

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO
FACULTAD DE MEDICINA
LICENCIATURA EN BIOINGENIERÍA MÉDICA
DEPARTAMENTO DE EVALUACIÓN PROFESIONAL



**SEGMENTACIÓN AUTOMÁTICA DE ESTRUCTURAS PRESENTES EN
OTOLITOS SAGITA A PARTIR DE IMÁGENES DE MICROFOTOGRAFÍA**
TESIS

QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN BIOINGENIERÍA MÉDICA

PRESENTA:

PL.B.M. JESÚS VÁZQUEZ OSORIO

DIRECTOR:

DR. JOSÉ JAVIER REYES LAGOS

ASESORA EXTERNA:

DRA. NIDIYARE HEVIA MONTIEL

REVISORES:

DRA. ADRIANA CRISTINA PLIEGO CARRILLO

MTRA. CLAUDIA IVETTE LEDESMA RAMÍREZ

TOLUCA, ESTADO DE MÉXICO; MAYO, 2023.

Índice general

1. Antecedentes	6
1.1. Relación entre el estudio morfométrico de otolitos y biología	6
1.1.1. Estudios morfológicos	8
1.1.2. Estudios automatizados de morfometría en otolitos de peces	9
2. Marco Teórico	11
2.1. Otolitos en peces	11
2.1.1. Microestructura del otolito	12
2.1.2. Objetivos del análisis de los otolitos	15
2.1.3. Imagenología por microscopia	16
2.2. Procesamiento de imágenes	16
2.2.1. Imágenes digitales	16
2.2.2. Transformación geométrica de imágenes	20
2.2.3. Relaciones básicas entre píxeles	20
2.3. Inteligencia Artificial (IA)	21
2.3.1. Aprendizaje automático	21
2.3.2. Aprendizaje profundo	22
2.3.2.1. Visión computarizada	22
2.3.2.2. Aumento de datos	23
2.3.2.3. Extracción de características	24
2.3.2.4. Detección de objetos	24
2.3.2.5. Segmentación semántica	24
2.3.2.6. Segmentación de instancias	25
2.3.2.7. Técnicas de segmentación de imágenes mediante aprendizaje profundo	25
2.3.2.8. Aplicaciones de la visión computarizada	26

3. Marco de referencia	28
3.1. Planteamiento del problema	28
3.1.1. Pregunta de investigación	29
3.2. Justificación	29
3.3. Hipótesis	31
3.4. Objetivos	31
3.4.1. Objetivo general	31
3.4.2. Objetivos específicos	32
4. Materiales y métodos	33
4.1. Etapa 1: Base de muestras	33
4.2. Etapa 2: Aumento de datos de microfotografía y segmentación manual	34
4.3. Etapa 3: Diseño de algoritmo de aprendizaje profundo	35
4.4. Etapa 4: Evaluación del desempeño del algoritmo diseñado	42
5. Resultados y discusión	46
6. Conclusión y trabajo futuro	58

Resumen

El objetivo de este estudio fue la propuesta de una red neuronal Mask R-CNN para la segmentación automática en imágenes de microfotografía, del sulco acústico en otolitos de peces, contemplando la relevancia del estudio de microestructuras presentes en estos otolitos para diferentes estudios posteriores, entre los cuales destacan; determinación de la edad de los peces, estudios de gestión y población pesquera, clasificación de especies. La propuesta de red logra un desempeño alto en cuanto a la precisión en esta tarea de segmentación.

Esta investigación fue aplicada, exploratoria y experimental. En la experimentación se trabajó con 104 microfotografías diferentes de otolitos *sagita* y consistió en el entrenamiento de una red convolucional con muestras previamente segmentadas manualmente por expertos y evaluar el rendimiento de la red con métricas específicas para aprendizaje profundo. Los resultados demuestran resultados que superan en diferentes aspectos a las técnicas comunes de segmentación, principalmente el tiempo y recursos necesarios.

Los resultados obtenidos en la evaluación final del trabajo, muestran porcentajes por encima del 80% en las métricas de evaluación de modelos de aprendizaje profundo: *precision*, *recall*, *F1 score* y *accuracy* comparando la segmentación propuesta por el modelo con segmentaciones manuales, incluso en imágenes no utilizadas en el proceso de entrenamiento del modelo.

En general, la investigación destaca el potencial de las técnicas de aprendizaje profundo, como Mask R-CNN, para la segmentación automática de estructuras complejas en imágenes de muestras biológicas, y abre nuevas posibilidades para futuras investigaciones en el campo de la gestión pesquera y la biología marina.

Abstract

The objective of this study was the proposal of a Mask R-CNN neural network for the automatic segmentation of the acoustic sulcus in fish otolith microphotograph images, considering the relevance of studying the microstructures present in these otoliths for different subsequent studies, including fish age determination, fisheries management and population studies, and species classification. The proposed network achieves high performance in terms of precision in this segmentation task.

This research was applied, exploratory, and experimental. The experimentation involved working with 104 different otolith microphotographs and consisted of training the convolutional network with samples previously manually segmented by experts and evaluating the network's performance using specific metrics for deep learning. The results demonstrate improvements over different aspects of common segmentation techniques, particularly in terms of time and resources required.

The results obtained in the final evaluation of the work show percentages above 80% in the evaluation metrics of deep learning models: *precision*, *recall*, *F1 score* and *accuracy* when comparing the segmentation proposed by the model with manual segmentations, even in images not used in the model training process.

Overall, the research highlights the potential of deep learning techniques such as Mask R-CNN for the automatic segmentation of complex structures in biological sample images and opens up new possibilities for future research in the field of fisheries management and marine biology.

Capítulo 1

Antecedentes

1.1. Relación entre el estudio morfométrico de otolitos y biología

A lo largo de los últimos años se han utilizado en diferentes áreas de estudio distintas técnicas para determinar y obtener mediciones morfométricas descriptivas de las imágenes que se utilicen en la investigación, este estudio permite la discriminación entre especies del mismo tipo en cuanto a sexo, edad, tamaño, población e incluso se han utilizado para determinar la época en la que algunas especies de peces vivieron al estudiar los restos fósiles y sus características, [1].

El estudio de características morfométricas descriptivas se puede llevar a cabo de manera automática mediante la implementación de algoritmos que comparan de manera matemática imágenes con características similares y que el ser humano a simple vista no podría determinar similitudes o diferencias de manera objetiva. El estudio matemático-computacional se realiza a nivel de píxel de las imágenes buscando características similares dentro de regiones u objetos que la conforman, es así como se obtienen patrones de discriminación entre especies e incluso dentro de la misma especie de peces.

Doering & Ludwig [2] realizaron un análisis morfológico de los otolitos izquierdos de anguila para deducir indirectamente la edad de estos peces, utilizaron imágenes digitales obtenidas de videos microscópicos de otolitos de anguila y aplicando un análisis Rápido de la Transformada de Fourier (*Fast Fourier Transform, FFT*) se obtuvieron 64 armónicos de Fourier limitantes que forman el contorno de los otolitos espaciados igualmente, los autores concluyeron que a mayor edad del pez, mayor será la irregularidad de la forma del otolito. Pocos años después [3], en un estudio en el cual se utilizaron imágenes de muestras de los 3

pares de otolitos (*sagita*, *lapillus* y *asteriscus*) de especies de peces presentes en el noreste del océano Atlántico y de un análisis morfométrico de área y perímetro, en conjunto con descriptores de Fourier, se buscó comprobar que la forma de los otolitos es un indicador de identidad para discriminar entre especies. Los autores de dicho estudio encontraron no sólo diferencias en la forma de los otolitos en cuanto a las especies, sino que dentro de estas también se discrimina entre sexo, edad y el número de crías producido durante un año, y es el otolito de tipo sagita quien provee más información para este tipo de exclusiones. Con el mismo objetivo de encontrar relaciones entre medidas morfológicas de los otolitos y la edad de los peces se han encontrado reportados otros trabajos de investigación similares, [4, 5].

Además de estudiar los patrones descriptivos y discriminar entre especies y dentro de la misma especie, también se han realizado comparaciones entre el par de otolitos de cada pez; en el estudio '*Morphometric analysis of otoliths of juvenile crucifix sea catfish *Sciades proops* (Valenciennes, 1840)*' [6], se comparan otolitos, izquierdo y derecho, de peces jóvenes con el propósito de identificar si existen diferencias morfométricas las cuales impedirían a futuro realizar estimaciones de la especie como peso o tamaño del pescado, se encontró que la variación de las medidas en los peces jóvenes de la especie estudiada no son relevantes para determinar las estimaciones mencionadas. Por otro lado, en evidencia proporcionada por Bostanci et al. [7], se compararon los otolitos izquierdo y derecho de una especie invasora presente en Turquía que amenaza a otras especies en su hábitat. El estudio analizó los otolitos de la especie presente en 3 diferentes lagos, y se evaluaron cambios dentro de la especie acorde al ecosistema en el que se encuentra, se encontró que las características morfométricas utilizadas en el estudio (longitud, peso, perímetro, ancho, área, redondez, circularidad, rectangularidad, elipticidad, factor de forma, relación de aspecto) varían según el hábitat en el que se encuentra el pez, concluyendo que el estudio morfométrico es una herramienta confiable para la discriminación entre especies invasoras y las diferencias entre otolitos izquierdo y derecho son mínimamente diferentes.

Yedier et al. [8], estudiaron en Turquía otolitos de especies de peces anormales de 3 mares que rodean el país, se compararon las características morfométricas anormales con las 'normales' de la especie determinadas en estudios anteriores [9], y se halló por primera vez, diferencias significativas en cuanto a la morfología y métricas de otolitos de ejemplares de *Diplodus puntazo*, *Pagellus acarne* y *Trachurus mediterraneus*, en Turquía; así como en este estudio los resultados son similares en otros trabajos, [10, 11, 12, 13, 14].

1.1.1. Estudios morfológicos

En cuanto al uso de las características morfométricas de otolitos y la determinación de edad, See et al. [15], estudiaron en Malasia la relación de las características morfométricas de otolitos de 8 especies de peces presentes en el mismo país para la determinación de la edad y características similares de las especies, en la cual encontraron gran relación entre las características morfométricas de los otolitos y el tamaño del pez, por lo que los autores concluyen que la relación entre las especies también tiene relación entre las medidas de los otolitos, en consecuencia la relación entre los tamaños de los peces, aunque destacan casos en los que debido a factores ambientales las medidas de algunos ejemplares resultan incongruentes al promedio de las demás. De igual manera, la conclusión de la relación entre el tamaño del otolito es correspondiente al del pez o especie se observa en otras investigaciones, [16, 17].

El estudio morfológico de los otolitos también se aplica en un ámbito reproductivo-evolutivo, Montanini [18], realizó un estudio con ejemplares de 2 especies en Italia, el objetivo era el análisis de los otolitos en estas especies en diferentes etapas de su vida, dicho estudio concluye que algunas estructuras del otolito efectivamente varían de acuerdo con la edad por la que atraviesa el ejemplar debido a factores endógenos, ambientales o individuales. Williams et al. [19], en su investigación realizada en Norteamérica '*Evaluating the performance of otolith morphometrics in deriving age compositions and mortality rates for assessment of data-poor tropical fisheries*', estudió a la especie de pargos de aguas profundas del océano Pacífico explotada por la pesca desmedida y que debido a esto han disminuido cada vez más el número de pargos con edad/tamaño adecuados para la pesca estudiando los otolitos de ejemplares de las principales especies de pargos para la pesca. Se tomaron las medidas morfométricas de los otolitos en conjunto con un análisis estadísticos de pesca y con ellas obtuvieron estimaciones de edad de los peces, rangos de mortalidad por causas naturales que explican la disminución de la especie para pesca.

La mayoría de la información morfométrica de los otolitos de las diversas especies alrededor del mundo, se indexan para servir como base a futuras investigaciones. de Carvalho [20], estudió en el Reino Unido si estas bases de datos son útiles como herramientas en estudios de ecología para discriminar entre 3 especies de peces de una misma familia antes y después de un proceso de digestión *in vitro*, se tomaron las medidas morfométricas y se compararon con las medidas tomadas en los ejemplares que se sometieron a la digestión *in vitro*; con esto se determinó que para 2 de las 3 especies estudiadas se logra la identificación de diferencias pre y post experimentación en las medidas, sin embargo ningún índice de los utilizados logró discriminar eficientemente otolitos cogenéricos, es decir, con características muy similares

después del experimento; además sugieren más investigaciones relacionadas con este tema para tener mayor información y que las futuras investigaciones sean más relevantes.

1.1.2. Estudios automatizados de morfometría en otolitos de peces

Soria [1], presenta en su tesis de doctorado en Barcelona un algoritmo para detección automática de características en los otolitos utilizado para discriminación entre especies y determinación de edad de los peces, la información utilizada en el trabajo fueron imágenes fotográficas y señales obtenidas con sensores electrónicos, centrándose en señales de una dimensión. En cada etapa de la investigación/experimento se encuentran limitantes; al principio para detección de características, la limitante más importante es determinar las irregularidades en los otolitos, el cual se resuelve mediante una medida discriminatoria de energía/densidad en el dominio frecuencial. Al término de esta caracterización, los resultados son favorables, ya que se realizó una interpretación automática de irregularidades, extensión en el eje, grado de relevancia y posicionamiento temporal- frecuencial. Para determinar la edad se utilizó una técnica basada en la intensidad de la imagen, sin embargo, existen factores que pueden afectar esta estimación, tal es el caso de las condiciones luminosas a las que se expuso la señal/imagen de entrada, por lo que se sugiere para estudios posteriores la realización de un algoritmo que normalice las variaciones en la intensidad/contraste en las imágenes para resultados más confiables ya que se logra obtener más precisión comparándolo con estimaciones manuales; además de ser una técnica relativamente fácil. Los autores concluyeron que para mejores resultados en futuros estudios de detección automática en otolitos se elimine la brecha entre clasificadores lineales y no lineales, es decir, que el algoritmo vaya aprendiendo conforme va trabajando, que se adapte mediante los discriminantes que se utilizan, priorizando aquellos que tengan una mayor relevancia/beneficio, y en cuanto a la estimación de edad se sugiere solamente un buen preprocesamiento de la imagen en cuanto al contraste e incluso un mejor método de adquisición de las imágenes. De igual manera, otros trabajos con métodos automáticos de detección de patrones para estimación de edad, tamaño o discriminación entre especies se han encontrado, [21, 22, 23, 24, 25, 26, 27].

En el trabajo titulado '*Digital imaging techniques in otolith data capture, analysis and interpretation*', en el Reino Unido, se realizó un estudio de las técnicas digitales para el estudio de los otolitos, la investigación abordó desde la adquisición de imágenes hasta el análisis e interpretación de los resultados obtenidos en otras investigaciones. Las técnicas que se estudiaron se basan en estimaciones de edad y crecimiento en otolitos mediante imágenes con asistencia computarizada, evaluando el costo-beneficio de utilizar estas técnicas, se

estudiaron técnicas de estudio de imágenes de una dimensión, 2 dimensiones y *machine learning*; de las cuales algunas son completamente automáticas. Los resultados muestran gran porcentaje de error de las técnicas utilizadas, no obstante, se menciona que estos resultados se pueden deber a la calidad de las imágenes, omisión de preprocesamiento de las imágenes, y falta de más ejemplares para el entrenamiento de las técnicas de visión computarizada, con esto se infiere que las técnicas de visión computarizada pueden ser una gran herramienta de apoyo para los investigadores, pero por el momento no se determina que sean un reemplazante de los investigadores para este tipo de estudios, además que para algunos laboratorios que ya han introducido estudios de morfometría relacionada a la edad/tamaño de especies para la pesca pueden ser de gran utilidad [28].

En el trabajo '*Invariant morphological descriptors from otolith shape in environment automatic classification*' se describe el reconocimiento de patrones de forma de otolitos en ictiología, con el objetivo de clasificar especies de peces, analizar patrones de crecimiento y evaluar la simetría para analizar la forma de los otolitos. Se propone introducir descriptores morfométricos invariantes como compacidad discreta, tortuosidad discreta, no circularidad y simetría especular, y comparar su desempeño con los descriptores existentes. Así como, implementar el algoritmo Random Forest (RF) para clasificar las especies de peces en función de su entorno (marino, salobre y de agua dulce). Los resultados indican que los descriptores invariantes brindan información adicional, mostrando una baja correlación con las características basadas en el área o el perímetro. El clasificador RF logró una precisión de clasificación positiva para la clasificación ambiental de las especies. Este análisis resulta valioso para estudiar la dinámica trófica y permite la inferencia del medio ambiente y la evolución de las comunidades en estudios arqueológicos y paleontológicos. En general, esta investigación ofrece información importante sobre el análisis de la forma de los otolitos y sus aplicaciones en varios campos [29].

El capítulo de anterior proporciona una descripción general de la importancia del estudio morfométrico de los otolitos de peces en la identificación y determinación de edades de especies de peces y las limitaciones de los métodos de segmentación del sulco acústico manual. También analiza las diversas técnicas de procesamiento de imágenes que se han utilizado en estudios previos para la segmentación de otolitos. En contraste, el capítulo que a continuación se presenta, aborda los términos y conceptos fundamentales para la comprensión del trabajo en todas las etapas de desarrollo del modelo de aprendizaje automático. Se abordan temas relacionados con la estructura y morfometría de los otolitos, procesamiento de imágenes, inteligencia artificial, aprendizaje automático y aprendizaje profundo.

Capítulo 2

Marco Teórico

La segmentación de microestructuras de los otolitos es importante en la gestión pesquera y los estudios ecológicos, puesto que permite la estimación más precisa actualmente de la edad y tasas de crecimiento de los peces. Los métodos tradicionales para esta tarea generalmente consumen mucho tiempo y requieren de expertos para su realización; sin embargo, los avances computacionales en cuanto al aprendizaje profundo muestran la posibilidad para automatizar este proceso. En este apartado, exploraremos las bases de la presente investigación sobre el desarrollo de un sistema de automatizado de segmentación del sulco acústico en otolitos basado en el aprendizaje profundo. Específicamente, se abordarán los conceptos de aprendizaje automático, aprendizaje profundo, redes neuronales artificiales, redes neuronales convolucionales, y cómo se pueden aplicar a la tarea de segmentación de microestructuras en otolitos de peces.

2.1. Otolitos en peces

Los otolitos (Figura 2.1), también llamados piedras del oído, son estructuras de carbonato de calcio que se encuentran dentro del oído interno de los peces teleósteos.

Dentro de cada oído (izquierdo y derecho) se encuentran 3 otolitos: *sagita*, *lapillus* y *asteriscus* (Figura 2.3), localizados en el sáculo, utrículo y lagena respectivamente, [31]. A pesar de estar ubicados dentro del cráneo de los peces, los otolitos no están adheridos/acoplados al cráneo, se mantienen en esa ubicación anatómica debido a la conexión con los canales del oído interno. Los otolitos tienen funciones sensitivas y motoras, proveen de sensibilidad a vibraciones dentro del agua para brindar equilibrio y orientación a los peces en su entorno, de igual forma apoyan en la audición; y son la estructura de los peces más utilizada para determinar la edad de los peces [32].



Figura 2.1: Imagen de microfotografía de otolito. Tomada de [30].

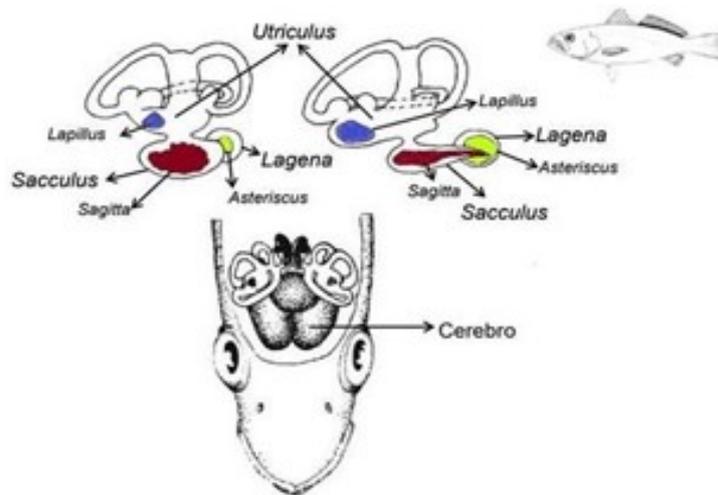


Figura 2.2: Ubicación anatómica de los otolitos en los peces. Tomada de [33].

