



Escuela
Politécnica
Superior

Identificación, tallado y segmentación de especies de pesca mediante visión artificial y deep learning en entornos con alto solapamiento



Máster Universitario en Ciencia de
Datos

Trabajo Fin de Máster

Autor:
Alejandro Galán Cuenca
Tutor/es:
Andrés Fuster Guilló
Jorge Azorín López



Universitat d'Alacant
Universidad de Alicante

Junio 2023

Identificación, tallado y segmentación de especies de pesca mediante visión artificial y deep learning en entornos con alto solapamiento

Autor

Alejandro Galán Cuenca

Directores

Andrés Fuster Guilló

Jorge Azorín López

Tecnología Informática y Computación



MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS



Escuela
Politécnica
Superior



Universitat d'Alacant
Universidad de Alicante

ALICANTE, 8 de junio de 2023

Information's strength lies in its trustworthy sources.

ChatGPT, OpenAI

Resumen

En este trabajo de fin de máster se propone una continuación a un sistema para la identificación y segmentación automática de especies en lonja mediante visión por computador y deep learning. Se abordan propuestas para un tallado de especies y cálculo de biomasa en entornos calibrados y no calibrados, además de la adaptación del sistema a un dominio de lonja mayorista, con una propuesta de cambio de arquitectura de red para mejorar los resultados en el nuevo dominio.

Los métodos propuestos se han probado con más de 3000 imágenes adquiridas en entornos reales de las lonjas del Campello y Altea, obteniendo resultados que permiten una implementación y un correcto desarrollo en un ambiente real. El presente trabajo pretende mejorar la gestión pesquera, contribuyendo a la preservación del ecosistema marino y favoreciendo la economía local.

Índice general

1	Introducción	1
1.1	Motivación y contexto	1
1.2	Estado del arte	7
1.2.1	Tallado de especies en entornos sin posibilidad de calibración	8
1.2.2	Segmentación de especies en entornos con alto nivel de solape	11
1.3	Objetivos	17
2	Antecedentes metodológicos	19
2.1	Tallado en entornos no calibrados	19
2.1.1	Errores en las medidas ground truth	20
2.1.2	Método propuesto	24
2.1.3	Parámetros de entrada del modelo de regresión	25
2.2	Identificación y segmentación de especies en entornos mayoristas. Método <i>KeyFish</i>	29
2.2.1	Método propuesto	32
2.2.2	Adaptación de la arquitectura de red	34
2.2.3	Adaptación del dataset	35
2.3	Obtención de biomasa	36
2.4	Tallado en un entorno calibrado y aplicación del método <i>KeyFish</i>	38
2.4.1	Diferencias respecto del entorno minorista no calibrado	38
2.4.2	Calibrado de imágenes	40

3	Experimentación	45
3.1	Tallado de especies en entorno no calibrado	45
3.2	Identificación y segmentación en entornos mayoristas	51
3.2.1	Comparativa de método base y no penalizado	58
3.3	Tallado en entorno calibrado y método de segmentación e identificación <i>KeyFish</i>	64
4	Resultados	65
4.1	Tallado en entorno no calibrado	65
4.2	Identificación y segmentación en entorno mayorista	67
4.3	Tallado en entorno calibrado con <i>KeyFish</i>	69
5	Conclusiones	71
5.1	Conclusión	71
5.2	Líneas futuras	74
	Bibliografía	82

Índice de figuras

1.1	Estructura edge de implementación del sistema en el entorno real (b) y su implementación en la lonja (a, c).	5
1.2	Ejemplos del trabajo de Fernandes et al. [1]. Etiquetado semántico de la especie, el fondo y aletas.	10
1.3	Arquitectura y ejemplos del modelo implementado en [2].	11
1.4	Ejemplos de <i>conjuntos homogéneos de objetos (HOC)</i> [3].	12
1.5	Etiquetado parcial de [4].	13
1.6	Ejemplos de imágenes para etiquetado parcial.	14
1.7	Ejemplo de capas amodales [5].	15
1.8	Ejemplo multi-etiquetado y su comparativa con otros métodos [6].	16
2.1	Muestra de una bandeja etiquetada mediante la versión modificada de [7]. Se marcarán las referencias reales correspondientes a las cuatro esquinas de la bandeja.	20
2.2	Dispersión e histograma de las especies con base en el tamaño ground truth obtenido mediante homografía.	21
2.3	Error de etiquetado humano, confusión en la selección de categoría de etiqueta.	22
2.4	Equina de bandeja mal etiquetada, contiene dos puntos juntos que dan lugar a una mala corrección de imagen.	22

2.5	Error de etiquetado humano, confusión en la ubicación de las referencias de la bandeja. Deben de estar en las asas, no en las esquinas.	23
2.6	Flujo de trabajo general de la metodología propuesta para la estimación del tamaño de los peces. La parte superior (zona azul) muestra el flujo de entrenamiento de la red neuronal [2] a partir del etiquetado de imágenes, y del modelo de regresión a partir de las salidas de la red y las imágenes corregidas mediante metrología visual. La parte inferior (amarillo) muestra el bloque de uso en entorno real, es decir, no se dispone de <i>ground truth</i> y la predicción de los dos bloques será la salida. A partir de estos cuatro elementos, se calculará la biomasa.	26
2.7	Área de segmentación (en azul) como conjunto de píxeles representando la zona perteneciente a una clase concreta. Se trata de una de las salidas de la red neuronal y de las entradas del modelo de regresión.	28
2.8	Representación del bounding box de una especie. Punto inicial (x, y) como esquina superior izquierda y punto final como esquina inferior derecha (x, y), ambos representados por una \mathbf{X} azul. Ancho y alto calculados a partir de estos puntos, representados a partir de las flechas verdes.	29
2.9	Diferentes retos encontrados en la lonja mayorista.	32
2.10	Comparativa del uso de la posición inicial del bounding box junto con ancho y alto (2.10a) o la supresión de este (2.10b).	39
2.11	Ejemplos de calibrado de cámara con chessboard.	41
2.12	Ejemplos de imágenes no calibradas (a, c, d, g, i, j) y sus respectivas calibradas (b, e, f, h, k, l).	42
2.13	Ejemplos de especímenes en la zona inferior (a) y superior (b) de la bandeja.	43
2.14	Imágenes del <i>chessboard</i> en zona alta y baja de la bandeja, preparadas para la obtención de los factores de conversión píxel-centímetro límite del rango de error de talla y biomasa.	44
3.1	Ejemplo de <i>Sphyraena sphyraena</i> en el dataset.	49
3.2	Entorno de etiquetado <i>Django Labeller</i> [7] modificado para este entorno.	52

3.3	Frecuencia de cada especie en el dataset.	53
3.4	Ejemplo de bandejas con silueta negra representando las instancias etiquetadas, el ground truth que recibirá la red.	53
3.5	Cantidad de bandejas por cada especie en el dataset ordenadas por frecuencia de aparición en este.	55
3.6	Media y desviación estándar de cada especie por bandeja.	56
3.7	Gráficas de validación en el entrenamiento. Las condiciones no son comparables, para el cálculo del método <i>No Penalización</i> se ha calculado la precisión con las N instancias coincidentes con el ground truth como estipula su funcionamiento. Para el método base se ha comparado la precisión con todas las predicciones, como se espera de este.	60
3.8	Ejemplo de <i>Illex condetii</i> y <i>Loligo vulgaris</i>	61
3.9	Comparativa entre <i>Yolact Base</i> y <i>No Penalización</i> para 50 % de umbral para el bounding box.	61
3.10	Comparativa entre <i>Yolact Base</i> y <i>No Penalización</i> para 50 % de umbral para la máscara de segmentación.	62
3.11	Comparativa entre <i>Yolact Base</i> y <i>No Penalización</i> para 75 % de umbral para la máscara de segmentación.	62
3.12	Comparativa entre <i>Yolact Base</i> y <i>No Penalización</i> para 25 % de umbral para la máscara de segmentación.	62
3.13	Precisión de <i>Yolact Base</i> bounding box (a) y máscara (b), y <i>No Penalización</i> bounding box (c) y máscara (d).	63
4.1	Ejemplos de tallado y estimación de biomasa sobre el entorno no calibrado.	66
4.2	Dos ejemplos de ejecución del método <i>Yolact Base</i> (a, b, e, f) y <i>No Penalización</i> (c, d, g, h) sobre el 50 % y 15 % de límite de confianza, respectivamente.	68
4.3	Resultados finales combinatorios de tallado y método <i>KeyFish</i> propuesto en lonja mayorista calibrada.	70

Índice de tablas

2.1	Tabla con datos para la obtención de biomasa regresiva de especies encontradas en la lonja minorista no calibrada de El Campello.	37
2.2	Tabla con datos para la obtención de biomasa regresiva de especies encontradas en la lonja mayorista calibrada de Altea.	37
3.1	Comparativa entre 25 modelos de regresión populares y muestra de sus resultados (se muestran únicamente los 20 mejores modelos).	47
3.2	Comparativa entre resultados de los seis mejores modelos considerados (y SVM), sin (<i>izquierda</i>) y con ajuste de hiperparámetros (<i>derecha</i>).	48
3.3	Comparativa del error absoluto medio, MAE, en centímetros entre los mejores modelos de regresión analizados y diferentes normalizaciones de datos de entrada y salida.	49
3.4	Resultados finales con mejores modelos de regresión analizados con las tres entradas originales (bounding box en píxeles, área de segmentación y etiqueta de especie), <i>izquierda</i> . Adición de parámetros de calibración, esquinas de la bandeja en píxeles (x, y), <i>derecha</i>	51
3.5	Instancias totales e imágenes de bandejas de dataset utilizado, y distribución entre los conjuntos de entrenamiento y validación.	52
3.6	Mínimo, máximo, media y desviación estándar de instancias en el dataset utilizado para cada imagen de una bandeja entre los conjuntos de entrenamiento y validación.	54

3.7	Parámetros utilizados tanto en el modelo base como el modificado.	56
3.8	Comparativa de mAP de bounding box entre <i>Yolact Base</i> y el método propuesto <i>No Penalización</i>	58
3.9	Comparativa de mAP de máscara entre <i>Yolact Base</i> y el método propuesto <i>No Penalización</i>	58

1 Introducción

En esta sección se va a mostrar la motivación que ha hecho posible este proyecto *sección 1.1*, el estado del arte correspondiente al área de trabajo explorada *sección 1.2* y los objetivos propuestos *sección 1.3*.

El resto del documento se estructura de la siguiente manera: En la *sección 2* se presenta la metodología utilizada y las consideraciones básicas para el entendimiento de los posteriores apartados, tanto del tallado como del segmentado. La *sección 3* describe los resultados experimentales, en la *sección 4* se presentan visualmente los resultados de la experimentación con su aplicación en el entorno real, y finalmente en la *sección 5* se extraen conclusiones y trabajos futuros de las tareas realizadas.

1.1. Motivación y contexto

Mi interés por el Machine Learning surgió a través del efecto *mágico* de asombro, con los resultados ofrecidos y el enorme potencial que esconde. ¿Cómo es posible realizar tareas tan complicadas a partir de recursos aparentemente tan simples? Recursos tan sencillos como imágenes de fácil adquisición, un ordenador encendido, y tiempo (mucho tiempo). Esa premisa, y especialmente el hecho de ser una tecnología utilizada por muchos a través de frameworks, pero comprendida realmente por pocos, es lo que verdaderamente llamó mi interés. ¿Cómo consigo que una máquina diferencie imágenes verdes

de amarillas con únicamente 20 fotos, y por qué yo no soy capaz?

Afortunadamente, en la carrera que cursé, Ingeniería Informática en la Universidad de Alicante, este paradigma es explorado. Además, participé en un proyecto de investigación asociado precisamente a Deep Learning y Visión por Computador. El proyecto es DeepFish. Este buscó la utilización de las técnicas nombradas para la identificación y tallado de especies marinas en lonja.

A través de este proyecto, realicé mi trabajo de final de grado. He continuado trabajando vinculado al proyecto, por lo tanto, voy a enfocar este trabajo de final de máster, TFM, como una continuación del trabajo de fin de grado, TFG [8]. Además, el trabajo que voy a presentar se encontrará también asociado al proyecto sucesor, DeepFish². Este va a recoger una de las dos propuestas realizadas en un inicio que no fue posible realizar por cuestiones temporales, el cálculo del tallado de las especies, y va a aumentar el número y tipología de lonjas a tratar, además de la implementación y funcionamiento día a día en un entorno real.

Por cuestiones de coherencia temática, en este TFM no se va a cubrir mi aportación completa al proyecto, únicamente las 2 ramas que he considerado como continuación del TFG y relacionadas entre sí. Concretamente, se va a tratar la adaptación del proyecto minorista a nuevas lonjas mayoristas, y el tallado de las especies.

El objetivo de ambos proyectos es revitalizar el sector pesquero en nuestro país, abaratando costes de gestión y aumentar (o crear) un registro de capturas para mantener o recuperar el equilibrio natural de las poblaciones de especies del ecosistema y mantenerlas dentro de los límites biológicos seguros. Con especial atención a especies no permitidas o tamaños fuera de los límites.

Sigo pensando que estos proyectos han sido apropiados para mí, porque he podido utilizar e investigar tecnologías que me crean interés y porque me considero una persona moralmente implicada socialmente y con la situación medioambiental. Aportar una

pequeña contribución es un aspecto muy relevante en mi vida.

Un aspecto que he aprendido a lo largo de estos meses de proyectos y conociendo a gente en el sector, es la enorme ayuda que podría aportar este trabajo a un nicho concreto, aunque abundante en nuestro país. Debido a la última crisis con la guerra de Ucrania sumado a la previa pandemia, las lonjas minoristas se han visto fuertemente afectadas. El precio del pescado se ha reducido, y el precio del combustible ha aumentado considerablemente. Esto solo ha contribuido a que muchas lonjas en nuestro país, especialmente minoristas, hayan cerrado y otras se encuentren en una situación muy precaria. He conocido de primera mano lo arrepentidos que se encuentran algunos días de salir a trabajar. Porque, pese a haber sido productivos, por motivos ajenos a ellos (bajada de precios en el mercado o subida del combustible) se encontrarían perdiendo dinero ese día de trabajo por haber salido a pescar. Estas lonjas, como la lonja de El Campello, tienen únicamente un tráfico de 5 o 6 pequeñas embarcaciones familiares, se encuentran en una situación muy vulnerable.

Una lonja mayorista va a soportar de mejor forma una crisis y el sistema propuesto mejoraría la gestión actual que tienen. En el caso de una minorista, el sistema informatizaría la gestión que, por falta de recursos, tanto económicos como espaciales, sigue siendo una gestión pesquera manual. Por lo tanto, los avances en este sector son altamente perceptibles debido a la fatiga, inexactitud o lentitud del factor humano.

Los objetivos de explotación de datos serían de gran utilidad para esta situación, como un cuadro de mando con posiciones geográficas de las zonas con mayor número de especies, rutas recomendadas para los barcos propiciando ahorro de combustible y mayores beneficios para este sector precarizado, predicciones de poblaciones... Pienso que el factor de mejora en esta situación es muy amplio, y existe una gran necesidad social. Este enfoque ha sido nuevo en esta segunda fase, respecto al claro y único objetivo inicial que existía en la primera fase de la conservación del ecosistema marino, el cual sigue muy vigente. La gran aportación será proporcionar datos de biomasa a la comunidad

científica del mar.

El sistema utilizado consistirá, en cada lonja, en una cámara estática conectada a un monitor y un pequeño computador que enviará las imágenes desde este nodo a un servidor central, como se puede ver en la *Figura 1.1*. Este procesará (etiquetará, medirá y pesará las especies) y almacenará las imágenes. Además, será capaz de detectar cuándo está viendo una bandeja de peces llena mediante el identificador, y en este momento realizará la imagen que será enviada. De esta forma, los trabajadores no ven reducido su ritmo de trabajo, al encontrarse el sistema instalado encima del peso, por el que debían de pasar la bandeja igualmente.

Mediante el identificador, también se almacenará información como la embarcación encargada de haber traído ese pescado o el número de lote.

Las tareas previstas para el proyecto de investigación son:

- Ampliación del dataset.
- Tallaje de especies.
- Adaptación del sistema de visión a nuevas lonjas mayoristas.
- Desarrollo del módulo de explotación de datos.
- Implantación del sistema de visión.

Debido a la actividad investigadora, me ha sido posible la creación de varios artículos asociados al trabajo que voy a exponer en las siguientes secciones:

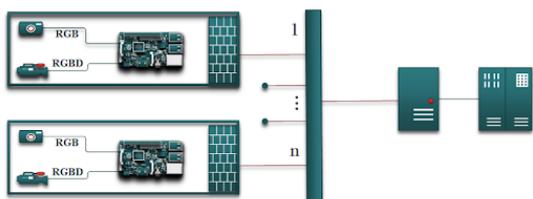
- *“Efficient instance segmentation using deep learning for species identification in fish markets”*.

Artículo publicado y presentado en el congreso IJCNN, Padua.

Contiene las modificaciones y comparativas iniciales a la red base “YOLACT” y “YOLACT++”.



(a) Implementación del sistema



(b) Arquitectura



(c) Utilización del sistema

Figura 1.1: Estructura edge de implementación del sistema en el entorno real (b) y su implementación en la lonja (a, c).

- “Automatic Fish Size Estimation from Uncalibrated Fish Market Images Using Computer Vision and Deep Learning” y “Tallado automático de especies de peces en lonja a partir de imágenes no calibradas mediante visión por computador y deep learning”.

Artículos publicados y presentados en el congreso SOCO, Salamanca y el Foro de científico de Pesca en el Mediterráneo, La Ràpita, respectivamente.

Tratan sobre el tallaje de especies en lonjas minoristas.

- “The DeepFish computer vision dataset for fish instance segmentation, classifica-

tion, and size estimation".

Artículo publicado en la revista Scientific Data, Nature.

Ocupa el dataset utilizado como base en el presente trabajo. El dataset resultado todavía no tiene artículo, pero se pretende realizar un dataset combinado con imágenes de diferentes topologías, pese a las diferencias sustanciales entre ambos.

- *“Arquitectura de adquisición de imágenes en lonja mediante edge computing”*.

Artículo publicado y presentado en las Jornadas SARTECO, Alicante.

Cubre la implementación del método mostrado en un entorno real y la arquitectura adoptada de *edge computing*.

- *“Increased accuracy for the best quality occurrences in a supervised instance segmentation”*.

Artículo enviado a revista IEEE Access.

Aborda el método *Key Fish* enfocado a lonjas mayoristas.

1.2. Estado del arte

Este apartado se subdivide en dos secciones principales, correspondientes a las dos ramas que serán tratadas en este trabajo, para su posterior unión. El presente *TFM* consistirá en la incorporación de dos líneas de investigación con punto de partida en el *TFG* [8] mencionado. Por lo tanto, el proyecto base, su elaboración y funcionamiento no será explorado en este trabajo, se consideran como punto de partida y no se va a realizar un estudio sobre el estado del arte, se recomienda consultar el material citado o la publicación asociada [2] si se desea profundizar más.

Los temas que sí se van a explorar corresponden, por un lado, al paradigma de una segmentación de instancias con alto nivel de solape y datasets con alta presencia de falsos positivos erróneos (que deberían ser verdaderos positivos, pero no han sido etiquetados) y, por otro lado, un tallado de instancias en un entorno no calibrado a partir de técnicas de metrología visual y un segmentado de instancias.

Tal y como se va a ver a continuación, son dos campos con baja presencia en la literatura. Se ha generalizado, porque, hasta lo que he podido ver, no hay estudios de esta índole en la aplicación al sector pesquero. Únicamente se pueden encontrar unos ejemplos más alejados, de un ámbito más generalista y menos de nicho, como [9, 10] sobre reconocimiento y tallado de clases de peces (en entorno calibrado) y para el primer caso son ejemplos más distantes como [11, 12, 13], los tres utilizando detección de objetos con *YOLO* [14] en diferentes condiciones (en el caso estudiado sería segmentación de instancias, con otra arquitectura y diferente dominio, también es más común encontrar estudios en el ambiente acuático que en una lonja de pescado).

