiento, que abarcó desde la época 0 hasta la época 25, se observaron tendencias alentadoras que indican una capacidad efectiva del modelo para aprender y generalizar.

Figura 4.1: Gráfica de precisión resultante del sistema propuesto de clasificación de imágenes.



Fuente: Elaboración propia.

La curva de precisión de entrenamiento, representada por la línea azul en la gráfica, mostró un ascenso significativo en las primeras épocas. Es particularmente destacable que entre las épocas 0 y 5, la precisión de entrenamiento se elevó de manera drástica, superando el umbral del 95%. Este rápido aumento sugiere que el modelo fue capaz de capturar de manera efectiva los patrones y características presentes en las imágenes de señales de tráfico del conjunto de entrenamiento.

La estabilidad y consistencia de la curva de precisión de validación, representada por la línea roja, también son indicativos de un buen rendimiento. La precisión de validación se mantuvo en niveles elevados a medida que avanzaban las épocas, lo cual es esencial ya que indica la capacidad del modelo para generalizar correctamente a nuevas instancias de señales de tráfico no vistas durante el entrenamiento.

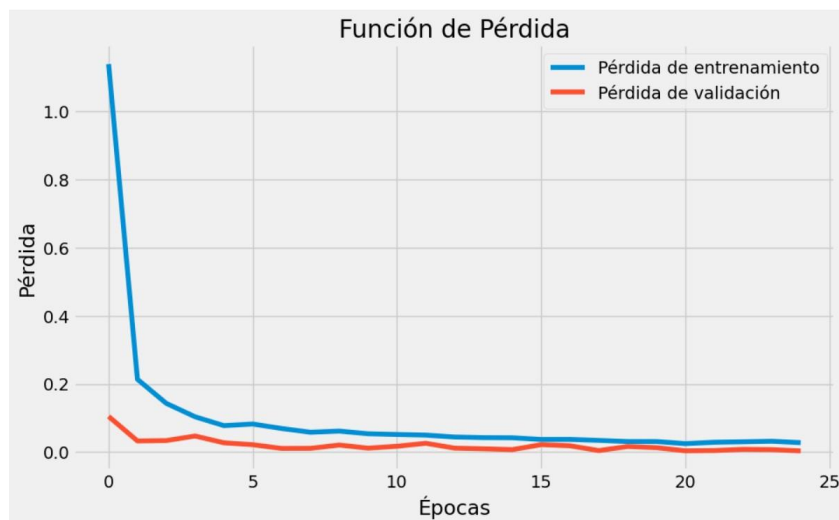
La evaluación visual de la eficacia del modelo de red neuronal convolucional, se encuentra en la Figura 4.2. Esta proporciona una representación visual del cambio en la pérdida del modelo a lo largo del tiempo. En el eje vertical, la pérdida representa la

discrepancia entre las predicciones del modelo y las etiquetas reales, mientras que el eje horizontal representa las épocas, o iteraciones a través del conjunto de datos.

Durante las 25 épocas de entrenamiento, se observó una disminución simultánea en las curvas de pérdida tanto para el conjunto de entrenamiento como para el conjunto de validación. Este comportamiento sugiere que el modelo ha logrado un ajuste efectivo a los datos de entrenamiento, adaptándose a patrones y características clave presentes en las imágenes de señales de tráfico. La sincronización en la reducción de la pérdida refleja una capacidad de generalización, indicando que el modelo no solo aprendió los detalles específicos del conjunto de entrenamiento, sino que también puede realizar predicciones precisas en datos no vistos.

Este fenómeno es crucial en la mitigación del sobreajuste, ya que la convergencia de las curvas de pérdida de entrenamiento y validación sugiere que el modelo no se limita a memorizar datos específicos, sino que ha adquirido la capacidad de reconocer patrones relevantes en nuevas instancias de señales de tráfico. Este equilibrio en la disminución de la pérdida respalda la afirmación de que el modelo ha alcanzado un buen ajuste y puede generalizar de manera efectiva a diversas situaciones.

Figura 4.2: Gráfica de pérdida resultante del sistema propuesto de clasificación de imágenes.



Fuente: Elaboración propia.