2.1.1. Microestructura del otolito

Los otolitos son acumulaciones de material calcáreo obtenido de la superficie que se forman alrededor de un núcleo, el cual aparece a muy temprana edad; estas acumulaciones crecen a lo largo de la vida del pez formando capas translúcidas y opacas que en conjunto forman *anillos*, que representan un año de crecimiento [34].

Los otolitos son estructuras muy pequeñas llegando a medir alrededor de 5nm, por lo cual el estudio de estas estructuras se vuelve un poco más complicado, aunado a esto las formas de contorno son variadas (Figura 2.3); sin embargo, gracias al avance tecnológico se pueden



Figura 2.3: Tipos de contornos de los otolitos. Tomada de [9].

estudiar estas estructuras con diferentes técnicas, por ejemplo; utilizando microscopia, microscopia electrónica, rayos X, entre otras técnicas [35].

De la tercia de otolitos de cada oído, el otolito más utilizado para estudios taxonómicos en peces es el otolito *sagita* debido a su morfología y la información que se puede extraer de ella, por ello se describe la morfología del otolito *sagita* por áreas (dorsal/superior y ventral/inferior), caras (interna/anterior y externa/posterior) y forma del contorno (Figura

2.4).

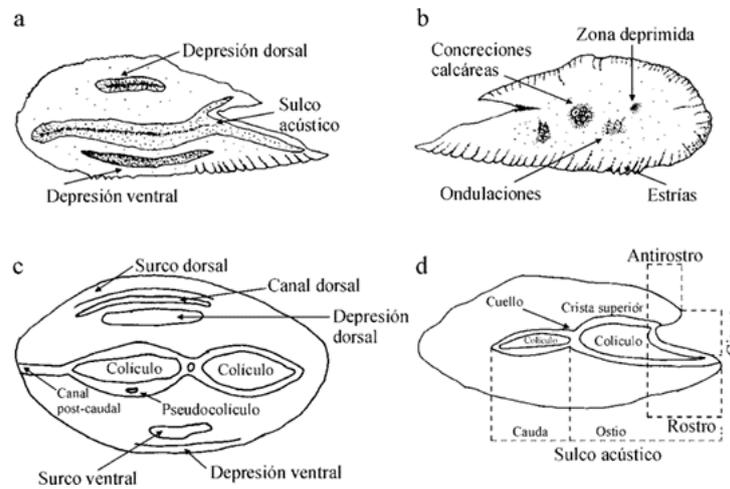


Figura 2.4: a), c), d) Cara interna otolito *sagita*. b) Cara externa otolito *sagita*. Tomada de [36].

Este trabajo se centra en el estudio de las microestructuras presentes en la cara interna de los otolitos *sagita*, debido a que ahí se encuentran las estructuras de interés; sulco acústico (ostio y cauda) como se muestra en la Figura 2.6, [9].

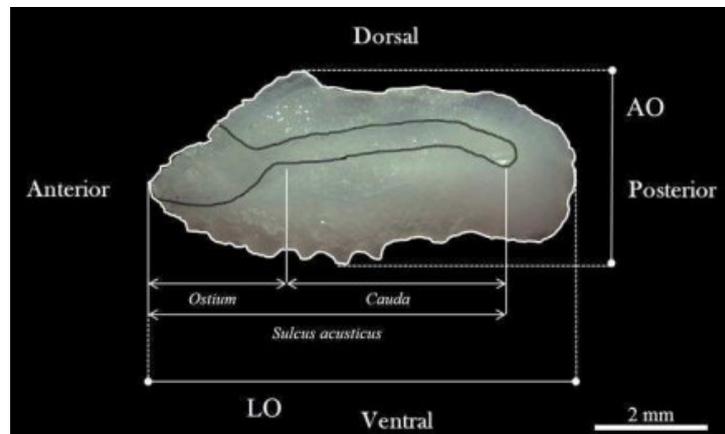


Figura 2.5: Delimitación del sulco acústico (ostio y cauda) en otolito de *mugilidae*. Tomada de [37].

El sulco acústico posee características muy importantes para la identificación de especies. Son especialmente importantes el tipo de apertura que presenta el sulco, la relación morfológica entre el ostio y la cauda, y la posición del surco [9]. En la Figura 2.6 se presentan los diferentes tipos de aperturas y posicionamiento de esta microestructura.

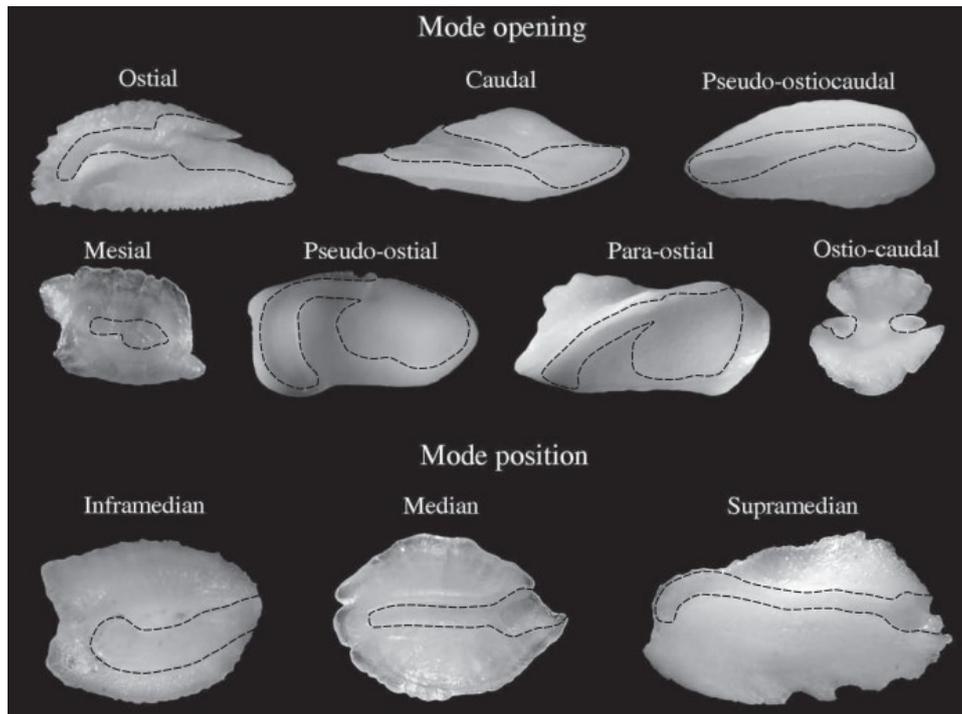


Figura 2.6: Tipos de aperturas y posicionamiento del sulco acústico. Tomada de [9].

2.1.2. Objetivos del análisis de los otolitos

El estudio de los otolitos tiene una amplia variedad de propósitos, dependiendo del análisis realizado; entre ellos destacan [38]:

- **Determinación de la edad:** Los otolitos crecen continuamente a lo largo de la vida de un pez, formando *anillos*, denominados *anillos anuales*, contando estos anillos los científicos pueden determinar la edad del pez.
- **Gestión pesquera:** Estudiando los otolitos de diferentes poblaciones de peces, los científicos pueden determinar las edades relativas de los peces en diferentes áreas, lo que puede ayudar en la toma de decisiones en cuanto a la gestión pesquera.
- **Monitoreo ambiental:** La composición química de los otolitos puede brindar información sobre la calidad del agua y las condiciones ambientales que experimentó el pez durante su vida.
- **Identificación de poblaciones:** Al analizar el material genético contenido en los otolitos, los científicos pueden identificar la población de origen de un pez, lo cual es importante para monitorear y administrar las poblaciones de peces.

- **Historia/evolución:** Con el estudio químico de los otolitos también se pueden determinar el comportamiento del pez, alimentación, ecosistema en el que vivió y patrones de migración.

2.1.3. Imagenología por microscopia

La microscopía es una herramienta fundamental utilizada para estudiar objetos/materiales a nivel microscópico. Con el uso de microscopios, se logran observar y analizar los detalles del objeto de estudio. La microscopía además, es capaz de producir imágenes de alta calidad que brindan información sobre la estructura y el comportamiento de las muestras en estudio. Las imágenes obtenidas por microscopía se han convertido en un elemento importante en la investigación científica de varios campos.

Las fotomicrografías permiten a los investigadores capturar imágenes de alta resolución de los otolitos, proporcionando información detallada sobre su estructura y composición. Estas imágenes se pueden usar para identificar y medir los incrementos de crecimiento en los otolitos, que luego se pueden usar para determinar la edad de un pez. Además, las fotomicrografías pueden revelar la forma, el tamaño y la posición de los otolitos, proporcionando información valiosa para la identificación taxonómica[39].

En resumen, las fotomicrografías son una herramienta esencial para el estudio de los otolitos y pueden proporcionar información valiosa sobre la edad, la tasa de crecimiento y la historia ambiental de las poblaciones de peces prueba de ello se muestra en el apartado anterior de antecedentes, donde la mayoría de las investigaciones referenciadas se basan en análisis de otolitos mediante microfotografía.

2.2. Procesamiento de imágenes

2.2.1. Imágenes digitales

Una imagen es una representación visual, que manifiesta la apariencia visual de un objeto real o imaginario captada por el ojo, espejo, placa fotográfica o cualquier otro aparato óptico, gracias a los rayos de luz que recibe y proyecta, [40]. Para procesar cualquier imagen mediante un ordenador primero debe realizarse un proceso de digitalización, esto quiere decir que una imagen analógica se convierte en una imagen digital; que a la vez se puede clasificar por:

- **Dimensión:**
 - 2D contempla ancho y largo de una imagen.

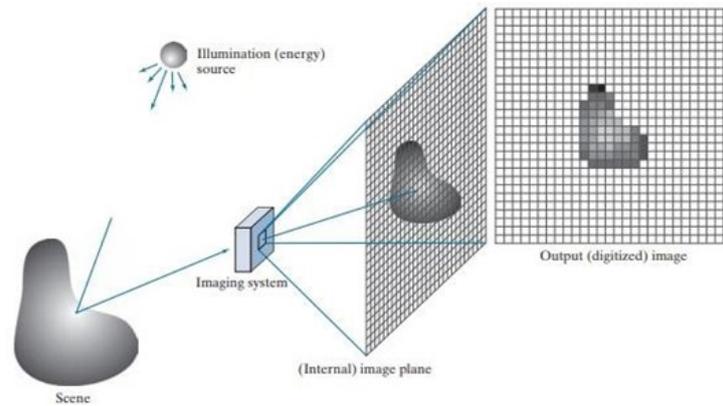


Figura 2.7: Proceso de digitalización de una imagen. Tomado de [41].

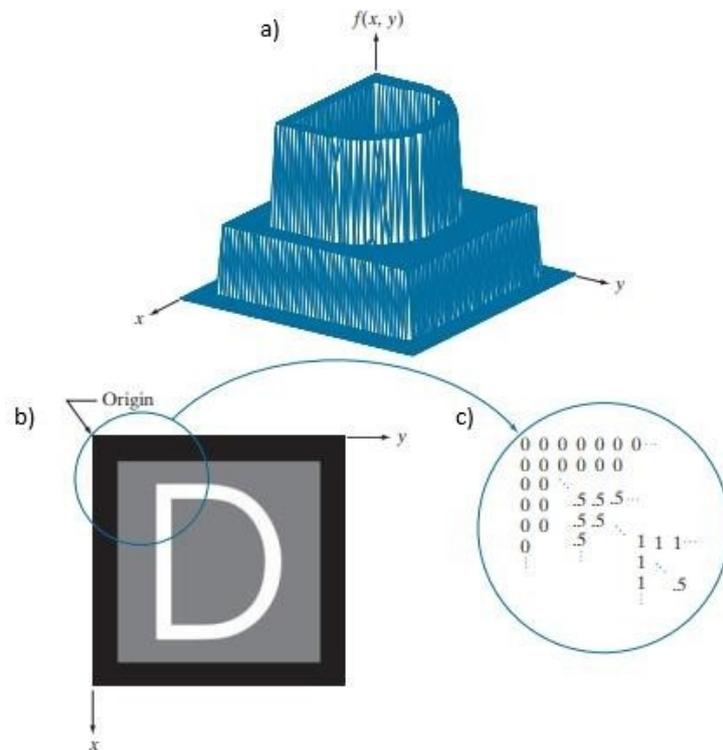


Figura 2.8: Representaciones de una imagen, a) imagen graficada en una superficie, b) imagen proyectada en arreglo visual de intensidades, c) arreglo numérico de la imagen (0=negro, .5=gris, 1=blanco). Tomada de [41].

- 3D contempla las dimensiones de 2D añadiendo profundidad a la imagen.
- 4D además de las dimensiones de 3D contempla el tiempo.
- **Color:**
 - Imagen binaria: sólo 2 valores de color posibles en la imagen.

- Escala de grises: se presentan valores de intensidad extendidos únicamente entre blanco y negro.
- Color: Existen varios modelos;
 - RGB (por las siglas en inglés Red, Green, Blue), es una imagen representada por la mezcla/adición de 3 colores primarios; azul, rojo y verde; principalmente usado en monitores y cámaras.
 - CMY (por las siglas en inglés Cyan, Magenta y Yellow), es la imagen representada por la resta de colores al color resultante de la combinación (negro) de los 3 colores mencionados; cyan (color opuesto al rojo), magenta (color opuesto al verde) y amarillo (color opuesto al azul), utilizado principalmente en impresión.
 - CMYK (por las siglas en inglés Cyan, Magenta, Yellow y Key), modelo muy similar al CMY, sin embargo, se agrega un color, el negro; principalmente usado en impresión.



Figura 2.9: Misma imagen con diferente resolución espacial; 930 dpi, 300 dpi, 150 dpi y 72 dpi. Tomada de [41].

Para la digitalización de una imagen se requiere una fuente de energía que ilumine la escena de la imagen, un sistema de captura que proyecta la imagen en un plano coordenado

obteniendo esa imagen digitalizada, (Figura 2.7).

Una imagen en escala de grises se define como una función bidimensional, $f(x, y)$ donde x y y son planos coordenados, y f es la amplitud de cada par de coordenadas (x, y) ; es decir el nivel de gris de la imagen o también llamada intensidad en ese punto, los valores de intensidad son todos discretos/finitos, la cual se puede procesar mediante un ordenador/computadora (Figura 2.8); el procesamiento de imágenes por computadora tiene como objetivo principal analizar imágenes que están fuera del espectro visible del humano dentro del espectro electromagnético, [41].

A cada valor/elemento de una imagen digital se le conoce como píxel, cada imagen está formada por un arreglo de píxeles los cuales cada uno tienen un valor numérico real, correspondiente a la intensidad en la coordenada (x, y) correspondiente para formar la imagen completa. La resolución espacial (Figura 2.9), es el más mínimo cambio perceptible en la imagen, en una imagen digital, los píxeles determinan la resolución espacial mediante la medida llamada *dots (pixels) per inch (dpi)* que mide el mayor número de pares de líneas perceptibles por unidad de distancia, [41].

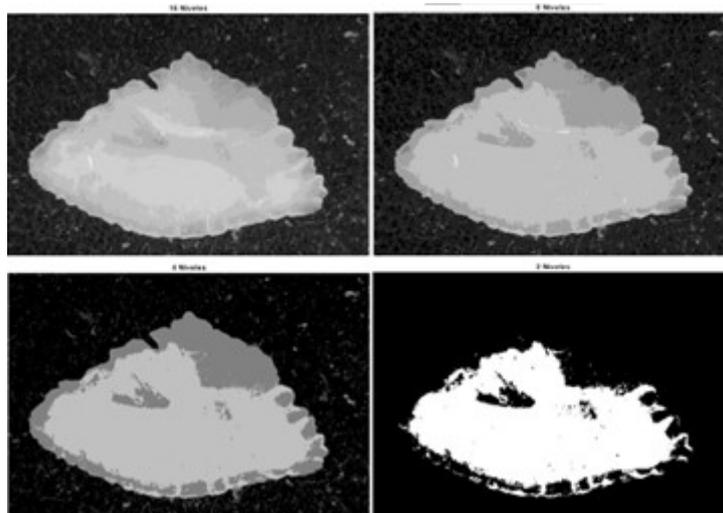


Figura 2.10: Misma imagen de otolito con diferentes niveles de intensidad; 16, 8, 4, 2 niveles de intensidad. Elaboración propia.

La resolución de intensidad es el más mínimo cambio en intensidad perceptible en la imagen (Figura 2.10), para una mejor resolución se consideran las características del equipo de cómputo utilizado para el procesamiento de imágenes y se cuantifica con números de potencia entera de 2; comúnmente 8-bits ($2^8 = 256$ valores de intensidad) y 16-bits ($2^{16} = 65,536$ valores de intensidad) [41].

2.2.2. Transformación geométrica de imágenes

La transformación geométrica en el procesamiento de imágenes se refiere al proceso de manipular la disposición espacial de los píxeles en una imagen, esto implica aplicar una transformación matemática a una imagen que cambia su tamaño, forma, orientación o ubicación; esta transformación puede ser lineal o no lineal.

Algunas de las principales transformaciones geométricas son escalamiento, rotación, traslación y reflexión [41]:

- Escalado: cambio del tamaño de una imagen estirándola o encogiéndola a lo largo de los ejes x y y .
- Rotación: rotar una imagen en un cierto ángulo alrededor de un punto específico.
- Traslación: movimiento de una imagen horizontal o verticalmente una cierta distancia.
- Reflexión: la imagen se voltea o refleja a lo largo de un eje particular. En imágenes bidimensionales, la reflexión se puede realizar a lo largo del eje x , el eje y o una línea diagonal.

2.2.3. Relaciones básicas entre píxeles

El estudio de las relaciones entre píxeles de una imagen se puede hacer mediante distintos métodos; por ejemplo [42]:

- Vecinos de un píxel:
 - 4-vecindad: Se compara el píxel con sus vecinos del arreglo, tiene 2 píxeles vecinos horizontales y 2 verticales (Figura 2.11).
 - 8-vecindad: Se compara el píxel con sus vecinos horizontales (2), verticales (2), y diagonales (4) (Figura 2.11).
- Conectividad: Establece límites de objetos y regiones que componen la imagen.
- Proximidad/distancia entre píxeles: Determina la cantidad de píxeles a recorrer para conectar 2 píxeles, considerando utilizar la conectividad de 4-vecindad u 8-vecindad.

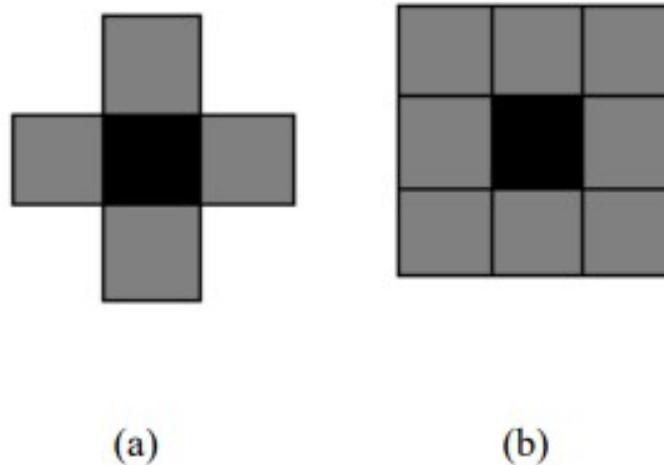


Figura 2.11: Píxel con a) 4-vecindad b) 8-vecindad. Elaboración propia.

2.3. Inteligencia Artificial (IA)

La inteligencia artificial (IA) es una rama de la informática que tiene como objetivo crear máquinas inteligentes que puedan simular el razonamiento humano y los procesos de toma de decisiones. Implica el desarrollo de algoritmos y modelos que pueden analizar y aprender de los datos, reconocer patrones y tomar decisiones o predicciones basadas en ese análisis. La IA abarca una amplia gama de subcampos, entre ellos, aprendizaje automático, aprendizaje profundo, procesamiento del lenguaje natural y robótica [43].

