



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Desarrollo de una plataforma de realidad aumentada de bajo coste
para la investigación en aplicaciones de visión artificial*

Grado en Ingeniería Mecánica

ALUMNO: D. Julio Egido Oliver

DIRECTORES: D. José Antonio González Prieto
D. José Ignacio Valles Cancela

CURSO ACADÉMICO: 2023-2024

Universida_{de}Vigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Desarrollo de una plataforma de realidad aumentada de bajo coste
para la investigación en aplicaciones de visión artificial*

Grado en Ingeniería Mecánica
Intensificación en Tecnología Naval
Cuerpo General

UniversidadeVigo

RESUMEN

La Realidad Aumentada (RA) es una tecnología en pleno auge, que se caracteriza por ser capaz de agregar información virtual a la realidad física.

En la actualidad, la seguridad portuaria enfrenta desafíos debido al incremento del tráfico marítimo. La integración de estas nuevas herramientas, junto con *Deep Learning* (DL), podrían mejorar la vigilancia y seguridad marítima. Se justifica, la elección del algoritmo YOLO para el presente trabajo, por su precisión y reducido tamaño, que favorecen respuestas rápidas en tiempo real.

En cuanto a la metodología, se contempla la creación de un *dataset*, un entrenamiento de modelos de visión artificial y validación de resultados para detección y seguimiento de embarcaciones. Además, cabe destacar cómo la integración de sistemas de seguimiento visual con datos AIS, enriquece la vigilancia marítima, permitiendo una interfaz más completa y precisa para la vigilancia virtual.

La combinación de RA y DL permite ofrecer una solución al problema planteado, permitiendo mejorar la eficiencia y precisión de un sistema de vigilancia portuaria.

PALABRAS CLAVE

Realidad Aumentada, visión artificial, *Deep Learning*, Tráfico marítimo, Sistema de Identificación Automática (AIS).

AGRADECIMIENTOS

Quiero expresar mi más sinceros agradecimientos, a todas las personas e instituciones que han puesto de su esfuerzo, tiempo y ganas para que la realización de este trabajo haya sido posible.

En primer lugar, agradezco a mis directores, D. José Antonio González, cuya inagotable paciencia, dedicado trabajo, guía constante y fuente de motivación, que junto con el CC D. José Ignacio Valles, que desde que era aspirante de primero me enseñó lo bonita que es esta vida de entrega a España y a la mar, han sido fundamentales en cada etapa de este proceso. Agradezco al Centro Universitario de la Defensa por enseñarme el camino de la ingeniería y dotarme de conocimientos profesionales de gran valor para mi paso por las unidades y a la Escuela Naval Militar y a mis comandantes de brigada, por brindarme la oportunidad de crecer como militar y sobre todo como persona, gracias a su valores y ejemplaridad, los cuales me sirven como timón en mi día a día y me ayudan a forjar el carácter del oficial de la Armada que me gustaría ser, ya que como siempre se ha dicho ‘*Mandar es servir*’.

A mi promoción 424-154, infinitamente agradecido por compartir este viaje lleno de desafíos, momentos y aprendizajes, de los que cada vez que mire atrás, siempre me dibujarán una enorme sonrisa contagiada por deseo de volver a esta etapa de mi vida.

A mi Galicia querida del alma, que, aunque mis raíces gaditanas de la Isla de Camarón están siempre presentes, le estoy enormemente agradecido por enseñarme tanto de su cultura, su tierra, su inmejorable gastronomía, por sus maravillosas gentes y que ‘*Se chove, ¡que chova!*’.

A mis padres Jose Julio y M^a del Mar (que se acuerdan del día del árbol), mis hermanas Marta y Paula, abuelos, primos, tíos, Carmen, amigos de la infancia, Lana y Trufi, les agradezco por su constante apoyo y amor incondicional desde el día que me ayudaron a entrar en esta familia que es la Armada, hasta el día 16 de julio de 2024, que estarán, todos celebrando el Carmen en la soleada villa de Marín.

Finalmente, quiero expresar mi profundo agradecimiento a la Armada Española por forjar mi futuro y mi camino, así como por influir en mi forma de pensar y afrontar los retos en entornos cambiantes y desafiantes.

Sin el respaldo y la inspiración de todos ustedes (que siempre estarán en mi memoria), este logro no habría sido posible. Es por ello por lo que, de aquí en adelante, todo lo que logre en el camino que me prosigue, también serán sus logros.

CONTENIDO

Contenido	1
Índice de Figuras	5
Índice de Tablas	9
1 Introducción y objetivos.....	11
1.1 Contexto y motivación.....	11
1.2 Objetivos y alcance.....	12
1.3 Estructura de la memoria.	12
2 Estado del arte	15
2.1 Descripción del apartado	15
2.2 Realidad Aumentada.....	15
2.2.1 Definición.....	15
2.2.2 Contexto histórico	16
2.2.3 Realidad Aumentada en la actualidad y sus tipos	18
2.2.4 Realidad Aumentada en el ámbito marino	18
2.2.5 Realidad Aumentada en la Armada Española.	20
2.3 Sistemas de aprendizaje automático y la visión artificial.	22
2.3.1 Inteligencia Artificial	23
2.3.2 Machine Learning.....	24
2.3.3 Deep Learning.	27
2.3.4 Visión artificial.....	28
2.3.5 Redes Neuronales Artificiales	30
2.3.6 El algoritmo de detección de objetos en imágenes (YOLO).....	33
2.4 Control y monitorización de tráfico marítimo	35
2.4.1 Datos AIS	35
2.4.2 ¿Qué buques pueden transmitir Datos AIS?.....	36
2.5 Datasets y tipos.....	36
2.5.1 China Maritime Datasets.	38
2.6 Aplicaciones para la Armada en el siglo XXI.	40
2.6.1 Estado actual de estas técnicas en la Armada.....	40
2.6.2 Deep Learning en la Armada.....	41
3 Desarrollo del TFG	43
3.1 Introducción.....	43
3.2 Aprendizaje y formación en la materia.....	43
3.3 Herramientas de Hardware	45

3.3.1 Cámara ELP 8 megapíxeles	45
3.3.2 Teléfono móvil	45
3.3.3 Receptor AIS inalámbrico QK-A024	45
3.4 Herramientas de software y entornos de trabajo.....	46
3.4.1 Sistema Operativo LINUX.....	46
3.4.2 Label-Studio	46
3.4.3 Ultralytics HUB.....	46
3.4.4 VS Codium	47
3.4.5 Entorno de trabajo VIDEO-IA-CUD	47
3.4.6 Google Colaboratory	48
3.5 Descripción de tareas.....	48
3.6 Justificación de Algoritmos	48
3.6.1 Algoritmo para la detección de objetos.....	49
3.6.2 Algoritmo para el seguimiento de objetos.....	49
3.7 Primera tarea: Conocer los mensajes AIS y sus tipos.....	49
3.7.1 Casos Particulares.....	53
3.8 Segunda tarea: grabación de vídeo y entrenamiento modelo inicial	54
3.8.1 Elección de la ubicación de grabación	54
3.8.2 Etiquetado de vídeos	55
3.8.3 Entrenamiento dataset MARIN_PORT en VS CODIUM.....	56
3.8.4 Obtención de vídeos con seguimiento de objetos	60
3.8.5 Entrenamiento dataset MARIN_PORT en Ultralytics HUB.....	60
3.9 Tercera tarea: integración de datos AIS y vídeo.....	64
3.9.1 Elección de ubicación de grabación.	65
3.9.2 Entrenamiento del modelo.....	65
3.9.3 Información temporal	66
3.9.4 Visualización de los puntos GPS en Google Maps	68
3.10 Cuarta tarea: elaboración interfaz Realidad Aumentada	70
3.10.1 Justificar los símbolos	71
3.10.2 Disposición de la ventana.....	71
4 Resultados / Validación / Prueba	73
4.1 Introducción.....	73
4.2 Resultados de la prueba realizada con el <i>dataset</i> de Escalera Monumental en <i>Ultralytics HUB</i> ..	73
4.2.1 Fallos comunes encontrados.....	74
4.2.2 Verificación de datos AIS en Google Maps.....	76
4.3 Validación del Interfaz Virtual de RA	78

4.4 Posibles mejoras de los resultados.....	79
5 Conclusiones y líneas futuras.....	81
5.1 Introducción.....	81
5.2 Conclusiones extraídas	81
5.3 Objetivos alcanzados	82
5.4 Tendencias futuras	83
6 Bibliografía	85
Anexo I: Implicaciones Sociales, y/o Económicas, y/o Ambientales.....	91
Anexo II: Reflexiones Éticas y Sociales	93
Anexo III: Glosario de siglas y acrónimos.....	95

ÍNDICE DE FIGURAS

Figura 1-1: Esquema resumen de la memoria del TFG [autoría propia]	13
Figura 2-1 Entorno de Realidad Extendida (XR) [5]	15
Figura 2-2 La espada de Damocles [6]	17
Figura 2-3 <i>Google Glass</i> [7]	17
Figura 2-4 Usuario trabajando con las <i>Apple Vision Pro</i> [10]	18
Figura 2-5 Usuario en puente llevando un dispositivo de visualización montado en la cabeza (HMD) [13]	19
Figura 2-6 Icebergs mostrados en interfaz de Realidad Aumentada [13]	20
Figura 2-7 Motor Diesel del buque de la Armada Española “Intermares” [autoría propia]	21
Figura 2-8 Dotación de un barco realizando un ejercicio de seguridad interior [14]	21
Figura 2-9 Almacén Parque y Centro de Abastecimiento de Mareal de Intendencia del Ejército de Tierra [15]	22
Figura 2-10 Diagrama que engloba la Inteligencia Artificial, <i>Machine Learning</i> , y <i>Deep Learning</i> [16]	23
Figura 2-11 Ajedrez visto en GNOME [19]	23
Figura 2-12 El equilibrio del aprendizaje [23]	25
Figura 2-13 Esquema Aprendizaje Supervisado [25]	25
Figura 2-14 Esquema Aprendizaje no Supervisado [25]	26
Figura 2-15 Segmentación del terreno con aprendizaje no supervisado [27]	26
Figura 2-16 Esquema Aprendizaje por refuerzo [29]	27
Figura 2-17 Cámara vista desde el vehículo, tomando medidas para ser capaz de ubicarse [31]	27
Figura 2-18 Diferencia entre <i>Machine Learning</i> y <i>Deep Learning</i> [32]	28
Figura 2-19 Diferencia entre visión humana y visión artificial [33]	28
Figura 2-20 Diferentes tareas a resolver en visión artificial [35]	29
Figura 2-21 Diferencia entre redes neuronales simples y profundas [37]	30
Figura 2-22 Redes Neuronales Monocapa y Redes Neuronales Multicapa [38]	30
Figura 2-23: Redes Neuronales Recurrentes [38]	31
Figura 2-24 Convolución CNN [40]	32
Figura 2-25 Línea temporal redes neuronales convolucionales [41]	32
Figura 2-26 Detección de objetos con YOLO [42]	34
Figura 2-27 Un equipo dotado de AIS muestra la distancia y el rumbo de los barcos cercanos en un formato similar al radar [44]	35
Figura 2-28 Velero de tres mástiles <i>Eendracht</i> con su respectivo IMO (8814275) siendo Róterdam su puerto de registro [45]	36
Figura 2-29 YOLOv3 aplicado a la vigilancia marítima en el río <i>Yangtse</i> , China [50]	37

Figura 2-30 <i>Dataset</i> en el ámbito submarino [51]	38
Figura 2-31 Diagrama de flujo del experimento [50]	39
Figura 2-32 Esquema del proceso de aprendizaje del modelo [50]	39
Figura 2-33 Funcionamiento de la Industria 4.0 [53]	40
Figura 2-34 Soldado Ucraniano adiestrándose para el uso de drones [55]	41
Figura 3-1 Fase de aprendizaje [autoría propia]	44
Figura 3-2 Cámara ELP 8 MP [56]	45
Figura 3-3 Receptor AIS inalámbrico QK-A024 [58]	45
Figura 3-4 Interfaz de etiquetado de tareas en <i>Label-Studio</i> [imagen importada de <i>Label Studio</i> de autoría propia]	46
Figura 3-5 Modelos con los que permite entrenar la herramienta software <i>Ultralytics HUB</i> [imagen importada de <i>Ultralytics</i>]	47
Figura 3-6 Ejemplos de tipos de transpondedores y receptores [65]	50
Figura 3-7: Lectura de datos AIS a través del <i>SCRIPT ais_capture_tool</i> [autoría propia]	52
Figura 3-8 Primer prototipo digital de interfaz de RA con datos relevantes [autoría propia]	53
Figura 3-9 Captura satélite de <i>Google Maps</i> para la generación del <i>dataset</i> [autoría propia]	54
Figura 3-10 A la izquierda vista desde la ventana la Torre del Patio de Aulas, a la derecha con un círculo rojo la ventana [autoría propia]	55
Figura 3-11: Etiquetado primer modelo Escuela Naval Militar, donde se pueden apreciar <i>military_boat</i> , <i>sailing_boat</i> , <i>teaching_boat</i> y <i>row_boat</i> [imagen extraída de <i>Label-Studio</i> de autoría propia]	56
Figura 3-12 Fragmento del script empleado para el entrenamiento del modelo [autoría propia]	57
Figura 3-13 Matriz de confusión generada durante el entrenamiento del modelo <i>MARIN_PORT</i> [autoría propia]	57
Figura 3-14 Gráfica que muestra <i>precision</i> para cada categoría de objeto detectado [autoría propia]	58
Figura 3-15 Gráfica que muestra <i>recall</i> para cada categoría de objeto detectado [autoría propia]	58
Figura 3-16 Gráfica que muestra la relación entre <i>recall</i> y <i>precision</i> [autoría propia]	59
Figura 3-17 Imágenes generadas por el modelo YOLOv8 que ha sido entrenado [autoría propia]	59
Figura 3-18 Captura de un <i>frame</i> de los vídeos obtenidos con <i>tracking</i> [autoría propia]	60
Figura 3-19 Distribución de las imágenes que componen el <i>dataset</i> "MARIN_PORT" [gráfica de Ultralytics HUB de autoría propia]	61
Figura 3-20 Representación visual proporcionada por Ultralytics HUB que muestra los resultados del mAP, recall y precision para cada uno de los epochs analizados [gráfica generada por Ultralytics HUB de autoría propia]	62
Figura 3-21 Gráfica del <i>box loss</i> para cada entrada evaluada [gráfica generada por <i>Ultralytics HUB</i> de autoría propia]	63
Figura 3-22 Gráfica del class loss para cada entrada evaluada [gráfica generada por Ultralytics HUB de autoría propia]	63
Figura 3-23 Gráfica del <i>object loss</i> para cada entrada evaluada [gráfica generada por <i>Ultralytics HUB</i>]	64

