

Universidad de Huelva

Departamento de Ciencias Integradas



Paradigmas de aprendizaje profundo: soluciones metodológicas de aplicación a problemas de interés

Memoria para optar al grado de doctor
presentada por:

Isaac Pérez Borrero

Fecha de lectura: 13 de mayo de 2022

Bajo la dirección de los doctores:

Manuel Emilio Gegúndez Arias

Diego Marín Santos

Huelva, 2022



Universidad de Huelva

Doctorado en Economía, Empresa, Finanzas y Computación



**Universidad
de Huelva**

**Paradigmas de aprendizaje
profundo: soluciones metodológicas
de aplicación a problemas de interés**

Memoria para optar al grado de doctor presentada por:

Isaac Pérez Borrero

Bajo la dirección de los doctores:

Manuel Emilio Gegúndez Arias

Diego Marín Santos

2022

*Dedicado a
mis padres*

Resumen

En esta tesis doctoral se ha llevado a cabo un estudio teórico y experimental en el campo del aprendizaje profundo. El objetivo de este estudio ha sido mejorar las técnicas actuales que se emplean en diferentes problemas de interés para la sociedad.

Los principales problemas que presentan los modelos de aprendizaje profundo a la hora de ser implementados en entornos reales son la falta de precisión y el elevado coste computacional de su ejecución. Existen problemas en los que es necesario que la precisión del modelo sea lo más alta posible para que este tipo de soluciones puedan ser utilizadas. Además, la demanda de procesamiento de los modelos puede impedir su uso en ciertos problemas en los que la capacidad de procesamiento se encuentra muy limitada o se precisa de una elevada velocidad de procesamiento, como en el caso de los sistemas que trabajan en tiempo real. Por lo tanto, se hace necesario encontrar nuevas técnicas y paradigmas que permitan mejorar los resultados de las soluciones actuales además de reducir el tiempo de procesamiento sin que ello suponga una pérdida de precisión.

Para la realización de esta tesis se han escogido dos problemas en los que se evidencian la necesidad de las mejoras mencionadas anteriormente. Por un lado, el primer problema consiste en la segmentación del árbol vascular en imágenes de fondo de ojo. Este problema tiene especial interés ya que permite crear una herramienta de apoyo al especialista para la monitorización del árbol vascular con el objetivo de detectar diferentes patologías. No obstante, al tratarse de un sistema destinado al ámbito médico, cualquier mejora de los resultados del modelo dota al especialista de una mayor confianza en esta herramienta. Por otro lado, el segundo problema consiste en la segmentación por instancias de fresas en imágenes. Este problema supone una pieza clave en la creación de cosechadoras automática de fresas y, por ende, es necesario utilizar modelos que puedan trabajar en tiempo real en equipos con gran limitación de procesamiento y memoria.

El trabajo llevado a cabo en esta tesis ha dado lugar a tres publicaciones en las que se han propuesto soluciones que abordan las principales limitaciones de los modelos de aprendizaje profundo en los dos problemas de interés: la mejora de los resultados y la velocidad de procesamiento. En concreto, para el caso de la segmentación del árbol vascular, se ha propuesto un nuevo modelo basado en U-Net (modelo de referencia para la segmentación semántica) y nuevas técnicas para el entrenamiento que, en su conjunto, logran mejorar los resultados del estado del arte. De este modo, el modelo propuesto, con una arquitectura mucho más eficiente que la del modelo original y sin necesidad de aplicar un procesamiento a la imagen antes ni después de ser procesada por el modelo, presenta unos valores de AUC superiores a los obtenidos por los modelos más representativos del estado del arte. En el caso de la segmentación por instancias de fresas, se ha propuesto una modificación de Mask R-CNN (uno de los modelos de referencia en segmentación por instancias) con el objetivo de mejorar la velocidad de procesamiento del modelo original de forma considerable sin que ello repercuta de forma notable en los resultados. Las modificaciones propuestas han permitido trabajar a 10 fps, lo que supone doblar la velocidad del modelo original sin que ello repercuta de forma notable en el valor de mAP. Además, se ha propuesto un nuevo paradigma para abordar el problema de la segmentación por instancias de fresas, así como un nuevo modelo, que por primera vez en este problema es capaz de trabajar en tiempo real (30 fps) con un incremento en el valor de mAP superior al 15 % respecto a Mask R-CNN.

Los resultados alcanzados en los trabajos realizados en esta tesis para los dos problemas de interés escogidos permiten considerar a las soluciones propuestas como las mejores alternativas para su implementación en entornos reales de trabajo.

Palabras clave: aprendizaje profundo, redes neuronales convolucionales, procesamiento de imágenes, segmentación de imágenes, segmentación por instancias.

Abstract

In this doctoral thesis a theoretical and experimental study has been carried out in the field of deep learning. The objective of this study has been to improve the current techniques used in problems from different fields of interest of society.

The main problems that deep learning models present when implemented in real environments are the lack of precision and the high computational cost of their execution. There are problems in which the accuracy of the model must be as high as possible for this type of solutions to be used. In addition, the processing demands of the models may prevent their use in certain problems where the processing capacity is very limited or a high processing speed is required, as in the case of systems working in real time. Therefore, it is necessary to find new techniques and paradigms to improve the results of the current solutions and to reduce the processing time without losing accuracy.

For the development of this thesis, two problems have been chosen in which the need for the mentioned above improvements is evident. On the one hand, the first problem consists of the segmentation of the vascular tree in fundus images. This problem is of special interest since it allows the creation of a tool to support the specialist in monitoring the vascular tree with the purpose of detecting different pathologies. However, as this is a system intended for use in the medical field, any improvement in the model results gives the specialist greater confidence in this tool. On the other hand, the second problem consists in the segmentation by instance of strawberries in images. This problem is a key part in the creation of automatic strawberry harvesters and, therefore, it is necessary to use models that can work in real time on equipment with processing and memory limitations.

The result of the work done in this thesis has led to three publications in which solutions have been proposed that address the main limitations of deep learning models in the two problems of interest: performance improvement and processing speed. Specifically, for the case of vascular tree segmentation, a new model based on U-Net (reference model for semantic segmentation) and new techniques for training have been proposed that, as a whole, manage to improve the state-of-the-art results. Thus, the proposed model, with a much more efficient architecture than the original model and without the need to apply processing to the image before or after being processed by the model, presents higher AUC values than those obtained by the most representative models of the state of the art. In the case of strawberry instance segmentation, a modification of Mask R-CNN (one of the reference models in instance segmentation) has been proposed with the goal of improving the processing speed of the original model considerably without notably affecting the results. The proposed modifications have made it possible to work at 10 fps, which means doubling the speed of the original model without notably affecting the mAP value. In addition, a new paradigm has been proposed to address the problem of strawberry instance segmentation, as well as a new model, which for the first time in this problem is able to work in real time (30 fps) with a 15% increase of mAP value compared to Mask R-CNN.

The results achieved in the work done in this thesis for the two chosen problems allow us to consider the proposed solutions as the best alternatives for implementation in commercial working environments.

Key words: *deep learning, convolutional neural networks, image processing, image segmentation, instance segmentation.*

Índice general

Índice general	VII
Índice de figuras	IX
Índice de tablas	IX

Índice general	1
1 Introducción	3
1.1 Motivación	3
1.2 Contribuciones científicas de la tesis	6
1.3 Estructura de la tesis	7
2 Objetivos y metodología	9
2.1 Objetivos	9
2.2 Metodología	11
3 Discusión de resultados	15
3.1 Primera publicación	15
3.2 Segunda publicación	16
3.3 Tercera publicación	17
4 Conclusiones	19

Apéndices	
A Copia completa de las publicaciones	25
A.1 Primera publicación	25
A.2 Segunda publicación	38
A.3 Tercera publicación	54
B Factor de impacto de las publicaciones	69
B.1 Primera publicación	69
B.2 Segunda publicación	75
B.3 Tercera publicación	81

Índice de figuras

Índice de tablas

Índice general

CAPÍTULO 1

Introducción

En este capítulo se hace una introducción al trabajo realizado en esta tesis doctoral. Para ello, este capítulo se ha dividido en tres secciones en las que se tratan de resumir los aspectos más relevantes de la introducción. En la primera, se realiza una motivación de la unidad temática de la tesis. En la segunda, se dan a conocer las contribuciones científicas obtenidas como resultado de la realización de esta tesis. Por último, en la tercera, se presentan el resto de los capítulos que componen la presente tesis.

1.1 Motivación

Esta tesis doctoral se enmarca en el campo del aprendizaje profundo [1] (*deep learning* en inglés). Este campo de conocimiento forma parte de la disciplina conocida como aprendizaje automático (*machine learning* en inglés), que se engloba dentro de la inteligencia artificial.