2.3.1. Aprendizaje automático

El aprendizaje automático es un campo de la IA que implica el desarrollo de algoritmos y modelos estadísticos que permiten a los sistemas informáticos mejorar su rendimiento en una tarea específica. Es un tipo de análisis de datos que implica el uso de técnicas estadísticas para dar a las computadoras la capacidad de aprender de los datos. Los algoritmos se entrenan en un gran conjunto de datos para reconocer patrones y relaciones entre variables. Este entrenamiento implica ajustar los parámetros del modelo para minimizar la diferencia entre los resultados previstos y los resultados reales. Una vez que se entrena el modelo, se puede usar para hacer predicciones sobre datos nuevos e 'invisibles' [43].

Existen varios tipos de aprendizaje automático; entre los cuales destacan [43]:

- Aprendizaje supervisado: El algoritmo se entrena en un conjunto de datos etiquetado, donde cada entrada se empareja con una salida correspondiente.

- **Aprendizaje no supervisado:** El algoritmo se entrena en un conjunto de datos sin etiquetar y debe encontrar patrones/relaciones por sí mismo.
- **Aprendizaje semisupervisado:** El algoritmo se entrena en una combinación de datos etiquetados y no etiquetados.
- **Aprendizaje por refuerzo:** El algoritmo aprende por ensayo y error, recibiendo retroalimentación en forma de recompensas o castigos por sus acciones.

A diferencia del aprendizaje automático tradicional, donde se necesita la ingeniería de características explícitas, el aprendizaje profundo puede descubrir automáticamente características y patrones complejos en los datos. Esto permite lograr un rendimiento destacado en una amplia gama de tareas, especialmente en visión computarizada y procesamiento del lenguaje natural. Por lo tanto, el aprendizaje profundo se ha convertido en una parte integral del campo del aprendizaje automático y ha mejorado significativamente su capacidad para realizar tareas complejas.

2.3.2. Aprendizaje profundo

El aprendizaje profundo es una rama del aprendizaje automático que implica entrenar redes neuronales artificiales con múltiples capas para aprender representaciones de datos. Estas representaciones se pueden usar para realizar diversas tareas, como la clasificación de imágenes, el reconocimiento de voz, el procesamiento del lenguaje natural y más. Los modelos de aprendizaje profundo pueden descubrir automáticamente características y patrones complejos en los datos, sin necesidad de una ingeniería de características explícita, esto les permite lograr un rendimiento destacado en una amplia gama de tareas, particularmente en visión artificial y el procesamiento del lenguaje natural [43].

2.3.2.1. Visión computarizada

La visión por computadora, también llamada visión computarizada o visión artificial; es el campo de estudio que se enfoca en desarrollar técnicas matemáticas para recuperar la forma tridimensional y la apariencia de objetos en imágenes, así como describir el mundo que vemos en una o más imágenes y reconstruir sus propiedades, como forma, iluminación y distribuciones de color. Es el proceso de permitir que las computadoras interpreten y entiendan la información visual del mundo de una manera similar a como la perciben los humanos [43].

La diferencia entre la visión computarizada y el procesamiento de imágenes digitales radica en la finalidad o el objetivo para el que se utilizan; mientras que el procesamiento de imágenes mejora la imagen para un uso posterior mediante su manipulación, la visión por computadora tiene la tarea de emular la visión humana para alguno de los fines antes mencionados [44].

La visión por computadora se basa en 3 niveles según David Marr, como se menciona en Stevens (2012), [45]:

1. Teoría computacional: ¿Cuál es el objetivo de la computación (tarea) y cuáles son las restricciones que se conocen o que se pueden aplicar al problema?
2. Representaciones y algoritmos: ¿Cómo se representa la información de entrada, salida e intermedia y qué algoritmos se utilizan para calcular el resultado deseado?
3. Implementación de hardware: ¿Cómo se mapean las representaciones y los algoritmos en hardware real, por ejemplo, un sistema de visión biológica o una pieza especializada de silicio? Por el contrario, ¿cómo se pueden usar las restricciones de hardware para guiar la elección de la representación y qué algoritmo? Con el uso cada vez mayor de chips gráficos (GPU) y arquitecturas de muchos núcleos para la visión por computadora se vuelve muy importante este nivel dentro de la tarea.

2.3.2.2. Aumento de datos

El aumento de datos es una técnica utilizada en el aprendizaje profundo para expandir artificialmente un conjunto de datos mediante la generación de versiones nuevas y modificadas de los datos existentes. El objetivo del aumento de datos es aumentar la diversidad de los datos de entrenamiento para mejorar los resultados de un modelo de aprendizaje automático. El proceso de aumento de datos implica aplicar varias transformaciones a los datos originales, como voltear, rotar, ampliar o disminuir el tamaño, recortar, eliminar o agregar ruido.

El aumento de datos es especialmente utilizado cuando el conjunto de datos original es pequeño o cuando el modelo es propenso a sobreajustarse (*overfitting*), lo que ocurre cuando el modelo memoriza los datos de entrenamiento en lugar de aprender a generalizar a nuevos datos [46].

2.3.2.3. Extracción de características

La extracción de características es el proceso de identificar y extraer las partes más relevantes e informativas de una señal, imagen o conjunto de datos para su posterior análisis.

En visión por computadora, la extracción de características se refiere al proceso de identificar atributos distintivos, o características, de una imagen que se puede usar para distinguirla de otras imágenes. Estas características pueden incluir bordes, geometría/formas, texturas, colores u otros patrones visuales. La extracción de características es un paso esencial en muchas tareas de visión artificial, como el reconocimiento de objetos, la detección de rostros, la segmentación de imágenes y el seguimiento. Al extraer características relevantes de las imágenes, los algoritmos de visión por computadora pueden analizar e interpretar automáticamente la información visual [47].

2.3.2.4. Detección de objetos

La detección de objetos identifica instancias de objetos en medios visuales. Los programas de detección de objetos dibujan un cuadro delimitador alrededor de una instancia/objeto detectado, emparejado con una etiqueta para representar el contenido del cuadro [48].

2.3.2.5. Segmentación semántica

La segmentación semántica es una técnica que permite asociar cada píxel de una imagen digital con una etiqueta de clase lo cual permite una comprensión más detallada de la imagen. Este tipo de segmentación clasifica píxeles de imagen de una o más clases en lugar de objetos que no son interpretables semánticamente, clasifica cada píxel en lugar de objetos, como es el caso de la detección de objetos.

El algoritmo utilizado para este tipo de segmentación tiene como objetivo extraer características antes de usarlas para formar categorías distintas en una imagen. El proceso para una segmentación semántica se puede resumir en [49]:

1. Análisis de los datos de entrenamiento para clasificar un objeto específico en la imagen.
2. Creación de una red de segmentación semántica para localizar los objetos y dibuje un contorno delimitador alrededor de ellos.
3. Entrenamiento de la red de segmentación semántica para agrupar los píxeles en una imagen localizada mediante la creación de una máscara de segmentación.

2.3.2.6. Segmentación de instancias

La segmentación de instancias es una forma única de segmentación de imágenes que se ocupa de detectar y delinear cada instancia distinta de un objeto que aparece en una imagen. Este tipo de segmentación, detecta todas las instancias de una clase con la funcionalidad adicional de delimitar instancias separadas de cualquier clase de segmento. Debido a su funcionamiento, refiere la incorporación de detección de objetos y funcionalidad de segmentación semántica en conjunto y logra obtener un resultado más completo [49].

2.3.2.7. Técnicas de segmentación de imágenes mediante aprendizaje profundo

Existen diferentes técnicas de visión computarizada con las cuales se puede realizar la segmentación semántica; entre las cuales destacan las arquitecturas de redes neuronales convolucionales (CNN) y las *encoder - decoder* (codificador-decodificador). A continuación se muestran las principales arquitecturas de redes neuronales utilizadas en segmentación de imágenes [50, 51, 52, 53, 54]:

- *Convolutional Neural Network (CNN)* (red neuronal convolucional): arquitectura que toma una imagen de entrada y la pasa a través de múltiples capas convolucionales, extrayendo características en cada capa. La salida de la capa convolucional final se muestrea para generar un mapa de segmentación, que asigna cada píxel en la imagen de entrada a una clase particular en función de sus características visuales.
- *Fully Convolutional Networks (FCN)* (redes totalmente convolucionales): FCN reemplaza las capas totalmente conectadas de una red neuronal convolucional tradicional con capas convolucionales, lo que permite que la red genere una máscara de segmentación de la misma resolución espacial que la imagen de entrada. Toma una imagen como entrada y genera una predicción por píxeles en forma de mapa de calor, que luego se convierte en una imagen segmentada.
- *U-Net*: arquitectura de red neuronal convolucional que se diseñó específicamente para la segmentación de imágenes biomédicas. Tiene una arquitectura en forma de U, con una ruta de contracción y una ruta de expansión, y utiliza conexiones de salto para combinar características de diferentes niveles, preservar información espacial y ayuda a recuperar detalles finos.
- *SegNet*: arquitectura de red neuronal con estructura de codificador-decodificador, donde el codificador se compone de capas convolucionales y de agrupación que extraen

características de la imagen de entrada, y el decodificador utiliza capas de muestreo ascendente para recuperar el tamaño de imagen original genera un mapa de segmentación por píxeles.

- *Faster R-CNN* (red neuronal convolucional basada en regiones): es una red neuronal de detección de objetos que utiliza una red de propuesta de región (RPN) para generar propuestas de regiones en la imagen y una red de detección para predecir clases de objetos y afinar los bordes de los objetos.
- *Mask R-CNN*: arquitectura de red neuronal que amplía la arquitectura de detección de objetos de *Faster R-CNN* para realizar la segmentación de instancias. Genera propuestas de objetos y luego utiliza una red neuronal convolucional basada en regiones para predecir máscaras de objetos para cada propuesta.

2.3.2.8. Aplicaciones de la visión computarizada

Existe una inmensa variedad de áreas y aplicaciones que se le da a la visión computarizada, y conforme el avance de la tecnología las aplicaciones siguen en constante crecimiento; a continuación se enlistan algunas de las principales áreas y aplicaciones actuales de la visión computarizada [51, 55, 56, 57]:

- Atención médica: la visión por computadora se utiliza en imágenes médicas para tareas como la detección de tumores, la segmentación de imágenes y el diagnóstico.
- Reconocimiento facial: reconocer e identificar a las personas en función de sus rasgos faciales.
- Industria/manufactura: la visión artificial se utiliza en la fabricación para tareas como el control de calidad, la detección de defectos y el seguimiento de productos.
- Robótica: la visión artificial se utiliza en robótica para tareas como el reconocimiento de objetos, la evitación de obstáculos y la localización.
- Entretenimiento: la visión por computadora se usa en el entretenimiento para tareas como efectos especiales, seguimiento de rostros y captura de movimiento.
- Deportes: la visión artificial se utiliza en los deportes para tareas como el seguimiento de jugadores, el análisis de patrones de movimiento y el suministro de análisis en tiempo real.

El capítulo sobre el marco teórico proporciona las bases cruciales para el desarrollo del modelo de segmentación automática del sulco acústico de los otolitos utilizando Mask R-CNN, se destaca la importancia del procesamiento de imágenes y las técnicas de IA en el análisis de otolitos. El capítulo marco de referencia define la razón y los objetivos de la investigación y describe el enfoque que se dará en las siguientes etapas.

Capítulo 3

Marco de referencia

3.1. Planteamiento del problema

Tomando en cuenta que los otolitos tienen medidas muy pequeñas, en promedio de 5 a 6 milímetros, [58], el análisis morfométrico de las microestructuras en otolitos se vuelve complicado para hacerlo con imágenes fotográficas con cámaras convencionales; es por eso por lo cual se utilizan imágenes fotográficas obtenidas con microscopio, también llamadas microfotografías. Los estudios presentados anteriormente, son estudios en los que las medidas morfométricas y demás técnicas de detección de estructuras en otolitos se realizan por personal humano, esto quiere decir que existe gran porcentaje de error debido a que prácticamente se estudia desde el punto de vista del investigador.

Los estudios automáticos mediante softwares y algoritmos diseñados por investigadores para este tipo de investigaciones con otolitos son más costosos, esto debido a las grandes complicaciones para obtención de muestras, procesamiento de las microfotografías, y análisis morfométrico, aunado a estas complicaciones se suma que la información sobre otolitos es muy poca y el entrenamiento/programación de algoritmos para la automatización de estos datos se complica aún más.

Las mediciones morfométricas en el ámbito biológico se utilizan en gran medida para obtener información de seres vivos a partir de inferencias que se hacen con las medidas obtenidas de estructuras, con ellas se puede deducir la edad, tamaño, sexo, especie entre otras variables; tal es el caso de los otolitos de peces de diferentes especies, [1]. De acuerdo con lo revisado se nota la importancia de la detección y conocimiento de la forma del otolito porque ayuda a realizar objetivamente la caracterización de estructuras internas del otolito *sagitta*: sulco acústico (cauda y ostio), las cuales brindan información importante para analizar datos sobre las especies de peces. El proceso de caracterización de estructuras requiere de una

automatización mediante algoritmos computacionales para cuantificar su morfometría y evitar la subjetividad en la caracterización actual realizada por los expertos.

Por lo tanto, el problema que se buscará resolver es eliminar la subjetividad de las delimitaciones manuales que actualmente se utilizan, además de automatizar este proceso y disminuir el tiempo de detección manual en grandes cantidades de imágenes para brindar información en estudios biológico-taxonómico.

3.1.1. Pregunta de investigación

¿Cuál es la efectividad de un modelo de aprendizaje automático para segmentar el sulco acústico en otolitos *sagita* de peces en términos de las métricas de *precision*, *recall*, *F1 score* y *accuracy* comparado con la segmentación manual?"

3.2. Justificación

Desde la perspectiva del bioingeniero médico, es importante abordar estudios que involucran otras especies biológicas y no sólo referentes a la especie humana, ya que como se ha mostrado en los demás apartados de la investigación, los conocimientos requeridos son los que un bioingeniero médico aborda y que genera un impacto taxonómico e incluso dentro de la especie humana para determinar/controlar las actividades que afectan a la cadena alimenticia, por ende, a todas las especies biológicas. Todos los años se extrae de los océanos más de 77,9 millones de toneladas de pescado y marisco. En su estudio, llamado SeafoodPrint («La huella de la pesca»), de Daniel Pauly, de la Universidad de Columbia Británica, y Enric Sala, *fellow explorer* de National Geographic, indica que la captura mundial no es estable ni se reparte equitativamente entre las naciones del mundo, dentro de este estudio, los investigadores proponen lo que a su juicio sería preciso hacer para salvar el mar; implementan una unidad de medida que llaman huella de la pesca, esta determina el impacto que tiene la pesca de las diferentes especies; o como lo menciona el propio Pauly «El problema es que cada pez es diferente. Un kilo de atún tiene una huella cien veces mayor que un kilo de sardinas.». Esto se refiere a que un atún grande tiene que comer el equivalente de su peso corporal cada diez días, todo pez grande depende para su subsistencia de varios eslabones de la cadena trófica. Los países poderosos suelen comprar mucho pescado, sobre todo depredadores de máximo nivel como el atún [59].

La Figura 3.1 muestra la producción primaria (organismos microscópicos situados en la base de la cadena trófica marina necesaria para producir un kilo de cada especie mari-

na), cantidad de peces totales que se requieren actualmente para atender la demanda de productos marinos en los 3 principales consumidores de peces en el mundo, y es que a pesar de que el consumo de peces es de alrededor de 9 millones en cada uno de estos países, para determinar el verdadero impacto/huella ecológica de este consumo deben tomarse en cuenta las especies de las que se alimentan y con las que logran crecer y desarrollarse estos peces más grandes para el consumo humano; aunado a esto, el consumo cada vez va más en aumento, por consecuente la pesca desmedida también crece así como la expansión de las zonas de alta mar que antes no se explotaban y hoy se han convertido en caladeros activos y monopolizados, y la venta al mejor postor de los recursos pesqueros de los países pobres. Un informe del Banco Mundial y la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) recientemente ha determinado que en el océano no hay suficientes peces para resistir el ritmo actual de explotación; de hecho, aunque se redujera a la mitad los barcos, anzuelos y redes, la pesca seguiría siendo excesiva [59].

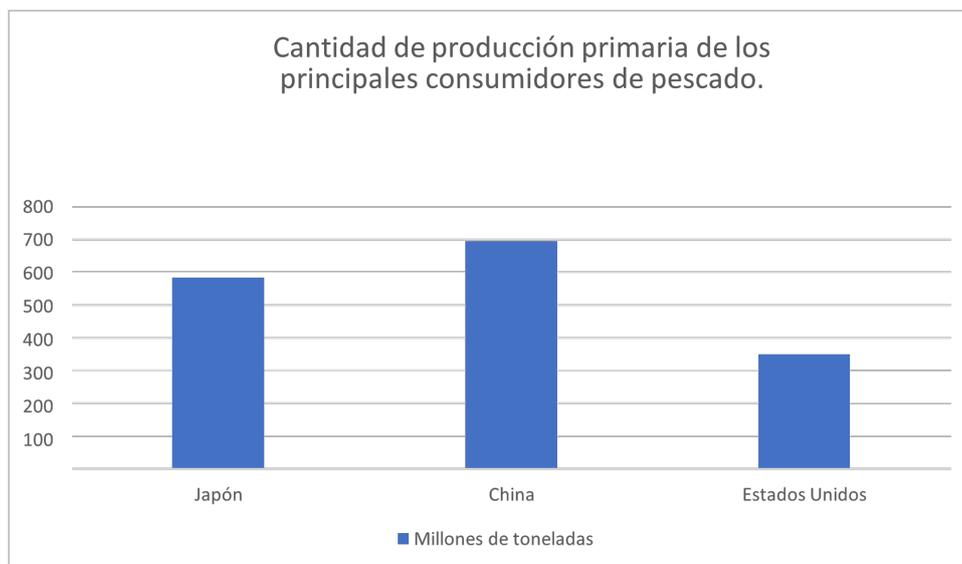


Figura 3.1: Gráfico de producción primaria en los principales consumidores de peces en el mundo [59]. Elaboración propia.

En general el impacto que genera el estudio de estructuras anatómicas automatizado mediante herramientas tecnológicas; como el estudio morfológico del otolito para identificar ciertos indicadores/características que brindan información sobre el pasado, presente y predicciones futuras, es relevante para que las diferentes ramas de la ciencia aumenten la información que se tiene, se busquen nuevas líneas de investigación que permitan mejorar la relación entre especies conservando y procurando cada componente de la cadena alimenticia para mejorar la calidad de vida de todas las especies generando un impacto tanto ecológico,

como social. Por lo tanto, este proyecto podría beneficiar a obtener datos más precisos de las especies marinas en los distintos océanos para determinar nuevas leyes que permitan la conservación de todas las especies y regular la pesca desmedida y así disminuir la huella ecológica, al conocer las especies de un ecosistema y la cadena para preservar las especies tanto predadores como de las especies de las que se alimenta. El impacto del proyecto radica en obtener métricas que permitan obtener datos de relevancia taxonómica como edad, sexo, especie; para el estudio de ecosistemas marinos, mediante la segmentación automática de los otolitos.

El estudio morfométrico automatizado en otolitos realiza un análisis objetivo, cuantitativo y más robusto en cuanto a las técnicas utilizadas en la actualidad tomando en cuenta sólo los componentes de las imágenes. No se ha explotado el estudio automatizado de otolitos debido a la desinformación acerca del impacto que tiene el estudio de estas estructuras, sin embargo, se ha demostrado que puede ser confiable y de gran utilidad para algunos laboratorios el automatizar estos procesos e utilizarlos para determinar edad, sexo, especie, tamaño entre otras inferencias, además de eliminar la subjetividad de expertos [28]. La utilización de la visión computarizada en diferentes áreas ha sido explotada en los últimos años, sin embargo, dentro del estudio de otolitos esto se ha quedado atrás debido a que los softwares que se diseñan tienen ciertas limitantes, es por eso por lo que se requiere más estudios en la materia para, de igual forma, aumentar la información al respecto y cada vez ir mejorando con cada estudio realizado.

3.3. Hipótesis

La segmentación automática de otolitos *sagita* para identificar las estructuras internas: sulco acústico (cauda y ostio) en peces, es posible mediante algoritmos de visión computacional para su análisis morfométrico.