Figura 4.3: Ejemplo de una imagen de señalamiento utilizada en las pruebas de entrenamiento.



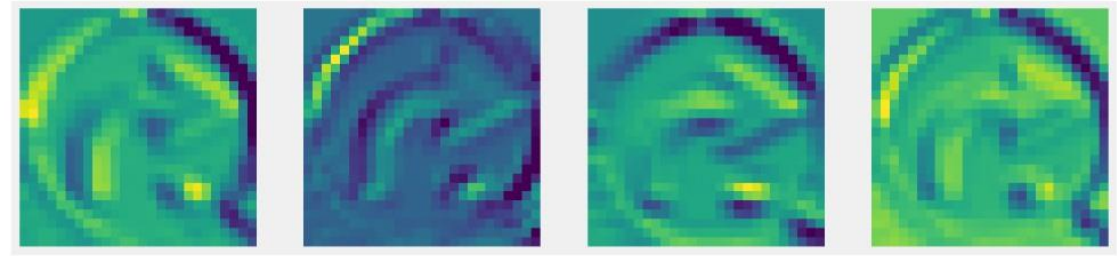
Fuente: Elaboración propia.

La aplicación práctica del modelo de red neuronal convolucional en la clasificación de señales de tráfico se destaca mediante dos ejemplos representativos. En la primera instancia, al presentar al modelo una imagen conteniendo diversas señales de tráfico, se logró una identificación precisa y certera. La red neuronal convolucional demostró su capacidad para discernir y clasificar con éxito las diferentes señales presentes en la imagen, reflejando su eficacia en el reconocimiento de patrones complejos y variados relacionados con señalización vial.

Para implementar y evaluar el rendimiento del modelo de red neuronal convolucional en la clasificación de señales de tráfico, se llevó a cabo un ejercicio práctico utilizando la Figura 4.3 como entrada al sistema. Esta figura representa un señalamiento vial que indica la dirección hacia la derecha y presenta una forma de flecha contenida dentro de un círculo, lo que añade complejidad a la tarea de reconocimiento.

La Figura 4.4 representa la progresión de la identificación de patrones a lo largo de las capas de la red neuronal convolucional al procesar el señalamiento de flecha de la Figura 4.3. Iniciando con la capa Conv2D inicial, la red detecta bordes y características básicas, evolucionando luego en capas subsiguientes para reconocer patrones más complejos y detalles específicos de la señal de tráfico. Las capas de agrupación máxima y normalización por lotes contribuyen a la reducción de dimensionalidad y estabilidad. Las capas finales de convolución y agrupación capturan detalles sutiles, mientras que las capas totalmente conectadas asocian patrones con clases específicas.

Figura 4.4: Parámetros identificados en distintas capas para flecha.



Fuente: Elaboración propia.

Se llevó a cabo otra prueba para evaluar la robustez y versatilidad del modelo de red neuronal convolucional en la clasificación de señales de tráfico. En este caso, se utilizó la Figura 4.5 como entrada al sistema, representando un señalamiento con una estructura diferente. La imagen presenta una forma similar a la de un rombo, con un rombo más pequeño en su interior, lo que introduce variabilidad y complejidad en la tarea de reconocimiento.

Figura 4.5: Ejemplo 2 de una imagen de señalamiento utilizada en las pruebas de entrenamiento.

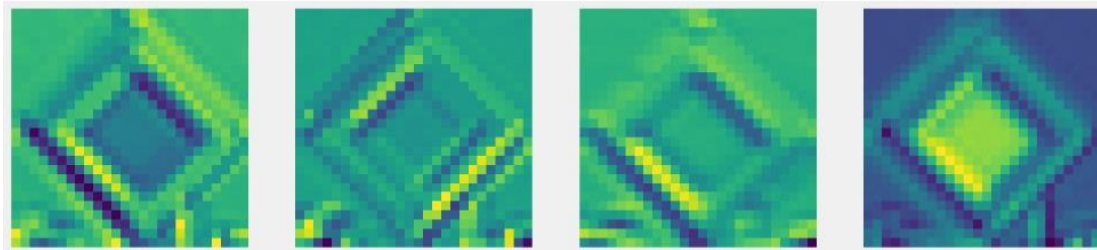


Fuente: Elaboración propia.

