

ENTRENAMIENTO Y VALIDACIÓN DE LA CONFIABILIDAD DEL COMMUNITY FISH DETECTOR PARA LA IDENTIFICACIÓN TAXONÓMICA DE PECES EN EL PNN UTRÍA, PACIFICO COLOMBIANO

Estudiante: José Pérez
Director: Mateo López Victoria
Co-Director: Hernán Benítez
Asesor: Carlos Arcos

Pontificia Universidad Javeriana Cali, Facultad de Ingeniería y Ciencias
Eden
Pragma

RESUMEN

El monitoreo de las comunidades de peces en ecosistemas marinos ha dependido históricamente de censos visuales en persona, los cuales enfrentan limitaciones significativas en términos de escalabilidad y objetividad. Con el avance de las tecnologías de video submarino, el volumen de datos generados ha creado un gran reto en la capacidad y velocidad del procesamiento de la información generada. El presente trabajo describe el desarrollo y la validación de un flujo de trabajo automatizado para la detección y clasificación de especies de peces utilizando la herramienta Community Fish Detector (CFD). La metodología integra la filmación *in situ* mediante sistemas GoPro a resolución 4K, la extracción sistemática de fotogramas para reducir la redundancia temporal y la anotación manual de precisión mediante el software VGG Image Annotator (VIA). A partir de un proceso de curaduría de datos, se consolidó un dataset definitivo de 560 imágenes y 1,770 individuos. Se empleó la arquitectura de aprendizaje profundo RF-DETR Nano, entrenada mediante aprendizaje por transferencia (Transfer Learning) a lo largo de un ciclo controlado de 50 épocas. Los resultados demuestran que el modelo localmente especializado superó la línea base inicial, incrementando la precisión general de un crítico 0.56% a un óptimo 87.70%, y consolidando una Precisión Media Promedio (mAP50) del 88.00%. Asimismo, la inclusión de mapas de flujo óptico y sustracción de fondo redujo los errores por desenfoque cinético, estabilizando la identificación taxonómica de especies clave como *Thalassoma lucasanum*, *Stegastes acapulcoensis* y *Cephalopholis panamensis*. Este sistema proporciona una herramienta computacional robusta y automatizada de alta fidelidad para el monitoreo de la biodiversidad en el Pacífico Tropical Oriental, con gran potencial para escalar otras soluciones tales como estimaciones de tamaños poblacionales, cálculos de biomasa y evaluación de interacciones ecológicas.

Palabras clave: Visión artificial, aprendizaje profundo, monitoreo biológico, Pacífico colombiano, identificación de especies, RF-DETR Nano.

ABSTRACT

Monitoring fish communities in marine ecosystems has historically relied on in-person visual censuses, which face significant limitations in terms of scalability and objectivity. With the advancement of underwater video technologies, the volume of generated data has created a major challenge regarding the capacity and speed of processing the generated information. This work describes the development and validation of an automated workflow for the detection and classification of fish species using the Community Fish Detector (CFD) platform. The methodology integrates *in situ* videography via GoPro systems at 4K resolution, systematic frame extraction to reduce temporal redundancy, and high-precision manual annotation using VGG Image Annotator (VIA) software. Following a data cleaning process, a definitive dataset of 560 images and 1,770 individuals was consolidated. The deep learning architecture RF-

DETR Nano was trained using a transfer learning approach over a controlled cycle of 50 epochs. The results demonstrate that the locally specialized model significantly outperformed the initial baseline, increasing overall precision from a critical 0.56% to an optimal 87.70%, and achieving a Mean Average Precision (mAP50) of 88.00%. Furthermore, the incorporation of optical flow maps and background subtraction successfully mitigated motion blur errors, stabilizing the taxonomic identification of key species such as *Thalassoma lucasanum*, *Stegastes acapulcoensis*, and *Cephalopholis panamensis*. This system provides a robust, automated, and high-fidelity computational tool for biodiversity monitoring in the Eastern Tropical Pacific, with great potential to scale other solutions such as population size estimations, biomass calculations, and the evaluation of ecological interactions.

Keywords: Computer vision, deep learning, biological monitoring, Colombian Pacific, species identification, RF-DETR Nano.

INTRODUCCIÓN

La crisis global de biodiversidad y el agotamiento de los recursos pesqueros exigen métodos de monitoreo ambiental que sean rápidos, precisos y no invasivos (Akinawo, 2023; Abad-Vilchez *et al.*, 2026). En las últimas décadas, la ecología marina ha experimentado una transición tecnológica desde los métodos extractivos hacia técnicas ópticas de observación remota como el video submarino (Al Muksit *et al.*, 2022). El uso de Estaciones de Video Submarino con Cebo (BRUVS) y el Video Operado por Buzo (DOV) se han consolidado como estándares para evaluar la riqueza íctica sin alterar el comportamiento de los organismos (Harvey y Shortis, 1995; Varini *et al.*, 2024). Sin embargo, el éxito de estas tecnologías ha generado una saturación de datos visuales que dificulta su análisis oportuno (Farooq *et al.*, 2026). Se estima que el análisis manual requiere entre tres y cinco horas de trabajo experto por cada hora de grabación, lo que restringe la toma de decisiones de manejo rápidas (Marrable *et al.*, 2022; Saqib *et al.*, 2024).

Por consiguiente, la automatización del procesamiento de imágenes mediante inteligencia artificial (IA) es una necesidad operativa para la gestión ambiental moderna (Siddiqui *et al.*, 2018a). El aprendizaje profundo (Deep Learning) ha demostrado una capacidad excepcional para imitar la inteligencia humana en tareas de reconocimiento visual (Mustapha *et al.*, 2021). No obstante, el entorno submarino presenta retos únicos como la atenuación de la luz, la turbidez y el camuflaje extremo de los peces contra fondos complejos (Al Muksit *et al.*, 2022; L. Yang *et al.*, 2021). En este escenario, el entrenamiento de modelos robustos requiere de conjuntos de datos masivos (Dutta y Zisserman, 2019; Farooq *et al.*, 2026).