1.2.1. Tallado de especies en entornos sin posibilidad de calibración

La gestión de la pesca es un proceso complejo en el que intervienen diversos agentes públicos y privados. Disponer de datos detallados sobre cuándo y dónde se captura una determinada especie es fundamental para mejorar la gestión. La evaluación de las poblaciones de peces puede evitar la sobrepesca y reducir la amenaza de la sostenibilidad de la industria pesquera. Bradley et al. [15] analizan la importancia de disponer de datos precisos y de información de retorno para mejorar la eficacia de la gestión pesquera, garantizando que el acceso compartido a los datos por parte de los pescadores y los gestores les ayude a trabajar hacia un objetivo mutuamente acordado.

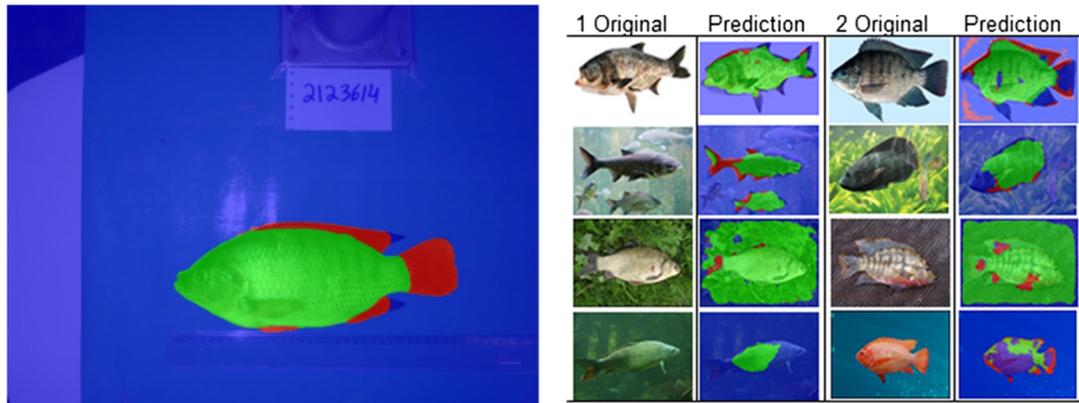
Según la FAO [16], los buques de pequeña escala representan más del 80 % de la flota en la zona del Mediterráneo. El Plan de Acción Regional para la Pesca a Pequeña Escala (RPOA-SSF) pide que se mejoren los conocimientos para garantizar un seguimiento adecuado del sector, incluidas las capturas. La mayoría de estas pesquerías a pequeña escala se gestionan en contacto directo con las autoridades pesqueras locales. D'Armengol et al. [17] subrayaron el interés de adoptar modelos de cogestión entre las partes interesadas, que deberían ser económicamente aceptables para los pescadores. Palmer et al. [18] destacan la importancia de contar con pruebas científicas para evaluar la población y, por tanto, la importancia de adoptar un seguimiento automático, frecuente y de alta resolución de los desembarques.

Las lonjas realizan un importante esfuerzo, a menudo manual, para contar las capturas por especies y medir las tallas de los peces [18]. El coste de este trabajo manual incorpora errores per se en los datos obtenidos, debidos al cansancio, a la capacidad limitada de la precisión humana, etc. Los sistemas de control basados en la visión por ordenador para recoger automáticamente los datos de las mediciones podrían reducir estos errores, aunque todavía son una excepción en la industria pesquera. El trabajo de Gladju et al. [19] revisa diferentes aplicaciones, incluyendo aspectos de la gestión pesquera, como

la monitorización de las capturas, que hacen más eficaz la pesca en la evaluación de la gestión de las poblaciones de peces. Los recientes avances en el área del aprendizaje profundo y la visión por ordenador están siendo aprovechados para desarrollar sistemas automáticos de identificación y tallado de especies basados en esta metodología [20, 21, 9].

En el contexto de la medición de objetos compaginado con la segmentación de instancias, dejando a un lado el sector pesquero, tampoco hay demasiada literatura. Se encuentran trabajos que no están enfocados en esta problemática, aunque sus enfoques cercanos podrían ser interesantes [22, 23, 24, 25], pertenecientes al campo de aplicación médico (medición de imágenes fetales o úlceras), más frecuente en esta problemática. Uno de los mejores ejemplos encontrados es el trabajo de Fernandes et al. [1], donde sí existe una aplicación al sector pesquero, con el fin principal de optimizar el proceso de gestión y evitar un malestar animal a la hora de las mediciones manuales. Plantean un dataset en un entorno calibrado (cámara fija a 0,5 m y pez fijo) con etiquetado semántico de la imagen completa, es decir, el fondo, cuerpos de peces y aletas. Cada imagen contendrá un único espécimen. Con estas tres etiquetas, se entrena un modelo de Deep Learning de segmentación de instancias, de la que se obtendrán valores como el alto y ancho de las especies. La conversión a centímetros será lineal dependiendo de la cantidad de píxeles de distancia obtenida en la máscara de segmentación. En la *Figura 1.2* se muestra un ejemplo del etiquetado y sus resultados con imágenes dentro y fuera del dominio.

Se va a centrar la atención en el problema de la identificación y medición automáticas de las especies en las pesquerías a pequeña escala. En estos mercados no siempre se dispone de sistemas de cámaras preinstalados, por lo que se utilizan sistemas de adquisición temporales y móviles sin un entorno calibrado. Por ello, se aborda la identificación y medición de las especies de pescado en las lonjas a partir de imágenes no calibradas. Para afrontar este problema, la propuesta recupera el tamaño del pescado en centímetros a partir de las imágenes no calibradas, utilizando un paso de segmentación de instancias

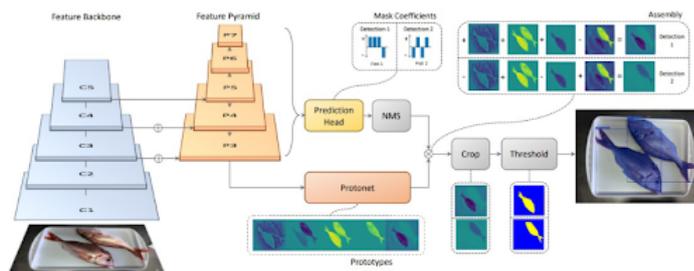


(a) Ground truth y su predicción. (b) Ejemplos externos al dataset (imágenes de internet) y su predicción.

Figura 1.2: Ejemplos del trabajo de Fernandes et al. [1]. Etiquetado semántico de la especie, el fondo y aletas.

y clasificación de especies realizado con la red presentada en [2], para luego utilizar esta información y calcular la estimación del tamaño. En la *Figura 1.3* se muestra la arquitectura y ejemplos de salida. Por otro lado, respecto al entrenamiento de modelos de regresión, es necesario obtener el *ground truth*, es decir, los valores reales que se pretenden predecir. Para estimarla se utilizarán técnicas de metrología visual y homografía, como se explica en [8].

La propuesta está pensada para trabajar en las lonjas minoristas del mundo real, donde el ritmo de trabajo y el pesaje de los peces suele ser acelerado, por lo que se requiere un tiempo de procesamiento con estrictas limitaciones de tiempo. Por lo tanto, la segmentación de instancias en tiempo real es de suma importancia. Esto se consigue mediante el uso de *Yolact++ (You Only Look At Coefficients)* [26], una arquitectura con grandes resultados sobre el estado del arte, especialmente en cuanto a rendimiento temporal, como se ha comentado anteriormente.



(a) Arquitectura.



(b) Predicción sobre una bandeja.



(c) Predicción sobre una bandeja.

Figura 1.3: Arquitectura y ejemplos del modelo implementado en [2].

1.2.2. Segmentación de especies en entornos con alto nivel de solape

Un correcto desarrollo del sector pesquero es crucial para la preservación marina y la economía local, como ha sido expuesto en la anterior *sección 1.2.1* con el foco en las lonjas minoristas. No obstante, el dominio mayorista presenta notables diferencias, especialmente en términos de solapamiento y heterogeneidad de cada bandeja de especies. Esto significa que el anterior enfoque no puede ser compatible, porque está esperando entornos donde todas las especies sean total o parcialmente visibles y el reto es la clasificación de cada ocurrencia. Por otro lado, en las bandejas de lonjas mayoristas se puede visualizar sobre un 10%-25% de las especies totales en esta, además de contener pocas instancias sin solapamiento. Se deberá utilizar una heurística para alcanzar dicha estimación, porque esa información no puede ser conocida mediante la directa detección de la

imagen. Como las bandejas presentarán una mayor homogeneidad de especies, será posible convertir el problema de clasificación de instancias en un problema de clasificación de bandejas, es decir, clasificar la especie que contienen las especies de cada bandeja.

En primer lugar, este problema se da en el contexto de la *segmentación de instancias* aplicada a *conjuntos homogéneos de objetos (HOC)* [3], en la *Figura 1.4* se muestran ejemplos de imágenes HOC. Las imágenes del dataset a trabajar encajarían en esta categoría, de no ser por una excepción. El dataset tiene la condición específica de no etiquetar cada instancia visible, sino de etiquetar únicamente las mejores ocurrencias u *objetos clave* de la imagen, lo cual, en lo que se ha investigado, apenas ha sido explorado. En este trabajo no se ha considerado un método de clúster como en [3], en el conjunto de datos analizado no está preparada ninguna estructura, por lo que será necesaria otra aproximación.



Figura 1.4: Ejemplos de *conjuntos homogéneos de objetos (HOC)* [3].

En el estado del arte, se pueden encontrar algunos trabajos sobre aprendizaje automático aplicado a imágenes con etiquetado incompleto, especialmente los aplicados únicamente a clasificación, que es lo más comúnmente estudiado. Xu et al. [4] modifican

la función de pérdida, restringiendo la penalización a al menos una correspondencia correcta en la imagen completa, para adaptar un SVM de uno contra todos a un conjunto de datos incompleto. Posteriormente, se realiza otra modificación en la función de pérdida basándose en la dependencia entre predicciones y etiquetas en la imagen, aprendiendo relaciones de etiquetas sin necesidad de una jerarquía semántica, con la intención de ser utilizado en anotación multi etiqueta. En la *Figura 1.5* se muestran dos ejemplos del etiquetado parcial de este dataset.

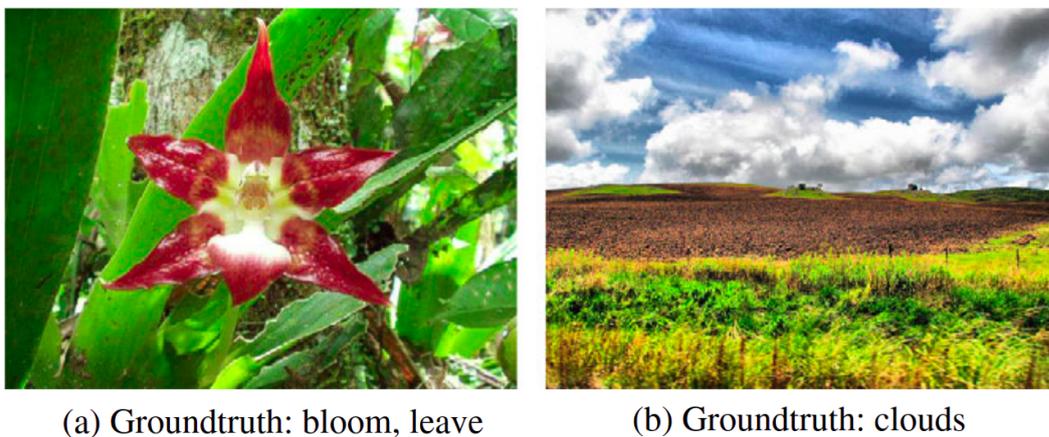


Figura 1.5: Etiquetado parcial de [4].

Sin embargo, un conjunto de datos incompleto tiende a tener un etiquetado parcial debido a la falta de recursos o a la baja calidad del proceso de etiquetado. Los métodos con estos objetivos tendrán una filosofía diferente a la del método propuesto en este trabajo, donde las imágenes tienen etiquetado parcial intencionadamente, porque el propósito es aprender a reconocer solo instancias de buena calidad de una clase, no todas las instancias de la misma. Además, las anotaciones parciales tienden a ser la omisión de clases, no de entidades. Por ejemplo, una imagen de un *mag*o en un prado con árboles, como la vista en la *Figura 1.6a*, con solo “*mag*o” como etiqueta para esa imagen, es un etiquetado parcial porque los árboles, cielo o prado no están, pero las clases que aparecen etiquetadas (*mag*o) aparecen al completo. Si la etiqueta fuese, por ejemplo, “*guerrero*”, como en la *Figura 1.6b*, deberán etiquetarse todos los *guerreros* aparecidos en la imagen, los cinco, de la misma forma que si apareciese la clase “*árbol*”, para mantener la

coherencia de estos métodos. En este caso de estudio, como aparece en la imagen, solo se habrían etiquetado los dos guerreros que mejor se ven de entre los cinco que aparecen en la imagen, y algunos de los árboles más nítidos. Con $N > m$, no es habitual encontrar conjuntos de datos con N instancias de una clase y solo m ocurrencias etiquetadas, si esto no es lo que se pretende. Es decir, el caso analizado en este TFM de tener etiquetados solo algunos guerreros de todos los que aparecen en la imagen, no ha sido encontrado en la literatura.



(a) *Ground truth*: Un mago.



(b) *Ground truth*: Dos guerreros.

Figura 1.6: Ejemplos de imágenes para etiquetado parcial.

La misma razón justifica la mayoría de las aproximaciones a conjuntos de datos compuestos por imágenes incompletas, que tienden a acercarse a la estrategia multi etiqueta con anotaciones de etiqueta única y son menos costosas [4, 27, 28]. Otro propósito común es el uso de métodos no supervisados. Sin embargo, el propósito está probablemente relacionado con algún enfoque de autoaprendizaje en el que la precisión y la calidad de cada etiqueta en las imágenes es baja. Un escenario típico en el que solo un pequeño porcentaje del conjunto de datos está etiquetado, pero con precisión, sería una situación completamente diferente.

Algunos trabajos hablan de *segmentación de instancias modales* [29, 5, 30]. Un enfoque para mejorar el reconocimiento en escenarios de alto solapamiento. Algunas técnicas utilizadas para mejorar los resultados son la predicción de la forma completa de un objeto solapado, la finalización de regiones solapadas de objetos o fondo, la división de la escena en múltiples objetos individuales y el establecimiento de un orden conocido de percepción

que tendrán los objetos en la imagen, ver *Figura 1.7*. Estas técnicas suelen requerir el aprendizaje de la estructura de los objetos y necesitan cantidades considerables de datos. La principal diferencia de este campo de estudio y el problema a tratar, es la clasificación de clases heterogéneas y el objetivo de etiquetar el mayor número de instancias posibles, aunque el porcentaje visto sea mínimo, que es lo contrario del objetivo principal tratado en este trabajo, etiquetar un pequeño conjunto de objetos maximizando la precisión de esas instancias sobre otras ocurrencias en la imagen.

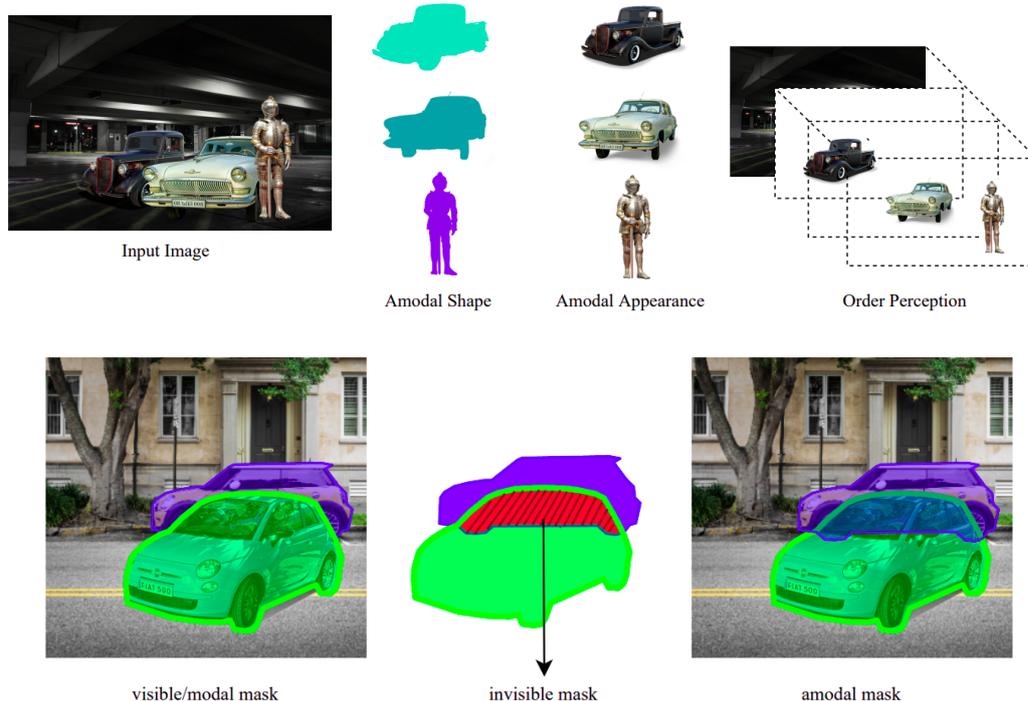


Figura 1.7: Ejemplo de capas amodales [5].

Como se menciona en [6], *“Entrenar una CNN con etiquetas parciales, por lo tanto, un pequeño número de imágenes para cada etiqueta, utilizando la pérdida de entropía cruzada estándar es propenso a sobreajuste y caída de rendimiento.”* Deberían introducirse algunas modificaciones en el método.

Junto con otros estudios, la visión es común sobre la importancia crucial de la función de pérdida [4, 27, 28, 6]. Dat Huynh et al. [6] regularizan la función de pérdida basándose

en la suavidad de las etiquetas. Esto se hace combinando el aprendizaje por similitud y el entrenamiento de la CNN. En otras palabras, el entrenamiento se realiza con CNN y simultáneamente se analizan las similitudes en los mini-lotes, y se realizan algunos cambios en las características en función de los resultados. Se puede ver un ejemplo en la *Figura 1.8*.



Figura 1.8: Ejemplo multi-etiquetado y su comparativa con otros métodos [6].

Por un lado, Thibaut Durant et al. [27] se centran en la conversión de etiqueta única a multi etiqueta. Se aplica una modificación en la pérdida de clasificación en función de la proporción de etiquetas conocidas por imagen. Por otro lado, algunos trabajos se centran en mecanismos de atención para abordar la mejora de la precisión. En concreto, Juan Carlos Ángeles et al. [31] aplican estas técnicas sobre una arquitectura *Yolact++* [26], la misma arquitectura base que se ha utilizado como punto de partida en el presente trabajo.

1.3. Objetivos

El objetivo general del proyecto de investigación DeepFish² en el que se enmarca este trabajo es, partiendo del prototipo de visión artificial para la identificación y tallaje de especímenes en lonjas desarrollado en DeepFish: avanzar en las capacidades del sistema, ampliando tanto el número de especies de peces objetivo, como de moluscos (pulpo, sepia y calamar), cubriendo una casuística mayor en tipologías de lonjas, introduciendo aspectos de trazabilidad mediante incorporación de información geo-referenciada y desarrollando un módulo de análisis y explotación de los datos proporcionados por DeepFish. También se pretende constituir un foro para la colaboración entre investigadores interesados en el reconocimiento y tallaje de especies mediante visión artificial.