La Figura 4.6 ilustra la progresión en la identificación de patrones durante el procesamiento de la red neuronal convolucional al analizar el señalamiento de rombo de la Figura 4.5. Desde la capa Conv2D inicial, la red comienza detectando bordes y características básicas, avanzando para reconocer patrones más complejos y detalles

específicos del rombo de la señal de tráfico. La presencia de un rombo interno en la figura introduce características geométricas únicas, desafiando la capacidad de la red para capturar detalles sutiles. Las capas finales de convolución y agrupación se centran en estos detalles geométricos, permitiendo una representación detallada de la señal de tráfico en su forma de rombo.

Figura 4.6: Parámetros identificados en distintas capas para rombo.



Fuente: Elaboración propia.

Este análisis destaca la adaptabilidad y precisión del modelo en la identificación de características distintivas en señales con formas geométricas específicas, como el rombo en este caso.

4.2 Resultados experimentales para el sistema con la CNN

Durante la fase de resultados experimentales del sistema, que integra una red neuronal convolucional y la comunicación con un vehículo aéreo no tripulado, se implementó una validación de los tipos de imágenes de señalamiento utilizadas para la transmisión de comandos al UAV a través de la conexión entre el sistema y el UAV.

Simultáneamente, se llevaron a cabo pruebas de comunicación para asegurar la efectividad en la transmisión de comandos de ROS desde el sistema hasta el UAV para la confiabilidad de la conexión, que es fundamental para garantizar una respuesta precisa y oportuna.

En la Tabla 4.1, contiene una representación detallada de la definición de imágenes y los comandos asociados para el UAV. Este esquema refleja cómo el sistema establece validaciones para cada salida generada por la CNN. En el proceso de identificación de una imagen específica, el sistema envía un comando correspondiente al UAV para ejecutar una acción determinada.

Cada fila de la tabla representa una imagen única que ha sido sometida al análisis de la CNN. En la columna Imágenes, se exhibe la representación visual correspondiente de la imagen procesada. Simultáneamente, la columna Comandos que proporciona descripciones textuales detalladas de los comandos que se activan como respuesta a la identificación de cada imagen por parte de la CNN.

En la Tabla 4.2 se presentan las imágenes adquiridas por el UAV, transmitidas mediante la conexión establecida entre el sistema y el vehículo aéreo no tripulado. Posteriormente, el sistema procesa estas imágenes y las dirige a la CNN para su clasificación. La CNN identifica la clase correspondiente y devuelve una respuesta al sistema. Esta respuesta es entonces procesada para determinar la validación, lo que permite la ejecución de comandos con instrucciones específicas. Estos comandos son enviados y recibidos por el UAV, completando así el ciclo de adquisición, procesamiento y envío basado en la inteligencia artificial integrada en el sistema.

