



**ESCUELA DE INGENIERÍA INFORMÁTICA DE OVIEDO**

**GRADO EN INGENIERÍA INFORMÁTICA DEL SOFTWARE**

**IDENTIFICACIÓN DE NEURONAS EN IMÁGENES OBTENIDAS  
DE CULTIVOS**

**PUERTA GONZÁLEZ, PAULA**

**TUTOR: DEL RÍO ÁLVAREZ, ÁNGEL FRANCISCO**

**FECHA: Julio de 2023**



## Resumen

En los últimos años, el conocimiento sobre las células neuronales y, por tanto, el estudio de las posibles causas de enfermedades neurológicas ha sufrido un importante aumento. En este escenario, la introducción de técnicas de Inteligencia Artificial al estudio de los diferentes elementos que pueden propiciar estas enfermedades puede suponer un importante salto cualitativo.

Los cultivos de neuronas ofrecen al investigador valiosa información del desarrollo neuronal. Cuando estos cultivos se exponen a un posible neurotóxico, el recuento posterior de neuronas vivas indicará el nivel de toxicidad de la sustancia. Sin embargo, los métodos tradicionales de conteo de neuronas vivas suelen afectar a su viabilidad, provocando modificaciones e incluso muerte de cultivo

En este TFG de investigación, se proponen métodos no invasivos de identificación y caracterización neuronal en imágenes obtenidas de cultivos, que mediante técnicas de Visión Artificial como la Detección de Objetos y la Segmentación de Instancias, permitan a los investigadores obtener información sobre la viabilidad de las neuronas en un cultivo, sin afectar a su desarrollo.

Los resultados obtenidos demostraron de manera concluyente la superioridad de la técnica de Segmentación de Instancias en la identificación de neuronas en imágenes de cultivos neuronales. El enfoque basado en instancias logró una mayor precisión y una capacidad superior para separar y distinguir células neuronales individuales, hasta en casos de superposición y conexiones entre células.

Los hallazgos obtenidos en esta investigación tienen un gran potencial para impulsar futuras investigaciones en el campo de la neurociencia y contribuir al avance del conocimiento en el estudio de las células neuronales y su participación en enfermedades neurológicas.

**Palabras clave:** neurona, Visión Artificial, *Machine Learning*, *Deep Learning*, Segmentación de Instancias, Detección de Objetos.

## Abstract

In recent years, knowledge about neuronal cells and, therefore, the study of the possible causes of neurological diseases has undergone through a significant increase. In this scenario, the introduction of Artificial Intelligence techniques to the study of the different elements that can cause these diseases can represent an important qualitative leap.

Cultures of neurons offer the researcher valuable information about neuronal development. When these cultures are exposed to a possible neurotoxic, the subsequent count of live neurons will indicate the level of toxicity of the substance. However, traditional methods of counting live neurons often affect their viability, causing modifications and even culture death.

In this final degree project, non-invasive methods of neuronal identification and characterization in images obtained from cultures are proposed, which through Artificial Vision techniques such as Object Detection and Instance Segmentation, allow researchers to obtain information on the viability of the neurons in a culture, without affecting their development.

The results obtained conclusively demonstrated the superiority of the Instance Segmentation technique in the identification of neurons in neuronal culture images. The instance-based approach achieved higher precision and a superior ability to separate and distinguish individual neuronal cells, even in cases of overlap and connections between cells.

The findings obtained in this research have great potential to promote future research in the field of neuroscience and contribute to the advancement of knowledge in the study of neuronal cells and their participation in neurological diseases.

**Keywords:** neuron, Artificial Vision, Machine Learning, Deep Learning, Instance Segmentation, Object Detection

## Índice

Resumen.....	3
Abstract .....	4
Índice de Ilustraciones .....	7
Índice de Tablas.....	9
1 Introducción .....	10
2 Hipótesis de partida y objetivos concretos .....	14
2.1 Hipótesis.....	14
2.2 Objetivos .....	14
3 Estado del arte .....	16
3.1 Antecedentes en Bioquímica .....	16
3.2 Análisis automático de imágenes.....	18
3.3 Procesamiento y análisis de imágenes.....	19
3.4 Problemática de la oclusión .....	20
3.5 Segmentación vs. Detección de Objetos.....	21
3.6 Detección de Objetos.....	22
3.7 Segmentación Semántica .....	24
3.8 Segmentación de instancias.....	25
4 Técnicas de Visión Artificial aplicadas a la caracterización de las células.....	26
4.1 VA tradicional vs <i>Deep Learning</i> .....	27
4.1.1 Redes neuronales.....	29
4.1.1.1 Redes neuronales convolucionales .....	31
4.1.1.2 Otras redes neuronales usadas en visión artificial.....	32
4.2 Clasificación, identificación de objetos y segmentación.....	33
4.2.1.1 Clasificación de imágenes .....	34
4.2.1.2 Detección de Objetos .....	35

4.2.1.3	Segmentación Semántica de imágenes.....	37
4.2.1.4	Segmentación basada en instancias.....	39
5	Experimentación .....	42
5.1	Dataset .....	42
5.2	Materiales .....	44
5.3	Métodos .....	44
5.3.1	Arquitectura YOLO .....	44
5.4	Métricas.....	46
5.4.1	Precisión .....	48
5.4.2	Recall .....	49
5.4.3	mAP (mean Average Precision) .....	49
5.5	Experimento .....	50
5.5.1	Preprocesamiento de las imágenes .....	50
5.5.2	Obtención de las anotaciones .....	51
5.5.3	Entrenamiento de un modelo de Segmentación de Instancias .....	51
5.5.4	Entrenamiento de un modelo de Detección de Objetos .....	52
6	Resultados .....	53
7	Conclusiones y trabajo futuro .....	57
7.1	Adecuación de la solución final .....	57
7.2	Evaluación de alcance .....	57
7.3	Trabajo futuro .....	59
8	Glosario .....	60
9	Referencias.....	67

## Índice de Ilustraciones

Ilustración 1: Esquema que muestra las distintas fases de ML.....	28
Ilustración 2: Diferencia entre ML y DL.....	29
Ilustración 3: (a) Red neuronal de deep learning, (b) Modelo de regresión lineal .....	30
Ilustración 4: Esquema de una red neuronal convolucional .....	32
Ilustración 5: Unidad recurrente de una red con arquitectura RNN.....	32
Ilustración 6: Esquema de la arquitectura de una red GAN [39] .....	33
Ilustración 7: Clasificación de una imagen con una clase presente .....	34
Ilustración 8: Detección de tres clases distintas de objetos .....	36
Ilustración 9: De izquierda a derecha: imagen original; segmentación basada en bordes; segmentación basada en regiones con un umbral bajo; segmentación en regiones con un umbral menos exacto. ....	38
Ilustración 10: (de izquierda a derecha y de arriba a abajo) Clasificación de imágenes, Detección de Objetos, Segmentación Semántica, Segmentación de Instancias.....	40
Ilustración 11: Imagen de las neuronas cultivadas (izquierda) y máscara correspondiente a la imagen original (derecha) .....	43
Ilustración 12: Neuronas en una etapa temprana (izquierda); Neuronas en estado avanzado de cultivo (derecha) .....	43
<i>Ilustración 13: Arquitectura de la red que usa YOLO .....</i>	<i>45</i>
<i>Ilustración 14: Ejemplo de cálculo y normalización de las predicciones dentro de una región en YOLO.....</i>	<i>46</i>
<i>Ilustración 15: Clasificación de una predicción en segmentación .....</i>	<i>47</i>
<i>Ilustración 16: Clasificación de una predicción para Detección de Objetos con un umbral de 0,5 .....</i>	<i>48</i>
<i>Ilustración 17: Cálculo de IOU en el caso de Detección de Objetos (izquierda) y en el caso de Segmentación Semántica (derecha).....</i>	<i>48</i>
<i>Ilustración 18: Área bajo la curva que se usa para calcular el mAP [47] .....</i>	<i>49</i>

*Ilustración 19: Aplicación del algoritmo Voronoi-Otsu a una máscara con neuronas superpuestas*  
..... 51

Ilustración 20: Curvas de Precision-Recall de los modelos de Detección de Objetos (rojo) y Segmentación de Instancias (azul)..... 54

Ilustración 21: Matrices de confusión de los modelos, a la izquierda, el modelo de Detección de Objetos, y, a la derecha, el modelo de Segmentación de Instancias..... 55

## Índice de Tablas

Tabla 1: Métricas obtenidas por el modelo de Segmentación de Instancias para la segmentación .....	53
Tabla 2: Métricas obtenidas por cada modelo entrenado para la localización de bounding boxes .....	53
Tabla 3: Tasas de pérdida de los modelos entrenados .....	55

## 1 Introducción

El problema de la caracterización de neuronas (conteo, forma, movimiento...) en las primeras fases de vida de estas, aporta información muy importante a los investigadores. Muchas enfermedades neurológicas pueden tener su origen, de muy diversa manera, en estos estados incipientes, que es el momento en el que las neuronas buscan su hueco y crean relaciones entre ellas a través de las neuritas.

Caracterizar estos comportamientos a nivel global, intentando generar un modelo común, es un problema difícil de resolver y que aún ahora plantea muchísimas dudas a los investigadores. De forma manual, resulta muy complicado hacer seguimientos individuales de los que extraer un comportamiento global. La Inteligencia Artificial abrió un nuevo abanico de posibilidades para el análisis celular, pero, aun así, en muchos casos las técnicas IA tradicionales no son suficientemente efectivas o requieren de estrategias adicionales.

Este Trabajo Fin de Grado forma parte de una investigación mayor, en la que se pretende caracterizar las neuronas de un cultivo mediante técnicas de aprendizaje profundo. En concreto, en este trabajo de investigación, se aborda el problema de la detección, caracterización morfológica, y conteo automático de neuronas vivas en imágenes de cultivos. Estas imágenes son tomadas, mediante microscopio convencional, en diferentes estados del cultivo y, se pretende con ellas, entrenar el modelo de *Deep Learning* (DL). La intención es que, una vez entrenado, el modelo sea capaz de identificar por sí solo las neuronas de forma precisa, así como delimitar su forma dentro de la imagen, de tal manera que se puedan realizar conteos que faciliten la observación de la evaluación de las neuronas cultivadas.

Desde que nacemos estamos expuestos a neurotóxicos que, absorbidos por distintos tipos de mucosas, entran en nuestro cuerpo y pueden influir en el desarrollo de distintas enfermedades neurodegenerativas. Por ello, el estudio del efecto de neurotóxicos en el desarrollo de cultivos de neuronas representa una pieza clave de la investigación de dichas enfermedades. Midiendo el número de neuronas, su forma y su volumen antes de agregar una sustancia a un cultivo de neuronas, y repitiendo la medición después, se puede medir el grado de neurotoxicidad de dicha sustancia.

El mayor problema, es que los métodos tradicionales de conteo y caracterización de neuronas suelen ser altamente invasivos y muchas de las técnicas utilizadas pueden provocar en sí mismas, cambios e incluso muertes neuronales. De hecho, la tinción de neuronas sigue siendo,

hoy en día, una de las técnicas más utilizadas para facilitar el conteo de neuronas vivas, a sabiendas de que el cultivo morirá después de la tinción, haciendo inviable un nuevo recuento futuro del mismo cultivo.

La solución adoptada tradicionalmente es el conteo de forma manual por personal capacitado que distingue entre neuronas vivas y muertas. Este proceso manual requiere una gran cantidad de trabajo y consume mucho tiempo. Además, puede generar variabilidad debido a discrepancias en los criterios de conteo entre los diferentes técnicos, lo que hace necesario el conteo del mismo conjunto de datos por varias personas, de forma que se asegure un consenso. La detección de células neuronales muertas introduce una problemática añadida, ya que una vez muertas las células tienden a agruparse a otras células independientemente de que las segundas estén vivas o muertas, tornando más difícil la detección de células muertas individuales [1].

Es por todo esto, que se hace necesaria la automatización de la identificación de las neuronas con técnicas como las que se tratarán en este TFG. Este problema de automatización supone una serie de retos, uno de los más importantes, es la técnica que se usará para entrenar el modelo que reconocerá las neuronas. Las principales opciones son la segmentación y la Detección de Objetos, de todas estas técnicas se hablará en profundidad a lo largo del trabajo.

Para tomar la decisión correcta, es importante tener en cuenta desde un principio las necesidades del proyecto, así como las dificultades a las que se tendrá que hacer frente.

Para comenzar, hay que tener en cuenta la complejidad y variabilidad de un cultivo neuronal. A diferencia de otros tipos de cultivos celulares, un cultivo de neuronas, según se va desarrollando va “tejiendo” una red de neuritas que van añadiendo complejidad a la imagen. Además, cualquier cultivo de este tipo, no solo contiene neuronas, sino que contiene células gliales, neuritas, neuronas muertas, y fragmentos de estas estructuras que complican enormemente la detección de neuronas vivas. Las neuronas son el tipo de células principalmente presentes en el sistema nervioso, seguido de las células gliales. Las primeras, son las encargadas de recoger la información del entorno, procesarla y emitir una respuesta, se consideran la unidad elemental de procesamiento y transmisión del sistema nervioso, de ahí la importancia que tienen para esta investigación. Por otro lado, están las células gliales, que tienen una misión auxiliar, regulan las transmisiones sinápticas entre las neuronas por lo que su existencia es primordial para el correcto funcionamiento de las comunicaciones neuronales [2], [3].

En los cultivos de neuronas que se han usado para la investigación descrita en este TFG están presentes ambos tipos de células, el personal que realiza el conteo manual de las neuronas vivas, está capacitado para distinguir ambos tipos de células, de igual forma, el modelo automatizado tiene que diferenciar una célula neuronal de cualquier otro tipo de célula que pueda estar presente en el cultivo.

Otra de las problemáticas que presentan las imágenes es la superposición de células neuronales, que complica la detección de las neuronas ya que los bordes de estas se ven ocluidos por la neurona adyacente.

Además de las dificultades expuestas, hay una serie de características, determinadas siempre por los expertos, que hacen que estos cultivos sean prácticamente únicos. Para empezar, las neuronas utilizadas, ya que los expertos han determinado que, para abordar este estudio, las neuronas utilizadas sean de cerebelo de rata. Las ratas usadas son de corta edad en el punto de madurez neuronal óptimo definido por los expertos. Además, se pretende que el sistema sea transversal, pudiendo identificar neuronas en cualquiera de las etapas de desarrollo del cultivo. Las neuronas presentan cambios morfológicos y movimiento en las primeras etapas, quedándose más estáticas y constantes en fases posteriores. En cambio, en estas fases posteriores ya existen grandes conexiones entre ellas -neuritas-, restos de células muertas, y otros objetos que dificultan la detección. Por lo que hacer un detector global, presenta una novedad respecto al estado del arte. Además, se utilizará concretamente, un microscopio óptico convencional en contraste de fase.

Con el objetivo de enfrentar estos desafíos, se han aplicado técnicas de aprendizaje profundo para crear un método de análisis automatizado que pueda ser empleado directamente a imágenes de cultivos neuronales capturadas con un microscopio de contraste de fase. Se han entrenado dos modelos distintos, uno de segmentación basada en instancias y otro de Detección de Objetos, ambos con arquitectura YOLO.

La segmentación mediante DL es una técnica que es capaz de realizar una clasificación a nivel de píxel, en este caso, estaríamos hablando de Segmentación Semántica. Otro tipo de segmentación, a nivel de instancia, consigue que el modelo utilice también la información de aislamiento de cada entidad recogida en las máscaras de entrenamiento, lo que dará como resultado una clasificación a nivel de píxel junto con una localización de cada objeto a nivel individual. Para los objetivos de caracterización de neuronas de este TFG, una Segmentación de Instancias proporcionaría toda la información necesaria. Sin embargo, para realizar conteo y

localización de objetos, tradicionalmente en el campo de la Visión por Computador, se utiliza Detección de Objetos. Por eso se considera interesante la experimentación con ambas técnicas para su posterior comparación de resultados.

Este trabajo fin de grado ha sido dividido en las siguientes secciones: una primera sección, la presente, en la que se introduce el tema del cuál trata el trabajo. La sección 2, presenta las hipótesis y objetivos establecidos. Una 3ª sección, ofrece un breve resumen de las investigaciones relacionadas que han sido llevadas a cabo hasta el momento. En la sección 4, se centra en realizar una explicación exhaustiva de las técnicas de visión artificial existentes y que, por tanto, podrían suponer una solución a la tarea que concierne a la investigación. En una 5ª sección, se detallan los recursos y técnicas empleados en el estudio descrito en este TFG, abarcando los materiales, el conjunto de datos, las estrategias y los procedimientos experimentales realizados. A lo largo de la 6ª sección, se exponen los resultados obtenidos de los dos enfoques analizados, incluyendo las evaluaciones cuantitativas del desempeño de los modelos de Segmentación de Instancias y Detección de Objetos aplicados a las células neuronales. Por último, en la 7ª sección, se pasan a debatir las conclusiones y se aportan algunas sugerencias para futuras investigaciones en este ámbito.

## 2 Hipótesis de partida y objetivos concretos

### 2.1 Hipótesis

En los últimos años, las soluciones basadas en DL han venido aportando resultados muy productivos en tareas de análisis de imágenes y cada vez se está aplicando más su uso al campo de biología y la medicina. Debido a ello se pretende usar estas tecnologías para la identificación de neuronas, pudiendo de esta forma realizar conteos de forma sencilla, para conseguir esto, es necesario explorar la viabilidad de las distintas técnicas existentes.

Para poder determinar la viabilidad de las neuronas, se atiende a múltiples rasgos morfológicos de la neurona, tan diversos que, en múltiples ocasiones, los expertos no son capaces de determinar la viabilidad de una neurona a simple vista. Esos rasgos incluyen, forma, reflexión de luz, permeabilidad de la membrana, etc. Un sistema de Inteligencia Artificial que ayudase en esta determinación debería de, no solo intentar contar las neuronas vivas, sino de ofrecer información morfológica que permitiera a los investigadores obtener una previsión de lo que pueda suceder con esa célula en un futuro. Las técnicas de Visión Artificial que pueden ayudarnos en esta labor son: La Detección de Objetos y la Segmentación. Dentro de las técnicas de Segmentación, se considerará la Segmentación de Instancias, técnica que por sí sola, ofrecería todos los datos necesarios para nuestra investigación.