Dado que el proyecto está constituido por más de una persona, se considerarán como objetivos de este TFM dos tareas realizadas únicamente por mí, correspondientes a los objetivos “*Adaptación de la arquitectura de red a nuevos entornos*” y “*Tallaje de especies mediante visión artificial*”. Los objetivos principales serán:

■ **Objetivo 1: Adaptación de la arquitectura de red a nuevos entornos.**

Para ello se realizarán las siguientes tareas:

- Captación de las imágenes para el nuevo dataset, su etiquetado y adaptación de los datos al formato de la red neuronal.
- Propuesta de método y modificación de la arquitectura de red.
- Diseño de una batería de experimentos para verificar la mejor configuración de parámetros.
- Comparativa y análisis de resultados respecto del modelo inicial.

■ **Objetivo 2: Tallaje de especies mediante visión artificial.**

Se desarrollarán las siguientes tareas:

- Propuesta de un método para el tallado de especies y obtención de biomasa.
- Experimentación del método propuesto en un entorno no calibrado.
- Calibración de imágenes en lonja mayorista. Prueba del método en entorno calibrado, junto a la adaptación de red realizada.
- Análisis de los resultados.

2 Antecedentes metodológicos

En esta sección se van a exponer los conocimientos básicos necesarios para abordar el desarrollo de este proyecto, de tal forma que sea posible el entendimiento de los siguientes experimentos, expuestos en la *sección 3*.

En los posteriores apartados se van a revisar las tecnologías y procedimientos utilizados para abordar las diferentes fases del proyecto, tallado en entornos tanto calibrados como no calibrados e identificación y segmentación en lonjas mayoristas. Se va a partir de la base ya establecida de un sistema de identificación y segmentación funcional de especies en lonjas minoristas, expuesto en mi trabajo de fin de grado [8].

2.1. Tallado en entornos no calibrados

Un entorno calibrado es una situación en la que tenemos el conocimiento de correspondencia sobre cuántos centímetros representa cada conjunto de píxeles de una imagen. En lonjas minoristas, como la analizada en este trabajo, no existe la situación de una cámara fija y las fotografías han sido tomadas mediante un teléfono móvil, variando en cada imagen la distancia y ángulo a la bandeja de peces. Para obtener la correspondencia o “calibración” de forma simulada, se han utilizado a posteriori marcas manuales, pero este método se encuentra sujeto a diferentes tipos de error del marcado de píxeles, precisión, métodos de corrección de imagen...

Ya se ha establecido un método de obtención de medidas reales, o ground truth, a partir de diferentes técnicas de metrología visual como la homografía, utilizando esas marcas realizadas a nivel de píxel en los puntos de las bandejas [8], como se puede ver en la *Figura 2.1*.



Figura 2.1: Muestra de una bandeja etiquetada mediante la versión modificada de [7]. Se marcarán las referencias reales correspondientes a las cuatro esquinas de la bandeja.

No obstante, el método de homografía utilizado [8] ha necesitado de un refinado para su mejor funcionamiento, y debido a las condiciones propias del dataset se encuentra a diferentes errores que perjudicarán el resultado final.

2.1.1. Errores en las medidas ground truth

En la *Figura 2.2a* se muestra la dispersión de los tamaños ground truth obtenidos a partir de homografía. Hay una clara constante cercana a los 20 centímetros, pero encontramos una gran cantidad de especímenes outliers con valores más grandes. Hay especies que pueden obtener un gran tamaño, no obstante sí que se encontrarán algunas como un error. También se pueden ver valores de orden de uno o tres centímetros, como

se mostrará a continuación, estos también se deben a errores en la adquisición. De forma complementaria, la *Figura 2.2b* muestra el histograma con los mismos datos. Se pueden intuir valores coherentes de tallado, excepto algunos valores muy extremos.

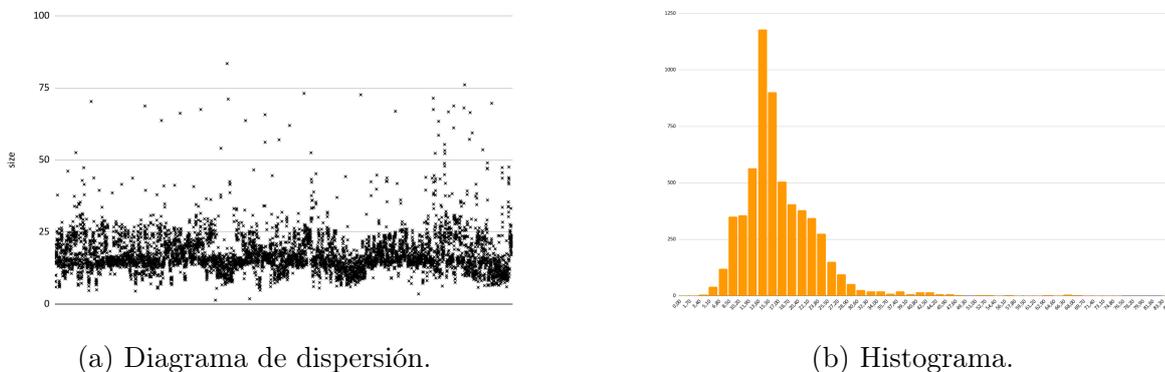
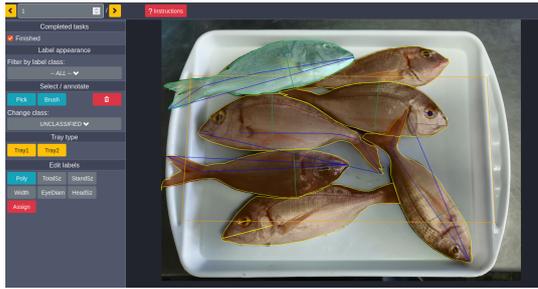


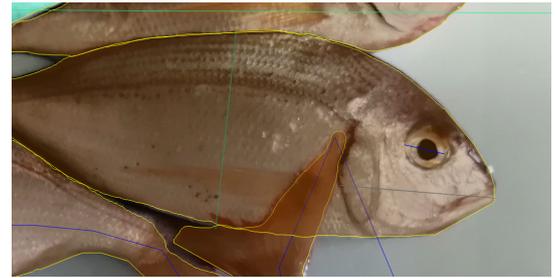
Figura 2.2: Dispersión e histograma de las especies con base en el tamaño ground truth obtenido mediante homografía.

Se han encontrado tres errores comunes en la adquisición de datos:

- **Equivocación en el tipo de medida.** En el entorno de etiquetado, existen diferentes tipos de etiqueta (especie, longitud total, longitud del ojo, ancho...), una equivocación en la selección de este tipo significará una medida incorrecta, como ocurre en la *Figura 2.3*. En este caso concreto, en el pez central de la derecha se ha utilizado la longitud total, color azul, (utilizada en el algoritmo de homografía para hallar la longitud ground truth de las especies) donde debería de haberse utilizado el diámetro de ojo. Como es lógico, el algoritmo ha arrojado un tamaño de 1 centímetro para esta instancia.
- **Incorrección en el número de aristas de la bandeja.** El algoritmo creado espera una cantidad de 4 esquinas para la bandeja, e incluso está sujeto a 3 esquinas en caso de pérdida de una de estas. Utilizará estas referencias para tomar constancia de las dispersiones de la imagen y corregirla (calibrarla artificialmente). Existen bandejas con un valor mayor de 4 esquinas, como la que se observa en la *Figura 2.4*. Dependiendo del orden seleccionado, es esperable que se pueda calcular mal la corrección de imagen y perjudicará a todas las instancias en esta.



(a) Bandeja con el error.



(b) Zoom sobre el error.

Figura 2.3: Error de etiquetado humano, confusión en la selección de categoría de etiqueta.

Si el número de esquinas intencionales son 4 y se han marcado más, la forma de solucionarlo es sencilla, filtrando por número de aristas. No obstante, si el usuario pretendía etiquetar tres esquinas y ha cometido algún clic extra, es posible que esa imagen con tallados incorrectos se encuentre en el dataset final.

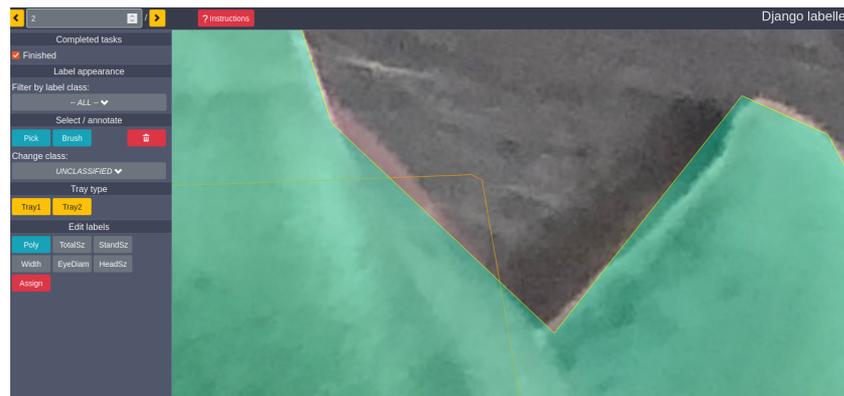
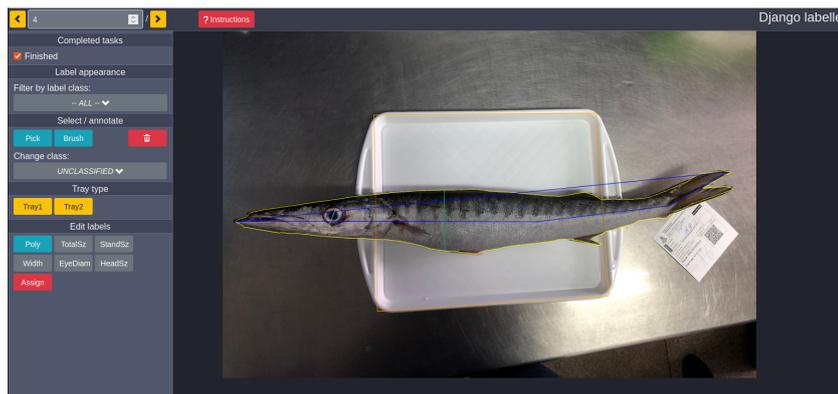


Figura 2.4: Esquina de bandeja mal etiquetada, contiene dos puntos juntos que dan lugar a una mala corrección de imagen.

- Incorrección en la ubicación de los puntos de la bandeja.** El algoritmo se encuentra preparado para corregir la imagen a partir de los puntos de referencia marcados, concretamente cada una de las aristas de las asas de la bandeja, como se puede ver en la *Figura 2.1*. Son comunes las imágenes con imprecisiones en estos puntos concretos, este error de imprecisión se va a propagar. También existe un conjunto imágenes con una elección incorrecta de la ubicación en las referencias, como pueden ser las esquinas de la bandeja (un lugar más impreciso y con mayor

lugar a error), visto en la *Figura 2.5*. En este ejemplo, debido a una mayor región de las marcas de referencia (cuadrado amarillo), la instancia de su interior tendrá un tamaño ligeramente menor del que debería si el marcado fuese más preciso. Este error también será difícil de detectar, debido al no tan notable cambio en talla, y necesitaría revisión manual. Es posible que haya pasado un número considerable de estos ejemplos al dataset.



(a) Bandeja con referencias incorrectas.



(b) Zoom sobre el error.

Figura 2.5: Error de etiquetado humano, confusión en la ubicación de las referencias de la bandeja. Deben de estar en las asas, no en las esquinas.

Finalmente, se han tenido en cuenta estos problemas y se han eliminado del dataset

todas las imágenes o instancias con errores que han sido posibles, delimitando los límites de tamaño (por ejemplo, cualquier especie menor de 5 centímetros se ha demostrado ser un error) o descartando especímenes concretos que se han detectado como mal etiquetados. Una ausencia de filtrado de estas instancias con errores perjudicará el modelo de regresión en su media y sesgo de especies. Un mayor tamaño de dataset reducirá estos efectos negativos y haría más consistente al sistema.

2.1.2. Método propuesto

Una vez establecidos los precedentes como base, el siguiente paso a realizar será la predicción de valores a partir de los datos ejemplos ya obtenidos. Al no tener un sistema de calibración, el objetivo principal será evitarlo mediante la simulación de este a un modelo, que aprenderá cómo “calibrar” sin la necesidad del costoso proceso real de tener un entorno calibrado.

El flujo de trabajo creado se encuentra en la *Figura 2.6*. La entrada al modelo de regresión del tamaño propuesto consistirá en la salida de la red neuronal [2] preentrenada para segmentación de instancias, además de su identificación de clase (especie de pescado, en este caso), y la caja delimitadora, o bounding box, en coordenadas de píxel indicando dónde se encuentra cada instancia.

En esta metodología propuesta, el tamaño de las especies se va a estimar con un modelo de regresión. Se trata de un método de aprendizaje supervisado, es decir, va a requerir de datos reales “ejemplo” con los que aprender (en este caso, las tallas de los peces). Estos datos reales se han obtenido mediante técnicas de metrología visual de forma automática a partir de referencias marcadas manualmente sobre puntos de la bandeja, como ha sido comentado, debido a que la medición manual de miles de bandejas y varios miles de especímenes no habría resultado viable para las condiciones de este proyecto. De esta manera, se ha calculado una función de transformación para cada imagen, de

forma que se pueda conocer la correspondencia a centímetros de cada píxel. También se transformarán los bounding box y segmentación de cada imagen. Esta información se utilizará como guía para el modelo de regresión durante su entrenamiento.

La principal contribución aportada sobre el tallado será, por tanto, la regresión precisa de los peces a partir de los datos de segmentación inferidos por la red neuronal, clase y bounding box, y el uso de los tamaños estimados mediante homografía como datos reales para el entrenamiento. Con el objetivo de abordar un problema de calibrado, sin la necesidad de calibrar.

La *Figura 2.6* muestra una visión general del flujo de trabajo completo para la metodología propuesta, incluyendo sus diferentes componentes. En la parte superior del esquema (recuadro azul), se representa el entrenamiento del modelo del flujo completo (la red y el modelo de regresión). La red utilizará como entrada el banco de imágenes y su etiquetado segmentado de forma manual. El modelo de regresión utilizará la salida de la red como entrada, y la calibración artificial de las imágenes como salida de datos reales para el entrenamiento.

Para la estimación final del tamaño de los peces en el entorno real (recuadro amarillo en la *Figura 2.6*), se contemplan diferentes modelos de regresión, en la *sección 3.1* se explorará el modelo más apropiado para este entorno. Finalmente, se obtendrá el tamaño en centímetros como salida del modelo de regresión y los bounding box, máscaras y especies concretas como salida la red. Ambas serán combinadas para la obtención final de la biomasa, como será explorado en la *sección 2.3*.

2.1.3. Parámetros de entrada del modelo de regresión

A partir de los datos de entrada, correspondientes a la salida de la red neuronal, se obtendrá una salida en centímetros del tallado estimado para cada especie.

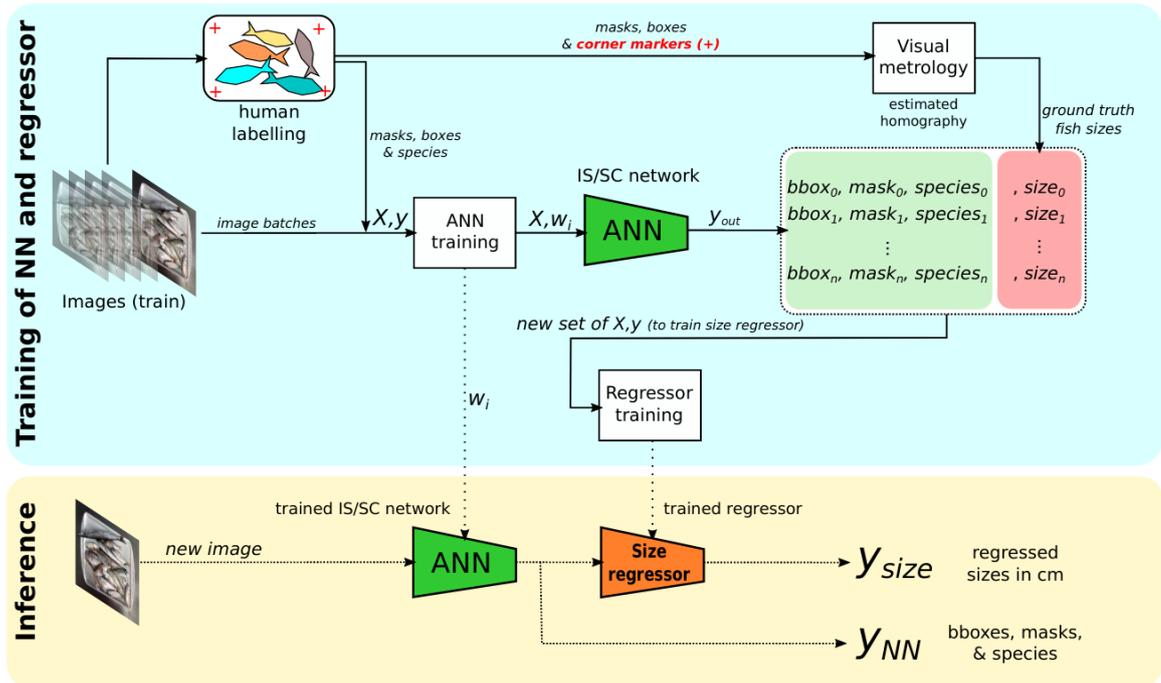


Figura 2.6: Flujo de trabajo general de la metodología propuesta para la estimación del tamaño de los peces. La parte superior (zona azul) muestra el flujo de entrenamiento de la red neuronal [2] a partir del etiquetado de imágenes, y del modelo de regresión a partir de las salidas de la red y las imágenes corregidas mediante metrología visual. La parte inferior (amarillo) muestra el bloque de uso en entorno real, es decir, no se dispone de *ground truth* y la predicción de los dos bloques será la salida. A partir de estos cuatro elementos, se calculará la biomasa.

Concretamente, las entradas del modelo de regresión son:

- **Bounding box.** El uso de la caja delimitadora será de gran significancia. No solo a mayor tamaño tendremos mayor longitud, también una caja del mismo perímetro con mayor diferencia en ancho y alto tendrá una talla mayor respecto de una caja más cercana a un cuadrado, con un valor de anchura cercano al valor de altura.
- **Segmentación.** Principal razón por la que se ha escogido una red neuronal de segmentación de instancias en lugar de detección de objetos. Dentro de un mismo bounding box, será muy relevante conocer si la especie es fina o ancha, ocupando gran parte de la caja o dejando más huecos en este, y de otro modo capturar

posibles especies con un enroscamiento, indicando una longitud mayor a la diagonal del bounding box.

- **Especie.** La clase servirá como un “*bias*”, sesgo, que ayudará al modelo a estimar cuáles suelen ser los valores de rango de dicha especie, contribuyendo a compensar la falta de calibración, aportando una guía de cuál podría ser la correspondencia píxel centímetro en esa imagen.

