

**Escuela Técnica Superior de Ingeniería**

**Universidad de Huelva**

Grado en Ingeniería Informática

**Trabajo Fin de Grado**

Aplicaciones de estrategias de Deep Learning para la detección de animales en imágenes de foto-trampeo

Carlos García Silva

¿?/06/2025

**Resumen**

La importancia de un futuro sostenible se refleja cada día con mayor claridad en nuestra sociedad. Es por ello por lo que los estudios de biodiversidad han cobrado un papel fundamental en la compresión, conservación y valoración de la vida en la Tierra.

Una de las técnicas ampliamente utilizada para la realización de dichos estudios es la técnica del foto-trampeo. Se basa en el despliegue de cámaras automáticas en el entorno a estudiar para poder capturar imágenes de animales en su hábitat natural. Uno de sus principales problemas es la gran facilidad para generar grandes volúmenes de imágenes, en las que realizar un análisis manual puede resultar costoso y laborioso.

Este Trabajo de Fin de Grado presenta una red neuronal convolucional capaz de clasificar imágenes de foto-trampeo para identificar aquellas donde existe presencia de animales.

El sistema propuesto ha sido desarrollado mediante el uso de la técnica de *transfer learning*, empleando un modelo previamente entrenado con los pesos de *ImageNet,* basado en la arquitectura *EfficientNet-B5*. La implementación de *MegaDetector* como preprocesamiento previo de las imágenes de entrada ha resultado en un rendimiento superior en comparación con otros modelos que no han hecho uso de dicha implementación.

El experimento se realizó utilizando una colección de 31.670 imágenes obtenidas de cámaras de foto-trampeo, de las cuales 21.670 presentaban evidencia de presencia animal. El proceso de entrenamiento, validación y evaluación de la red se llevó a cabo utilizando una proporción del 70%, 15% y 15% del conjunto de imágenes, respectivamente.

**[TO-DO] [PÁRRAFO DE RESULTADOS]**.

**[TO-DO] [PÁRRAFO FINAL DE CONCLUSIONES Y VALORACIÓN]**.

**Abstract**

The significance of a sustainable future is becoming increasingly evident within our society. Consequently, the study of biodiversity has assumed a pivotal role in the comprehension, preservation and estimation of life on Earth.

A prevalent technique in the execution of such research is the photo-trapping technique. This method involves the implementation of automated cameras within the designated study environment, with the objective of capturing images of animals within their natural habitat. A salient challenge associated with this approach pertains to the generation of voluminous image data, which renders manual analysis to be both costly and time-consuming.

The present Final Degree Project proposes a convolutional neural network with the capacity to classify photo-trapping images to identify those in which animals are present.

The proposed system has been developed using the transfer learning technique, employing a model previously trained with ImageNet weights, based on the EfficientNet-B5 architecture. The implementation of MegaDetector as a pre-processing of the input images has resulted in a superior performance compared to other models that have not made use of such implementation.

The experiment was conducted using a collection of 31.670 images obtained from photo-trapping cameras, of which 21.670 showed evidence of animal presence. The network training, validation and evaluation process was carried out using a proportion of 70%, 15% and 15% of the image set, respectively.

**[TO-DO] [PÁRRAFO DE RESULTADOS]**.

**[TO-DO] [PÁRRAFO FINAL DE CONCLUSIONES Y VALORACIÓN]**.

**Índice**

[Capitulo 1 7](#_Toc192153241)

[1.1 Motivación 7](#_Toc192153242)

[1.2 Objetivos 8](#_Toc192153243)

[1.3 Competencias 8](#_Toc192153244)

[1.4 Hardware y Software 8](#_Toc192153245)

[1.5 Organización de la memoria 8](#_Toc192153246)

[Capítulo 2: Introducción y Estado del arte 9](#_Toc192153247)

[Capítulo 3: Materiales 10](#_Toc192153248)

[3.1 Base de datos 10](#_Toc192153249)

[3.2 Conjuntos de datos 10](#_Toc192153250)

[3.3 Métricas de evaluación 10](#_Toc192153251)

[Capítulo 4: Metodología 11](#_Toc192153252)

[4.1 Arquitecturas de la red 11](#_Toc192153253)

[4.2 Hiperparámetros 11](#_Toc192153254)

[4.3 Fase de entrenamiento 11](#_Toc192153255)

[Capítulo 5: Resultados y Discusión 12](#_Toc192153256)

[5.1 Resultados en el conjunto de test 12](#_Toc192153257)

[5.2 Análisis y Discusión 12](#_Toc192153258)