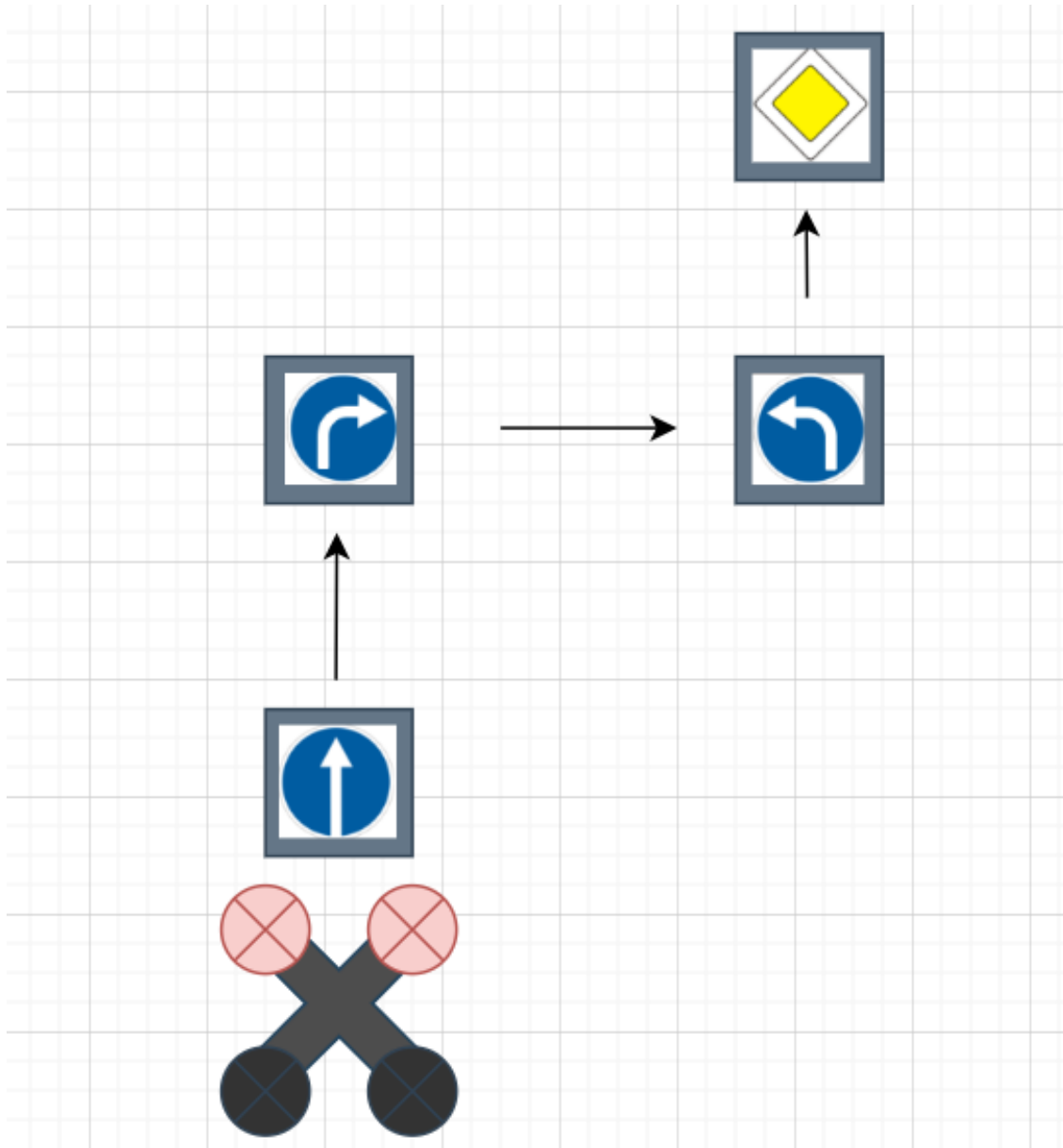
Tabla 4.1: Tabla de imágenes y sus acciones utilizadas en el sistema propuesto.

Imágenes	Acciones
	Avanzar un metro hacia delante
	Girar 90 grados a la derecha y avanzar un metro.
	Girar 90 grados a la izquierda y avanzar un metro.
	Aterrizar

Fuente: Elaboración propia.


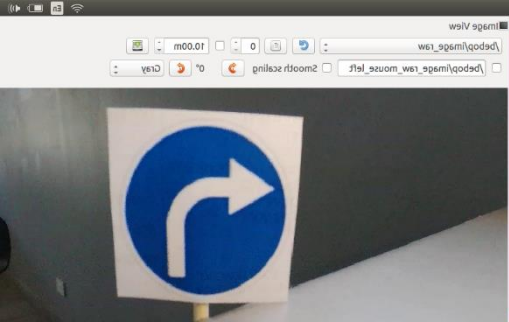


Los comandos transmitidos al vehículo aéreo no tripulado (UAV) consistirán en instrucciones para sus actuadores, con el objetivo de llevar a cabo una misión de recorrido entre las imágenes. Tal como se ilustra en la Figura 4.7, el sistema es capaz de reconocer las imágenes mediante la CNN y simula un recorrido, como se evidencia en el diagrama de flujo de la misión del UAV.

Figura 4.7: Diagrama del flujo de la misión del UAV.



Fuente: Elaboración propia.

Tabla 4.2: Ejemplo de imágenes de entrada adquiridas por el UAV y su respectivo comando de salida.