3.4. Objetivos

3.4.1. Objetivo general

Implementar un algoritmo computacional para detección/segmentación automática de microestructuras internas: sulco acústico (cauda y ostio), en otolitos *sagita* de peces.

3.4.2. Objetivos específicos

- Seleccionar imágenes 2D de bases de imágenes de microfotografía electrónica de barrido de otolitos y segmentarlas manualmente con validación de expertos biólogos, de otolitos *sagita* de peces.
- Realizar aumento de muestras de la base de datos mediante transformaciones geométricas de las imágenes originales y segmentarlas manualmente para el proceso de entrenamiento y validación del algoritmo de detección automática.
- Diseñar e implementar algoritmo que segmente las estructuras internas: sulco acústico (cauda y ostio), automáticamente en imágenes de microfotografía electrónica de barrido de otolitos *sagita*.
- Evaluar el desempeño del algoritmo y validar funcionamiento de detección y segmentación del sulco acústico (cauda y ostio) en imágenes de microscopía electrónica de barrido de otolitos *sagita* mediante la comparación cuantitativa de métricas que evalúan el desempeño de algoritmos de aprendizaje automático.

Al describir los pasos prácticos tomados a implementar en este capítulo, el capítulo siguiente: materiales y métodos, describe detalladamente la configuración experimental y los procedimientos utilizados para lograr los objetivos de la investigación. Se aborda minuciosamente la base de datos utilizada, aumento de datos, el diseño de la red neuronal Mask R-CNN y la evaluación del rendimiento de la red neuronal. Asimismo, proporciona la base para el capítulo siguiente de resultados y discusión.

Capítulo 4

Materiales y métodos

4.1. Etapa 1: Base de muestras

Las imágenes de microfotografía que se utilizaron en esta investigación, se obtuvieron de bases de datos de la Universidad Nacional Autónoma de México (UNAM); principalmente del “Catálogo de Otolitos Sagita de Peces del Golfo de México”, a la cual se obtuvo acceso gracias a la participación en esta investigación de la Doctora Nidiyare Hevia Montiel, [60]. Dicha base de datos contiene imágenes de otolitos sagita derecho e izquierdo de especies marinas, salobres y de agua dulce de peces adultos y juveniles. Algunas especies son especies comercialmente importantes, mientras que otras tienen un valor ecológico ya que son endémicas y se encuentran en varias categorías de riesgo en el Golfo de México y Península de Yucatán, la lista ordenada filogenéticamente se encuentra en el Apéndice A.

La captura de especies se realizó bajo permisos de recolección de la SAGARPA (no. DGO-PA/04031/310510.1940, no. PPF/DGOPA-319/17), y el material biológico se encuentra resguardado en una colección registrada por la SEMARNAT (DGVS- CC-305-18). Las imágenes microscópicas se adquirieron con un dispositivo Axio Zoom V16 Zeiss, muestran una descripción estandarizada de la forma y las estructuras del otolito *sagita* utilizando características morfométricas adquiridas con el software ZEN 2 Pro con una incertidumbre del 5%.

El fin de utilizar imágenes de bases de datos ya establecidas es no generar un impacto ambiental en la extracción de muestras de otolitos de peces; la información de las bases de datos se utilizará únicamente con fines académicos y de investigación. Se utilizaron imágenes de microfotografía de 104 diferentes otolitos *sagita* de peces.

4.2. Etapa 2: Aumento de datos de microfotografía y segmentación manual

Para el estudio se utilizarán específicamente imágenes de microfotografía electrónica de barrido *scanning* como la presentada en la Figura 4.1

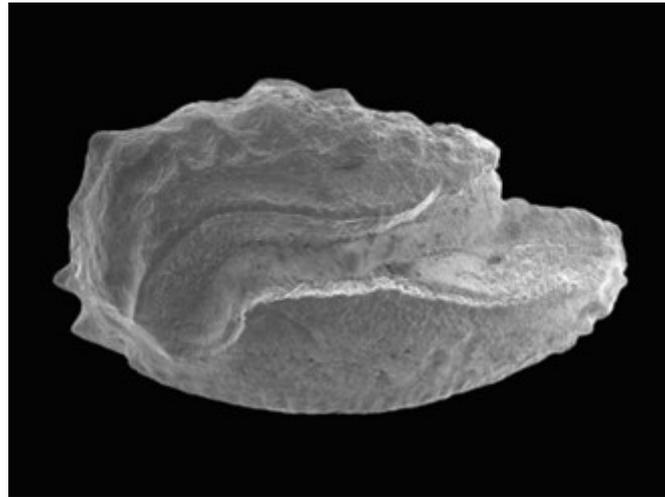


Figura 4.1: Imagen de otolito de *Acanthurus chirurgus* obtenida con microscopio electrónico de barrido (*scanning*). Tomada de [30].

Las imágenes de microfotografía utilizadas fueron, como se mencionó 104 diferentes, las cuales se dividieron de la siguiente manera:

Uso de imágenes	Número de imágenes
Aumento de datos	90
Prueba/ <i>test</i> final	14

Tabla 4.1: División de las imágenes originales de otolitos.

A las 100 imágenes utilizadas para aumento de datos se les realizaron transformaciones geométricas en conjunto (por ejemplo; reflexión+traslación+rotación) de manera aleatoria y con diferentes valores para cada transformación en los rangos mostrados en Tabla 5.1, para finalmente obtener como resultado 1000 imágenes. Posteriormente se realizó la segmentación manual de las 1000 imágenes mediante el software VGG Image Annotator (VIA) versión 2.0.11 (<https://www.robots.ox.ac.uk/vgg/software/via/via-2.0.11.html>), obteniendo como resultado un archivo con extensión JavaScript Object Notation (.json).

De igual manera, el proceso de segmentación manual mediante el software VIA se realizó a las 14 imágenes para prueba final.

Transformación geométrica	Rango de valores
Escalado	Factor entre 0 y 1
Rotación	Factor entre -40° y 40°
Desplazamiento horizontal	Factor entre -0.2 y 0.2 del ancho total
Desplazamiento vertical	Factor entre -0.2 y 0.2 del alto total
Cizallamiento	Factor entre -0.2 y 0.2
Zoom	Factor entre 0.8 y 1.2
Reflexión horizontal	Factor de 0.5
Reflexión vertical	Factor de 0.5
Ajuste de brillo	Factor entre 0.2 y 0.8

Tabla 4.2: Rango de valores para transformaciones geométricas.

4.3. Etapa 3: Diseño de algoritmo de aprendizaje profundo

Para el diseño del algoritmo de segmentación automática, primero se realizó una búsqueda sobre los trabajos relacionados a esta investigación, de ellos se obtuvieron diferentes propuestas de arquitecturas de aprendizaje profundo. En segunda instancia, teniendo las opciones de arquitectura de algoritmos utilizados en tareas semejantes a la nuestra, se realizó de nuevo investigación sobre las opciones que se tenían; con esto se determinaron los pros y contras de la utilización de una u otra arquitectura, para finalmente determinar que se utilizaría una arquitectura de red neuronal convolucional denominada Mask R-CNN desarrollada por el equipo de Kaiming He y su equipo de Facebook AI Researches (FAIR) en 2018 [61]. En la Figura 4.2, se puede ver gráficamente a grandes rasgos la arquitectura que utiliza esta red neuronal y en la tabla siguiente se describen estas etapas que la conforman [62].

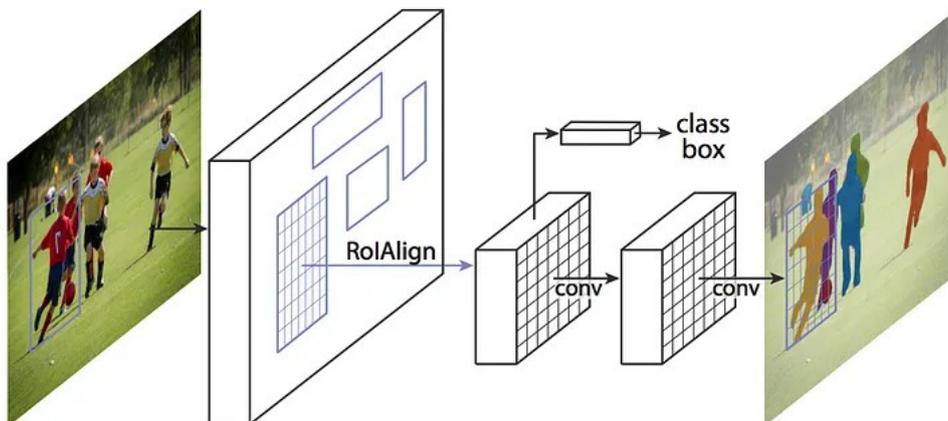


Figura 4.2: Estructura de Mask R-CNN. Tomada de [62].

Mask R-CNN	
Etapas	Descripción
1 <i>Backbone</i> (columna vertebral/red troncal)	Red neuronal convolucional estándar, sirve como extractor de características. Las primeras capas detectan características de bajo nivel, las capas posteriores detectan características de nivel superior. La imagen se convierte en un mapa de características que será la entrada para las siguientes etapas. <i>Feature Pyramid Network</i> (FPN) mejora la representación de objetos de múltiples escalas al agregar una segunda pirámide que pasa las características de alto nivel a las capas inferiores.
2 <i>Region Proposal Network</i> (RPN) (red de región propuesta)	Escanea imágenes en busca de objetos utilizando cuadros de anclaje (cuadros distribuidos en el área de la imagen). Superpone alrededor de 200,000 anclajes para cubrir la imagen, los escanea en paralelo utilizando la naturaleza convolucional de la red y reutiliza las características extraídas para evitar cálculos duplicados. Aquí se generan 2 resultados; predice la clase de primer plano o de fondo y refina los cuadros delimitadores. Selecciona los anclajes principales, los refina y realiza la supresión para generar propuestas finales.
3 <i>ROI Classifier and Bounding Box Regressor</i> (clasificador de regiones de interés y regresor de cuadro delimitante)	Etapas funcionales sobre las regiones de interés (ROIs) propuestas por el RPN; genera dos salidas para cada ROI, clase que clasifica las regiones con mayor profundidad, no sólo fondo o primer plano y el refinamiento del cuadro delimitador que precisa aún más la ubicación del objeto detectado.
3.1 <i>ROI pooling</i> (agrupamiento ROI)	Recorta y cambia el tamaño de los mapas de características, pero tiene requisitos de tamaño fijos. Se sugiere, por los autores de Mask R-CNN, el uso de <i>ROI Align</i> (muestreo e interpolación bilineal).

4 <i>Segmentation masks</i> (máscaras segmentadas)	Esta etapa es una red convolucional que genera máscaras blandas para regiones positivas seleccionadas por el clasificador ROI. Con un tamaño pequeño (28x28 píxeles), las máscaras de baja resolución mantienen la luz de la rama de la máscara. Durante el entrenamiento, las máscaras de verdad se reducen al mismo tamaño para calcular la pérdida en el entrenamiento y, durante la inferencia, las máscaras pronosticadas se escalan hasta el tamaño del cuadro delimitador de ROI para obtener las máscaras finales.
--	--

Tabla 4.3: Etapas del funcionamiento de Mask R-CNN [62].

Se codificó el algoritmo en lenguaje Python para la ejecución de la red neuronal seleccionada (Mask R-CNN), tomando como base el código de Canu (2021) [63] el cual está basado en el repositorio de GitHub de Matterport Inc. [64] y se establecieron sus hiperparámetros [65];

Hiperparámetros de Mask R-CNN		
Hiperparámetro	Descripción	Valor asignado
<i>Backbone</i> (columna vertebral/red troncal)	Mask R-CNN requiere una arquitectura convolucional tradicional; ResNet50, ResNet101 y ResNext101 son las principales. ResNet50 es más rápido con pesos pre-entrenados, mientras que ResNet101 y ResNext101 son más precisos con un alto poder de cómputo.	ResNet101 añadiendo mejora con FPN
<i>Train ROIs per image</i> (regiones de interés para entrenamiento por imagen)	Número máximo de ROIs generados por la red de propuestas regionales (RPN) para cada imagen, que se procesará para su clasificación y enmascaramiento en la siguiente etapa.	200
<i>Maximum get instances</i> (número máximo de instancias en una imagen)	Número máximo de instancias que se pueden detectar en una imagen. Si la cantidad de instancias en las imágenes es limitada, se reducirá el número de falsos positivos así como el tiempo de entrenamiento.	100

<i>Detection minimum confidence</i> (confianza mínima de detección)	El umbral del nivel de confianza determina la clasificación de una instancia. Se puede ajustar según la cantidad de instancias detectadas, con un umbral más bajo detectaríamos todo generando más falsos positivos, y un umbral más alto genera menos falsos positivos y una mayor precisión.	0.7
<i>Image minimum and maximum dimension</i> (dimensión mínima y máxima de las imágenes)	Estos ajustes controlan el tamaño de la imagen, el tamaño es cuadrado. El tamaño predeterminado es 1024x1024, pero se pueden usar tamaños más pequeños para reducir los requisitos de memoria y el tiempo de entrenamiento. Lo ideal, es comenzar con tamaños más pequeños y aumentarlos gradualmente durante el entrenamiento e ir aumentando el tamaño para obtener los parámetros finales del modelo.	Mínimo=800 Máximo=1024

Tabla 4.4: Hiperparámetros utilizados para el funcionamiento de Mask R-CNN [62].

Al igual que los hiperparámetros para el entrenamiento de la red, se establecieron hiperparámetros para calcular el valor de pérdida en los pesos de esta arquitectura; Mask R-CNN utiliza una función de pérdida para medir la diferencia entre la salida prevista del modelo y la salida objetivo real durante el entrenamiento. Esta función es compleja y consta de varias pérdidas diferentes, cada una correspondiente a una etapa diferente del modelo. Posteriormente, las pérdidas en cada etapa se ponderan utilizando hiperparámetros que determinan la importancia de cada etapa en el cálculo de la pérdida general. Al ajustar estos hiperparámetros, podemos controlar la medida en que cada etapa del modelo contribuye a la pérdida final y, por lo tanto, al proceso de entrenamiento general (4.3) [65].

Hiperparámetros para cálculo de la función de pérdida de Mask R-CNN		
Hiperparámetro	Descripción	Valor asignado
<i>RPN class loss</i> (pérdida en la clasificación de las regiones propuestas por la red)	Corresponde a la pérdida que se le asigna a la clasificación indebida de los anclajes (presencia/ausencia de algún objeto) por la red de propuesta de regiones. Al aumentar este valor, se asegura que la red de propuestas de regiones capturará más objetos.	Este valor es el resultado del cálculo de la pérdida de entropía cruzada entre las clases de anclaje objetivo y los <i>logits</i> (modelo de elección binaria) del clasificador RPN previstos para fondo/objeto. La función de pérdida selecciona los anclajes positivos y negativos para el cálculo, y excluye los anclajes neutrales (valor de coincidencia = 0). El valor de pérdida se obtiene tomando la media de los valores de pérdida calculados, o 0 si no hay filas seleccionadas.
<i>RPN bounding box loss</i> (pérdida del cuadro delimitador de la red de regiones propuestas)	Hiperparámetro que influye en la precisión de localización del RPN. Este peso se ajusta en caso de que se detecte el objeto, pero debe corregirse el cuadro delimitador.	Se calcula seleccionando solo los anclajes positivos de la comparación entre anclajes y cuadros delimitadores obtenidos en la imagen así como sus deltas (<i>offsets</i> del cuadro delimitador previsto para las ROIs propuestas por la RPN). Más adelante, los deltas del cuadro delimitador objetivo se recortan a la misma longitud que los <i>offsets</i> . Finalmente, <i>smooth L1 loss</i> (pérdida uniforme de L1) se calcula entre el cuadro delimitador objetivo y el propuesto por RPN.

<p><i>Mask R-CNN classification loss</i> (pérdida en la clasificación de Mask R-CNN)</p>	<p>Corresponde a la pérdida que se asigna a la clasificación indebida del objeto en la RPN. Debe aumentarse en caso de que el objeto sea detectado, pero se clasifique incorrectamente.</p>	<p>Se calcula la pérdida de entropía cruzada entre las probabilidades de clase predichas y las etiquetas de clase verdaderas. La función toma como entrada las etiquetas de clase verdaderas codificadas como números enteros, y los <i>logits</i> de clase predichos, que son las probabilidades de clase no previstas generadas por el clasificador de Mask R-CNN. Además, la función borra las pérdidas de predicciones de clases que no están en las clases activas de la imagen; se promedian los números de predicciones activas en el lote para el valor final.</p>
<p><i>Mask R-CNN bounding box loss</i> (pérdida del cuadro delimitador de Mask R-CNN)</p>	<p>Pérdida asignada a la localización del cuadro delimitador de la clase identificada en la red; se aumenta si se realiza la clasificación correcta del objeto, pero la localización no es precisa.</p>	<p>Para este cálculo, sólo los ROIs positivos son tomados en cuenta junto con el nombre de su clase perteneciente. Los deltas (predichos y verdaderos) que contribuyen a la pérdida se recopilan y se calcula <i>smooth-L1</i> con estos y se obtiene la media de la pérdida igual al valor de pérdida.</p>
<p><i>Mask R-CNN mask loss</i> (pérdida en las máscaras de Mask R-CNN)</p>	<p>Corresponde a la pérdida de entropía cruzada binaria en las máscaras creadas sobre los objetos identificados. Si la identificación de máscara (a nivel de píxel) es relevante, se debe aumentar este peso.</p>	<p>Se seleccionan los ROIs positivos, sus máscaras y clase. Las máscaras pronosticadas y las máscaras verdaderas se recopilan utilizando los índices de los ROIs positivos y clase. Finalmente, la pérdida de entropía cruzada binaria se calcula con las máscaras recopiladas. Si no hay ROIs positivos, la pérdida es igual a cero; se promedian los ROIs positivos y ese es el valor.</p>

Tabla 4.5: Hiperparámetros utilizados para el cálculo de pérdida de Mask R-CNN [65].

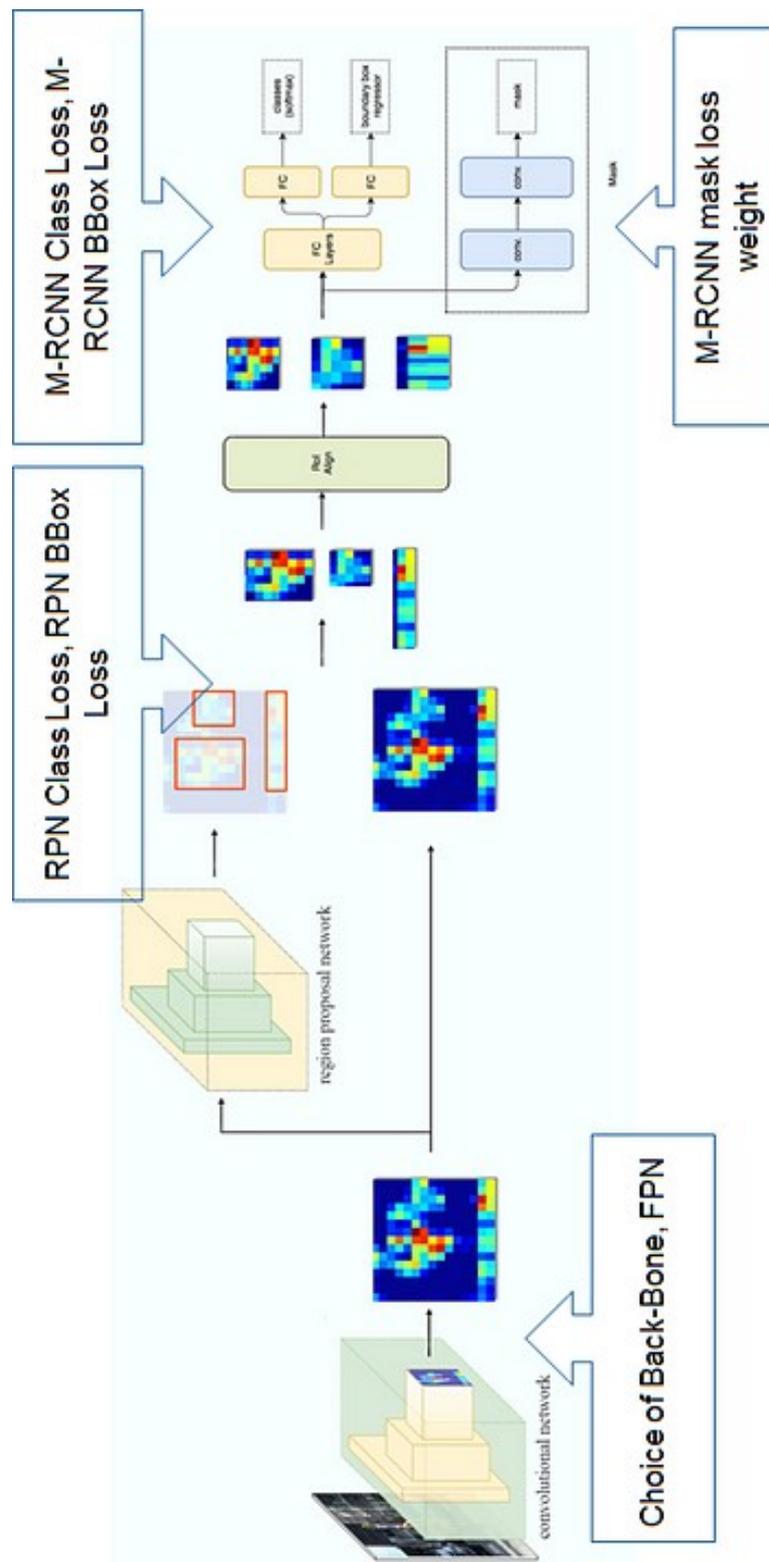


Figura 4.3: Representación gráfica de valores de pérdida dentro de Mask R-CNN. Tomada de [62].