Desde sus comienzos, la inteligencia artificial ha tratado de emular las habilidades de los seres humanos mediante algoritmos diseñados a medida para cada problema. Sin embargo, existen multitud de problemas para los que no existe, de forma evidente, una secuencia de pasos que permitan su resolución. Este es el caso, por ejemplo, del reconocimiento de personas en imágenes. Mientras que un ser humano es capaz de resolver este problema fácilmente, para un ordenador, esta tarea resulta ser extremadamente complicada.

Debido a esta dificultad de diseñar algoritmos que dieran solución a ciertos problemas, comenzó a surgir interés por nuevos paradigmas, que proponen obtener el algoritmo que resuelve un problema de forma automática a partir de los datos disponibles para el problema en cuestión. Este es el objetivo del aprendizaje automático (y del aprendizaje profundo): ajustar (entrenar) modelos para resolver un problema concreto en base a un conjunto de datos. De esta forma, no es necesario definir los pasos que debe dar el modelo, al ser este el encargado de obtenerlos a partir del conjunto de datos durante su entrenamiento. Esto permite abordar problemas en los que no es evidente la secuencia de pasos necesarias para resolverlos.

La principal diferencia entre los modelos de aprendizaje automático y aprendizaje profundo radica en el uso del conjunto de datos. Mientras que los modelos de aprendizaje automático precisan de una etapa previa en la que un experto extrae las características más relevantes del conjunto de datos con las que entrenar el modelo, los modelos de aprendizaje profundo son capaces de trabajar con los datos en bruto del problema y extraer, por sí mismos, las características más relevantes para resolverlo. Esto hace que el

desarrollo de modelos de aprendizaje profundo sea mucho más sencillo y menos costoso que el de modelos de aprendizaje automático. Sin embargo, como contrapartida, la cantidad de datos necesaria para entrenar correctamente los modelos de aprendizaje profundo es mucho más elevada (y, por ende, el tiempo necesario para entrenar los modelos) que la necesaria para los modelos de aprendizaje automático, lo que dificultó en sus comienzos el auge del aprendizaje profundo.

Existen diversos motivos por los que los modelos de aprendizaje profundo se han impuesto actualmente a los modelos de aprendizaje automático. En primer lugar, el aumento de la conectividad y la popularización de las redes sociales, así como el abaratamiento de los dispositivos de almacenamiento, han propiciado que se disponga de grandes cantidades de datos con los que entrenar estos modelos. Este hecho es de suma importancia, ya que para los modelos de aprendizaje automático, el aumento del conjunto de datos no suele suponer una mejora de los resultados; sin embargo, los modelos de aprendizaje profundo son capaces de seguir mejorando los resultados a medida que el conjunto de datos crece. En segundo lugar, el aumento de la potencia de cómputo, más concretamente el auge del procesamiento paralelo en GPU, ha solucionado gran parte de las limitaciones de procesamiento que supone el entrenamiento de los modelos de aprendizaje profundo. Y, por último, el motivo más evidente por el que los modelos de aprendizaje profundo se han popularizado es la demostración práctica de que son capaces de superar los resultados de los modelos de aprendizaje automático en problemas complejos de diversa naturaleza.

El hito que supuso un punto de inflexión en el uso de los modelos de aprendizaje automático en favor de los modelos de aprendizaje profundo se produjo en el año 2012 cuando una red neuronal convolucional (un modelo de aprendizaje profundo), llamada AlexNet [2], logró ganar la competición ImageNet [3]. Esta competición, con más de 14 millones de imágenes para el entrenamiento de los modelos, consiste en la clasificación de imágenes en mil clases diferentes. Hasta la llegada de AlexNet, los modelos de aprendizaje automático no eran capaces de bajar del 25 % de error. Sin embargo, AlexNet consiguió un error del 15.3 %, 10.8 puntos menos que el mejor modelo de aprendizaje automático hasta la fecha. Los resultados alcanzados por AlexNet en un problema tan complejo, como es la competición ImageNet, captaron automáticamente el interés de toda la comunidad científica.

Actualmente, los modelos de aprendizaje profundo han logrado resultados inimaginables en multitud de campos de aplicación. Merecen una mención especial su aplicación al procesamiento del lenguaje natural [4], plegado de proteínas [5] o a la segmentación por instancias en imágenes [6]. Los resultados obtenidos por los modelos de aprendizaje profundo en los problemas anteriores han supuesto avances tan significativos que sitúan a esta tecnología como la más firme candidata a ser capaz de solucionarlos de forma definitiva. En problemas más específicos, como la conducción autónoma, empresas como Tesla apuestan por esta tecnología para su desarrollo [7]. Además, los modelos de aprendizaje profundo están presentes en muchos de los servicios que prestan las grandes compañías, como Google [8] o Facebook [9], y se pueden encontrar ejemplos de uso en campos tan diversos como la medicina [10], la biología [11], la economía [12] o la agricultura [13].

A pesar del fuerte impacto que ha tenido la implantación de los modelos de aprendizaje profundo en todos los ámbitos de la sociedad, estos modelos son susceptibles de mejora por diversos motivos. Por un lado, los modelos de aprendizaje profundo con mejores resultados son aquellos que tienden a requerir una gran capacidad de procesamiento para su uso [14]. Esto supone una limitación a la hora de implantar estos modelos en entornos reales con capacidades limitadas de procesamiento. Por lo tanto, existe un gran

interés en el desarrollo de nuevos modelos de aprendizaje profundo que reduzcan la demanda de procesamiento sin que perjudique de forma notable a los resultados. Por otro lado, en ciertos problemas, como la segmentación por instancias en imágenes, los modelos de referencia, como Mask R-CNN [6], cuentan con una arquitectura demasiado compleja y con diferentes componentes y procesos para los resultados que alcanza. En este sentido, el desarrollo de nuevos modelos de menor complejidad que aborden estos problemas manteniendo un nivel de precisión similar es también un campo abierto de gran interés.

De esta forma, el diseño de nuevos modelos de aprendizaje profundo que den solución a las limitaciones descritas tendría gran transcendencia práctica, debido a la gran variedad de campos susceptibles de su aplicación. Por ello, la demostración práctica de estas soluciones pasa por su aplicación en problemas relevantes, donde se demuestre su viabilidad en términos de coste computacional y precisión. En esta tesis doctoral, se proponen nuevos modelos de aprendizaje profundo que muestran su potencial en dos problemas específicos de diferente naturaleza. El primero se encuadra en el ámbito médico, con el consecuente beneficio que supone para la sociedad; el segundo se sitúa en el campo de la agricultura y presenta un fuerte impacto en la economía local de la provincia donde se ha realizado la tesis (Huelva). A continuación se realiza una breve presentación de los problemas reseñados.

En el campo de la medicina, se ha abordado un problema de especial relevancia, la segmentación del árbol vascular en imágenes de fondo de ojo. Esto permite el desarrollo de herramientas de apoyo al especialista en la monitorización del árbol vascular con el objetivo de detectar diferentes patologías. Un posible uso de este tipo de herramientas es la detección precoz de la retinopatía diabética. Este problema afecta actualmente a una gran parte de la población adulta. Actualmente, se estima que 415 millones de diabéticos tienen riesgo de sufrir una ceguera a causa de esta enfermedad. La detección precoz de esta enfermedad supone un elemento clave para su control. Sin embargo, muchos países en desarrollo, como India (que cuentan con 70 millones de diabéticos), tienen problemas a la hora de llevar a cabo revisiones oculares a causa de la escasez de especialistas. Por este motivo, actualmente se está trabajando en el desarrollo de sistemas que sirvan de ayuda al especialista mediante el uso de modelos de aprendizaje profundo para la segmentación del árbol vascular en imágenes de fondo de ojo con el objetivo de detectar la retinopatía diabética y de reducir el tiempo que emplea el especialista en la revisión de los pacientes [15]. Estos sistemas se beneficiarán en gran medida de nuevos modelos que reduzcan el tiempo de procesamiento y mejoren los resultados alcanzados por las soluciones actuales, ya que permitirán un mejor diagnóstico a los especialistas y aumentar el número de pacientes que son capaces de examinar.