Imágenes de entrada	Comando de salida
	<pre data-bbox="813 411 1442 520"> armando@armando-VirtualBox: ~ rostopic pub --once bebop/cmd_vel geometry_msgs/Twist '{linear: {x: 1.0, y: 0.0, z: 0.0}, angular: {x: 0.0, y: 0.0, z: 0.0}}'</pre> <p data-bbox="813 552 1260 583">Avanzar un metro hacia delante</p>
	<pre data-bbox="813 646 1458 877"> armando@armando-VirtualBox: ~ : rostopic pub -1 bebop/cmd_vel geometry_msgs/Twist linear: {x: 0.0, y: 0.0, z: 0.0} angular: {x: 0.0, y: 0.0, z: -1.57} sleep 5 rostopic pub -1 bebop/cmd_vel geometry_msgs/Twist linear: {x: 1.0, y: 0.0, z: 0.0} angular: {x: 0.0, y: 0.0, z: 0.0}</pre> <p data-bbox="813 911 1489 982">Girar 90 grados a la derecha y avanzar un metro.</p>
	<pre data-bbox="813 1071 1458 1302"> armando@armando-VirtualBox: ~ rostopic pub -1 bebop/cmd_vel geometry_msgs/Twist linear: {x: 0.0, y: 0.0, z: 0.0} angular: {x: 0.0, y: 0.0, z: 1.57} sleep 5 rostopic pub -1 bebop/cmd_vel geometry_msgs/Twist linear: {x: 1.0, y: 0.0, z: 0.0} angular: {x: 0.0, y: 0.0, z: 0.0}</pre> <p data-bbox="813 1331 1489 1402">Girar 90 grados a la izquierda y avanzar un metro.</p>
	<pre data-bbox="813 1612 1442 1686"> armando@armando-VirtualBox: ~ rostopic pub -1 bebop/land_std_msgs/Empty</pre> <p data-bbox="813 1724 935 1755">Aterrizar</p>

Fuente: Elaboración propia.

Capítulo 5

Conclusiones

Este trabajo de tesis propone un sistema de navegación autónoma de vehículos aéreos utilizando referencia visual. El sistema propuesto aborda de manera precisa la clasificación de imágenes de señales para la navegación mediante un planteamiento de arquitectura neuronal convolucional.

La implementación exitosa de la red neuronal convolucional para la clasificación de señales de tráfico representa un logro significativo. La elección estratégica de esta arquitectura y la utilización de un extenso conjunto de datos han posibilitado la captura eficiente de patrones y características.

El modelo, entrenado con 27,446 imágenes de señales, exhibe una mejora rápida en la precisión de entrenamiento, superando el 95% en las primeras épocas. La consistencia en la precisión de validación indica una notable capacidad de generalización a instancias nuevas.

En términos prácticos, la aplicación del modelo ha destacado su eficacia al identificar señales de tráfico con formas y características distintivas, incluso aquellas con diferencias notables, como señales circulares de dirección a la derecha y señales de forma romboidal o cuadrada. La progresión a través de las capas subraya la habilidad del modelo para reconocer desde detalles fundamentales hasta patrones más complejos en cada señal.

Adicionalmente, la elección acertada de ROS para la manipulación del vehículo aéreo no tripulado ha demostrado ser altamente efectiva. La plataforma proporciona una conexión robusta que garantiza la transmisión confiable de comandos, tanto para el control preciso de los actuadores como para la adquisición de imágenes. Este aspecto es fundamental para la ejecución eficiente de las tareas asignadas al UAV.

La integración de la red neuronal convolucional con el sistema de navegación del vehículo aéreo se realiza de manera apropiada mediante la elección de ROS como

plataforma para la manipulación del vehículo aéreo no tripulado. La plataforma ROS proporciona una conexión robusta que garantiza la transmisión confiable de comandos, facilitando tanto el control preciso de los actuadores como la adquisición de imágenes. Esta integración efectiva entre la red neuronal convolucional y el sistema de navegación del vehículo aéreo contribuye significativamente a mejorar la capacidad del UAV para adaptarse y responder de manera autónoma a su entorno, fortaleciendo su autonomía y eficiencia en operaciones futuras.