4.4. Etapa 4: Evaluación del desempeño del algoritmo diseñado

Teniendo el código del algoritmo, se determinó que se utilizarían de las 1000 imágenes resultantes del aumento de datos; 889 imágenes para la etapa de entrenamiento de la red neuronal y 111 para la etapa de validación. A continuación, debido al requerimiento computacional para estos procesos se optó por ejecutar el algoritmo en Google Colab en la versión Pro para lograr la ejecución del código completa como se determinó en los hiperparámetros. El proceso de ejecución se repitió 5 veces para hacer validación cruzada con estas ejecuciones, obteniendo las métricas que evalúan el desempeño de la red con las imágenes de validación (111); estas métricas son:

Métricas para evaluar el funcionamiento Mask R-CNN		
Métrica	Descripción	
<i>True positives</i> (TP) (verdaderos positivos)	Número de píxeles positivos (píxeles blancos) correctamente identificados en una imagen binaria.	
<i>True negatives</i> (TN) (verdaderos negativos)	Número de píxeles negativos (píxeles negros) correctamente identificados en una imagen binaria.	
<i>False positives</i> (FP) (falsos positivos)	Número de píxeles positivos (píxeles blancos) incorrectamente identificados en una imagen binaria.	
<i>False negatives</i> (FN) (falsos negativos)	Número de píxeles negativos (píxeles negros) incorrectamente identificados en una imagen binaria.	
Métrica	Descripción	Fórmula
<i>Precision</i> (precisión)	Métrica de proporción de verdaderos positivos del total de predicciones positivas realizadas por el modelo. En otras palabras, mide qué tan precisas son las predicciones positivas. Se calcula como la proporción de casos positivos verdaderos con respecto al número total de predicciones positivas. Alta precisión significa que cuando el modelo predice un positivo, por lo general es correcto.	$Precision = TP / (FP + TP)$

<i>Recall</i> (sensibilidad)	Representación de la proporción de verdaderos positivos del total de casos positivos en los datos verdaderos. En otras palabras, mide cuántos píxeles positivos de los datos de entrenamiento se predijeron correctamente. Se calcula como la relación entre los píxeles positivos verdaderos y el número total de píxeles positivos reales. Un <i>recall</i> alto significa que el modelo puede encontrar la mayoría de los casos positivos reales.	$Recall = TP / (FN + TP)$
<i>F1 score</i> (puntuación f1)	Esta métrica resulta de la media armónica de <i>precision</i> y <i>recall</i> , lo cual proporciona un balance entre las dos métricas.	$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$
<i>Accuracy</i> (precisión global)	Proporción de predicciones correctas, tanto verdaderos positivos como verdaderos negativos, entre todas las predicciones hechas por el modelo.	$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$

Tabla 4.6: Métricas de evaluación del funcionamiento de Mask R-CNN [66].

De las 5 ejecuciones se obtuvieron los archivos ejecutables resultantes de 3 épocas; época 10, 20 y 35; esto con el objetivo de realizar una validación cruzada más sofisticada.

Para la elección del ejecutable con mejor desempeño, se determinó una jerarquía entre las métricas *precision*, *recall*, *F1 score* y *accuracy*, la jerarquía se hizo acorde al objetivo de la implementación de la red; 1 *Precision*, 2 *Recall*, 3 *F1 score*, 4 *Accuracy*.

La razón de esta jerarquía es para dar prioridad a los valores de las métricas en la elección del mejor ejecutable, las 2 métricas con mayor peso, (*precision* y *recall*), es debido al objetivo principal, que es detectar sólo los píxeles pertenecientes al sulco acústico (verdaderos positivos) y dejando los píxeles no pertenecientes al sulco acústico como fondo (píxeles negativos).

Por último, al obtener el archivo ejecutable con mejor desempeño con las imágenes de validación, se procedió a ejecutar el archivo con las imágenes de prueba final (14) y de igual modo obtener las métricas de desempeño con estas imágenes que son nuevas para el algoritmo, debido a que no se incluyeron ni en la etapa de entrenamiento ni en la de validación. También se añadieron las máscaras de verdad, las máscaras generadas por la red y la superposición de ambas máscaras para calcular los índices de similitud en imágenes de

la Tabla 4.7.

Índice	Descripción
<i>Jaccard</i>	También conocido como intersección sobre unión (IoU). Se define como el tamaño de la intersección de dos conjuntos dividido por el tamaño de la unión de los conjuntos. Se usa comúnmente para evaluar la precisión de los modelos de segmentación y detección de objetos.
<i>Dice</i>	Se calcula y se define igual que <i>F1 score</i> . En segmentación de imágenes, se utiliza para evaluar la similitud entre la segmentación predicha y la segmentación real.
<i>Cosine</i>	Mide la similitud entre dos vectores calculando el coseno del ángulo entre ellos. Se usa comúnmente para medir la similitud entre dos imágenes representándolos como vectores de características de alta dimensión. En el contexto del procesamiento de imágenes, la similitud del coseno se utiliza para comparar la similitud entre las características de la imagen extraídas por diferentes modelos.

Tabla 4.7: Índices de similitud en imágenes.

Este capítulo presentó la metodología utilizada para desarrollar el sistema automático de segmentación automática del sulco acústico de otolitos con ayuda del algoritmo de Mask R-CNN. En el próximo capítulo, se reportan los resultados para cada etapa del capítulo de materiales y métodos, proporcionando una explicación detallada del rendimiento y la precisión del modelo propuesto, de igual forma, se discuten las limitaciones del enfoque. Específicamente, la sección de resultados presenta la evaluación cuantitativa y cualitativa del método propuesto, incluyendo métricas como *accuracy*, *precision*, *recall* y *F1 score*. Además, el capítulo proporciona una discusión sobre los resultados alcanzados, destacando la eficacia y eficiencia de la metodología propuesta comparándolo con los métodos existentes, y se discuten posibles mejoras futuras.

Capítulo 5

Resultados y discusión

Las imágenes de microfotografía utilizadas (104), tienen propiedades en común, sin embargo, el tamaño varía demasiado entre estas imágenes. La siguiente tabla indica las propiedades en común de las imágenes, en seguida la Figura 5.1 es un ejemplo:

Propiedad	Valor
Resolución	96 <i>dpi</i>
Formato	.jpg (<i>Joint Photographic Experts Group</i>)
Profundidad	32 <i>bits</i>
Modelo de color	CMYK

Tabla 5.1: Propiedades de imágenes de la base de muestras.



Figura 5.1: Ejemplo de imagen de la base de muestras. Especie *Calamus nodosus*.

Aplicando aleatoriamente las transformaciones geométricas de la Tabla 5.1 a las imágenes

originales, se obtuvieron 1000 imágenes como la presentada en 5.2, aproximadamente se obtuvieron 11 imágenes secundarias/nuevas de cada imagen original. Las 1000 imágenes obtenidas se normalizaron con las especificaciones de la Tabla 5.2.

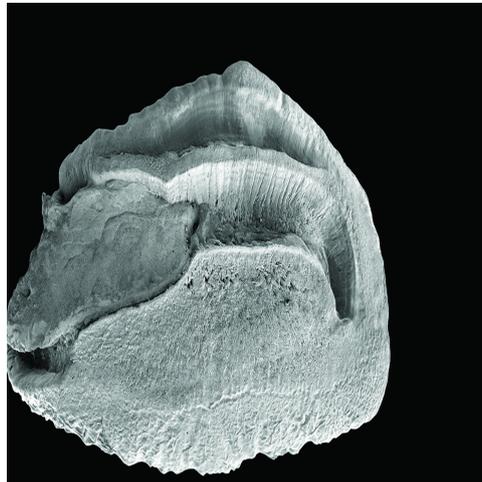


Figura 5.2: Ejemplo de imagen obtenida en el aumento de datos. Especie *Calamus nodosus*.

Propiedad	Valor
Dimensiones	2000 x 2000 pixeles
Modelo de color	RGB
Resolución	96 <i>dpi</i>
Formato	.jpg (<i>Joint Photographic Experts Group</i>)
Profundidad	24 <i>bits</i>
Modelo de color	RGB

Tabla 5.2: Propiedades de imágenes resultantes del aumento de datos.

Como se mencionó, existen diferentes contornos de otolitos, así como tipos de aperturas y posicionamientos del sulco acústico, la Figura 5.3 muestra 6 segmentaciones manuales del sulco acústico realizadas con el software VIA a otolitos con diferente contorno, apertura y posicionamiento utilizadas en este trabajo.

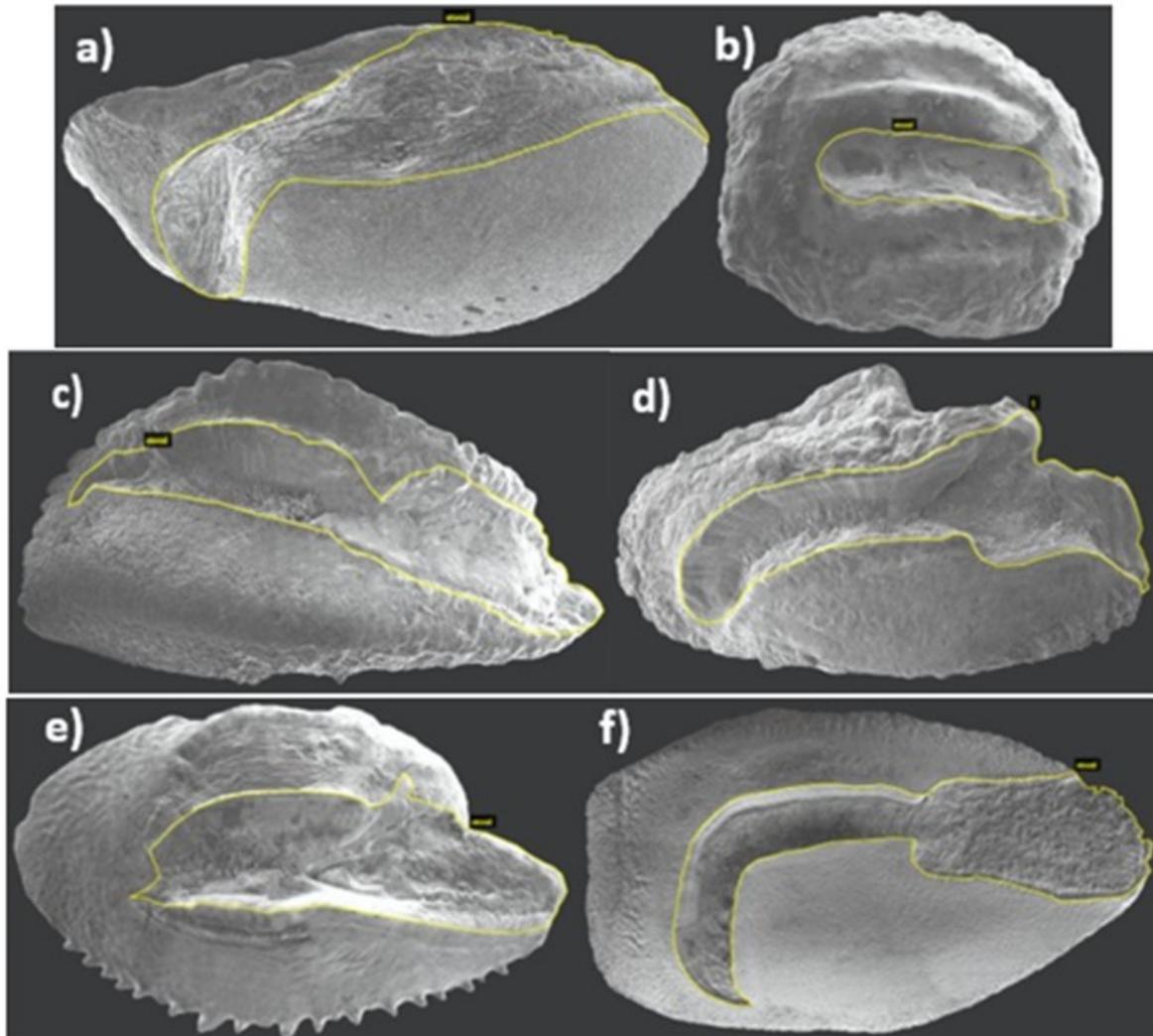


Figura 5.3: Otolitos segmentados de las especies: a) *Albula vulpes*, b) *Achirus lineatus*, c) *Calamus campechanus*, d) *Abudedefduf saxatilis*, e) *Anchoa cayorum*, f) *Anisotremus virginicus*.

En la imagen 5.4, se muestra el diagrama completo de la arquitectura utilizada en el modelo Mask R-CNN propuesto. En ella podemos ver las 4 etapas de Mask R-CNN descritas en la Tabla 4.3. La entrada inicial (input) es la imagen con *padding* (relleno) de ceros, la etapa 1 de *Backbone* está representada en la parte superior del esquema con las capas convolucionales de Resnet101 de la Tabla 5.3 (Figura 5.5), además de la adición de la pirámide de mapas de características de FPN para mejorar la extracción de características a diferentes escalas con el mismo conjunto de mapas de ResNet101. Cabe destacar que en los bloques de ResNet101 se realiza el proceso común de 3 capas, de esta arquitectura: 1 capas convolucionales, 2 capas de normalización por lotes y 3 capas de activación no lineal (Unidad lineal rectificada ReLU),

y la entrada a cada capa es la salida de la capa anterior.

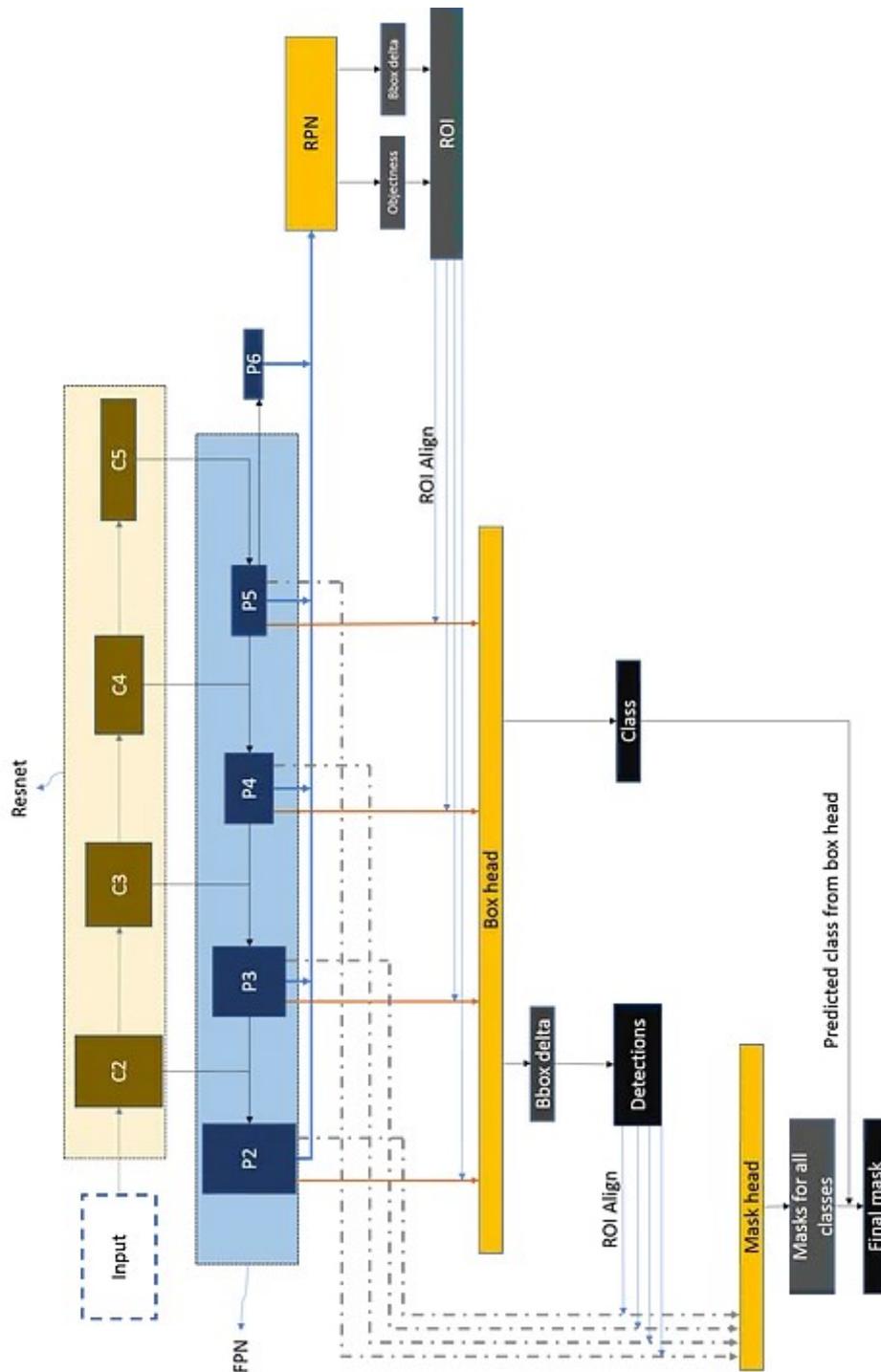


Figura 5.4: Diagrama de la arquitectura del modelo Mask R-CNN utilizado.

Capas de <i>Backbone</i>	
ResNet101	
Capa	Acción
Input (C1)	La imagen de entrada pasa a través de una capa convolucional con 64 filtros, tamaño de kernel de (7, 7) y un <i>strides</i> (zancada/paso) de (2, 2) que reduce las dimensiones espaciales en un factor de 2. El mapa de características de salida luego pasa a través de un capa de <i>pooling</i> (agrupación) máxima con 64 filtros, tamaño de kernel de (3, 3) y <i>strides</i> de (2, 2) reduciendo las dimensiones espaciales por un factor de 2 nuevamente.
C2	El mapa de características de la etapa anterior se pasa a través de tres bloques de filtros; los primeros 2 con 64 filtros y el tercero con 256 filtros, cada bloque contiene capas convolucionales con un tamaño de kernel de (3, 3) y un <i>strides</i> de (1, 1), con lo cual no cambian las dimensiones espaciales.
C3	El mapa de características de la etapa anterior pasa a través de cuatro bloques de filtrado; el primero con 128 filtros y los demás con 512 filtros, cada bloque contiene capas convolucionales con un tamaño de kernel de (3, 3) y un <i>strides</i> de (2, 2) reduciendo las dimensiones espaciales a la mitad.
C4	El mapa de características de la etapa anterior pasa a través de seis bloques de filtros; 256 filtros el primero y los restantes con 1024, cada bloque contiene capas convolucionales con un tamaño de kernel de (3, 3) y un <i>strides</i> de (2, 2) que reduce las dimensiones espaciales por un factor de 2.
C5	El mapa de características de la etapa anterior se pasa a través de tres bloques de filtrado; 512 filtros el primero y los demás 2048 filtros, cada bloque contiene capas convolucionales con un tamaño de kernel de (3, 3) y un <i>strides</i> de (2, 2). Esto reduce las dimensiones espaciales a la mitad.