[Capítulo 6: Conclusiones y Trabajos Futuros 13](#_Toc192153259)

[6.1 Conclusiones técnicas 13](#_Toc192153260)

[6.2 Trabajos futuros 13](#_Toc192153261)

[6.3 Valoración personal 13](#_Toc192153262)

[1.- Introducción al Deep-Learning 15](#_Toc192153263)

[2.- Redes Neuronales 16](#_Toc192153264)

[2.1.- Fundamentos 16](#_Toc192153265)

[2.2.- Desarrollo de redes neuronales 16](#_Toc192153266)

[2.2.1- Etapa de desarrollo 16](#_Toc192153267)

[2.2.2.- Etapa de entrenamiento: aspectos generales 16](#_Toc192153268)

[2.2.3.- Etapa de entrenamiento: Hiperparámetros 16](#_Toc192153269)

[2.2.4.- Etapa de entrenamiento: Control y seguimiento 16](#_Toc192153270)

[2.2.5.- Etapa de inferencia 16](#_Toc192153271)

[3.- Redes Neuronales Convolucionales 17](#_Toc192153272)

[3.1.- Introducción 17](#_Toc192153273)

[3.2.- Tipos de capas 17](#_Toc192153274)

[3.3.- Arquitecturas populares 17](#_Toc192153275)

[3.4.- Arquitectura utilizada en este trabajo 17](#_Toc192153276)

[4.- MegaDetector 18](#_Toc192153277)

# Capitulo 1

**Propuesta de Proyecto**

En el presente capítulo introductorio, se exponen las motivaciones que han fundamentado la realización de este trabajo, los objetivos que se han establecido y el sistema implementado para su consecución. Asimismo, se detalla el desarrollo de dichos objetivos y las diversas tecnologías empleadas.

## Motivación

En años recientes, se han observado significativos avances en el ámbito de la tecnología, lo cual ha generado un impacto sustancial en la sociedad contemporánea. En cierta medida, se puede afirmar que la dependencia hacia esta área es una realidad en el contexto cotidiano. Diversos analistas consideran que la actual situación representa una nueva revolución industrial. En este sentido, se puede afirmar que, del mismo modo que sucedió con la industria textil en la primera revolución industrial, la electricidad en la segunda y la electrónica en la tercera, hoy en día se puede constatar que la informática se erige como uno de los pilares que han permitido identificar a esta nueva revolución como la denominada *Industria 4.0*.

En el campo de la informática, se ha observado un aumento significativo en la implementación de algoritmos y técnicas de inteligencia artificial en los últimos años. Este fenómeno se atribuye al incremento en la capacidad de procesamiento y al vasto volumen de datos disponibles en la actualidad, en contraste con la limitada disponibilidad de hace unos años. En particular, nos referimos a los algoritmos de aprendizaje profundo (*Deep Learning*)[[1]](#footnote-1), que se fundamentan principalmente en redes neuronales artificiales.

Dentro del ámbito de la inteligencia artificial, los algoritmos han experimentado una notable evolución en su capacidad de procesamiento de datos. La implementación de estos algoritmos se lleva a cabo durante el proceso de entrenamiento, mediante la utilización de vastas cantidades de datos. Además, estos algoritmos se ejecutan en sistemas con capacidades de cómputo que, hasta hace una década, se consideraban inalcanzables para las máquinas. Los resultados obtenidos recientemente han superado las expectativas, logrando niveles de rendimiento que anteriormente eran impensables para las máquinas y, en algunos casos, incluso superando a los humanos en ciertas tareas. Este avance representa una significativa apertura a la capacidad de abordar nuevos y complejos problemas que, anteriormente, solo podían resolverse mediante el ingenio humano.

El propósito de este estudio es examinar y desentrañar la implementación del *Deep Learning* en un amplio espectro de aplicaciones tecnológicas. En este sentido, se centrará en la visión artificial para concebir un sistema de detección de animales en imágenes de foto-trampeo. El enfoque metodológico seleccionado se fundamenta en la utilización de *MegaDetector[[2]](#footnote-2)*, aprovechando sus capacidades y descartando las áreas de escaso interés de las imágenes. Posteriormente, se implementa una red neuronal convolucional (CNN)[[3]](#footnote-3) para la clasificación de las imágenes, permitiendo distinguir entre aquellas que contienen animales y las que no. Este procedimiento se lleva a cabo con el propósito de mejorar la precisión y la eficiencia del proceso de monitoreo de la fauna en comparación con un sistema similar que no incorpora el preprocesamiento de las imágenes de entrada.