5.1 Futuros trabajos

- *Precisión de vuelo*

La implementación cuidadosa de algoritmos de control autónomo y sistemas de estabilización, ya que proporcionarían la capacidad de realizar ajustes dinámicos y continuos durante el vuelo, permitiendo una adaptación precisa a cambios en las condiciones ambientales y la topografía.

- *Planificación de trayectorias*

Explorar la aplicación de algoritmos de aprendizaje profundo y técnicas de *machine learning* podría permitir que el UAV adapte dinámicamente su ruta en función de las condiciones del entorno y de datos en tiempo real.

- *Seguridad*

Implementación de seguridad del sistema, incluyendo la capacidad de resistir posibles interferencias y el uso de medidas de seguridad para evitar situaciones de riesgo durante la operación del UAV.

Bibliografía

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283.
- Ahad, M. T., Li, Y., Song, B., & Bhuiyan, T. (2023). Comparison of cnn-based deep learning architectures for rice diseases classification. *Artificial Intelligence in Agriculture*, 9, 22–35. <https://doi.org/https://doi.org/10.1016/j.aiia.2023.07.001>
- Alshehri, A., Bazi, Y., Ammour, N., Almubarak, H., & Alajlan, N. (2019). Deep attention neural network for multi-label classification in unmanned aerial vehicle imagery. *IEEE Access*, 7, 119873–119880. <https://doi.org/10.1109/ACCESS.2019.2936616>
- Arshad, M. A., et al. (2022). Drone navigation using region and edge exploitation-based deep cnn. *IEEE Access*, 10, 95441–95450. <https://doi.org/10.1109/ACCESS.2022.3204876>
- Bah, M. D., Hafiane, A., & Canals, R. (2020). Crownnet: Deep network for crop row detection in uav images. *IEEE Access*, 8, 5189–5200. <https://doi.org/10.1109/ACCESS.2019.2960873>
- Beeharee, A., & Clare, L. (2006). A kalman filter-based algorithm for imu/magnetometer sensor fusion using an adaptive likelihood function. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

- Behera, T. K., Bakshi, S., Nappi, M., & Sa, P. K. (2023). Superpixel-based multiscale cnn approach toward multiclass object segmentation from uav-captured aerial images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 1771–1784. <https://doi.org/10.1109/JSTARS.2023.3239119>
- Bengio, Y. (2016). *Deep learning*. MIT Press.
- Beul, M., Benning, L., Bodenmüller, T., & Scherer-Negenborn, N. (2020). A survey of visual and visual-inertial navigation for small unmanned aerial vehicles. *Journal of Field Robotics*, 37(4), 553–589.
- Bradski, G., & Kaehler, A. (2008a). *Learning opencv: Computer vision with the opencv library*. O'Reilly Media, Inc.
- Bradski, G., & Kaehler, A. (2008b). *Learning opencv: Computer vision with the opencv library*. O'Reilly Media, Inc.
- Brown, A., & Wang, L. (2018). Advanced gyroscopes for uavs: A review. *Sensors*, 18(6), 1746.
- Choi, Y. J., Rahim, T., Ramatryana, I. N. A., & Shin, S. Y. (2021). Improved cnn-based path planning for stairs climbing in autonomous uav with lidar sensor. *2021 International Conference on Electronics, Information, and Communication (ICEIC)*, 1–7. <https://doi.org/10.1109/ICEIC51217.2021.9369805>
- Downey, A. (2012). *Think python: How to think like a computer scientist*. Green Tea Press.
- Elitaş, M., & Zengin, A. (2019). Evaluation of barometric pressure sensors for altitude estimation in unmanned aerial vehicles. *Sensors*, 19(9), 1980.
- Farfán, F. (2005, January). *Control cerebral de interfases: Análisis exploratorio de técnicas paramétricas digitales para la detección y cuantificación de estados mentales* [Doctoral dissertation]. <https://doi.org/10.13140/RG.2.1.1649.1123>
- Fikri, M. Y., et al. (2019). Clustering green openspace using uav (unmanned aerial vehicle) with cnn (convolutional neural network). *2019 International Symposium on Electronics and Smart Devices (ISESD)*, 1–5. <https://doi.org/10.1109/ISESD.2019.8909502>