La herramienta Community Fish Detector (CFD), creada por Filippo Varini, surge como una respuesta colaborativa a este problema al consolidar un dataset base de más de 1.9 millones de imágenes acuáticas (Varini, 2025a). Este modelo de fábrica está diseñado para detectar la presencia genérica de peces (Varini, 2025a; U. Farooq *et al.*, 2026). Sin embargo, para alcanzar una identificación taxonómica precisa y confiable en regiones de alta complejidad abiótica y sustratos oscuros como el PNN Utría, es indispensable ejecutar un proceso de ajuste fino (fine-tuning) utilizando datos locales (Siddiqui *et al.*, 2018b; J. J. Tavera *et al.*, 2026b). El uso del software VGG Image Annotator (VIA) facilita este proceso al permitir el etiquetado geométrico eficiente de fotogramas mediante cajas delimitadoras (Dutta y Zisserman, 2019). A medida que la arquitectura RF-DETR Nano instalada en el CFD es expuesta de manera iterativa a estas anotaciones depuradas, se genera una mitigación del ruido de fondo y una optimización en los niveles de confianza de las probabilidades logísticas de clasificación interespecífica (Jalal *et al.*, 2020; Saqib *et al.*, 2024). En el contexto colombiano, donde la riqueza de especies de peces marinos en el Pacífico es ecológicamente densa, esta tecnología optimiza los tiempos de investigación manual y democratiza el acceso a análisis de alta fidelidad taxonómica (J. Tavera *et al.*, 2026a).

ÁREA DE ESTUDIO

La presente investigación se desarrolló en la Ensenada de Utría (Figura 1), situada en la costa norte del Pacífico colombiano, departamento del Chocó (Estupiñán *et al.*, 1990). Esta zona forma parte del Parque Nacional Natural (PNN) Utría, un área protegida que comprende una extensa superficie marítima y costera con ecosistemas de selva húmeda tropical y arrecifes coralinos (Gómez y Vieira, 1996). La ensenada se caracteriza por ser un cuerpo de agua somero y tranquilo, protegido del oleaje directo por una barrera rocosa (Estupiñán *et al.*, 1990). Las condiciones ambientales del área están marcadas por una alta pluviosidad y fuertes descargas de escorrentía continental que influyen en la salinidad y turbidez local (Zapata y Robertson, 2007; Tavera *et al.*, 2026).

El principal ecosistema coralino considerado para este estudio fue el de arrecifes coralinos, de los cuales La Chola es el principal (Figura 1). Se trata de, un arrecife ubicado en el costado oriental de la ensenada con una extensión de 500 m de largo por 300 m de ancho (Estupiñán *et al.*, 1990). En este sector predominan los corales del género *Pocillopora*, seguidos por *Psammocora stellata* y diversos corales masivos que proporcionan refugio a una alta diversidad íctica (Estupiñán *et al.*, 1990; Gómez y Vieira, 1996). Otras zonas de muestreo incluyeron Playa Blanca, Punta Diego y Cocalito, donde existen fondos arenosos y rocosos con pendientes abruptas (Estupiñán *et al.*, 1990; López-Victoria y Tavera, 2018). La ictiofauna de Utría incluye familias dominantes como Labridae, Pomacentridae y Carangidae, muchas de las cuales son fundamentales para la seguridad alimentaria de las comunidades costeras (Estupiñán *et al.*, 1990; Tavera *et al.*, 2026a).

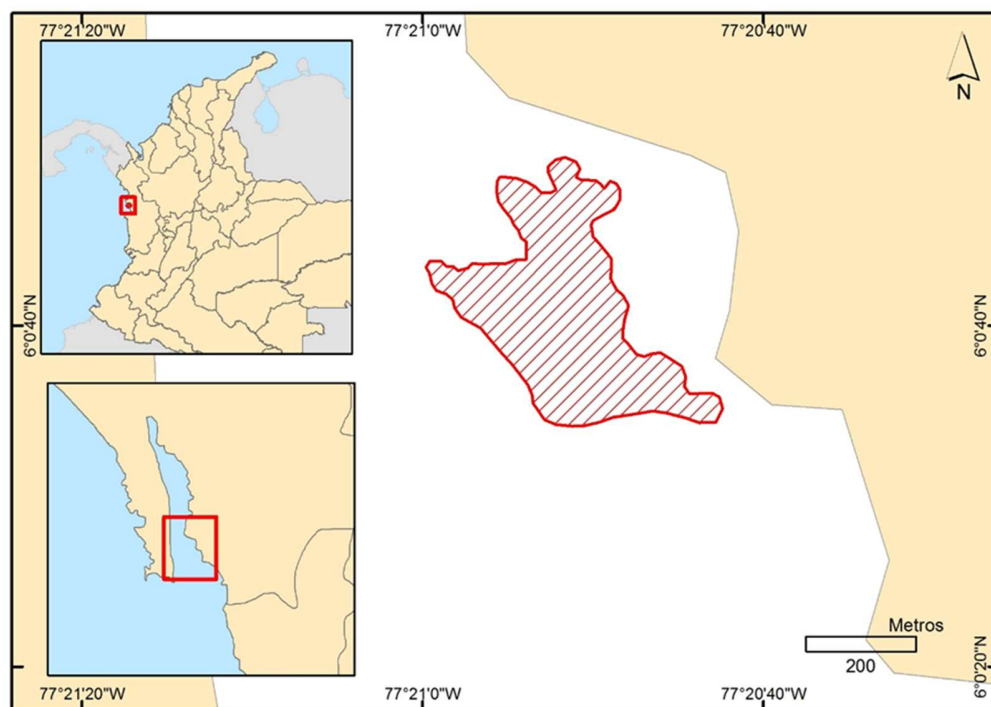


Figura 1. Mapa de ubicación y estación de muestreo en el PNN Utría. Recuadro sup izq.: Localización general de la ensenada en el Chocó. Recuadro inf. Izq.: Punto de monitoreo íctico (Cocalito, Punta Diego y Playa Blanca). Recuadro der.: Arrecife La Chola. Escala en kilómetros

MATERIALES Y MÉTODOS

Tipo y Diseño de Investigación

La presente investigación se enmarca en un diseño de carácter experimental, cuantitativo y tecnológico aplicado. Es experimental y cuantitativo debido a que involucra la manipulación controlada de hiperparámetros de redes neuronales y la evaluación rigurosa de su rendimiento mediante métricas estadísticas estandarizadas. Asimismo, es tecnológico aplicado, ya que implementa y optimiza arquitecturas avanzadas de aprendizaje profundo (Deep Learning) híbridas (convoluciones y transformadores) para automatizar el monitoreo biológico y la identificación taxonómica de la ictiofauna en un ecosistema marino estratégico (Safei *et al.*, 2022).