En el ámbito biomédico, la Segmentación basada en Instancias ha venido demostrando su potencial respecto al resto al resto de técnicas, esto se debe a que mezcla las ventajas de las principales técnicas usadas hasta el momento, la segmentación y la Detección de Objetos. La demostración de la superioridad de esta técnica resultaría determinante para el análisis de neuronas, ya que permitiría una simplificación, no sólo del conteo de estas, sino también el estudio de su morfología a lo largo de su vida, además de permitir la realización de seguimientos de neuronas de forma individualizada. Sin embargo, este campo de estudio está supeditado a las complicaciones inherentes al propio desarrollo de las neuronas en el cultivo, que provocan que las imágenes sean muy distintas en edades temprana y en fases más avanzadas. Y por ello, aún está por demostrar si una Segmentación combinada con Detección de Objetos ofrecería, en este contexto mejores resultados que la Segmentación de Instancias por si sola.

### 2.2 Objetivos

La presente investigación nace con el propósito de permitir un análisis automatizado de los cultivos neuronales, siendo el principal objetivo poder llevar a cabo un recuento, rápido y sin

grandes esfuerzos, de las neuronas presentes en un cultivo para poder estudiar el desarrollo de estas. Es por ello, por lo que se propone el uso de técnicas de DL para lograr este fin.

De forma explícita, los objetivos de la investigación presente en este TFG son:

- Investigación del estado del arte de Visión Artificial
- Estudio de las diferentes técnicas de Visión Artificial que podrían ser aplicadas al problema de detección
- Consenso de las técnicas más adaptadas a la solución que se busca
- Aplicación de algoritmos de preprocesamiento de imágenes
- Entrenamiento de modelos de *Deep Learning*
- Análisis de los resultados obtenidos con las técnicas propuestas y comparación de estas
- Determinación de la técnica más efectiva para la identificación de las neuronas

La realización de este presente trabajo tiene otros objetivos para el alumno, estos son:

- Aumentar los conocimientos del alumno en el área de la Inteligencia Artificial
- Comprender los pasos que implican la realización de una investigación científica
- Adquirir facilidad en la búsqueda de información a través de recursos de carácter científico

## 3 Estado del arte

El problema de la caracterización celular, y en concreto neuronal, ha sido ampliamente estudiado desde distintos ámbitos. Para poder entender la necesidad de las últimas herramientas basadas en DL, como la presentada en este trabajo de investigación, es necesario relatar todos los antecedentes y la motivación de este estudio desde el punto de vista de la Bioquímica.

### 3.1 Antecedentes en Bioquímica

Actualmente, alrededor de un 12% de las muertes a nivel mundial son causadas por trastornos neurológicos. Además, estas enfermedades tienen un gran impacto en los años de vida saludables de las personas que las padecen. Sin embargo, se sabe poco acerca de muchos de ellos. Entre estas neuropatologías, se pueden encontrar cefaleas, enfermedades cerebrovasculares, esclerosis múltiple, infecciones que afectan al sistema nervioso, epilepsia, así como otras denominadas neurodegenerativas. De estas últimas, las más populares son el Parkinson, Alzheimer y otras demencias, y la esclerosis lateral amiotrófica [4].

Las enfermedades neurodegenerativas afectan negativamente al sistema nervioso central, atacando tanto a las células madre del cerebro, encargadas de la generación de nuevas neuronas, como a las neuronas existentes, lo que eventualmente conduce a su muerte [5]. Por lo tanto, es crucial desarrollar métodos cuantitativos que permitan una evaluación precisa y ágil de la calidad de las células neuronales, así como de las alteraciones estructurales que ocurren con el tiempo[5], [6].

Desde la etapa prenatal, los seres humanos están expuestos a neurotoxinas que ingresan al sistema nervioso a través de las mucosas, lo que perturba las funciones fundamentales de las células neuronales [7]. Se cree que esta exposición puede estar relacionada con el desarrollo de estas enfermedades neurodegenerativas. Por lo tanto, evaluar los efectos químicos de las neurotoxinas en modelos neuronales in vitro de enfermedades como el ictus o las enfermedades neurodegenerativas, se considera una fuente importante de resultados [8]. Los enfoques bioquímicos más comunes para estudiar la neurotoxicidad se basan en el impacto de las cascadas de muerte celular en la actividad de enzimas, la permeabilidad de la membrana celular y la función mitocondrial de las neuronas. Sin embargo, estos métodos presentan desafíos en términos de dificultad y costos, y no permiten la observación de los efectos de la neurotoxicidad a nivel celular individual [9].

Debido a las limitaciones inherentes que presenta la metodología basada en la cascada de muertes, se propone el estudio de las células neuronales vivas mediante una técnica de observación, considerada altamente efectiva en el análisis celular llamada “microscopía de contraste de fases”, en la cual se usa un microscopio concreto que se basa en la refracción de los elementos observados [9]. La aplicación de esta técnica es factible ya que la exposición a los agentes neurotóxicos es observable en las neuronas cultivadas a través de los cambios en su morfología, su crecimiento y la integridad de las neuritas [9].

Cuando una célula se empieza a morir, entra en fase de apoptosis, se puede detectar mediante la óptica de contraste de fases que una célula entra en esta fase cuando esta se encoje y se vuelve más brillante [1], sin embargo, una vez la célula está muerta, no es posible diferenciar de forma precisa mediante óptica de contraste de fases las células muertas de las sanas, es por esta razón que, para lograr una evaluación cuantitativa precisa de la viabilidad neuronal, resulta necesaria la utilización de métodos complementarios a las imágenes de contraste de fase. Estos métodos adicionales pueden incluir técnicas de tinción celular o marcadores fluorescentes que permitan detectar y distinguir de manera confiable entre células vivas y células muertas, al combinar estas herramientas con las imágenes de contraste de fase, se obtiene una evaluación más completa y precisa de la viabilidad neuronal.

Cómo se ha comentado en la introducción, una de las técnicas más utilizadas para distinguir entre células vivas y células muertas es la tinción diferencial con compuestos como el diacetato de fluoresceína, la calceína AM o el bromuro de etidio. Estas sustancias permiten obtener una observación detallada de las estructuras celulares, revelando aspectos clave de la viabilidad neuronal [10], [11]. Cada tipo de tinción proporciona información específica sobre las células en estudio. Por ejemplo, el diacetato de fluoresceína puede colorear las células de diferentes colores según su estado de vida o muerte, mientras que la calceína AM puede reflejar la viabilidad de las células en función de la intensidad de un mismo color [10], [11].

Aunque esta metodología de tinción es sencilla, económica y brinda información confiable y relevante sobre la viabilidad celular a nivel individual, también presenta inconvenientes significativos. Uno de ellos es la naturaleza destructiva de la técnica, ya que el uso de disolventes orgánicos para la tinción puede causar la permeabilización de la membrana celular neuronal, lo cual restringe la posibilidad de realizar estudios a largo plazo sin el uso de distintas muestras de un mismo cultivo en cada análisis. Además, se requiere un seguimiento preciso e individual de las neuronas viables para garantizar una evaluación precisa [9].

### 3.2 Análisis automático de imágenes

El análisis celular basado en imágenes desempeña un papel fundamental en la investigación biomédica, al proporcionar una herramienta que permite obtener información detallada y cuantitativa sobre las células estudiadas. Mediante el uso de microscopios, es posible capturar imágenes que proporcionan información valiosa sobre la morfología celular y los cambios sufridos por estas durante el cultivo, estas imágenes permiten examinar la estructura y los procesos celulares, lo que brinda una comprensión más profunda de los fenómenos biológicos en estudio. Gracias al análisis de imágenes, es posible identificar características y patrones específicos, evaluar la viabilidad celular y detectar posibles anomalías o alteraciones en tiempo real.

Desde el descubrimiento de la estructura neuronal por Ramon y Cajal, la neurociencia es una de las disciplinas más ricas en información, por lo que el manejo de tanta información se vuelve una tarea compleja. A esta complejidad, se suma la falta de consenso en los criterios para clasificar neuronas según su tipo. A pesar de existir una gran cantidad de información acerca del sistema nervioso, aún no existen estrategias sistemáticas para estudiar las células en profundidad. Debido a esto, en [12] proponen un sistema informático que, tras establecer un estándar de terminología, se pueda alimentar de distintas descripciones de neuronas que posteriormente serán estudiadas usando una caracterización ontológica. Aunque esta propuesta representa un punto de partida interesante, la automatización del análisis de los cultivos neuronales se vuelve cada vez más necesario, ya que se ha demostrado que las imágenes obtenidas mediante microscopía ofrecen datos de gran importancia acerca de la forma y el comportamiento de las células cultivadas. En este sentido, la integración de técnicas de procesamiento y análisis de imágenes se presenta como una solución prometedora para agilizar y mejorar el estudio de las células neuronales.

Como ya se ha mencionado anteriormente, el campo de la neurociencia recoge una vasta cantidad de datos y la información que reportan las nuevas técnicas de imagen imposibilitan el análisis manual de estos. La creación de nuevos sistemas de caracterización y *tracking* puede tener un gran impacto en las metodologías usadas hasta el momento y por ende, en la investigación de las enfermedades neurodegenerativas [13]. Los escasos métodos de seguimiento de células neuronales disponibles actualmente, como NeuronJ [14], presentan grandes limitaciones ya que requieren de identificaciones manuales y examinan un reducido conjunto de características, además, suelen estar específicamente diseñados para ser aplicados

en imágenes de microscopía de fluorescencia, la cual ya se ha visto que implica un problema de muerte celular.

Para abordar estos desafíos, se requiere el desarrollo de nuevos enfoques que permitan el análisis automatizado y robusto de las imágenes de células neuronales. La adopción de estos nuevos sistemas de caracterización y seguimiento de células en la investigación neurocientífica puede proporcionar una visión más profunda de los procesos celulares, abriendo una puerta a una comprensión más completa de los mecanismos subyacentes.

### 3.3 Procesamiento y análisis de imágenes

Los avances en el análisis automatizado de imágenes y el seguimiento celular ofrecen una oportunidad única para comprender mejor las células neuronales y su participación en enfermedades neurológicas. Sin embargo, es importante tener en cuenta las problemáticas que presentan las imágenes obtenidas por microscopio. Una de las problemáticas más importantes está en la pérdida de calidad de las imágenes, que pueden ocasionar la presencia de zonas borrosas, las cuales a su vez pueden provocar la pérdida de la claridad de los contornos de las células. Además, existe un concepto llamado anisotropía que agrega otra capa de complejidad, esta propiedad expone que la materia observada puede cambiar sus propiedades en función de la dirección en las que esta está siendo observada [15]. A pesar de que esta última es imposible de evitar, sí es posible mejorar el ruido en una imagen, de esto se encarga el procesamiento de imágenes. El procesamiento de imágenes se encarga de mejorar el contraste, eliminar el ruido con la finalidad de optimizar la calidad de la imagen y, en ocasiones, aplicar técnicas de escalado si las imágenes de entrada tienen tamaños dispares. El procesamiento puede llevarse a cabo mediante el ajuste de parámetros como el contraste, la claridad, la exposición, etc., pero en casos en los que eso no es suficiente puede recurrirse al uso de redes GAN [16]. Estas redes GAN, si son correctamente entrenadas, pueden resolver los problemas de calidad y resolución que pueda tener una imagen. Es habitual confundir el concepto de procesamiento con el de análisis, sin embargo son dos etapas diferentes e igualmente necesarias para poder realizar estudios basados en imágenes.

El análisis de imágenes implica el uso de métodos de extracción de características para detectar y clasificar objetos en función de sus atributos como textura, color, escala de gris, entre otros. En el campo del análisis de imágenes, la clasificación y la segmentación son dos tareas fundamentales para comprender y extraer información de las imágenes, ambas técnicas se utilizan para organizar y etiquetar los objetos presentes en una imagen, pero difieren en su

enfoque y nivel de detalle. Mientras que la clasificación asigna una etiqueta predefinida a una imagen completa, la segmentación de imágenes divide la imagen en segmentos más pequeños con el fin de delimitar los objetos individuales presentes en la imagen.

### 3.4 Problemática de la oclusión

Realizar un análisis preciso de las células neuronales mediante imágenes microscópicas es un desafío importante en el campo de la visión por computadora y la neurociencia. Una de las problemáticas más significativas en este contexto es la oclusión de células, donde múltiples células se superponen y dificultan su distinción y análisis individual. Se ha observado que este fenómeno puede generar un rendimiento poco confiable en algunos algoritmos de aprendizaje profundo basados en CNN no consideran la información temporal, con el fin de superar esta problemática y lograr un rendimiento óptimo en la segmentación de las imágenes, se han propuesto enfoques diferentes. Arbelle et al. [17] propusieron un algoritmo actualizado de una red generativa antagónica (GAN) para realizar una segmentación, esta red GAN estaba compuesta de un estimador basado en una CNN y un discriminador. Tradicionalmente, las redes GAN están compuestas por un generador y un discriminador, sin embargo, en este estudio los investigadores sustituyeron el generador por un estimador. En este modelo se ataca la problemática de las células adyacente dividiendo el dominio de la imagen en tres regiones: fondo, primer plano (núcleo de la célula) y contorno de las células. Este enfoque innovador demuestra la importancia de considerar la información temporal en la segmentación de imágenes microscópicas de neuronas, ya que al evitar la generación de imágenes a partir de vectores aleatorios ruidosos y enfocarse en estimaciones más precisas, se obtiene un mejor rendimiento en la segmentación, incluso en casos de oclusión celular.

Otros planeamientos para resolver la superposición han sido sistemas de análisis combinados, que integran múltiples algoritmos de aprendizaje profundo (DL) para lograr una identificación y segmentación exhaustiva de los componentes en imágenes. Siguiendo esta línea, Arbelle et al. exploraron un nuevo enfoque en una publicación posterior [18], donde proponen un sistema que incorpora bloques convolucionales de *Long Short-Term Memory* (LSTM) en la arquitectura U-Net. Estos investigadores propusieron tres escenarios de aplicación para el sistema: la integración de LSTM en la sección del codificador, en la sección del decodificador y en el conjunto completo de U-Net (incluyendo tanto el codificador como el decodificador). Al incorporar LSTM, se busca aprovechar la capacidad de los bloques de memoria a corto y largo plazo para capturar la información temporal y contextual relevante en la segmentación de células superpuestas.

Isola et al. [19] propusieron otro enfoque interesante al emplear U-Net como generador dentro de una estructura GAN condicional. El objetivo de este enfoque es el de facilitar la transferencia de información de bajo nivel a lo largo de la red y generar imágenes más detalladas y realistas, considerando el desafío de la superposición celular. Esta estrategia busca aprovechar la capacidad de las redes generativas adversariales para aprender y generar imágenes segmentadas de alta calidad.

Todos estos enfoques innovadores que se han expuesto demuestran un significativo progreso en la investigación y desarrollo de métodos avanzados para abordar el desafío de la superposición de células en la segmentación de imágenes microscópicas. La combinación de arquitecturas de redes neuronales profundas con elementos como los bloques convolucionales de LSTM y las redes GAN ha proporcionado nuevas herramientas y oportunidades para mejorar la precisión y eficacia en el análisis de imágenes de neuronas.

La superposición de las células biológicas dificulta la identificación y delimitación precisa de las células individuales, esto ha servido de impulso para el desarrollo de algoritmos y enfoques avanzados basados en aprendizaje profundo para abordar la solución al problema. En los últimos años, se han propuesto varios algoritmos de segmentación basados en aprendizaje profundo para la identificación directa de neuronas en imágenes sin necesidad de crear bounding boxes. Siguiendo esta línea, Schmidt et al. exploró un enfoque innovador en el cual se usaban polígonos estrellados (*Star-Convex Polygons*) [20], en sustitución de las cajas, para localizar las células y realizaron modificaciones en la reconocida red U-Net para la segmentación precisa de las células. En este modelo, los polígonos se forman desde el centro de la célula y trazan radios hasta encontrarse con un píxel de otra clase, de esta forma cada radio podrá usarse para delimitar los contornos de la célula. Otros estudios [21], plantearon un enfoque de Segmentación de Instancias basado en cajas que buscaba generar *bounding boxes* utilizando puntos clave. El modelo presentado en esta investigación se servía de un algoritmo de detección de puntos clave y de una red troncal basada en ResNet50, gracias a esta unión se logró reconocer el total de cinco puntos que conforman un *bounding box*, es decir, las cuatro esquinas y el punto central.

### 3.5 Segmentación vs. Detección de Objetos

La segmentación y la detección (o reconocimiento) de imágenes son dos métodos ampliamente utilizados en el análisis de imágenes de microscopía, y se han aplicado en diversos estudios de detección basados en imágenes.

La Segmentación es simplemente una separación a nivel de píxel en función a las características de estos. Cuando es realizada mediante DL y tiene como objetivo la asignación de cada píxel a diferentes clases semánticas, se denomina Segmentación Semántica. En este tipo de Segmentación, se agruparán los píxeles en base a características extraídas, separando los pertenecientes a la(s) clase(s) buscada del fondo de la imagen. Esta técnica ha reportado resultados robustos y con alta capacidad de caracterizar las neuronas [15].

Si bien la Segmentación Semántica se puede considerar una clasificación a nivel de píxel, la Detección de Objetos es una clasificación a nivel del propio objeto. En Detección de Objetos se buscan apariciones completas de los objetos a clasificar por distintas partes de la imagen con el objetivo de determinar la posición de las distintas apariciones del objeto buscado. Estos enfoques han demostrado ser efectivos en el reconocimiento de células normales y anormales en imágenes microscópicas [22].

Mientras la Segmentación Semántica puede ofrecer resultados de la morfología de las neuronas, pero no es capaz de proporcionar información sobre las neuronas a nivel individual; la Detección de Objetos permite delimitar y contar las neuronas a nivel individual.

Existe una segmentación más avanzada, denominada Segmentación de Instancias. Esta técnica, además de clasificar a nivel de píxel, es capaz de reconocer las diferentes instancias de las clases contempladas. A diferencia de lo que pueda parecer, este tipo de segmentación no es una combinación de Segmentación Semántica + Detección de Objetos, sino que utiliza una arquitectura más compleja donde la información a nivel de píxel se realimenta en cada iteración con la información a nivel de objetos.

Estas técnicas serán vistas en mayor detalle en el apartado [“Técnicas de Visión Artificial aplicadas a la caracterización de las células”](#). Ahora se expondrán trabajos destacados publicados recientemente en el campo de estudio de detección de células utilizando estas técnicas.

### 3.6 Detección de Objetos

Tanto la Clasificación de imágenes como la Detección se centran en buscar apariciones de objetos dentro de una imagen, sin embargo, su forma de buscarlos y reportarlos es bien distinta.

La Clasificación de imágenes se centra en decir si los objetos de una clase están o no presente en una imagen, pero no se encarga de devolver las localizaciones de estos ni de realizar un

recuento. Ya que la clasificación se limita a reportar las clases de los objetos presentes en la imagen, su uso en casos en los que varios objetos de una misma clase están presentes dentro de la imagen, suele reportar información escasa. En cambio, la Detección de Objetos se encarga de buscar trozos de la imagen que puedan coincidir con el objeto a detectar, además, la Detección si almacena la información de la localización y el tamaño del objeto detectado.