Por otro lado, en el campo de la agricultura, existe un especial interés en el desarrollo de cosechadoras automáticas que permitan reducir los costes de recolección de las explotaciones agrícolas. Este es el caso de cultivos como la fresa, que se encuentra ampliamente expandido alrededor del mundo, y muy presente en la provincia de Huelva, donde juega un papel esencial en su economía. Sin embargo, la recolección de la fresa supone una parte significativa del coste total de la cosecha y uno de sus mayores problemas, dada la baja demanda de este trabajo debido a las duras condiciones laborales. Por este motivo, la fresa es un cultivo que se beneficiará de la creación de cosechadoras automáticas. Para ello, es necesario la localización y análisis de la calidad de cada fresa de manera individual. Actualmente, este problema se ha comenzado a abordar mediante modelos de aprendizaje profundo que realizan la segmentación por instancias de las fresas facilitando su localización específica y el análisis de su calidad [16]. No obstante, las soluciones propuestas se basan en modelos complejos (como Mask R-CNN) que necesitan de equi-

pos informáticos con gran capacidad de cómputo para poder ejecutar estos modelos. Por lo tanto, es necesario diseñar nuevos modelos que cuenten con arquitecturas más simples que las actuales a la vez que logren mejorar sus resultados facilitando así el desarrollo de cosechadoras automáticas de fresas.

Estos dos problemas, la segmentación del árbol vascular en imágenes de fondo de ojo y la segmentación por instancias de fresas, son un entorno real en el que poder estudiar y demostrar el potencial de cualquier solución que se proponga para mejorar los resultados de las metodologías existentes en la literatura. Además, las limitaciones anteriormente descritas de los modelos de aprendizaje profundo para su aplicación en estos problemas, así como el fuerte impacto de esta tecnología en la industria, justifican el interés y desarrollo de esta propuesta de tesis doctoral. En este contexto, el objetivo general de la misma es contribuir al avance del aprendizaje profundo, por un lado, proponiendo soluciones a sus principales limitaciones y, por otro, creando nuevos modelos que mejoren los resultados del estado del arte en problemas de interés.

1.2 Contribuciones científicas de la tesis

Esta tesis doctoral se presenta como un compendio de publicaciones. La investigación llevada a cabo durante esta tesis doctoral ha dado lugar a la publicación de tres artículos científicos en revistas con un alto índice de impacto en el ámbito del aprendizaje profundo que avalan el trabajo realizado. Una copia de las publicaciones y un informe con el factor de impacto de estas se encuentra en los anexos **A** y **B** respectivamente.

Las principales contribuciones de esta tesis doctoral tienen que ver con la creación de nuevos modelos de aprendizaje profundo para la mejora de los resultados y la reducción de la demanda de procesamiento. En concreto, estas contribuciones se han enfocado a los problemas de segmentación del árbol vascular en imágenes de fondo de ojo y la segmentación por instancias de fresas.

A continuación, se detallan las tres publicaciones obtenidas con un breve resumen de estas y sus principales contribuciones:

- **Primera publicación.**

- **Título:** *A fast and accurate deep learning method for strawberry instance segmentation.*
- **Autores:** Perez-Borrero, Isaac; Marin Santos, Diego; Gegundez-Arias, Manuel E.; Cortés-Ancos, Estefanía.
- **Revista:** *Computers and electronics in agriculture.*
- **ISSN:** 0168-1699.
- **DOI:** 10.1016/j.compag.2020.105736.

En este trabajo se aborda el problema de la segmentación por instancias de fresas para su integración en un sistema de recolección automática. Para ello, en primer lugar, se presenta un nuevo conjunto de datos (StrawDI_Db1) de acceso libre y una nueva métrica de evaluación para el entrenamiento y la comparación de diferentes soluciones bajo las mismas condiciones. En segundo lugar, se toma uno de los modelos del estado del arte para segmentación por instancias, Mask R-CNN, como referencia para proponer un nuevo modelo, basado en el anterior, que reduzca la complejidad para conseguir un buen compromiso entre el tiempo de procesamiento y los resultados obtenidos.

■ Segunda publicación.

- **Título:** *A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and U-Net model.*
- **Autores:** Gegundez-Arias, Manuel E.; Marin Santos, Diego; Perez-Borrero, Isaac; Vasallo Vázquez, Manuel J.
- **Revista:** *Computer Methods and Programs in Biomedicine.*
- **ISSN:** 0169-2607.
- **DOI:** 10.1016/j.cmpb.2021.106081.

En este trabajo se presenta un modelo para la segmentación del árbol vascular en imágenes de fondo de ojo. El modelo es evaluado en las principales bases de datos de imágenes de fondo de ojo.

El modelo propuesto se basa en la modificación de la red U-Net para lograr un buen compromiso en cuanto al tiempo de procesamiento, el número de parámetros y los resultados del modelo. Además, se propone el uso de la distancia al árbol vascular para penalizar los píxeles que cometen errores durante el entrenamiento en zonas alejadas de este. También se propone dividir las imágenes de entrenamiento en partes para mejorar los resultados del entrenamiento al poder utilizar en un mismo lote más imágenes diferentes que si se utiliza la imagen original.

■ Tercera publicación.

- **Título:** *A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network.*
- **Autores:** Perez-Borrero, Isaac; Marin-Santos, Diego; Vasallo-Vazquez, Manuel J.; Gegundez-Arias, Manuel E.
- **Revista:** *Neural Computing and Applications.*
- **ISSN:** 1433-3058.
- **DOI:** 10.1007/s00521-021-06131-2.

En este trabajo se vuelve a abordar el problema de la segmentación por instancias de fresas para su recolección. Esta vez, se propone un nuevo paradigma para realizar la segmentación por instancias de fresas, que consiste en la predicción del centroide de cada fresa mediante una red neuronal convolucional que no precisa de una red neuronal.

Mediante este nuevo enfoque, se modifica la red U-Net para la predicción del centroide de cada fresa y se modifica su arquitectura para que se ajuste a las limitaciones computacionales del problema. De la misma forma, se elabora un algoritmo eficiente para obtener las segmentaciones de las fresas a partir de la salida de la red. Estas modificaciones han permitido conseguir mejores resultados que los modelos del estado del arte a la vez que reducir de forma considerable el tiempo de procesamiento y la complejidad del modelo.

1.3 Estructura de la tesis

Además de este capítulo inicial de introducción, este documento de tesis doctoral se compone de tres capítulos adicionales y dos anexos con el siguiente contenido:

- **Introducción.** En este capítulo se ha realizado una introducción a la tesis doctoral. Se ha descrito la motivación detrás de esta, sus principales contribuciones científicas y la estructura de la tesis.
- **Objetivos y metodología.** En este capítulo se detallan los objetivos que se persiguen cumplir con la realización de esta tesis y la metodología a seguir para cada objetivo planteado.
- **Discusión de resultados.** En este capítulo se hace una discusión de los resultados obtenidos a través de las publicaciones realizadas. De esta forma, se pretende poner en valor y dar a conocer con más detalle el alcance de los mismos y el potencial de las soluciones propuestas.
- **Conclusiones.** En el último capítulo de esta tesis se presentan las conclusiones principales de la tesis doctoral.
- **Anexo A.** Copia completa de las publicaciones de la tesis doctoral.
- **Anexo B.** Informe con el factor de impacto de las publicaciones de la tesis doctoral.

CAPÍTULO 2

Objetivos y metodología

El presente capítulo se divide en dos secciones en las que se presentan, por un lado, los objetivos propuestos en esta tesis y, por otro, la metodología propuesta para cumplir los objetivos planteados.

2.1 Objetivos

A pesar de tratarse de tecnología incipiente, existen multitud de problemas en los que los modelos de aprendizaje profundo se han situado como la mejor alternativa a la hora de abordarlos gracias a sus resultados. No obstante, la implantación de soluciones basadas en modelos de aprendizaje profundo en entornos reales sigue planteando varios desafíos. En primer lugar, muchos de los entornos donde se integra esta tecnología requieren soluciones que trabajen en tiempo real. Es decir, los modelos cuentan con limitaciones de procesamiento y memoria, por lo que se deben buscar soluciones aptas para su uso en este tipo de entornos. En segundo lugar, existen problemas, como la segmentación por instancias, en los que los modelos de aprendizaje profundo no consiguen un buen compromiso en términos de resultados y complejidad del modelo. Por estos motivos, los modelos de aprendizaje profundo que son utilizados hoy día en ciertos problemas de interés cuentan aún con un amplio margen de mejora.

En esta tesis doctoral se plantea la realización de una investigación teórica y experimental para la mejora de los modelos actuales de aprendizaje profundo con un doble objetivo. Por un lado, dotar a los modelos de la precisión requerida en ciertos ámbitos en términos de velocidad y resultados. Por otro, proponer nuevos modelos de aprendizaje profundo capaces de mejorar los resultados actuales en aquellos problemas en los que exista cierto margen de mejora de los resultados alcanzados por las soluciones del estado del arte.

Tal y como se indicó en el capítulo de introducción, la investigación a realizar en esta tesis se enmarca en el campo de la medicina y la agricultura. En concreto, en el campo de la medicina, en el problema de la segmentación del árbol vascular en imágenes de fondo de ojo; y, en el campo de la agricultura, en la segmentación por instancias de fresas en imágenes.