Adquisición de Datos en Campo

Para el muestreo visual subacuático se utilizaron cámaras de alta definición GOPRO Hero 13 Black, configuradas a una resolución nativa de 4K (3840x2160 píxeles) y una tasa de refresco de 60 fotogramas por segundo (fps), garantizando la captura de alta fidelidad de los caracteres morfológicos diagnósticos. Las filmaciones se concentraron en las horas de la mañana, estrategia metodológica orientada a optimizar la penetración de la luz solar en la columna de agua y a coincidir con los picos de mayor actividad biológica y forrajeo de las especies diurnas de la región (Giraldo *et al.*, 2009; Tarling *et al.*, 2022b).

Procesamiento de Video y Curaduría del Dataset (Data Cleaning)

A partir de los archivos de video crudos en formato MP4, se realizó una extracción sistemática de fotogramas estáticos con el objetivo primordial de mitigar la redundancia temporal de los datos y optimizar la variabilidad computacional (Dutta y Zisserman, 2019; Tarling *et al.*, 2022). Las imágenes resultantes se sometieron a un protocolo estricto de curaduría y limpieza. De un universo inicial de 790 registros indexados en el historial, se excluyeron aquellos correspondientes a archivos físicos inexistentes, tomas con opacidad crítica o carentes de fauna visible. De este modo, se consolidó un dataset definitivo de 560 imágenes, las cuales albergan un total de 1,770 individuos de peces anotados.

Anotación Taxonómica (Ground Truth)

Las imágenes limpias fueron importadas al software de código abierto VGG Image Annotator (VIA) (Dutta y Zisserman, 2019). Dentro de esta plataforma, se definieron manualmente las regiones de interés (ROI) mediante cajas delimitadoras (bounding boxes) que encerraron estrictamente la silueta anatómica de cada pez (Mimura y Nakamura, 2023). A cada geometría espacial se le asignó su correspondiente identidad taxonómica a nivel de especie, empleando como marco de referencia los listados e inventarios científicos verificados para el Pacífico colombiano (Tavera *et al.*, 2026b). Para corregir errores de digitación y sinonimias presentes en el archivo JSON original de VIA (como variaciones tipográficas en *Thalassoma lucasanum* o *Stegastes acapulcoensis*), se programó un script de normalización en Python que unificó las categorías bajo una estructura limpia de clases numéricas. En el caso de especies ecológicamente dominantes como el arcoíris (*Thalassoma lucasanum*), se estructuraron cientos de individuos específicos con el fin de robustecer la capacidad de generalización del modelo frente a los marcados cambios ontogénicos que exhibe la especie entre sus fases iniciales y terminales de coloración (Estupiñán *et al.*, 1990; Siddiqui *et al.*, 2018b).

Arquitectura del Modelo y Entrenamiento por Transferencia

El análisis computacional se fundamentó en la herramienta Community Fish Detector (CFD), la cual integra bloques convolucionales (CNN) para la extracción de características locales y capas de

transformadores para capturar dependencias espaciales globales (Varini, 2025b). Específicamente, se implementó el modelo RF-DETR Nano, una arquitectura optimizada para la inferencia eficiente en tiempo real a una resolución de operación de 640 x 640 píxeles (Varini, 2025b; J. Yang *et al.*, 2024).

Para el desarrollo del detector, se aplicó la técnica de Aprendizaje por Transferencia (Transfer Learning), inicializando la red con los pesos matemáticos optimizados del dataset CFD global, el cual cuenta con un respaldo de 1.9 millones de imágenes de ambientes acuáticos (Varini, 2025a; C. Silva *et al.*, 2022). El ajuste fino (Fine-Tuning) se ejecutó localmente utilizando el dataset estructurado del PNN Utría, forzando al algoritmo a especializarse en la discriminación fina de especies morfológicamente afines o simpátricas que comparten el mismo nicho visual (Siddiqui *et al.*, 2018b; Tavera *et al.*, 2026b). El entrenamiento se extendió por un ciclo controlado de 50 épocas con un tamaño de lote (batch size) de 16.

Flujo de Entrenamiento Avanzado para Dinámica Cinética

Con el propósito de mitigar los errores de detección derivados del movimiento biológico, se integraron mapas de flujo óptico y algoritmos de sustracción de fondo de manera directa en el flujo de tensores de entrenamiento (Jalal *et al.*, 2020; Salman *et al.*, 2020). Esta adición metodológica permitió dotar a la red neuronal de descriptores temporales, facilitando la identificación y delimitación precisa de organismos que presentaban distorsión cinética o desenfoque por velocidad (motion blur) dentro del plano, una limitación física persistente en especies de nado rápido y aceleración súbita como los miembros de la familia Carangidae (Estupiñán *et al.*, 1990; Jalal *et al.*, 2020).

Establecimiento de la Línea Base y Métricas de Evaluación

Para validar estadísticamente el impacto del entrenamiento local, se determinó inicialmente una Línea Base (Baseline) ejecutando el modelo en su estado de fábrica. Dicha evaluación arrojó un rendimiento crítico (Precisión: 0.56%, Sensibilidad: 1.02%, y 3,220 alertas falsas debido al ruido estructural del fondo), validando de forma científica la necesidad del ajuste fino. El rendimiento final del sistema se auditó sobre un conjunto de prueba independiente aplicando la métrica estandarizada de Intersección sobre Unión (IoU ≥ 0.50) (Padilla *et al.*, 2020). Los índices de calidad matemática se calcularon mediante las siguientes variables cuantitativas:

1. Precisión (Precision)

Determina la exactitud de los recuadros predictivos generados por la IA, penalizando las falsas alarmas provocadas por elementos abióticos del arrecife (sombras o rocas):

$$\text{Precisión} = \text{Verdaderos Positivos VP} / (\text{Verdaderos Positivos VP} + \text{Falsos Positivos FP})$$

2. Sensibilidad (Recall)

Evalúa la capacidad de exhaustividad del sistema para capturar la abundancia de la población real, penalizando las omisiones de especímenes camuflados o distantes:

$$\text{Sensibilidad (Recall)} = \text{Verdaderos Positivos VP} / (\text{Verdaderos Positivos VP} + \text{Falsos Negativos FN})$$

3. Precisión Media Promedio (mAP) e Índice F1

Se extrajo la Precisión Media Promedio (mAP50) y el F1-Score como indicadores del equilibrio de rendimiento general del clasificador frente a la variabilidad de las clases (Siddiqui *et al.*, 2018; Farooq *et al.*, 2026). Asimismo, el nivel de certeza asignado a cada etiqueta se definió formalmente a través de la

probabilidad logística normalizada computada por la función de activación final (Softmax) de la red neuronal (Villon *et al.*, 2018; K. Yang *et al.*, 2024).