La Detección de Objetos ha demostrado ser más rápida y precisa que la Segmentación para objetos cotidianos. Pang et al. [23] lo probaron mediante el desarrollo de una combinación del método de la ventana deslizante con máquinas de vector soporte lineales para detectar manos, caras y peatones en vídeos. En esta investigación demostraron que este método tenía una mayor habilidad de generalización y era mucho más rápido que los ya desarrollados. También se han aplicado técnicas de Detección de Objetos en el contexto de imágenes de células, como se evidenció en [24], donde aplican Detección de Objetos con diferentes extractores a imágenes de citologías de fluido pleural, los resultados obtenidos demostraron una rapidez y precisión significativa.

Los métodos de Detección de Objetos se enfocan en generar cajas delimitadoras que abarcan las células, también conocidas como *bounding boxes*. En el uso de estas, se muestrean las imágenes de entrenamiento, encuadrando cada neurona, y a partir de ahí se procede a entrenar los modelos. Este método tiene como desventaja que ubicar todos los *bounding boxes* puede ser bastante laborioso y llevar mucho tiempo, mientras que, la Segmentación de imágenes tiene el potencial de reducir la carga computacional. Esto se debe a que la Segmentación introduce algoritmos que infieren las áreas donde es probable que se encuentren las células.

La aplicación de técnicas de aprendizaje profundo a la Detección de Objetos ya ha demostrado en numerosos casos su eficiencia para identificar objetos cotidianos, sin embargo, su uso para detectar imágenes biológicas no está tan estudiado [25]. Estos métodos tienen la capacidad de identificar dentro de una imagen aquellos objetos que se le habían enseñado previamente, devolviendo la posición en la que encuentra. La detección basada en imágenes de microscopio tiene sus propios desafíos peculiares, como son la resolución de células superpuestas, o distinguir entre células de distintas morfologías y tipos. En estudios realizados hasta el momento, también han resultado problemáticas las escalas usadas en la obtención de las imágenes, ya que las redes neuronales han demostrado resultados insatisfactorios frente a imágenes tomadas a distintas escalas [26].

El enfoque de la investigación tratada en este trabajo fin de grado se ha centrado en la detección de neuronas utilizando imágenes de cultivos neuronales primarios, este tipo de cultivos son ampliamente utilizados en estudios de neurotoxicología y neurociencia. En las investigaciones realizadas hasta la fecha en las que se aplican algoritmos de DL, abundan las que usan imágenes obtenidas con microscopios de campo claro, seguidas de las tomadas haciendo uso de microscopía fluorescente o electrónicos, mientras que tan sólo un 10% de las publicaciones han sido realizadas con microscopía de contraste de fases. Además también hay una clara dominancia de las redes neuronales convolucionales (CNN) respecto al uso de otro tipo de redes neuronales [25].

### 3.7 Segmentación Semántica

Como ya se ha explicado con anterioridad, la Segmentación radica en asignar una clase semántica a cada uno de los píxeles que forman una imagen, y cuando esta asignación se realiza gracias al uso de métodos de DL, se denomina Segmentación Semántica.

Para asignar una clase a cada píxel el proceso que se sigue es fraccionando una imagen en diferentes regiones basándose en las características de los píxeles que componen la imagen. Esta división, puede llevarse a cabo mediante la detección de nuevas regiones gracias a los contornos, a los cuales se les atribuyen cambios abruptos en los valores de píxeles contiguos. La Segmentación de imágenes desempeña un papel crucial en esta etapa del análisis, permitiendo establecer una distinción semántica entre las diferentes componentes de una imagen.

La Segmentación se puede clasificar en tres grupos principales [27]: segmentación manual, segmentación semiautomática y segmentación completamente automática. En la segmentación manual, expertos etiquetan todos los componentes de la imagen pasando por cada uno de ellos. Aunque este proceso es lento, dependiente del experto y requiere una identificación precisa de las regiones de interés (ROI) para cada objeto, se utiliza ampliamente como una verdad terrenal o "*Ground Truth*" (GT) para evaluar los resultados de los modelos. La segmentación semiautomática implica la interacción del usuario para realizar tres tareas, inicialización (definir ROI), retroalimentación y evaluación. Las corrientes actuales de esta metodología semiautomática utilizan un enfoque de clasificación, sumando de este modo los inconvenientes de la clasificación. En este método, las regiones generalmente se clasifican (anotan) utilizando algoritmos de aprendizaje automático. Por otro lado, el método de segmentación completamente automática no requiere interacción del usuario. Debido a la diversidad de

formas y tamaños de las ROI en imágenes biomédicas, este tipo de segmentación presenta desafíos significativos.

### 3.8 Segmentación de instancias

A pesar de que tanto la Segmentación como la Detección pueden llegar a aportar resultados satisfactorios para distintos dominios de problemas, estos siempre acaban derivando en las mismas problemáticas, de esta forma se intentó combinar ambas técnicas para poder sacar el máximo rendimiento de las ventajas que cada uno de ellos presenta, a esta nueva técnica se la llamó Segmentación basada en Instancias.

Girshick et al. [28] fueron unos de los primeros en presentar este enfoque que combina la detección y la segmentación, en su estudio proponían usar un novedoso método que usaba la proposición de regiones en conjunción con una red CNN (denominaron a este método R-CNN). Los investigadores testearon la valía de su modelo comparándolo con un potente detector basado en una ventana deslizante y pudieron comprobar que, como era de esperar, sobrepasaba al detector con amplitud. No obstante, este nuevo modelo tenía sus desventajas, el entrenamiento consumía mucho tiempo y recursos, problema que, naturalmente, seguía presente en las etapas de prueba del modelo. Girshick et al. siguieron trabajando sobre el modelo R-CNN con el objetivo de mejorarlo, así surgió Fast R-CNN [29], que resultó más preciso y eficiente multiplicando por doscientos la velocidad de prueba de su antecesor.

Cheng et al. [30] presentaron un método de Segmentación de Instancias basado en cajas, para el cuál aunaron una red piramidal de características (FPN) y una U-Net, que demostró sobresalir en precisión y en velocidad. Aunque el enfoque de este estudio presenta un problema y es que el uso de *bounding boxes* en imágenes biológicas dificultan la diferenciación del objeto detectado con el de sus vecinos, es por eso, que se propuso un enfoque orientado a objetos [31], en el que la segmentación usa las características de los objetos como una guía para separar los objetos de sus adyacentes. Una gran parte de las investigaciones realizadas que aplican Segmentación de Instancias a imágenes microscópicas, se sirven de redes CNN en combinación con otras técnicas de detección, un ejemplo es el desarrollado para detectar células neuronales [32], que crean una red neuronal que combina un detector de una única pasada (SSD) con la red U-Net.

## 4 Técnicas de Visión Artificial aplicadas a la caracterización de las células

La inteligencia artificial (IA), es una disciplina de la ciencia de la computación cuya misión es crear sistemas que repliquen la inteligencia humana para realizar tareas, estos sistemas están conformados por combinaciones de algoritmos que tratan de imitar las capacidades cognitivas e intelectuales que posee el ser humano. El objetivo de la inteligencia artificial es desarrollar sistemas capaces de procesar grandes cantidades de datos, ver, comprender el lenguaje hablado o escrito, identificar patrones, aprender de la experiencia y/o actuar de forma autónoma, de forma precisa y eficiente, con resultados equiparables a los que reportaría la tarea al ser realizada por un humano [33], [34]. Los problemas que puede resolver la inteligencia artificial se agrupan en seis disciplinas[35], estas son: procesamiento del lenguaje natural, representación del conocimiento, razonamiento automatizado, aprendizaje automático, visión artificial y robótica.

Existen distintas categorías de técnicas de Inteligencia Artificial que se pueden clasificar según la retroalimentación que reciben para evaluar la validez de las nuevas hipótesis. Estas categorías incluyen el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje reforzado.

En el aprendizaje supervisado, se presenta al modelo una serie de pares de datos de entrada y salida, a partir de los cuales se infiere una función que relaciona ambos. Por ejemplo, en el contexto de la Visión Artificial, las imágenes se utilizan como datos de entrada, y las etiquetas de los objetos que se desean identificar se consideran como las salidas esperadas.

En cambio, en el aprendizaje no supervisado, el modelo solo recibe datos de entrada sin ningún tipo de retroalimentación explícita. En estos casos, los modelos a menudo emplean técnicas de agrupamiento para identificar patrones o similitudes entre los objetos presentes en las imágenes. El objetivo es agrupar objetos que puedan pertenecer a la misma clase o categoría.

Por último, en el aprendizaje reforzado, el modelo recibe *feedback* basado en la validez de las nuevas hipótesis que genera. Utilizando esta retroalimentación, el modelo realiza ajustes para mejorar sus predicciones en futuras situaciones. En esencia, el modelo aprende a través de la interacción con el entorno, recibiendo indicaciones sobre si sus acciones son correctas o no.

Cuando se trabaja con problemas tan específicos como el que se pretende resolver en este TFG, se suelen utilizar técnicas de aprendizaje supervisado. De esta manera, los expertos en

Bioquímicos involucrados en este proyecto definirán en los sets de imágenes de entrenamiento, que objetos son o no son una neurona viva. El proceso de entrenamiento les dará a los modelos generados el conocimiento necesario para hacer nuevas predicciones.

Por otro lado, los algoritmos de Inteligencia Artificial se pueden clasificar en función a la tipología de sus datos. Este TFG, en el que se analizará y procesarán imágenes, se corresponde con la rama de Visión Artificial (VA) o Visión por computadora, concretamente en la identificación y segmentación de objetos. La VA además es la rama de la IA que dota a las máquinas de la capacidad de ver, analizar y comprender imágenes, es decir, su misión es replicar la percepción visual humana para poder extraer de las imágenes información que pueda ser analizada. Mediante el uso de esta disciplina, se podrán detectar y analizar las imágenes de las neuronas en cultivo, y por tanto, estudiar su comportamiento, la VA no sólo puede tratar la detección de las neuronas, sino que, con el uso de las técnicas correctas, también permitiría realizar un seguimiento de estas a lo largo del tiempo.

#### 4.1 VA tradicional vs *Deep Learning*

La VA es un campo de estudio que busca dotar a las máquinas de la capacidad de interpretar y comprender imágenes o videos, de manera similar a como lo hace el ser humano.

La carga computacional siempre ha sido un factor limitante en la VA. Las imágenes son datos matriciales de varias dimensiones y con una gran cantidad de datos por dimensión. Esto dificultó, tradicionalmente, que los modelos de IA funcionales para otro tipo de datos se pudieran utilizar con imágenes. Para superar esta problemática, se introducía un paso previo al entrenamiento llamado “descripción” o “extracción de características”, donde se filtraba la imagen para quedarse con la información más relevante. Este paso, introducía una variable más en el correcto desarrollo del modelo, pues ese filtrado de características podía dejar información relevante fuera del modelo. Cuando, gracias al alto desarrollo computacional de las máquinas en los últimos años, es posible la utilización de modelos DL que se alimenten de imágenes completas, podemos hablar de una nueva época para la VA. Por ello, y a pesar de que la fase de descripción se sigue utilizando hoy en día para casos muy concretos, nos referiremos a ellos como tecnologías de VA tradicionales, para diferenciarlas de las que se basan en DL.

La VA tradicional, como se acaba de exponer, se basa en la descripción y extracción manual de características de los objetos presentes en las imágenes. En este enfoque, se requiere que los investigadores identifiquen y definan las características relevantes para la tarea específica que se desea realizar. Estas características pueden incluir información sobre la forma, el color, la

textura u otros aspectos visuales de los objetos. Una vez extraídas estas características, se utilizan algoritmos clásicos de *Machine Learning* (ML) para clasificar y reconocer los objetos en las imágenes.

Este enfoque tradicional ha sido utilizado con éxito en muchos problemas de VA. Sin embargo, presenta limitaciones importantes. La extracción manual de características puede ser un proceso laborioso y requiere de un conocimiento experto para identificar las características más relevantes. Además, este enfoque depende en gran medida de la calidad y adecuación de las características seleccionadas, lo que puede limitar su rendimiento en situaciones complejas o con gran variabilidad visual. En la Ilustración 1 se puede ver un esquema de las fases típicas de la generación de un modelo de Detección de Objetos mediante ML.

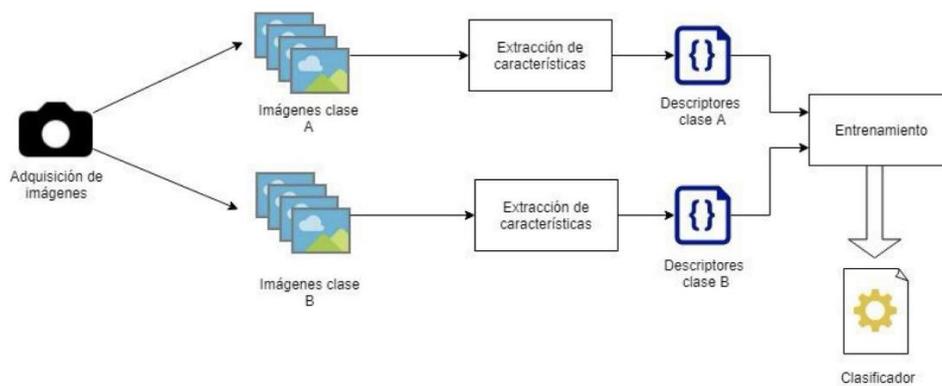


Ilustración 1: Esquema que muestra las distintas fases de ML

Por otro lado, el *Deep Learning* ha revolucionado el campo de la VA en los últimos años. Esta técnica se basa en el uso de redes neuronales profundas, que son capaces de aprender automáticamente las características y patrones de los datos de entrada, sin la necesidad de una extracción manual previa. En lugar de definir y seleccionar características, el DL se basa en el entrenamiento con grandes conjuntos de datos para aprender representaciones internas complejas. En la Ilustración 2 [36] se puede ver de forma esquemática las diferencias entre ambos procesos.

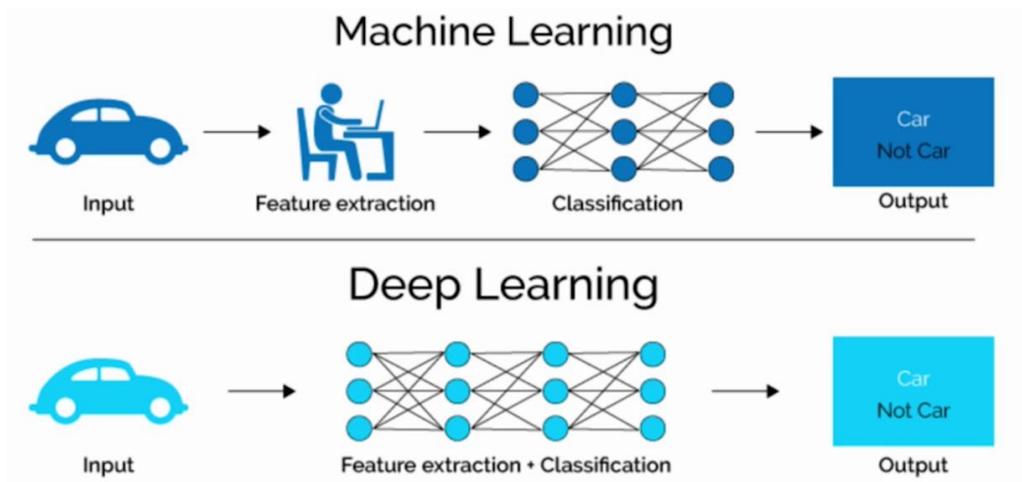


Ilustración 2: Diferencia entre ML y DL

La principal ventaja del DL es su capacidad para extraer características altamente discriminativas de las imágenes. Las redes neuronales profundas pueden aprender patrones sutiles y complejos en los datos, lo que les permite realizar tareas de reconocimiento de objetos con una precisión sorprendente. Además, el DL ha demostrado ser especialmente efectivo en problemas donde los datos de entrada son altamente complejos o presentan una gran variabilidad visual.

Sin embargo, el DL también presenta algunos desafíos. Requiere grandes conjuntos de datos para el entrenamiento, lo que puede ser costoso y requerir una gran capacidad de almacenamiento. Además, el entrenamiento de redes neuronales profundas es computacionalmente intensivo y requiere de hardware y recursos computacionales adecuados.

En situaciones en las que los datos de entrada presentan mayor complejidad, los modelos de DL demuestran una precisión destacada, ya que pueden extraer descriptores altamente discriminativos. Esta característica del DL resulta especialmente beneficiosa en el problema de identificación de neuronas, donde las imágenes de entrada pueden contener células que se asemejan a neuronas pero que pertenecen a otros tipos celulares. Gracias a los descriptores potentes que el DL es capaz de generar, se espera lograr una diferenciación muy precisa entre las células neuronales y los demás tipos de células.

#### 4.1.1 Redes neuronales

En los modelos de DL, las hipótesis conforman un complejo circuito de operaciones algebraicas con pesos ajustables entre ellas, el uso de la palabra “*deep*”, en inglés: “profundo”, se refiere a que esos circuitos están dispuestos en capas que conforman los diferentes caminos que pueden tomar cada una de las entradas.

A diferencia de los modelos tradicionales de ML, como la regresión, donde cada una de las entradas colabora de forma independiente a crear una salida, provocando que el camino a recorrer entre las entradas y la salida sea realmente corto, lo que únicamente permite representar problemas lineales. Mientras que, los modelos de DL pretenden crear caminos más largos con múltiples capas y más complejos con la relación de todas las entradas con cada una de las capas intermedias, estos caminos hacen que los modelos sean suficientemente completos como para representar la complejidad de los problemas del mundo real, no sólo los lineales. La diferencia entre estos modelos puede observarse en la Ilustración 3 [37]. Estos métodos de DL son también conocidos como redes neuronales, esto se debe a que su origen viene de tratar de imitar las conexiones neuronales presentes en el sistema nervioso.