Figura 3-24 Imagen sometida a <i>test</i> tras entrenamiento del modelo [imagen generada por <i>Ultralytics HUB</i> de autoría propia]	64
Figura 3-25 A la izquierda vista desde la Escalera Monumental, a la derecha con un círculo rojo la escalera [autoría propia]	65
Figura 3-26 Representación visual proporcionada por <i>Ultralytics HUB</i> que muestra los resultados del <i>mAP</i> , <i>recall</i> y <i>precision</i> para cada uno de los <i>epochs</i> analizados [gráfica generada por <i>Ultralytics HUB</i> de autoría propia]	65
Figura 3-27 Captura de un <i>frame</i> del vídeo capturado, a la izquierda el número correspondiente y a la derecha la fecha y hora [autoría propia]	66
Figura 3-28 Cámara izquierda durante el <i>track</i> con ambas cajas en cada embarcación [autoría propia]	67
Figura 3-29 Captura del <i>video_track_3d_tool</i> en su función distancia entre embarcaciones [autoría propia]	68
Figura 3-30 Captura del <i>video_track_3d_tool</i> en su función de interfaz virtual [autoría propia]	68
Figura 3-31 Imagen extraída de <i>Google Maps</i> donde se ven los puntos GPS base utilizados para crear el sistema de medición de distancias [Alfárez de fragata Don Víctor Recio]	69
Figura 3-32 Captura del <i>frame</i> número 350 en donde se puede apreciar como el dato AIS coincide en la posición de la lancha de instrucción [autoría propia]	69
Figura 3-33 Captura de las embarcaciones en el <i>frame</i> 350 en un plano 2 dimensiones [Plano de <i>Google My Maps</i> de autoría propia]	70
Figura 3-34 Distancia entre la lancha de instrucción y la coordenada GPS que devuelve su transpondedor AIS [Plano de <i>Google My Maps</i> de autoría]	70
Figura 3-35 Datos que ofrece la interfaz virtual para vigilancia portuaria [autoría propia]	72
Figura 4-1 Representación visual proporcionada por <i>Ultralytics HUB</i> que muestra los resultados del <i>mAP</i> , <i>recall</i> y <i>precision</i> para cada uno de los <i>epochs</i> analizados [gráfica generada por <i>Ultralytics HUB</i> de autoría propia]	74
Figura 4-2 Captura del script <i>video_track_tool</i> [autoría propia]	74
Figura 4-3 Lancha de instrucción que el sistema no detecta correctamente [autoría propia]	75
Figura 4-4 Pérdida de embarcaciones cuando se produce una situación de cruce [autoría propia]	75
Figura 4-5 Diferencia de luminosidad debido a las condiciones meteorológicas [autoría propia]	76
Figura 4-6 Captura del <i>video_track_3d_tool</i> en su función distancia entre embarcaciones [autoría propia]	76
Figura 4-7 Vista en 2 dimensiones de la trayectoria de las embarcaciones en <i>Google Maps</i> [autoría propia]	77
Figura 4-8 Diferencia de vista en 2 dimensiones en <i>Google Maps</i> respecto de la interfaz virtual [autoría propia]	77
Figura 4-9 Captura del <i>video_track_3d_tool</i> en su función de interfaz virtual [autoría propia]	78

ÍNDICE DE TABLAS

Tabla 1 Procesos de la visión artificial [34].....	29
Tabla 2 Tasas de actualización en función de la velocidad sobre fondo [65].....	51
Tabla 3: Tipos de datos en función de clase de transpondedor A o B/B+ [65]	52
Tabla 4: Tipos de Mensajes AIS más relevantes para el trabajo [67]	53
Tabla 5 Justificación de los símbolos escogidos para la interfaz virtual [autoría propia]	71

1 INTRODUCCIÓN Y OBJETIVOS

1.1 Contexto y motivación

Numerosas tecnologías emplean la Inteligencia Artificial (IA) para realizar diferentes tareas presentes en nuestra vida cotidiana, desde algoritmos para reconocimiento de voz hasta pilotos automáticos en automóviles. Este panorama se amplía aún más cuando consideramos técnicas como el *Machine Learning* (ML) y el DL. La aplicación de estas técnicas en diversos campos ha crecido exponencialmente en los últimos años, siendo el uso de la visión artificial en la creación de entornos de RA uno de los más activos.

Más concretamente, en el ámbito marítimo y dentro de la Armada Española, estas técnicas son de especial interés para el desarrollo de nuevas tecnologías aplicables a bordo de las unidades, como menciona el AJEMA en las líneas generales de la Armada 2022. Este documento se emite cada 5 años y aborda los intereses generales que la Armada enfrentará en los próximos años, tanto en aspectos tecnológicos como otros de especial interés [1].

Esta transformación digital busca una Armada más ágil simple y flexible. Como además reflejó el actual Subdirector General de Enseñanza Naval, nos preparamos para alcanzar los objetivos de la reciente bautizada Armada 4.0¹, entrenándonos para un nuevo entorno multidominio que favorezca la implementación de IA. Un entorno que habilite a las dotaciones para gestionar eficazmente la fusión de datos, mejorando así la conciencia general ante cualquier situación y facilitando la toma de decisiones [2]. Además, los datos correctamente estructurados presentan una gran ventaja competitiva, siempre y cuando sean interpretados por personal cualificado.

En la misma línea, la vigilancia marítima junto con la monitorización del tráfico marítimo, tanto en España como a nivel mundial, están adquiriendo una importancia crucial en los últimos años. Más concretamente, las cámaras de vigilancia portuarias, como ya ocurre en puertos como Barcelona, Valencia o Algeciras, están comenzando a implementar sistemas integrados con IA. Esto demuestra su capacidad para mejorar la seguridad, la eficiencia y la protección del medio ambiente [3]. Asimismo, al combinar la información que devuelven los transpondedores AIS de las embarcaciones, se logra una verificación más confiable de la información sobre una embarcación que se encuentre en las inmediaciones de un puerto, gracias a posibles interfaces virtuales que facilitarían la inspección de los entornos a vigilar.

Por lo tanto, este estudio se centra en el *edge computing*, una línea de investigación centrada en el desarrollo de sistemas de IA distribuidos, que cuentan con diversos sensores, como cámaras o receptores

¹ La Armada 4.0 en la Armada Española implica la modernización y adaptación tecnológica para fortalecer su capacidad operativa mediante la integración de tecnologías avanzadas.

AIS en nuestro caso, que procesan datos de manera inteligente para reducir la cantidad de información a transmitir. Esto permite realizar la integración de la información sin necesidad de enviar todos los datos de los sensores. Además, evita la centralización y la necesidad de disponer de grandes centros de datos, permitiendo la creación de sistemas capaces de aprender y ejecutar tareas repetitivas de manera altamente automatizada. Además, al procesar más cerca de los sensores, se logra un ahorro significativo y se evitan problemas asociados al transporte de información., Por otra parte, facilita el futuro escalado del sistema, así como la integración de datos de nuevos tipos de sensores (radar, lidar, etc).

1.2 Objetivos y alcance

En este trabajo, se propone como objetivo crear una interfaz virtual que aplique técnicas de DL y ML a la detección y seguimiento de objetos en un entorno portuario, utilizando medios de grabación simples, así como un receptor AIS. El propósito final del sistema será detectar, clasificar y hacer el seguimiento de una serie de embarcaciones, estableciendo su posición (sin necesidad de recibir datos AIS) y realizando estimaciones sobre su cinemática, es decir, sus velocidades y rumbos. Por otra parte, los datos AIS obtenidos de una embarcación, se emplean para poder validar las posiciones en la interfaz de RA a desarrollar, siempre y cuando la embarcación esté emitiendo dichos datos.

Para lograr este objetivo, se propone utilizar el algoritmo YOLO, específicamente para resolver las tareas relacionadas con detección, clasificación y seguimiento de embarcaciones en vídeos, ya que se considera el más adecuado para el uso de técnicas de aprendizaje supervisado. Para implementar estas soluciones de visión artificial se llevarán cabo los entrenamientos de los modelos de DL tanto a nivel local como empleando la herramienta *online Ultralytics-HUB*., Esta herramienta permitirá analizar los parámetros que devueltos por los entrenamientos para evaluar si el modelo es capaz de cumplir con las tareas demandadas. *Por otra parte*, para el etiquetado de imágenes para el entrenamiento de las redes neuronales convolucionales (CNN) en las que se basa el algoritmo YOLO, se empleará la herramienta *Label-Studio*

Para que alcance del objetivo planteado pueda ser ajustado al desarrollo de este TFG, se plantea el uso de los datos AIS para validar los resultados del posicionamiento de las embarcaciones, para lo cual es necesario resolver una serie de problemas que se plantean, ya que se trata de integrar los datos obtenidos de dos cámaras de vídeo y un receptor AIS cuyos flujos de información deben ser sincronizados. En resumen, el principal desafío consiste en correlacionar los datos AIS recibidos con las embarcaciones que se encuentren en seguimiento en el puerto, lo cual requiere sincronizar los datos en el momento exacto de su captura y establecer relaciones coherentes entre ellos.

1.3 Estructura de la memoria.

Una vez contextualizados, y presentados los objetivos del presente trabajo, a continuación, se detalla la estructura que sigue la memoria, como se aprecia en la Figura 1-1, incluyendo los avances realizados en cada etapa del proyecto, los resultados alcanzados y las conclusiones extraídas. De este modo, la estructura del trabajo se divide en los siguientes 5 capítulos:

- **Introducción y objetivos:** En este tema se expone la importancia de la RA en el mundo actual, así como las posibilidades que ofrece en combinación con técnicas de *DL* y *ML* para su implementación en el entorno marítimo, especialmente en la vigilancia y seguridad portuaria, además de la integración con datos AIS. Por otra parte, se detallan los objetivos y el alcance del trabajo y finalmente, se describe la organización de la memoria.
- **Estado del arte:** Para este segundo capítulo se analiza el contexto histórico y teórico de la RA y la visión artificial, destacando además la importancia que ambas tienen en el mundo actual y en la vida cotidiana, así como sus posibilidades de explotación. Además, se aborda el uso del DL para la resolución de tareas y cómo trasladarlo a la Armada Española. Los contenidos que se revisan en él son los siguientes:

- Introducción e historia de la RA, su empleo en la actualidad, ámbito marino y la Armada Española.
- Visión artificial (visión por computador), incluye el escalón de IA, ML y DL. También se explican qué son las redes neuronales haciendo especial énfasis en la convolucionales, y por último el algoritmo de detección de objetos YOLO.
- Control y monitorización del tráfico marino, pasando por los famosos datos AIS que emiten las embarcaciones y estaciones en forma de señal de radio VHF, junto con la existencia de los *Marine datasets* y *ABOships onboard and onshore*.
- Aplicaciones para la Armada del siglo XXI, desde las que ya existen en la actualidad y sus posibilidades en distintos campos y entornos.
- **Desarrollo del TFG:** Este capítulo, antes de comenzar con las distintas etapas y pasos para elaborar la interfaz, cuenta con un proceso previo de aprendizaje al trabajo en los campos de investigación abordados, y una descripción de todos los elementos de *hardware* y *software* que se han utilizado, además de los entornos de trabajo y algoritmos. Posteriormente, se dan a conocer los problemas que se presentan a lo largo del desarrollo del TFG, desde la detección y seguimiento de objetos en imágenes, hasta la elaboración de una interfaz que sincroniza a su vez datos AIS con vídeos.
- **Resultados obtenidos y validación:** Se exponen los resultados que se han obtenido tras el desarrollo del trabajo, indicando que objetivo se han logrado alcanzar y a qué porcentaje de cumplimiento de los problemas planteados se corresponde. Además, se indican cuáles han sido los objetivos no alcanzados y por qué no han podido lograrse o cuales serían los posibles motivos.
- **Conclusiones y líneas futuras:** Por último, tras la validación de los resultados obtenidos se plantean las conclusiones extraídas, para dar lugar a una serie de líneas furias en las que podría desembocar el desarrollo complementario del proyecto. También se tiene en cuenta como se podrían optimizar los resultados obtenidos.

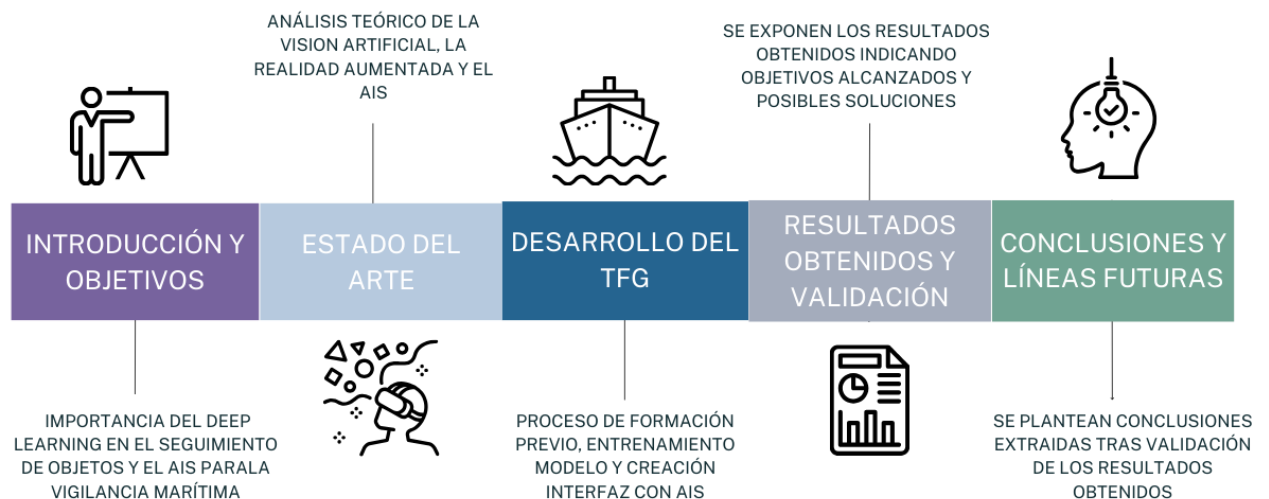


Figura 1-1: Esquema resumen de la memoria del TFG [autoría propia]

2 ESTADO DEL ARTE

2.1 Descripción del apartado

Este capítulo contextualiza los contenidos desarrollados durante este trabajo, dando a conocer las definiciones de las principales materias a estudiar, tales como son la RA, la visión artificial, y el DL. Adicionalmente, se añade un breve contexto histórico de dichas materias, así como una descripción de cuáles son sus aplicaciones en la actualidad. A continuación, se introduce al algoritmo principal que ha sido empleado durante el desarrollo del TFG, conocido como YOLO, encuadrándose en el proyecto, como uno de los pilares fundamentales ya que resuelve la tarea de la localización, clasificación y seguimiento de objetos en vídeos. Finalmente, se trata la aplicación de la información de datos AIS, en el entorno marítimo que corresponde a la Armada Española, gracias a los instrumentos estudiados y cuáles son sus posibles líneas de desarrollo.

2.2 Realidad Aumentada

2.2.1 Definición

Desde una perspectiva general, y ante el desafío de abordar el presente trabajo, se deberá responder la siguiente pregunta ¿Cómo se entiende la RA? Es importante destacar, desde un principio, que no debe confundirse con la realidad virtual (RV), la cual reproduce contenidos mediante dispositivos como gafas y trasladan al usuario a una realidad completamente digital, ni tampoco con la Realidad Mixta (MR), que combina elementos del mundo real con elementos virtuales.

Por su propia naturaleza, la RA muestra a los usuarios, un entorno digital, proyectado y superpuesto sobre una vista real del espacio de trabajo [4]. Es decir, se trata de la integración de información digital sobre el escenario real que el usuario visualiza en tiempo real. La mezcla de los tres escenarios anteriormente mencionados da lugar a la realidad extendida (XR) [5], como se aprecia en la Figura 2-1.

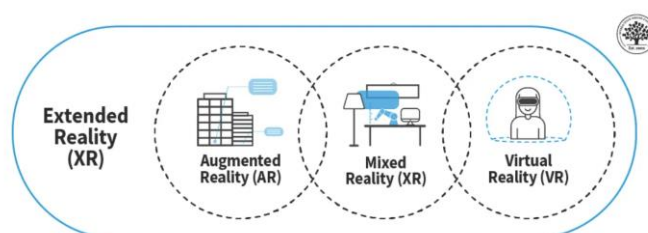


Figura 2-1 Entorno de Realidad Extendida (XR) [5]

La RA tiene ciertas aplicaciones; desde proporcionar información adicional, combinando componentes tridimensionales (en 3 dimensiones), con los del entorno real; hasta ser capaces de ofrecer entornos completamente diferentes de cómo son en realidad.