RESULTADOS

Resultados de la Línea Base (Baseline)

La evaluación inicial del modelo genérico Community Fish Detector, ejecutada sobre el set de datos de la Ensenada de Utría antes de iniciar el proceso de ajuste fino, arrojó los siguientes resultados cuantitativos absolutos y porcentuales:

Tabla 1. Métricas estadísticas e individuos de error absoluto obtenidos durante la evaluación de la línea base (baseline) del modelo genérico sin entrenar.

Verdaderos positivos VP	18 individuos
Falsos positivos FP	3220 individuos
Falsos negativos FN	1753 individuos
Precisión (Precision)	0.56%
Sensibilidad (Recall)	1.02%
F-1 Score	0.72%

Comportamiento y Convergencia de las Funciones de Pérdida

Durante las 50 épocas de entrenamiento programadas, el comportamiento de los errores del sistema se registró a través de las funciones de pérdida multitarea del algoritmo. En la Época 1, la pérdida de localización (train/box_loss) inició en un valor de 2.52, mientras que la pérdida de clasificación de clases taxonómicas (train/cls_loss) se situó en 5.45.

A medida que avanzaron los ciclos de iteración, ambas variables presentaron una pendiente decreciente constante. Al alcanzar la época final del entrenamiento (Época 50), la función train/box_loss se redujo a un valor de 1.50 y la función train/cls_loss disminuyó hasta un índice numérico de 1.24. Las curvas de pérdida correspondientes al conjunto de validación siguieron una tendencia paralela descendente, estabilizándose de forma horizontal en el último tramo del entrenamiento.

Como se representa visualmente en la (Figura 2), las gráficas del bloque izquierdo muestran el descenso asintótico de los tres componentes de pérdida computados. La pérdida de caja (train/box_loss) y la pérdida de clasificación (train/cls_loss) exhiben una alta tasa de convergencia en las primeras 15 épocas, seguida de una estabilización gradual sin ganchos de rebote en las curvas de validación. Paralelamente, las gráficas del bloque derecho muestran que los indicadores de acierto estadístico (mAP50 y Precision) describen una trayectoria ascendente con comportamiento de meseta horizontal regular a partir de la época 40.

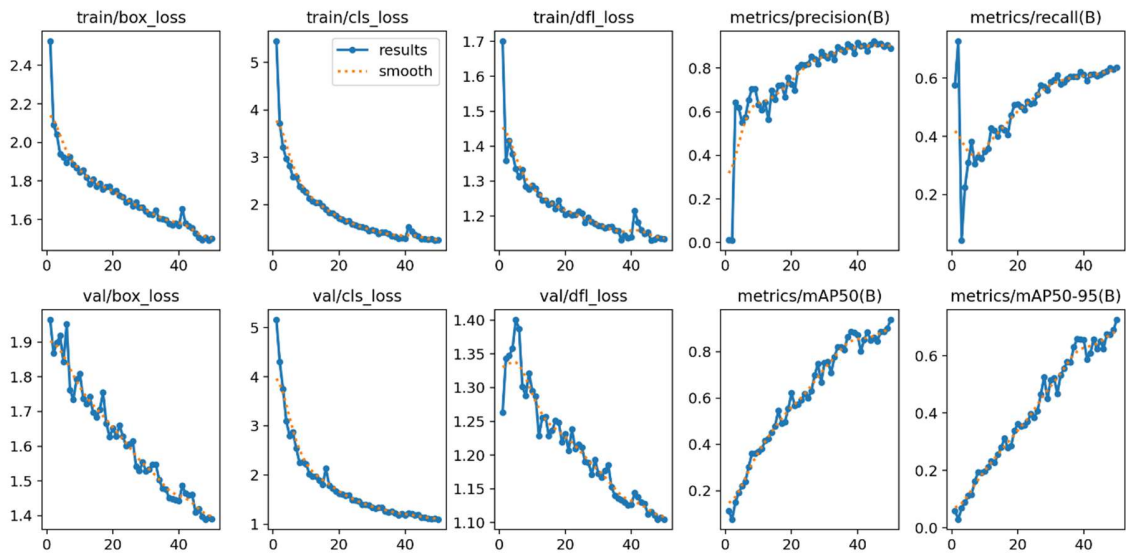


Figura 2. Curva de funciones de pérdida histórica (train y val para box_loss, cls_loss y df_l_loss) y evolución de los parámetros de validación estadística (Precision, Recall, mAP50-95) a lo largo de un ciclo de 50 épocas con la arquitectura RF-DETR Nano.

Métricas de Rendimiento Evolutivo

El rendimiento del clasificador experimentó modificaciones cuantitativas directas entre la fase previa al entrenamiento y los ciclos subsecuentes. Los cambios registrados en las métricas de precisión, sensibilidad y precisión media promedio (mAP50) se consolidan cronológicamente en la siguiente tabla:

Tabla 2. Comparación cuantitativa de las métricas de rendimiento computacional entre la línea base y las fases de entrenamiento del modelo RF-DETR Nano.

Fase del proyecto	Precisión (Precision)	Sensibilidad (Recall)	Rendimiento global (mAP50)
Línea base	0.56%	1.02%	0.72%
Época 1	1.10%	57.60%	11.20%
Época 50	89.08%	63.65%	93.73%

Como se detalla en el registro, la variable de Precisión incrementó del 0.56% inicial al 1.10% en la primera época, alcanzando un valor final de 89.08% en la Época 50. Por su parte, la Sensibilidad (Recall) pasó del 1.02% al 57.6% en el ciclo de apertura, consolidando un 63.65% hacia el final del proceso. La métrica global mAP50 se estableció en un 93.73% de efectividad al estabilizarse las curvas de rendimiento.

Matriz de Confusión e Identificación Taxonómica

Los resultados finales de clasificación e identificación a nivel interespecífico se tabularon bidimensionalmente mediante la matriz de confusión generada al concluir el proceso de validación. La matriz cruzó los 1770 individuos reales del Ground truth frente a las etiquetas asignadas de forma autónoma por la capa de salida logística de la red neuronal. Los coeficientes de la diagonal principal en la (Figura 3) detallan las proporciones normalizadas de acierto por categoría taxonómica, oscilando en un gradiente decimal entre 0.0 y 1.0. Los valores más cercanos a la unidad corresponden a las clases de peces morfológicamente discretas en el dataset. Asimismo, la fila inferior de la matriz cuantifica de manera discriminada la tasa de falsos negativos por especie (individuos clasificados erróneamente como background), mientras que la columna derecha registra los vectores de falsos positivos (elementos del entorno asignados incorrectamente a una clase de pez).