A partir del estudio de las limitaciones de los modelos actuales de aprendizaje profundo y sus posibles mejoras centradas en los problemas elegidos, se han planteado los siguientes objetivos generales para esta tesis doctoral:

- **Objetivo general 1.**

Proponer nuevos modelos de redes neuronales convolucionales que puedan integrarse en sistemas en tiempo real, con limitaciones de procesamiento y memoria, para problemas de segmentación y segmentación por instancias en imágenes.

Una de las principales limitaciones de las redes neuronales convolucionales es el tiempo que necesitan para procesar la ingente cantidad de información con la que trabajan, haciendo inviable su uso en aplicaciones que cuentan con recursos de procesamiento limitados. En esta tesis se propone desarrollar modelos simplificados de redes neuronales convolucionales que reduzcan de manera considerable el tiempo de procesamiento de las soluciones actuales a la vez que mantiene un nivel equiparable en cuanto a los resultados.

■ **Objetivo general 2.**

Desarrollar nuevas metodologías basadas en redes neuronales convolucionales para abordar problemas de segmentación por instancias.

La principal limitación de las metodologías actuales de segmentación por instancias radica en su alta complejidad, que dificulta la obtención de buenos resultados y su implantación en determinados entornos con recursos de procesamiento limitados. El objetivo que se persigue es mejorar los resultados de las metodologías actuales reduciendo su complejidad.

■ **Objetivo general 3.**

Aplicar las soluciones propuestas para los objetivos generales 1 y 2 anteriores en problemas reales de medicina y agricultura.

En el campo de la medicina, se aplicará al problema de la segmentación del árbol vascular en imágenes de fondo de ojo. En el campo de la agricultura, se aplicará a la segmentación por instancias de fresas en imágenes.

A continuación, se presentan los objetivos específicos de esta tesis, enmarcados en cada uno de los objetivos generales anteriores.

■ **Relativos al objetivo general 1:**

● **Objetivo específico 1.**

Proponer modificaciones a la arquitectura de la red neuronal convolucional de referencia en problemas de segmentación semántica (U-Net) y segmentación por instancias (Mask R-CNN) para hacer posible su uso en sistemas con limitaciones de tiempo y capacidad de procesamiento.

● **Objetivo específico 2.**

Proponer nuevos algoritmos de agrupamiento para generar de forma más eficiente las segmentaciones finales de las redes de detección y segmentación por instancias.

■ **Relativos al objetivo general 2:**

● **Objetivo específico 3.**

Proponer metodologías alternativas de segmentación por instancias que reduzcan la complejidad del problema de optimización.

● **Objetivo específico 4.**

Proponer una nueva arquitectura de red neuronal convolucional que realice la segmentación por instancias de forma simple y eficiente mediante una red

neuronal convolucional basada en el uso exclusivo de capas de convolución y *pooling*.

- **Relativos al objetivo general 3:**

- **Objetivo específico 5.**

Aplicar las metodologías y modelos desarrollados en los objetivos específicos anteriores para mejorar los resultados del estado del arte en problemas de diversa naturaleza:

- Medicina: segmentación del árbol vascular en imágenes de fondo de ojo como herramienta de apoyo al especialista para la detección de patologías.
- Agricultura: segmentación por instancias de fresas para su recolección automática o análisis de la calidad.

La amplia variedad de casos de uso del aprendizaje profundo y su adopción en multitud de soluciones utilizadas en industrias de diferentes ámbitos garantizan que la consecución de los objetivos de esta tesis doctoral tendrá un impacto apreciable en la sociedad. En este sentido, los objetivos de esta tesis doctoral están alineados con la Estrategia Española de I+D+I en Inteligencia Artificial y la Estrategia Europea para la IA que buscan situar a España y Europa a la vanguardia de esta tecnología.

2.2 Metodología

En esta sección se indica la metodología a seguir para cumplir cada uno de los objetivos propuesto en esta tesis doctoral. A continuación, se detalla la metodología concreta para cada objetivo específico:

- **Metodología para el objetivo específico 1.**

Para reducir el coste computacional asociado a la gran mayoría de modelos de referencia para diferentes problemas (Mask R-CNN en segmentación por instancias o U-Net en segmentación semántica p. ej.) y hacer posible su uso en sistemas de tiempo real, es necesario un estudio previo de la arquitectura para detectar los cuellos de botella de esta y dónde se produce el mayor consumo de recursos. Una vez se ha detectado las partes del modelo que consumen más recursos, se propondrán nuevas arquitecturas y algoritmos teniendo en cuenta el estado del arte actual y el problema concreto en el que se aplicará el nuevo modelo.

Actualmente, la mayoría de los modelos de referencia del estado del arte fueron desarrollados para competiciones complejas como es ImageNet. Al utilizarse en problemas específicos, con un entorno más controlado en términos de variabilidad de los datos, mediante un proceso de ensayo y error, puede encontrarse modelos con arquitecturas mucho más reducida que consigan igualar o superar a los modelos del estado del arte en problemas más específicos, pero con una demanda de procesamiento mucho menor.

Para atender a los requisitos de precisión y velocidad, será necesario comparar el modelo propuesto con el modelo de referencia en el mismo campo de aplicación, bajo las mismas métricas y sobre el mismo conjunto de datos.

- **Metodología para el objetivo específico 2.**

Una de las etapas que más tiempo consume en modelos de segmentación por instancias como Mask R-CNN es la agrupación de las múltiples detecciones que ocu-

rren sobre un mismo objeto. Es por ello que cualquier propuesta de mejora del algoritmo de agrupamiento de las detecciones repercutirá notablemente en la reducción del tiempo de procesamiento total del modelo. Para ello, será necesario estudiar el algoritmo actual y detectar los bucles de procesamiento que lo ralentizan. A continuación, se propondrán posibles alternativas a los bucles originales que puedan ser paralelizables fácilmente utilizando las librerías actuales de procesamiento paralelo en GPU.

El algoritmo propuesto deberá producir resultados similares a los del algoritmo original, pero reduciendo de manera considerable el tiempo de procesamiento. Por lo tanto, se comparará el tiempo de procesamiento medio de Mask R-CNN cuando utiliza uno u otro para poder determinar la mejora alcanzada.

■ Metodología para el objetivo específico 3.

En los modelos actuales de referencia para segmentación por instancias, como Mask R-CNN, se ha de optimizar diferentes valores de pérdida que atienden a diversos objetivos. Por un lado, se optimiza la detección mediante una regresión de las coordenadas de los recuadros. Por otro lado, se optimiza la clasificación mediante la regresión logística. Por último, la segmentación se optimiza mediante una regresión logística a nivel de píxel. Esta forma de abordar el problema (mediante la optimización de diferentes subproblemas) supone un desafío añadido, ya que se debe calibrar correctamente la definición de la función de pérdida de forma que no se perjudique a alguno de los objetivos más que a otro y esto se traduzca en malos resultados en algunas de las tareas que realiza la red. Por lo tanto, el estudio del estado del arte y la propuesta de nuevos paradigmas para abordar el problema puede suponer una reducción de la complejidad del problema de optimización a la vez que permita utilizar modelos con arquitecturas mucho más simples que las actuales.

Una vez se proponga una nueva definición del problema de optimización para la segmentación por instancias y se entrene el modelo, se ha de comparar los resultados con el de los modelos del estado del arte para demostrar que la solución propuesta consigue igualar o superar los resultados de soluciones alternativas a la vez que mejoran de forma considerable el tiempo de procesamiento. Esto precisa que se haga sobre el mismo conjunto de datos y bajo las mismas condiciones de evaluación.

■ Metodología para el objetivo específico 4.

Para realizar segmentación por instancias con modelos de aprendizaje profundo se suele combinar redes neuronales convolucionales junto con otros algoritmos para conseguir las segmentaciones individuales de cada elemento de interés. Sin embargo, el empleo de diferentes elementos en el sistema hace que sea más difícil de optimizar y precise de más capacidad de procesamiento. Por este motivo, diseñar un modelo que utilice únicamente capas de convolución y *pooling* reduciría de manera considerable el tiempo de procesamiento del sistema y facilitaría el problema de optimización. En este sentido, es necesario trabajar en el desarrollo de una red neuronal convolucional que no necesite de redes neuronales en su arquitectura para lograr la segmentación por instancias.

La red neuronal convolucional a desarrollar se entrenará y validará sobre el mismo conjunto de datos que las redes del estado del arte del problema escogido. Mediante una comparativa en términos de precisión y tiempo de procesamiento se podrá demostrar si el modelo propuesto supone una mejora a las soluciones existentes.

- **Metodología para el objetivo específico 5.**

Todas las metodologías y modelos desarrollados para los diferentes objetivos específicos serán entrenadas y probadas en conjuntos de datos públicos, que sean representativos del problema en cuestión.