- González, R. C., Woods, R. E., & Eddins, S. L. (2018). *Procesamiento digital de imágenes*. Pearson Educación.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016a). *Deep learning* (Vol. 1). MIT press Cambridge.
- Guclu, U., & van Gerven, M. A. (2017). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 37(13), 3410–3418.
- Gwon, G.-H., Lee, J. H., Kim, I.-H., & Jung, H.-J. (2023). Cnn-based image quality classification considering quality degradation in bridge inspection using an unmanned aerial vehicle. *IEEE Access*, 11, 22096–22113. <https://doi.org/10.1109/ACCESS.2023.3238204>
- Hagen, T., & Ruotsalainen, L. (2018). Comparison of smartphone and dedicated sensors in measuring magnetic field for indoor positioning. *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 1–8.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with numpy. *Nature*, 585(7825), 357–362.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Kellenberger, B., Marcos, D., Courty, N., & Tuia, D. (2018). Detecting animals in repeated uav image acquisitions by matching cnn activations with optimal transport. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 3643–3646. <https://doi.org/10.1109/IGARSS.2018.8519012>

- Kim, J., Kang, J.-K., & Kim, Y. (2022). A low-cost fully integer-based cnn accelerator on fpga for real-time traffic sign recognition. *IEEE Access*, *10*, 84626–84634. <https://doi.org/10.1109/ACCESS.2022.3197906>
- Kim, S., Philip, O. A., Tullu, A., & Jung, S. (2023). Development and verification of a ros-based multi-dof flight test system for unmanned aerial vehicles. *IEEE Access*, *11*, 37068–37081. <https://doi.org/10.1109/ACCESS.2023.3267128>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks.
- Kulesza, W. J., Nitschke, G., & Sikora, A. (2018). The pressure field as a source of information for an autonomous flying robot. *Sensors*, *18*(7), 2189.
- Kyrkou, C., & Theocharides, T. (2020). Emergencynet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 1687–1699. <https://doi.org/10.1109/JSTARS.2020.2969809>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
- Li, L., Zhang, W., & Li, Y. (2019). Attitude estimation of uavs using a quaternion-based complementary filter. *Sensors*, *19*(6), 1329.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60–88.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*(4), 115–133.
- McKinney, W. (2012). *Python for data analysis*. O'Reilly Media, Inc.

- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*.
- Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. Parrot. (n.d.). *Parrot official website*. <https://www.parrot.com/en>
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., & Ng, A. Y. (2009). Ros: An open-source robot operating system. *ICRA workshop on open source software*.
- Rahman, M. T., Choudhury, I. A., & Rahman, M. A. (2016). Optimization of performance parameters of a brushless dc motor. *2016 IEEE International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–4.
- Rajagopal, A., et al. (2020). A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles. *IEEE Access*, 8, 135383–135393. <https://doi.org/10.1109/ACCESS.2020.3011502>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks.
- Rivas-Asanza, W., Mazon-Olivo, B., & Mejia, F. (2018, June). Capítulo 1: Generalidades de las redes neuronales artificiales.
- Rohan, A., Rabah, M., & Kim, S.-H. (2019). Convolutional neural network-based real-time object detection and tracking for parrot ar drone 2. *IEEE Access*, 7, 69575–69584. <https://doi.org/10.1109/ACCESS.2019.2919332>
- ROS. (December 6th, 2023). *Core components* [En línea]. <https://www.ros.org/core-components/>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

- Silva, J., & Santos, T. (2018). Three-axis accelerometer calibration using dynamic model of rigid body rotation. *Measurement*, 117, 13–22.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.
- Tavanaei, A., Ghodrati, M., & Mousavi, P. (2019). Deep learning in magnetoencephalography: Shedding light on the dark data. *Neuroinformatics*, 17(1), 57–77.
- Yang, J., Zheng, P., & Li, Y. (2017). Research on a new control method of a brushless dc motor. *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1030–1035.
- Zhu, L., & Mourikis, A. I. (2017). Multi-sensor fusion for 6-dof mav state estimation: A comparison of filter-based and optimization-based methods. *IEEE Transactions on Robotics*, 33(2), 397–404.