Además, para todos los métodos de regresión utilizados en la experimentación, se han tenido en cuenta algunas consideraciones comunes con la intención de la mejora sobre los resultados:

- **Conversión de la segmentación.** La segmentación de instancias se obtiene como un conjunto de coordenadas (x, y) , de una magnitud tan grande como píxeles contenga esa especie, en otras palabras, la sombra correspondiente a esta, como se puede ver en la *Figura 2.7*. Una cantidad tan grande de parámetros no es recomendable para un modelo de regresión, que trabajan mejor con un número más bajo de parámetros. De hecho, existe una tendencia a necesitar un mayor número de datos para tener un buen funcionamiento si el número de parámetros crece. No solo esto, el número de parámetros también sería variable, puesto que cada instancia tendrá un número de puntos diferente, correspondiente a su silueta. Por lo tanto, será necesario comprimir esta información. La solución utilizada ha sido la conversión de este polígono en un número, calculando el área encerrada en este.
- **Transformación del bounding box.** La mayoría de las imágenes del conjunto total van a presentar diferentes ángulos y distancia desde la toma de imagen hasta la bandeja entre sí. Al tratarse de este entorno no calibrado, una misma caja delimitadora con las mismas cuatro coordenadas (puntos iniciales x, y , y puntos finales x, y , ver \mathbf{X} en la *Figura 2.8*) en dos imágenes diferentes pueden representar un tamaño real totalmente distante, pudiendo confundir a la red. Para evitar esta



Figura 2.7: Área de segmentación (en azul) como conjunto de píxeles representando la zona perteneciente a una clase concreta. Se trata de una de las salidas de la red neuronal y de las entradas del modelo de regresión.

situación, se ha utilizado el punto inicial (x, y) , la anchura y la altura, obtenidas a partir del punto final e inicial.

Si se busca una independencia sobre la posición de una instancia en la imagen, y un enfoque en su tamaño, lo ideal será utilizar únicamente el ancho y el alto, como ocurre en los entornos calibrados, y lo que se realizará en la *sección 2.4*. No obstante, se ha comprobado, en esta situación no calibrada, cómo la aportación extra del punto inicial proporcionará información relevante al modelo sobre la distorsión, deformaciones o escala. Este conocimiento es suficientemente significativo como para justificar su inclusión, incluso teniendo en cuenta su dependencia negativa respecto de la posicionalidad, por lo que acaba siendo beneficioso para el modelo finalmente.



Figura 2.8: Representación del bounding box de una especie. Punto inicial (x, y) como esquina superior izquierda y punto final como esquina inferior derecha (x, y) , ambos representados por una X azul. Ancho y alto calculados a partir de estos puntos, representados a partir de las flechas verdes.

2.2. Identificación y segmentación de especies en entornos mayoristas. Método *KeyFish*

El paso de un entorno minorista a uno mayorista tiene las siguientes diferencias principales:

- **Tipos y cantidad de clientes.** La principal diferencia se trata del tipo de clientes al que va asociado este negocio. Mientras que la lonja minorista va a tener una gran cantidad de consumidores particulares, comprando baja cantidad de pescado; la lonja mayorista va a tener unos pocos compradores que representarán cadenas de supermercados o restaurantes, con pedidos de grandes cantidades. La mayoría de diferencias dependerán de este detalle.
- **Tipología de las bandejas.** Al tratarse de una atracción turística, las bandejas

de una lonja minorista van a encontrarse bien cuidadas estéticamente, las especies se podrán contar visualmente y tendrán buena visibilidad y variabilidad. En una lonja mayorista la tendencia será de bandejas grandes, con una gran cantidad de oclusión, pero un bajo ratio de variabilidad de especies, es decir, la mayoría de bandejas tendrán alta homogeneidad de especies. Se muestra un ejemplo de ambas tipologías de bandeja en la *Figura 2.9f* y la *Figura 2.9g*, respectivamente.

- **Volumen de tráfico.** Al día aparecerá un mayor número de instancias por la lonja mayorista, por lo que será más rápido llegar a determinado número de imágenes para el dataset.
- **Mayor poder adquisitivo.** La tendencia será a tener unas instalaciones más grandes y con más recursos, puesto que el volumen de negocio es mayor. En lo que respecta al proyecto, esto significará mayor espacio y dinero para informatizar el entorno, por ejemplo, realizar un calibrado. También es más probable que ya existan sistemas de visión, por lo que sería necesario adaptarse a estos, en lugar de construir uno desde cero.

A partir de las diferencias expuestas, también cabe destacar nuevos problemas o retos a los que se enfrenta el proyecto. Principalmente son:

- **Alto nivel de solape.** Todas las instancias que no sean visibles y no se puedan identificar o contar, no serán reconocibles por el sistema mediante la arquitectura actual. Ver *Figura 2.9a*.
- **Posiciones cambiantes.** No existe un patrón de colocación de especies en la bandeja, puede resultar confuso para el sistema de visión y tampoco se podría utilizar como una ayuda. Se pueden observar dos bandejas con cantidades similares de la misma especie, colocadas de forma diferente en la *Figura 2.9c* y la *Figura 2.9d*.
- **Heterogeneidad de tamaños.** Cualquier tipo de aproximación sobre el tamaño

de especies se verá dificultada por la inconsistencia en la longitud de los especímenes en algunas bandejas. Ver *Figura 2.9e*.

- **Heterogeneidad de especies.** Dejará de ser un problema tan relevante como en el entorno de lonja minorista, pero seguirá existiendo un porcentaje de bandejas con heterogeneidad de especies, como la vista en la *Figura 2.9b*.
- **Artefactos.** Los más comunes que reducirán la calidad de la imagen son la presencia de hielo, agua, suciedad o humedad en la cámara o brillo de la estancia, dependiendo de la hora del día y la climatología. Ver *Figura 2.9h* y *Figura 2.9i*.
- **Baja resolución.** Si la situación es de adaptación al método de visión existente, es muy probable que este no se encuentre preparado para tareas de visión por computador, por lo tanto, la resolución de imagen tenderá a ser inferior respecto de la vista anteriormente, donde el sistema de visionado ha sido realizado por el proyecto con este objetivo en mente. Concretamente, en la lonja minorista hay imágenes de “4032x3024”, mientras que la lonja mayorista analizada son de “1280x800” píxeles, como se puede ver en la *Figura 2.9f* y la *Figura 2.9g*.
- **Visibilidad incompleta.** Capturar la bandeja entera no fue el objetivo cuando se implantó el sistema de visión. La imagen puede dejar especies parcial o totalmente fuera de la imagen, como en la *Figura 2.9j*.

Se ha tenido en cuenta este número de nuevos problemas que no cubre la actual arquitectura de red, reduciendo su precisión, y los aspectos como la prioridad de clasificación de especies dentro de la misma bandeja, que ya no será tan relevante en este escenario (existe una mayoría de bandejas homogéneas), para realizar la modificación de la arquitectura.

(a) *Alto grado de solape.*(b) *Bandeja multi especie.*(c) *Posicionalidad 1.*(d) *Posicionalidad 2.*(e) *Heterogeneidad de tamaños.*(f) *Resolución lonja minorista.*(g) *Resolución lonja mayorista.*(h) *Artefactos: plástico y hielo.*(i) *Artefactos: nombre embarcación.*(j) *Bandeja incompleta.*

Figura 2.9: Diferentes retos encontrados en la lonja mayorista.

2.2.1. Método propuesto

Se ha propuesto un nuevo método pretendiendo abordar estos nuevos retos y prescindir de objetivos no tan necesarios en este entorno. El principal objetivo de este nuevo

enfoque, llamado *KeyFish*, es la detección de la segmentación más precisa en toda la imagen, sin importar la cantidad de predicciones realizadas o la calidad de estas. Con el subconjunto k de instancias predichas con mejor confianza, se calculará la talla y peso, y el número de especies totales puede ser extraído mediante el peso total de la bandeja y el peso medio de los *KeyFish*. No obstante, en este trabajo no se va a cubrir dicha estimación sobre el conteo de instancias totales, la lonja no ha facilitado el peso de las bandejas todavía.

Como se ha comentado anteriormente, la arquitectura ya utilizada en el entorno minorista no está preparada para este problema, se basa en contar las instancias que puede detectar y hace esfuerzos por diferenciar cada una. Si solo se puede visualizar una pequeña fracción del total de instancias en una imagen, el número de instancias detectadas será mucho menor. Además, cada predicción errónea supondrá información incorrecta o pérdida de información, decreciendo la precisión del método. Como el nuevo entorno presentará más ruido, tanto de resolución como artefactos ya comentados, la precisión debería ser menor, por lo tanto, se cometerán más errores que repercutirán directamente en información errónea.

El nuevo método contempla los problemas de solape, posiciones cambiantes y artefactos. Perder visibilidad de zonas de la imagen tal vez no afecten a la precisión, siempre y cuando se pueda ver al menos una especie con suficiente calidad en otra zona como para ser considerada *KeyFish*. No obstante, si el conjunto de *KeyFish* no es representativo de la población total de la bandeja, las bandejas multi especie o con tamaños heterogéneos pueden presentar peores resultados de los esperados, en caso de que el conjunto utilizado para aproximar el total se encuentre incompleto o la distribución sea alejada a la de la bandeja total.

El objetivo principal ha cambiado de la detección de todas las instancias posibles a la detección de las mejores instancias, optimizando la precisión de estas. Para adecuarse a esta nueva metodología, será necesario adaptar el dataset utilizado y la arquitectura de

red a este enfoque.

2.2.2. Adaptación de la arquitectura de red

El principal cambio realizado en la arquitectura de red es la función de pérdida. Esta función tiene tres componentes principales [2]: un error de clasificación L_{cls} , un error de bounding box L_{box} , y un error de la máscara de segmentación L_{mask} . El error de L_{cls} y L_{box} se calcula como se establece en [32]. Para el cálculo de la máscara de segmentación se utiliza una entropía cruzada binaria (BCE) sobre el conjunto de máscaras predichas y el conjunto de máscaras ground truth.

La contribución principal del método es la siguiente: para cada uno de los tres parámetros de pérdida, se utilizará la misma función de loss, pero la importancia que esta aporte al algoritmo va a depender de la similitud que esta instancia tenga respecto del ground truth más cercano. En otras palabras, solo las M instancias predichas con mayor intersección sobre la unión (IoU) sobre los N ejemplos ground truth disponibles serán penalizadas, como se puede ver en la *Ecuación 2.1*. El parámetro α se utiliza como umbral para determinar si la predicción será penalizada, dependiendo de su *IoU* con el ground truth. χ representa todas las predicciones en la imagen, y será utilizado para expresar la mejor coincidencia de predicción para cada ejemplo de ground truth.

$x \cap y$ representa la cantidad de IoU de la instancia predicha con su correspondiente ground truth. De tal forma, una instancia recibirá penalización si la cantidad de solape cumple un mínimo, y esta es la instancia con mejor solape para ese ground truth de entre todas las predicciones de la imagen.

$$f(x, y) = \begin{cases} x - y & \text{if } (x \cap y) > \alpha \wedge (x \cap y) = \text{máx}(\chi \cap y) \\ 0 & \text{if } (x \cap y) \leq \alpha \vee (x \cap y) \neq \text{máx}(\chi \cap y) \end{cases} \quad (2.1)$$

Esta fórmula será aplicada a cada predicción. Dependiendo del parámetro β , la nueva función de pérdida puede variar su impacto y la función original de error puede permanecer. Ver *Ecuación 2.2*.

$$L = (1 - \beta)\|f(y, \hat{y})\| + \beta\|y - \hat{y}\| \quad (2.2)$$

El caso más restrictivo de mínima penalización será el utilizado en la *sección de experimentación 3.2.1* como método con mejores resultados. La configuración es la siguiente:

- La cantidad de instancias penalizadas M será igual a la cantidad de ejemplos ground truth N en esa imagen.
- Parámetro α es igual a 0. Se penalizarán las N mejores predicciones, independientemente de lo “malas” que sean.
- Parámetro β igual a 0. No tendrá relevancia la función de pérdida original, y se utilizará únicamente la propuesta en la *Ecuación 2.1*.

2.2.3. Adaptación del dataset

El objetivo principal seguirá siendo el mismo, detectar la biomasa presente en una bandeja y contar el número de instancias, su talla y especie. No obstante, la forma de abordar el problema ha cambiado, por lo que será más efectivo si también se crea el dataset teniendo estas metas en cuenta. Debido al nuevo entorno, se ha creado un nuevo banco de imágenes desde cero. Contiene las especies ya contempladas en el anterior dataset [33], además de 14 nuevas incorporaciones, ver *Figura 3.3*. La norma de creación será etiquetar únicamente las instancias de la imagen que presenten una gran calidad de visibilidad, por lo que van a haber imágenes donde no haya etiquetas. No solo favorece el entrenamiento de esta nueva aproximación al problema, también resulta menos costoso

etiquetar cada imagen y el ritmo de obtención de estas se ve aumentado. Este nuevo dataset se encuentra disponible con acceso abierto [34].

2.3. Obtención de biomasa

La obtención de la biomasa de las especies (volumen de los peces y el peso total de las capturas) se considera una magnitud crucial para la conservación de las especies en el entorno marino, y actualmente no se tiene esta información en la zona geográfica explorada, o la información disponible es muy escasa.

Por lo tanto, la estimación de esta sería de gran valor para la comunidad científica y ayudaría a la preservación del ecosistema marino. Para llevar a cabo esta predicción, se utilizará el tamaño estimado de los peces por el modelo de regresión, junto con la clase predicha a partir de fórmulas de regresión proporcionadas por expertos en el campo de la biología marina, aspecto que se encuentra fuera del alcance de este trabajo.

Se aprovechará la obtención de la talla en centímetros para la obtención de la biomasa, junto con la especie de cada instancia. Mediante estudios realizados por expertos en el área, han sido proporcionadas para el proyecto diferentes tablas de regresión con estos dos parámetros, como se muestra en la *ecuación 2.3*, para las distintas clases de especies encontradas en el estudio de la zona mediterránea, con salida en gramos.

$$W = a * L^b \tag{2.3}$$

W será el peso en gramos de la instancia, a y b se obtienen a partir de la especie dependiendo de las tablas, y L representa la talla del espécimen.

Las tablas utilizadas han sido la *Tabla 2.1* para la lonja minorista no calibrada de *El Campello*, y la *Tabla 2.2* para la lonja mayorista calibrada de *Altea*.

ESPECIE	a	b	REFERENCIA	ZONA DE ESTUDIO
<i>Dentex dentex</i>	0,018	3,029	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Diplodus annularis</i>	0,0217	2,84	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Diplodus sargus</i>	0,0175	2,921	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Mullus surmuletus</i>	0,0104	3,0617	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Mullus barbatus</i>	0,0077	3,1095	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Merluccius merluccius</i>	0,0046	3,1191	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Pagellus acarne</i>	0,0069	3,221	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Pagellus erythrinus</i>	0,0096	3,118	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Pagrus pagrus</i>	0,0179	2,95	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Sarda sarda</i>	0,001	3,593	Evangelopoulos et al., 2020	Mar Jónico, Grecia.
<i>Scorpaena porcus</i>	0,061	2,652	Evangelopoulos et al., 2020	Mar Jónico, Grecia.
<i>Symphodus tinca</i>	0,0212	2,8351	Soldo 2020	Mar Adriático
<i>Sphyaena sphyraena</i>	0,0066	2,89	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Spicara maena</i>	0,013	2,988	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Serranus scriba</i>	0,012	2,963	Ailane et al., 2021	Mediterráneo sur oeste. Algeria
<i>Seriola dumerili</i>	0,019	2,8726	Nesma et al., 2018	Mediterráneo sur. Libia
<i>Sparus aurata</i>	0,0109	3,091	Mehanna et al., 2021	Mediterráneo este. Costa de Egipto.
<i>Sepia officinalis</i>	0,1884	2,81	Adamidou et al., 2020	Norte del Mar Egeo

Tabla 2.1: Tabla con datos para la obtención de biomasa regresiva de especies encontradas en la lonja minorista no calibrada de El Campello.

ESPECIE	a	b	REFERENCIA	ZONA DE ESTUDIO
<i>Trachurus sp</i>	0,00892	2,951	Rodríguez Castañeda, J.C et al., 2021	Mediterráneo español
<i>Lophius sp</i>	0,0117	3,03	Adamidou, A. et al., 2020	Mar Egeo
<i>Phycis blennoides</i>	0,0019	3,46	Romdhani, A. et al., 2016	Golfo de Túnez
<i>Trisopterus minutus</i>	0,0011	2,981	Šantić, M. et al., 2010	Mar Adriático
<i>Boops boops</i>	0,0025	3,47	Adamidou, A. et al., 2020	Mar Egeo
<i>Zeus faber</i>	0,0205	2,915	Vrgoč, N. et al., 2006	Mar Adriático
<i>Citharus linguatula</i>	0,0027	3,33	Adamidou, A. et al., 2020	Mar Egeo
<i>Micromesistius poutassou</i>	0,0038	3,187	Mir-Arguimbau, J. et al., 2020	Mediterráneo español
<i>Loligo vulgaris</i>	0,1309	2,418	Šifner, S. et al., 2004	Mar Adriático
<i>Illex coindetii</i>	0,0365	2,95	Ceriola, L. et al., 2006	Mar Adriático

Tabla 2.2: Tabla con datos para la obtención de biomasa regresiva de especies encontradas en la lonja mayorista calibrada de Altea.

2.4. Tallado en un entorno calibrado y aplicación del método *KeyFish*

Un entorno calibrado va a suponer una simplificación del problema ya visto en la *sección 2.1*, además de una mejora en los resultados, puesto que se realizarán menos aproximaciones y cálculos, con una mayor variedad de datos reales disponibles.

2.4.1. Diferencias respecto del entorno minorista no calibrado

En el dataset de entrenamiento, la diferencia respecto a estructura del dataset utilizado en el entorno no calibrado, será que esta vez no se han capturado los píxeles de referencia correspondientes a las esquinas de bandeja, no serán necesarios. En su lugar, se han calibrado todas las imágenes como se verá a continuación. De esta forma, la correspondencia píxel a centímetro será constante para cada conjunto de píxeles de la imagen, y también será equivalente para todas las imágenes.

Por lo tanto, el método utilizado será el mismo que en la *Figura 2.6*, con la diferencia que el ground truth no se va a obtener mediante metrología visual. Al tratarse de imágenes calibradas, la salida predicha y el ground truth se encontrarán en píxeles en lugar de en centímetros, y posteriormente se realizará la conversión para cada imagen.

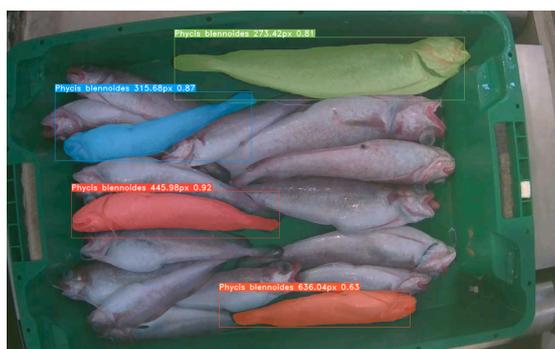
Respecto a los parámetros de entrada del modelo de regresión, el bounding box tendrá únicamente el ancho y alto, ambos no dependientes de la posicionalidad, ya no será necesario aportar la coordenada inicial, incluso sería negativo para el aprendizaje.

En la *Figura 2.10* se demuestra la utilidad de prescindir de esa coordenada inicial, con una comparativa entre el modelo de regresión utilizando ese parámetro de entrada y otro ejemplo sin utilizarlo. Por casualidades del entrenamiento, si a medida que disminuye el valor de y (la altura) crece el valor de predicción (como se observa en la imagen,

independientemente del tamaño real), mayor importancia tendrá en la talla respecto a otros valores como ancho, alto o área.

En el ejemplo *2.10a*, *Phycis blennoides* de 635 píxeles es el espécimen que debería tener una menor talla, y de la misma forma el espécimen de la zona superior, debería ser el de mayor longitud. Se puede observar claramente cómo, la disposición espacial de las especies en el dataset ha condicionado en gran manera el resultado, en el resultado visto no parece coherente.

Al tratarse de muestras de entrenamiento con clases poco representadas e imágenes con pocas instancias repartidas de forma aleatoria, no tiene por qué ser equilibrada la posicionalidad junto con una disparidad de longitudes, este ruido está afectando negativamente a la predicción de forma muy significativa, la cual será diferente si la misma instancia se encontrase en la zona inferior de la imagen. Por lo tanto, será recomendable prescindir de estos dos valores de entrada. En el ejemplo *2.10b* se ve una mayor independencia a la posicionalidad de las instancias, tan solo se han utilizado los valores de ancho y alto.