Este proyecto constituye una oportunidad para profundizar en el conocimiento y la aplicación de técnicas avanzadas de *Deep Learning*, que no se abordan en el plan de estudios del Grado en Ingeniería Informática. Además, posee una clara relevancia práctica en el ámbito de la conservación de la biodiversidad. La implementación de CNN en la clasificación de imágenes de foto-trampeo permite optimizar el monitoreo de la fauna, facilitando el análisis de datos con mayor eficiencia y precisión. Este avance no solo contribuye al desarrollo de nuevas metodologías en el procesamiento de imágenes, sino que también abre la posibilidad de generar herramientas de apoyo en estudios ecológicos y en la preservación de especies, lo que destaca la importancia de la inteligencia artificial en la resolución de problemas del mundo real.

## Objetivos

En el marco de la presente investigación, se han planteado dos objetivos primordiales a alcanzar:

* Adquisición de conocimiento: Introducción y estudio teórico de las técnicas de *Deep Learning*, con el propósito de comprender su naturaleza, su evolución hasta la actualidad y los fundamentos necesarios para su aplicación en el problema específico de interés.
* Implementación: En segundo lugar, se implementará un sistema basado en *CNN* que permitirá llevar a cabo una detección y clasificación eficaces de la presencia de animales en imágenes de foto-trampeo.

Para una mayor profundización en el tema, se propone una subdivisión de los objetivos principales en los siguientes puntos: En primer lugar, es necesario realizar una exhaustiva revisión bibliográfica y un análisis de los conceptos teóricos básicos acerca del *Deep Learning*. En segundo lugar, se obtendrá, analizará y preparará un conjunto de imágenes[[4]](#footnote-4) para el entrenamiento, la validación y la evaluación de los modelos implementados. En tercer lugar, se procederá al diseño, implementación y evaluación de varios modelos entrenados para la detección de animales en imágenes de foto-trampeo. Por último, se llevará a cabo una búsqueda e implementación de métricas para la evaluación de los modelos, que permitirá cuantificar objetivamente la calidad de los resultados obtenidos.

## Competencias

El desarrollo de este proyecto ha permitido la aplicación y el refuerzo de diversas competencias del Grado en Ingeniería Informática, particularmente aquellas vinculadas con el aprendizaje computacional y el diseño e implementación de sistemas basados en inteligencia artificial. Especificamente, se ha concentrado en el estudio y la aplicación de las CNN para clasificación de imágenes de foto-trampeo. Este proceso ha implicado una revisión exhaustiva del estado del arte en *Deep Learning*, así como la experimentación con modelos diseñados específicamente para la detección de fauna en imágenes digitales.

En contraste, la imperativa de operar con *datasets* de un entorno real ha posibilitado la evolución de competencias en la extracción automática de información y filtrado de imágenes para potenciar la precisión de los modelos. En este sentido, el trabajo no solo ha servido como un ejercicio de profundización en técnicas de *Deep Learning* que no se abordan en profundidad en el plan de estudios, sino también ha permitido aplicar estos conocimientos a un caso práctico de relevancia en el ámbito de la conservación de la fauna, evidenciando la importancia del aprendizaje computacional en la solución de problemas en situaciones reales.

## Hardware y Software

Para poder alcanzar las metas anteriormente establecidas de manera efectiva, resulta necesario el uso de un sistema cuyo hardware posea la capacidad de ejecutar algoritmos de *Deep Learning.* Además, dicho sistema debe estar equipado con un software especializado en el diseño y creación de modelos que hagan uso de estos algoritmos.

En lo que respecta al software empleado, la implementación del modelo se realizará en el lenguaje de programación Python 3.8 debido a su amplia popularidad en aplicaciones de *Machine Learning[[5]](#footnote-5)*. (ver Figura 1).

Logotipo, nombre de la empresa

Descripción generada automáticamente

Figura 1: Principales frameworks de desarrollo de modelos de Deep Learning

El marco de trabajo seleccionado para la ejecución de este proyecto es *TensorFlow*, debido a su exhaustiva documentación y activa comunidad de usuarios, lo que facilitará la implementación del proyecto. Desde su versión 2.0, TensorFlow incorpora de forma nativa la librería Keras, que permite la creación de redes neuronales desde un alto nivel gracias a su estructura de capas de abstracción. Esta ingración proporciona una metolodogía eficiente para la concepción de modelos de *Deep Learning*, permitiendo descender en la abstracción solo cuando sea necesario y de forma más precisa.