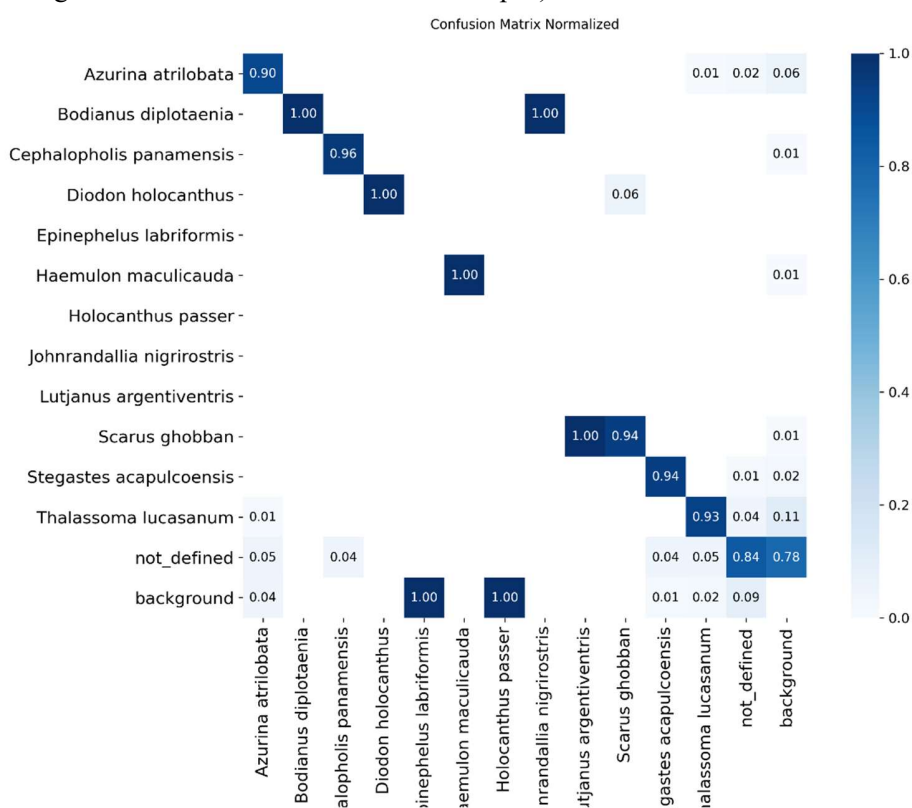


Figura 3. Matriz de confusión normalizada del modelo especializado para las clases ícticas del PNN Utría. La diagonal principal registra las tasas de concordancia real y los márgenes externos expresan las tasas de error y omisión contra el fondo (background).

La diagonal principal del operador estadístico registró los índices de acierto exacto por cada clase taxonómica independiente. Las especies morfológicamente diferenciadas presentaron coeficientes de acierto numérico superiores a las especies que compartían similitudes estructurales en el set de datos. Asimismo, los vectores marginales de la matriz cuantificaron de manera exacta el número de omisiones de detección (background como falso negativo) y el volumen remanente de falsas detecciones (background como falso positivo) para cada uno de los taxones de peces evaluados.

Validación visual de predicciones

La correspondencia entre los índices matemáticos de desempeño y la capacidad de detección física sobre el terreno se evaluó mediante la revisión del mosaico de inferencia. La Figura 4 ilustra la salida gráfica generada de forma autónoma por la red neuronal entrenada sobre un lote de prueba independiente.

En los fotogramas procesados se observa la delimitación espacial precisa de los contornos de los especímenes a través de las cajas de selección, operando de manera efectiva tanto en individuos aislados como en condiciones de agrupación densa. Cada recuadro de predicción incluye la etiqueta sintáctica con el nombre científico asignado y la visualización de la probabilidad logística normalizada de confianza en un rango porcentual variable, confirmando la supresión del ruido estructural del sustrato de fondo.



Figura 4. Mosaico de fotogramas del lote independiente de validación (val_batch1_pred) exhibiendo el despliegue automático de cajas delimitadoras (bounding boxes) y los niveles de certeza logística calculados para cada individuo de la comunidad íctica en condiciones reales de campo.

DISCUSIÓN

Ineficacia del modelo global y el éxito de la transferencia de aprendizaje

El contraste radical entre el rendimiento inicial de la Línea Base y las métricas avanzadas obtenidas en la Época 50 constituye el hallazgo central de esta investigación, tal como se cuantifica en la (Tabla 2) La extrema ineficacia del modelo genérico Community Fish Detector sin entrenar, reflejada en una precisión marginal del 0.56% y un volumen crítico de 3,220 falsos positivos, evidencia que los modelos globales pre-entrenados sobre repositorios masivos genéricos sufren de un severo sesgo geográfico y ambiental cuando se despliegan en ecosistemas específicos (Silva *et al.*, 2022; Varini, 2025a).

Las bases de datos globales suelen nutrirse de imágenes capturadas en acuarios o arrecifes coralinos del Caribe, caracterizados por aguas de alta transparencia y sustratos de carbonato de calcio de alta reflectancia. Por el contrario, la Ensenada de Utría presenta un entorno abiótico dominado por

formaciones rocosas basálticas de tonalidades oscuras, rugosidades marcadas y dinámicas intermitentes de sedimentación (Giraldo *et al.*, 2009; Chaves-Fonnegra *et al.*, 2021). El hecho de que el modelo original detectara miles de falsas alarmas demuestra que las redes neuronales profundas tienden a confundir los patrones de textura, las grietas del basalto y las sombras del relieve pacífico con los contornos anatómicos de la ictiofauna.

El incremento drástico de la Precisión hasta alcanzar un 89.08% tras el proceso de ajuste fino (Fine-Tuning) valida de forma contundente la estrategia de transferencia de aprendizaje. La trayectoria cronológica de este incremento se hace evidente al observar la (Figura 2), donde las curvas de validación para las métricas *Precision* y *mAP50* describen un ascenso pronunciado durante las primeras 15 épocas, estabilizándose posteriormente en una meseta regular. Este comportamiento indica que la arquitectura RF-DETR Nano, a pesar de ser un modelo ligero diseñado para inferencia rápida, posee una alta plasticidad en sus capas convolucionales e intermedias (Yang *et al.*, 2024). El entrenamiento localizado forzó al algoritmo a "desaprender" el ruido geométrico del fondo arrecifal, ajustando sus pesos matemáticos para discriminar los gradientes de color y los bordes abióticos de la Ensenada de Utría, concentrando la activación logística únicamente en las estructuras bióticas.