(a) Punto inicial, ancho y alto



(b) Ancho y alto

Figura 2.10: Comparativa del uso de la posición inicial del bounding box junto con ancho y alto (*2.10a*) o la supresión de este (*2.10b*).

2.4.2. Calibrado de imágenes

A diferencia de las lonjas minoristas exploradas, donde se necesitaban puntos de referencia a marcar, al disponer de una cámara fija en los nuevos entornos, no se encontrarán variaciones en cuanto a diferentes ángulos y rotaciones, por lo tanto, no habrá necesidad de corrección y se reducirá el error final.

Si se realiza un calibrado de la cámara se podrá obtener una correspondencia píxel-centímetro aplicable a todas las imágenes capturadas en este entorno.

Para realizar el calibrado, se ha utilizado la librería *opencv* [35], concretamente las funciones de calibración ya preparadas para estas situaciones de esta librería. Han sido necesarias varias imágenes de calibración para calcular las distorsiones concretas que posee la cámara y obtener la matriz de calibración. Para ello, se han tomado estas imágenes con un mismo objeto (en este caso un *chessboard* de cartón) en múltiples posiciones y rotaciones, concretamente han sido 35 imágenes, con la intención de cubrir el mayor campo de localizaciones y rotaciones de objetos para afinar la precisión del método. En la *Figura 2.11* se muestran varios ejemplos.

Una vez se han obtenido los parámetros de distorsión intrínsecos de la cámara, será posible la transformación de las imágenes, es decir, su calibración. En la *Figura 2.12* se muestran varias imágenes de bandejas y sus correspondencias etiquetadas. De la misma forma que se han transformado las imágenes, se deberá calibrar también los datos asociados a cada una, como la máscara de segmentación y los bounding box. Para ello, se les ha aplicado la misma matriz de calibrado.

Una vez se tiene el modelo de regresión y la red neuronal entrenados con estas imágenes y estos datos calibrados, el sistema ya se encontrará preparado para ser ejecutado sobre imágenes calibradas (únicamente).

No obstante, la salida del tallado se encuentra en píxeles. Concretamente, la salida

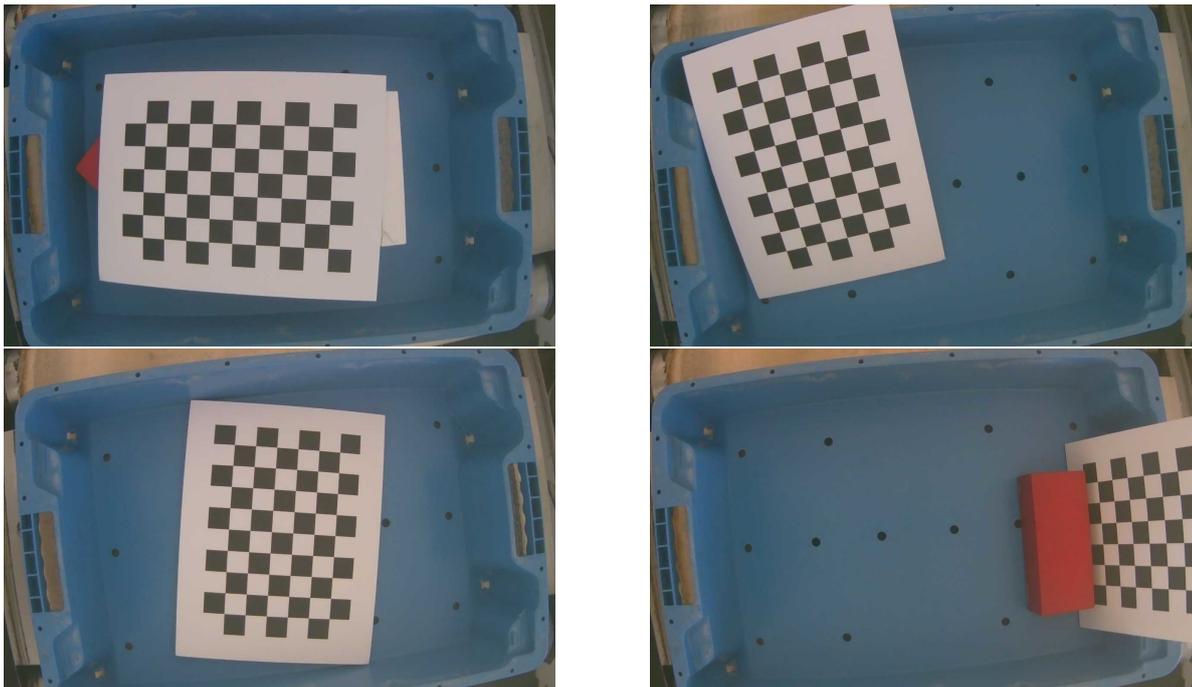


Figura 2.11: Ejemplos de calibrado de cámara con chessboard.

corresponderá a la longitud del espécimen en píxeles, como se ha visto en la *Figura 2.10*. Para su posterior conversión, se ha utilizado el tablero mostrado en la *Figura 2.11* y las imágenes de calibración obtenidas con este. Se conocen las dimensiones de cada cuadrado negro, concretamente de 250×250 milímetros. Mediante la suma de cuadrados y diferentes combinaciones comparativas pretendiendo minimizar errores, se halla un factor de conversión píxel a centímetro, el cual se utilizará directamente sobre la salida del modelo de regresión de píxeles, como será mostrado en la *sección de resultados 4.3*.

Un aspecto relevante a tener en cuenta es el error de calibrado. Pese a encontrarnos en un entorno con menor propensión a errores al tener una cámara estática, es necesario tener en cuenta diferentes factores que podrían propagar estas imprecisiones.



Figura 2.12: Ejemplos de imágenes no calibradas (*a, c, d, g, i, j*) y sus respectivas calibradas (*b, e, f, h, k, l*).

Error de calibrado

La situación actual tendría una mayor precisión si fuese un entorno en dos dimensiones, donde la profundidad fuese siempre la misma (por ejemplo, si se midiesen trozos de papel encima de una mesa). No obstante, se están tratando especies de peces con volumen considerable en una bandeja con volumen también. Al tomar imágenes de color únicamente, el volumen no se está teniendo en cuenta.

Debido a la profundidad de las bandejas, es posible encontrar especímenes apoyados en la base inferior de estas, como en la *Figura 2.13a*, normalmente debido a escasez de especies o pequeño tamaño de estas. También serán comunes bandejas con una mayor densidad, donde los especímenes visibles se encuentren más cerca de la parte alta de la bandeja, con una profundidad diferente a la anterior, por lo que podría dar errores en el calibrado, ver *Figura 2.13b*.



(a)

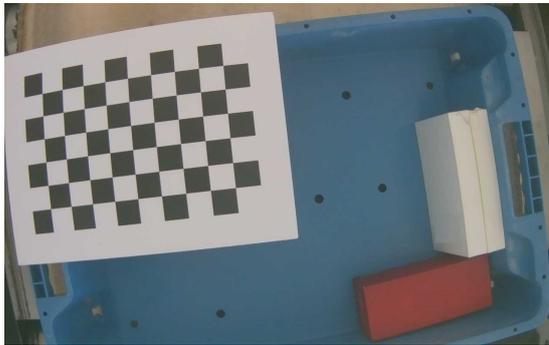


(b)

Figura 2.13: Ejemplos de especímenes en la zona inferior (a) y superior (b) de la bandeja.

Para abordar esta problemática, se ha asumido este error y se mostrará un rango esperado de talla y biomasa como error. Estos son mostrados en la *sección de resultados 4.3*. Se aportará una medida asumiendo que la especie se encuentra en la zona media de la bandeja, y estará comprendida en un intervalo entre la predicción si estuviese en la zona alta de la bandeja y si estuviese en la zona baja de la bandeja. Se han obtenido estos

nuevos factores de conversión píxel-centímetro a partir de las referencias de las imágenes del *chessboard* en estas zonas, como se puede ver en la *Figura 2.14*.



(a) Zona alta de la bandeja



(b) Zona baja de la bandeja

Figura 2.14: Imágenes del *chessboard* en zona alta y baja de la bandeja, preparadas para la obtención de los factores de conversión píxel-centímetro límite del rango de error de talla y biomasa.

3 Experimentación

Como ha sido realizado en otras secciones, la experimentación se va a dividir en tres grupos. Por un lado, se va a analizar la obtención de talla en entornos no calibrados (lonjas minoristas), por otro lado, en calibrados (lonjas mayoristas) y, finalmente, se comprobará la eficacia del método propuesto para abordar la identificación y segmentación de especies en lonjas mayoristas.

3.1. Tallado de especies en entorno no calibrado

El método propuesto es aplicable tanto para lonjas mayoristas como para minoristas, pero el mayor reto se va a encontrar en el entorno no calibrado, donde hay un mayor número de dificultades y ha sido menos explorado en la literatura. Por lo tanto, se hará un mayor análisis en el dominio minorista.

Dataset utilizado: Para llevar a cabo los experimentos, se ha utilizado el dataset de DeepFish [33], con imágenes recogidas en la lonja minorista de El Campello (Alicante, España), entorno no calibrado. Concretamente, se han utilizado 1,291 imágenes con bandejas de pescado (7,339 especímenes en estas), con 59 especies diferentes, con hasta 60 clases (debido al dimorfismo sexual se ha dividido el *Symphodius Tinca* en macho y hembra). Cada imagen contendrá información sobre la especie, el segmentado de cada instancia y marcas de la bandeja realizadas manualmente, como referencia para cálculos

de homografía.

Configuración experimental: El objetivo de la siguiente experimentación consistirá en seleccionar el modelo de regresión más apropiado para la estimación del tamaño de los peces, aprendiendo la calibración durante el entrenamiento a partir de los datos reales, o ground truth, generados mediante técnicas de metrología visual. Se han realizado cinco experimentos:

- Selección de un set de los mejores modelos de regresión en base a su rendimiento.
- Reducción del set con los hiperparámetros ajustados.
- Selección del algoritmo con mayor rendimiento en base a normalización de parámetros.
- Verificación del algoritmo ganador con bucles 10 k -fold.
- Comparación de resultados finales con mismo modelo utilizando esquinas de bandeja (simulación de entorno calibrado).

Entre los modelos de regresión utilizados en los siguientes experimentos, los que han obtenido unos mejores resultados son *GBR* [36], *Extra Trees* [37] y *CatBoost* [38]. Además, se ha utilizado el algoritmo *SVM* [39] como método comparativo, al tratarse de un algoritmo estándar utilizado en diferentes registros con buenos resultados, con intencionalidad de ver cuánta mejoría están obteniendo los diferentes modelos. Se ha utilizado *SVM* con kernel radial como ejemplo de buen rendimiento general, y con kernel lineal como método que consigue solucionar el problema, pero con bajos resultados de precisión, sirve como comparativa de resultados como límite inferior, esperablemente mucho más bajos.

Batería de modelos de regresión

En el primer experimento, se ha probado el comportamiento sobre el dataset de 25 diferentes modelos con buenos resultados en otros problemas. Se pueden observar los diferentes parámetros de error obtenidos de los mejores veinte modelos en la *Tabla 3.1*. A partir de estos, se van a elegir los seis mejores modelos para la siguiente comparativa.

Modelo	MAE (cm)	MSE	R2	MAPE	TT (Seg)
Extra Trees Regressor	1.8613	8.7115	0.7694	0.1173	0.101
CatBoost Regressor	1.8506	8.8161	0.7668	0.1172	1.211
Gradient Boosting Regressor	1.8504	9.3102	0.7544	0.1166	0.075
Random Forest Regressor	1.8830	9.5934	0.7474	0.1175	0.201
Light Gradient Boosting Machine	1.9224	9.5624	0.7471	0.1201	0.021
Extreme Gradient Boosting	1.9853	9.8369	0.7409	0.1252	0.076
K Neighbors Regressor	2.0806	10.1672	0.7312	0.1331	0.005
Linear Regression	2.5980	15.2516	0.6071	0.1656	0.127
Ridge Regression	2.5973	15.2517	0.6071	0.1655	0.003
Bayesian Ridge	2.5962	15.2524	0.6071	0.1655	0.003
Least Angle Regression	2.6365	15.518	0.5993	0.1676	0.003
Huber Regressor	2.4311	16.4486	0.5823	0.1585	0.005
Decision Tree Regressor	2.6236	17.0694	0.5577	0.1617	0.007
Lasso Regression	2.7333	19.2231	0.5162	0.1769	0.004
Elastic Net	2.7526	19.3541	0.5145	0.1768	0.003
Orthogonal Matching Pursuit	2.6763	21.8864	0.4456	0.1753	0.003
AdaBoost Regressor	4.3506	29.9264	0.1917	0.3018	0.035
Passive Aggressive Regressor	3.8979	31.7804	0.1534	0.2338	0.004
Lasso Least Angle Regression	4.3817	39.4312	-0.0028	0.2706	0.003
Dummy Regressor	4.3817	39.4312	-0.0028	0.2706	0.002

Tabla 3.1: Comparativa entre 25 modelos de regresión populares y muestra de sus resultados (se muestran únicamente los 20 mejores modelos).

Análisis de hiperparámetros

A partir de los resultados previos, los seis métodos con mejor rendimiento han sido seleccionados para comprobar su precisión con ajuste de parámetros. Como ha sido establecido, las máquinas de vectores de soporte (SVM) han sido incluidas como método comparativo base. Los resultados son mostrados en la *Tabla 3.2*

Regressor	Sin ajuste de parámetros				Con ajuste de parámetros			
	MAE (cm)	MSE	R ²	MAPE	MAE (cm)	MSE	R ²	MAPE
Extra Trees	1.8613	8.7115	0.7694	0.1173	1.8108	8.8154	0.769	0.1152
GBR	1.8504	9.3102	0.7544	0.1166	1.8339	8.7386	0.7705	0.116
CatBoost	1.8506	8.8161	0.7668	0.1172	1.8033	8.6005	0.7742	0.1144
LightBM	1.9224	9.5624	0.7471	0.1201	1.8780	8.9229	0.7649	0.1188
Random Forest	1.8830	9.5934	0.7474	0.1175	1.8329	9.0251	0.7645	0.1161
XGBoost	1.9853	9.8369	0.7409	0.1252	1.8452	8.8572	0.7662	0.1158
SVM	2.0654	157.179	0.6136	0.1311	1.9343	10.1436	0.736	0.1244

Tabla 3.2: Comparativa entre resultados de los seis mejores modelos considerados (y SVM), sin (*izquierda*) y con ajuste de hiperparámetros (*derecha*).

Nivel de normalización

Se ha considerado un experimento completo para la normalización debido al gran impacto que tiene en este conjunto concretamente. Los resultados se reflejan en la *Tabla 3.3*. Se compararán los mejores tres modelos de la prueba anterior, y SVM como modelo referencia. Intuitivamente, los modelos de regresión tienen una tendencia a obtener mejores resultados con el uso de normalización.

La razón detrás de un mejor rendimiento de *MinMax* por encima de la normalización *Standard* puede ser causada por las características de los valores atípicos de este dataset. Comúnmente, estos valores atípicos surgen debido a errores en la medida o instancias muy poco comunes. No obstante, en este dataset, hay algunas especies como *Sphyraena sphyraena*, con una presencia de 124 instancias en el dataset (correspondiente al 2% del total), con valores de tamaño oscilando entre 25 cm a 83 cm, con una longitud media de $45 \pm 12,62$ cm. Es decir, se trata de una especie con gran tamaño y capacidad de crecimiento. Como se puede ver en la *Figura 3.1*, es común que esta especie no quepa dentro de la bandeja. Respecto al dataset completo, los valores de tamaño son de 5 cm a 83 cm, con una media de longitud de $17 \pm 6,91$. Esto significa que cada instancia de *Sphyraena sphyraena* es un valor atípico en el dataset. No solo el mayor valor pertenece a esta clase, la instancia más pequeña encontrada, 25 cm, supera la media del dataset entero. Cualquier instancia de esta especie será más grande que la mayoría de otras especies recogidas en el dataset.



Figura 3.1: Ejemplo de *Sphyraena sphyraena* en el dataset.

Debido a estas características de valores atípicos fuera de lo común, donde una clase entera se encuentra representada únicamente por estos, no se va a querer reducir la importancia que aportan al modelo. En contraposición con otros algoritmos, que se basan en disminuir la significancia de los valores atípicos, *MinMax* no cambia el tamaño de la distribución, por lo que no reduce el peso que aportan estos valores.

Modelo de regresión	Sin scaling	Standard scaling input	MinMax scaling input	MinMax scaling I/O
GBR k -fold	1.8564	1.8564	1.8539	1.854
Extra Trees k -fold	2.0052	1.9969	1.9857	2.0119
SVM Tuned k -fold	4.3581	1.8471	1.8195	21.5391
CatBoost k -fold	1.7954	1.7710	1.7920	1.7824

Tabla 3.3: Comparativa del error absoluto medio, MAE, en centímetros entre los mejores modelos de regresión analizados y diferentes normalizaciones de datos de entrada y salida.

Mejores resultados

Una vez se ha identificado el modelo con mejores resultados a partir de las diferentes pruebas, se ha obtenido una vez más un conjunto de resultados, ver *Tabla 3.4* izquierda,

con los datos de entrada escalados mediante el algoritmo MinMax, y una validación 10 k -fold.

Los resultados mostrados pueden diferir ligeramente de la anterior prueba. En esta ocasión, como ya se ha decidido un modelo, se ha realizado el entrenamiento con una hiperparametrización más rigurosa, tiempos de computación más exigentes y se muestra la media de un bucle con 10 iteraciones de diferentes semillas para 10 k -fold, con el objetivo de evitar medidas mejores, o peores precisiones de las que deberían ser debido a la causalidad, y conseguir unos resultados más robustos.

Una vez más, se ha utilizado SVM. Esta vez, también con kernel lineal para tomar mejor perspectiva de la mejora en el desempeño realizado por los otros modelos.

Comparativa entorno no calibrado y calibrado

Finalmente, el último experimento será un comprobante del funcionamiento del modelo. Se van a comparar los resultados ya vistos en la *Tabla 3.4* izquierda y con el uso de parámetros adicionales de entrada en la *Tabla 3.4* derecha. Estos parámetros serán las coordenadas de los píxeles donde se encuentran las esquinas de las bandejas en las imágenes, datos que han sido utilizados para obtener las medidas “reales”, ground truth, utilizadas como salida esperada.

Utilizar estos parámetros extra va a simular un entorno calibrado, porque le estamos dando al modelo información de la situación 3D de la cámara en la imagen concreta. Estos cuatro puntos 2D (x, y) , ayudarán al modelo de regresión a tener un mejor entendimiento de los datos y reducirán errores en bounding box y áreas de segmentación ocasionados por perspectiva, distorsiones y variaciones en ángulos.

Esta configuración va a funcionar mejor, pero no va a ser posible su utilización en tiempo real en el sistema actual, los parámetros de entrada utilizados han sido obtenidos

a partir de etiquetados manuales. No obstante, sí que son muy útiles porque muestran el límite máximo de precisión al que aspira llegar el modelo no calibrado donde, idealmente, tendrá resultados cercanos pero no mejores. Es decir, poseer esta información ayuda a mejorar los resultados, será recomendable utilizarla siempre que sea posible.