Para la implementación y el entrenamiento del modelo de clasificación se utilizarán aplicaciones como Jupyter Notebook y Visual Studio Code, junto con el entorno de desarrollo proporcionado por Anaconda, que nos permitirá instalar y gestionar fácilmente las librerías necesarias, como Tensorflow, OpenCV, Numpy y otras más.

Para la implementación del proyecto, se ha utilizado un MacBook Pro de 14 pulgadas, equipado con el chip Apple M3 (ver Figura 2)[[6]](#footnote-6), 16 GB de memoria unificada y un almacenamiento SSD de 512 GB. Este dispositivo ha permitido la ejecución eficiente de los experimentos, gracias a su CPU de 8 núcleos, optimizada para tareas de alto rendimiento y eficiencia, así como su GPU de 10 núcleos, que incorpora tecnologías avanzadas como Dynamic Caching, Mesh Shading y Ray Tracing acelerado por hardware. Además, la librería TensorFlow ha experimentado actualizaciones que han posibilitado su compatibilidad nativa con la nueva arquitectura de procesadores Apple Silicon, lo que ha conducido a un notable incremento en su rendimiento. La autonomía del dispositivo, con una duración de hasta 22 horas de batería, ha facilitado la realización de entrenamientos de larga duración sin interrupciones, optimizando el flujo de trabajo durante el desarrollo del proyecto.

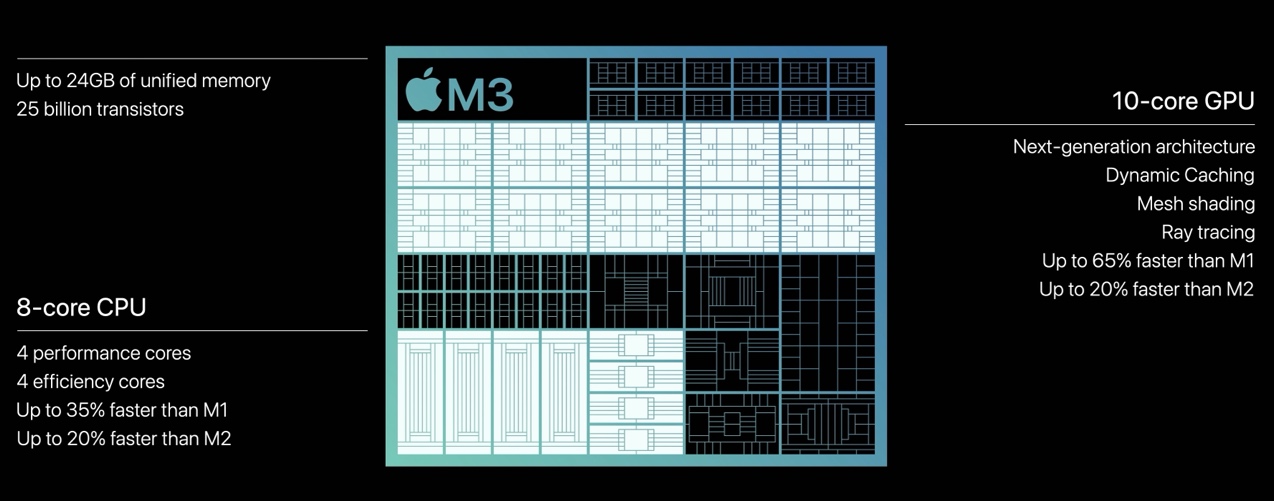


Figura 2: Arquitectura de procesador M3 de Apple

## Organización de la memoria

Además del primer capítulo, en el que se presenta la propuesta, la motivación y los objetivos del trabajo, así como las competencias desarrolladas y los recursos hardware y software empleados, esta memoria se organiza en los siguientes capítulos:

* **Capítulo 2 Introducción y estado de arte:** Se presenta un análisis exhaustivo de la implementación de técnicas de *Deep Learning* en el ámbito de la clasificación de imágenes. Este capítulo examina estudios previos y tecnologías relevantes en el contexto del reconocimiento de fauna mediante foto-trampeo.
* **Capítulo 3 Materiales:** Se aborda la descripción de los materiales empleados en el desarrollo del modelo. En este capítulo se describen los datos utilizados en la implementación del modelo, incluyendo la base de datos empleada, los conjuntos de datos generados para el entrenamiento, validación y evaluación, y las métricas utilizadas para medir el desempeño del sistema.
* **Capítulo 4 Metodología:** Como su propio nombre indica, se centra en la metodología empleada, detallando las estrategias implementadas en el desarrollo del modelo. En este sentido, se aborda la selección de la arquitectura de la red neuronal, los hiperparámetros utilizados y el proceso de entrenamiento del modelo.
* **Capítulo 5 Resultados y discusión:** Presenta los resultados obtenidos en el conjunto de evaluación, analizando su rendimiento en términos de las métricas establecidas y discutiendo las fortalezas y limitaciones del enfoque adoptado.
* **Capítulo 6 Conclusiones y trabajos futuros:** Ofrece una reflexión final sobre las implicaciones del estudio realizado y plantea posibles líneas de trabajo para futuras investigaciones. En este capítulo se presentan las conclusiones técnicas extraídas del estudio, se identifican posibles mejoras y líneas de trabajo futuro, y se incluye una valoración personal sobre la experiencia adquirida durante la realización del proyecto.