En el campo de la medicina se utilizarán los conjuntos de datos públicos más relevantes de la literatura con los que se entrenará y evaluará la solución propuesta para el problema de la segmentación del árbol vascular en imágenes de fondo de ojo. En concreto, se utilizarán los tres conjuntos de datos más representativos del problema: DRIVE [17], STARE [18] y CHASE_Db1 [19].

En el campo de la agricultura se abordará el problema de la segmentación por instancias de fresas. En este caso, al no existir un conjunto de datos público y representativo de este problema, se preparará un conjunto de datos representativo de imágenes de fresas en el que se marcarán todas las fresas presentes en cada imagen del conjunto de manera individual. Este conjunto de datos será accesible de forma gratuita y será utilizado para entrenar y evaluar los diferentes modelos propuestos para este problema.

CAPÍTULO 3

Discusión de resultados

La realización de esta tesis doctoral ha dado como resultado tres publicaciones. En este capítulo se discuten individualmente los resultados principales de las mismas.

3.1 Primera publicación

En la primera publicación (ver anexo A.1), titulada *A fast and accurate deep learning method for strawberry instance segmentation*, se ha abordado el problema de segmentación por instancias de fresas mediante modelos de aprendizaje profundo. Este problema ha sido abordado previamente en [16]. En este trabajo, se utiliza la red Mask R-CNN con una de las configuraciones propuesta por los autores originales de esta red: la red ResNet-50 como *backbone*. La red fue entrenada en un conjunto de datos privado y, como sus autores señalan, a pesar de disminuir el número de anclas utilizadas y la resolución de las imágenes, la exigencia computacional del modelo es tan grande que no es posible conseguir trabajar en tiempo real.

Para afrontar los problemas descritos en [16] de Mask R-CNN en la segmentación por instancias de fresas, en esta primera publicación se propone un nuevo modelo basado en esta red. Además, para facilitar la comparación entre diferentes soluciones, se ha creado un nuevo marco comparativo para el entrenamiento y evaluación de los modelos. Concretamente, se ha creado una base de datos de imágenes de fresas que incluye la segmentación y etiquetado manual de las mismas; por otra parte, además de las principales métricas de evaluación ampliamente utilizadas en este tipo de problemas (*frames por seconds* o mAP), el artículo incluye una nueva métrica de evaluación, *mean instance intersection over union* (mI²oU). Junto con el modelo propuesto, tanto el conjunto de datos como la métrica mI²oU forman parte de las contribuciones de este trabajo y permitirán replicar y comparar modelos de segmentación por instancias, independientemente del enfoque del problema, bajo las mismas condiciones.

Las modificaciones introducidas a Mask R-CNN para reducir el tiempo de procesamiento original han consistido en la construcción de un nuevo *backbone* para la extracción de características, así como reemplazar el algoritmo *non-maximum suppression* y las redes de clasificación y detección por un nuevo algoritmo que genera las clasificaciones finales del modelo. Estas modificaciones han permitido reducir de manera considerable el tiempo de procesamiento al utilizar un *backbone* con un número de parámetros de varios órdenes de magnitud inferior al del *backbone* utilizado en [16] (ResNet-50). Además, el nuevo algoritmo propuesto para el filtrado y agrupamiento de las detecciones ha contribuido

también a reducir el tiempo total de procesamiento y la complejidad del modelo de manera considerable.

Los resultados conseguidos por el modelo propuesto demuestran que es posible utilizar un modelo más simple y rápido que Mask R-CNN para la segmentación por instancias de fresas introduciendo modificaciones en sus componentes más complejos, sin sacrificar precisión en las segmentaciones. En concreto, los resultados conseguidos por el modelo propuesto son similares a los conseguidos por Mask R-CNN para las métricas mAP (43.85 vs. 45.36) y mI²oU (87.27 vs. 87.70), pero aumentando la velocidad de procesamiento pasando de 5 fps con Mask R-CNN a 10 fps con el modelo propuesto. Esto significa que se ha doblado el número de imágenes que la red es capaz de procesar en el mismo tiempo sin que ello repercuta en los resultados de forma notable.

3.2 Segunda publicación

En la segunda publicación (ver anexo A.2), titulada *A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and U-Net model*, se aborda el problema de la segmentación semántica del árbol vascular en imágenes de fondo de ojo.

Para validar y entrenar el modelo propuesto, se han utilizado las principales bases de datos públicas de este problema disponibles: DRIVE, STARE y CHASE_Db1. Al disponerse de diferentes bases de datos, la evaluación del modelo se ha realizado siguiendo los dos escenarios más habituales encontrados en el estado del arte:

- Entrenar y evaluar sobre cada base de datos por separado. Tanto el conjunto de entrenamiento como el de prueba pertenecen a la misma base de datos.
- Hacer el entrenamiento y la evaluación mediante bases de datos cruzadas. El conjunto de entrenamiento proviene de una base de datos diferente a la del conjunto de pruebas.

La evaluación del modelo se ha llevado a cabo utilizando las métricas de evaluación más utilizadas en el estado del arte para este problema: *sensitivity* (Se), *specificity* (Sp), *accuracy* (Acc) y *area under the receiver operating characteristic* (ROC) *curve* (AUC). Dado que no todos los trabajos del estado del arte reportan todas estas métricas, la discusión de resultados se ha centrado en las métricas más discutidas en el estado del arte: Acc y AUC.

La solución propuesta en este trabajo ha consistido en la modificación de la red U-Net para reducir su complejidad y evitar el sobreajuste del modelo dada la escasez de imágenes para el entrenamiento (solo existen 98 imágenes segmentadas en las tres bases de datos). En concreto, se han introducido capas de *batch normalisation* y conexiones residuales que han permitido reducir de forma considerable el número de convoluciones del modelo y el nivel mínimo de resolución de los datos. Además, se ha propuesto una nueva función de pérdida que tiene en cuenta la distancia al píxel más cercano de la clase de interés para penalizar el error cometido en la segmentación.

Al comparar la solución propuesta con el modelo original y sin utilizar *batch normalisation* en el escenario de entrenamiento y evaluación mediante bases de datos cruzadas, se ha comprobado que consigue mejores resultados en términos de Acc y AUC (ver Anexo A.2, tabla 9). También, se ha observado el efecto significativo que tiene el uso de la nueva función de pérdida en la mejora de los resultados (ver Anexo A.2, tabla 10). Por otro lado, en términos de complejidad, el modelo propuesto cuenta con 87 mil parámetros y realiza

4.93 GFLOPS mientras que el modelo original cuenta con 31 millones de parámetros y realiza 96.40 GFLOPS.

El modelo propuesto es también el que mejores resultados ofrece en las tres bases de datos utilizadas al ser comparado con otros modelos de la literatura basados en la red U-Net (ver Anexo A.2, tabla 11). En la comparación con el resto de los modelos del estado del arte se concluye que únicamente en [20] y [21] se alcanzan unos resultados similares al del modelo propuesto (ver Anexo A.3, tabla 6, 7 y 8). Sin embargo, cabe destacar que, en [20], el modelo cuenta con una red neuronal, lo que supone una mayor complejidad en la arquitectura del modelo comparada con la del modelo propuesto, así como un número mayor de parámetros que ajustar. También es importante notar que este modelo realiza un proceso previo de pre-procesado de la imagen, mientras que en el modelo propuesto no es necesario y se utiliza la imagen original directamente. Por otro lado, la metodología propuesta en [21] utiliza una arquitectura con un mayor número de capas, que se traduce en una mayor demanda de cómputo respecto al modelo propuesto. Además, este modelo también requiere de una etapa previa de pre-procesamiento de la imagen original antes de ser procesada por la red. En ambos trabajos, los modelos propuestos requieren de una etapa final de post-procesamiento para generar los resultados. En el caso del modelo propuesto, la segmentación generada por el modelo se obtiene directamente a partir de su salida.

Otro elemento importante a tener en cuenta durante la comparación de los modelos es la elección del punto de operación para obtener las métricas. El modelo propuesto en [21] obtiene el punto de operación como aquel que consigue un mayor valor de Acc en la base de datos que se está evaluando, lo que requiere de la existencia de una segmentación para poder calcular correctamente este punto. Este hecho hace imposible la elección de este punto en un entorno práctico de aplicación real. Por lo tanto, es necesario que el modelo trabaje correctamente en todo el rango de posibles puntos de operación. Al comparar estos modelos con el modelo propuesto en términos de AUC, el método propuesto es el que alcanza unos valores más altos en todas las bases de datos, lo que facilita la elección del punto de operación para el modelo. De hecho, el modelo propuesto obtiene el máximo *accuracy* en un rango reducido para el valor de umbral de 0.56-0.60 para tres bases de datos completamente independientes y con características diferentes. Esta es una de las principales ventajas de la metodología propuesta respecto al resto de los modelos del estado del arte: con una arquitectura más simple, que no precisa de una etapa de pre y post procesamiento, los valores AUC son superiores en la experimentación a los reportados por los modelos del estado del arte.