Son numerosos los formatos en los que pueden presentarse, como el uso cotidiano en *smartphones* y tabletas, o el uso de gafas que ya están siendo desarrolladas por algunas empresas como *Apple*. Sus principales elementos para su desarrollo combinan *hardware* de diferentes tipos, como cámaras, acelerómetros, dispositivos de entrada o pantallas. Adicionalmente, muchos de estos sistemas cuentan con GPS (sistemas de posicionamiento global), y brújulas digitales, que combinados ofrecen al usuario la posibilidad de determinar de manera precisa, su ubicación para una mejor experiencia en el entorno de RA [5].

Por otra parte, la RA permite contar con un programa de alto rendimiento especializado en el campo tridimensional, que permite a los desarrolladores conectar información digital a los marcos espaciales correspondientes de la realidad de los escenarios de trabajo en la que se encuentra el usuario o, en otros casos, que el usuario se encuentra monitorizando. En otras palabras, cuando la aplicación recibe datos digitales de un marcador conocido, ejecuta su código y superpone las imágenes adecuadas a los elementos de la realidad correspondientes.

El proceso implica tres pasos:

1. **Detección y seguimiento:** Se trata de, a través del dispositivo en cuestión, detectar el entorno que le rodea, mediante herramientas como cámaras, acelerómetros, giróscopos, GPS e incluso láseres para rastrear la posición y orientación de usuario y su dispositivo.
2. **Procesamiento y reconocimiento de imágenes:** Una vez que se han recibido todos los datos de los diferentes sensores el sistema los analiza, y utilizando algoritmos de procesamiento y reconocimiento de imágenes y de fusión de datos con sensores de otro tipo, identifica los elementos a tiempo real.
3. **Renderizado y visualización:** Para finalizar, se genera y muestra el contenido generado por ordenador sobre el entorno del mundo real. Esto, lo hace en la perspectiva y posición correctas en relación con el punto de vista del usuario. Es lo que comúnmente se conoce como un holograma.

Si volvemos al primer punto planteado, estas son las diferencias principales entre RA y Realidad Virtual.

- La principal y más notable, es la naturaleza del entorno, siendo totalmente simulado para la RV, y una mezcla de entorno existente e información superpuesta en forma de capas a tiempo real, para la RA.
- Los dispositivos utilizados por la RA son menos restrictivos, ya que se tratan de gafas, proyecciones o HUD (*head-up display*) en automóviles, mientras que la RV utiliza visores que se colocan sobre la cabeza del usuario que no les permiten ver nada real, además de contar con auriculares que aumentan la inmersión del usuario.

2.2.2 Contexto histórico

Para definir los orígenes de la RA, tenemos retroceder a través de 50 décadas de innovación tecnológica, para dar paso a esta potente y prometedora tecnología pionera, que ha cambiado por completo el modo en el que consumimos contenido del mundo real [6].

Remontándonos a la década de los años sesenta, la RA instaura sus raíces en el año 1968, cuando *Ivan Sutherland*, un destacado científico informático, desarrolla el primer sistema de visualización de RA con montura para la cabeza en la universidad de Harvard. Este fue bautizado con el nombre ‘La espada de Damocles’, en la Figura 2-2 se puede apreciar el sistema. Este dispositivo permitía al usuario mejorar su percepción del mundo a través de gráficos generados por ordenador. Es así como se asientan las bases para la exploración en este campo.

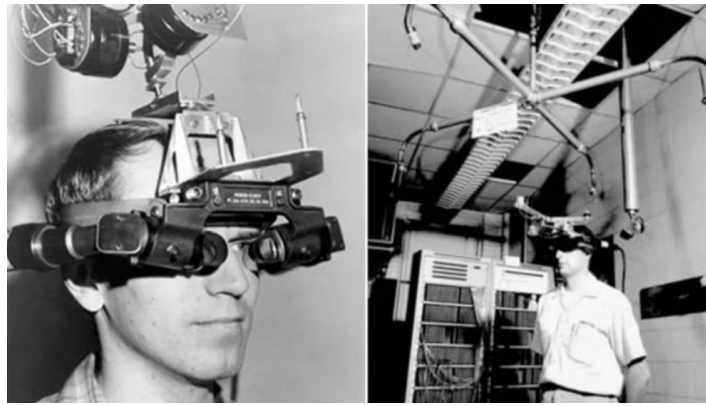


Figura 2-2 La espada de Damocles [6]

Posteriormente, en la década de los ochenta y noventa, se llevan a cabo numerosos avances en laboratorios universitarios, empresas y agencias militares. No es hasta el año 1990, que el término RA aparece a manos del investigador de *Boeing*, *Tom Caudell*. Dos años más tarde, en el laboratorio de investigación *Armstrong* de la Fuerza Aérea de los Estados Unidos, ‘Virtual Fixtures’ es creada por *Louis Rosenburg*, siendo uno de los primeros sistemas de RA que ya podían considerarse completamente funcionales. Este, permitía entrenar a pilotos en prácticas de vuelo, para que fueran más seguros posteriormente.

En el mundo de los deportes, también tuvo una de sus primeras apariciones, con el sistema gráfico ‘virtual 1st & Ten’, que proyecta una línea amarilla superpuesta en la transmisión, para que los espectadores pudieran ver cuanta distancia había avanzado el equipo para anotar un punto.

De los años 2000 en adelante, la RA se va implementando para los consumidores cotidianos. En el año 2009, la revista de actualidad masculina ‘Esquire’, utilizó esta tecnología para dar vida a sus páginas, de esta manera, cuando los lectores escaneaban la portada, podían ver a *Robert Downey Jr.* dirigiéndose hacia ellos.

No podía quedarse atrás *Google* quien, en 2014, presentó sus internacionalmente conocidas ‘*Google Glass*’, tratándose de un par de gafas que permitían a los usuarios experiencias inmersivas, pudiendo visualizar mensajes a la vez que hablaban con el dependiente de un café [7]. En la Figura 2-3 se puede ver a un usuario mostrando una de las lentes de dichas gafas.

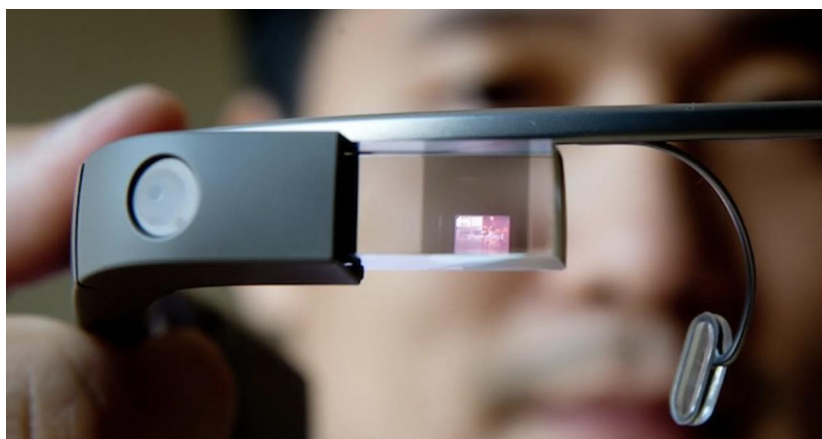


Figura 2-3 Google Glass [7]

En el mundo de la fabricación, venta de muebles y artículos para el hogar, *Ikea* lanza una aplicación para los móviles inteligentes, en 2017, que permite a los compradores, ver como quedarían los muebles en sus habitaciones, ahorrando mucho tiempo, a la hora de consultar medidas y espacios.

A medida que dependemos cada vez más de los dispositivos móviles, más casos de esta tecnología se irán viendo con el paso de los años, de manera exponencial, en nuestra vida cotidiana, cambiando así el modo en el que interaccionamos con el medio que nos rodea.

2.2.3 Realidad Aumentada en la actualidad y sus tipos

Hoy en día, la RA se está utilizando en numerosos campos, ya que se encuentra en desarrollo junto con las ciencias en las que se introduce. Además, deben recalcar los tipos que existen.

- **Basada en Marcadores o Imágenes:** En este método, solo se necesitan un código o marca (como puede ser un código QR) y una cámara. Suelen emplearse en móviles inteligentes, dando como resultado que el contenido se superpone en el código, ya que esta marca está preparada para ofrecer orientación y posición [8].
- **Sin Marcadores:** Este tipo, aún más sofisticado, permite prescindir de los marcadores gracias al GPS y acelerómetro integrados en nuestro dispositivo, es capaz de discernir la posición y ubicación en la que se encuentra el usuario. En el año 2016, salió al mercado una aplicación de entretenimiento llamada ‘*Pokémon GO*’, que se provechaba de esta tecnología [9].
- **Basada en la proyección:** Tratan de renderizar objetos 3D virtuales dentro del espacio físico del usuario. Su principio básico es la proyección de luz artificial sobre superficies del mundo real, como cristales acrílicos. La interacción del ser humano se realiza mediante periféricos táctiles o incluso gestos con sus manos o mediante la vista en el caso de tratarse de gafas como las nuevas desarrolladas por *Apple*. En la Figura 2-4 [10], se puede apreciar a un usuario haciendo uso de las mismas.

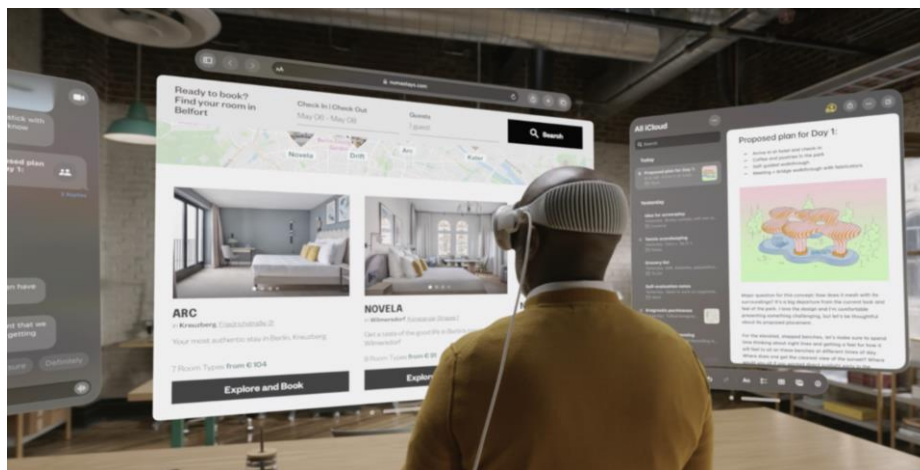


Figura 2-4 Usuario trabajando con las *Apple Vision Pro* [10]

- **Basada en superposición:** Reemplaza parcial o totalmente la vista original de un objeto, con la vista aumentada del mismo. El reconocimiento de objetos es crucial en este tipo de RA, ya que, si no son capaces de detectarlos, no pueden proyectarlos.

2.2.4 Realidad Aumentada en el ámbito marino

Como bien se ha visto, esta tecnología emergente tiene muchísimo potencial llegando también al ámbito marino, mejorando la seguridad en la navegación, la eficiencia durante la misma, y una mejor formación y aprendizaje de los marineros en el puente de navegación.

Se han realizado numerosos experimentos hasta la fecha en los que se persigue implementar la RA en los barcos. A medida que la industria marina se intensifica, en las operaciones marítimas, la seguridad se vuelve un factor más relevante que puede llegar a condicionar gravemente el flujo del tráfico marítimo si sugestión no es la adecuada. Diversos estudios, en los que se investigan una gran cantidad de accidentes marítimos, establecen que estos accidentes tienen en común una serie de factores

desencadenantes tales como la falta de comunicación entre el capitán y el personal a bordo, una observación inadecuada, o la poca anticipación ante situaciones de peligro, entre otros.

Como resultado, la RA supondría una gran ayuda como herramienta de apoyo a la navegación y al control del tráfico marítimo en general, ya que permitiría evitar incidentes, tanto humanos como materiales, y en caso de producirse un incidente, permitiría analizar de forma óptima sus causas para evitar futuros accidentes del mismo tipo. Este objetivo cumple directamente con los compromisos de la Organización Marítima Internacional (OMI) para afrontar el cambio climático y eficiencia energética [11].

Se realizan experimentos con el propósito de minimizar los costos y evitar la ejecución de pruebas en entornos reales. Para ello, como se aprecia en la Figura 2-5, se recurre a la simulación de espacios mediante Realidad Virtual, donde se implementa, de manera adicional, la tecnología de RA.



Figura 2-5 Usuario en puente llevando un dispositivo de visualización montado en la cabeza (HMD) [13]

En una de las aplicaciones existentes de la RA, se procedió a llevar a cabo un experimento donde el usuario, equipado con un dispositivo de Visión “HTC Vive Pro”[12], fue introducido en un escenario con icebergs, siendo el propósito principal navegar sin colisionar con ninguno de ellos. Este entorno simulado ofrecía ventanas superpuestas con información adicional, simulando la experiencia de utilizar un dispositivo de RA. Además, se incluyó una línea que representaba la estela que el barco seguiría con el rumbo introducido [13].

Estas ventanas proporcionaban detalles como el régimen de máquinas (velocidad), rumbo adoptado, dirección y velocidad del viento real y relativo, así como los contactos detectados en el radar de navegación. La combinación de todos estos elementos permitía al usuario anticipar el avance del buque sobre los icebergs, brindando una amplia ventana temporal para maniobrar el barco con mayor seguridad. Esto contrastó con otro sujeto que completó el mismo escenario, pero sin los elementos de RA, demostrando así la eficacia y utilidad de esta tecnología en entornos de navegación y seguridad marítima.

Como resultado del experimento, las conclusiones fueron bastante claras y concisas. Las características de la RA facilitaron, en gran medida, la labor en puente del usuario y sobre todo la seguridad en la navegación, siendo las siguientes las principales conclusiones obtenidas:

- **Comunicación en puente:** La información era compartida por los operadores de manera mucho más fácil, solo tenían que compartir los datos del Iceberg en cuestión y de manera automática el resto del personal lo podía ver directamente. En situaciones críticas ayuda a reducir el estrés y poder reaccionar mejor.
- **Movilidad del operador y comprensión de la situación:** Les permitía mantener la conciencia situacional mientras se movían por el puente, algo especialmente importante en buques grandes, cuyo desplazamiento es considerable, y las maniobras en canales restringidos debe ser preciso.

2.2.5 Realidad Aumentada en la Armada Española.

La RA, como se ha mencionado con anterioridad, habilita la superposición de datos digitales en el mundo real, posibilitando que los usuarios interactúen con la información de forma más orgánica y espontánea. Es por ello por lo que se presenta como una tecnología innovadora con el potencial de cambiar la forma en la que la Armada Española lleva a cabo sus operaciones.

En el contexto de la Armada, la aplicación de la RA abarca diversas áreas, desde la formación de la tripulación, hasta la planificación de operaciones, navegación, identificación de objetos y gestión de mantenimiento. Su utilidad se extiende incluso a la mejora de la seguridad y eficiencia en la vigilancia del tráfico marítimo, una de sus mayores responsabilidades. De esta manera la RA es de gran importancia para la Armada, elevando así su eficiencia en las futuras misiones.

Acto seguido, se presentan algunas de las aplicaciones que podría explotar la Armada para un mejor desempeño de sus labores en la mar:

2.2.5.1 Identificación de contactos asistida.

El oficial de marina, cuando se encuentra navegando en el puente de un buque debe ser conocedor de la situación que le rodea en todo momento. Esto incluye condiciones meteorológicas, régimen de máquinas y, lo más importante, los contactos de superficie. Estos pueden estar parados o llevando su propio rumbo y velocidad, pero en cualquiera de los casos nos encontraremos con problemas cinemáticos de movimiento, que hacen que el oficial de guardia en puente tenga que dar una serie de órdenes para gobernarlos.

Otra de las situaciones de riesgo es cuando se produce un hombre al agua, siendo esta una de las más indeseadas, por la posibilidad de perder a un hombre en la mar. Durante el día no sería difícil localizarlo gracias al color naranja del chaleco salvavidas cuando está abierto, pero si esto ocurre durante la noche o en situaciones de baja visibilidad, resulta muchas veces, casi imposible encontrarlo.