Tabla 5.3: Capas convolucionales de *Backbone*.

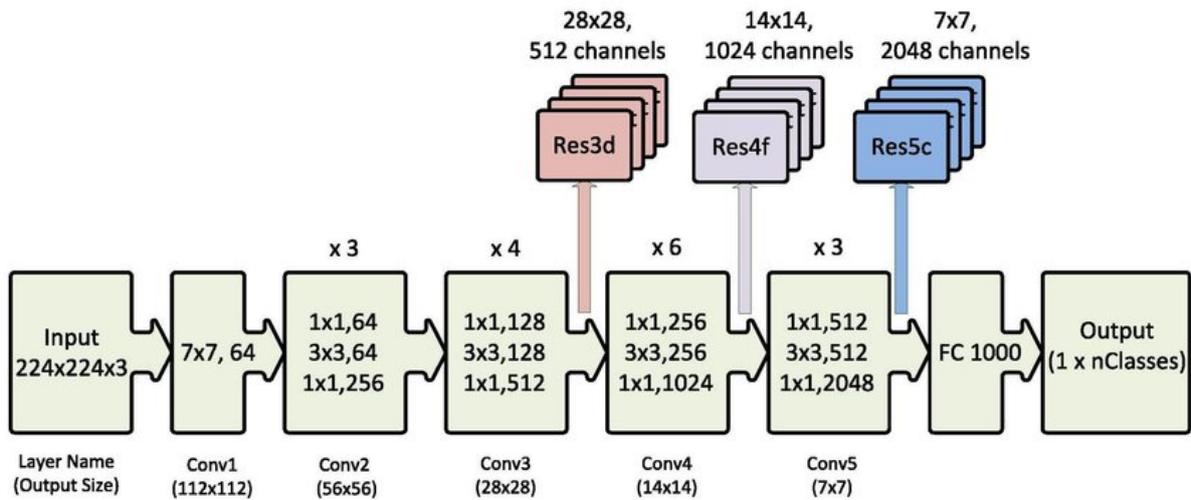


Figura 5.5: Diagrama del funcionamiento de ResNet101. Tomada de [67].

La etapa 2 (RPN) se encuentra en la parte derecha del diagrama en la que, como se explicó, reutiliza la extracción de características de FPN final (P6) para evitar cálculos duplicados y obteniendo los 2 resultados: objetos y cuadro delimitador que en conjunto forman las ROIs y pasar a la etapa 3 (parte media de 5.4, debajo de *Box head*) en la que se obtienen los 2 resultados: 1 deltas del cuadro delimitador y 2 clase a la que pertenece más refinados que los ROIs generados en la etapa 2. A las deltas del cuadro delimitador se les realiza la etapa 3.1 agrupando los ROIs con *ROI Align* que para cada propuesta de RPN, realiza interpolación bilineal para muestrear las ubicaciones exactas del mapa de características y permitir la alineación a nivel de subpíxel de la máscara. De esta forma en la última etapa se obtienen las máscaras encontradas y su clase perteneciente, en nuestro caso sólo sulco acústico escaladas al tamaño original de la imagen de entrada.

Los parámetros de entrenamiento utilizados en este trabajo, Tabla 5.4, se utilizaron para todas las instancias de la validación cruzada (K1, K2, K3, K4, K5). El cambio fue en los datos que se tomaron para validación y entrenamiento.

Párametro de entrenamiento	Valor
<i>Learning rate</i> (tasa de aprendizaje)	0.001
<i>Batch size</i> (tamaño de lote)	8
<i>Number of epochs</i> (Número de épocas)	35
<i>Number of classes</i> (número de clases)	1 (sulco acústico) + 1 (fondo)

Tabla 5.4: Parámetros de entrenamiento.

En las Tablas 5.5, 5.6, 5.7, 5.8 y 5.9 se muestran los resultados de las métricas en las 3

diferentes épocas (10, 20, 35) de las 5 ejecuciones (K's) del algoritmo resaltando los valores más altos de las métricas calculadas.

De K1 (Tabla 5.5) se destaca la estabilidad de las métricas jerarquizadas durante la época 35, sin olvidar que la época 10 tiene mejor *recall* y dentro de nuestra jerarquía es la segunda prioridad.

K1								
Época	TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10	9.387%	86.449%	2.759%	1.406%	0.791 ± 0.158	0.885 ± 0.078	0.821 ± 0.105	0.958 ± 0.028
20	9.272%	86.829%	2.379%	1.520%	0.810 ± 0.158	0.873 ± 0.083	0.827 ± 0.106	0.961 ± 0.027
35	9.362%	87.084%	2.123%	1.431%	0.824 ± 0.143	0.879 ± 0.076	0.840 ± 0.093	0.964 ± 0.024

Tabla 5.5: Promedio de las métricas de la primera ejecución (K1) del algoritmo con imágenes de validación

En la segunda prueba de entrenamiento (Tabla 5.6) se destaca gran inestabilidad en los resultados, sin embargo, de igual forma que en K1 la época 35 sobresale de las demás.

K2								
Época	TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10	9.136%	87.201%	2.007%	1.656%	0.834 ± 0.145	0.862 ± 0.081	0.835 ± 0.093	0.963 ± 0.025
20	9.224%	87.191%	2.017%	1.568%	0.832 ± 0.147	0.868 ± 0.077	0.838 ± 0.096	0.964 ± 0.024
35	9.259%	87.099%	2.108%	1.533%	0.828 ± 0.146	0.870 ± 0.075	0.837 ± 0.094	0.964 ± 0.025

Tabla 5.6: Promedio de las métricas de la segunda ejecución (K2) del algoritmo con imágenes de validación

Durante el tercer y cuarto entrenamiento, los resultados en la Tabla 5.7 y Tabla 5.8 revelan que son los 2 entrenamientos con resultados más variados, esto quiere decir que no muestran

regularidad en cuanto a los valores obtenidos. Es preciso mencionar de K4 que, a diferencia de los entrenamientos anteriores, la época 20 muestra mayor estabilidad.

K3								
Época	TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10	9.164%	86.934%	2.274%	1.628%	0.821 ± 0.152	0.863 ± 0.096	0.829 ± 0.109	0.961 ± 0.031
20	9.081%	87.060%	2.148%	1.711%	0.830 ± 0.160	0.857 ± 0.098	0.828 ± 0.113	0.961 ± 0.031
35	9.219%	86.955%	2.253%	1.573%	0.816 ± 0.157	0.867 ± 0.100	0.827 ± 0.111	0.962 ± 0.027

Tabla 5.7: Promedio de las métricas de la tercer ejecución (K3) del algoritmo con imágenes de validación

K4								
Época	TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10	9.225%	86.962%	2.246%	1.567%	0.818 ± 0.143	0.873 ± 0.083	0.833 ± 0.094	0.962 ± 0.026
20	9.246%	87.029%	2.179%	1.546%	0.825 ± 0.154	0.868 ± 0.080	0.834 ± 0.105	0.963 ± 0.029
35	9.337%	86.887%	2.321%	1.455%	0.814 ± 0.155	0.878 ± 0.083	0.832 ± 0.102	0.962 ± 0.026

Tabla 5.8: Promedio de las métricas de la cuarta ejecución (K4) del algoritmo con imágenes de validación

Po último, en K5 (Tabla 5.9), podemos observar que resulta el mejor entrenamiento, y con mayor coherencia, muestra que en la época 35 se tienen los mayores valores de las métricas, lo cual nos dice que hay lógica debido a que si se tiene mayor número de TP y TN, lógicamente se debería contar con el menor número de FP y FN; esto traducido a los mayores valores en las métricas de evaluación del modelo jerarquizadas *precision*, *recall*, *f1 score* y *accuracy*.

K5								
Época	TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10	9.148%	86.726%	2.481%	1.644%	0.808 ± 0.163	0.864 ± 0.097	0.821 ± 0.119	0.959 ± 0.034
20	9.230%	87.121%	2.087%	1.562%	0.829 ± 0.146	0.870 ± 0.083	0.837 ± 0.095	0.964 ± 0.025
35	9.326%	87.146%	2.062%	1.466%	0.832 ± 0.146	0.877 ± 0.078	0.842 ± 0.096	0.965 ± 0.026

Tabla 5.9: Promedio de las métricas de la quinta ejecución (K5) del algoritmo con las imágenes de validación

Después de analizar los resultados de la evaluación a los 5 entrenamientos de la red, se concluyó que el ejecutable de la época 35 de K5 es el modelo con mejor funcionamiento entre los comparados, por lo tanto sería en el que se probarían las imágenes de test (14), recordando que son nuevas para la red. Al igual que durante la etapa de entrenamiento, se obtuvieron las métricas de evaluación de la red (Tabla 5.10), tanto para cada imagen como un promedio de los 14 conjuntos de resultados. También se añadieron las máscaras de verdad, las máscaras generadas por la red y la superposición de ambas máscaras para calcular los índices de similitud en imágenes de la Tabla 4.7. En estos resultados vemos un gran desempeño en la segmentación automática del sulco acústico generada la red, y que a pesar de ser imágenes nuevas, los resultados son similares a los obtenidos durante la validación del entrenamiento. Se obtuvo la matriz de confusión Figura 5.6 junto con las métricas promediadas.

Resultados de las métricas con imágenes de test							
TP	TN	FP	FN	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Accuracy</i>
10.928%	84.307%	2.359%	2.406%	0.822 ± 0.108	0.823 ± 0.194	0.806 ± 0.143	0.952 ± 0.029

Tabla 5.10: Promedio de las métricas del ejecutable elegido con las 14 imágenes de validación

Mean results from 14 test images
 True positives: 203769.929 pixels, equal to: 10.928%
 True negatives: 2342716.429 pixels, equal to: 84.307%
 False positives: 36367.500 pixels, equal to: 2.359%
 False negatives: 160258.214 pixels, equal to: 2.406%
 Mean Precision: 0.822 +/- 0.108
 Mean Recall: 0.823 +/- 0.194
 Mean F1 Score: 0.806 +/- 0.143
 Mean Accuracy: 0.952 +/- 0.029
 Mean Confusion Matrix:
 [[203769.929 36367.5]
 [160258.214 2342716.429]]

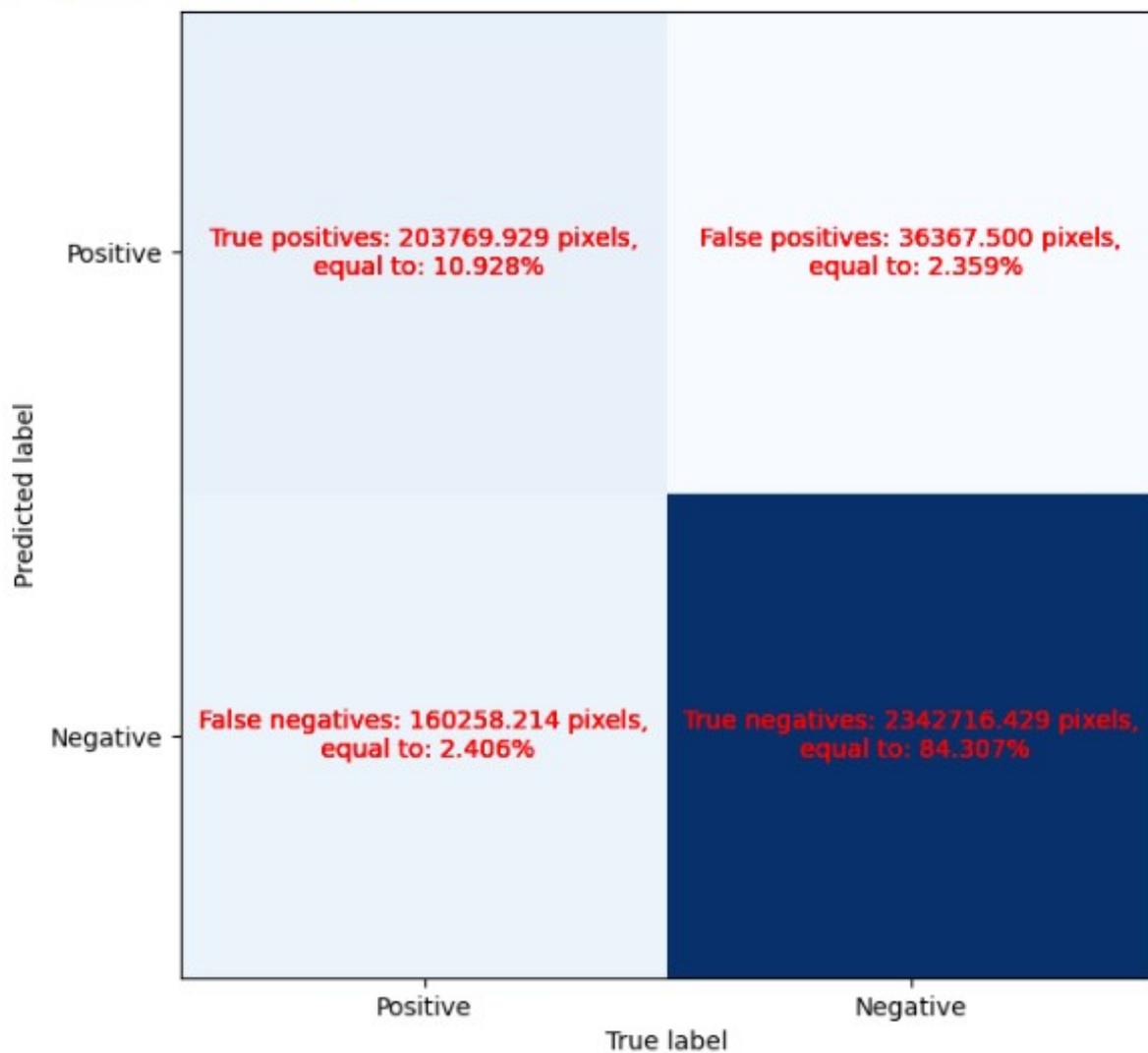


Figura 5.6: Resultados obtenidos en Google Colab del ejecutable con imágenes de test.

En la página 67 se presenta el Anexo A, el cual incluye los resultados de las 14 segmentaciones automáticas y sus métricas calculadas (Figura 3). Ahí podemos observar que 9 de las 14 segmentaciones (imágenes 1a, 1b, 1d, 2a, 2c, 3a, 3b, 3c y 3d) son muy precisas comparadas con la máscara de verdad, además de tener valores de todas las métricas similares entre sí y con el promedio de la validación durante el entrenamiento de la Tabla 5.9.

Las imágenes restantes del Anexo A; Figuras 1c, 1e, 1f, 2b, 3e, desde la vista gráfica (máscara y superposición) se logra apreciar un resultado no óptimo, aunque algo a destacar de todas las segmentaciones tanto en los procesos de validación durante el entrenamiento como en las de test, es que la métrica *accuracy* es superior, en la mayoría de los casos, a un valor de 0,900, lo cual nos indica que a pesar de no encontrar exactamente la máscara, la posición de la máscara (*bounding box*) está bien delimitada porque no encuentra píxeles fuera del sulco acústico pertenecientes a este.

Los errores en la detección exacta del sulco pueden deberse a diferentes factores, conforme a la experiencia en este trabajo, se mencionan a continuación, las principales causas y una posible solución a manera de trabajo futuro.

Posible causa del error	Posible solución
Falta de muestras diferentes (imágenes originales)	Se debe ampliar el catálogo de otolitos debido a que como se explico existen diferentes tipos de contorno, así como posiciones y aperturas del sulco acústico
Calidad de las imágenes	Las imágenes contienen ruido, además de la iluminación es y contraste de las imágenes utilizadas en este trabajo son diferentes desde su adquisición. Se podría normalizar la iluminación, mejorar contraste y eliminar posible ruido.
Arquitectura errónea	Al ser tan sofisticada, incluyendo segmentación semántica y detección de objetos. Se debería entrenar otra arquitectura de red para observar funcionamiento y compararlo con este trabajo.
Hiperparámetros	Los hiperparámetros definidos en los entrenamientos pueden afectar el rendimiento, deberían probarse diferentes valores de hiperparámetros.
Diferentes formas de otolito y sulco acústico	Relacionado con el primer punto de esta tabla, debería realizarse un aumento de datos mas igualitario entre las formas, aperturas y posiciones de las estructuras.

Tabla 5.11: Posibles adecuaciones para mejorar el funcionamiento del modelo Mask R-CNN.

La sección de resultados y discusión de esta investigación mostró que la segmentación automática del sulcus acusticus en otolitos *sagita* usando modelos de aprendizaje automático es posible y se obtuvieron resultados satisfactorios. El siguiente capítulo brinda una conclusión al trabajo y posibles trabajos a futuro, en relación a esta investigación.

Capítulo 6

Conclusión y trabajo futuro

Se concluye mediante el trabajo realizado que la segmentación del sulco acústico en otolitos *sagita* es posible realizarse mediante modelos de aprendizaje automático, logrando resultados adecuados de esta tarea. El uso de estos modelos para tareas que actualmente se realizan de forma manual, reduce en tiempo de realización cuando se trata de grandes cantidades de información (imágenes de muestras). No obstante se deben realizar mejoras como las planteadas en la sección de resultados 5.11 para lograr mejor desempeño en la segmentación.

Con la realización de este trabajo se logra comprender el trabajo futuro que se tiene:

- Mejorar el rendimiento obtenido, mediante la experimentación con otras arquitecturas o valores de los hiperparámetros.
- Abrir paso a segmentación de otras microestructuras, no sólo de los otolitos, sino también de otros objetos/materiales de tamaño microscópico.
- Desarrollando lo planteado al principio, se pueden comparar los resultados obtenidos e ir combinando métodos; de igual forma se puede comparar con segmentación manual.
- Para disminuir costos, se puede realizar un modelo pero utilizando otras técnicas de adquisición de imágenes microscópicas, como óptica, confocal, entre otras.
- Transferencia de aprendizaje, el ejecutable final se puede usar como entrada para nuevos entrenamientos.
- En cuanto a la arquitectura de Mask R-CNN, se puede estudiar profundamente para identificar posibles mejoras a la arquitectura, estudiando las características, ROIs, o decisiones que toma el modelo.

Referencias

- [1] Soria JA. On the automatic detection of otolith features for fish species identification and their age estimation. Universitat Politècnica de Catalunya; 2012. Available from: <http://www.tdx.cat/handle/10803/116814>.
- [2] Doering P, Ludwig J. Shape Analysis of Otoliths—a Tool for Indirect Ageing of Eel, *Anguilla anguilla* (L.). *Internationale Revue der gesamten Hydrobiologie und Hydrographie*. 1990;75:737-43. Available from: <https://doi.org/10.1002/iroh.19900750607>.
- [3] Campana SE, Casselman JM. Stock Discrimination Using Otolith Shape Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*. 1993 mayo;50(5):1062-83. Available from: <https://doi.org/10.1139/f93-123>.
- [4] McDougall A. Assessing the use of sectioned otoliths and other methods to determine the age of the centropomid fish, barramundi (*Lates calcarifer*) (Bloch), using known-age fish. *Fisheries Research*. 2004 abril;67(2):129-41. Available from: <https://doi.org/10.1016/j.fishres.2003.09.044>.
- [5] Steward CA, DeMaria KD, Shenker JM. Using otolith morphometrics to quickly and inexpensively predict age in the gray angelfish (*Pomacanthus arcuatus*). *Fisheries Research*. 2009 agosto;99(2):123-9. Available from: <https://doi.org/10.1016/j.fishres.2009.05.011>.
- [6] Freire KM, Oliveira CS, Rosa LC. Morphometric analysis of otoliths of juvenile crucifix sea catfish *Sciades proops* (Valenciennes, 1840). *Journal of Applied Ichthyology*. 2017 junio;33(3):485-90. Available from: <https://doi.org/10.1111/jai.13321>.
- [7] Bostanci D, Yilmaz M, Yedier S, Kurucu G, Kontas S, Darcin M, et al. Sagittal Otolith Morphology of Sharpsnout Seabream *Diplodus puntazzo* (Walbaum, 1792) in the Aegean Sea. *International Journal of Morphology*. 2016 junio;34(2):484-8. Available from: <https://doi.org/10.4067/S0717-95022016000200012>.