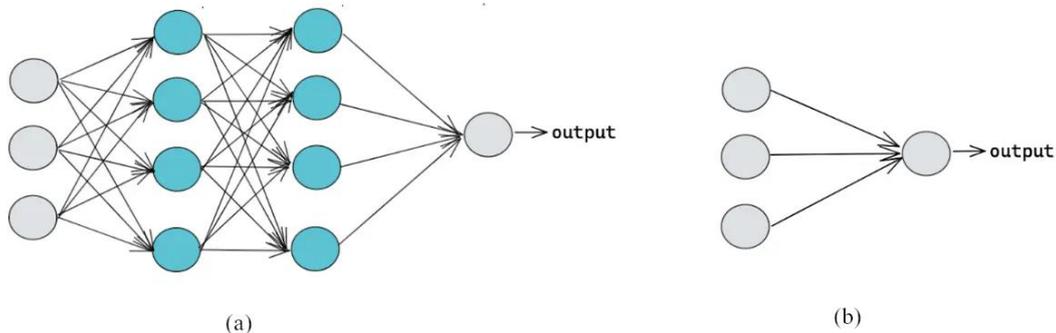


Ilustración 3: (a) Red neuronal de deep learning, (b) Modelo de regresión lineal

Una red neuronal está constituida por unas entradas, unas capas ocultas y una salida, las cuales a su vez están formadas por nodos que conectan cada capa con la siguiente, estos nodos se llaman neuronas o unidades. A cada una de las entradas que llegan a una unidad se les aplica un factor de peso, a continuación, se suman todas ellas y finalmente a ese resultado se le aplica una función de activación no lineal que tiene como resultado la salida de esa unidad. Es muy importante que la función de activación sea no lineal, ya que es esta la que le confiere la arbitrariedad necesaria para resolver problemas complejos y en este problema concreto será la que permita a la red adaptarse para identificar de forma precisa las células neuronales cultivadas.

La forma en la que se disponen y conectan los nodos de las diferentes capas, da lugar a diferentes tipos de redes neuronales. Cada tipo de red resuelve mejor diferentes problemas, ya que diferentes disposiciones de las unidades recogen de mejor forma la información de diferentes dominios.

Cuando el dominio de los datos de entrada son imágenes surgen algunas complicaciones sobre cómo alimentar la red neuronal. Por ejemplo, en una imagen RGB de  $N \times M$  píxeles, la información se dispone en una matriz  $N \times M$  de *arrays* de dimensión 3, en los cuales se almacenan los valores de rojo, verde y azul mediante un entero entre 0 y 255, si se considerase una entrada por cada uno de los colores (resultando un total de  $3 \times N \times M$  entradas a la red) se perderían tanto la pertenencia de cada valor del color a un mismo píxel como la adyacencia de los píxeles. Es por esto por lo que, normalmente, las redes neuronales que se usan en el tratamiento de imágenes siguen una topología similar a la de los *arrays*, concretamente, suelen usarse redes neuronales convolucionales, ya que la relación de las unidades permite mantener la información de los píxeles adyacentes.

#### 4.1.1.1 Redes neuronales convolucionales

Una red neuronal convolucional (CNN, del inglés “*Convolutional Neural Network*”) es aquella cuyas conexiones en las primeras capas ocultas son locales, para ser considerada una CNN la red debe tener al menos una capa de convolución. Una CNN se caracteriza por que los pesos que se aplican a las conexiones se replican en cada una de las secciones locales, cuando existe un patrón de pesos, a este se le denomina *kernel* y al proceso de aplicar este se le denomina convolución, de ahí el nombre de este tipo de redes.

Las capas de convolución tienen como objetivo extraer las características principales de las distintas regiones de la imagen, en estas capas el *kernel* actúa como filtro de extracción de características, en estas capas también se usa una función de actuación no lineal (habitualmente ReLU) con el fin de que las salidas de esta capa sean no lineales.

En las CNN, también se usan capas de *pooling*, estas capas tienen como misión reducir el tamaño de las imágenes, el *pooling* es un tipo de convolución pero con distinto propósito, además el *kernel* aplicado en las capas de convolución se aprende durante el entrenamiento de la red, mientras que el que se aplica en la capa de *pooling* es fijo. En la Ilustración 4 se puede observar la aplicación de dos pares de capas de convolución y *pooling*, seguidas de una etapa clásica de clasificación.

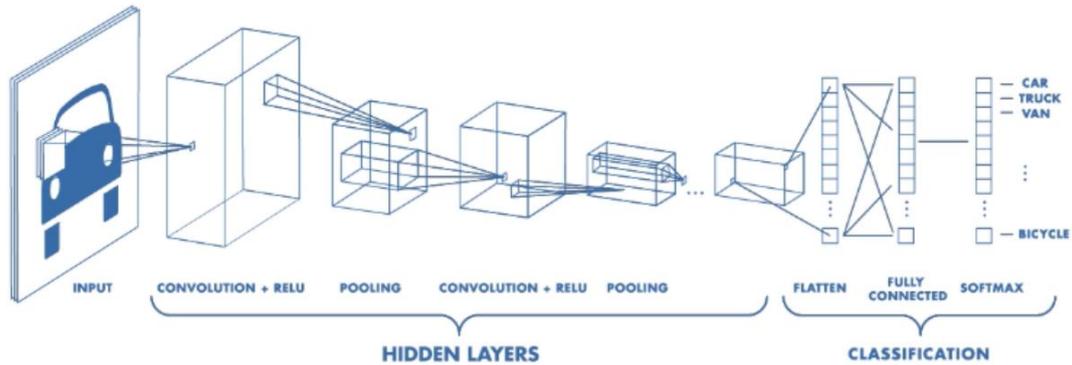


Ilustración 4: Esquema de una red neuronal convolucional

#### 4.1.1.2 Otras redes neuronales usadas en visión artificial

A pesar de que en este TFG no se utilizan, es necesario nombrar otros tipos de redes neuronales muy utilizadas en VA. Una de estas redes que pueden ser útiles en el análisis de secuencias de imágenes son las redes neuronales recursivas (RNN, del inglés “*Recurrent Neural Network*”), mientras que en el resto de las redes neuronales la dirección de las conexiones entre neuronas va de una capa a la siguiente, en las redes recursivas las neuronas pueden tener conexiones con ellas mismas, esto permite que la red tenga una memoria en forma de estado interno almacenado durante un entrenamiento anterior. La Ilustración 5 [38] contiene un esquema de una unidad recurrente. Debido a esta memoria, esta arquitectura da muy buenos resultados en *datasets* secuenciales, como imágenes tomadas en instantes temporales próximos, ya que esa memoria les permite comparar los distintos estados por los que van pasando los fotogramas. En este tipo de redes, la conexión recursiva da lugar a un problema que se denomina desvanecimiento del gradiente, las redes de memoria corta a largo plazo o *Long Short-Term Memory* (LSTM) permiten que la red tenga una memoria a largo plazo sin que la información se corrompa.

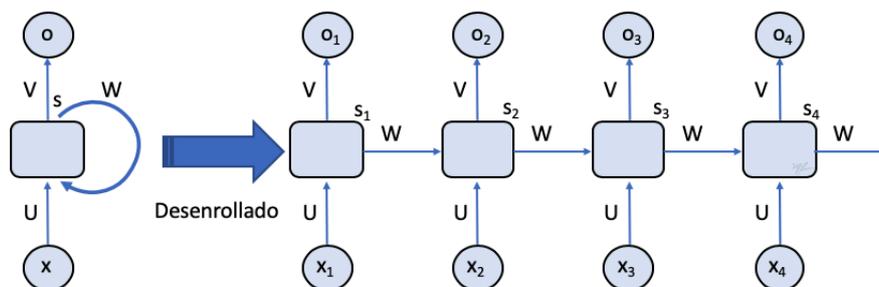


Ilustración 5: Unidad recurrente de una red con arquitectura RNN

Finalmente, las redes GAN (redes generativas adversas), son unas redes muy utilizadas en el campo de la VA, aunque mayoritariamente se usan para la creación y reconstrucción de imágenes, también han sido usadas en tareas de segmentación en conjunto con otras técnicas. Una red GAN, está constituida por dos redes que trabajan de forma conjunta: un generador y un discriminador. El generador crea nuevas imágenes tratando que sean lo más parecidas posibles a las imágenes reales de entrenamiento con el objetivo de confundir al discriminador para que no logre diferenciar entre la imagen real y la creada. A su vez, el discriminador se entrena tratando de diferenciar cuál es la imagen real y cual la resultante del generador, de esta forma el generador y el discriminador se mejoran el uno al otro en el entrenamiento. La Ilustración 6 muestra es esquema de composición de una red GAN.

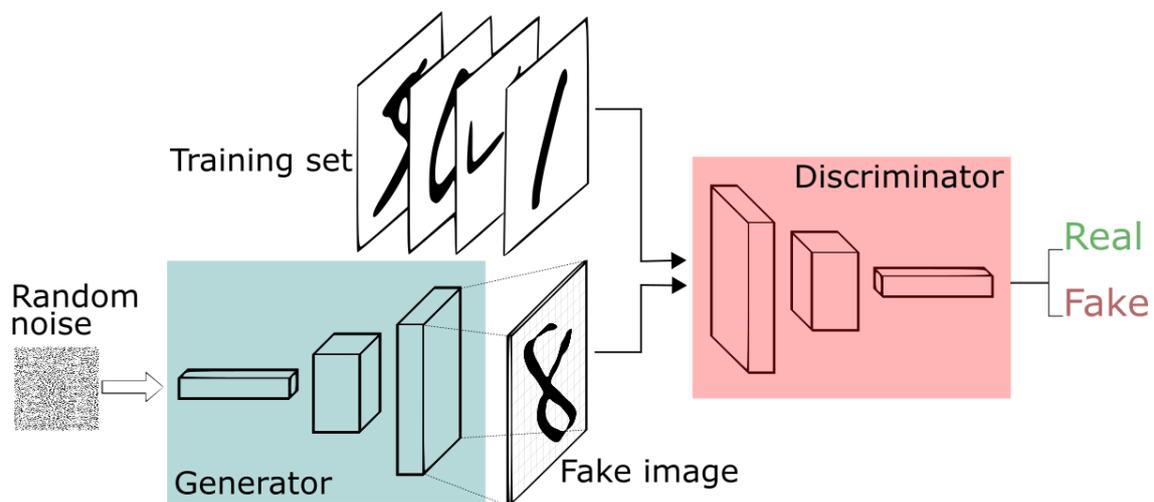


Ilustración 6: Esquema de la arquitectura de una red GAN [39]

## 4.2 Clasificación, identificación de objetos y segmentación

Los modelos de DL en Visión Artificial son capaces de aprender patrones y características complejas de las imágenes, lo que los convierte en herramientas poderosas para el análisis y la comprensión de datos visuales. Las técnicas de DL más utilizadas para el análisis de imágenes son la Clasificación de imágenes, Detección de objetos y Segmentación, tanto semántica como de instancias.

La clasificación es la capacidad de un modelo de IA de asignar una etiqueta como resultado de la predicción, clasificando la nueva muestra en una clase determinada. En VA, la clasificación permite asignar una categoría a una imagen. Cuando solamente tenemos una clase, como es el caso de este TFG, se podría identificar una imagen como la clase positiva “neurona” o como clase negativa.

La Detección de Objetos es un tipo de algoritmos que en última instancia utiliza la clasificación para determinar si una parte de la imagen se corresponde con alguna de las clases buscadas, pero esta clasificación se combina con otras técnicas de recorrido de la imagen, para en vez de clasificarla en su totalidad, buscar en ella los diferentes objetos que pueda contener. Esto significa que localiza y delimita mediante una casilla las distintas apariciones de objetos de una clase en una imagen, siendo capaz de distinguir las apariciones del objeto del resto de la imagen.

Por otro lado, se denomina segmentación a la técnica de clasificación a nivel de píxel, es decir, se le otorga una categoría a cada uno de los píxeles que conforman la imagen, si esta segmentación es estándar se conoce como Segmentación Semántica y si se busca reconocer grupos de píxeles con identidad propia se denomina Segmentación de Instancias.

En los próximos apartados se exponen en profundidad cada una de estas técnicas y se estudia las ventajas e inconvenientes que puede tener cada una de ellas para la resolución del problema de reconocimiento y análisis de células neuronales.

#### 4.2.1.1 Clasificación de imágenes

La Clasificación es una técnica que, aunque es muy usada dentro del DL, no es exclusiva de la VA. El objetivo de la clasificación es asignar una clase o etiqueta predefinida a una imagen de entrada con la finalidad de distinguir diferentes clases. Un ejemplo de clasificación se puede ver en la Ilustración 7 [48].

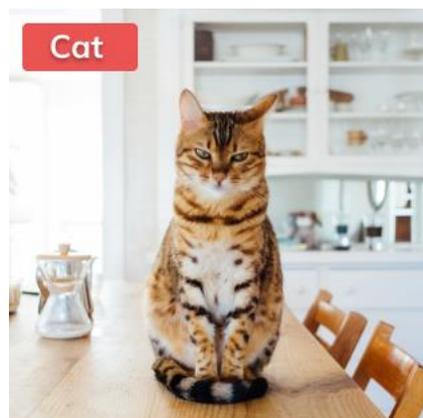


Ilustración 7: Clasificación de una imagen con una clase presente

Los sistemas modernos de Clasificación de imágenes se basan en la apariencia de los objetos, tomando como “apariencia” las características relacionadas con morfología, texturas y colores. Sin embargo, las imágenes cuentan con una serie de factores de los que surgen ciertas

problemáticas que complican la clasificación: la luz de una imagen puede hacer que el brillo y los colores de la imagen capturadas se vean alterados, el ángulo en el que se toma la imagen puede hacer que el objeto que se desean clasificar tenga una apariencia distinta a la real, la dirección en la que se toma la fotografía puede ocasionar que el objeto parezca otro que realmente no es, la oclusiones (por otros objetos o por el propio objeto) pueden hacer que características importantes del objeto a detectar no se observen en la imagen y por último, el movimiento del propio objeto, que puede ocasionar que este se vean distorsionado de su morfología habitual.

Las ya mencionadas redes convolucionales son excelentes clasificadores de imágenes, sus resultados son difíciles de imitar por otras soluciones, esto se debe a que las capas de convolución con funciones de activación ReLU actúan como un mecanismo de detección de patrones locales, ya que comprueban el grado de similitud entre el kernel y la imagen de entrada. A pesar de los buenos resultado que reporta la Clasificación de imágenes con redes CNN, no es útil en el problema de análisis de cultivos de neuronas ya que precisa de mucha más exactitud que discernir entre si la imagen contiene o no objetos de la clase neurona.

#### *4.2.1.2 Detección de Objetos*

La Detección de Objetos es una técnica que hace uso de la clasificación, pero en vez de clasificar como un objeto de una determinada clase a la imagen completa, el algoritmo selecciona subimágenes donde realiza una clasificación, y de esta forma es capaz de encontrar y localizar diferentes objetos en una misma imagen. En la Ilustración 8 [49] se puede observar un ejemplo de resultado de Detección de Objetos, la imagen analizada presenta tres objetos de tres clases distintas.

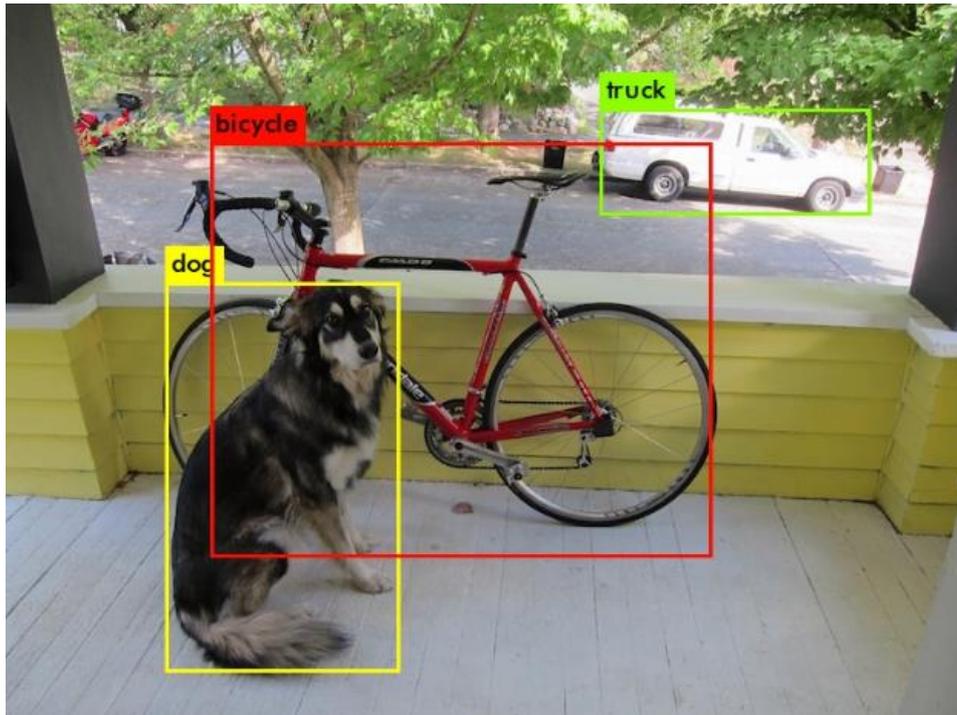


Ilustración 8: Detección de tres clases distintas de objetos

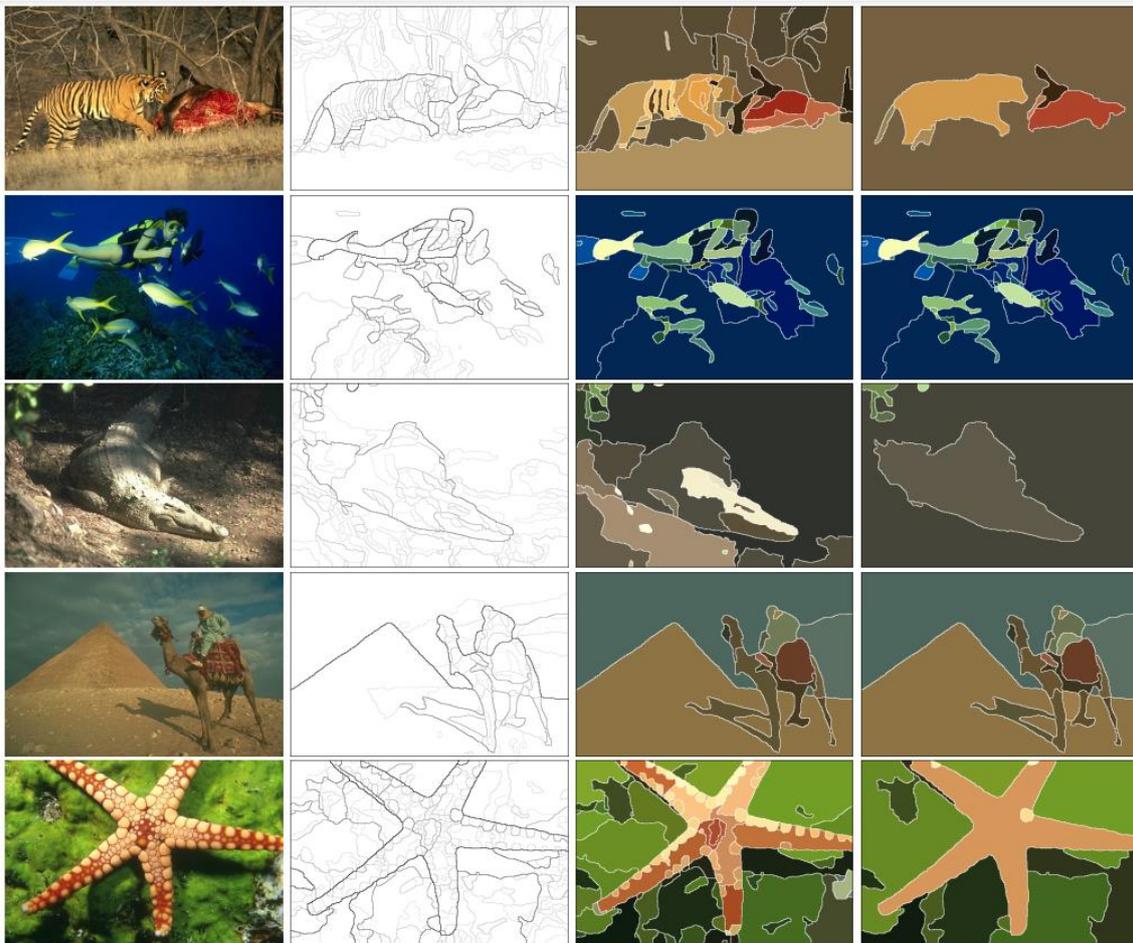
Para recorrer la imagen, se suele utilizar una ventana deslizante que recorre con un cierto paso o *step* (número de píxeles) toda la imagen para encontrar múltiples objetos dentro de una imagen. Normalmente este algoritmo de ventana deslizante se combina con algoritmo de pirámide, que por cada recorrido completo de la ventana deslizante, modifica secuencialmente el tamaño de la imagen (o de la ventana) para permitir la Detección de Objetos de diferente tamaño.