Los oficiales son instruidos durante su formación en la Escuela Naval Militar, de diferentes tácticas y procedimientos que les permiten recoger al naufrago hasta en las situaciones más desfavorables. Pero si contasen con la ayuda de la RA, podrían acudir a rescatarlo conociendo el punto GPS en el que se cayó al agua.

Como medida de solución, se propone la implementación de sistemas integrados en el buque que aprovechen la información disponible. Se sugiere la instalación de una cámara omnidireccional capaz de reconocer los contactos previamente identificados por los sistemas a bordo o basados en datos GPS, de forma que estos contactos podrían visualizarse a través de la ventana del puente, como se aprecia en la Figura 2-6 utilizando dispositivos de RA, como gafas especializadas o una pantalla acrílica con proyector, con el objetivo de minimizar la necesidad de movilidad por el puente mientras se utilizan dichos dispositivos.

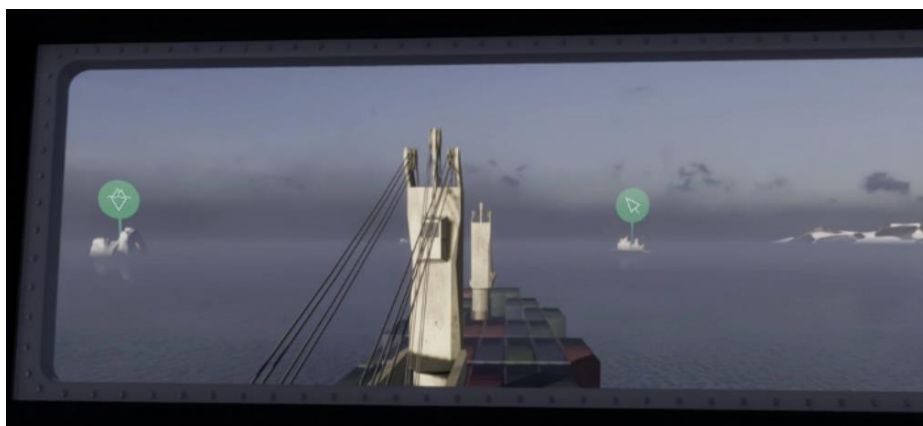


Figura 2-6 Icebergs mostrados en interfaz de Realidad Aumentada [13]

2.2.5.2 Incidencias en la sala de máquinas. Seguridad interior.

La sala de máquinas de un buque representa el núcleo mecánico de la embarcación, albergando motores, como el presente en la Figura 2-7 y sistemas esenciales como bombas contraincendios o calderas que proporcionan agua caliente a la dotación. La importancia de mantener una vigilancia constante en este aérea radica en la posibilidad de detectar cualquier irregularidad en los sistemas de propulsión o maquinaria que pudieran tener consecuencias para la seguridad a bordo.

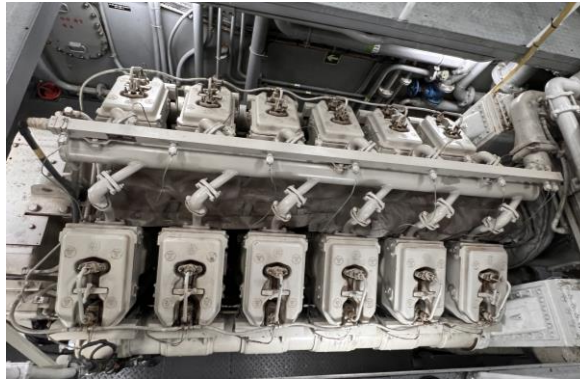


Figura 2-7 Motor Diesel del buque de la Armada Española ‘Intermares’ [autoría propia]

La monitorización continua de la sala de máquinas desde la cámara de control es llevada a cabo por los miembros del buque, a través de unas pantallas que muestran las temperaturas de los equipos vitales. Si alguno de ellos alcanzase valores fuera del rango de seguridad, se activa una alarma y una persona designada al mantenimiento de las máquinas se acerca para comprobar el estado en el que se encuentran.

Muchas veces el personal embarca en las unidades recién salido de las escuelas de formación y nunca ha tenido la oportunidad previa de estar en una sala de máquinas real o similar a las que podrían encontrarse en los barcos. Como resultado, en situaciones de estrés, no son capaces de encontrar con rapidez la ubicación de los instrumentos.

Aplicando la tecnología RA el personal puede recibir información de donde está ubicada la avería y el estado en el que se encuentra en tiempo real. De esta manera cualquier persona a bordo, independientemente de su labor a bordo, podría ser capaz de localizar la incidencia.

Muy relacionado con esto último, los Oficiales de marina son formados en Seguridad Interior. La seguridad interior es un equipo de la dotación, como se aprecia en la Figura 2-8, encargado de minimizar los daños de una emergencia para no perder el buque [14]. Sus componentes principales son el Oficial de Seguridad Interior y el jefe de trozo que, junto con el jefe de escena, montan un circuito de mangueras para apagar el incendio.



Figura 2-8 Dotación de un barco realizando un ejercicio de seguridad interior [14]

A la hora de preparar la maniobra, hay que apagar los cuadros eléctricos de la zona afectada para que el incendio sea menos peligroso, y posteriormente buscar las salidas de mangueras correspondientes, seleccionadas por el jefe de escena para atacar el fuego. En estas situaciones, al ser espacios pequeños y cerrados, el humo impide ver con claridad la ubicación de las boquillas de incendios. Implementando los dispositivos de RA, el usuario sería capaz de encontrarlos instantáneamente, y guiarse por el buque con mayor facilidad, además de saber dónde se encuentra el resto del personal.

2.2.5.3 Asistencia logística en almacenes.

Empresas a la vanguardia de sus respectivas industrias como *Amazon* o *Aliexpress*, ya se aprovechan del uso de la RA. Para garantizar la distribución global y que las entregas lleguen lo más rápido posible, necesitan que sus almacenes tengan trabajadores altamente cualificados y experimentados. Sin embargo, incluso contando con ellos, mediante la implementación gafas de RA son capaces de conocer la localización exacta de un paquete dentro del almacén, y puedan recuperarlo de manera eficiente y rápida, porque visualmente tienen acceso al diseño del almacén y pueden navegar rápidamente a la ubicación específica del paquete deseado.



Figura 2-9 Almacén Parque y Centro de Abastecimiento de Mareal de Intendencia del Ejército de Tierra [15]

En la Armada Española, cuando un barco entra en PIP para su restauración, debe cumplir una lista extensa de reparaciones y actualizaciones de equipos. Se levanta un parte de mantenimiento, y el arsenal responsable de dicho buque debe proporcionar los necesarios con la mayor antelación posible. A menudo al buscar un repuesto en los arsenales, como el de la Figura 2-9, puede ser una tarea difícil debido a que no se encuentra disponible o a que no es fácil de encontrar por la gran cantidad de sistemas almacenados. Sin embargo, con la ayuda de un equipo de RA, los trabajadores podrían encontrarlo al instante y enviarlo al buque que lo necesita en cuestión de horas [15].

2.3 Sistemas de aprendizaje automático y la visión artificial.

En el próximo apartado es importante situar la visión artificial dentro de un marco contextual que ayude a su entendimiento mediante la definición de tres conceptos: DL, ML e IA.

Cada uno de estos conceptos desempeña una función única y esencial, sentando las bases sobre las cuales se desarrollan las capacidades fundamentales de la visión artificial. El DL permite la extracción de características complejas y abstractas de los datos, el ML facilita el aprendizaje a partir de ejemplos y la toma de decisiones basada en modelos estadísticos, mientras que la IA que incluye a DL y ML como se muestra en el diagrama de la Figura 2-10 [16], engloba el diseño y la implementación de sistemas capaces de realizar tareas que requieren inteligencia humana, como el reconocimiento de patrones en imágenes. Estos pilares conceptuales son fundamentales para comprender el funcionamiento y las

aplicaciones de la visión artificial en diversos campos, incluyendo la vigilancia marítima y la navegación.

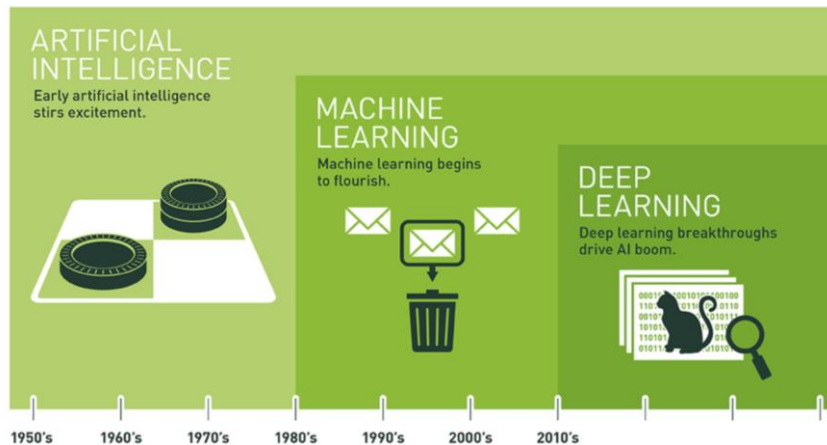


Figura 2-10 Diagrama que engloba la Inteligencia Artificial, Machine Learning, y Deep Learning [16]

2.3.1 Inteligencia Artificial

En primer lugar, antes de comenzar con la visión artificial debemos tratar la IA, que es el campo del que procede. En el diccionario de la Real Academia de la Lengua Española nos la define la “inteligencia” de la siguiente manera: “Capacidad de entender o comprender”[17]. Partiendo de la definición anterior la IA se puede precisar trasladándola al mundo de las máquinas.

Nosotros vamos a enfocar este concepto en una definición más cercana a la realidad, de la mano de *Marvin Lee Minsky*, científico estadounidense considerado como uno de los padres de la IA [18]. Decía:

“La inteligencia artificial es la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos, requerirían inteligencia”.

En otras palabras, se trata de crear sistemas que puedan imitar de manera autónoma ciertos aspectos del pensamiento y comportamiento humano, incorporando el conocimiento adquirido a los procesos para que tengan éxito al ejecutar las tareas que le han sido encomendadas.

A modo de ilustración se podría considerar el juego del ajedrez, como en la Figura 2-11 [19]. El ordenador podría tardar bastante tiempo en evaluar todas y cada una de las jugadas posibles para el próximo movimiento. Para evitar esta situación, se incorpora conocimiento previo sobre el juego en el proceso de búsqueda para batir al jugador en forma de procedimientos de evaluación inteligentes [20].



Figura 2-11 Ajedrez visto en GNOME [19]

Lo verdaderamente desafiante radica en el desarrollo de sistemas que no solo posean capacidad de retención de información, sino que también sean capaces de percibir y comprender su entorno, ofreciendo soluciones a las tareas asignadas. Esto implica definir patrones preestablecidos en una gran variedad de situaciones.

Principalmente existe una clasificación genérica con dos tipos de IA:

- **IA Integrada:** En la actualidad, en casi todas las fábricas del mundo se utilizan elementos mecánicos o máquinas, que reemplazan la acción humana. Estos dispositivos requieren software específico para su funcionamiento eficiente y seguro.
- **Software:** Se trata del núcleo de procesamiento de la máquina, para desarrollar algoritmos de reconocimiento de imágenes (uno de los temas centrales de este TFG), sistemas de identificación de voz y asistentes virtuales, entre otros.

De esta forma la IA se ha convertido en una disciplina crucial en el desarrollo de nuevas tecnologías digitales. Su importancia se refuerza gracias a la cantidad de bases de datos disponibles en la actualidad junto con el creciente desarrollo que ha tenido la informática en los últimos años, automatizando la resolución de problemas y permitiendo realizar el procesado y posterior análisis a partir de diferentes fuentes de información disponibles [21].

2.3.2 Machine Learning

Ahora que ya se ha introducido la IA debemos entender una de sus ramas fundamentales, el ML. Esta viene a resolver la incógnita principal que surge de la IA; ser capaces de entrenar y enseñar a las máquinas, a aprender de los datos introducidos para posteriormente optimizar los resultados y ejecutar procesos correctamente. En otras palabras, encontrar soluciones que discriminen los elementos que conforman la entrada, para encontrar correlaciones en grandes bases de datos (*datasets*). Para lograrlo surgen los denominados algoritmos de entrenamiento.

La herramienta persigue la mejora del análisis de datos con el propósito de establecer definiciones en etapas posteriores, ya sea optimizando los sistemas preexistentes o implementando nuevos, a través del uso estos algoritmos mencionados previamente.

Además, esta materia es prometedora en ciencias como mecánica, medicina, robótica, pero preocupa a algunos usuarios que afirman que crear computadoras que puedan volverse más inteligentes de manera progresiva sin necesitar de la intervención humana, podría volvernos dependientes de estos sistemas en el futuro [22].

2.3.2.1 Underfitting y Overfitting: el problema de la generalización.

Antes de explicar los diferentes modelos disponibles, hay que explicar dos términos importantes que se utilizan cuando se validan los resultados obtenidos en el proceso de entrenamiento en un entorno de ML: *Underfitting* y *Overfitting*. Cuando se entrena un modelo se intenta ‘‘encajar’’ los datos de entrada con los de salida, término que en inglés se indica como *fitting*. Por lo tanto, los términos más apropiados son ‘‘subajuste’’ y ‘‘sobreajuste’’, denotando la insuficiencia y la excesiva adaptación del modelo a los datos, respectivamente [23].

Cuando un modelo exhibe un desempeño óptimo en la fase de entrenamiento al ajustarse demasiado a los datos específicos, pero su habilidad para generalizar a nuevos datos se ve comprometida, se presenta un caso de sobreajuste, conocido como *overfitting*. Por otro lado, si el modelo no logra identificar de manera adecuada los patrones en los datos de entrenamiento, se está frente a un caso de ajuste deficiente, denominado *underfitting*. En la Figura 2-12 se aprecia el equilibrio entre ambos términos.



Figura 2-12 El equilibrio del aprendizaje [23]

Por otro lado, la generalización es una de las cuestiones más difíciles e importantes a resolver en el campo de la IA, ya que se trata de determinar como el modelo creado es capaz de extrapolar los resultados a unos datos con los que no ha sido entrenado, es decir, la capacidad del modelo de generar respuestas satisfactorias cuando los datos de entrada corresponden a una zona del espacio de trabajo de la que no se emplearon datos para su entrenamiento. La capacidad de generalización es una forma de comprobar la robustez del modelo creado, de forma que, si el modelo intenta aproximar demasiado los datos de entrenamiento (generando *overfitting*), no es capaz de generar respuestas correctas cuando los datos son nuevos. Por otra parte, cuando el entrenamiento no es suficiente, el modelo deriva en *underfitting*, de forma que ni siquiera es capaz de aproximar con precisión una respuesta para los datos de entrenamiento.

2.3.2.2 Aprendizaje supervisado

El primer modelo para abordar de los tres que disponemos es el aprendizaje supervisado, en el que es el usuario el que etiqueta y clasifica los datos manualmente, para que los patrones puedan ser detectados y posteriormente aprovechados por la máquina para etiquetar información nueva [24]. El sistema correlaciona datos y establece similitudes y diferencias entre ellos constantemente hasta que finalmente es capaz de hacerlo sin ayuda del usuario. En la Figura 2-13 [25], se aprecia un esquema de dicho aprendizaje.



Figura 2-13 Esquema Aprendizaje Supervisado [25]

Dentro de este aprendizaje tenemos dos tipos diferenciados:

- **Aprendizaje supervisado por clasificación:** Consiste en un algoritmo que trata de etiquetar a los ejemplos con 2 o más clases, es decir, utiliza la información previamente aprendida para elegir la correcta.
- **Aprendizaje supervisado por regresión:** En cambio el algoritmo por regresión es el entrenamiento para predecir un resultado en función de un rango de valores posibles. Estima un valor real y continuo mediante el análisis de datos previos.