La memoria prosigue con el capítulo destinado a la bibliografía, en el cual se efectúan las referencias de los artículos, libros y recursos empleados para la elaboración del trabajo.

Además de los capítulos principales, la memoria contiene un anexo teórico adicional que profundiza en los conceptos fundamentales del trabajo. Organizado de la siguiente manera:

* **Apartado 1 Introducción al Deep Learning:** Se presentan los principios básicos del *Deep Learning*, destacando su evolución y relevancia en la actualidad.
* **Apartado 2 Redes Neuronales:** En este apartado se abordan los fundamentos del funcionamiento de las redes neuronales, desde su desarrollo hasta las diferentes etapas del entrenamiento e inferencia, detallando aspectos como la configuración de hiperparámetros y el control del proceso de entrenamiento.
* **Apartado 3 Redes Neuronales Convolucionales:** Se analiza la estructura y funcionamiento de las CNN, describiendo los tipos de capas utilizadas, las arquitecturas más populares y la arquitectura específica empleada en este trabajo.
* **Apartado 4 MegaDetector:** Finalmente se presenta MegaDetector, utilizado para el preprocesamiento de las imágenes de foto-trampeo utilizadas como entrada para nuestro modelo clasificador, explicando su funcionamiento y su integración dentro del sistema desarrollado.

# Capítulo 2: Introducción y Estado del arte

# Capítulo 3: Materiales

## 3.1 Base de datos

## 3.2 Conjuntos de datos

## 3.3 Métricas de evaluación

# Capítulo 4: Metodología

## 4.1 Arquitecturas de la red

## 4.2 Hiperparámetros

## 4.3 Fase de entrenamiento

# Capítulo 5: Resultados y Discusión

## 5.1 Resultados en el conjunto de test

## 5.2 Análisis y Discusión

# Capítulo 6: Conclusiones y Trabajos Futuros

## 6.1 Conclusiones técnicas

## 6.2 Trabajos futuros

## 6.3 Valoración personal

**Bibliografía**

**Anexo Teórico**

# 1.- Introducción al Deep-Learning

# 2.- Redes Neuronales

## 2.1.- Fundamentos

## 2.2.- Desarrollo de redes neuronales

### 2.2.1- Etapa de desarrollo

### 2.2.2.- Etapa de entrenamiento: aspectos generales

### 2.2.3.- Etapa de entrenamiento: Hiperparámetros

### 2.2.4.- Etapa de entrenamiento: Control y seguimiento

### 2.2.5.- Etapa de inferencia

# 3.- Redes Neuronales Convolucionales

## 3.1.- Introducción

## 3.2.- Tipos de capas

## 3.3.- Arquitecturas populares

## 3.4.- Arquitectura utilizada en este trabajo

# 4.- MegaDetector

1. Debido a la popularidad del término Deep Learning, durante el resto del trabajo lo utilizaremos en sustitución del término aprendizaje profundo. [↑](#footnote-ref-1)
2. Enlace oficial al modelo de detección: <https://github.com/microsoft/CameraTraps/tree/main> [↑](#footnote-ref-2)
3. Siglas del término inglés Convolutional Neural Network. Por motivos de simplicidad, durante el resto del trabajo utilizaremos dichas siglas en sustitución del término red neuronal convolucional. [↑](#footnote-ref-3)
4. Debido a su popularidad, a partir de ahora el término inglés *Dataset* sustituirá el término conjunto de datos / imágenes. [↑](#footnote-ref-4)
5. Término inglés cuyo significado se corresponde con el concepto de aprendizaje automático. [↑](#footnote-ref-5)
6. Fuente: <https://www.notebookcheck.org/fileadmin/Notebooks/Apple/apple-m3-stats.jpg> [↑](#footnote-ref-6)