También es necesario que las técnicas de extracción de ventanas comentadas en párrafos anteriores se combinen con un algoritmo que descarte ventanas en las que se repitan un mismo objeto y se quede con la ventana que mejor captura el objeto identificado, este paso es usual que se resuelva mediante el uso de redes RPN (*Regional Proposal Network*). Para facilitar este paso es importante definir adecuadamente el paso que se va a utilizar para deslizar la ventana por la imagen, si el paso es muy pequeño, el contenido que abarca la ventana apenas varía, pero si se escoge un paso demasiado grande es posible que ninguna de las ventanas abarque el objeto completo sino que se tengas dos ventanas cada una con una mitad del objeto. Una vez se descartan las ventanas innecesarias, el detector sólo tiene las ventanas en las que potencialmente hay un objeto que coincida con una de las clases preestablecidas, todas aquellas ventanas en la que se detectan objetos son reportadas en forma de *bounding box* por el detector junto a su ubicación dentro de la imagen.

Esta técnica reporta muy buenos resultados cuantitativos en el problema de las neuronas y lidia bien con las oclusiones si se entrena correctamente, sin embargo, no permiten definir de forma clara los bordes de cada una de las neuronas encontradas, ya que sólo las identifica mediante las casillas del *bounding box*, también impide llevar registros en los cambios de la morfología de las neuronas

#### 4.2.1.3 Segmentación Semántica de imágenes

La Segmentación Semántica de imágenes, también conocida simplemente como Segmentación de imágenes, es una técnica de análisis de imágenes que consiste en dividir la imagen en secciones con la finalidad de asignar una categoría a cada uno de los píxeles que componen una imagen. Se basa en la idea de que características como el brillo, el color y la textura asociadas a los píxeles que componen un mismo objeto tienen valores significativamente similares, mientras que los bordes que separan estos objetos del resto de la imagen sí provocan grandes variaciones en esas mencionadas características. Aunque esta técnica se basa en el cambio de las características en los píxeles que comprenden un objeto respecto a los que conforman sus bordes, es importante no confundir la segmentación con la detección de bordes, de hecho la segmentación puede llevarse a cabo mediante la detección de bordes o mediante la detección de regiones. En la Ilustración 9 [40] puede observarse la diferencia entre una segmentación basada en detección de bordes y una basada en regiones, en la detección de regiones es posible definir el grado de diferencia que conforman regiones.



*Ilustración 9: De izquierda a derecha: imagen original; segmentación basada en bordes; segmentación basada en regiones con un umbral bajo; segmentación en regiones con un umbral menos exacto.*

El enfoque de segmentación de detectar bordes se puede abordar de dos maneras: la primera de una forma simple en la que se buscan cambios abruptos del gradiente entre los píxeles vecinos y una segunda que reporta mejores resultados, pero que también tiene mayor complejidad, esta solución se basa en que la curva de un borde localizada en un punto  $(x, y)$  tiene una orientación  $\theta$ , se traza entonces un disco con centro en  $(x, y)$  que se divide en dos con un diámetro con la misma orientación que  $\theta$ , de esas dos mitades del disco, se comprueban sus características y se saca la probabilidad  $P(x, y, \theta)$  de que el píxel  $(x, y)$  sea parte de un contorno. Aunque esta segunda solución consigue mejores resultados, sigue teniendo un problema: que los píxeles con una alta probabilidad de que sea un contorno no acaben formando una curva cerrada. Debido a todas estas complicaciones, nace el enfoque de detectar regiones, esta solución agrupa los píxeles vecinos con características similares, para que esta solución dé buenos resultados es necesario que el algoritmo sepa qué tipo de objetos se puede encontrar, ya que de lo contrario podría omitir regiones, si esto no fuese posible, habría que realizar una sobresegmentación de la imagen, la cual podría introducir falsos positivos.

Como tanto el enfoque de detectar bordes como el de detectar regiones tienen ventajas e inconvenientes, es usual que ambos enfoques se combinen, dando así lugar a una segmentación bastante precisa y que, en combinación con otras técnicas de seguimiento de objetos, logra resultados satisfactorios para el tracking de células. Sin embargo, la Segmentación Semántica no da buenos resultados en problemas que tratan con superposiciones, como es el caso del análisis de neuronas. Esto se debe a dos factores, por un lado, la segmentación requiere de los detalles de los objetos para identificarlos por lo que podrían clasificar de forma errónea los píxeles que los componen, por otro lado, la Segmentación Semántica no trata los objetos de una misma clase de forma individual y daría como resultado que dos neuronas juntas fuesen difíciles de tratar, ya que el conjunto de ambas sería etiquetado como neurona, este último factor además impide el análisis cuantitativo de células neuronales y habría que combinar la Segmentación Semántica con alguna otra técnica para poder realizar un conteo.

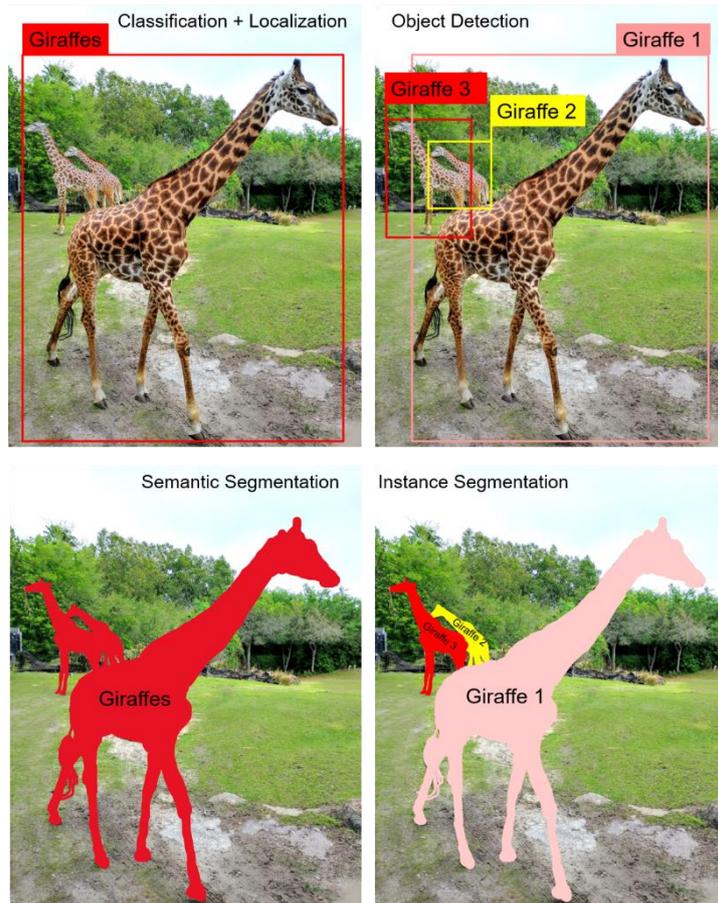
#### 4.2.1.4 Segmentación basada en instancias

La segmentación basada en instancias es una evolución de la Segmentación Semántica en la cual entra en juego la identidad de cada objeto reconocido. De esta forma, a nivel de resultados, se obtendrían la misma información que de una Segmentación Semántica junto con una Detección de Objetos. Pero esto no quiere decir que sea una mezcla de ambas técnicas, pues la arquitectura necesaria en una Segmentación de Instancias es bastante más compleja, teniendo especial relevancia la segmentación a nivel de píxel en la determinación de la identidad de las diferentes instancias.

Esta arquitectura se compone por una red neuronal Mask R-CNN, que es un subtipo de red CNN, que tiene primero una red CNN que extra las características principales de las imágenes, después estas pasan por una red RPN que identifica las regiones candidatas a contener objetos y finalmente una tercera red de segmentación produce una máscara binaria para cada objeto con la información de las dos redes anteriores para juntarlas todas en una máscara de segmentación.

En esta técnica cada objeto se localiza mediante un *bounding box* (como en Detección de Objetos) y dentro de la casilla se detectan los bordes del objeto. Se llama “basada en instancias” debido a que, a diferencia de la Segmentación Semántica, cada nuevo objeto detectado de una clase se etiqueta de forma independiente al resto de objetos detectados de esa clase. Es decir, cada objeto de una clase se trata como una nueva instancia de ella, instancia la cuál viene acompañada de la posición en la que se localiza el objeto detectado [41] e información relacionada con el tamaño y la morfología del objeto concreto identificado [42]. La Ilustración

10 permite visualizar de manera clara la diferencia entre todas las técnicas por las que se ha pasado hasta el momento.



*Ilustración 10: (de izquierda a derecha y de arriba a abajo) Clasificación de imágenes, Detección de Objetos, Segmentación Semántica, Segmentación de Instancias*

La Segmentación de Instancias es una técnica que por sí sola, ofrecería toda la información que se precisa obtener sobre las neuronas en un determinado momento del cultivo: forma e identidad individual.

Aun así, es importante señalar que a nivel de reconocimiento de instancias -información muy importante para investigación, pues de ella se podrá extraer el conteo neuronal-, es una completa incógnita si un modelo de Segmentación de Instancias va a superar un modelo de Detección de Objetos. El mejor rendimiento de uno u otro depende de la complejidad del escenario, y de si la información a nivel de píxel puede ayudar o complicar la identificación de instancias. Por ello, aunque esta técnica ofrezca toda la información necesaria, y dada la relevancia que tiene el conteo de neuronas en nuestra investigación, se realizará en este TFG un

exhaustivo experimento que tratará de determinar la técnica más precisa para la obtención de dicha información en este tipo de cultivos complejos.

## 5 Experimentación

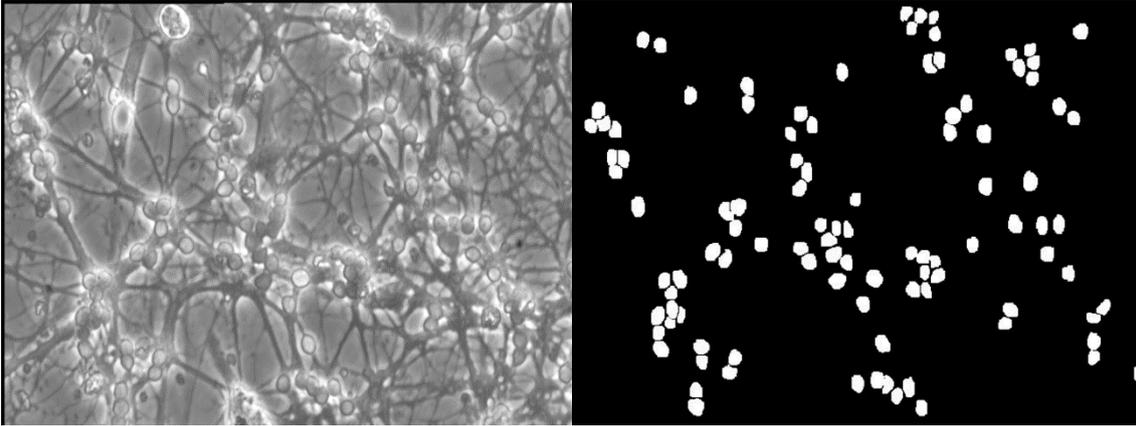
Como ya se ha venido mencionando, los cultivos que se desean estudiar presentan múltiples retos: la presencia de otros elementos, como neuritas, células gliales o restos de todas estas, dificultan la identificación de las células de interés, las neuronas, además están varían en forma, tamaño y posición a medida que se va desarrollando el cultivo.

En las próximas subsecciones, se explicarán las metodologías usadas para enfrentar los múltiples retos que supone el análisis visual de las células neuronales. Se incluye una descripción del dataset que ha sido usado para entrenar y probar el modelo, se detalla también cuáles han sido los materiales necesarios para obtener las imágenes de los cultivos que forman dicho dataset. Se continúa con una exposición del modelo utilizado para realizar el entrenamiento, así como las métricas que se han usado para cuantificar la calidad de este. Por último se expondrán todos los procedimientos realizados en el experimento como las actividades de preprocesamiento y el entrenamiento realizado.

### 5.1 Dataset

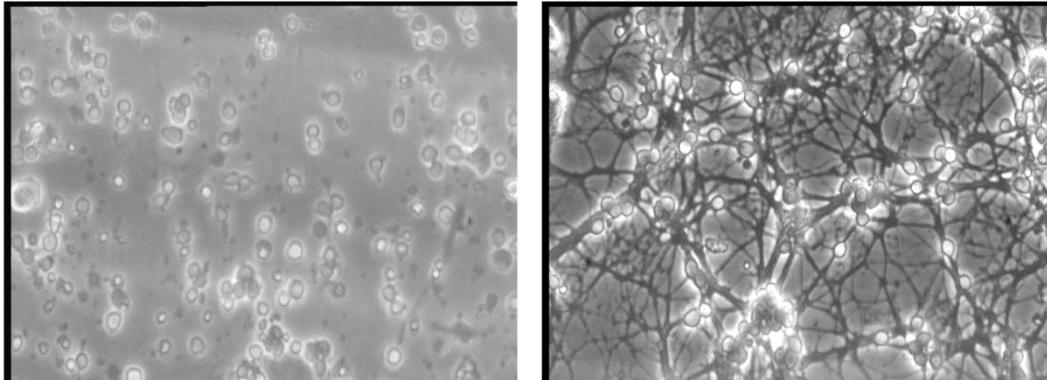
En este estudio se han usado 292 imágenes tomadas de cultivos primarios de neuronas de cerebelo de rata, de estas imágenes 205 fueron usadas para el entrenamiento, 58 para validación y 29 para test. Estas imágenes fueron capturadas en distintos estadios de cultivo haciendo uso de un microscopio convencional de contraste de fases.

Cada una de estas imágenes fueron debidamente anotadas manualmente por expertos en la materia, por lo que la calidad de las muestras está garantizada. Las anotaciones son en formato máscara, en las cuales sobre un fondo negro, se identifica la presencia de neuronas con color blanco, la media de neuronas por imagen es de 50. Puede verse la correspondencia de una imagen con su máscara en la Ilustración 11.



*Ilustración 11: Imagen de las neuronas cultivadas (izquierda) y máscara correspondiente a la imagen original (derecha)*

La anotación manual es una tarea desafiante debido a la complejidad y la gran variedad en la morfología de las neuronas, es por ello por lo que se requiere de personal con amplia experiencia para poder realizar la identificación y segmentación de las neuronas de una forma correcta, la adecuación de las anotaciones a la realidad es vital para la investigación. Con el fin de obtener una buena precisión en las métricas de los modelos evaluados y, en consecuencia, poder sacar conclusiones que se ajusten a la realidad, es necesario que las anotaciones sean correctas.



*Ilustración 12: Neuronas en una etapa temprana (izquierda); Neuronas en estado avanzado de cultivo (derecha)*

Este dataset fue seleccionado de forma premeditada para asegurar que representa de forma correcta las distintas fases de desarrollo de las neuronas cultivadas, las imágenes fueron tomadas en diferentes momentos del crecimiento desde las etapas más tempranas hasta su madurez completa. En cada una de estas etapas el tamaño y la apariencia de las neuronas es distinto, al inicio del desarrollo las neuronas pasan por cambios físicos de forma rápida y según va avanzando esta velocidad de cambio se va ralentizando, además a medida que estas se desarrollan van creando conexiones entre ellas, puede observarse en la Ilustración 12. Este amplio

abánico de imágenes ha permitido evaluar los algoritmos de segmentación aplicados para todas las etapas de maduración por las que puede pasar el cultivo y, por tanto, del cual se pueden extraer imágenes que se deseen estudiar con los modelos propuestos.

## 5.2 Materiales

Para realizar la observación de los cultivos de neuronas se usó un microscopio de contraste de fases, concretamente el microscopio Olympus IMT-2, y para capturar las imágenes se requirió de una cámara fotográfica, el dispositivo utilizado fue una cámara de vídeo en blanco y negro, de la marca Sony, el modelo era el SPT-M308CE.

## 5.3 Métodos

En este experimento se ha usado dos arquitecturas YOLO para entrenar dos modelos diferentes. En el modelo de Detección de Objetos se ha decidido usar la versión 5 de YOLO ya que actualmente más utilizada, y por tanto la más comprobada, para ese fin. Por otra parte, en el modelo de Segmentación de Instancias, se ha usado la versión 7 de YOLO, esta versión es muy similar a la 5, pero está optimizada para realizar segmentaciones basadas en instancias. A pesar de que YOLO tenga diferentes versiones que se adapten mejor a determinadas técnicas de identificación, su arquitectura no varía.

### 5.3.1 Arquitectura YOLO

YOLO (*You Only Look Once*) es una arquitectura para Detección de Objetos, es ampliamente usada en sistemas de detección mediante vídeo ya que su gran rapidez permite su uso en tiempo real. Esta arquitectura toma la forma de una red neuronal totalmente convolucional (FCN), esta está formada por 24 capas, en las cuales se intercalan capas convolucionales con capas de pooling, seguidas de 2 capas densamente conectadas, siendo la última capa formada por neuronas selectoras y neuronas clasificadoras, todas las capas aplican una función de activación *leaky ReLU*, a excepción de la última capa que usa una ReLU. La función de activación *leaky ReLU* rectifica los valores negativos, mientras que aquellos que son positivos no se les aplica ninguna operación. Un esquema de la red puede encontrarse en la Ilustración 13 [43].