Cuando se trate de un entorno más complejo, la solución brindada por el sistema son múltiples tipos de salidas. Así, por ejemplo, tratando de detectar objetos en imágenes, se les atribuyen cuatro números

para determinar la posición y proporción del rectángulo que lo enmarca (regresión múltiple). Además, el algoritmo tendrá que ser capaz de identificar el tipo de objeto, es decir, de realizar una clasificación.

2.3.2.3 Aprendizaje NO supervisado

Como bien indica este tipo de aprendizaje, solo se va a disponer de datos de entrada para que el algoritmo sea capaz de extraer y conocer los patrones que se repiten en los valores introducidos. Descubren agrupaciones de datos ocultos sin necesidad de intervención del usuario en datos que no han sido etiquetados a priori. Además, a medida que más datos se van introduciendo, la capacidad de la máquina de identificarlos se vuelve más intuitiva. En la Figura 2-14, se aprecia un esquema de dicho aprendizaje.



Figura 2-14 Esquema Aprendizaje no Supervisado [25]

Se dispone como ejemplo de los algoritmos de clusterización que buscan relaciones de semejanza entre los distintos datos, para determinar el número de categorías que existen en la información que recibe la máquina [26].

Por otro lado, un ejemplo bastante claro de este modelo de aprendizaje reside en los sistemas topográficos cuya entrada es una fotografía aérea que posteriormente da resultado un mapa. Se logra mediante el análisis de patrones como la distribución de las carreteras o ciudades, y la segmentación, como se aprecia en la Figura 2-15 (concepto que se explicará más adelante junto con la visión artificial) [27].

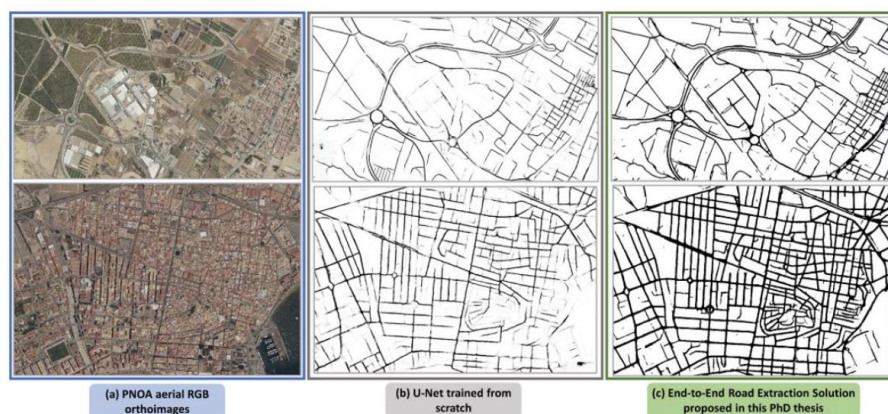


Figura 2-15 Segmentación del terreno con aprendizaje no supervisado [27]

2.3.2.4 Aprendizaje por refuerzo (Reinforcement Learning)

En último lugar, y a diferencia de los algoritmos anteriores, el aprendizaje por refuerzo sigue un modelo basado en “acción-recompensa”. Se trata de una máquina que irá recibiendo recompensas o penalizaciones del entorno en función de las acciones de tome. Podría asemejarse a la selección natural del científico de la evolución biológica, Charles Robert Darwin [28]. Solo aquellos caminos del algoritmo que desemboquen en el buen comportamiento de la máquina son los que continuarán

actualizando la IA del software. En la Figura 2-16 [29], se puede apreciar un esquema de dicho aprendizaje.

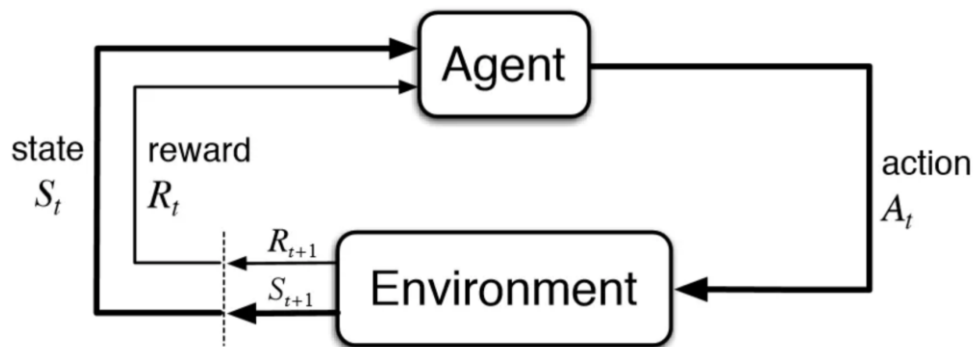


Figura 2-16 Esquema Aprendizaje por refuerzo [29]

En Trackmania [30], que es un popular juego de carreras en el que el jugador asume el papel de un piloto y se enfrenta a diversas pistas complejas y desafiantes en forma de obstáculos, reside un claro ejemplo de este tipo de algoritmo.

Un estudiante de la Universidad de Twente, desarrolló un programa con el fin de lograr que el vehículo sea manejado por la IA para que sea capaz de batir al usuario, o al menos, para dificultar su victoria [31]. Para lograrlo, el programa utiliza un tipo de aprendizaje en el que se eligen los modelos de vehículos que pueden avanzar más en el circuito, y luego se los somete de nuevo al proceso de aprendizaje. Al repetir este proceso, el modelo es finalmente capaz de hacer el recorrido completo de principio a fin. En la Figura 2-17, se muestra como el sistema mide las distancias hasta los márgenes de la carretera para poder ubicar el vehículo en ella, proceso similar al piloto automático presente en algunos vehículos que cuentan con conducción autónoma.

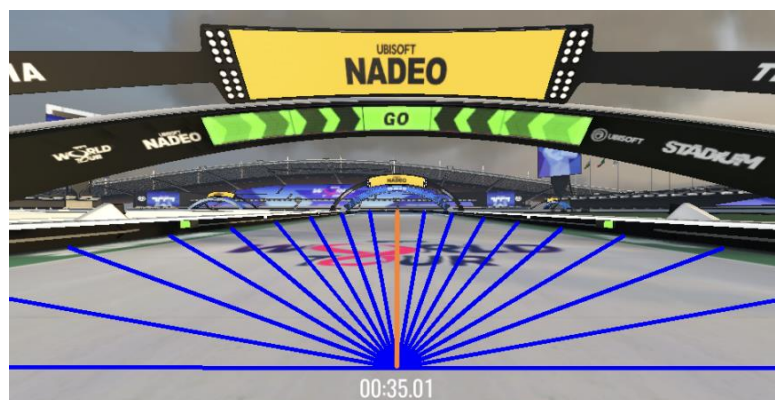


Figura 2-17 Cámara vista desde el vehículo, tomando medidas para ser capaz de ubicarse [31]

2.3.3 Deep Learning.

El DL aparece como evolución del ML que basa su algoritmo en árboles decisión, mientras que su predecesor se asemeja más al aprendizaje humano, como se aprecia en la Figura 2-18. Baza su funcionamiento en un algoritmo automático que imita la percepción humana del cerebro sus conexiones nerviosas diferenciadas a nivel jerárquico [32].

Al comienzo la red aprende tareas básicas de bajo nivel, cuyo resultado genera información más compleja que traslada a los niveles superiores para que consecuentemente, cada capa aumente su capacidad de abstracción y mejore los resultados. Esto requiere en su contra ordenadores de gran

potencia que sean capaces de solventar exigencias relacionadas con el crecimiento exponencial del número parámetros del modelo a entrenar.

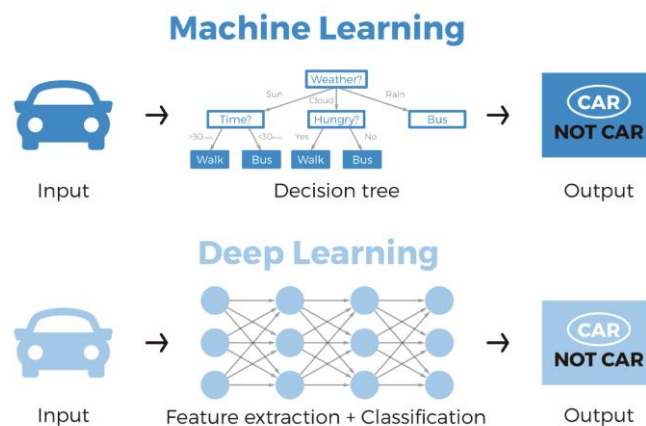


Figura 2-18 Diferencia entre *Machine Learning* y *Deep Learning* [32]

A la hora de trasladarlo a un problema real podría ser la diferenciación entre un racimo de uvas y un par de cerezas. En primer lugar, se estudiaría una capa que fuera capaz de distinguir los bordes entre zonas claras y oscuras, para que posteriormente otra recogiese esa información y pudiera ver el número de siluetas juntas en cada zona. Como resultado, una tercera capa determinaría si finalmente se trata de las uvas o de las cerezas.

Claramente la gran ventaja que ofrece el DL es la facilidad de resolver problemas más complejos por la fragmentación en capas que ofrece, para abordarlos de manera más sencilla. Muchas son las empresas que han descubierto su enorme potencial y se encuentran apostando por este tipo de tecnología.

2.3.4 Visión artificial

Tras habernos situado en el mundo de la IA, ahora es turno de explicar una de las áreas principales del TFG, la visión artificial. Esta intenta emular la capacidad que poseen los seres vivos de observar una escena para posteriormente interpretarla y actuar de manera conforme. Dota a las máquinas de un sistema de visión que permite automatizar la capacidad de reconocimiento visual a través del procesamiento de imágenes visuales. En la Figura 2-19 [33], se muestra la diferencia entre ambas perspectivas de visión.

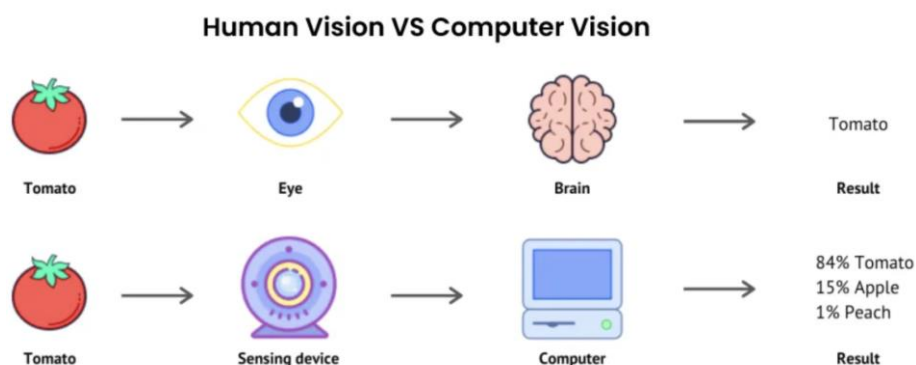


Figura 2-19 Diferencia entre visión humana y visión artificial [33]

Por lo tanto, la visión artificial tiene como tarea recoger la información a través una serie de sensores (cámaras), procesarla y finalmente obtener conclusiones de los datos de salida generados.

La visión artificial puede ser definida como los procesos de obtención, caracterización e interpretación de información de imágenes tomadas desde un mundo tridimensional a partir de imágenes bidimensionales [34]. A continuación, en la Tabla 1 se muestran los seis procesos que existen en la visión artificial:

Procesos	Nivel de Visión	Entrada	Salida	Área
Captura	Bajo	Imagen	Imagen	Procesamiento de imágenes
Pre-Procesamiento				
Segmentación				
Descripción	Medio	Objetos o regiones	Grupos de píxeles en bruto (objetos o regiones)	Análisis de Imágenes
Reconocimiento (clasificación)			Información cuantitativa de los objetos o regiones	
Interpretación	Alto	Objetos clasificados en categorías	Objetos clasificados en categorías	Visión por Computadora

Tabla 1 Procesos de la visión artificial [34]

También debemos considerar los distintos tipos de problemas a los que puede enfrentarse la visión artificial, como se muestra en la Figura 2-20, [35]:

- **Clasificación y detección:** Son los modelos más simples que tratan de determinar en primer lugar la clase de objeto presente en la entrada respecto de lo que la máquina ya conoce en su base de datos y en segundo lugar la detección aplica las *bounding boxes*. Estas encierran a los objetos en cajas para señalarlos.
- **Segmentación:** Es el siguiente escalón que como se indica en la tabla superior, atribuye una clasificación a cada píxel de la imagen para poder clasificarlos a cada objeto de la imagen y ofrecer una detección del objeto correcta. Sin una buena resolución de imágenes resulta muy difícil alcanzar este problema.

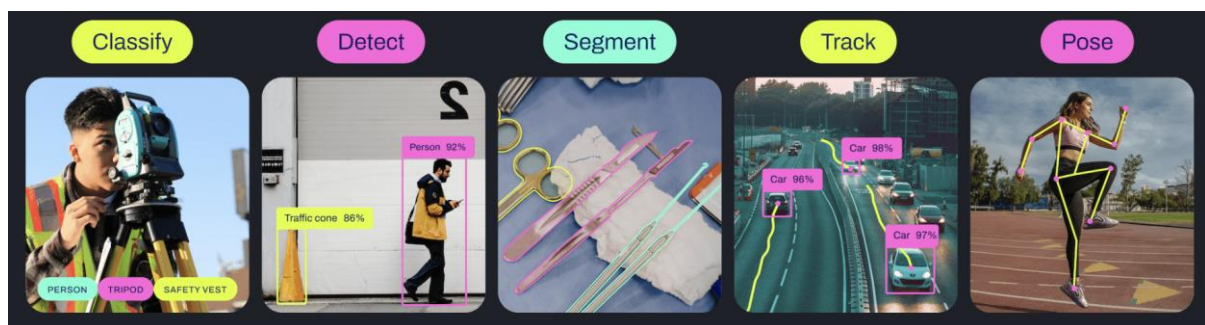


Figura 2-20 Diferentes tareas a resolver en visión artificial [35]

- **Tracking:** En algunos vídeos se puede hacer seguimiento de vehículos, o personas. Supone un gran reto porque requiere de la asignación de identificadores únicos a cada objeto y una base de datos muy determinada. De no ser así puede intercambiar el identificador entre objetos por similitudes entre ellos. En el presente
- **Keypoint detection y Gesture detection:** Es importante también ser capaces de detectar puntos clave de los objetos (como pueden ser articulaciones) para posteriormente el mayor reto, detectar el cambio de gestos humanos.

2.3.5 Redes Neuronales Artificiales

A la hora de poder enseñar a las máquinas distintos patrones o a procesar datos se ha extendido el uso de redes neuronales como modelos para resolver estas tareas. Comúnmente conocidas como Red Neuronal Artificial (RNA), son modelos matemáticos que se inspiran en el comportamiento biológico de las neuronas y el cerebro humano para así resolver problemas. Las neuronas van a ser nodos que se agrupan en numerosas capas que operan en serie, siendo clara la distinción entre la capa de entrada, capas ocultas que son las encargadas de procesar los datos y en último lugar la capa de salida que se encarga de generar los resultados [36].