- [8] Yedier S, Bostanci D. Aberrant otoliths in four marine fishes from the Aegean Sea, Black Sea, and Sea of Marmara (Turkey). *Regional Studies in Marine Science*. 2020 febrero;34. Available from: <https://doi.org/10.1016/j.rsma.2019.101011>.
- [9] Tuset VM, Lombarte A, Assis CA. Otolith atlas for the western Mediterranean, north and central eastern Atlantic. *Scientia Marina*. 2008 julio;72(S1):7-198. Available from: <https://doi.org/10.3989/scimar.2008.72s17>.
- [10] Jaramillo AM, Tombari AD, Dura VB, Rodrigo ME, Volpedo AV. Otolith eco-morphological patterns of benthic fishes from the coast of Valencia (Spain). *Thalassas: An International Journal of Marine Sciences*. 2014 enero;30(1):57-66. Available from: https://www.researchgate.net/publication/262182362_Otolith_eco-morphological_patterns_of_benthic_fishes_from_the_coast_of_Valencia_Spain.
- [11] Lychakov DV, Rebane YT, Lombarte A, Fuiman LA, Takabashi A. Fish otolith asymmetry: Morphometry and modeling. *Hearing Research*. 2006 septiembre;219(1-2):1-11. Available from: <https://doi.org/10.1016/j.heares.2006.03.019>.
- [12] Tuset VM, Lozano IJ, González JA, Pertusa JF, García-Díaz MM. Shape indices to identify regional differences in otolith morphology of comber, *Serranus cabrilla* (L., 1758): Shape indices to identify regional differences in otolith. *Journal of Applied Ichthyology*. 2003 abril;19(2):88-93. Available from: <https://doi.org/10.1046/j.1439-0426.2003.00344.x>.
- [13] Vignon M. Ontogenetic trajectories of otolith shape during shift in habitat use: Interaction between otolith growth and environment. *Journal of Experimental Marine Biology and Ecology*. 2012 junio;420-421:26-32. Available from: <https://doi.org/10.1016/j.jembe.2012.03.021>.
- [14] Vignon M, Morat F. Environmental and genetic determinant of otolith shape revealed by a non-indigenous tropical fish. *Marine Ecology Progress Series*. 2010 julio;411:231-41. Available from: <https://doi.org/10.3354/meps08651>.
- [15] See M, Marsham S, Chang CW, Chong VC, Sasekumar A, Dhillon SK, et al. The Use of Otolith Morphometrics in Determining the Size and Species Identification of Eight Mulletts (Mugiliformes: Mugilidae) from Malaysia. *Sains Malaysiana*. 2016 mayo;45(5):735-43. Available from: https://www.researchgate.net/publication/305176788_The_use_of_otolith_morphometrics_in_determining_the_size_and_species_identification_of_eight_mulletts_Mugiliformes_Mugilidae_from_Malaysia.

- [16] Parisi-Baradad V, Manjabacas A, Lombarte A, Olivella R, Chic O, Piera J, et al. Automated Taxon Identification of Teleost fishes using an otolith online database—AFORO. *Fisheries Research*. 2010 junio;105(1):13-20. Available from: <https://doi.org/10.1016/j.fishres.2010.02.005>.
- [17] Salimi N, Loh KH, Dhillon SK, Chong VC. Fully-automated identification of fish species based on otolith contour: using short-time Fourier transform and discriminant analysis (STFT-DA). *PeerJ*. 2016 febrero;4:e1664. Available from: <https://doi.org/10.7717/peerj.1664>.
- [18] Montanini S, Stagioni M, Benni E, Vallisneri M. Ontogenetic changes in otolith morphology and shape analyses in *Chelidonichthys cuculus* (Linnaeus, 1758) and *Chelidonichthys lucerna* (L., 1758). *Journal of Applied Ichthyology*. 2017 abril;33(33):217-20. Available from: <https://doi.org/10.1111/jai.13299>.
- [19] Williams AJ, Newman SJ, Wakefield CB, Brunel M, Halafih T, Kaltavara J, et al. Evaluating the performance of otolith morphometrics in deriving age compositions and mortality rates for assessment of data-poor tropical fisheries. *ICES Journal of Marine Science*. 2015 octubre;72(7):2098-109. Available from: <https://doi.org/10.1093/icesjms/fsv042>.
- [20] de Carvalho BM, Spach HL, Vaz-Dos-Santos AM, Volpedo AV. Otolith shape index: is it a tool for trophic ecology studies? *Journal of the Marine Biological Association of the United Kingdom*. 2019 noviembre;99(7):1675-82. Available from: <https://doi.org/10.1017/S0025315419000729>.
- [21] Bermejo S, Monegal B, Cabestany J. Fish age categorization from otolith images using multi-class support vector machines. *Fisheries Research*. 2007 abril;84(2):247-53. Available from: <https://doi.org/10.1016/j.fishres.2006.11.021>.
- [22] Fablet R. Semi-local extraction of ring structures in images of biological hard tissues: application to the Bayesian interpretation of fish otoliths for age and growth estimation. *Canadian Journal of Fisheries and Aquatic Sciences*. 2006 junio;63(6):1414-28. Available from: <https://doi.org/10.1139/f06-001>.
- [23] Fablet R. Statistical learning applied to computer-assisted fish age and growth estimation from otolith images. *Fisheries Research*. 2006 noviembre;81(2):219-28. Available from: [10.1016/j.fishres.2006.07.013](https://doi.org/10.1016/j.fishres.2006.07.013).

- [24] Mahe K, Parisi V, Carbini S, Soria JA, Harbitz A, de Pontual H, et al. Automatic Fish Ageing (AFISA). In: 4th International Otolith Symposium. Archive institutionnelle de l'Ifremer; 2009. p. 24-8.
- [25] Caillault EP, Vanhelst F, Bastian C, Mahe K. Otolith age estimation by machine learning approaches. In: 6th International Otolith Symposium 2018m. HAL Open Science; 2018. .
- [26] Moen E, Handegard NO, Allken V, Albert OT, Harbitz A, Malde K. Automatic interpretation of otoliths using deep learning. PLOS ONE. 2018 diciembre;13(12):e0204713. Available from: <https://doi.org/10.1371/journal.pone.0204713>.
- [27] Muñoz A, Suárez A. In: Maglogiannis I, Iliadis L, Hammer B, Manolopoulos Y, Kurková V, editors. Directional Data Analysis for Shape Classification. Cham: Springer International Publishing; 2018. p. 598-607. Available from: https://doi.org/10.1007/978-3-030-01418-6_59.
- [28] Fisher M, Hunter E. Digital imaging techniques in otolith data capture, analysis and interpretation. Marine Ecology Progress Series. 2018 junio;598:213-31. Available from: <https://doi.org/10.3354/meps12531>.
- [29] Nidiyare Hevia-Montiel AGTMBAXCC Jorge Perez-Gonzalez. Invariant morphological descriptors from otolith shape in environment automatic classification. Wiley Online Library. 2021 abril;37(4):534-44. Available from: <https://doi.org/10.1111/jai.14207>.
- [30] IIMAS. Base Datos Tesis Jesús - Archivos - Nextcloud; 2022. Available from: <https://datayucatan.iimas.unam.mx/index.php/apps/files/?dir=/Base%20Datos%20Tesis%20Jes%C3%BA&fileid=256543>.
- [31] Kasumyan A. Structure and function of auditory system in fish. Journal of Ichthyology. 2005 diciembre;45(2):S223-70. Available from: https://www.researchgate.net/publication/264977023_Structure_and_function_of_auditory_system_in_fish.
- [32] Campana SE. Photographic Atlas of Fish Otoliths of the Northwest Atlantic Ocean. NRC Research Press; 2004.
- [33] Moreno AP. Los Otolitos: Las piedras del Equilibrio; 2021. Accessed: 2022-05-05. <http://bloglemu.blogspot.com/2021/08/los-otolitos-las-piedras-del-equilibrio.html>.

- [34] Leta HR, Keim A. Lectura de edades en otolitos de peces teleosteos;. Accessed: Febrero 27, 2023. <https://www.fao.org/3/AC567S/AC567S03.htm>.
- [35] Stevenson D, Campana S. Otolith Microstructure Examination and Analysis. Canadian Special Publication of Fisheries and Aquatic Sciences; 1992. Available from: <https://doi.org/10.13140/RG.2.2.22258.61127>.
- [36] Mascarenas I, Aburto-Oropeza O, Balart E. Otolitos de peces de arrecife del Golfo de California. Universidad Autónoma de Baja California Sur Centro de Investigaciones Biológicas del Noroeste; 2003. Available from: https://www.researchgate.net/publication/285649113_Otolitos_de_peces_de_arrecife_del_Golfo_de_California.
- [37] Fortunato RGC. Aplicación de los otolitos en el estudio de la interconectividad de stocks pesqueros de mugílidos de zonas costeras de la provincia de Buenos Aires, Argentina y de la costa Valenciana, España. Universitat Politècnica de València; 2017. Available from: <https://riunet.upv.es/handle/10251/79549>.
- [38] Luell S. The importance of otoliths in fisheries biology;. Accessed: Febrero 27, 2023. <https://habitat.fisheries.org/the-importance-of-otoliths-in-fisheries-biology/>.
- [39] EasyMicroscopes. What Are the Uses and Applications of Microscopy in Science?;. Accessed: Febrero 27, 2023. <https://easymicroscopes.com/blog/what-are-the-uses-and-applications-of-microscopy-in-science/#:~:text=Microscopy%20is%20considered%20an%20important%20field%20as%20it,of%20crucial%20compounds%20in%20plants%20or%20other%20organisms>.
- [40] RAE A. imagen | Diccionario de la lengua española;. Accessed: Mayo 6, 2022. <https://dle.rae.es/imagen>.
- [41] Gonzalez RC, Woods RE, Masters BR. Digital Image Processing. Journal of Biomedical Optics. 2009;14(2):029901. Available from: <https://doi.org/10.1117/1.3115362>.
- [42] Rojas R. Multimedia: Unidad de Competencia IV. Imagen.;. Accessed: Mayo 6, 2022. https://ri.uaemex.mx/bitstream/handle/20.500.11799/103085/secme-35716_1.pdf?sequence=1&isAllowed=y.

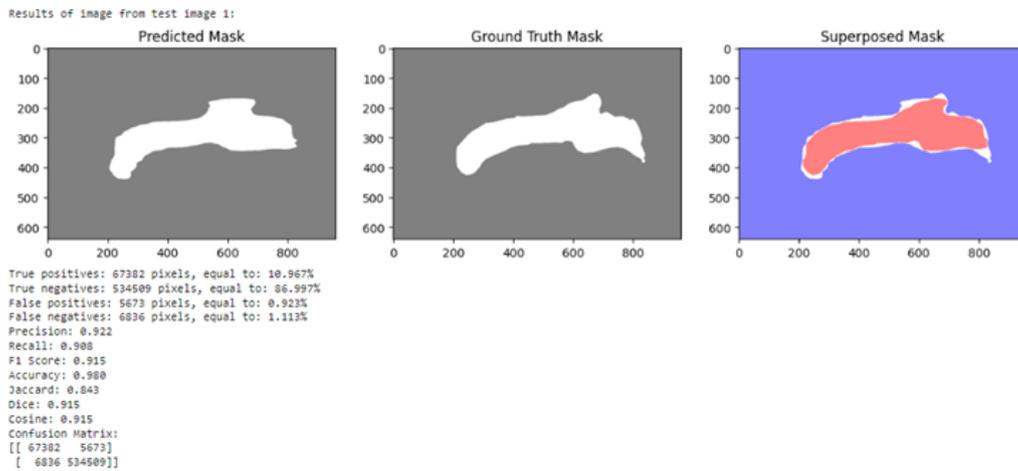
- [43] Sebastian Raschka VM. Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2, 3rd Edition. Packt Publishing; 2003. Available from: https://www.researchgate.net/publication/285649113_Otolitos_de_peces_de_arrecife_del_Golfo_de_California.
- [44] Szeliski R. Computer Vision: Algorithms and Applications. Springer Cham; 2022.
- [45] Stevens K. The Vision of David Marr. Perception. 2012 septiembre;41:1061-72. Available from: <https://doi.org/10.1068/p7297>.
- [46] Shorten C, Khoshgofaar TM. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data. 2019 julio;6(1):60. Available from: <https://doi.org/10.1068/p7297>.
- [47] Wamihd KM, Shaker KA, Zahoor MA, Bahaa HT. Feature Extraction Methods: A Review. Journal of Physics: Conference Series. 2020:1591. Available from: <https://doi.org/10.1088/1742-6596/1591/1/012028>.
- [48] Solawetz J. Object detection for computer vision;. Accessed: Febrero 27, 2023. <https://blog.roboflow.com/object-detection/>.
- [49] Walia M. Difference Between Semantic and Instance Segmentation;. Accessed: Febrero 27, 2023. <https://blog.roboflow.com/difference-semantic-segmentation-instance-segmentation/>.
- [50] Mitate E, Inoue K, Sato R, Shimomoto Y, Ohba S, Ogata K, et al. Application of the sliding window method and Mask-RCNN method to nuclear recognition in oral cytology. Diagnostic Pathology. 2022 agosto;17(1):62. Available from: <https://doi.org/10.1186/s13000-022-01245-0>.
- [51] Li YZ, Wang Y, Fang KB, Zheng HZ, Lai QQ, Xia YF, et al. Automated meniscus segmentation and tear detection of knee MRI with a 3D mask-RCNN. European Journal of Medical Research. 2022 noviembre;27(1):247. Available from: <https://doi.org/10.1186/s40001-022-00883-w>.
- [52] Alam M, Wang JF, Guangpei C, Yunrong LV, Chen Y. Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. Mobile Networks and Applications. 2021 febrero;26(1):200-15. Available from: <https://doi.org/10.1007/s11036-020-01703-3>.

- [53] Ulku I, Akagunduz E. A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images. Taylor Francis Group. 2022 febrero;36(1). Available from: <https://doi.org/10.1080/08839514.2022.2032924>.
- [54] Guo Y, Liu Y, Georgiou T, Lew MS. A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval. 2018 junio;7(2):87-93. Available from: <https://doi.org/10.1007/s13735-017-0141-z>.
- [55] Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J. In: Tolosana R, Vera-Rodriguez R, Rathgeb C, Busch C, editors. An Introduction to Digital Face Manipulation. Cham: Springer International Publishing; 2022. p. 3-26. Available from: https://doi.org/10.1007/978-3-030-87664-7_1.
- [56] Rombach R, Esser P, Blattmann A, Ommer B. In: Fingscheidt T, Gottschalk H, Houben S, editors. Invertible Neural Networks for Understanding Semantics of Invariances of CNN Representations. Cham: Springer International Publishing; 2022. p. 197-224. Available from: https://doi.org/10.1007/978-3-031-01233-4_7.
- [57] Hou L, Pandey HM. Research on the Positioning Technology of Sports 3D Teaching Action Based on Machine Vision. Mobile Networks and Applications. 2022 diciembre;27(6):2419-28. Available from: <https://doi.org/10.1007/s11036-021-01885-4>.
- [58] Viera M. Características ecomorfológicas de los otolitos sagitta de *Genypterus Blacodes* y *Genypterus Brasiliensis* provenientes de la zona común de pesca Argentino-Uruguaya. Conocimiento Libre Repositorio Institucional. 2011. Available from: <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/1364>.
- [59] National-Geographic. Las consecuencias del exceso de pesca;. Accessed: Junio 13, 2022. https://www.nationalgeographic.com.es/mundo-ng/grandes-reportajes/exceso-pesca-consecuencias_3459.
- [60] Pérez JAM, León MRKM, Tafolla BF, Alemán MB, Torres AG, Chiappa-Carrara X. Catálogo de Otolitos Sagitta de peces del Golfo de México;. Accessed: Mayo 17, 2022. <http://www.biocon.unam.mx/grupo-de-trabajo/listado-de-publicaciones/libros/martinez-perez-ja-morquecho.html>.
- [61] FractaIScale. Mask R-CNN Unmasked;. Accessed: Febrero 27, 2023. <https://medium.com/@fractal.ai/mask-r-cnn-unmasked-28419817a990>.

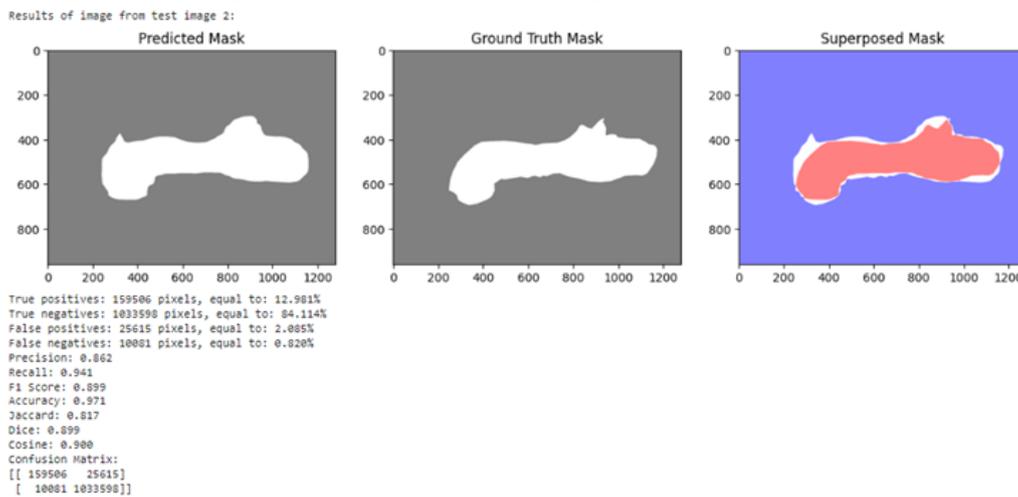
- [62] Abdulla W. Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow;. Accessed: Febrero 27, 2023. <https://acortar.link/Jm06tI>.
- [63] Canu S. Train Mask R-CNN for Image Segmentation (online free gpu);. Accessed: Mayo 6, 2022. <https://pysource.com/2021/08/10/train-mask-r-cnn-for-image-segmentation-online-free-gpu/>.
- [64] Abdulla W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. Github; 2017. https://github.com/matterport/Mask_RCNN.
- [65] Bobba R. Taming the Hyper-Parameters of Mask RCNN;. Accessed: Febrero 27, 2023. <https://medium.com/analytics-vidhya/taming-the-hyper-parameters-of-mask-rcnn-3742cb3f0e1b>.
- [66] Kumar A. Accuracy, Precision, Recall F1-Score - Python Examples;. Accessed: Febrero 27, 2023. <https://vitalflux.com/accuracy-precision-recall-f1-score-python-example/>.
- [67] Fang D. Answer to Where is faster rcnn resnet101 defined in tensorflow. Stackoverflow; 2019. <https://stackoverflow.com/questions/55680426/>.

Anexos

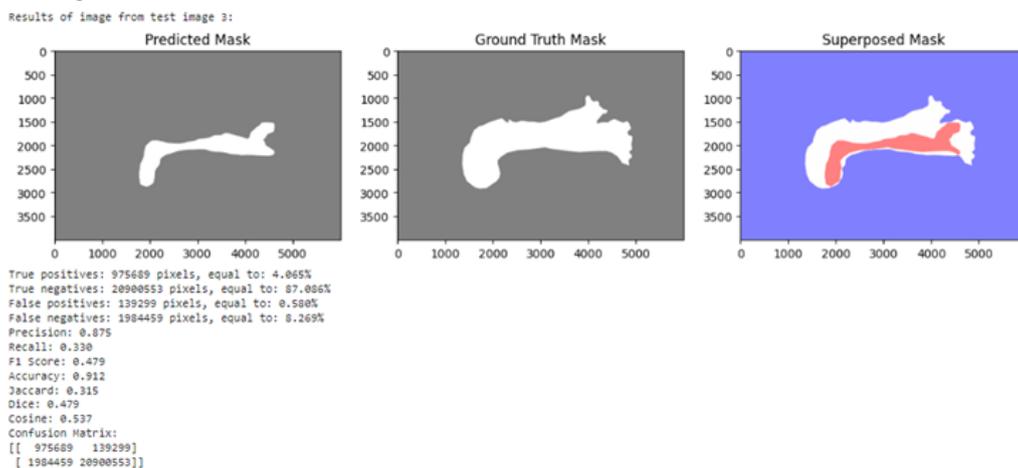
Anexo A



(a) Resultado de la segmentación automática de la imagen de test 1. Especie *Abudefduf saxatilis*.