3.3 Tercera publicación

En la tercera publicación (ver anexo A.3), titulada *A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network*, se aborda el problema de segmentación por instancias de fresas mediante aprendizaje profundo. Este problema ha sido abordado previamente en la primera publicación de esta tesis donde se realizaron modificaciones al modelo de referencia de este problema (Mask R-CNN) para reducir el tiempo de procesamiento sin que repercuta en los resultados. En el presente trabajo, el modelo propuesto se ha entrenado y evaluado utilizando el mismo conjunto de datos y utilizando las mismas métricas que se utilizaron en la primera publicación de esta tesis.

En este trabajo, se ha propuesto una nueva metodología para abordar el problema de segmentación por instancias de fresas mediante modelos más sencillos, que no precisen

de diferentes etapas de procesamiento en la que se involucren diferentes procesos como en el caso de Mask R-CNN. En concreto, esta nueva metodología ha permitido el uso de modelos de segmentación semántica, como redes *fully convolutional neural network* (redes neuronales convolucionales que no cuentan con capas de redes neuronales), para realizar la segmentación por instancias de fresas.

Para obtener la segmentación por instancias de fresas con modelos de segmentación semántica se aborda el problema mediante la predicción del centroide de cada fresa en la imagen para asignar a cada píxel clasificado como fresa la etiqueta asociada al centroide más cercano. De esta forma, el problema de la segmentación por instancias se traduce en un problema de clasificación y regresión a nivel de píxel. El problema de clasificación es tratado como un problema de segmentación semántica mientras que el problema de regresión consiste en predecir el desplazamiento a aplicar a las coordenadas del píxel para obtener las coordenadas de su centroide. Mediante esta nueva definición, es posible utilizar redes de segmentación semántica, como U-Net, para la segmentación por instancias. Una vez se ha obtenido la salida del modelo, es necesario un algoritmo de agrupamiento y filtrado para obtener las segmentaciones finales. En este trabajo, se ha presentado también un algoritmo eficiente para este fin, que añade 10 milisegundos al tiempo total de procesamiento y representa únicamente un 30 % del tiempo total necesario para procesar una imagen.

Dada la naturaleza eminentemente práctica de este problema, el modelo propuesto se ha diseñado de forma que cumpla los requisitos esperados para modelos que serán utilizados en entornos reales de producción. Para ello, en este trabajo, se ha modificado la arquitectura de la red U-Net para reducir los requisitos de procesamiento reduciendo el número de convoluciones de manera considerable gracias al uso de capas de *batch normalisation* y a reemplazar las capas de *transposed convolution* por interpolaciones bilineales, que reducen el número de parámetros a ajustar.

Como resultado de las modificaciones introducidas, el modelo propuesto cuenta con un total de 300 mil parámetros, lo que supone una reducción del número de parámetros del 96 % respecto al modelo original. Además, los resultados alcanzados por el modelo propuesto en el conjunto de pruebas son superiores a los conseguidos por los modelos del estado del arte. En términos de mAP y mI²oU, el modelo propuesto consigue unos valores de 52.61 y 93.38 respectivamente. Esto supone una mejora considerable respecto al modelo Mask R-CNN original (45.36 y 87.70 para mAP y mI²oU respectivamente), que es el modelo del estado del arte con los valores más altos de estas métricas. Por otro lado, en términos de velocidad, el modelo propuesto en esta publicación supera de forma notable la velocidad de trabajo de los modelos del estado del arte al conseguir 30 fps, lo que supone el triple de velocidad que el modelo más rápido del estado del arte (el modelo propuesto en la primera publicación de esta tesis consigue trabajar a 10 fps mientras que Mask R-CNN trabaja a 5).

CAPÍTULO 4

Conclusiones

En esta tesis doctoral se ha desarrollado una investigación teórica y experimental sobre aprendizaje profundo con el principal objetivo de mejorar los resultados y la velocidad de procesamiento del estado del arte en problemas de detección, segmentación y segmentación por instancias en imágenes. En concreto, este trabajo se ha enfocado en los problemas de segmentación del árbol vascular en imágenes de fondo de ojo y la segmentación por instancias de fresas.

La investigación realizada ha dado lugar a tres publicaciones con las que se han cumplido los objetivos propuestos. A partir de estas publicaciones se derivan las conclusiones principales de este trabajo que se detallan a continuación.

- Conclusiones relativas al objetivo específico 1 (*propuesta de modificaciones a la arquitectura de la red neuronal convolucional de referencia en problemas de segmentación semántica y segmentación por instancias*).

Por un lado, para el problema de segmentación del árbol vascular en imágenes de fondo de ojo, en la segunda publicación de este trabajo (Anexo A.2), se propone una nueva arquitectura para la red U-Net (red en la que se basan la mayoría de los modelos del estado del arte de este problema). En concreto, se introducen conexiones residuales y *batch normalization*, lo que ha permitido reducir el número de capas y convoluciones. Estas modificaciones han reducido de forma considerable el tiempo de procesamiento del modelo. Además, se han propuesto nuevas técnicas para el entrenamiento del modelo que consisten en usar partes de la imagen en los lotes de entrenamiento en lugar de toda la imagen y penalizar el error de segmentación de cada píxel en base a la distancia de este al píxel más cercano de la clase de interés. Todas estas modificaciones han permitido reducir el tiempo de procesamiento de la red original además de mejorar sus resultados.

Por otro lado, en el problema de segmentación por instancias de fresas en imágenes, en la primera publicación (Anexo A.1), se ha logrado mejorar la velocidad de procesamiento de Mask R-CNN (modelo de referencia en el estado del arte), sin que repercuta de manera considerable en los resultados. Para ello, se ha propuesto un nuevo *backbone*, con una arquitectura mucho más reducida que el original, y un algoritmo de filtrado y agrupamiento que ha permitido eliminar las redes neuronales y el algoritmo *non-maximum suppression*. Este nuevo algoritmo ha supuesto cumplir el segundo de los objetivos específicos de este trabajo, donde se planteaba la propuesta de nuevos algoritmos de agrupamiento para Mask R-CNN, y ha sido esencial para reducir el tiempo de procesamiento del modelo original. Gracias a las modificaciones introducidas en Mask R-CNN, el modelo propuesto dobla la velo-

cidad de procesamiento de Mask R-CNN mientras que mantiene unos resultados competitivos en términos de mAP.

- Conclusiones relativas a los objetivos específicos 2 (*propuesta de nuevos algoritmos de agrupamiento para generar de forma más eficiente las segmentaciones finales*), 3 (*propuesta de metodologías alternativas de segmentación por instancias que reduzcan la complejidad del problema de optimización*) y 4 (*propuesta de una nueva arquitectura de red neuronal convolucional que realice la segmentación por instancias de forma simple y eficiente*).

En la tercera publicación de este trabajo (Anexo A.3), se ha propuesto una nueva metodología para la segmentación por instancias basada en el uso de redes neuronales convolucionales. Además, se ha implementado un nuevo algoritmo eficiente para generar las segmentaciones finales que supone únicamente el 30 % del tiempo total del modelo. Esta nueva metodología de segmentación por instancias ha permitido reducir la complejidad del problema de optimización de redes como Mask R-CNN (en el que se emplean multitud de términos con objetivos y funciones de pérdidas diferentes) a un problema de regresión de 4 valores. Esto ha permitido también reducir el número de hiperparámetros que se necesitan para el entrenamiento del modelo. Con la nueva definición del problema, es posible utilizar redes tradicionales de segmentación semántica para realizar segmentación por instancias. En este trabajo, se ha utilizado un modelo basado en la red U-Net, que cuenta únicamente con capas de convolución y *pooling*, lo que reduce la complejidad de modelos como Mask R-CNN, que cuentan con diferentes componentes y algoritmos en su arquitectura. Todas estas modificaciones, han permitido a la red propuesta alcanzar mejores resultados que el resto del estado del arte y triplicar la velocidad del modelo más rápido.

- Conclusiones relativas al objetivo específico 5 (*aplicar las metodologías y modelos desarrollados en los objetivos específicos anteriores para mejorar los resultados del estado del arte en problemas de diversa naturaleza*).