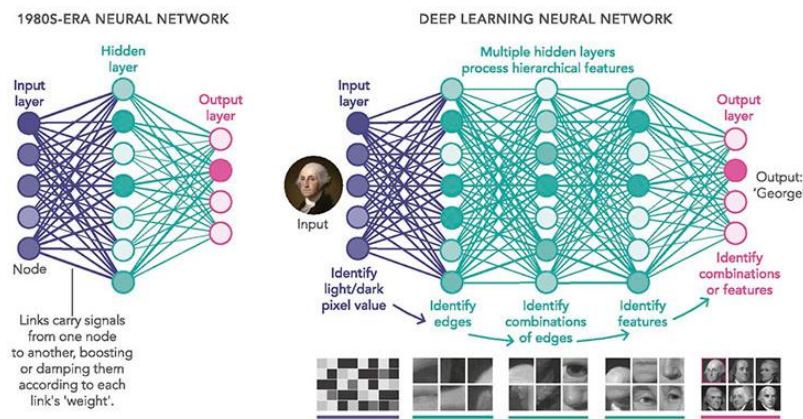


Figura 2-21 Diferencia entre redes neuronales simples y profundas [37]

A diferencia de las RNA, el DL cuenta con múltiples capas ocultas que procesan características jerárquicas, lo que les permite realizar tareas más complejas como se aprecia en la Figura 2-21 [37]. Existen numerosas RNA en función de su arquitectura, pero para el presente TFG nos vamos a centrar en las Redes Neuronales Convolucionales (CNN), de forma que otras arquitecturas se mencionarán de manera más breve.

2.3.5.1 Redes Neuronales Monocapa y Multicapa

En primer lugar, las redes neuronales monocapa son el escalón más básico de todo que se corresponde con redes neuronales, compuesta por una capa de neuronas que a su vez envían información a otra capa para realizar los diferentes cálculos. [38]



Figura 2-22 Redes Neuronales Monocapa y Redes Neuronales Multicapa [38]

Sin embargo, las redes multicapas son una evolución de las monocapa, cuya diferencia principal reside en que la segunda está compuesta por una capa de neuronas de entrada y otra capa para la salida, con capas intermedias que se encuentran ocultas entre ambas. En la Figura 2-22 se aprecia la diferencia entre ambas redes.

2.3.5.2 Redes Neuronales Recurrentes

Continuando con las redes neuronales recurrentes, estas se distinguen gracias a su estructura interconectada, permitiendo conexiones libres entre neuronas, incluso generando ciclos. Esto permite la incorporación de la temporalidad y la memoria en la red.[38]

Los datos ingresados en el tiempo t se procesan y circulan a través de la red, continuando en los instantes $t + 1$, $t + 2$, y así sucesivamente, como se muestra en la Figura 2-23.

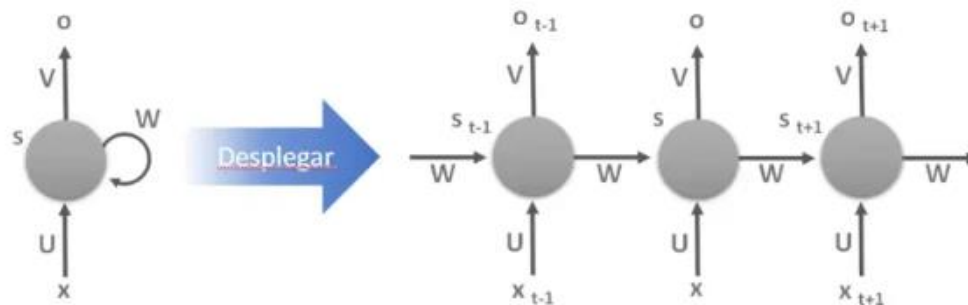


Figura 2-23: Redes Neuronales Recurrentes [38]

2.3.5.3 Redes Neuronales Auto Encoders

Este tipo de redes neuronales aprende a codificar y decodificar los datos con aprendizaje no supervisado. Su objetivo es aprender a representar los datos de entrada de manera más compacta, mediante un proceso de comprensión y reconstrucción. Sería como enseñar a un ordenador a simplificar la información para después poder recuperarla correctamente. La mayor ventaja que ofrecen estas redes es que evita que la máquina memorice los datos de entrada, evitando así el *overfitting*, que no es más que el exceso de modelado al que puede llegar una red cuando entrena en exceso para un conjunto de datos específico.

2.3.5.4 Redes Neuronales Convolucionales

Se trata de una red neuronal multicapa que se basa en la biología de la corteza visual animal, de forma que se pretende copiar para dotar de vista a las máquinas. Al tratarse de una red profunda, las primeras capas reconocen características como los bordes de objetos, líneas, etc. A su vez, las capas posteriores recombina estas capas en atributos de entrada de nivel superior [39].

Su principal ventaja reside en que todas y cada una de las partes de la red está entrenada para realizar una tarea reduciendo el número de capas ocultas y a su vez el tiempo de entrenamiento. Son muy potentes ya que pueden identificar patrones simples en imágenes y a partir de ellos generar tareas más complejas para ser finalmente capaces de identificar los objetos.

Para entender mejor este tipo de redes necesitamos desarrollar el concepto de convolución, que es una operación matemática que permite transformar numéricamente unos datos de entrada en unos datos de salida diferentes. Esto se consigue aplicando la matriz *kernel* junto con operaciones de productos y sumas entre la capa de partida y una serie de filtros (o núcleos) generando así un mapa de características. Esta convolución se puede apreciar como ejemplo en la Figura 2-24 [40].

La misma neurona sirve para extraer la misma característica de cualquier punto de la entrada, reduciendo así el número de conexiones.

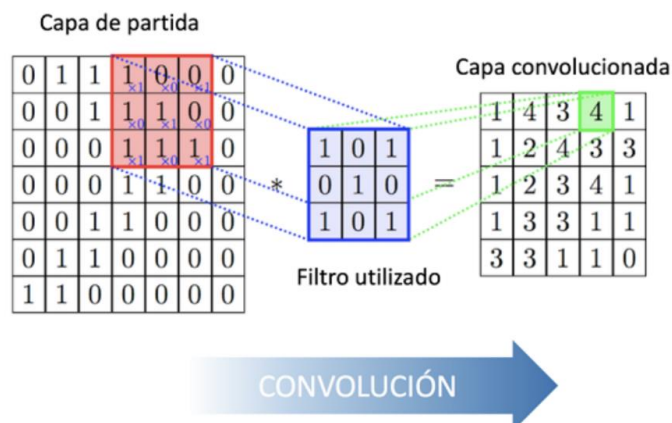


Figura 2-24 Convolución CNN [40]

2.3.5.5 Contexto histórico de las redes neuronales convolucionales

Haciendo un breve repaso a cerca del contexto en el que se desarrolló este modelo de red neuronal, es importante remontarse al año 1998, momento en el cual, *Yann LeCun* introduce la arquitectura *LeNet*, que ha perdurado como una de las más comunes en la actualidad. Sin embargo, debido a las limitaciones que tenían los ordenadores de la época a la hora de procesar cantidades masivas de datos, como los generados de conjuntos de imágenes o vídeos, su acogida final no fue tan exitosa [41].

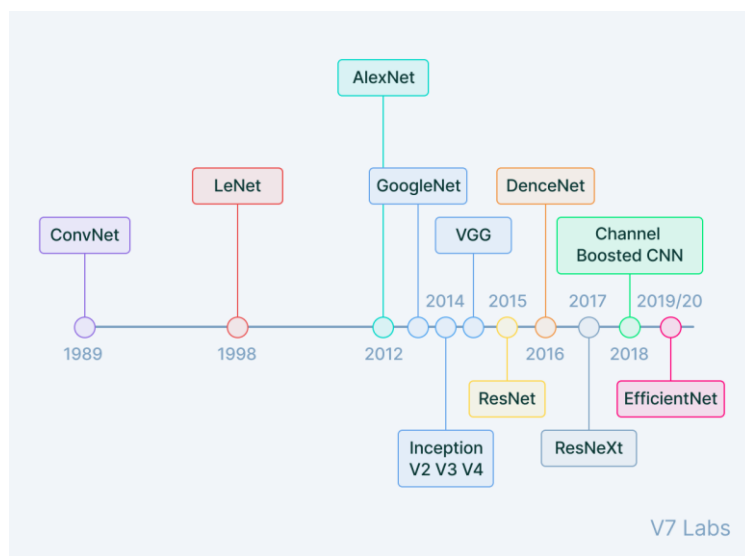


Figura 2-25 Línea temporal redes neuronales convolucionales [41]

No obstante, el verdadero avance en el desarrollo de las redes convolucionales tuvo lugar a partir de 2010, evidenciando por la sucesión casi anual de nuevos modelos, como se ilustra en la Figura 2-25. Destacando entre ellos, *AlexNet* nace en el año 2012, siendo considerada la pionera en el ámbito y un punto de referencia para desarrollos subsiguientes. Esta red neuronal consistía únicamente en cinco capas convolucionales y tres capas totalmente conectadas, logrando ser la primera entrenada exitosamente en *ImageNet*.²

² *ImageNet* es una base de datos gratuita y de libre acceso para cualquier usuario, que cuenta con miles de imágenes para entrenar a una red.

2.3.5.6 Tipos de capas en las redes neuronales convolucionales

A pesar de que el principio básico de este tipo de redes es que están compuestas por capas dónde tienen lugar una serie de convoluciones, no todas son del mismo tipo, es decir, algunas capas que contribuyen a su funcionamiento no son puramente convolucionales. A continuación, se exponen los tipos de capas que suelen emplearse en este tipo de modelos:

- **Capa Convolutiva:** En este tipo de capas, la red se limita a llevar a cabo operaciones de convolución, las cuales consisten en operaciones matemáticas que generan un resultado de salida. Como se apreciaba en la Figura 2-24, el filtro opera en todas las posiciones posibles de la matriz, generando así una nueva matriz como salida. Este procedimiento se repite para todas las capas convolucionales de la red, sin importar si es del mismo tipo o no. Como resultado de todas las operaciones se obtiene una matriz denominada *feature map*, obtenida para cada filtro o *kernel* empleado. Por consiguiente, a este tipo de estratos de procesamiento basado en convoluciones se les suele referir como *feature extractors*.
- **Capa ReLu:** En esta capa, a través de la función de activación ReLu (*Rectified Linear Unit*), se procede a la transformación de los valores negativos a cero, teniendo como propósito evitar que la suma resultante sea anulada debido a la presencia de números negativos.
- **Capa Pooling:** Tienen como único objetivo la reducción del tamaño de la imagen representada, situándose entre dos capas convolucionales. Su funcionamiento consiste en aplicar el filtro a cada posición posible de la matriz, y según el propósito asignado, seleccionar el mínimo, máximo o valor medio de los valores detectados en ese momento por el filtro.
- **Capa de normalización:** Como bien su nombre sugiere, esta capa se orienta a normalizar el valor de salida de capas precedentes, imponiendo a la red un aprendizaje más autónomo y evitar el *overfitting* en el modelo. Estas capas se insertan entre las capas convolucionales y las Capas *Pooling*, fomentando la generalización de la red neuronal.
- **Capa Fully-connected:** Tras procesar las capas de convolución y las capas *Pooling*, la información es “rebajada” gracias a la capa *Fully-connected*. Estas últimas se encargan de restaurar la estructura original de la red neuronal, asegurando que todas las neuronas de una misma capa reciban datos de entrada provenientes de la totalidad de la capa precedente. Este proceso es esencial para mantener la coherencia en la transmisión y procesamiento de la información dentro de la red.

2.3.6 El algoritmo de detección de objetos en imágenes (YOLO)

Tras haber aclarado lo que son las redes neuronales, debemos explicar cómo se aplica el DL para la detección de objetos en imágenes, una rama que pertenece a la visión artificial. Si bien es cierto que hace diez años aún se aplicaba el ML y sus técnicas convencionales para la detección, fue a partir del año 2013 que comenzaron las técnicas de DL. Estas se basan en las arquitecturas de redes neuronales que optimizan al proceso de detección. A simple vista puede parecer un método complejo, pero está siendo aplicados en diversos campos que nos rodean de nuestro mundo cotidiano, como cámaras de detección de tráfico o vigilancia en puertos.

Para llevar a cabo esta detección existen dos tipos de aproximaciones en función del número de pasos; *two-stage* y el *one-stage*. En el presente TFG el algoritmo que se va a utilizar es YOLO (cuyas siglas son “*You Only Look Once*” haciendo referencia a que solo requiere un paso), que pertenece al *one-stage*. Este algoritmo crea “cajas detectoras” en la zona de interés de la imagen y las clasifica al mismo tiempo [42], valorando las diferencias del objeto detectado, como se aprecia en la Figura 2-26.

Este proceso de detección del algoritmo YOLO define un cuadro delimitador alrededor del objeto que se desea detectar y designa cinco números: dos números para las coordenadas en las que se encuentra el centro del cuadro, dos números para las dimensiones del cuadro, y un último número para clasificar el cuadro (clase a la que pertenece el objeto en cuestión).

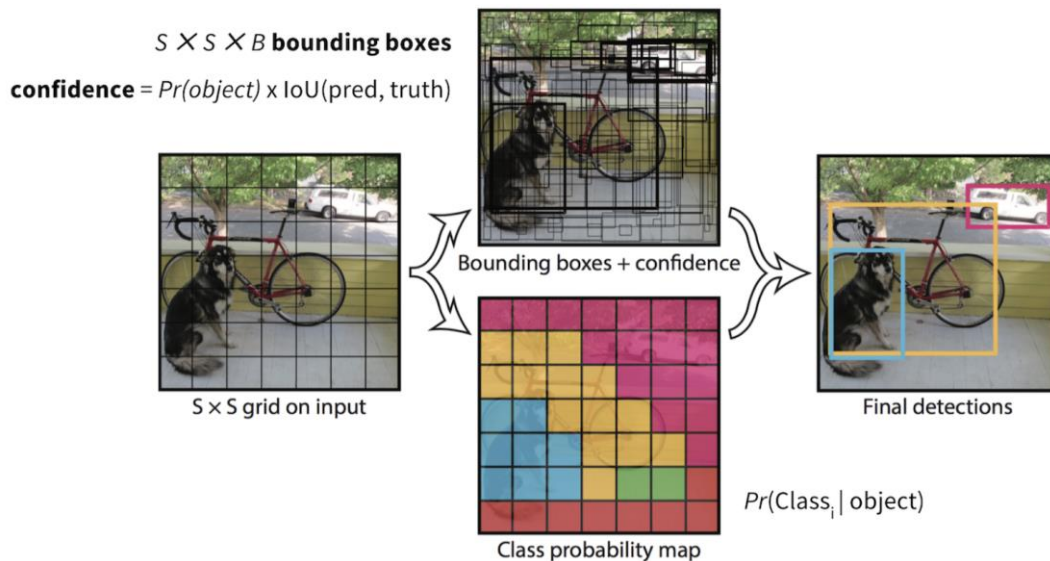


Figura 2-26 Detección de objetos con YOLO [42]

Los parámetros que se van a monitorizar durante el entrenamiento son:

- **Intersection over Union (IoU):** Este primer parámetro representa la convergencia en la intersección de dos ventanas, evaluando el área compartida entre ellas y dividiéndola por el total del área combinada de ambas ventanas. En el caso de una coincidencia perfecta entre ambas, el índice resultante es 1, mientras que, en ausencia de solapamiento, este será de 0. Considerando sistemas diseñados para la localización de objetos, cuando el valor exceda 0,5, estaremos hablando de una superposición significativa. Por el contrario, un índice negativo señala la inexistencia de superposición entre las ventanas evaluadas.
- **Precision:** El presente parámetro constituye una medida cuantitativa que evalúa el porcentaje de objetos detectados de manera acertada por la red neuronal. En otras palabras, representa el cociente entre el número de objetos identificados correctamente y el número total de objetos propuestos por el sistema de detección.
- **Recall:** Este indicador evalúa la proporción entre la cantidad de objetos identificados de manera precisa y el total de objetos que efectivamente deberían haber sido identificados.
- **Mean Average Precision:** Este último parámetro conocido por *mAP* derivado de sus siglas en inglés, constituye un indicador fundamental para evaluar la eficacia de los sistemas dedicados a la detección y localización de objetos. Surge de la combinación de variables previamente mencionadas, obtenidas a partir de la recopilación de resultados bajo múltiples condiciones, como, por ejemplo, la fijación de un umbral mínimo de *IoU*, el cual descarta detecciones que no cumplen con los criterios de similitud preestablecidos. Estos resultados, obtenidos en diferentes condiciones, nos devuelven el *mAP* como una media global del rendimiento del modelo en proceso de entrenamiento [43].