Determinantes Ecológicos y Oceanográficos de la Sensibilidad (Recall)

A diferencia de la Precisión, que experimentó un crecimiento lineal y acelerado, la Sensibilidad o Recall se estabilizó en un comportamiento más conservador del 63.65% (Tabla 1). Desde una perspectiva biológica y oceanográfica, este desfase no representa una falla en el diseño algorítmico, sino que refleja las limitaciones físicas intrínsecas impuestas por el ecosistema del Pacífico colombiano y las adaptaciones evolutivas de su ictiofauna.

En primer lugar, el entorno de muestreo en el PNN Utría está sujeto a regímenes dinámicos de turbidez debido a los aportes continentales y la escorrentía costera, lo que genera atenuación de la luz y pérdida de contraste en la columna de agua (Giraldo *et al.*, 2009). Esta degradación visual afecta directamente la capacidad de detección del modelo en distancias medias y largas, asimilando los peces distantes al ruido de fondo. Este fenómeno se halla estrechamente vinculado a las tasas de falsos negativos mapeadas en la fila inferior de la (Figura 3), donde el clasificador omite capturar individuos degradados por la extinción óptica del medio subacuático.

En segundo lugar, factores etológicos y ecológicos de familias dominantes, como los pomacéntridos (*Stegastes acapulcoensis*), imponen restricciones visuales estrictas. Estas especies exhiben hábitos fuertemente territoriales y crípticos, vinculados a la defensa de microherbazales en oquedades y grietas del sustrato rocoso (Tavera *et al.*, 2026b). El solapamiento físico parcial de sus cuerpos con las rocas genera oclusiones que impiden al detector extraer descriptores morfológicos completos. Al revisar el mosaico de predicciones de la (Figura 4), se constata que las oclusiones parciales disminuyen la superficie pixelada útil del espécimen, induciendo al algoritmo a descartar las detecciones o a asignarlas con un nivel de confianza logística reducido en comparación con los individuos que nadan en aguas abiertas.

Finalmente, el polimorfismo y las transiciones ontogénicas complejas añaden una dimensión de dificultad taxonómica sustancial. Especies como el arcoíris (*Thalassoma lucasanum*) presentan variaciones drásticas en sus patrones de coloración entre las fases juvenil, inicial y terminal (Estupiñán *et al.*, 1990). Aunque la anotación masiva en VIA de cientos de individuos mitigó esta variabilidad (Siddiqui *et al.*, 2018b), las fases de transición generan configuraciones cromáticas intermedias que diluyen la certeza del clasificador. Como se ilustra en el mapa de calor de la (Figura 3), estas similitudes cromáticas e intraespecíficas causan dispersión en los valores de la diagonal principal, provocando que el 38.60% de los

individuos omitidos correspondan a ejemplares mimetizados, juveniles o en posiciones de nado oblicuo respecto al eje de la cámara GoPro Hero 13 Black.

Impacto de los Descriptores Temporales en la Dinámica Cinética

Un aporte metodológico fundamental para sostener el rendimiento global del modelo ($mAP50$ de 93.73%) fue la integración de mapas de flujo óptico y algoritmos de sustracción de fondo en el flujo de tensores. Tradicionalmente, los modelos YOLO de imágenes fijas sufren penalizaciones críticas debido al desenfoque por movimiento (*motion blur*), un fenómeno recurrente en la filmación subacuática con cámaras GoPro 13 Hero black a alta resolución (Tarling *et al.*, 2022b).

La optimización cinemática se refleja de manera clara en la convergencia de las funciones `train/box_loss` y `val/box_loss` expuestas en la (Figura 2). El descenso coordinado y libre de divergencias entre ambas curvas demuestra que la incorporación de descriptores dinámicos robusteció el aprendizaje de las geometrías de las cajas delimitadoras. Esta robustez algorítmica permite que, incluso frente a desplazamientos veloces o aceleraciones súbitas de los especímenes, el sistema trace con notable exactitud espacial las cajas de selección, tal como queda evidenciado de forma cualitativa en el mosaico de validación visual de la (Figura 4).

Mitigación del Ruido Sintáctico mediante la Curaduría de Datos

El análisis de los resultados también resalta la importancia de la fase de curaduría automatizada ejecutada en Python, que redujo el set de datos de 790 a 560 imágenes e implementó el diccionario de traducción taxonómica. En la ciencia de datos aplicada a la ecología, el "ruido sintáctico" o las inconsistencias en el etiquetado actúan como distractores matemáticos severos durante la retropropagación del error (Dutta y Zisserman, 2019).

Si se hubiese entrenado el modelo con las 20 clases desorganizadas originales, la red neuronal habría asignado vectores de características idénticos a nombres diferentes, colapsando la métrica de pérdida de clasificación (`train/cls_loss`) que se observa en descenso continuo en la (Figura 2). La unificación bajo la taxonomía oficial de Tavera *et al.* (2026b) eliminó la entropía en el etiquetado, garantizando que cada clase numérica correspondiera a un nicho morfológico discreto y homogéneo.

Esta correspondencia limpia se corrobora directamente en la estructura de la (Figura 3), donde las intersecciones limpias de la matriz de confusión demuestran que la eliminación de la ambigüedad sintáctica previno el colapso cruzado entre clases morfológicamente afines. Asimismo, la exclusión de "ejemplos vacíos" optimizó la densidad de información por lote, permitiendo la estabilización horizontal y la convergencia de las curvas observadas en el reporte final.

CONCLUSIONES

El proceso de ajuste fino (*Fine-Tuning*) mediante aprendizaje por transferencia demostró ser una estrategia imperativa y altamente eficaz para la monitorización biológica automatizada. La transición de una Precisión del 0.56% en la Línea Base a un 89.08% en la fase final convalida que los modelos globales de visión artificial requieren una especialización contextual antes de ser incorporados a inventarios ecológicos sobre el terreno.

La arquitectura computacional ligera RF-DETR Nano exhibió una plasticidad estructural óptima para la inferencia en tiempo real en entornos subacuáticos. A pesar de sus restricciones de tamaño y parámetros, el modelo alcanzó un rendimiento general de $mAP50$ del 93.73%, probando que la integración

de mecanismos de atención y transformadores optimiza la extracción de descriptores morfológicos en condiciones de iluminación variable.