Modelo de regresión	Sin calibración		Incluyendo calibración	
	MAE (cm)	R ²	MAE (cm)	R ²
GBR 10 <i>k</i>-fold	1,8501 ± 3,0099	0.7613	1,3304 ± 2,0937	0.8740
Extra Trees 10 <i>k</i>-fold	1,9715 ± 3,0396	0.7462	1,4098 ± 2,3239	0.8531
SVM Lineal 10 <i>k</i>-fold	2,6711 ± 4,4582	0.4746	1,8234 ± 3,3598	0.6996
SVM Radial 10 <i>k</i>-fold	1,8741 ± 3,1885	0.7307	1,2994 ± 2,2449	0.8620
CatBoost 10 <i>k</i>-fold	1.7614 ± 2.7633	0.7926	1.2713 ± 2.0616	0.8840

Tabla 3.4: Resultados finales con mejores modelos de regresión analizados con las tres entradas originales (bounding box en píxeles, área de segmentación y etiqueta de especie), *izquierda*. Adición de parámetros de calibración, esquinas de la bandeja en píxeles (x , y), *derecha*.

3.2. Identificación y segmentación en entornos mayoristas

Se va a justificar por medio de experimentación la validez del método propuesto para lonjas mayoristas, en comparativa con el método desarrollado para lonjas minoristas y adaptado a este entorno.

Dataset utilizado: Para llevar a cabo los experimentos, se ha utilizado el dataset DeepFish² [34]. Este contiene 876 imágenes de bandejas de pescado (con 3,154 especies etiquetadas en total) con 27 especies diferentes de la lonja mayorista de Altea (Alicante, España). Las imágenes han sido etiquetadas a nivel de pixel, utilizando una versión modificada de *Django Labeller* [7], ver ejemplo en la *Figura 3.2*.

El dataset completo ha sido dividido con un 80% de instancias para el conjunto

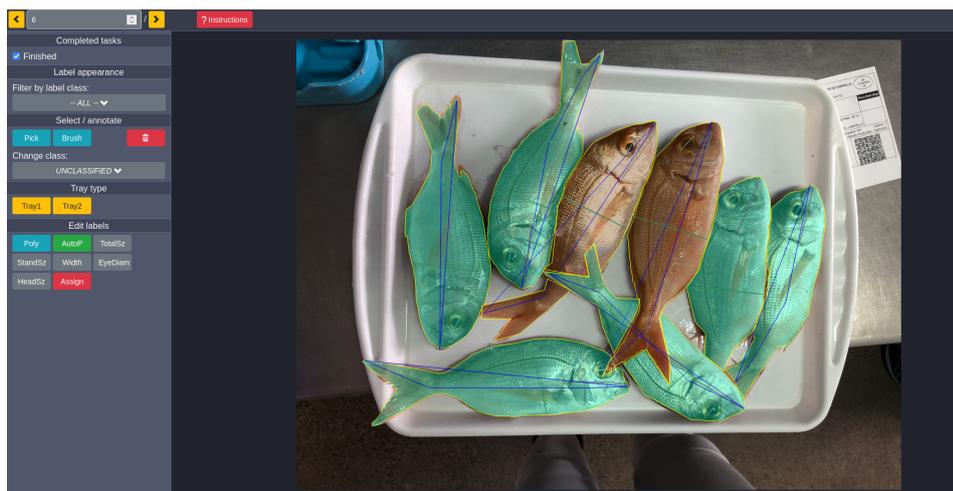


Figura 3.2: Entorno de etiquetado *Django Labeller* [7] modificado para este entorno.

Distribución	Bandejas	Anotaciones
Entrenamiento	648	2461
Validación	192	693

Tabla 3.5: Instancias totales e imágenes de bandejas de dataset utilizado, y distribución entre los conjuntos de entrenamiento y validación.

de entrenamiento y un 20% para el conjunto de validación. Como se puede observar en la *Figura 3.3*, el dataset presenta un gran desbalanceo entre clases. Debido a este desequilibrio, la división ha sido realizada por cada clase, procurando asignar el mismo porcentaje de instancias al conjunto de entrenamiento y validación. La distribución final se muestra en la *Tabla 3.5*.

En este dataset, el número total de instancias etiquetadas no va a coincidir con la cantidad real de instancias presentes en la imagen la mayor parte de las veces. De hecho, la cantidad real va a ser todavía mayor respecto de las instancias que aparecen en la fotografía.

Esta afirmación puede resultar contra intuitiva, porque algunas especies tienen tendencia de aparición constante, con una cantidad similar cada vez, mayormente debido al tamaño de estas. Por ejemplo, algunos peces grandes como *Seriola dumerilli* tienen bandejas con presencia de únicamente dos instancias, mientras que otras clases como

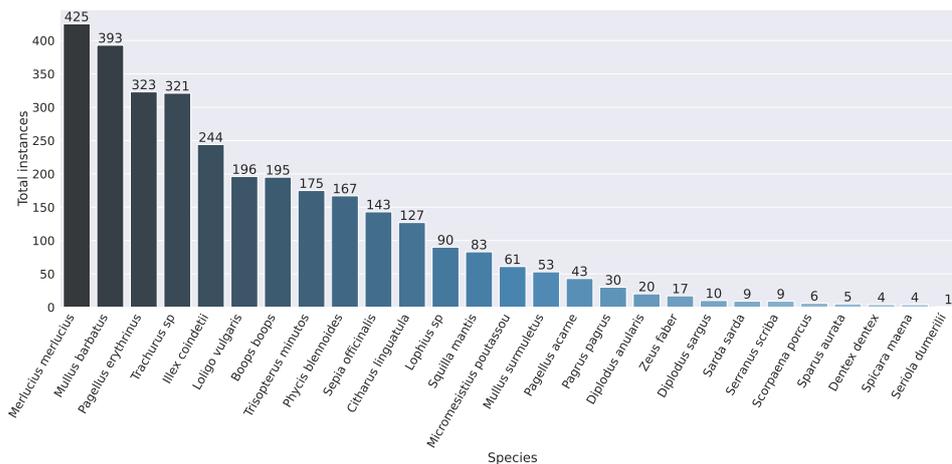


Figura 3.3: Frecuencia de cada especie en el dataset.

Phycis blennoides pueden tener bandejas con cincuenta o más instancias debido al menor tamaño, ver *Figura 3.4*.



(a) *Seriola dumerillii*



(b) *Phycis blennoides*

Figura 3.4: Ejemplo de bandejas con silueta negra representando las instancias etiquetadas, el ground truth que recibirá la red.

A pesar de la irregularidad en la frecuencia de aparición por cada especie en una imagen, la *Tabla 3.6* demuestra el desajuste entre la cantidad de especies por bandeja y el número de especies etiquetadas, donde se mantiene constante un rango entre una y cinco entidades etiquetadas por imagen, mientras que el ratio de aparición es mucho mayor.

A partir de lo expuesto anteriormente, la información mostrada en la *Figura 3.5* y la *Figura 3.6* indican el ratio de precisión esperada para algunas clases a partir de su

Distribución	Mínimo	Máximo	Media	Desviación estándar
Entrenamiento	1	22	3.598	2.3289
Validación	1	14	3.6094	2.2322

Tabla 3.6: Mínimo, máximo, media y desviación estándar de instancias en el dataset utilizado para cada imagen de una bandeja entre los conjuntos de entrenamiento y validación.

relación entre número de apariciones y número de imágenes.

Un ejemplo es la *Sepia officinalis*, con una de las mayores medias de aparición por bandeja del dataset, con 7 instancias por imagen. Esto significa que esa especie tiene siete entidades con gran calidad por imagen, porque solo han sido etiquetadas instancias únicamente si son muy reconocibles y tienen gran visibilidad. Debido a esto, el algoritmo necesitará un menor número de imágenes, que vendrán con gran densidad de información, y necesitará menos instancias para el aprendizaje, porque estas serán más fácilmente identificables. Esta clase representa la décima con mayor número de instancias (143 especímenes) y la número 14 con respecto del número de bandejas, con únicamente 20 imágenes. No obstante, como será mostrado en la *sección 3.2.1*, más precisamente en la *Figura 3.10*, un mayor número de instancias no estará correlacionado con una mayor precisión que obtendrá una clase. Puede favorecer, pero no será el único factor a tener en cuenta, y hay más que afectarán a la salida y se deberían tener en cuenta.

La “calidad” de los especímenes dependerá altamente en las características propias de estas. En este ejemplo, *Sepia officinalis* tiene un color, forma e incluso textura suficientemente diferente al resto de especies como para justificar mayores resultados con un menor número de imágenes.

Configuración experimental: Por motivos de capacidad computacional y especialmente temporal, se han llevado a cabo los primeros experimentos con un subconjunto del dataset reducido, para decidir las primeras direcciones a tomar y descartar posibles aproximaciones sin ser necesario ejecutar el dataset completo, lo cual habría llevado

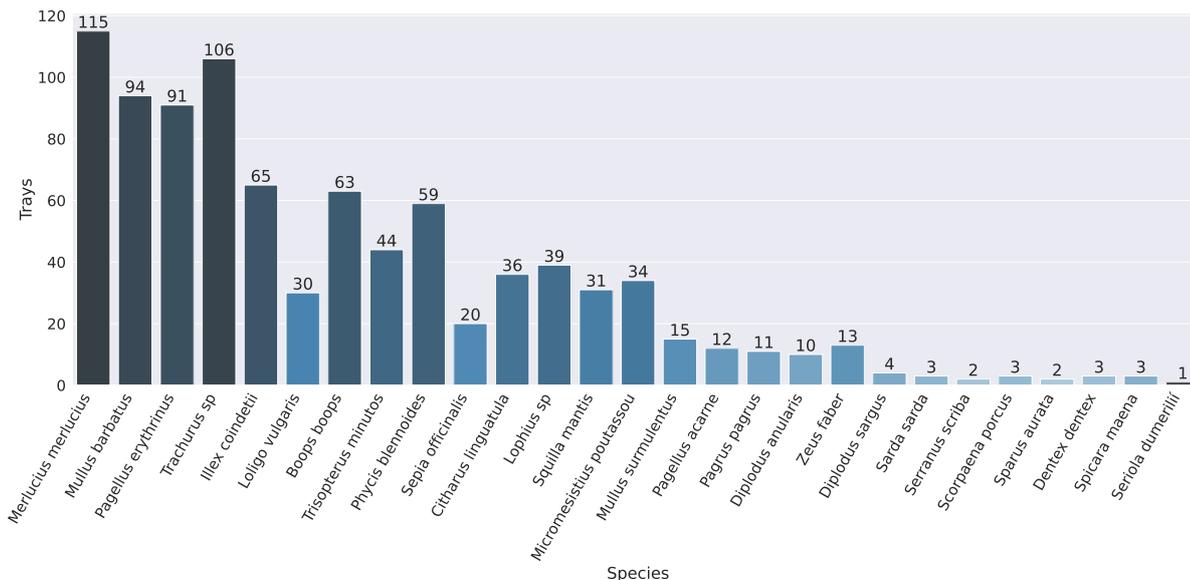


Figura 3.5: Cantidad de bandejas por cada especie en el dataset ordenadas por frecuencia de aparición en este.

tiempos no asumibles si se hubiese querido ejecutar la misma cantidad de pruebas.

El subconjunto del dataset principal, consistirá de 127 imágenes de bandejas de peces (con 556 especímenes etiquetados). El principal objetivo va a ser la obtención más apropiada del valor de los parámetros α y β , descritos en la *sección 2.2*. Una vez han sido clarificados esos parámetros, los siguientes experimentos han sido realizados con el dataset completo.

El objetivo de estas pruebas es la verificación de la hipótesis inicial de reducción de la penalización en contraste con el modelo original, esperando un mejor comportamiento en un entorno con alta presencia de instancias no etiquetadas, como el dataset sobre el que este trabajo se enfoca.

Se han llevado a cabo diferentes experimentos utilizando la mejor configuración de YOLACT++ analizada en [2] adaptada a este entorno, con modificaciones en la función de pérdida. La *Tabla 3.7* sintetiza los diferentes parámetros utilizados tanto en el modelo base como en el modificado. Se puede encontrar una descripción más exhaustiva de los parámetros en [2].

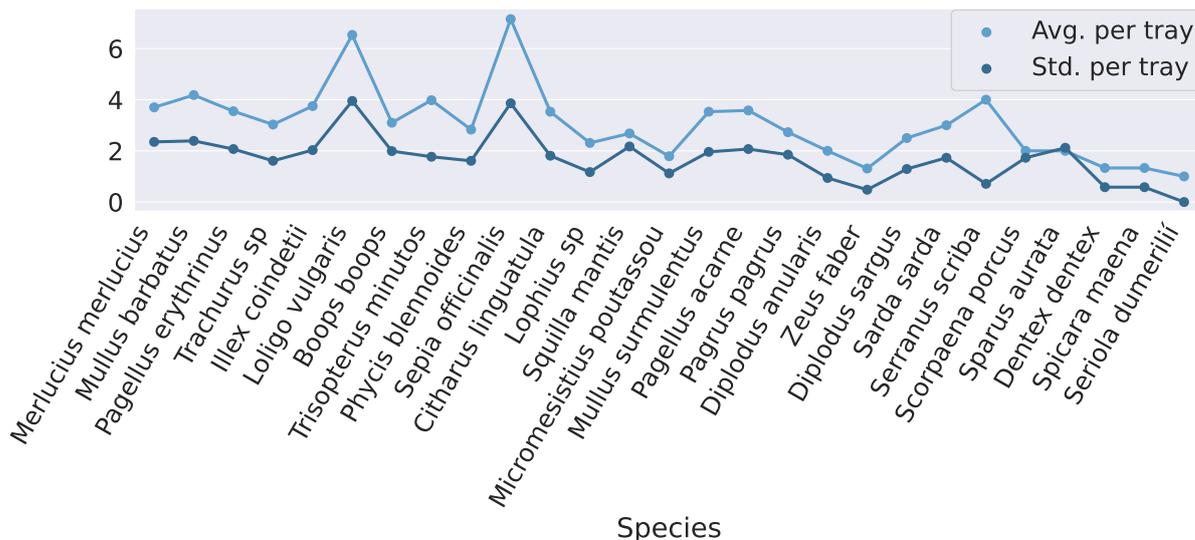


Figura 3.6: Media y desviación estándar de cada especie por bandeja.

Parámetro	Valor
Tamaño de entrada	550 × 550 píxeles
Tamaño de batch	8
Iteraciones	400,000
Optimizador	SGD
SGD (γ ; momentum; decay)	(0.1; 0.9; $5 \cdot 10^{-4}$)
Learning rate (inicial)	$1 \cdot 10^{-4}$
Learning rate (iteración 200k)	$1 \cdot 10^{-5}$
Learning rate (iteración 275k)	$1 \cdot 10^{-6}$

Tabla 3.7: Parámetros utilizados tanto en el modelo base como el modificado.

Elección de metaparámetros en dataset reducido:

Todos los test realizados sobre los metaparámetros α y β en el subconjunto reducido del dataset, han confirmado la hipótesis inicial. Esto demuestra que una penalización mínima (α y β igual a 0) reduce proporcionalmente la cantidad de mejora en cada paso, haciendo más lento el proceso de aprendizaje (es lógico, puesto que la red está recibiendo menos cantidad de información por paso). Este aspecto va a implicar un mayor requerimiento computacional, pero también se ha comprobado que aumentará el límite máximo de precisión media que puede alcanzar tanto el bounding box como la

segmentación de instancias. En resumen, se podría decir que “*aprende más lento, pero mejor*”.

Además, se han probado diferentes cambios en los parámetros α y β y sus restricciones. Por ejemplo, descartar predicciones cuya correspondencia de IoU no sea superior a cierto umbral, y penalizar cada instancia con mayor solape que cierto límite (no necesariamente el mismo umbral). Otro ejemplo ha sido la adición de la penalización original con menor peso, o incluso un peso ajustable a cada instancia, dependiendo del IoU que esta tiene sobre el ground truth.

Todos estos tests llevados a cabo se basan en penalizar de alguna forma falsos positivos que deberían ser verdaderos positivos debido a la falta de etiquetado, pero con menor importancia que otros, puesto que son incorrectos. De esta manera, se potencia las expectativas iniciales sobre un uso recomendado de ausencia total en la penalización, ignorando esos falsos positivos incorrectos, aunque los correctos también se ignoren.

Experimentación sobre dataset completo:

Por las razones mostradas en la anterior sección, los siguientes experimentos se van a realizar con el dataset completo ejecutado con *Yolact++* y las configuraciones de [2] adaptado a este entorno como el método base y la versión modificada de este con la versión de mínima penalización como comparativa.

Para una mejor visualización y consistencia de los resultados debido al desbalanceo de estos y la infra representación de algunas clases, como se puede ver en la *Figura 3.3*, el número de especies a utilizar se va a reducir de 27 a 17, manteniendo las clases con al menos 30 instancias en el dataset. De esta forma, se descartan diez especies con baja representación, la cual puede dar lugar a resultados demasiado buenos o negativos que pueden confundir, puesto que ciertas casualidades pueden influir altamente en el resultado, además de permitir un mayor enfoque en la red al tener menor número de clases, aumentando la precisión de las restantes clases.

		mAP medio					
		25 %	35 %	50 %	60 %	70 %	80 %
Y. Base		0.9336	0.9086	0.8597	0.8317	0.7531	0.6110
No Pen.		0.9651	0.9371	0.8969	0.8346	0.7030	0.4723

Tabla 3.8: Comparativa de mAP de bounding box entre *Yolact Base* y el método propuesto *No Penalización*.

		mAP medio					
		25 %	35 %	50 %	60 %	70 %	80 %
Y. Base		0.8148	0.7973	0.7403	0.6844	0.5789	0.4438
No Pen.		0.8958	0.8760	0.8142	0.7432	0.6348	0.4543

Tabla 3.9: Comparativa de mAP de máscara entre *Yolact Base* y el método propuesto *No Penalización*.

3.2.1. Comparativa de método base y no penalizado

Para realizar la comparativa, se ha utilizado una función del cálculo de la precisión en base a las N predicciones de los N ground truth únicamente (la función de precisión asociada al método *Key Fish*). De esta manera, se han obtenido datos comparables. Si se utilizase una función comparativa donde todas las predicciones intervienen en el cálculo de la precisión, como establece el método base, la diferencia en rendimiento sería mayor en favor del método propuesto, como se puede ver en la *Figura 3.7*. De forma más precisa, se va a comparar el método *Yolact Base* adaptado al nuevo entorno y con una fórmula de precisión adaptada, al método propuesto.

La principal conclusión extraída de las presentes pruebas ha sido la mejora en desempeño tanto para el cálculo de bounding box como de máscara de segmentación, perdiendo importancia en la mejora con un mayor límite de solape. De hecho, para altos IoU en bounding box, *Yolact Base* obtiene mejores resultados, como se puede ver en la *Tabla 3.8*. La predicción de la máscara de segmentación parece tener mejor ajuste comparativo y consistencia para este entorno, como se muestra en la *Tabla 3.9*.

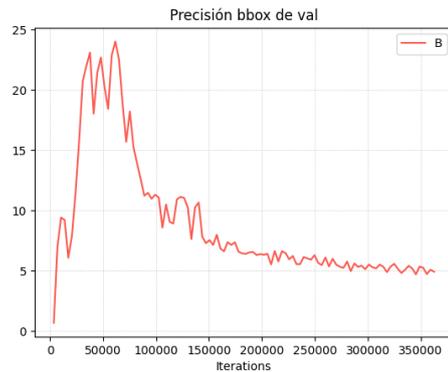
Se van a mostrar resultados con un 50% de solape, puesto que es un valor utiliza-

do en diferentes aplicaciones de visión y detección de objetos, como pueden ser otras líneas del proyecto DeepFish. Esta comparativa sobre el bounding box se puede ver en la *Figura 3.9* con un ligero incremento en desempeño desde el **85.97 %** al **89.69 %** en mAP. No obstante, *Trisopterus minutus*, *Loligo vulgaris*, *Trachurus sp* y *Merluccius merluccius* no mejoran su precisión para el método *No Penalización*, probablemente debido al incremento en el número de instancias de entrenamiento necesarias, con motivo de las peculiaridades de cada especie concreta. El método *No Penalización* sufrirá más la escasez de datos en una especie, puesto que aprende más lento y han sido entrenados con el mismo número de iteraciones para una comparativa más fiel.