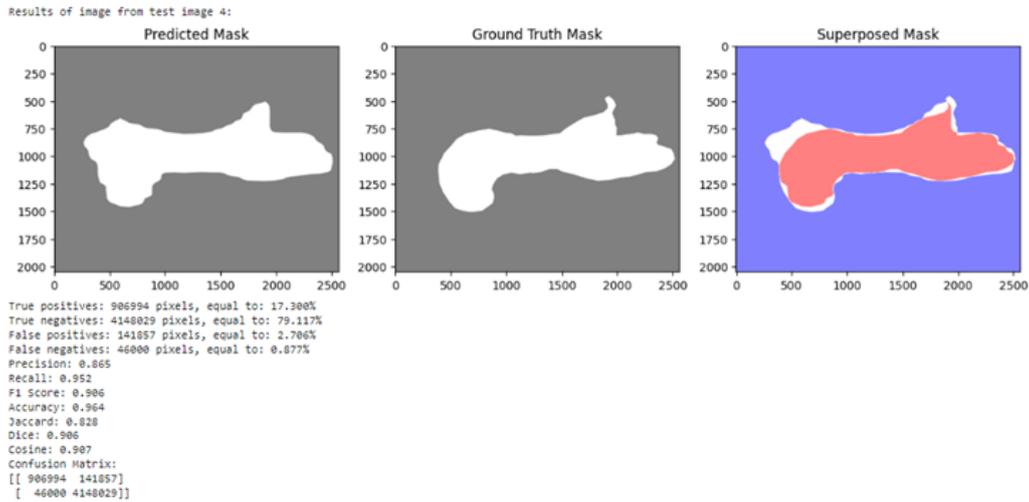


(b) Resultado de la segmentación automática de la imagen de test 2. Especie *Acanthurus chirurgus*.

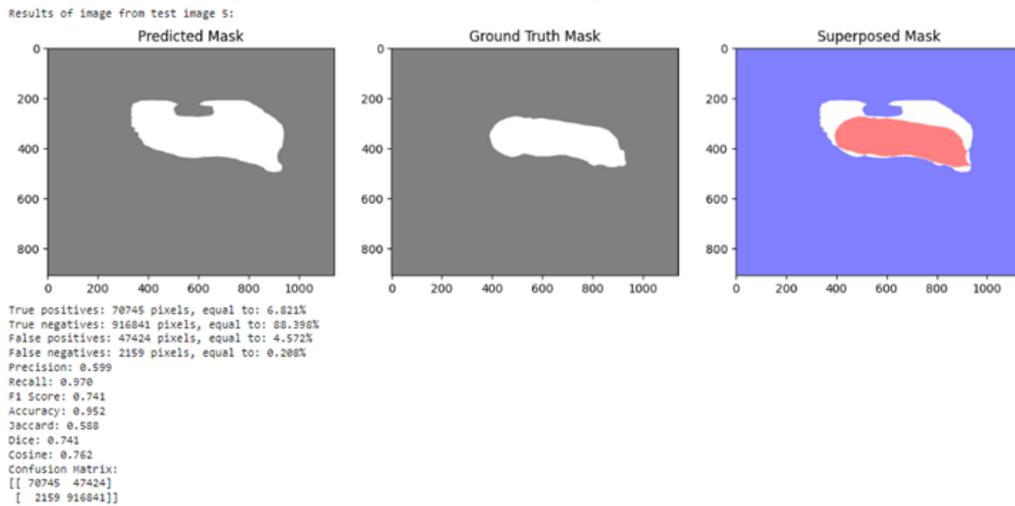


(c) Resultado de la segmentación automática de la imagen de test 3. Especie *Acanthurus coeruleus*.

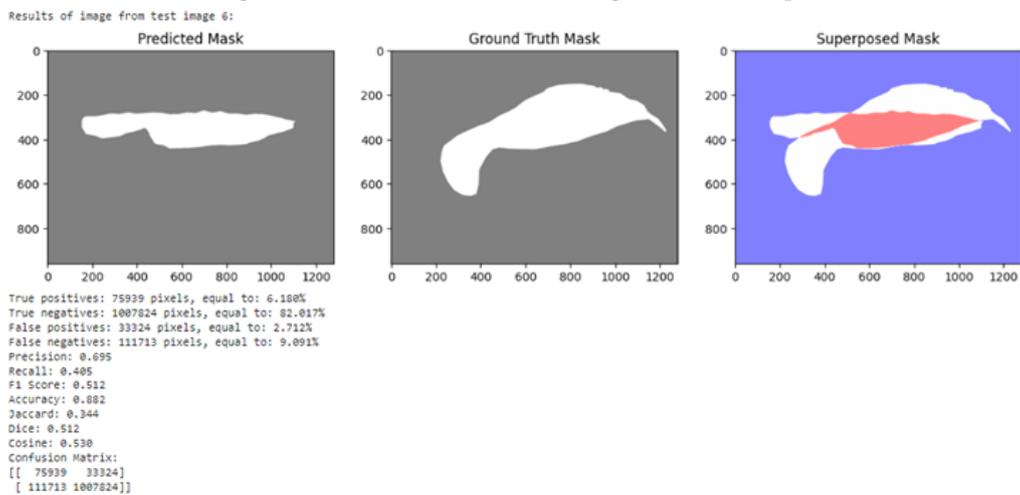
Figura 1: Resultados de la segmentación automática de las 14 imágenes de test en el ejecutable.



(d) Resultado de la segmentación automática de la imagen de test 4. Especie *Acanthurus tractus*.

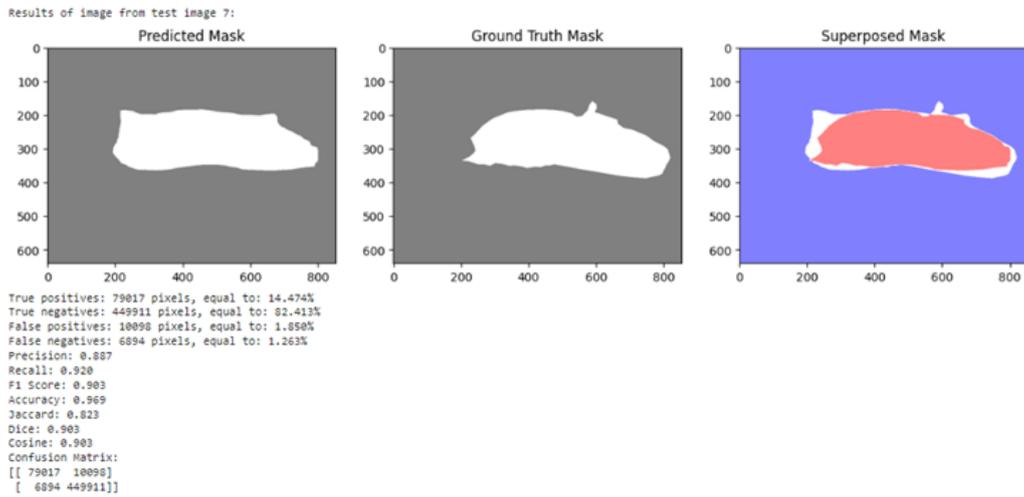


(e) Resultado de la segmentación automática de la imagen de test 5. Especie *Achirus lineatus*.

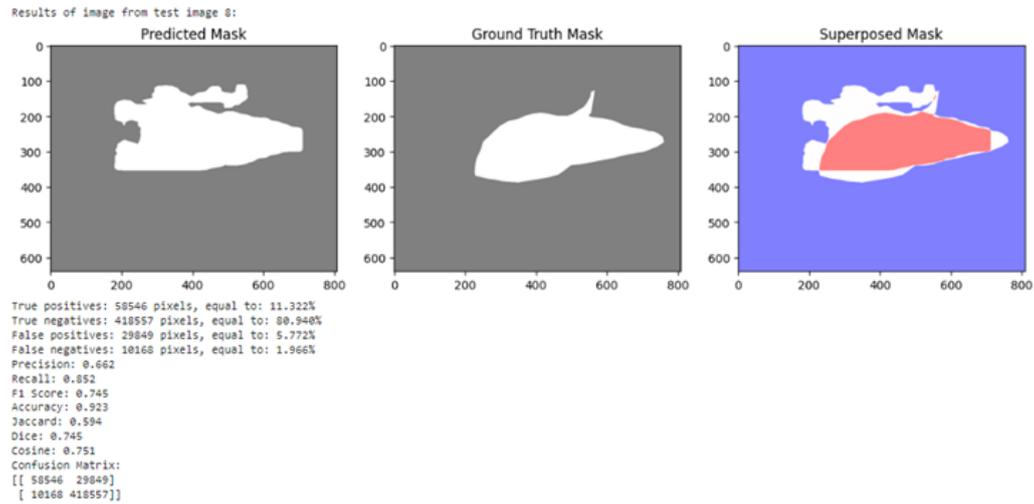


(f) Resultado de la segmentación automática de la imagen de test 6. Especie *Albula vulpes*.

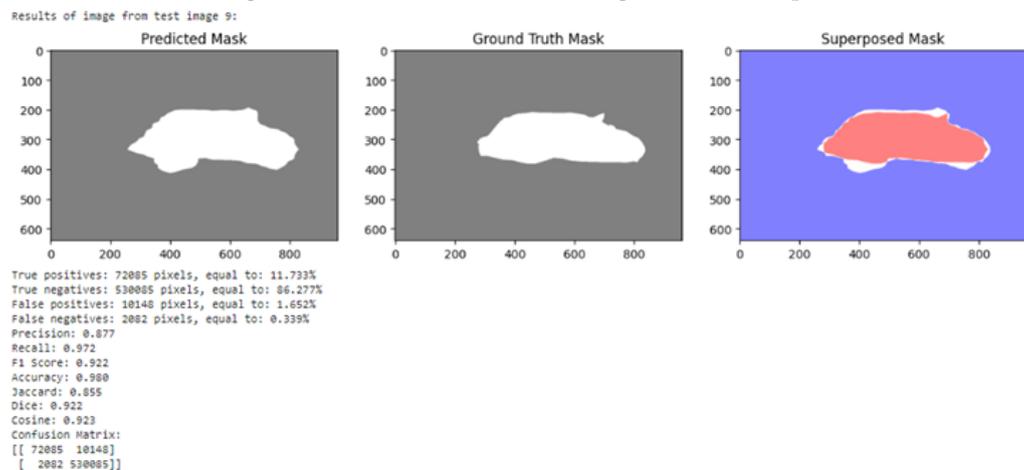
Figura 1: Resultados de la segmentación automática de las 14 imágenes de test en el ejecutable. (continuación).



(a) Resultado de la segmentación automática de la imagen de test 7. Especie *Anchoa cayorum*.

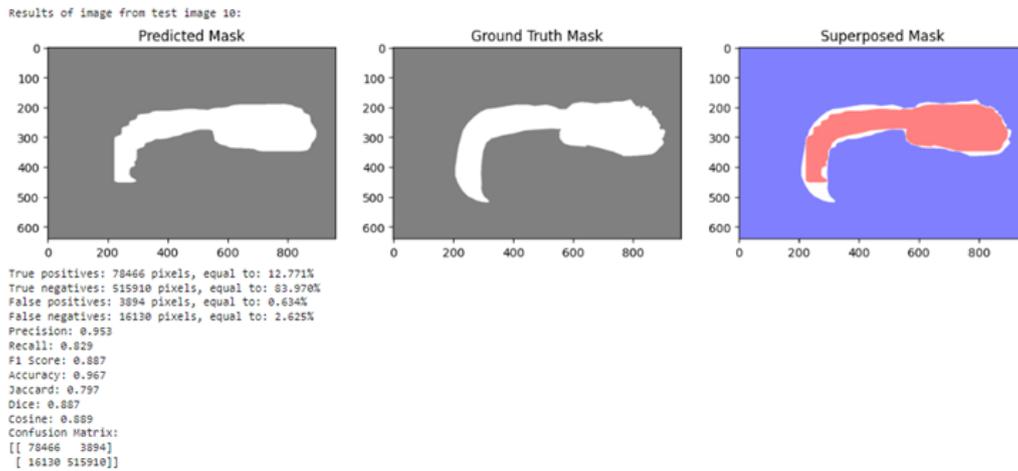


(b) Resultado de la segmentación automática de la imagen de test 8. Especie *Anchoa cubana*.

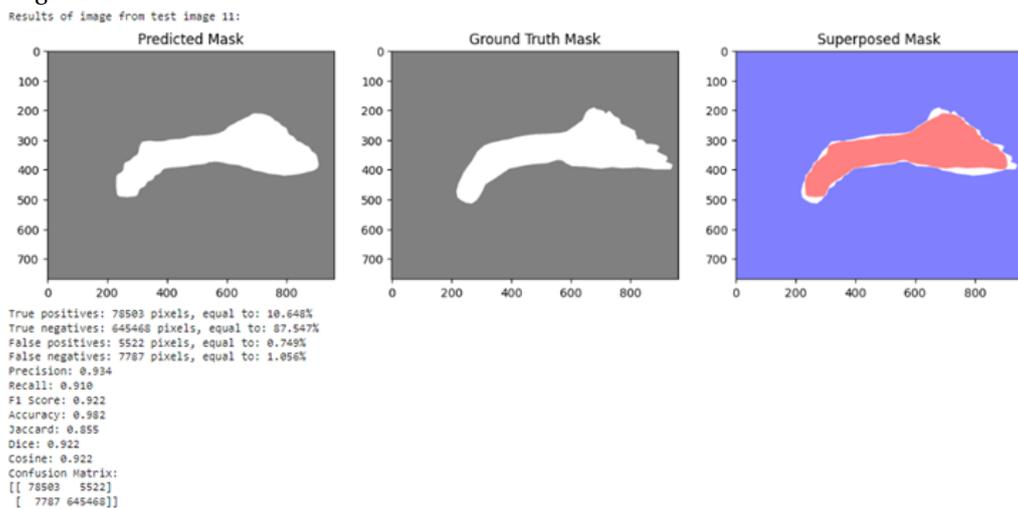


(c) Resultado de la segmentación automática de la imagen de test 9. Especie *Anchoa lamprotaenia*.

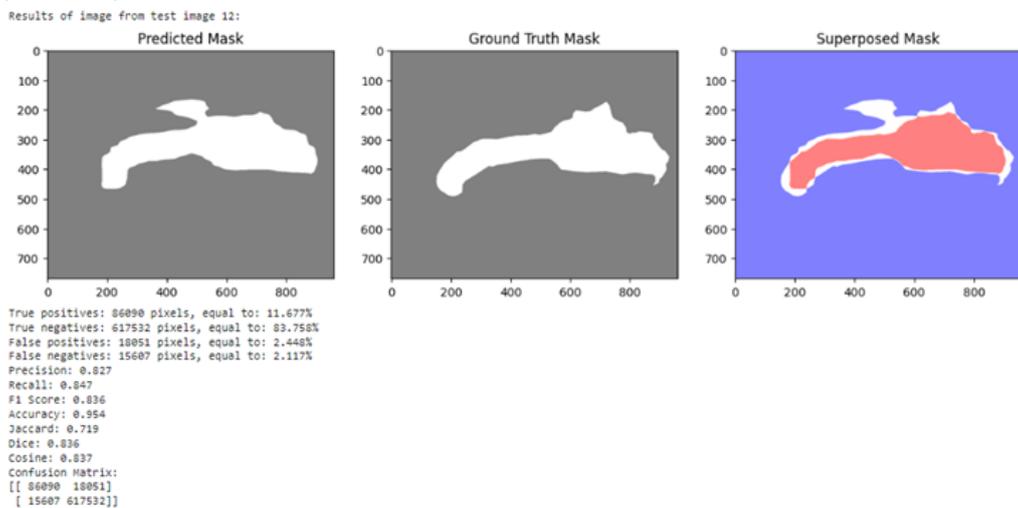
Figura 2: Resultados de la segmentación automática de las 14 imágenes de test en el ejecutable.



(a) Resultado de la segmentación automática de la imagen de test 10. Especie *Anisotremus virginicus*.

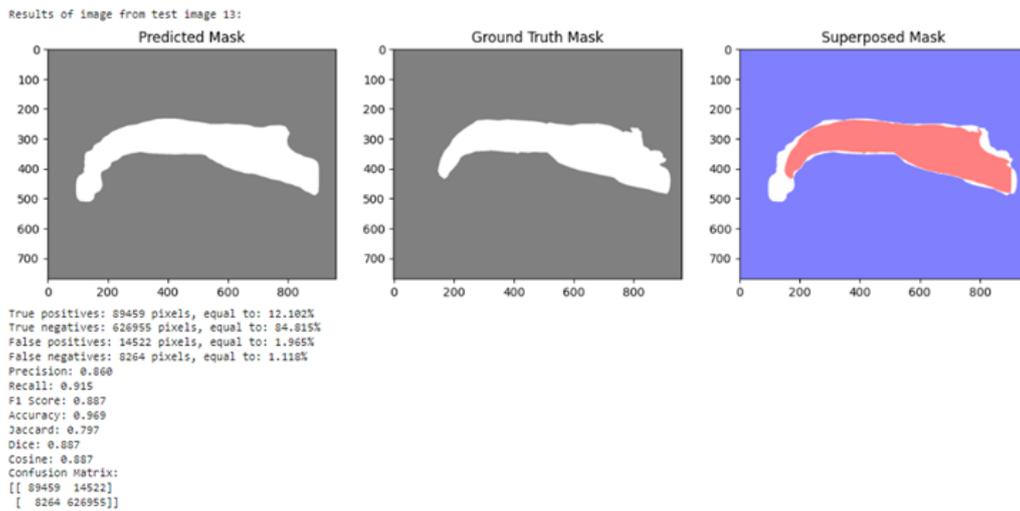


(b) Resultado de la segmentación automática de la imagen de test 11. Especie *Archosargus probatocephalus*.

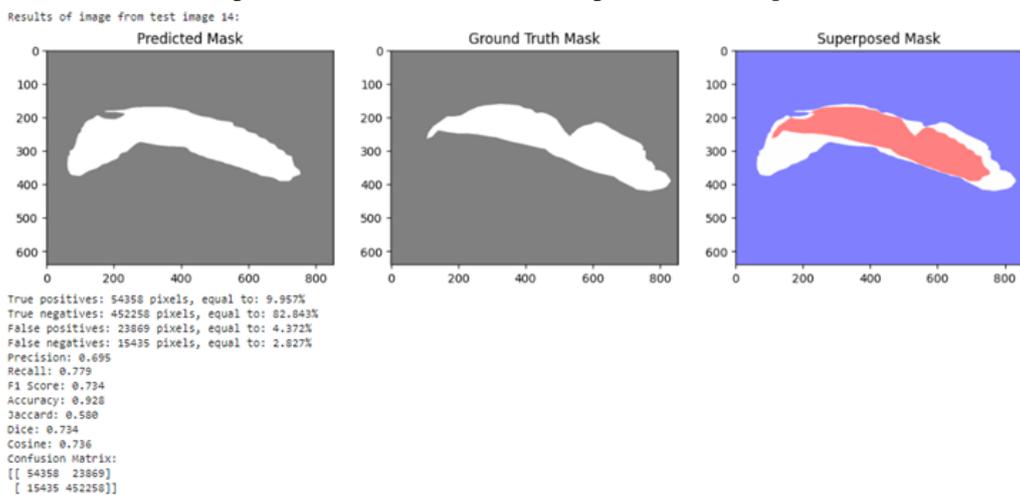


(c) Resultado de la segmentación automática de la imagen de test 12. Especie *Arcosargus rhomboidalis*.

71
Figura 3: Resultados de la segmentación automática de las 14 imágenes de test en el ejecutable.



(d) Resultado de la segmentación automática de la imagen de test 13. Especie *Calamus calamus*.



(e) Resultado de la segmentación automática de la imagen de test 14. Especie *Calamus campechanus*.

Figura 3: Resultados de la segmentación automática de las 14 imágenes de test en el ejecutable. (continuación).