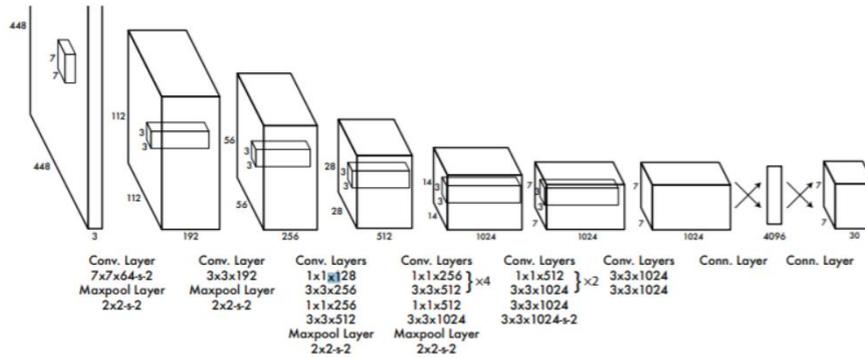


Ilustración 13: Arquitectura de la red que usa YOLO

Esta arquitectura fue creada inicialmente por Joseph Redmond en 2015, este se inspiró en el modelo GoogleNet, que se usa en Clasificación de imágenes. Este autor fue el creador de todas las versiones hasta la v3, posteriormente Alexey Bochkovsky creó la versión 4 y, de la versión 5 en adelante, todas ellas están atribuidas a Glenn Jocher y su empresa Ultralytics.

El funcionamiento de esta arquitectura es sencillo, primero se realiza un preprocesamiento de la imagen, en este la imagen se divide con una cuadrícula de  $S \times S$  casillas, a cada una de estas casillas se le aplica una extracción de características y con ello cada casilla hace una predicción de los bounding boxes que contiene. La predicción da como resultado un total de  $B$  bounding boxes o ventanas, a las cuales se les asocia una confianza, la cual indica la fiabilidad de que la celda contenga un objeto, así como la precisión de la detección, esta precisión se obtiene mediante el cálculo de la IOU (*Intersection Over Union*). La confianza de la predicción se calcula con la fórmula indicada en la ecuación siguiente, donde  $P(O)$  indica la probabilidad de que la ventana contenga un objeto de una determinada clase [44].

$$\text{Confianza} = P(O) * IOU$$

Cada una de las  $B$  ventanas predicen 5 valores:  $x$ ,  $y$ ,  $w$ ,  $h$ , además de la confianza ya mencionada. El par  $(x, y)$  es el centro asociado a la ventana, mientras que  $w$  es el ancho de la ventana y  $h$  es la altura de esta. Las coordenadas que indican al centro se normalizan respecto al tamaño de la región en la que se encuentra, mientras que, tanto la altura como el ancho se normalizan respecto a las medidas de la imagen al completo. Este proceso se puede observar en la Ilustración 14.

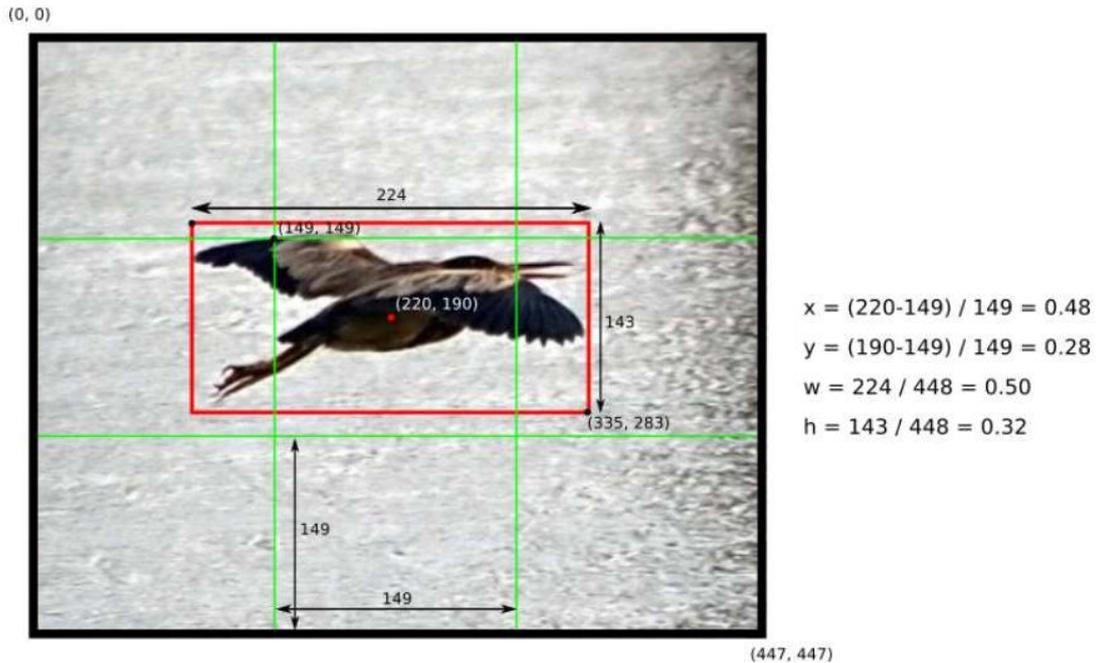


Ilustración 14: Ejemplo de cálculo y normalización de las predicciones dentro de una región en YOLO

Si un objeto ocupa más de una de las regiones en las que está dividida la imagen, habrá más de una detección de un mismo objeto, es por ello por lo que cuando sucede esto, al finalizar la etapa de detecciones se aplica una supresión de no máximos, es decir el sistema se quedará con aquella detección en la que el valor del IOU sea mayor y descartará el resto de las detecciones del objeto.

Como ya se ha comentado el uso principal de YOLO es la Detección de Objetos, pero a lo largo de las versiones se han ido integrando otros métodos de análisis de imágenes, en el caso de la versión 5, hay 4 subversiones distintas: la original que utiliza Detección de Objetos, una que usa bounding boxes con diferentes orientaciones para detectar objetos con diferentes ángulos de posición, versión de Segmentación de Instancias y, por último, versión de Clasificación. En siguientes versiones, como han sido la 7 y la 8, sólo se han desarrollado la versión original de Detección de Objetos y la de segmentación basada en instancias [45].

## 5.4 Métricas

El propósito de este TFG es valorar un modelo de Segmentación basada en Instancias para la tarea de identificación de neuronas en imágenes tomadas de cultivos, como ya se ha expuesto, la Segmentación de Instancias es una extensión de la Detección de Objetos, debido a ello los resultados de entrenar un modelo de este tipo se cuantificarán con métricas que evalúan la

detección. Con el fin de identificar cuál es mejor, se compararán sus resultados con los obtenidos por un modelo de Detección de Objetos, ambos independientes y para los cuales se han obtenido las métricas a estudiar.

Para evaluar los dos modelos se utilizarán tres métricas: la precisión, *recall* o exhaustividad y mAP. Para poder explicar estas métricas, previamente se expondrán algunos conceptos importantes para comprender las métricas que evaluarán el experimento.

Primero es necesario saber qué significado tienen los conceptos TP, FP, FN y FP; se explicarán los conceptos en base al dominio en el que se está trabajando, que tiene una sola clase: neurona. Estos cuatro conceptos clasifican si una predicción está bien hecha o no en función de la clasificación real (*Ground Truth* - GT) que debería de tener asignada.

- *True Positive* (TP): se predice que hay una neurona y, efectivamente, la hay.
- *False Positive* (FP): se predice que hay una neurona presente y, sin embargo, no la hay.
- *False Negative* (FN): no se predice que haya una neurona y, sin embargo, la hay.
- *True Negative* (TN): se predice que no hay una neurona presente y, efectivamente, no hay.

Esta clasificación de las predicciones realizadas por el modelo se trata diferente en Detección de Objetos y en Segmentación Semántica. En el caso de la segmentación, se considera la adecuación de la predicción a nivel de píxel, es decir, si el píxel tiene asignada la misma clase en la predicción que en la GT. Sin embargo, para la Detección de Objetos, se define un umbral (habitualmente se usa 0,5) a partir del cual, si el bounding box predicho coincide con el de la GT, entonces la predicción es correcta [46]. La Ilustración 15 contiene un ejemplo de clasificación de las predicciones para la Segmentación y la Ilustración 16 un ejemplo para la Detección.



Ilustración 15: Clasificación de una predicción en segmentación

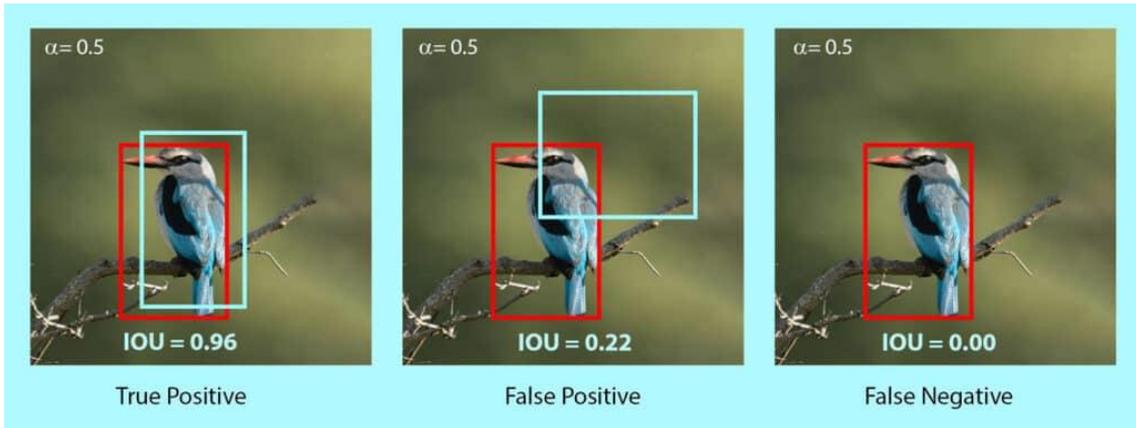


Ilustración 16: Clasificación de una predicción para Detección de Objetos con un umbral de 0,5

Otro concepto importante es el IOU (Intersection Over Union), esta métrica define el porcentaje del área de solapamiento entre la predicción y lo esperado, respecto al área de la unión de ambas. Para el caso de detección, se refiere al área que encierran los bounding boxes y, para la segmentación, se toma el área que encierra los contornos del objeto. Este concepto se entiende más fácilmente de forma visual, como se muestra en la Ilustración 17.

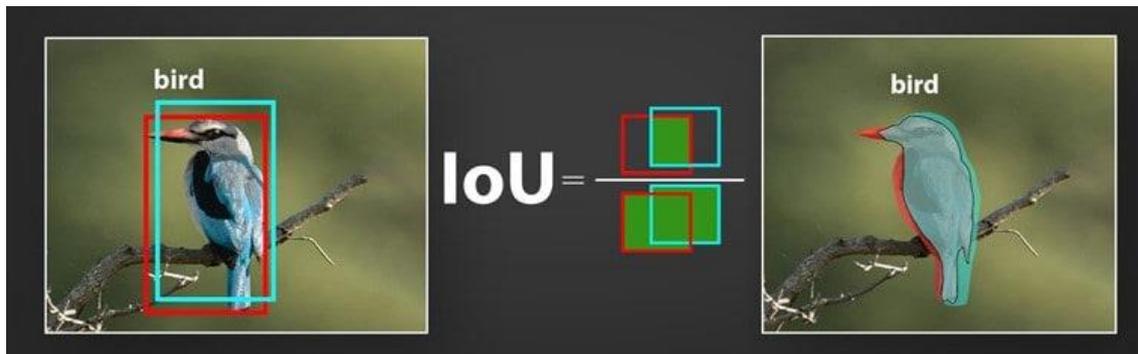


Ilustración 17: Cálculo de IOU en el caso de Detección de Objetos (izquierda) y en el caso de Segmentación Semántica (derecha)

Una vez aclarada la terminología necesaria se pueden explicar con mayor facilidad las métricas que se usarán para estudiar el modelo usado en el experimento.

#### 5.4.1 Precisión

La precisión indica la cantidad de predicciones positivas que han sido correctas y se calcula como la relación entre la cantidad de verdaderos positivos predichos entre todas las predicciones que han sido realizadas.

$$Precisión = \frac{TP}{TP + FP}$$

#### 5.4.2 Recall

El recall o exhaustividad muestra la cantidad de positivos que fueron correctamente predichos, esta métrica se calcula mediante la relación entre los verdaderos positivos entre el total que realmente debería de haber sido detectado.

$$Recall = \frac{TP}{TP + FN}$$

#### 5.4.3 mAP (mean Average Precision)

Esta métrica mide el área encerrada bajo la curva de la gráfica que enfrenta precisión y exhaustividad para cada una de las predicciones realizadas, la Ilustración 18 contiene un ejemplo del área que encierra dicha curva.

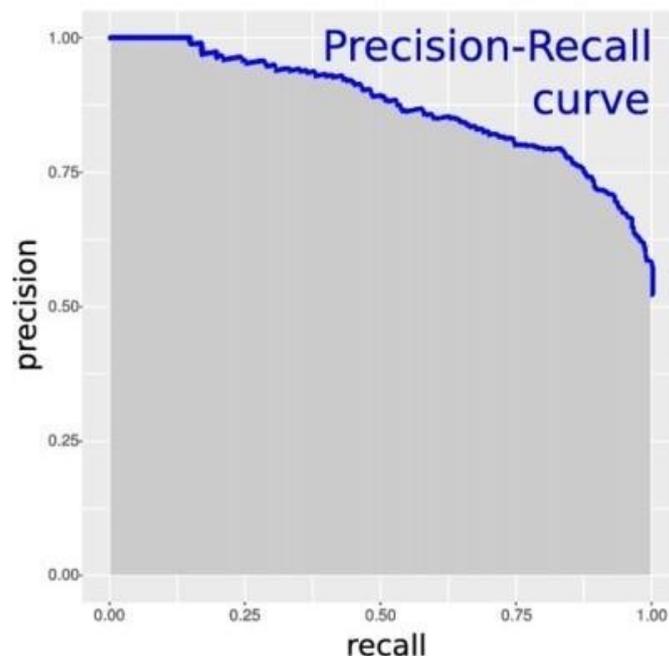


Ilustración 18: Área bajo la curva que se usa para calcular el mAP [47]

Existen dos variantes de esta métrica: mAP@0.5 y mAP@[0.5:0.05:0.95]. La primera de ellas se calcula simplemente con el área que encierra la curva ya mencionada, habiendo tomado para la consideración de TP un umbral de IOU=0.5. Sin embargo, la segunda métrica, hace la media de los mAP, tomando como referencia un IOU desde 0.5 hasta 0.95, con un paso de 0.05; es decir, se calcula el mAP para esos 10 umbrales diferentes de IOU y se hace la media de ellos.

La diferencia de esta medida respecto a otras parecidas, como AUC, radica en que, para problemas en los que el problema presenta más de una clase, el mAP se calcula haciendo la media de cada una de las curvas que se dibujaría para cada una de las clases.

## 5.5 Experimento

En el experimento que se aborda en este TFG, se ha realizado un entrenamiento de dos modelos: uno de Segmentación basada en Instancias y otro de Detección de Objetos, con la finalidad de enfrentar los resultados obtenidos por cada uno de los modelos, en ambos se han utilizado las imágenes de cultivo del dataset descrito con anterioridad, las cuáles han tenido que pasar un preprocesamiento para asegurar el rendimiento de los modelos.

Para cada uno de los modelos entrenados, y debido a la naturaleza de cada modelo, se han tenido que utilizar diferentes versiones de YOLO. Cada una de estas versiones requiere del uso de anotaciones en formatos distintos, por lo que se ha tenido que realizar un proceso de convertir las máscaras en las anotaciones que necesita cada uno de los modelos.

Conviene enfatizar que todos los algoritmos usados en este experimento han sido desarrollados en el lenguaje de programación *Python*, este lenguaje es uno de los más extendidos en ML.

### 5.5.1 Preprocesamiento de las imágenes

Debido a las complicaciones que presenta el dataset, ha sido necesario realizar un preprocesamiento de las imágenes para poder asegurar que la calidad de las imágenes de entrada no afectase al rendimiento de los modelos.

A las imágenes originales, se les han aplicado unos filtros con el fin de mejorar el contraste, facilitando así que los modelos detecten los bordes de las neuronas, primeramente, se ha aplicado un filtro gaussiano, cuya finalidad es reducir el ruido de toda la imagen, por último, se ha aplicado un filtro Sobel, enfatiza los bordes de las neuronas.

En las máscaras, ha sido necesario tratar el problema de superposición, ya que las neuronas se mostraban agrupadas unas sobre otras sin límites claros entre ellas, el uso de las máscaras sin este preprocesamiento sería insatisfactorio ya que no muestra los contornos de las células y, por tanto, sería imposible para los modelos aprender los contornos de las células en estos casos de solapamiento. Para solucionar este problema se ha aplicado un algoritmo de Voronoi-Otsu, que es capaz de colorear en diferentes tonalidades las neuronas a partir de las diferencias de intensidad entre las células, a pesar de que estas estén superpuestas, puede observarse en la

Ilustración 19. Después, otro algoritmo recorre todos los píxeles de la imagen en busca de pares de píxeles que no pertenecen al fondo (negro) cuyo color sea distinto entre sí, cuando encuentra estos casos establece un único píxel a negro. De esta forma, se logra crear entre las neuronas solapadas una separación de un píxel de ancho, suficiente para que los modelos puedan obtener información de los bordes de las células.