2.4 Control y monitorización de tráfico marítimo

2.4.1 Datos AIS

Los Datos AIS (*Automatic Identification System*) son fundamentales en el ámbito marino al ofrecernos información de manera automática, sobre la posición de un barco, su rumbo, velocidad y puerto de destino, para que los demás buques que se encuentren a sus alrededores puedan evitar colisiones y conocer más acerca de sus intenciones. Además, en caso de encontrarse en la bocana de un puerto (paso estrecho de mar que sirve de entrada a una bahía, fondeadero o puerto), pueden indicar que se encuentran fondeados, es decir, amarrados por su ancla al fondo marino.

Esto se logra mediante transpondedores instalados en las embarcaciones, es decir, un estándar para las comunicaciones marítimas ubicado en la banda VHF marina, transmitiendo datos que obtienen del GPS a bordo. Se podría incluso convertir un barco en estación en alta mar si lo equipamos con receptor y emisor AIS. En la Figura 2-27 [44], se muestra un visualizador AIS que presenta los datos transmitidos por los buques próximos, con el propósito de informar a la tripulación a bordo de una embarcación.



Figura 2-27 Un equipo dotado de AIS muestra la distancia y el rumbo de los barcos cercanos en un formato similar al radar [44]

Cuando una embarcación quiere incorporarse al sistema AIS necesita de una matrícula única y diferenciada, siendo este un número denominado MMSI (*Maritime Mobile Service Identity*). Este consta de un código de 9 dígitos provisto por la autoridad marítima del país de origen de la embarcación. La Dirección General de la Marina Mercante es la responsable en España que concede dicho código tras pagar la correspondiente tasa. Nuestro pabellón español empieza con 224 o 225.

También se dispone del número IMO a nivel internacional, como se aprecia en la Figura 2-28 [45], que a diferencia del MMSI es de obtención gratuita y consta de las siglas IMO seguido de siete dígitos. Este código identifica a la embarcación permanentemente a lo largo de su vida útil, independientemente del cambio de bandera. Fue introducido en el año 1987, como medida de prevención del fraude marítimo y seguridad para los buques mercantes.

Como breve comparativa entre ambos códigos expuestos, el MMSI se utiliza principalmente para la comunicación y la identificación de embarcaciones con sistemas radio (llamada selectiva digital) y navegación, mientras que el número IMO es un identificador permanente e invariable.



Figura 2-28 Velero de tres mástiles *Eendracht* con su respectivo IMO (8814275) siendo Róterdam su puerto de registro [45]

2.4.2 ¿Qué buques pueden transmitir Datos AIS?

Debido a la útil herramienta en la que se ha convertido, la Organización Marítima Internacional (OMI) exige el uso obligatorio del AIS en las siguientes situaciones:

- En todos los buques cuyo desplazamiento sea mayor de 500GT (*Gross Tonnage*, Toneladas de Registro Bruto en español, unidad utilizada en la industria marítima para hacer referencia a la capacidad de carga que se posee, en este caso equivaldría a 500 toneladas).
- En caso de que pese más de 300GT, pero además se encuentre en un viaje internacional también deberá llevar el AIS.
- Buques de pasajeros.

Es importante tener en cuenta que las regulaciones pueden variar según la zona en la que se navegue y las regulaciones marítimas de cada país. A pesar de ello los buques de la lista anterior que se encuentren navegando o fondeados van a tener que llevarlo en funcionamiento siempre. Bajo una situación de peligro, si el capitán lo considera necesario para la seguridad de su barco, podrá apagarlo de manera inminente.

Cada año y de manera global, son más de 400.000 dispositivos AIS transmiten información que es recopilada por satélites y estaciones terrestres para rastrear su movimiento incluso en áreas oceánicas lejanas. En algunas plataformas como *Marine Traffic* cualquier usuario puede ver a nivel mundial los buques transmitiendo AIS gracias a las estaciones terrestres y satélites que los recojan.

Los buques militares no están obligados a llevar AIS en las mismas condiciones de los buques civiles. Muchas veces se encuentran en ejercicios con planes de comunicaciones restrictivos, por lo que deben llevarlo apagado, o simplemente por cuestiones de inteligencia. Sin embargo, cuando se encuentran entrando o saliendo de puerto lo llevan encendido para favorecer la seguridad en la navegación y porque la normativa IMO lo exige.

2.5 Datasets y tipos

Los denominados conjuntos de datos, comúnmente referidos como *datasets* en inglés, son generalmente colecciones de datos utilizados en informática y análisis de datos, que están tabulados en filas y columnas. Pueden presentarse en diversas formas, desde números hasta imágenes dependiendo de área de estudio o su propósito [46]. De esta manera, para analizar grandes volúmenes de datos son necesarios sistemas especializados para procesarlos. No es lo mismo analizar los coches de cada vecino del barrio en el que vivimos, que el *dataset* conjunto que utiliza *Toyota* [47] para recopilar información del rendimiento, registros de mantenimiento y reseñas de los clientes de sus vehículos. Por consiguiente,

la calidad de las bases de datos es fundamental en los análisis posteriores, ya que los algoritmos y soluciones utilizados en nuestros sistemas se derivan de la experiencia de trabajar con estos *datasets*.

Desde un punto inicial se dispone del conjunto de datos COCO usado principalmente como herramienta para el reconocimiento de imágenes. Se trata de un conjunto de datos publicado por *Microsoft* para la segmentación y detección de objetos a gran escala, que incluye una amplia gama de elementos cotidianos que la comunidad de visión artificial utiliza para entrenar y probar sus modelos, de forma que entre sus características cabe destacar que contiene 330.000 imágenes y 80 categorías de objetos [48]. En líneas generales se disponen de los siguientes tipos de datos en los *datasets* disponibles *on-line*:

- **Archivo:** Se trata de un fichero único e independiente que contiene información para que el sistema los ejecute y pueda trabajar con ellos. Al encontrarse almacenados de manera local, cuentan con la ventaja de ser más rápidos. Sin embargo, cuentan la limitación a la hora de establecer conexiones con otros *datasets* que no están integrados dentro del mismo sistema.
- **Folder:** Una carpeta se refiere a un conjunto de datos que se han guardado en el mismo directorio. Generalmente cuentan con una estructura similar, por lo que pueden procesarse juntos con facilidad y están vinculados entre sí.
- **Database:** Son bastante similares a los archivos, se almacenan en sistemas de gestión de bases de datos, que permiten almacenar, modificar y extraer los datos de manera eficiente y segura, además de poder monitorizarlos de manera digital.
- **Dataset Web:** Pertenecen a información accesible a través de direcciones web específicas, y su obtención implica acceso a internet.

Es cierto que aunque no sea común encontrar *datasets* enfocados a entornos marítimos, ya existen aplicaciones en la actualidad que sí que cuentan e incorporan esta forma de trabajo con datos para fines como la vigilancia de puertos, batimetría, biología marina, etc. [49]. Se muestran algunos ejemplos de *datasets* para entornos marítimos a continuación:

- **Vigilancia de puertos:** La vigilancia marítima tradicionalmente ha constado de personal profesional capacitados para la gestión del tráfico marítimo para garantizar la seguridad en la navegación, sin embargo, en la actualidad, el campo de visión artificial ha introducido una notable revolución en este ámbito. Un claro ejemplo de ello viene de la mano del estudio realizado en el río *Yangtze*, China, donde se realizó un experimento para detectar barcos en tiempo real utilizando en el algoritmo YOLOv3 original para barcos, detectando mercantes portacontenedores como se muestra en la Figura 2-29 [50].

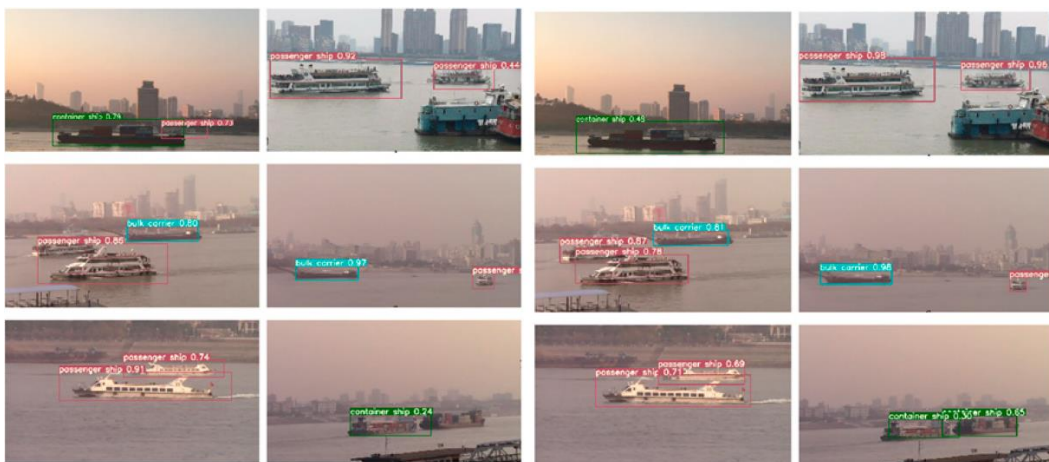


Figura 2-29 YOLOv3 aplicado a la vigilancia marítima en el río *Yangtze*, China [50]

- **Vigilancia Marítima:** El interés por la vigilancia marítima ha aumentado notablemente durante los últimos años, emergiendo como un asunto de gran importancia para asegurar la integridad del comercio internacional y las operaciones de defensa. La visión artificial supone una herramienta de gran potencial para abordar diversos problemas directamente relacionados con este entorno marino, desde la contaminación ocasionada por actividades humanas como derrames de combustible, hasta la detección y localización de incidentes en el mar, o embarcaciones sospechosas.
- **Biología submarina:** En el ámbito submarino esta aplicación es de relevante utilidad, contando *datasets* que documentan la vida marina en ecosistemas específicos. Estos recursos son de gran utilidad para investigar los patrones migratorios de determinadas especies, sus hábitos de alimentación o el control y seguimiento de animales en peligro de extinción como tiburones o ballenas [51]. En la Figura 2-30, se puede ver un ejemplo de este tipo de *datasets* marinos.

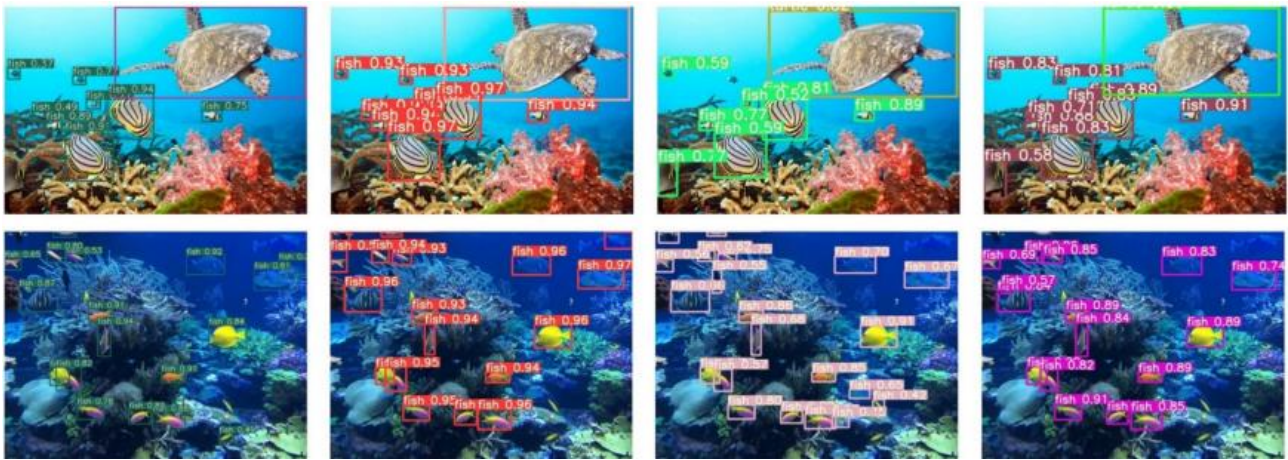


Figura 2-30 *Dataset* en el ámbito submarino [51]

2.5.1 China Maritime Datasets.

Son numerosos los puertos a lo largo del mundo que cuentan con cámaras para poder cumplir con sus objetivos nacionales de monitorización del tráfico marítimo. Un gran ejemplo es el experimento realizado con un *dataset* creado en el río Yangtze, China, como se indicaba anteriormente, donde un grupo de investigadores lo crearon desde cero, con la ayuda de YOLOv3 y el algoritmo DeepSORT que se utiliza para el seguimiento de objetos en tiempo real [50].

En la Figura 2-31 se muestra el diagrama de flujo del proceso seguido que se divide principalmente en *ship detection* y *ship tracking*, que serán explicados a continuación para entender mejor su funcionamiento:

- **Ship detection:** El *ship detection* implica la identificación de buques en imágenes o vídeos, mediante el algoritmo YOLOv3, que incluye la optimización de los valores iniciales de los marcos de referencia mediante algoritmos de agrupamiento como *Kmeans*, la elección de clasificadores adecuados, como *Softmax*, y la optimización del algoritmo de supresión de no máximos mediante enfoques como *Soft-NMS*, que suavizan la función de puntuación para reducir los errores de detección.
- **Ship tracking:** El seguimiento de embarcaciones en el algoritmo DeepSort consiste en mejorar la precisión mediante la combinación de información de movimiento y características visuales. Además, utiliza el filtro *Kalman*, un algoritmo de estimación que predice el futuro estado de un sistema dinámico utilizando mediciones del pasado y de la

actualidad. De esta manera es capaz de predecir la posición de la embarcación y prioriza el seguimiento de objetivos más frecuentes.

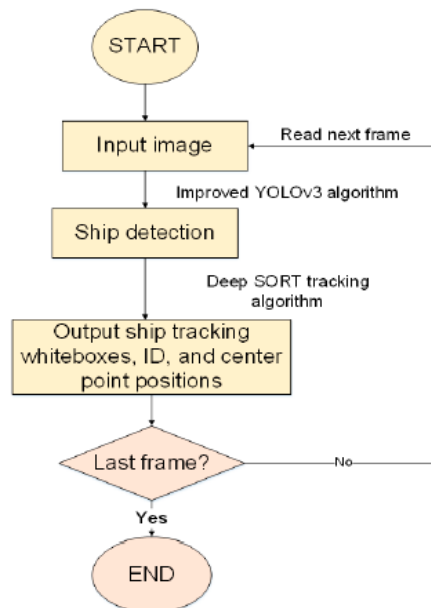


Figura 2-31 Diagrama de flujo del experimento [50]

Como se aprecia en la Figura 2-32, el diagrama de flujo muestra el entrenamiento completo. El proceso comienza con la creación de un *dataset* que contiene imágenes y etiquetas de diferentes tipos de barcos. A continuación, se utiliza un algoritmo de filtrado guiado para mejorar la calidad de las imágenes. Luego, se utiliza un algoritmo de extracción de características para extraer elementos clave de las imágenes, como el tamaño, la forma y el color del barco. Estas características se utilizan luego para entrenar una red neuronal para clasificar los barcos en diferentes categorías que una vez esté entrenada, se puede utilizar para clasificar barcos en un vídeo.

El sistema se evalúa utilizando un *dataset* que contiene imágenes de barcos que no se utilizaron para entrenar la red neuronal denominado *test*.