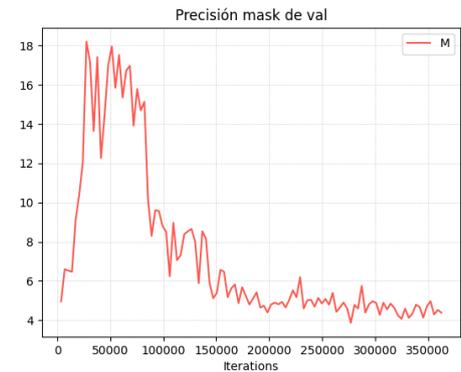
El experimento se ha llevado a cabo también para la segmentación de máscara, mostrado en la *Figura 3.10*. Cada una de las clases muestra consistencia y mejora los resultados respecto a *Yolact Base*. Los falsos positivos incorrectos tendrán un mayor impacto en la máscara de segmentación, donde el método *No Penalización* obtiene mejores resultados, al no estar influenciado negativamente por esta información confusa.

La *Figura 3.11* muestra la comparativa para la máscara de segmentación con mayores índices de solape, concretamente 75 %. El método propuesto obtiene una ligera mejora global sobre el método base, con únicamente *Illex condetii* y *Loligo vulgaris* como las especies con peores resultados, de entre las especies con al menos 100 instancias de entrenamiento, para poder tomar mejores conclusiones. Ambas son especies marrones de *calamares*. Estas dos especies tienen características únicas que les permiten un reconocimiento fácil respecto de otras para la diferenciación de un algoritmo, como se puede apreciar en la *Figura 3.8*. Además, combinado a un uso más eficiente de la información provista por el método base (no descarta datos en cada paso), y la existencia de similares condiciones de entrenamiento para ambos métodos, puede llevar a pensar que nos podemos encontrar en un escenario donde los dos métodos no han alcanzado el punto más alto de desempeño. Y, en el caso de que lo haya conseguido solo uno, lo más probable es que este sea el método base. Se puede ver un ejemplo de esto en las gráficas de precisión en el entrenamiento, en la *Figura 3.7*, parece que el método base ha alcanzado su pico,

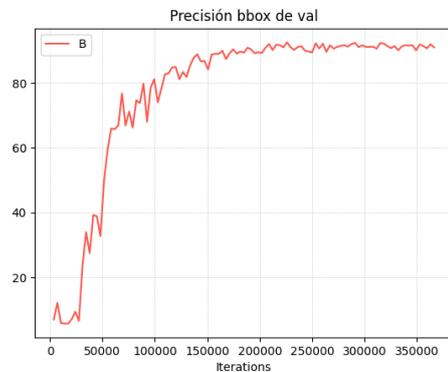
mientras que el método propuesto puede que tenga margen de crecimiento todavía.



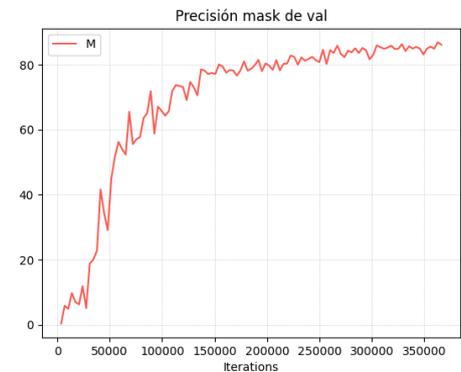
(a) Validación bounding box *Yolact Base*.



(b) Validación mask *Yolact Base*.



(c) Validación bounding box *No Penalización*.



(d) Validación mask *No Penalización*.

Figura 3.7: Gráficas de validación en el entrenamiento. Las condiciones no son comparables, para el cálculo del método *No Penalización* se ha calculado la precisión con las N instancias coincidentes con el ground truth como estipula su funcionamiento. Para el método base se ha comparado la precisión con todas las predicciones, como se espera de este.

Como se ha establecido en los párrafos anteriores, para menores índices de solape, la diferencia de rendimiento va a aumentar en favor del método propuesto. En la *Figura 3.12* se puede ver una comparativa con el 25% de solape.

Se han analizado más porcentajes de solape además de los ya comentados, más ejemplos se encuentran en la *Figura 3.13*. No hay más diferencias notables a destacar fuera de los ya analizados en los párrafos anteriores. Se mantiene la constante de un mejor

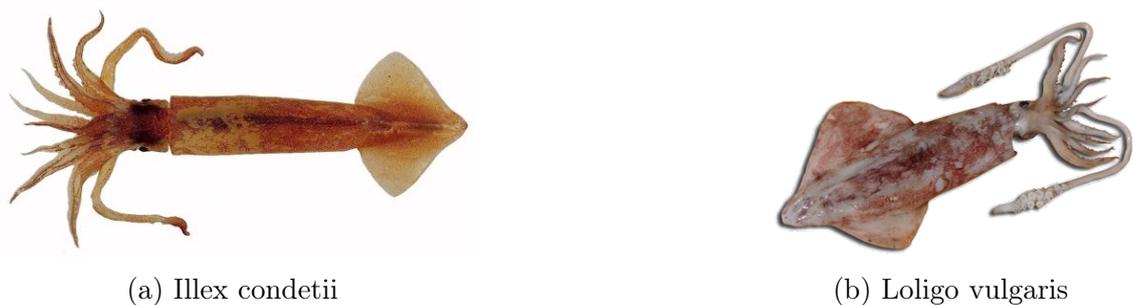


Figura 3.8: Ejemplo de *Illex condeitii* y *Loligo vulgaris*.

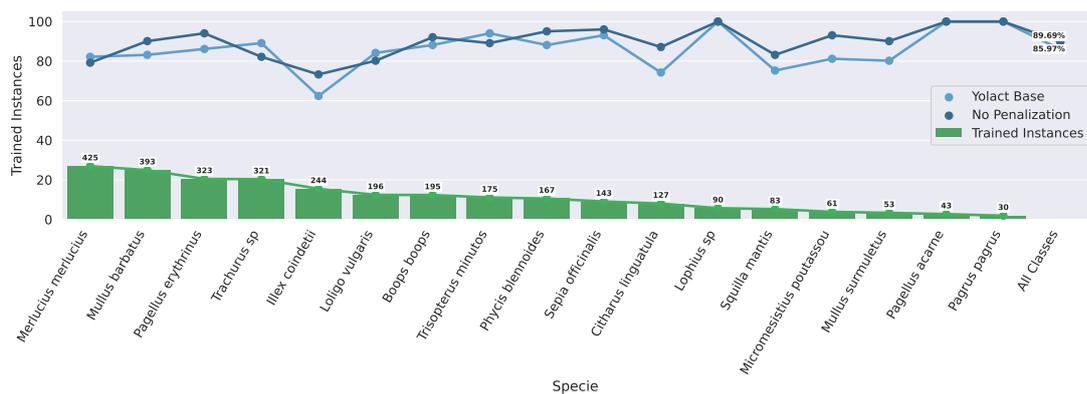


Figura 3.9: Comparativa entre *Yolact Base* y *No Penalización* para 50% de umbral para el bounding box.

desempeño del método propuesto incremental a medida que el límite de solape es menor, y esta diferencia es más notable en la segmentación de la máscara también.

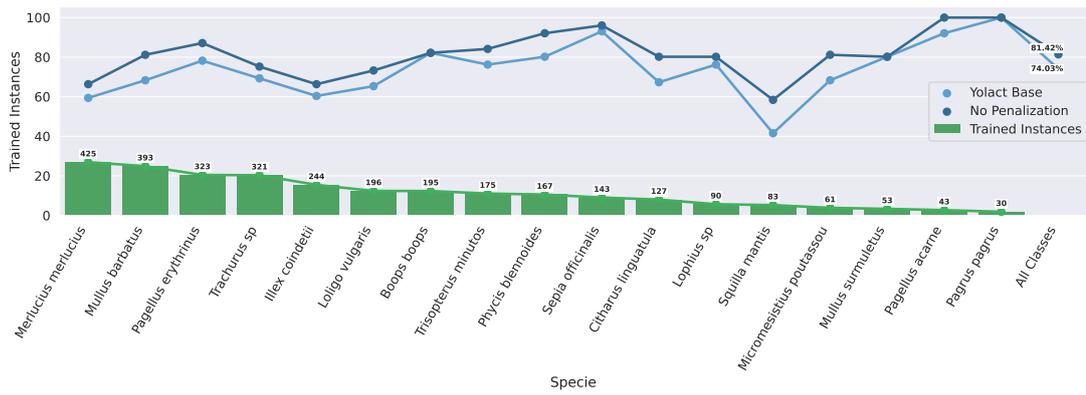


Figura 3.10: Comparativa entre *Yolact Base* y *No Penalización* para 50% de umbral para la máscara de segmentación.

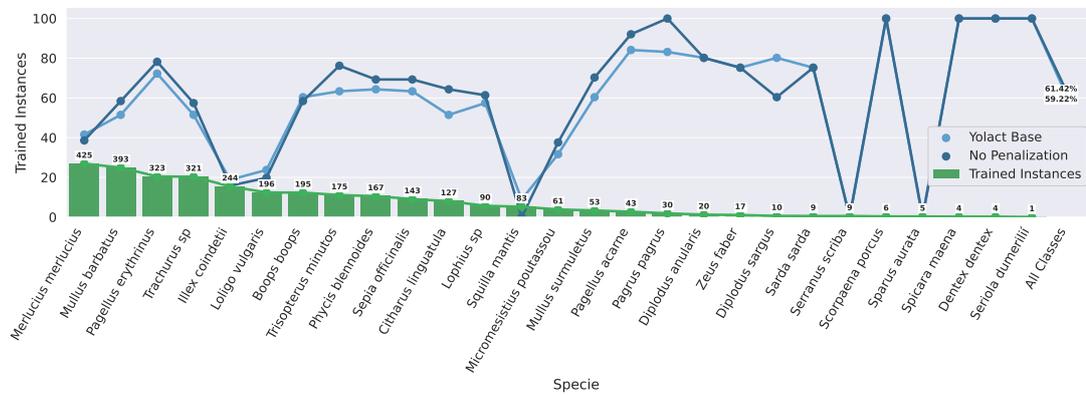


Figura 3.11: Comparativa entre *Yolact Base* y *No Penalización* para 75% de umbral para la máscara de segmentación.

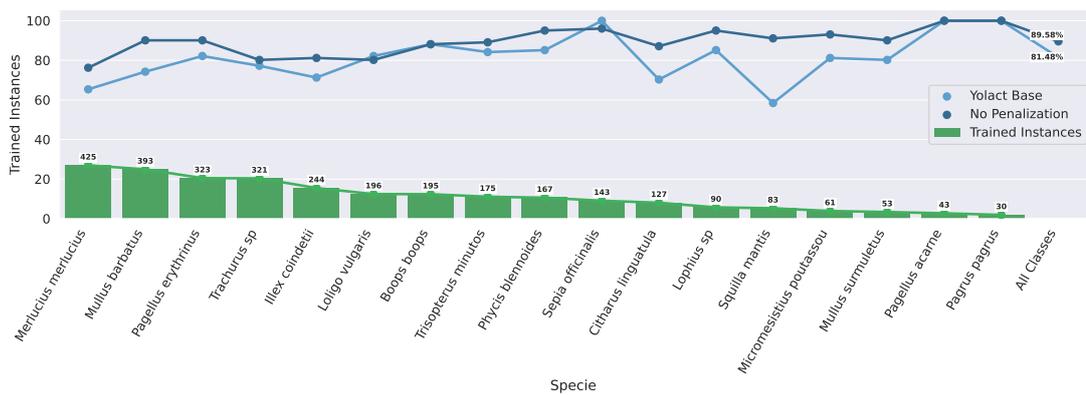
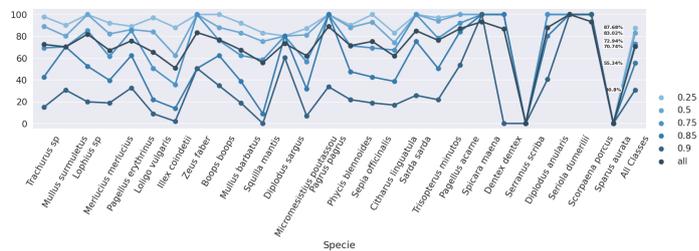
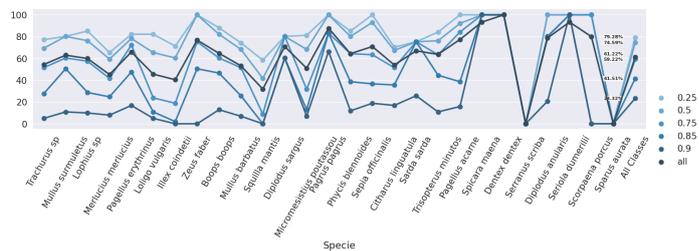


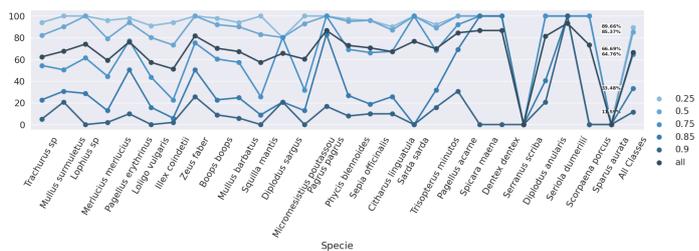
Figura 3.12: Comparativa entre *Yolact Base* y *No Penalización* para 25% de umbral para la máscara de segmentación.



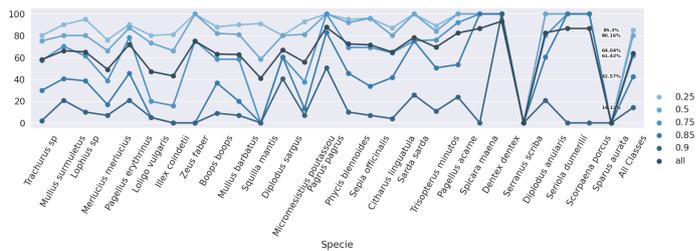
(a) *Yolact Base* bbox.



(b) *Yolact Base* mask.



(c) *No Penalización* bbox.



(d) *No Penalización* mask.

Figura 3.13: Precisión de *Yolact Base* bounding box (a) y máscara (b), y *No Penalización* bounding box (c) y máscara (d).

3.3. Tallado en entorno calibrado y método de segmentación e identificación *KeyFish*

Se va a adoptar el mismo sistema desarrollado para el entorno no calibrado, pero utilizando datos de calibración como entrada. En este caso se han calibrado las imágenes de entrada, tal cual se muestra en la *Figura 2.12*, como ha sido también expuesto en la *Tabla 3.4* realizado para las lonjas mayoristas.

El método de cálculo artificial de la talla ya no se ha obtenido a partir de técnicas de metrología visual. Si se trata de un entorno calibrado, una distancia X en píxeles siempre se corresponderá a la misma longitud Y en centímetros. Por lo tanto, no tendría sentido utilizar como ground truth tamaños obtenidos de esta forma. Afortunadamente, el método ya ha sido probado en la anterior *sección 3.1*. Volver a realizar una experimentación solo aportaría una confirmación de los resultados ya mostrados. Además, no se encuentran datos ground truth reales de las medidas en el dataset, puesto que son imágenes tomadas en un entorno real de la lonja, no ha sido posible su medición manual cuando se tomaron las fotografías, representaría un gran coste de recursos, temporal y económico, no sería asumible por el proyecto.

Se muestran resultados de la aplicación del método al entorno calibrado de lonjas mayoristas en la *sección resultados 4.3*. Estos se podrían considerar como los resultados finales, puesto que se combinan ambas ramas exploradas en este trabajo, se va a tallar el nuevo entorno mayorista.

4 Resultados

A continuación, se van a mostrar ejemplos de los resultados finales obtenidos y su comportamiento para las dos principales tareas en las que se ha enmarcado este trabajo. Además de la combinación de estos, el cual sería el tallado en el entorno mayorista con la aproximación propuesta de *KeyFish*, en un entorno calibrado.

4.1. Tallado en entorno no calibrado

Ejemplos son mostrados en la *Figura 4.2*. Se indica tanto el tallado en centímetros como su estimación de biomasa en gramos. Debido a la imposibilidad de calibración, la especie aportará una ayuda sobre el margen de tamaño que tendrá una especie, pero también resultará como una lacra para instancias lejanas a la media, como en la *Figura 4.1a*. En esta figura, a simple vista, la especie verde es más pequeña que el resto, pero parece que la diferencia en tamaño debería ser mayor de lo que se indica.



(a)



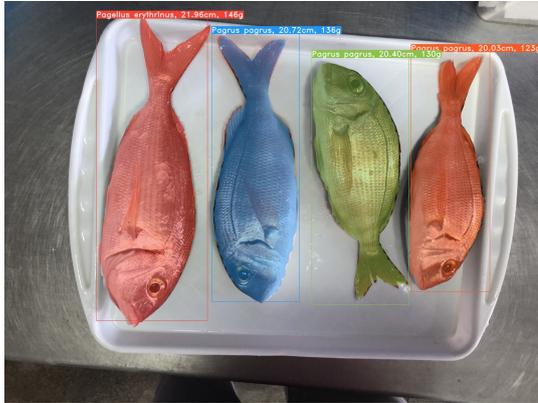
(b)



(c)



(d)



(e)



(f)



(g)



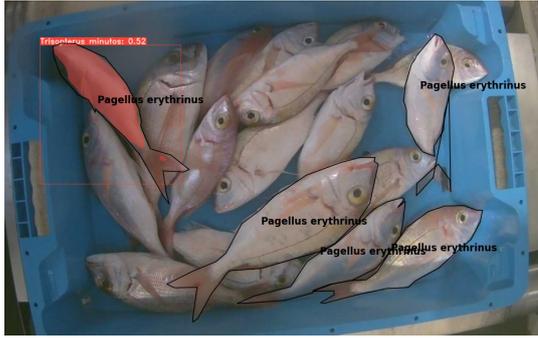
(h)

Figura 4.1: Ejemplos de talleo y estimación de biomasa sobre el entorno no calibrado.

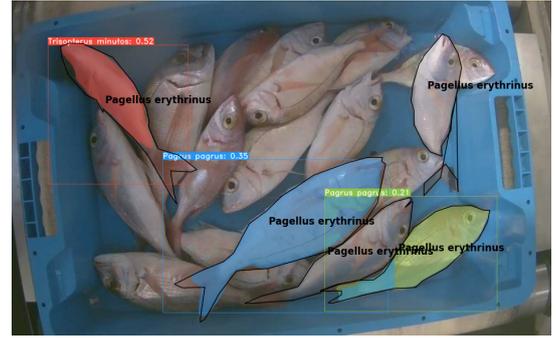
4.2. Identificación y segmentación en entorno mayorista

Se puede apreciar una comparativa en la *Figura 4.2*. El método base se encuentra más *ajustado* al dataset. Al entrenar con información errónea, aprenderá a cometer esos errores en lugar de generalizar. Se muestra en la *Figura 4.2a* y la *Figura 4.2b* donde las predicciones son pocas y presentarán una tendencia a coincidir con los ground truth (siluetas negras). Estableciendo que el algoritmo ha aprendido los ejemplos provistos con calidad pero no la abstracción de dichas especies, una evidencia de la confusión debida a entrenar con falsos positivos que deberían ser verdaderos positivos.