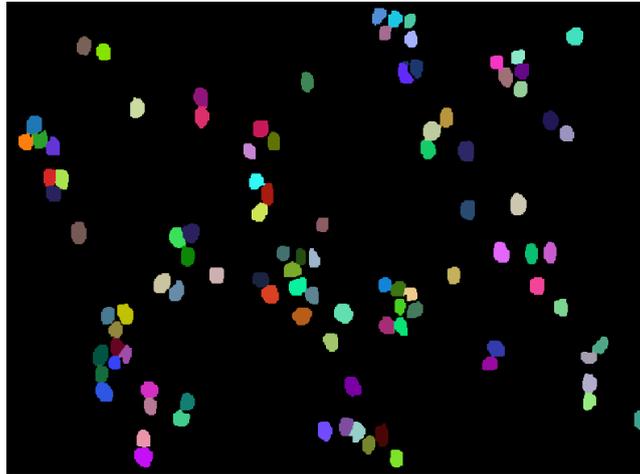


Ilustración 19: Aplicación del algoritmo Voronoi-Otsu a una máscara con neuronas superpuestas

### 5.5.2 Obtención de las anotaciones

Ya se ha comentado que las anotaciones que contiene el dataset son máscaras binarias, sin embargo, YOLO requiere de unas anotaciones específicas. Para la versión 5 se necesita un archivo de anotaciones por cada una de las imágenes, en este fichero se especifica la posición de una neurona por línea, siguiendo el formato de un *bounding box*: “clase coord\_x\_centro coord\_y\_centro ancho alto”. En cambio, la versión 7 de YOLO, tan solo admite anotaciones en formato COCO, esto es, un único fichero *JSON* que contiene todas las anotaciones específicas de cada una de las máscaras, las anotaciones de cada neurona especifican la clase a la que pertenece el objeto, el bounding box que los contiene y las coordenadas que identifican el perímetro de cada objeto. Para poder obtener los dos tipos de anotaciones, ha sido necesario aplicar dos algoritmos (uno para las anotaciones de YOLOv5 y otro para las de YOLOv7) a todo el conjunto de máscaras, el cual a partir de los contornos de cada neurona ubicada en la máscara, extrae las notaciones correspondientes.

### 5.5.3 Entrenamiento de un modelo de Segmentación de Instancias

Para llevar a cabo el entrenamiento del modelo que identificará las neuronas mediante Segmentación de Instancias (YOLOv7) se ha utilizado el dataset anteriormente descrito junto

con las anotaciones COCO recién mencionadas, para llevar a cabo la ejecución del entrenamiento se ha usado *Google Colaboratory*, una herramienta de *Google* para ejecutar código *Python* en línea. Tras lanzar varios entrenamientos, se ha estimado que el número óptimo de *epochs* necesarias para obtener la mayor calidad del modelo era de 65.

#### 5.5.4 Entrenamiento de un modelo de Detección de Objetos

El entrenamiento del modelo de Detección de Objetos se ha realizado con YOLOv5, este ha usado como imágenes de entrada las mismas que para el modelo previo y las anotaciones en formato YOLOv5. En este caso se han estimado necesarias 200 *epochs* para obtener las mejores métricas posibles.

## 6 Resultados

Tras realizar el entrenamiento de ambos modelos, se obtienen las métricas explicadas en el apartado “[Métricas](#)”. El modelo de Segmentación basada en Instancias reporta resultados tanto de segmentación como de detección, pero en la comparación con el modelo de Detección de Objetos, se utilizarán tan sólo las métricas de obtención de bounding boxes.

Los resultados de segmentación nos indica la capacidad del modelo para obtener los contornos de las neuronas, es decir, cuánto de eficaz es acotando cada neurona respecto al resto de la imagen. A nivel de segmentación, el modelo de Segmentación de Instancias ofrece los resultados expuestos en la Tabla 1.

*Tabla 1: Métricas obtenidas por el modelo de Segmentación de Instancias para la segmentación*

	Precisión	Recall	mAP@0.5
Segmentación de Instancias (Segmentación)	0,909	0,8895	0,9338

Las métricas que se han recogido suelen considerarse buenas cuando su valor está por encima de 0,5 y pasan a considerarse muy buenas cuando superan el 0,7. Por tanto, se puede decir que las métricas recogidas del proceso de segmentación son altamente satisfactorias.

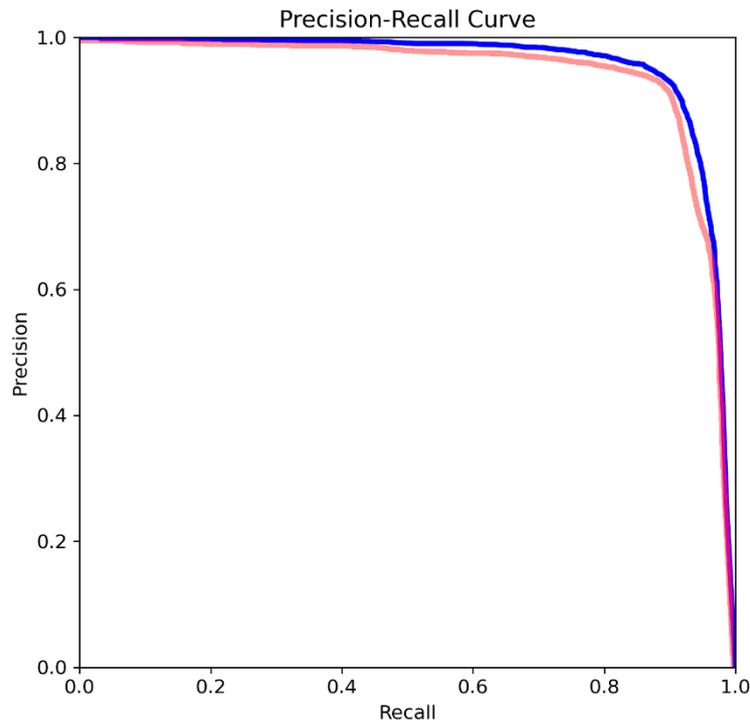
Con el objetivo de realizar la comparativa a nivel de identificación de las instancias celulares, a continuación, se compararán los resultados del mismo modelo de segmentación con los del modelo de Detección de Objetos, para el caso de la obtención de los *bounding boxes*.

*Tabla 2: Métricas obtenidas por cada modelo entrenado para la localización de bounding boxes*

	Precisión	Recall	mAP@0.5
Detección de Objetos	0,9217	0,8950	0,9451
Segmentación de Instancias ( <i>BBox</i> )	0,9236	0,9108	0,9559

En la Tabla 2, se puede observar las métricas obtenidas por cada uno de los modelos entrenados. La precisión que, como ya se había comentado, determina el número de predicciones correctas respecto a todas las predicciones realizadas ha resultado ser un 0,206% superior para el modelo de Segmentación de Instancias. En el recall es posible ver una mejoría más clara, siendo la

Segmentación de Instancias un 1,77% mayor que en el modelo de Detección de Objetos. Por último, en el mAP, la mejoría de usar segmentación basada en instancias respecto a Detección de Objetos se muestra un 1,144% mejor. Hay que tener también en cuenta, que el modelo de Detección de Objetos necesitó un total de 200 epochs para alcanzar estos resultados, mientras que el modelo de Segmentación de Instancias logró los resultados obtenidos en la mitad de epochs.



*Ilustración 20: Curvas de Precision-Recall de los modelos de Detección de Objetos (rojo) y Segmentación de Instancias (azul)*

En la Ilustración 20 es posible observar la diferencia entre las gráficas de Precisión-Recall de los dos modelos entrenados. Si el modelo fuese perfecto, la curva debería de dibujar el vértice superior izquierdo. Por ello, a la hora de evaluar un modelo, se considera que cuanto más se acerca la curva a dicho vértice, mejor es este. En la comparativa entre ambas puede observarse que la curva que representa al modelo de Segmentación de Instancias está más cercana del vértice, además puede verse que esta curva es más continua respecto a la del modelo de Detección de objetos que se ve más errática. En base a esta curva, es posible decir que el modelo de Segmentación de Instancias es ligeramente superior al modelo de Detección de Objetos.

En cuanto a la tasa de positivos verdaderos, es posible decir, a partir de las matrices de confusión de la Ilustración 21 que el modelo de Detección de Objetos reconoce el 94% de las neuronas

presentes en las imágenes, mientras que el modelo de Segmentación de Instancias tiene una tasa un poco superior, reconociendo el 95% de las neuronas reales. En lo que respecta al reconocimiento del fondo, ambos modelos son capaces de reconocer al 100% todo aquello que no son neuronas. A partir, de las matrices de confusión, podemos decir que ambos modelos son igual de eficientes reconociendo el fondo, sin embargo, a la hora de detectar neuronas, el modelo de Segmentación de Instancias muestra ser un 1% mejor que el modelo de Detección de Objetos.

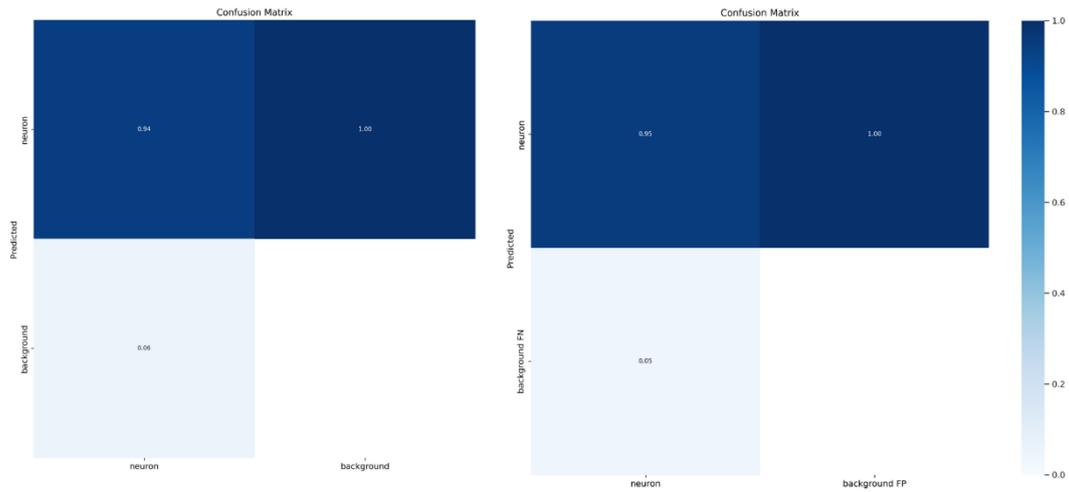


Ilustración 21: Matrices de confusión de los modelos, a la izquierda, el modelo de Detección de Objetos, y, a la derecha, el modelo de Segmentación de Instancias

Por último, la tasa de pérdida representa la parte del objeto identificado que queda sin identificar, por lo tanto, cuanto más bajo es ese valor, mejor se considera la detección. Se pueden observar las tasas de pérdida de cada modelo en la Tabla 3.

Tabla 3: Tasas de pérdida de los modelos entrenados

	Media	Mínimo
Detección de Objetos	0,2774873	0,245
Segmentación de Instancias	0,2186023	0,15349

La media de la tasa de pérdida del modelo de Segmentación basada en Instancias es ligeramente menor que en el modelo de detección, sin embargo, se puede observar una mayor diferencia en

los mínimos que alcanzó la tasa de pérdida, siendo mucho mejor para el modelo de segmentación.

En base a los resultados obtenidos, es posible determinar que el modelo de Segmentación de Instancias ha resultado, no sólo más eficaz, sino también más veloz en cuanto a su fase de entrenamiento. Es por ello, que se puede concluir que el modelo de Segmentación de Instancias resulta más adecuado para la resolución del problema de conteo de células neuronales dentro de un cultivo.

## 7 Conclusiones y trabajo futuro

Para finalizar este trabajo fin de grado, se expondrán unas conclusiones, en ellas se analizará la adecuación de la solución final adoptada para la resolución del problema y se hará un repaso de los objetivos finales fijados inicialmente para comprobar si estos han sido cubiertos mediante la elaboración de la presente investigación.

En última instancia, se hará una proposición de tareas que, tomando como punto de partida la investigación llevada a cabo en este TFG, se podrían llevar a cabo en un futuro para expandir el estudio de los cultivos de neuronas.

### 7.1 Adecuación de la solución final

En el apartado de “[Resultados](#)” ha sido posible evaluar dos modelos que se adecuaban correctamente a la tarea de conteo de neuronas y se ha determinado que el modelo que aporta mejores resultados es el de Segmentación de Instancias. Este modelo ha aportado mejores resultados cuantitativos y también ha demostrado consumir menos tiempo en el entrenamiento.

El modelo de Segmentación de Instancias permitirá realizar un conteo de las neuronas gracias a la cuantificación de los *bounding boxes* detectados en las imágenes. Además, como ya se había comentado, facilita tareas futuras que quieran automatizarse como son el estudio de la morfología y tamaño de las células neuronales y el seguimiento de neuronas individuales en distintas imágenes que se obtengan de un mismo cultivo.

### 7.2 Evaluación de alcance

A continuación, se hará un repaso de los objetivos establecidos en el apartado “[Hipótesis de partida y objetivos concretos](#)”, pasando uno por uno para comentar si el objetivo se ha alcanzado y en qué términos.

- **Investigación del estado del arte de la Visión Artificial.** Se ha realizado un estudio del estado del arte en lo referente a VA, se han descubierto interesantes investigaciones científicas que aplican soluciones de DL a la resolución de problemas de análisis de imágenes.
- **Estudio de las diferentes técnicas de Visión Artificial que podrían ser aplicadas al problema de detección.** Se han explorado y valorado todas las opciones disponibles para realizar el análisis de imágenes y de forma concreta, las técnicas existentes para

identificación de objetos, habiendo explorado las ventajas y desventajas de aplicar estas al problema de identificación de neuronas.

- **Consenso de las técnicas más adaptadas a la solución que se busca.** Una vez se conocían todas las opciones posibles, se optó por dos opciones, habiendo una que traería más ventajas a la larga.
- **Aplicación de algoritmos de preprocesamiento de imágenes.** Se han construido y aplicado algoritmos de preprocesamiento de las imágenes, necesarios antes de realizar los entrenamientos, para tener la seguridad de que los resultados de los modelos eran los mejores.
- **Entrenamiento de modelos de *Deep Learning*.** Se han entrenados dos modelos distintos de DL, cada uno basado en una tecnología de identificación diferente y se han obtenido métricas de los entrenamientos realizados.
- **Análisis de los resultados obtenidos con las técnicas propuestas y comparación de estas.** Con las métricas resultantes de haber entrenado sendos modelos, se han comparado las métricas de cada uno de ellos, así como la duración de los entrenamientos para tratar de observar la superioridad de uno de ellos.
- **Determinación de la técnica más efectiva para la identificación de las neuronas.** Tras haber analizado las métricas obtenidas del entrenamiento del modelo, se ha observado que uno de los modelos trabajaba mejor en todos los aspectos posibles, demostrando de esta manera la superioridad de la Segmentación de Instancias respecto a la Detección de Objetos.
- **Aumentar los conocimientos de la alumna en el área de la Inteligencia Artificial.** Gracias a todo el estudio necesario para llevar a cabo esta investigación presente, la alumna a conseguido aumentar sus conocimientos en Inteligencia Artificial, habiendo sido capaz de entender y razonar la diversidad de técnicas de DL que se pueden aplicar a problemas de VA.
- **Comprender los pasos que implican la realización de una investigación científica.** Mediante la realización de esta investigación, la alumna a descubierto las dinámicas de una investigación, como los pasos a seguir para redactar un documento de investigación y herramientas a usar.
- **Adquirir facilidad en la búsqueda de información a través de recursos de carácter científico.** A lo largo de esta investigación, se ha tenido que realizar una gran búsqueda de información, manteniendo el cuidado de que esta fuera veraz. Se han buscado y leído numerosas publicaciones científicas, suponiendo de esta forma que la alumna pudiera adaptarse a estudio de este tipo de fuentes.

### 7.3 Trabajo futuro

Siguiendo las líneas de esta investigación, se propone una comparativa de la técnica de Segmentación de Instancias frente a la Segmentación Semántica, teniendo de esta forma la posibilidad de asegurar la superioridad de la Segmentación de Instancias frente a cualesquiera de las técnicas que podrían llevar a cabo la tarea de conteo de neuronas.

Una vez resuelto el problema de conteo de neuronas, el siguiente paso de investigación sería el seguimiento de neuronas de forma individualizada, pudiendo de esta forma rastrear cómo se desarrolla neuronas concretas dentro del cultivo, esto permitiría realizar análisis de cambios en el tamaño y morfología de estas, permitiendo de esta manera tener información mucho más exhaustiva de las células neuronales de un cultivo.

## 8 Glosario

### B

**Bioquímica:** Rama de la ciencia que estudia las reacciones químicas y procesos que ocurren en los organismos vivos.

**Biomedicina:** Campo de estudio interdisciplinario que se centra en la aplicación de principios y métodos científicos para comprender las enfermedades humanas.

***Bounding box:*** Cuadro delimitador rectangular que encierra un objeto o región de interés en una imagen.

**Bromuro de etidio:** Un colorante fluorescente utilizado para detectar ácidos nucleicos en experimentos de biología molecular.

### C

**Calceína AM:** Un tinte fluorescente que se utiliza para evaluar la viabilidad celular y el metabolismo en estudios de laboratorio.

**Célula glial:** Tipo de célula del sistema nervioso que brinda soporte y protección a las neuronas.

**Cerebelo:** Región del encéfalo responsable del control del movimiento y el equilibrio.

**Clasificación de Imágenes:** Tarea de asignar una categoría a una imagen determinada.

**Clasificador:** Algoritmo o modelo utilizado para clasificar objetos o datos en diferentes clases.

**CNN (*Convolutional Neural Network*):** Red neuronal convolucional utilizada comúnmente en tareas de visión por computadora para procesar imágenes y extraer características relevantes.

**COCO (formato):** Formato de anotación comúnmente utilizado para describir objetos y segmentos en imágenes.

**Convolución:** Operación matemática que combina dos funciones para producir una tercera función que representa cómo una de las funciones se superpone a la otra.

**Cultivo:** Método de crecimiento controlado de células o tejidos en un ambiente de laboratorio.

### D

**Dataset:** Conjunto de datos utilizados para entrenar y evaluar modelos de aprendizaje automático.

**Deep Learning:** Subcampo del *Machine Learning* que se basa en redes neuronales artificiales para extraer características y aprender representaciones de alto nivel de los datos.

**Detección de Objetos:** Tarea de identificar y localizar objetos de una clase específica en una imagen.

**Diacetato de fluoresceína:** Un tinte fluorescente utilizado para evaluar la viabilidad celular y la función de la membrana.

## E

**Enfermedad neurodegenerativa:** Trastorno caracterizado por la progresiva degeneración y pérdida de las células del sistema nervioso.

**Enfermedad neurológica:** Trastorno que afecta al sistema nervioso, incluido el cerebro, la médula espinal, los nervios periféricos y los músculos.

**Epoch:** Un ciclo completo de entrenamiento en el que el modelo ha procesado todos los ejemplos de entrenamiento una vez.

**Extracción de características:** Proceso de identificar y capturar características relevantes de los datos para su posterior análisis.

## F

**False Negative:** Resultado de un modelo de clasificación que predice incorrectamente una muestra negativa como positiva.

**False Positive:** Resultado de un modelo de clasificación que predice incorrectamente una muestra positiva como negativa.

**Feedback:** Retroalimentación que se proporciona a un sistema para ajustar su rendimiento.

**Filtro gaussiano:** Filtro utilizado en el procesamiento de imágenes para reducir el ruido de una imagen mediante una distribución gaussiana.