Todas las propuestas presentadas en las tres publicaciones de este trabajo se han aplicado a los problemas de interés de esta tesis: segmentación del árbol vascular en imágenes de fondo de ojo y la segmentación por instancias de fresas en imágenes. En la segmentación del árbol vascular en imágenes de fondo de ojo, el modelo propuesto consigue mejores resultados que el resto de los modelos del estado del arte con los mejores valores de *accuracy* para todos los posibles puntos de operación. Además, logra reducir el tiempo de procesamiento del modelo original y simplifica el flujo de trabajo respecto a otros modelos del estado del arte al procesar directamente la imagen en bruto, sin etapa de pre/post procesamiento. En la segmentación por instancias de fresas, se ha conseguido un modelo que es capaz de triplicar la velocidad de procesamiento y mejorar significativamente los resultados del mejor modelo del estado del arte.

Las principales aportaciones de este trabajo pueden resumirse en la propuesta de tres nuevos modelos y metodologías de entrenamiento que permiten mejorar los resultados y velocidad de procesamiento del resto de modelos del estado del arte para los problemas que se han abordado en este trabajo. Por este motivo, los modelos propuestos pueden considerarse como los principales candidatos para su implantación en cosechadoras automáticas de fresas o en sistemas de apoyo al especialista que integren módulos de segmentación de árbol vascular.

En resumen, en este trabajo se ha demostrado que las propuestas de modificación de los modelos del estado del arte suponen una mejora cuantitativa de los resultados y de la velocidad de procesamiento. Igualmente, se ha conseguido aplicar con éxito un nuevo

paradigma para abordar la segmentación por instancias mediante aprendizaje profundo, que, tal y como se ha observado en la experimentación realizada, consigue unos resultados y una velocidad de procesamiento inalcanzable mediante el enfoque tradicional de los modelos del estado del arte. Además, este nuevo enfoque es aplicable a cualquier problema de segmentación por instancias, no se limita a la segmentación por instancias de fresas. Esto significa que puede utilizarse para aquellos problemas en los que se trabaja en entornos con limitaciones de procesamiento (como las cosechadoras automáticas) y mejorar de forma considerable los resultados conseguidos por los modelos que se emplean actualmente.

Finalmente, como futuras líneas de investigación que se abren tras la realización de esta tesis cabría destacar las siguientes:

- Aplicar técnicas de aumento artificial del conjunto de datos. Actualmente, los conjuntos de imágenes de fondo de ojo son escasos y cuentan con un número muy limitado de imágenes para entrenar los modelos. También servirá para evitar el trabajo que supone etiquetar manualmente las imágenes, tanto en el caso de la segmentación del árbol vascular como la segmentación individual de las fresas en una imagen.
- Modificar la función de pérdida de los problemas para seguir mejorando los resultados. Tanto en el caso de la segmentación del árbol vascular como en la segmentación por instancias de fresas, la modificación de la función de pérdida ha supuesto una mejora considerable de los resultados.
- Probar nuevas arquitecturas para los modelos. La obtención de una arquitectura concreta para un modelo y un problema se consigue mediante un proceso de ensayo y error. Por lo tanto, seguir realizando este proceso y probando nuevas combinaciones de hiperparámetros podría dar lugar a una mejora de los resultados.

Bibliografía

- [1] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. "Deep learning". En: *nature* 521.7553 (2015), págs. 436-444.
- [2] Alex Krizhevsky, Ilya Sutskever y Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". En: *Advances in neural information processing systems* 25 (2012), págs. 1097-1105.
- [3] Jia Deng y col. "Imagenet: A large-scale hierarchical image database". En: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, págs. 248-255.
- [4] Li Deng y Yang Liu. *Deep learning in natural language processing*. Springer, 2018.
- [5] John Jumper y col. "Highly accurate protein structure prediction with AlphaFold". En: *Nature* (2021), pág. 1.
- [6] Kaiming He y col. "Mask r-cnn". En: *Proceedings of the IEEE international conference on computer vision*. 2017, págs. 2961-2969.
- [7] Shantanu Ingle y Madhuri Phute. "Tesla autopilot: semi autonomous driving, an uptick for future autonomy". En: *International Research Journal of Engineering and Technology* 3.9 (2016), págs. 369-372.
- [8] Jacob Devlin y col. "Bert: Pre-training of deep bidirectional transformers for language understanding". En: *arXiv preprint arXiv:1810.04805* (2018).
- [9] Jongsoo Park y col. "Deep learning inference in facebook data centers: Characterization, performance optimizations and hardware implications". En: *arXiv preprint arXiv:1811.09886* (2018).
- [10] Geert Litjens y col. "A survey on deep learning in medical image analysis". En: *Medical image analysis* 42 (2017), págs. 60-88.
- [11] Binhua Tang y col. "Recent advances of deep learning in bioinformatics and computational biology". En: *Frontiers in genetics* 10 (2019), pág. 214.
- [12] Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek y Omer Berat Sezer. "Deep learning for financial applications: A survey". En: *Applied Soft Computing* 93 (2020), pág. 106384.
- [13] Andreas Kamilaris y Francesc X Prenafeta-Boldú. "Deep learning in agriculture: A survey". En: *Computers and electronics in agriculture* 147 (2018), págs. 70-90.
- [14] Simone Bianco y col. "Benchmark analysis of representative deep neural network architectures". En: *IEEE Access* 6 (2018), págs. 64270-64277.
- [15] Varun Gulshan y col. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs". En: *Jama* 316.22 (2016), págs. 2402-2410.
- [16] Yang Yu y col. "Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN". En: *Computers and Electronics in Agriculture* 163 (2019), pág. 104846.

-
- [17] Joes Staal y col. "Ridge-based vessel segmentation in color images of the retina". En: *IEEE transactions on medical imaging* 23.4 (2004), págs. 501-509.
- [18] AD Hoover, Valentina Kouznetsova y Michael Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response". En: *IEEE Transactions on Medical imaging* 19.3 (2000), págs. 203-210.
- [19] Christopher G Owen y col. "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program". En: *Investigative ophthalmology & visual science* 50.5 (2009), págs. 2004-2010.
- [20] Américo Oliveira, Sergio Pereira y Carlos A Silva. "Retinal vessel segmentation based on fully convolutional neural networks". En: *Expert Systems with Applications* 112 (2018), págs. 229-242.
- [21] Zhexin Jiang y col. "Retinal blood vessel segmentation using fully convolutional network with transfer learning". En: *Computerized Medical Imaging and Graphics* 68 (2018), págs. 1-15.

APÉNDICE A

Copia completa de las publicaciones

En este anexo se encuentra una copia completa de cada una de las publicaciones que forman parte de esta tesis doctoral.

A.1 Primera publicación

APÉNDICE A. Copia completa de las publicaciones

Los artículos que forman parte del apartado “Apéndice A”, debido a restricciones relativas a derechos de autor, han sido retirados de la tesis. En sustitución de los artículos ofrecemos la siguiente información: referencia bibliográfica, enlace a la revista y resumen.

- Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M. E., & Cortés-Ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. In *Computers and Electronics in Agriculture* (Vol. 178, p. 105736). Elsevier BV. <https://doi.org/10.1016/j.compag.2020.105736>

Enlace al texto completo del artículo: <https://doi.org/10.1016/j.compag.2020.105736>

RESUMEN:

Fruit instance segmentation is a key element for autonomous picking systems. This paper proposes a methodology for instance segmentation of strawberries using deep learning techniques. This methodology presents the following changes in relation to the well-known Mask R-CNN network: a new architecture is designed for the backbone and the mask network; the object classifier and the bounding-box regressor are removed; the non-maximum suppression algorithm is replaced with a new region grouping and filtering algorithm without increasing the order of complexity.

For this research, a new high-resolution data set of 3100 strawberry images is obtained, along with the corresponding manually-segmented ground-truth images (StrawDI_Db1 database). In addition to using widely-used metrics based on Average Precision (AP), a new performance metric is proposed: the Instance Intersection Over Union (I2oU), to assess different options of instance segmentation techniques. This material, which is published and available to the research community, allows the rigorous performance comparison between future methodologies.

The results obtained in the 200 images included in the test set show that the proposed methodology significantly reduces the inference time (10 fps vs. 5 fps) while maintaining competitive results to the original Mask R-CNN for mean AP (43.85 vs. 45.36) and mean I2oU (87.27 vs. 87.70) metrics. Thus, the proposed method can be considered as the most promising one in view of its possible integration in automatic strawberry picking systems.

- Gegundez-Arias, M. E., Marin-Santos, D., Perez-Borrero, I., & Vasallo-Vazquez, M. J. (2021). A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model. In *Computer Methods and Programs in Biomedicine* (Vol. 205, p. 106081). Elsevier BV. <https://doi.org/10.1016/j.cmpb.2021.106081>

Enlace al texto completo del artículo: <https://doi.org/10.1016/j.cmpb.2021.106081>

RESUMEN:

Background and Objective: Automatic monitoring of retinal blood vessels proves very useful for the clinical assessment of ocular vascular anomalies or retinopathies. This paper presents an efficient and accurate deep learning-based method for vessel segmentation in eye fundus images.