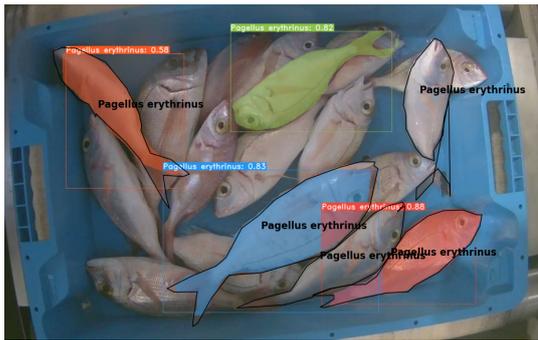
Por otro lado, el mismo ejemplo se muestra para el método propuesto *No Penalización* en la *Figura 4.2e* y la *Figura 4.2f* con una predicción más precisa en algunos ejemplos, e incluso etiquetando más instancias, especialmente instancias fuera de los ejemplos del ground truth. En general, *No Penalización* tiene más, y peores predicciones, pero el objetivo será quedarse únicamente las k predicciones con mayor confianza, convirtiendo este escenario uno más apropiado respecto a la aproximación con el método *Yolact Base*.



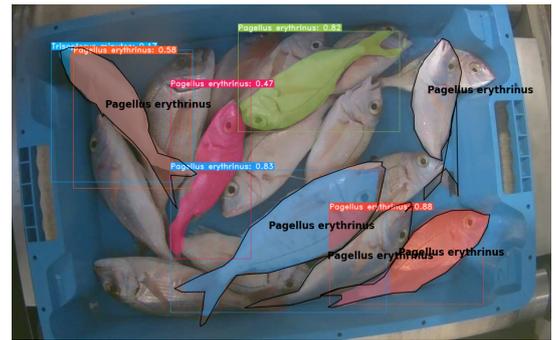
(a) *Yolact Base* 50% confianza.



(b) *Yolact Base* 15% confianza.



(c) *No Penalización* 50% confianza.



(d) *No Penalización* 15% confianza.



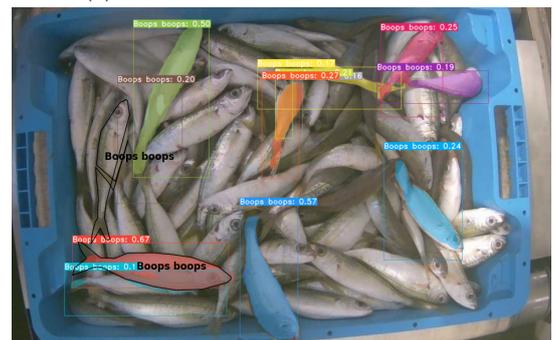
(e) *Yolact Base* 50% confianza.



(f) *Yolact Base* 15% confianza.



(g) *No Penalización* 50% confianza.



(h) *No Penalización* 15% confianza.

Figura 4.2: Dos ejemplos de ejecución del método *Yolact Base* (a,b,e,f) y *No Penalización* (c, d, g, h) sobre el 50% y 15% de límite de confianza, respectivamente.

4.3. Tallado en entorno calibrado con *KeyFish*

Se podrían considerar estos resultados como el objetivo final del trabajo, donde se unificarán los avances obtenidos en el tallado aplicados al nuevo entorno mayorista y el enfoque adaptado con los *KeyFish*. La *Figura 4.3* muestra ejemplos de bandejas de entorno mayorista con el sistema de tallado y cálculo de biomasa aplicado a un margen de error para cada especie, mediante el calibrado mencionado en la *sección 2.4*.



(a)



(b)



(c)



(d)



(e)



(f)

Figura 4.3: Resultados finales combinatorios de tallado y método *KeyFish* propuesto en lonja mayorista calibrada.

5 Conclusiones

En este capítulo se resumen los objetivos alcanzados en el proyecto resaltando las contribuciones resultantes del trabajo. Finalmente, se identifican las principales líneas futuras en las que se continuará trabajando en el contexto del proyecto.

5.1. Conclusión

En el presente trabajo se han planteado una serie de objetivos relacionados con la continuación de un proyecto basado en la identificación y tallado de especies de peces en lonja. El punto de partida ha sido un sistema funcional con identificación y segmentación en lonja minorista. Para su continuación, se han mejorado dos aspectos clave: el tallado de especies, el cambio de dominio a una lonja mayorista y la combinación de ambos, es decir, el tallado de especies en lonja mayorista.

Tallado de especies: Se ha elaborado un método consistente en la combinación de salidas para una predicción de medida en un entorno no calibrado con la principal intención de ahorrar costes. De forma comparativa, también se ha realizado esta implementación en el nuevo dominio, el cual se trata de un entorno calibrado.

La predicción de la talla consistirá en un modelo de regresión, con tres entradas: el bounding box, la máscara de segmentación y la clase. Estas serán obtenidas como salida

de la red neuronal sobre las imágenes. Además, se utilizarán estos cuatro datos para obtener la biomasa de las especies.

Si el método es entrenado en un entorno no calibrado, aprenderá a obtener el tallado directamente. Es decir, aprenderá a “calibrar” sin la necesidad de instalar un sistema de calibrado. En algunas situaciones, la instalación de cámaras fijas totalmente calibradas es inviable, como ocurre en algunas lonjas que carecen de espacio o recursos, y en las que resulta mucho más práctico obtener imágenes de cámaras móviles no calibradas tomadas desde diversos ángulos y distancias.

Utilizando el conjunto de datos DeepFish [33] para la validación, se ha demostrado que el modelo más adecuado en términos de capacidad de generalización y mínimo error es el regresor *CatBoost*, especialmente para el caso en que la entrada se normaliza con el enfoque *MinMax* y utilizando un ajuste de hiperparámetros. Se ha demostrado que los resultados pueden mejorarse aún más cuando se utilizan los puntos de esquina de la bandeja como entradas para el modelo de regresión, ya que esta información puede ser útil para corregir la perspectiva.

Sin embargo, dado que el sistema propuesto se basa principalmente en imágenes no calibradas, solo las imágenes de *ground truth* disponen de estos marcadores; es decir, las nuevas bandejas capturadas por el sistema en tiempo real carecen de estas referencias. En este entorno no calibrado y su simulación calibrada (el uso de puntos de referencia como entrada), el rendimiento varía de $1,7614 \pm 2,7633$ a $1,2713 \pm 2,0616$ como error medio en centímetros.

Por lo tanto, el método propuesto es adecuado para situaciones donde un calibrado no sea posible por cuestiones económicas o espaciales, pero se ha demostrado que un mismo entorno con capacidad de calibración mejora notablemente los resultados. Será recomendable siempre que sea posible, se obtendrá un rendimiento más fiable y consistente.

Método KeyFish: Se ha adoptado una nueva filosofía de etiquetado para abordar el nuevo dominio, se ha llamado *KeyFish*. El principal objetivo es la detección de un pequeño conjunto de especímenes con la mayor precisión posible. Para lograrlo, dejan de ser prioridad los falsos positivos en el entrenamiento y se admite un gran número de predicciones erróneas o de baja calidad, teniendo en cuenta que, tras un filtrado de las de mejor confianza, son resultados que mejoran el anterior enfoque adoptado, como se puede ver en la *Tabla 3.8* y la *Tabla 3.9*. Con esta metodología en mente, se ha creado un nuevo dataset y se ha modificado la arquitectura de red.

El método propuesto de *No Penalización* obtiene una mejora de rendimiento en la detección del bounding box si el umbral de solape es menor del 75%, mientras que la detección de la máscara obtiene mejores y consistentes resultados hasta un 80% de umbral.

Para unos resultados finales del 50% de solape (umbral estándar para diferentes aplicaciones) se obtiene una mejora del 3.72% y 7.39% respecto del 85.97% y 74.03% de bounding box y segmentación obtenidos, respectivamente, mediante el método base de *Yolact++*.

Con los experimentos realizados satisfactoriamente, se puede concluir demostrando el impacto negativo que aporta un dataset con una gran cantidad de falsos positivos erróneos, y el mayor impacto dañino que este tiene sobre la predicción de la máscara de segmentación respecto del bounding box.

Como se puede ver en la *Figura 3.3*, el dataset presenta un gran desbalanceo y la mayoría de especies presentan un bajo número de instancias. Dado que el método *No Penalización* es más exigente en cuanto a requerimiento de información, un aumento en el tamaño del dataset de entrenamiento podría incrementar más la diferencia de rendimiento entre el método base y el método propuesto.

5.2. Líneas futuras

Dado que este trabajo de fin de máster se enmarca en un proyecto de investigación, he considerado como líneas prioritarias ideas de continuación relativas a los temas expuestos. Estos son:

- **Explotación de datos.** Ya se ha desarrollado un sistema para capturar y analizar datos. Conforme se vayan almacenando, sería apropiado tratarlos y realizar diferentes visualizaciones y predicciones. No sería útil a nivel biológico para la conservación del ecosistema únicamente, este sistema podría ahorrar tiempo y dinero a los pescadores, puesto que sabrían en qué zonas pueden esperar cada día más peces de determinada especie y acudir directamente, ahorrando combustible y obteniendo mayor cantidad de pesca. Las lonjas también podrían beneficiarse de esto, al obtener predicciones de cuántos barcos y bandejas llegarán en determinado momento del año y tener una mejor publicidad u organización, tal vez contratando más personal en un periodo concreto por aumento en la carga de trabajo.
- **Conteo de instancias en lonja mayorista.** Mediante los *KeyFish* se ha establecido la posibilidad de realizar la estimación de especímenes totales en una bandeja a partir de los pesos de las instancias y la bandeja total. No obstante, no se han realizado pruebas en este trabajo porque la lonja no ha facilitado estos datos todavía. Se esperan unos resultados satisfactorios de esta predicción.
- **Unificación de bloques.** Se ha creado una nueva metodología de red para nuevos dominios, pero ahora mismo hay dos arquitecturas distintas. Se podrían unificar ambos datasets, con un parámetro indicando si la imagen pertenece a una lonja minorista o mayorista. De la misma manera, se podría entrenar una arquitectura de red conjunta que tal vez obtenga mejores resultados a partir de la generalización de ambos casos.
- **Ampliación del sistema.** Ya se ha desarrollado el sistema en las dos tipologías

de lonja existentes. La tarea de expansión a diferentes lonjas, tanto mayoristas como minoristas, debería resultar un desafío menor.

- **Uso de otras tecnologías.** Probar otra arquitectura que aparezca en el estado del arte o incluso para su asistencia en el etiquetado, pudiendo acelerar la segmentación o acercarla a un etiquetado automático únicamente con la necesidad de identificación y revisión humana, abaratando y acelerando el proceso. Esto se podría llevar a cabo con tecnologías recientes como Segment Anything Model (SAM) [40].
- **Entorno no calibrado.** Para mejorar la situación de las lonjas minoristas que carecen de un entorno calibrado, considero diferentes opciones:
 - Calibrar el entorno siempre que sea posible. Se migrará el sistema a un entorno calibrado, como ha sido realizado en las lonjas mayoristas que ya tenían un entorno con posibilidades de calibrado. El peor caso será donde no exista una diferencia notable, una mayor muestra de información relevante no podrá perjudicar al sistema. El caso más habitual será una mejora sustancial de los resultados.
 - Obtención de los puntos de la bandeja de forma automática. Actualmente, se están realizando de forma manual. Esta automatización permitiría calibrar el sistema de forma artificial donde no sea posible un calibrado físico. Como ya se ha comprobado en la *sección resultados 3.1*, es una información relevante para el modelo.
 - Utilización de las coordenadas de silueta y estimación de volumen para los modelos de regresión con otro tipo de compresión. Actualmente, se está reduciendo la información de segmentación a un número, el área, para el cálculo del tamaño y biomasa, mientras que el volumen no se está utilizando. Un mayor aprovechamiento de estos recursos podría mejorar los resultados.

Bibliografía

- [1] A. F. Fernandes, E. M. Turra, Érika R. de Alvarenga, T. L. Passafaro, F. B. Lopes, G. F. Alves, V. Singh, and G. J. Rosa, “Deep learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia,” *Computers and Electronics in Agriculture*, vol. 170, p. 105274, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169919311561>
- [2] N. E. Garcia-D’Urso, A. Galan-Cuenca, P. Climent-Pérez, M. Saval-Calvo, J. Azorin-Lopez, and A. Fuster-Guillo, “Efficient instance segmentation using deep learning for species identification in fish markets,” in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–8.
- [3] Z. Wu, R. Chang, J. Ma, C. Lu, and C.-K. Tang, “Annotation-free and one-shot learning for instance segmentation of homogeneous object clusters,” 2018. [Online]. Available: <https://arxiv.org/abs/1802.00383>
- [4] X. Xu, A. Shimada, H. Nagahara, R.-i. Taniguchi, and L. He, “Image annotation with incomplete labelling by modelling image specific structured loss,” *IEEE Transactions on Electrical and Electronic Engineering*, vol. 11, no. 1, pp. 73–82, 2016. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/tee.22190>

-
- [5] J. Ao, K. A. Ehinger, and Q. Ke, “Image amodal completion: A survey,” 2022. [Online]. Available: <https://arxiv.org/abs/2207.02062>
- [6] D. Huynh and E. Elhamifar, “Interactive Multi-Label CNN Learning With Partial Labels,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2020.
- [7] G. French, M. Fisher, and M. Mackiewicz, “Django labeller,” 2021. [Online]. Available: <https://github.com/Britefury/django-labeller>
- [8] A. Galán Cuenca, “Identificación de especies de pesca artesanal mediante visión artificial y deep learning,” 2021-09-29.
- [9] C. Vilas, L. Antelo, F. Martin-Rodriguez, X. Morales, R. Perez-Martin, A. Alonso, J. Valeiras, E. Abad, M. Quinzan, and M. Barral-Martinez, “Use of computer vision onboard fishing vessels to quantify catches: The iobserver,” *Marine Policy*, vol. 116, p. 103714, 2020.
- [10] A. Álvarez Ellacuría, M. Palmer, I. A. Catalán, and J.-L. Lisani, “Image-based, unsupervised estimation of fish size from commercial landings using deep learning,” *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1330–1339, 11 2019.
- [11] M. Sung, S.-C. Yu, and Y. Girdhar, “Vision based real-time fish detection using convolutional neural network,” in *OCEANS 2017 - Aberdeen*. IEEE, jun 2017, pp. 1–6.
- [12] M. Zhang, S. Xu, W. Song, Q. He, and Q. Wei, “Lightweight Underwater Object Detection Based on YOLO v4 and Multi-Scale Attentional Feature Fusion,” *Remote Sensing*, vol. 13, no. 22, p. 4706, nov 2021.
- [13] M. Pedersen, J. Bruslund Haurum, R. Gade, and T. B. Moeslund, “Detection of Marine Animals in a New Underwater Dataset with Varying Visibility,” in *Procee-*

- dings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019, pp. 18 – 26.
- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
- [15] D. Bradley, M. Merrifield, K. M. Miller, S. Lomonico, J. R. Wilson, and M. G. Gleason, “Opportunities to improve fisheries management through innovative technology and advanced data systems,” *Fish and Fisheries*, vol. 20, no. 3, pp. 564–583, may 2019.
- [16] FAO, *The State of Mediterranean and Black Sea Fisheries 2020*. FAO, dec 2020.
- [17] L. d’Armengol, M. Prieto Castillo, I. Ruiz-Mallén, and E. Corbera, “A systematic review of co-managed small-scale fisheries: Social diversity and adaptive management improve outcomes,” *Global Environmental Change*, vol. 52, pp. 212–225, 2018.
- [18] M. Palmer, A. Álvarez Ellacuría, V. Moltó, and I. A. Catalán, “Automatic, operational, high-resolution monitoring of fish length and catch numbers from landings using deep learning,” *Fisheries Research*, vol. 246, p. 106166, 2022.
- [19] J. Gladju, B. S. Kamalam, and A. Kanagaraj, “Applications of data mining and machine learning framework in aquaculture and fisheries: A review,” *Smart Agricultural Technology*, vol. 2, p. 100061, 2022.
- [20] A. Álvarez-Ellacuría, M. Palmer, I. A. Catalán, and J. L. Lisani, “Image-based, unsupervised estimation of fish size from commercial landings using deep learning,” *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1330–1339, 2020.
- [21] G. French, M. Mackiewicz, M. Fisher, H. Holah, R. Kilburn, N. Campbell, and C. Needle, “Deep neural networks for analysis of fisheries surveillance video and

- automated monitoring of fish discards,” *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1340–1353, 08 2019.
- [22] J. C. Prieto, H. Shah, A. J. Rosenbaum, X. Jiang, P. Musonda, J. T. Price, E. M. Stringer, B. Vwalika, D. M. Stamilio, and J. S. A. Stringer, “An automated framework for image classification and segmentation of fetal ultrasound images for gestational age estimation,” in *Medical Imaging 2021: Image Processing*, I. Išgum and B. A. Landman, Eds., vol. 11596, International Society for Optics and Photonics. SPIE, 2021, p. 115961N. [Online]. Available: <https://doi.org/10.1117/12.2582243>
- [23] X. Chen, M. He, T. Dan, N. Wang, M. Lin, L. Zhang, J. Xian, H. Cai, and H. Xie, “Automatic measurements of fetal lateral ventricles in 2d ultrasound images using deep learning,” *Frontiers in Neurology*, vol. 11, 2020. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fneur.2020.00526>
- [24] M. T. Cazzolato, J. S. Ramos, L. S. Rodrigues, L. C. Scabora, D. Y. Chino, A. E. Jorge, P. M. de Azevedo-Marques, C. Traina, and A. J. Traina, “The utrack framework for segmenting and measuring dermatological ulcers through telemedicine,” *Computers in Biology and Medicine*, vol. 134, p. 104489, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482521002833>
- [25] M. T. Cazzolato, J. S. Ramos, L. S. Rodrigues, L. C. Scabora, D. Y. T. Chino, A. E. S. Jorge, P. M. Azevedo-Marques, C. Traina, and A. J. M. Traina, “Semi-automatic ulcer segmentation and wound area measurement supporting telemedicine,” in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, pp. 356–361.
- [26] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact++ better real-time instance segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

- vol. 44, no. 2, pp. 1108–1121, 2022.
- [27] T. Durand, N. Mehrasa, and G. Mori, “Learning a Deep ConvNet for Multi-Label Classification With Partial Labels,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2019.
- [28] T. Cour, B. Sapp, C. Jordan, and B. Taskar, “Learning from ambiguously labeled images,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 919–926.
- [29] L. Qi, L. Jiang, S. Liu, X. Shen, and J. Jia, “Amodal instance segmentation with kins dataset,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3009–3018.
- [30] W. Gu, S. Bai, and L. Kong, “A review on 2d instance segmentation based on deep neural networks,” *Image and Vision Computing*, vol. 120, p. 104401, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0262885622000300>
- [31] J. C. Ángeles Cerón, L. Chang, G. O. Ruiz, and S. Ali, “Assessing yolact++ for real time and robust instance segmentation of medical instruments in endoscopic procedures,” pp. 1824–1827, 2021.
- [32] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [33] A. Fuster-Guilló, J. Azorín-López, N. E. García-D’urso, A. Galán-Cuenca, G. Soler-Capdepón, M. Vicedo-Maestre, J. E. Guillén-Nieto, and P. Pérez-Sánchez, “The DeepFish dataset,” Oct. 2021. [Online]. Available: <https://zenodo.org/record/6475675>
- [34] A. G. Cuenca, P. P. Sánchez, A. F. Guilló, and J. A. López, “Deepfish2 dataset,”

- Sep. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.7108754>
- [35] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [36] R. Zemel and T. Pitassi, “A gradient-based boosting algorithm for regression problems,” in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. MIT Press, 2000.
- [37] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, Apr 2006.
- [38] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “Catboost: unbiased boosting with categorical features,” in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018.
- [39] S. Suthaharan, *Support Vector Machine*. Boston, MA: Springer US, 2016, pp. 207–235.
- [40] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” 2023.