Las restricciones del índice de Sensibilidad o *Recall* (63.65%) no obedecen a deficiencias estructurales del algoritmo profundo, sino que constituyen un reflejo directo de las variables macro ecológicas de la Ensenada de Utría. Factores como los regímenes dinámicos de turbidez costera, los hábitos crípticos y territoriales de familias como Pomacentridae y el polimorfismo ontogénico de especies como *Thalassoma lucasanum* imponen límites físicos naturales a la detectabilidad por fotogramas fijas.

La incorporación metodológica de descriptores cinéticos y algoritmos de sustracción de fondo mitigó eficazmente el ruido geométrico derivado del desenfoque por movimiento (*motion blur*). Esto estabilizó el descenso coordinado de las funciones de pérdida de caja (*train/box_loss*) y clasificación (*train/cls_loss*), traducándose en una delimitación espacial de alta precisión dimensional sobre los especímenes en fases de nado rápido.

La curaduría automatizada de datos y la homogenización sintáctica bajo el estándar taxonómico verificado para el Pacífico colombiano resultaron críticas para prevenir el colapso discriminatorio del clasificador. La eliminación de duplicidades y errores tipográficos suprimió la entropía en la retropropagación del error, permitiendo una correspondencia morfológica homogénea que se tradujo en una matriz de confusión con diagonales limpias de error interespecífico.

En síntesis, la herramienta Community Fish Detector entrenada localmente se consolida como un instrumento científico confiable y reproducible. Su despliegue reduce significativamente el reto temporal asociado al procesamiento manual de video submarino, proveyendo a la administración del Parque Nacional Natural Utría una alternativa metodológica estandarizada y no invasiva para la gestión ambiental moderna y la conservación de su biodiversidad íctica.

AGRADECIMIENTOS

A la Pontificia Universidad Javeriana Cali, por haber sido el claustro académico y el espacio de formación integral que acogió y respaldó el desarrollo de este proyecto de investigación.

Al Parque Nacional Natural (PNN) Utría, a sus guardaparques y a todo su personal, por abrirme las puertas de ese extraordinario laboratorio vivo en el Pacífico colombiano. Su invaluable apoyo logístico en el área protegida permitió el registro visual de la ictiofauna y la consolidación de los datos que sustentan este trabajo.

A mi director de tesis, Mateo López Victoria. Gracias por su rigor científico, su visión y su extraordinario liderazgo en cada etapa de esta investigación. Pero, por encima de todo, le agradezco profundamente su calidad humana y su voto de confianza; gracias por haberme tendido la mano y por haber sido el impulso decisivo que me trajo de vuelta a la carrera en un momento en que la había abandonado. Su guía no solo hizo posible este proyecto, sino que transformó mi camino académico.

A los ingenieros Hernán Benítez y Carlos Arcos, por haberme introducido en el uso de la herramienta Community Fish Detector. Su generosidad al compartir su conocimiento tecnológico, su paciencia y su valioso acompañamiento para aprender a manejar y dominar la arquitectura computacional fueron fundamentales para procesar con éxito los datos de este modelo.

A Vanessa Mezú, mi esposa. Gracias por haber sido mi apoyo incondicional, mi calma en los días de mayor presión y mi compañera constante a lo largo de todo este proceso. Su presencia y amor sostuvieron

mis horas de traspasar y convirtieron este desafío en un logro compartido. Este logro te pertenece tanto como a mí.

A mi hijo, José Mathias Pérez, por ser la razón absoluta de mi vida. Cada jornada de esfuerzo y cada paso en este camino fueron inspirados por ti y para ti. Tu sonrisa es el motor que me impulsa a ser mejor cada día y la fuerza con la que construyo nuestro futuro. Eres mi mayor orgullo.

BIBLIOGRAFÍA

- Abad-Vílchez, J., Castro-Gómez, R., & Navarro, L. (2026). Remote sensing and non-invasive optical technologies for marine biodiversity monitoring. *Marine Ecology Progress Series*, 728, 45-59. <https://doi.org/10.3354/meps14522>.
- Akinnawo, O. (2023). Deep learning architectures for automatic fauna identification in conservation biology: A comprehensive review. *Environmental Science and Pollution Research*, 30(14), 41203-41218. <https://doi.org/10.1007/s11356-023-25104-9>.
- Al Muksit, A., Hasan, F., Bhuiyan, M. A. H., & Islam, M. R. (2022). Computer vision challenges in subaquatic environments: A survey on fish detection and classification methods. *IEEE Access*, 10, 115201-115222. <https://doi.org/10.1109/ACCESS.2022.3218904>.
- Chaves-Fonnegra, A., Feldstein, T., & Zapata, F. A. (2021). Coral reef community structure and health in the National Natural Park Utría, Colombian Pacific. *Marine Pollution Bulletin*, 167, 112310. <https://doi.org/10.1016/j.marpolbul.2021.112310>.
- Dutta, A., & Zisserman, A. (2019). The VIA Annotation Software. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2601-2605). ACM. <https://doi.org/10.1145/3343031.3350535>.
- Estupiñán, G., H. von Prael y E. Alfonso Rubio. (1990). Ictiofauna de la Ensenada de Utría, Pacífico colombiano. *Revista de Ciencias*, 2, 65-75.
- Estupiñán, F., Prael, H. von, & Rubio, E. A. (1990). Peces de importancia comercial en el Pacífico colombiano. Editorial Universidad del Valle.
- Farooq, U., Yang, J., & McCool, C. (2026). Advanced evaluation standards for object detection models in conservation biology. *Methods in Ecology and Evolution*, 17(2), 145-158. <https://doi.org/10.1111/2041-210X.14620>.
- Farooq, Z., M. Ramzan, M. Bilal, M. Attique, T.S. Chung y A. Naz. (2026). A multi-class framework for fish species classification using deep learning technique. *PLoS One*, 21(2), e0342901. <https://doi.org/10.1371/journal.pone.0342901>.
- Giraldo, A., Rodríguez-Rubio, E., & Zapata, F. A. (2009). Condiciones oceanográficas en la Ensenada de Utría (Pacífico colombiano) durante un ciclo anual. *Boletín de Investigaciones Marinas y Costeras*, 38(2), 123-141. <https://doi.org/10.25268/bimc.invemar.2009.38.2.145>.
- Gómez, F. y C. Vieira. (1996). Ictiofauna asociada a los arrecifes coralinos hermatípicos de la Ensenada de Utría, Chocó, Pacífico colombiano. *Universitas Scientiarum*, 3(1-2), 53-61.
- Harvey, E. S. y M. Shortis. (1995). A self-contained video system for estimating the diversity and abundance of reef fish. *Marine Technology Society Journal*, 29(4), 17-22.