Methods: The approach consists of a convolutional neural network based on a simplified version of the U-Net architecture that combines residual blocks and batch normalization in the up- and downscaling phases. The network receives patches extracted from the original image as input and is trained with a novel loss function that considers the distance of each pixel to the vascular tree. At its output, it generates the probability of each pixel of the input patch belonging to the vascular structure. The application of the network to the patches in which a retinal image can be divided allows obtaining the pixel-wise probability map of the complete image. This probability map is then binarized with a certain threshold to generate the blood vessel segmentation provided by the method.

Results: The method has been developed and evaluated in the DRIVE, STARE and CHASE_Db1 databases, which offer a manual segmentation of the vascular tree by each of its images. Using this set of images as ground truth, the accuracy of the vessel segmentations obtained for an operating point proposal (established by a single threshold value for each database) was quantified. The overall performance was measured using the area of its receiver operating characteristic curve. The method demonstrated robustness in the face of the variability of the fundus images of diverse origin, being capable of working with the highest level of accuracy in the entire set of possible points of operation, compared to those provided by the most accurate methods found in literature.

Conclusions: The analysis of results concludes that the proposed method reaches better performance than the rest of state-of-art methods and can be considered the most promising for integration into a real tool for vascular structure segmentation.

- Perez-Borrero, I., Marin-Santos, D., Vasallo-Vazquez, M. J., & Gegundez-Arias, M. E. (2021). A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. In *Neural Computing and Applications* (Vol. 33, Issue 22, pp. 15059–15071). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00521-021-06131-2>

Enlace al texto completo del artículo: <https://doi.org/10.1007/s00521-021-06131-2>

RESUMEN:

Instance segmentation is one of the image processing problems where deep learning techniques are beginning to show potential. In agriculture, one of its main application is automatic fruit harvesting. This study focuses on its application on strawberry crops, where the development of automatic harvesting machines is of particular interest. At present, the reference methodology to deal with instance segmentation is Mask R-CNN. However, Mask R-CNN requires a large processing power which limits its implementation in real-time systems. This work proposes a new methodology to carry out instance segmentation of strawberries based on the use of a fully convolutional neural network. Instance segmentation is achieved by adding two new channels to the network output so that each strawberry pixel predicts the centroid of its strawberry. The final segmentation of each strawberry is obtained by applying a grouping and filtering algorithm. The methodology was tested using the publicly available StrawDI_Db1 database. The evaluation results show values of mean average precision (mAP) and mean instance intersection over union (I22oU) of 52.61 and 93.38, respectively, with a processing speed of 30 fps. These figures mean an increase in precision higher than 15% and a fps rate six times higher than those obtained in the reference methodologies based on Mask R-CNN. Therefore, the methodology presented in this paper can be considered as the latest reference methodology for strawberry segmentation, meeting the precision and speed requirements needed for it to be used in the automatic strawberry harvesting systems that work in real time.

APÉNDICE B

Factor de impacto de las publicaciones

Este anexo recoge los informes con el factor de impacto de las publicaciones presentadas en esta tesis doctoral.

B.1 Primera publicación

	Title20	Year	Edition	ISO_ABBREV	Country	IssuesYear	Frequency	ISSN
1	COMPUT ELECTRON AGR	2020	SCI	Comput. Electron. Agric.	ENGLAND	12	M	0168-1699
2	COMPUT ELECTRON AGR	2020	SCI	Comput. Electron. Agric.	ENGLAND	12	M	0168-1699

	Title20	Language	1stYrPub	Categories	Pubcode	Publisher
1	COMPUT ELECTRON AGR	ENGLISH	1983	AH, EV	JM451	COMPUTERS AND ELECTRONICS IN AGRICULTURE
2	COMPUT ELECTRON AGR	ENGLISH	1983	AH, EV	JM451	COMPUTERS AND ELECTRONICS IN AGRICULTURE

	Title20	CATEGORY_DESCRIPTION	CATEGORY_CODE	TOT_CITES	IMPACT_FACTOR
1	COMPUT ELECTRON AGR	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	EV	17657	5,565
2	COMPUT ELECTRON AGR	AGRICULTURE, MULTIDISCIPLINARY	AH	17657	5,565

	Title20	Number of Journals in Category	ARTL_INFLUENCE	CATEGORY_RANKING	IMMEDIACY_INDEX
1	COMPUT ELECTRON AGR	N/D	0,846	Q1	1,181
2	COMPUT ELECTRON AGR	N/D	0,846	Q1	1,181

	Title20	CITED_HALF_LIFE	5YR_IMPACT_FACTOR	EIGENFACTOR	Rank by Impact Factor	QUARTILE_RANK
1	COMPUT ELECTRON AGR	4,4	5,494	0,01646	N/D	19/112
2	COMPUT ELECTRON AGR	4,4	5,494	0,01646	N/D	2/58

B.2 Segunda publicación

	Title20	Year	Edition	ISO_ABBREV	Country	IssuesYear	Frequency	ISSN
1	COMPUT METH PROG BIO	2020	SCI	Comput. Meth. Programs Biomed.	NETHERLANDS	12	M	0169-2607
2	COMPUT METH PROG BIO	2020	SCI	Comput. Meth. Programs Biomed.	NETHERLANDS	12	M	0169-2607
3	COMPUT METH PROG BIO	2020	SCI	Comput. Meth. Programs Biomed.	NETHERLANDS	12	M	0169-2607
4	COMPUT METH PROG BIO	2020	SCI	Comput. Meth. Programs Biomed.	NETHERLANDS	12	M	0169-2607

	Title20	Language	1stYrPub	Categories	Pubcode	Publisher
1	COMPUT METH PROG BIO	ENGLISH	1985	EX, IG, EV, PT	JN201	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE
2	COMPUT METH PROG BIO	ENGLISH	1985	EX, IG, EV, PT	JN201	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE
3	COMPUT METH PROG BIO	ENGLISH	1985	EX, IG, EV, PT	JN201	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE
4	COMPUT METH PROG BIO	ENGLISH	1985	EX, IG, EV, PT	JN201	COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE

	Title20	CATEGORY_DESCRIPTION	CATEGORY_CODE	TOT_CITES	IMPACT_FACTOR
1	COMPUT METH PROG BIO	MEDICAL INFORMATICS	PT	12277	5,428
2	COMPUT METH PROG BIO	COMPUTER SCIENCE; INTERDISCIPLINARY APPLICATIONS	EV	12277	5,428
3	COMPUT METH PROG BIO	ENGINEERING, BIOMEDICAL	IG	12277	5,428
4	COMPUT METH PROG BIO	COMPUTER SCIENCE; THEORY & METHODS	EX	12277	5,428

	Title20	Number of Journals in Category	ARTL_INFLUENCE	CATEGORY_RANKING	IMMEDIACY_INDEX
1	COMPUT METH PROG BIO	N/D	0,95	Q1	2,674
2	COMPUT METH PROG BIO	N/D	0,95	Q1	2,674
3	COMPUT METH PROG BIO	N/D	0,95	Q1	2,674
4	COMPUT METH PROG BIO	N/D	0,95	Q1	2,674

	Title20	CITED_HALF_LIFE	5YR_IMPACT_FACTOR	EIGENFACTOR	Rank by Impact Factor	QUARTILE_RANK
1	COMPUT METH PROG BIO	4,5	5,034	0,01119	N/D	5/30
2	COMPUT METH PROG BIO	4,5	5,034	0,01119	N/D	22/112
3	COMPUT METH PROG BIO	4,5	5,034	0,01119	N/D	19/90
4	COMPUT METH PROG BIO	4,5	5,034	0,01119	N/D	13/110

B.3 Tercera publicación

	Title20	Year	Edition	ISO_ABBREV	Country	IssuesYear	Frequency	ISSN
1	NEURAL COMPUT APPL	2020	SCI	Neural Comput. Appl.	USA	8	B	0941-0643

	Title20	Language	1stYrPub	Categories	Pubcode	Publisher
1	NEURAL COMPUT APPL	ENGLISH	1993	EP	YK501	NEURAL COMPUTING & APPLICATIONS

	Title20	CATEGORY_DESCRIPTION	CATEGORY_CODE	TOT_CITES	IMPACT_FACTOR
1	NEURAL COMPUT APPL	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	EP	18417	5,606

	Title20	Number of Journals in Category	ARTL_INFLUENCE	CATEGORY_RANKING	IMMEDIACY_INDEX
1	NEURAL COMPUT APPL	N/D	0,783	Q1	1,771

	Title20	CITED_HALF_LIFE	5YR_IMPACT_FACTOR	EIGENFACTOR	Rank by Impact Factor	QUARTILE_RANK
1	NEURAL COMPUT APPL	2,7	5,573	0,01651	N/D	31/140