ANEXOS



MAESTRÍA EN INGENIERÍA E INNOVACIÓN FORMATO DE AUTORIZACIÓN DE PROYECTO DE APLICACIÓN

Matrícula y Nombre del Alumno:	Especialidad (LGAC):	CAMPUS	Empresa:
041298 FRANCISCO ARMANDO CASTILLO VAZQUEZ	TECNOLOGÍAS DE LA INFORMACIÓN	MEXICALI	CETYS UNIVERSIDAD
Nombre del proyecto:			
NAVEGACIÓN AUTÓNOMA DE VEHÍCULOS AÉREOS POR REFERENCIA VISUAL			
Describir LA PROBLEMÁTICA QUE se propone resolver con el proyecto:			
LA NAVEGACIÓN AUTÓNOMA DE VEHÍCULOS AÉREOS NO TRIPULADOS EN ENTORNOS DINÁMICOS Y CAMBIANTES PRESENTA DESAFÍOS SIGNIFICATIVOS, COMO OBTENER Y PROCESAR INFORMACIÓN DEL ENTORNO. TAMBIÉN PARA LOS SENSORES INERCIALES, COMO EL GPS, PUEDEN SER LIMITADOS EN ENTORNOS DONDE LA SEÑAL GPS ES DÉBIL O INTERMITENTE, Y PUEDE NO SER SUFICIENTEMENTE PRECISA PARA MANIOBRAS DE ALTA COMPLEJIDAD. POR ESTO, LA REFERENCIA VISUAL APOYARÁ LA NAVEGACIÓN AUTÓNOMA DE VEHÍCULOS AÉREOS NO TRIPULADOS CON REFERENCIA VISUAL PARA EVITAR OBSTÁCULOS, SEGUIR RUTAS PREDEFINIDAS Y LOGRAR UN VUELO MÁS SEGURO Y EFICIENTE.			
Describir LA INNOVACIÓN que se propone en el proyecto:			
DESARROLLAR Y EVALUAR UN SISTEMA DE NAVEGACIÓN AUTÓNOMA PARA VEHÍCULOS AÉREOS NO TRIPULADOS BASADO EN REFERENCIA VISUAL, HACIENDO USO DE APRENDIZAJE PROFUNDO Y DE REDES NEURONALES CONVOLUCIONALES PARA MEJORAR LA PERCEPCIÓN Y TOMA DE DECISIONES DE LOS VEHÍCULOS AÉREOS NO TRIPULADOS EN BASE A LAS IMÁGENES CAPTADAS COMO PUNTOS DE REFERENCIA.			
Indique que mecanismo de protección de PROPIEDAD INTELECTUAL utilizará:	<input checked="" type="checkbox"/> Derecho de autor (software, documento, libro)	<input type="checkbox"/> Patente (registro IMPI)	
	<input type="checkbox"/> Diseño industrial (Dibujos, planos 3D)	<input type="checkbox"/> Secreto industrial (documento no abiertos)	
	<input type="checkbox"/> Modelo de utilidad (documento)	<input type="checkbox"/> No se pretende generar propiedad intelectual	
Valor Potencial de la Innovación			
Ganancia estimada por la innovación: (Estimación económica de la reducción de costos en el proceso o monto de incremento en ventas, anual)	N/A		
Valor Social			
Explique cómo el proyecto impacta de manera positiva a la sociedad (sustentabilidad, nivel de vida empleados, mejora de la comunidad, medio ambiente, educación, etc.)			
EL IMPACTO POSITIVO DEL PROYECTO ES MEJORAR EL TIEMPO DE RESPUESTA A LA COMUNIDAD ANTE CATÁSTROFES NATURALES, APOYANDO A LAS COMPLICACIONES EMERGENTES PARA EL EQUIPO DE RESCATE AL IDENTIFICAR Y AUXILIAR A LAS PERSONAS QUE SE ENCUENTREN EN PELIGRO.			
AUTORIZACIÓN			
Nombre:	DR. ULISES OROZCO ROSAS	Firma:	
Puesto:	PROFESOR DE TIEMPO COMPLETO		
Fecha:	5 DE DICIEMBRE DE 2023		
Teléfono:	+52 (664) 329 2494		
Email:	ULISES.OROZCO@CETYS.MX		