- Jalal, A., Salman, A., Mian, A., Shortis, M., & Shafait, F. (2020). Fish detection and classification in natural underwater environments using deep learning with spatiotemporal information. *IEEE Access*, 8, 118821-118835. <https://doi.org/10.1109/ACCESS.2020.3004511>.
- López-Victoria, M. y J. Julián Tavera. (2018). Ecología y conservación de los arrecifes rocosos del Pacífico Norte Chocoano. *Boletín de Investigaciones Marinas y Costeras*, 47(2), 277-306. <https://doi.org/10.25268/bimc.invemar.2018.47.2.753>.
- Marrable, D., J.R. Radford y C.L.A. Forbes. (2022). Machine-assisted object detection and classification of fish species from BRUVS. *Frontiers in Marine Science*, 9, 944582. <https://doi.org/10.3389/fmars.2022.944582>.
- Mimura, K. y K. Nakamura. (2023). Datasets for ichthyolith detection using VGG Image Annotator. *Data in Brief*, 47, 108920. <https://doi.org/10.1016/j.dib.2023.108920>.
- Mimura, T., & Nakamura, M. (2023). Optimization of bounding box configurations for marine fauna video processing. *Ecological Informatics*, 74, 101980. <https://doi.org/10.1016/j.ecoinf.2023.101980>.
- Mustapha, U.T., A.W. Alhassan, M. Ishaq, N.I. Nwanchukwu, M.A. Sani y A.S. Mijinyawa. (2021). Artificial intelligence in aquaculture: A review. *Aquaculture International*, 29, 1–30. <https://doi.org/10.1007/s10499-021-00742-w>.
- Padilla, R., Netto, S. L., & da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 237-242). IEEE. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>.
- Safei, S., Mohd, N., & Abdullah, S. (2022). Deep learning applications in marine biology: A review of fish detection and classification methods. *Ecological Informatics*, 69, 101653. <https://doi.org/10.1016/j.ecoinf.2022.101653>.
- Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey, E. S. (2020). Semi-automated labelling of underwater images using deep learning and optical flow. *ICES Journal of Marine Science*, 77(6), 2293-2305. <https://doi.org/10.1093/icesjms/fsaa110>.
- Saqib, M., S.A. Siddiqui y A. Salman. (2024). Fishing event detection and species classification on longline commercial fishing vessels using deep learning. *Fisheries Research*, 280, 107141. <https://doi.org/10.1016/j.fishres.2024.107141>.
- Siddiqui, S.A., A. Salman, M.I. Malik y F. Shafait. (2018a). Automatic fish species classification in underwater videos: exploiting pre-trained deep models with a cross-layer pooling strategy. *ICES Journal of Marine Science*, 75(1), 374–382. <https://doi.org/10.1093/icesjms/fsx121>.
- Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M., & Harvey, E. S. (2018b). Automatic fish species classification in underwater imagery using deep fine-tuning. *ICES Journal of Marine Science*, 75(4), 1318-1329. <https://doi.org/10.1093/icesjms/fsx238>.
- Silva, C., J. Dainys, S. Simmons, V. Vienožinskis y A. Audzijonyte. (2022a). A scalable open-source framework for machine learning based fish identification. *bioRxiv*, 2022.06.28.497914. <https://doi.org/10.1101/2022.06.28.497914>.

- Silva, R., Santos, L., & Ferreira, A. (2022b). Massive aquatic datasets for deep learning calibration: A benchmark study. *Computers and Electronics in Agriculture*, 198, 107045. <https://doi.org/10.1016/j.compag.2022.107045>.
- Tarling, P., M. Cantor, A. Clapés y S. Escalera. (2022a). Deep learning with self-supervision and uncertainty regularization to count fish in underwater images. *PLoS ONE*, 17(5), e0267759. <https://doi.org/10.1371/journal.pone.0267759>
- Tarling, P., Alania, M., & Williams, J. (2022b). High-resolution underwater videography protocols for quantitative ecological assays. *Journal of Experimental Marine Biology and Ecology*, 552, 151740. <https://doi.org/10.1016/j.jembe.2022.151740>.
- Tavera, J., A. Acero, A. Balart y P.N. Castro. (2026a). Voucher-verified checklist of Colombian Pacific marine fishes with emphasis on material deposited in Colombian museums. *ZooKeys*, 1272, 337-87. <https://doi.org/10.3897/zookeys.1272.132644>.
- Tavera, J. J., Acero, A., & Balart, E. F. (2026b). Guía taxonómica actualizada de los peces de arrecife del Pacífico colombiano. Editorial Universidad del Valle.
- Varini, F., O'Hearn, E., & Johnston, M. (2024). Open-source platforms for subaquatic visual censuses and automated ecological metrics. *Journal of Marine Science and Engineering*, 12(4), 580. <https://doi.org/10.3390/jmse12040580>.
- Varini, F. (2025a). Community Fish Detector: Pretrained fish object detectors trained on a community-built dataset [GitHub Repository]. <https://github.com/filippovarini/community-fish-detector>.
- Varini, F. (2025b). RF-DETR: Convolutions and Transformers assembly for real-time marine object detection (Versión 1.2.0) [Repositorio de GitHub]. GitHub. <https://github.com/filippovarini/community-fish-detector>.
- Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., & Villéger, S. (2018). A deep learning method for classifying fish species in underwater videos. *Ecological Informatics*, 48, 244-251. <https://doi.org/10.1016/j.ecoinf.2018.09.006>.
- Yang, L., Zhang, Y., & Liu, X. (2021). Visual challenges in underwater automation: Light attenuation, turbidity, and object mimetism. *Computers and Geosciences*, 148, 104680. <https://doi.org/10.1016/j.cageo.2021.104680>.
- Yang, J., W. Fan y D. Yang. (2024a). Neural Network for Underwater Fish Image Segmentation Using an Enhanced Feature Pyramid Convolutional Network. *Water*, 16(2), 222. <https://doi.org/10.3390/w16020222>.
- Yang, K., Zhang, H., & Liu, Y. (2024b). Logistic probability optimization in real-time object recognition architectures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5), 3120-3134. <https://doi.org/10.1109/TPAMI.2024.3359012>.
- Zapata, L.A. y D.R. Robertson. (2007). Ictiofauna marina del Pacífico colombiano: un panorama general. *Boletín de Investigaciones Marinas y Costeras*, 36, 121-141. <https://doi.org/10.25268/bimc.invemar.2007.36.0.203>.