**Filtro Sobel:** Filtro utilizado para enfatizar bordes en una imagen mediante la convolución con núcleos Sobel.

**FPN (*Feature Pyramid Network*):** Red neuronal utilizada para la Detección de Objetos que extrae características en diferentes escalas y niveles de abstracción.

**Función de activación:** Función matemática que determina la salida de una neurona en una red neuronal.

## G

**GAN (*Generative Adversarial Network*):** Arquitectura de red neuronal utilizada para generar datos nuevos y realistas a partir de un conjunto de datos de entrenamiento.

**Google Colaboratory:** Plataforma de Google que proporciona acceso gratuito a recursos de computación en la nube para ejecutar proyectos de software.

**GoogleNet:** Una arquitectura de red neuronal convolucional desarrollada por Google para tareas de visión por computadora.

**Ground Truth:** Datos etiquetados por expertos que representan lo correcto en un problema de aprendizaje automático.

## I

**Inteligencia Artificial:** Campo de estudio que se enfoca en desarrollar sistemas y algoritmos que pueden realizar tareas que repliquen la inteligencia humana.

**IOU (*Intersection Over Union*):** Métrica utilizada para evaluar la superposición entre dos regiones o máscaras, se calcula dividiendo el área de intersección entre el área de unión.

## J

**JSON:** Formato de intercambio de datos basado en texto utilizado para representar estructuras de datos como objetos y listas.

## K

**Kernel:** Una matriz utilizada en una operación de convolución para extraer características específicas en una imagen.

## L

**Leaky ReLU:** Función de activación utilizada en redes neuronales que evita la desaparición del gradiente al permitir un pequeño flujo de información para valores negativos.

**LSTM (*Long Short-Term Memory*):** Un tipo de célula recurrente en redes neuronales utilizada para modelar dependencias a largo plazo en secuencias de datos.

## M

**Machine Learning:** Campo de estudio que se enfoca en desarrollar algoritmos y modelos que pueden aprender y mejorar automáticamente a partir de datos sin ser programados explícitamente.

**mAP (*mean Average Precision*):** Métrica utilizada para evaluar la precisión de los modelos al calcular el promedio de las precisiones en diferentes umbrales de confianza.

**Máscara binaria:** Representación de un objeto en una imagen en forma de una matriz binaria donde los píxeles dentro del objeto son blancos y los píxeles fuera del objeto son negros.

**Mask R-CNN:** Arquitectura de red neuronal que combina la Detección de Objetos con la segmentación de instancias al agregar una máscara binaria a las detecciones.

**Microscopía de fluorescencia:** Técnica de imagen que utiliza tinciones fluorescentes para visualizar y estudiar muestras biológicas bajo un microscopio.

**Microscopio de contraste de fases:** Tipo de microscopio que resalta las diferencias de fase en una muestra para producir imágenes de alta resolución de estructuras transparentes.

## N

**Neurita:** Término utilizado para referirse a las proyecciones largas y delgadas de las neuronas, como las dendritas y los axones.

**Neurociencia:** Campo de estudio que se enfoca en la estructura y función del sistema nervioso, incluyendo el cerebro, la médula espinal y los nervios.

**Neurona:** Unidad fundamental del sistema nervioso que transmite señales eléctricas y químicas.

**NeuronJ:** Software utilizado para el análisis y trazado de neuronas en imágenes neuronales.

**Neurotoxina/neurotóxico:** Sustancia o compuesto que puede dañar o alterar la función normal de las células o tejidos nerviosos.

**Píxel:** Unidad básica de una imagen digital, que representa un punto en la cuadrícula de píxeles.

**Pooling:** Operación utilizada en las redes neuronales para reducir el tamaño de las características al tomar el valor máximo o promedio de un grupo de píxeles.

**Precisión:** Métrica que mide la proporción de predicciones positivas correctas en relación con el total de predicciones positivas realizadas.

**Preprocesamiento de imágenes:** Proceso de manipulación y transformación de imágenes para mejorar los resultados de su posterior análisis.

**Python:** Lenguaje de programación ampliamente utilizado en el ámbito del aprendizaje automático y la ciencia de datos.

## R

**Recall:** Métrica que mide la proporción de muestras positivas correctas que se han identificado en relación con el total de muestras positivas.

**Red neuronal:** Modelo computacional inspirado en el sistema nervioso que consiste en capas de nodos interconectados, llamados neuronas, que procesan y transmiten información.

**ReLU:** Función de activación llamada *Rectified Linear Unit* que produce la salida igual a cero para valores negativos y lineal para valores positivos.

**RGB (imagen):** Modelo de representación una imagen como una combinación de los colores rojo, verde y azul.

**RNN (Recurrent Neural Network):** Tipo de red neuronal que tiene conexiones recurrentes, lo que le permite procesar secuencias de datos.

**ROI (Region Of Interest):** Región específica de una imagen que se selecciona para su posterior análisis o procesamiento.

**RPN (Regional Proposal Network):** Red neuronal utilizada en modelos de Detección de Objetos para generar propuestas de regiones candidatas.

## S

**Sistema nervioso:** Sistema complejo de tejidos y órganos que coordina y controla las funciones corporales y permite la comunicación entre diferentes partes del cuerpo.

**Segmentación Semántica:** Tarea de asignar una etiqueta o categoría a cada píxel en una imagen para identificar y distinguir diferentes objetos o regiones.

**Segmentación de Instancias:** Tarea de identificar y separar cada objeto individual en una imagen mediante la asignación de una etiqueta única a cada instancia.

**SSD (*Single Shot Detector*):** Arquitectura de red neuronal utilizada para la Detección de Objetos en imágenes que es rápida y precisa.

## T

**Tinción diferencial:** Método de tinción utilizado para resaltar y visualizar diferentes componentes o estructuras en una muestra biológica.

**Toxina:** Sustancia venenosa o dañina que puede afectar negativamente a los organismos vivos.

**True Negative:** Resultado de un modelo de clasificación que predice correctamente una muestra negativa como negativa.

**True Positive:** Resultado de un modelo de clasificación que predice correctamente una muestra positiva como positiva.

## U

**U-Net:** Arquitectura de red neuronal convolucional utilizada para la segmentación de imágenes, especialmente en problemas médicos.

## V

**Ventana deslizante:** Técnica utilizada en la Detección de Objetos que consiste en deslizar una ventana sobre la imagen para identificar posibles regiones de interés.

**Voronoi-Otsu:** Método de segmentación de imágenes que combina el algoritmo de Otsu con la transformación de Voronoi para segmentar objetos.

**Visión Artificial:** Campo de estudio que se enfoca en desarrollar sistemas y algoritmos capaces de entender, interpretar y analizar imágenes y videos.

## Y

**YOLO (*You Only Look Once*):** Arquitectura de red neuronal utilizada para la Detección de Objetos en tiempo real que propone detecciones de objetos directamente a partir de una única pasada por la red.

## 9 Referencias

- [1] S. Huh, D. Fei, E. Ker, H. Su, y T. Kanade, «Apoptosis Detection for Adherent Cell Populations in Time-lapse Phase-contrast Microscopy Images».
- [2] A. Martínez-Gómez, «Comunicación entre células gliales y neuronas I. Astrocitos, células de Schwann que no forman mielina y células de Schwann perisinápticas», *Revista de Medicina e Investigación*, vol. 2, n.º 2, pp. 75-84, jul. 2014, doi: 10.1016/S2214-3106(15)30002-9.
- [3] «TEMA 1: CELULAS DEL SISTEMA NERVIOSO».
- [4] «Organización Mundial de la Salud».
- [5] «¿Qué significa tener una enfermedad neurodegenerativa? - Alzheimer Catalunya». <https://alzheimercatalunya.org/es/que-significa-tener-una-enfermedad-neurodegenerativa/> (accedido 6 de junio de 2023).
- [6] «El cerebro tiene capacidad para crear nuevas neuronas durante toda la vida». <https://www.lavozdegalicia.es/noticia/ciencia/2021/10/21/cerebro-capacidad-crear-nuevas-neuronas-durante-vida/00031634836656145685320.htm> (accedido 6 de junio de 2023).
- [7] «Neurotoxinas – Clínica Marco Rived». <https://clinicamarcorived.com/neurotoxinas/> (accedido 6 de junio de 2023).
- [8] A. Rosca *et al.*, «Impact of environmental neurotoxic: current methods and usefulness of human stem cells», *Heliyon*, vol. 6, n.º 12, dic. 2020, doi: 10.1016/J.HELIYON.2020.E05773.
- [9] «MICROSCOPIO DE CONTRASTE DE FASES». <https://www.equiposylaboratorio.com/portal/articulo-ampliado/microscopio-de-contraste-de-fases> (accedido 6 de junio de 2023).
- [10] M. T. Fernández, V. Zitko, S. Gascón, y A. Novelli, «The marine toxin okadaic acid is a potent neurotoxin for cultured cerebellar neurons», *Life Sci*, vol. 49, n.º 19, pp. PL157-PL162, ene. 1991, doi: 10.1016/0024-3205(91)90398-U.

- [11] P. Sánchez Álvarez, I. Ochoa Garrido Raquel Moreno-Loshuertos Jose M<sup>a</sup> Ayuso, y C. Rio Ebro Grado en BIOTECNOLOGÍA, «Trabajo Fin de Grado Desarrollo de sistemas microfluídicos biomiméticos aplicados a la investigación tumoral Autor», Accedido: 6 de junio de 2023. [En línea]. Disponible en: <http://zaguan.unizar.es>
- [12] D. J. Hamilton, G. M. Shepherd, M. E. Martone, y G. A. Ascoli, «An ontological approach to describing neurons and their relationships», *Front Neuroinform*, vol. 6, n.º APRIL 2012, abr. 2012, doi: 10.3389/FNINF.2012.00015.
- [13] C. H. Huang, S. Sankaran, D. Racoceanu, S. Hariharan, y S. Ahmed, «Online 3-D tracking of suspension living cells imaged with phase-contrast microscopy», *IEEE Trans Biomed Eng*, vol. 59, n.º 7, pp. 1924-1933, 2012, doi: 10.1109/TBME.2012.2194487.
- [14] S. Y. Ho, C. Y. Chao, H. L. Huang, T. W. Chiu, P. Charoenkwan, y E. Hwang, «NeurphologyJ: An automatic neuronal morphology quantification method and its application in pharmacological discovery», *BMC Bioinformatics*, vol. 12, n.º 1, pp. 1-18, jun. 2011, doi: 10.1186/1471-2105-12-230/FIGURES/13.
- [15] F. He, X. Huang, X. Wang, S. Qiu, F. Jiang, y S. H. Ling, «A neuron image segmentation method based Deep Boltzmann Machine and CV model», *Computerized Medical Imaging and Graphics*, vol. 89, p. 101871, abr. 2021, doi: 10.1016/J.COMPMEDIMAG.2021.101871.
- [16] «Five GANs for Better Image Processing | by Martin Isaksson | Towards Data Science». <https://towardsdatascience.com/five-gans-for-better-image-processing-fabab88b370b> (accedido 22 de junio de 2023).
- [17] A. Arbelle y T. R. Raviv, «MICROSCOPY CELL SEGMENTATION VIA ADVERSARIAL NEURAL NETWORKS», Accedido: 6 de junio de 2023. [En línea]. Disponible en: <https://github.com/arbellea/DeepCellSeg.git>
- [18] A. Arbelle y T. R. Raviv, «Microscopy Cell Segmentation via Convolutional LSTM Networks», *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, pp. 1008-1012, may 2018, doi: 10.1109/ISBI.2019.8759447.
- [19] P. Isola, J. Y. Zhu, T. Zhou, y A. A. Efros, «Image-to-Image Translation with Conditional Adversarial Networks», *Proceedings - 30th IEEE Conference on Computer Vision and*

- Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 5967-5976, nov. 2016, doi: 10.1109/CVPR.2017.632.
- [20] U. Schmidt, M. Weigert, C. Broaddus, y G. Myers, «Cell Detection with Star-convex Polygons».
- [21] J. Yi *et al.*, «Multi-scale Cell Instance Segmentation with Keypoint Graph based Bounding Boxes», *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11764 LNCS, pp. 369-377, jul. 2019, doi: 10.1007/978-3-030-32239-7\_41.
- [22] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, y J. Yao, «DeepPap: Deep convolutional networks for cervical cell classification», *IEEE J Biomed Health Inform*, vol. 21, n.º 6, pp. 1633-1643, nov. 2017, doi: 10.1109/JBHI.2017.2705583.
- [23] Y. Pang, K. Zhang, Y. Yuan, y K. Wang, «Distributed object detection with linear SVMs», *IEEE Trans Cybern*, vol. 44, n.º 11, pp. 2122-2133, 2014, doi: 10.1109/TCYB.2014.2301453.
- [24] E. Baykal, H. Dogan, M. E. Ercin, S. Ersoz, y M. Ekinici, «Modern convolutional object detectors for nuclei detection on pleural effusion cytology images», *Multimed Tools Appl*, vol. 79, n.º 21-22, pp. 15417-15436, jun. 2020, doi: 10.1007/S11042-019-7461-3.
- [25] J. Hung *et al.*, «Applying Faster R-CNN for Object Detection on Malaria Images», *Conf Comput Vis Pattern Recognit Workshops*, vol. 2017, p. 808, 2017, doi: 10.1109/CVPRW.2017.112.
- [26] F. Xing, Y. Xie, H. Su, F. Liu, y L. Yang, «Deep Learning in Microscopy Image Analysis: A Survey», *IEEE Trans Neural Netw Learn Syst*, vol. 29, n.º 10, pp. 4550-4568, oct. 2018, doi: 10.1109/TNNLS.2017.2766168.
- [27] «Segmentación de imágenes - MATLAB & Simulink - MathWorks España». [https://es.mathworks.com/help/images/image-segmentation.html?s\\_tid=CRUX\\_lftnav](https://es.mathworks.com/help/images/image-segmentation.html?s_tid=CRUX_lftnav) (accedido 6 de junio de 2023).
- [28] R. Girshick, J. Donahue, T. Darrell, y J. Malik, «Rich feature hierarchies for accurate object detection and semantic segmentation», *Proceedings of the IEEE Computer Society*

*Conference on Computer Vision and Pattern Recognition*, pp. 580-587, nov. 2013, doi: 10.1109/CVPR.2014.81.

- [29] R. Girshick, «Fast R-CNN», Accedido: 22 de junio de 2023. [En línea]. Disponible en: <https://github.com/rbgirshick/>
- [30] Z. Cheng y A. Qu, «A fast and accurate algorithm for nuclei instance segmentation in microscopy images», *IEEE Access*, vol. 8, pp. 158679-158689, 2020, doi: 10.1109/ACCESS.2020.3020393.
- [31] J. Yi *et al.*, «Object-Guided Instance Segmentation with Auxiliary Feature Refinement for Biological Images», *IEEE Trans Med Imaging*, vol. 40, n.º 9, pp. 2403-2414, sep. 2021, doi: 10.1109/TMI.2021.3077285.
- [32] J. Yi, P. Wu, M. Jiang, Q. Huang, D. J. Hoepfner, y D. N. Metaxas, «Attentive neural cell instance segmentation», *Med Image Anal*, vol. 55, pp. 228-240, jul. 2019, doi: 10.1016/J.MEDIA.2019.05.004.
- [33] «¿Qué es la inteligencia artificial o IA? | Google Cloud | Google Cloud». <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=es-419> (accedido 7 de junio de 2023).
- [34] «¿Qué es la Inteligencia Artificial? - Iberdrola». <https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial> (accedido 7 de junio de 2023).
- [35] Stuart. J. Russell y P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th Global Edition. Pearson Education, 2022.
- [36] R. Zambrano, «Diferencias entre Machine Learning y Deep Learning», 2019. <https://openwebinars.net/blog/diferencias-entre-machine-learning-y-deep-learning/> (accedido 27 de junio de 2023).
- [37] «Data Science vs Machine Learning vs Deep Learning: The Difference - StrataScratch». <https://www.stratascratch.com/blog/data-science-vs-machine-learning-vs-deep-learning-the-difference/> (accedido 12 de junio de 2023).

- [38] «introduccion\_rnn.md».  
[http://personal.cimat.mx:8181/~mriviera/cursos/aprendizaje\\_profundo/RNN\\_LTSM/introduccion\\_rnn.html](http://personal.cimat.mx:8181/~mriviera/cursos/aprendizaje_profundo/RNN_LTSM/introduccion_rnn.html) (accedido 16 de junio de 2023).
- [39] «A Short Introduction to Generative Adversarial Networks - Thalles' blog».  
<https://sthalles.github.io/intro-to-gans/> (accedido 16 de junio de 2023).
- [40] P. Arbeláez, M. Maire, C. Fowlkes, y J. Malik, «From Contours to Regions: An Empirical Evaluation \*», Accedido: 14 de junio de 2023. [En línea]. Disponible en: <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/papers/amfm-cvpr2009.pdf>
- [41] «What Is Instance Segmentation? [2023 Guide & Tutorial]».  
<https://www.v7labs.com/blog/instance-segmentation-guide> (accedido 15 de junio de 2023).
- [42] «Instance segmentation for digital pathology tissue image analysis».  
<https://digitalpathologyplace.com/instance-segmentation-for-digital-pathology-tissue-image-analysis-what-is-it-and-why-do-we-need-it/> (accedido 15 de junio de 2023).
- [43] «Introducción a YOLO: detección de objetos en tiempo real - HashDork».  
<https://hashdork.com/es/Yolo/> (accedido 26 de junio de 2023).
- [44] «Detección de objetos I: YOLO • Un artículo de La Máquina Oráculo».  
<https://lamaquinaoraculo.com/computacion/deteccion-de-objetos/> (accedido 26 de junio de 2023).
- [45] «Explore Top Computer Vision Models». <https://roboflow.com/models> (accedido 26 de junio de 2023).
- [46] «Mean Average Precision (mAP) in Object Detection». <https://learnopencv.com/mean-average-precision-map-object-detection-model-evaluation-metric/> (accedido 27 de junio de 2023).
- [47] «a Example of Precision-Recall curve, with the precision score on the y... | Download Scientific Diagram». [https://www.researchgate.net/figure/a-Example-of-Precision-Recall-curve-with-the-precision-score-on-the-y-axis-and-the\\_fig1\\_321672019](https://www.researchgate.net/figure/a-Example-of-Precision-Recall-curve-with-the-precision-score-on-the-y-axis-and-the_fig1_321672019) (accedido 27 de junio de 2023).

- [48] «Object Detection: Models, Architectures & Tutorial [2023]». <https://www.v7labs.com/blog/object-detection-guide> (accedido 16 de junio de 2023).
- [49] «YOLO (You Only Look Once): Detección de Objetos en Tiempo Real - Mi Diario Python». <https://pythondiario.com/2018/09/yolo-you-only-look-once-deteccion-de.html> (accedido 16 de junio de 2023).