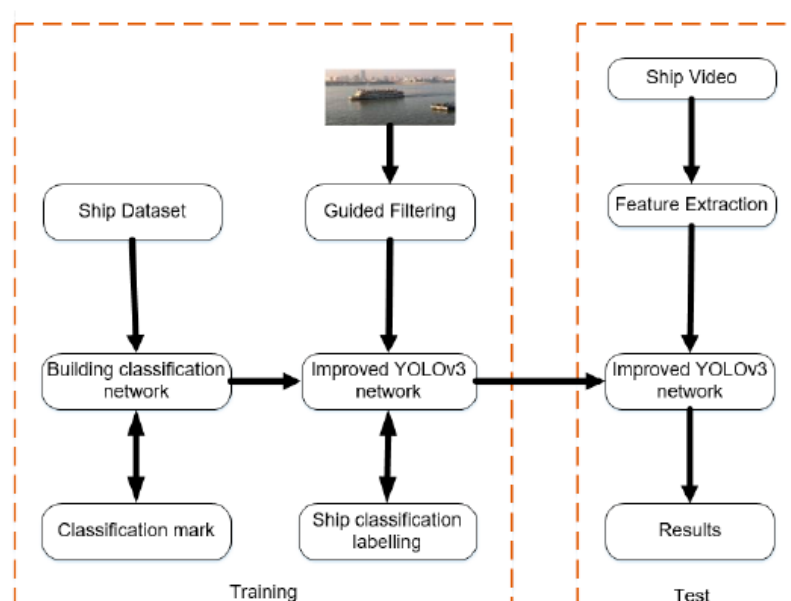


Figura 2-32 Esquema del proceso de aprendizaje del modelo [50]

2.6 Aplicaciones para la Armada en el siglo XXI.

Como se había comentado en la introducción del presente trabajo, en las Líneas Generales para la Armada según el AJEMA [1], la institución no contempla distanciarse de la transformación digital a la que se está sometiendo la sociedad en general, y más particularmente en la industria. Como se evidencia a continuación, la implementación de técnicas de inteligencia en la Armada se ha convertido en una cuestión de actualidad, destacándose el CESADAR [52] y el CIA2 como sus principales instituciones actuales. Además, se muestra una serie de áreas que podrían ser objeto de más desarrollo en el futuro para las situaciones a las que la Armada se enfrenta diariamente y como siempre, ligada a ese nuevo concepto de Armada 4.0, muy similar al de Industria 4.0 ilustrado en la Figura 2-33 [53]. Contexto y motivación



Figura 2-33 Funcionamiento de la Industria 4.0 [53]

2.6.1 Estado actual de estas técnicas en la Armada

Como ya se venía introduciendo, tanto el CESADAR como la CIA2 son los dos organismos principales con los que cuenta la Armada para el desarrollo e investigación de las técnicas abordadas en el presente trabajo.

Uno de los avances más significativos dentro del ámbito de la IA, es la reciente creación del CIA2 en el año 2011 y encuadrado en la Escuela Naval Militar, centro que cuenta con el objetivo de promover la investigación de la IA, estableciendo sólidas relaciones entre la industria y centros de investigación, y contribuyendo al desarrollo doctrinal de este campo científico. Hasta el momento, cuenta con varios casos de éxito, como es el proyecto SEDA, que trata de la obtención de imágenes a través de satélite, contando además con una aplicación web de inteligencia geoespacial que permite la detección automática de anomalías en dichas imágenes.

Por otro lado, la Armada cuenta, además con el CIA2 ubicado en el Arsenal de Cartagena. Este centro se enfoca en el mantenimiento predictivo de los motores de los buques de la Armada, es decir, crear una base de datos que relacione las anomalías detectadas a lo largo de los años, con las presentes para agilizar la detección de averías y solucionarlas de manera más eficiente. Otro campo donde se podría apreciar la aplicación de estas técnicas 4.0, es en las acciones de medidas contra minas, ya que la IA junto con esas grandes bases de datos, podrían analizar la actividad humana en una determinada área del entorno marino y disminuir los riesgos de las operaciones al poder estimar el porcentaje de riesgo que pudiera existir por la presencia de minas [54].

En resumen, el interés y desarrollo de la IA en la Armada son evidentes, y los resultados indican que esta tendencia se mantendrá en los próximos años.

2.6.2 Deep Learning en la Armada

Existen unas líneas de desarrollo para la Armada relacionadas con las técnicas de DL abordadas en el presente trabajo. A continuación, se muestran cuáles podrían ser dichas líneas (relacionadas con el presente trabajo) y por qué son de importancia para la Armada. Las aplicaciones son las siguientes:

- En primer lugar, la creación de una interfaz virtual, que pudiera incorporarse en la ventana del puente de navegación de un buque, permitiendo al oficial de guardia tener la información de las diferentes embarcaciones a tiempo real, a través de un *display* integrado sobre los cristales. Es decir, a través de los diferentes equipos a bordo, como radares de navegación, transpondedores AIS, y los datos cinemáticos del buque, facilitar la información de manera que el oficial pueda tomar decisiones de manera más ágil y sencilla.
- Como se indica en el Anexo II: Reflexiones Éticas y Sociales, en la guerra actual entre Rusia y Ucrania se está empleando el uso de drones para enfrentar buques, de manera que resulte un golpe sorpresa para el enemigo. De esta manera, integrando a los sistemas de combate de los buques, junto con cámaras, un sistema que aplicara un modelo cuyo aprendizaje permitiera la detección y seguimiento de estos contactos de manera automática, estos podrían defenderse ante estas nuevas amenazas asimétricas. Incluso podrían emplear gafas de realidad virtual, tal como ya lo hacen algunos soldados ucranianos, como ilustra en la como en la Figura 2-34 [55].



Figura 2-34 Soldado Ucraniano adiestrándose para el uso de drones [55]

3 DESARROLLO DEL TFG

3.1 Introducción

En este tercer capítulo se exponen las etapas fundamentales que han dado forma a este proyecto, seguidas por las conclusiones obtenidas. En primer lugar, la inmersión en nuevos conceptos ha sido esencial para una comprensión adecuada y profunda de nuestros objetivos. Desde el entrenamiento de una IA capaz de identificar las diferentes embarcaciones, hasta la habilidad para asociarles los datos AIS correspondientes, cada paso ha sido crucial para el desarrollo de este estudio.

A continuación, se expondrán las diversas herramientas de software y entornos de trabajo empleados en el presente proyecto, con especial énfasis en las herramientas para crear *datasets* y entrenar modelos. Además, se detallarán los puntos de origen de los vídeos capturados, especificando los periféricos utilizados durante el proceso de obtención de los datos.

El propósito inicial de este estudio es la creación de un *dataset* denominado "*MARINE_PORT*" que será empleado posteriormente para entrenar un modelo de visión artificial basado en el algoritmo YOLO, con el fin de realizar un seguimiento preciso de las embarcaciones presentes en los vídeos. En esencia, este proyecto se centra en la aplicación del DL, una técnica de IA que permite el aprendizaje automático a partir de datos, haciendo énfasis en su adaptación para entornos marinos complejos. Este enfoque posibilita la creación de modelos altamente precisos y adaptables, los cuales pueden mejorar sus prestaciones a medida que se les suministra nueva información.

Sin embargo, este Trabajo Fin de Grado no se centra exclusivamente en el entrenamiento de este sistema. Su verdadera proa pone rumbo hacia el desarrollo de una solución aplicable a las cámaras instaladas en puertos que participan en la vigilancia marítima con el objetivo de mejorar sus prestaciones mediante sistemas de RA. A través de conjuntos de datos y bases de datos específicas, junto con la lectura de datos AIS, se busca identificar embarcaciones en tiempo real y verificar su información correspondiente. En resumen, se pretende crear una interfaz aplicable al mundo de la RA para la vigilancia portuaria.

3.2 Aprendizaje y formación en la materia.

En la introducción de este capítulo se ha resaltado la importancia de un periodo de formación en los conceptos clave abordados durante el trabajo, particularmente enfatizándose en la relevancia actual que tienen las técnicas de *DL*, las cuales están en constante evolución y cada vez más integradas en diversos aspectos de nuestra vida cotidiana. De esta manera, en los siguientes apartados se parte desde una comprensión general del concepto de IA hasta la especialización en la tarea específica de detección, clasificación y seguimiento de objetos en vídeos, además de conocer cómo trabajan los sistemas que

emiten y reciben datos AIS, así como los tipos existentes en la actualidad. En la Figura 3-1 se presenta un esquema que abarca todas las fases de aprendizaje.

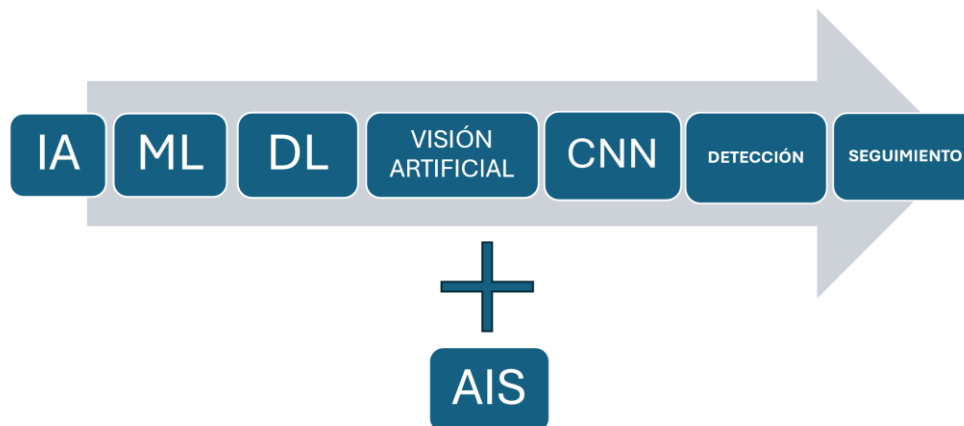


Figura 3-1 Fase de aprendizaje [autoría propia]

De esta forma, para comprender a fondo los conceptos abordados en este trabajo, es necesario comenzar desde el principio, desde los fundamentos más básicos. En primer lugar, se ha investigado acerca de la IA y sus aplicaciones, centrándonos en las técnicas de ML, que capacitan a las máquinas para realizar tareas progresivamente más complejas, hasta el límite deseado, pudiendo incluso llegar a comprender conceptos abstractos. Más específicamente se han explorado las propiedades de los sistemas basados en técnicas de DL, que dividen el aprendizaje en múltiples capas representando un modelo jerárquico de aprendizaje. Por ejemplo, en el reconocimiento facial, las primeras capas identifican patrones sencillos (líneas verticales, horizontales, curvas, etc.), utilizados en las capas posteriores son empleadas para establecer los patrones asociados a la detección de los ojos o la nariz y otros elementos faciales, para, en las capas finales, ser finalmente capaz de distinguir un rostro humano. Se trata, por tanto, de un proceso en el que la información se vuelve más abstracta a medida que pasa por cada capa, desde los píxeles iniciales que representan la entrada de información, hasta el establecimiento de la detección y análisis de una cara humana, un concepto mucho más complejo y abstracto.

Posteriormente, en el proceso de formación, es crucial conocer los problemas comunes que aparecen en la detección y seguimiento de objetos en imágenes en general (oclusiones, cambios de luz, deformaciones en la vista obtenida, etc.) y, específicamente, aquellos relacionados con el uso de las redes convolucionales, las cuales se especializan en la búsqueda de patrones en conjuntos de datos estructurados, como son las imágenes.

Asimilados dichos conceptos, es necesario seleccionar algoritmos adecuados para las tareas del presente trabajo. Para la detección de objetos en imágenes se cuenta con el algoritmo YOLO, mientras que para el seguimiento de objetos en vídeos se empleará el algoritmo *ByteTrack*, los cuales serán detallados más adelante.

Finalmente, tras haber comprendido los conceptos de visión artificial, redes convoluciones, detección de objetos en imágenes y seguimiento de los mismos, es importante conocer los distintos mensajes AIS que se pueden recibir, qué significado tiene cada uno, y cómo funcionan los transpondedores y especialmente en nuestro caso, los receptores.

A continuación, se introducen los recursos empleados durante la ejecución de este proyecto.

3.3 Herramientas de Hardware

3.3.1 Cámara ELP 8 megapíxeles

El dispositivo de grabación seleccionado es una cámara web de la marca ELP como se puede apreciar en la Figura 3-2 [56], en su versión de 8 megapíxeles con un enfoque múltiple de alta definición, específicamente 1080P a 30 fotogramas por segundo. Este dispositivo ha sido elegido por su capacidad para ajustar manualmente el zoom en un rango de 5-50 mm, así como la cantidad de luz con la que se toman vídeos (ISO). Este último aspecto es crucial, ya que garantiza la consistencia de los parámetros de captura necesarios para el entrenamiento de algoritmos de IA.



Figura 3-2 Cámara ELP 8 MP [56]

Adicionalmente, su reducido tamaño y capacidad *plug and play*, han permitido su instalación en diversos entornos para la captura de imágenes. Cabe destacar que la cámara cuenta con una rosca en su parte inferior para acoplar diferentes trípodes según el lugar de grabación

3.3.2 Teléfono móvil

Para las primeras tomas, dado que no teníamos la cámara mencionada previamente, se quiso poder contar con algún vídeo inicial para poder ir haciendo pruebas iniciales, por lo que se utilizó el *iPhone 13 PRO-MAX* [57]. Este dispositivo ofrece una calidad de imagen excepcional gracias a su potente conjunto de lentes en tres cámaras diferentes. Destaca por sus 12 MP que, combinados con su teleobjetivo, gran angular y su zoom digital de hasta 15 aumentos, es capaz de ofrecer capturas de vídeo a larga distancia sin comprometer la calidad.

3.3.3 Receptor AIS inalámbrico QK-A024

El receptor AIS inalámbrico QK-A024 es un dispositivo diseñado para capturar y decodificar señales emitidas por embarcaciones o estaciones radio, lo que permite el seguimiento preciso de su ubicación en tiempo real. Es de gran utilidad a la hora de facilitar la recopilación de datos fundamentales para el análisis del tráfico marítimo.



Figura 3-3 Receptor AIS inalámbrico QK-A024 [58]

Este periférico destaca por su capacidad multiplexora, es decir, puede combinar señales de identificación automática (AIS) con datos de navegación electrónica estándar (NMEA 0183) para dispositivos electrónicos marinos GPS, sondas, radares, entre otros.

Ofrece la posibilidad de trabajar de manera inalámbrica, pero junto con una antena como se aprecia en la Figura 3-3 [58], y su salida de datos AIS por USB se pueden integrar los mensajes y leerlos en LINUX utilizando un *script* en lenguaje *Python*.

3.4 Herramientas de software y entornos de trabajo.

Para este proyecto, se requieren varias etapas de etiquetado de imágenes, según el número de entrenamientos deseados para generar nuestro modelo. Por lo tanto, es esencial utilizar herramientas de software que simplifiquen estas tareas. A continuación, se describen de manera concisa las herramientas elegidas y algunas de sus capacidades destacadas:

3.4.1 Sistema Operativo LINUX

El sistema operativo elegido para trabajar ha sido LINUX, siendo este de código abierto y conocido por su gran estabilidad, seguridad y flexibilidad. Es ampliamente utilizado por programadores debido a su gran capacidad para trabajar libremente con diversas librerías y entornos de trabajo optimizados. También soporta una gran variedad de lenguajes de programación y herramientas de desarrollo respaldadas a su vez por la gran comunidad activa que le proporciona constantemente recursos y soluciones a problemas técnicos que pudieran presentarse [19].

3.4.2 Label-Studio

Para etiquetar cada fotograma de los vídeos capturados, se ha decidido utilizar *Label-Studio*. Esta herramienta permite la generación de anotaciones precisas y escalables, exportables luego como *datasets* en diversos formatos compatibles con nuestras necesidades de trabajo, siendo YOLO de especial interés en este contexto. Cuenta con una interfaz intuitiva y personalizable como se aprecia en la Figura 3-4, y permite personalizar el código para emplear las etiquetas específicas requeridas [59].

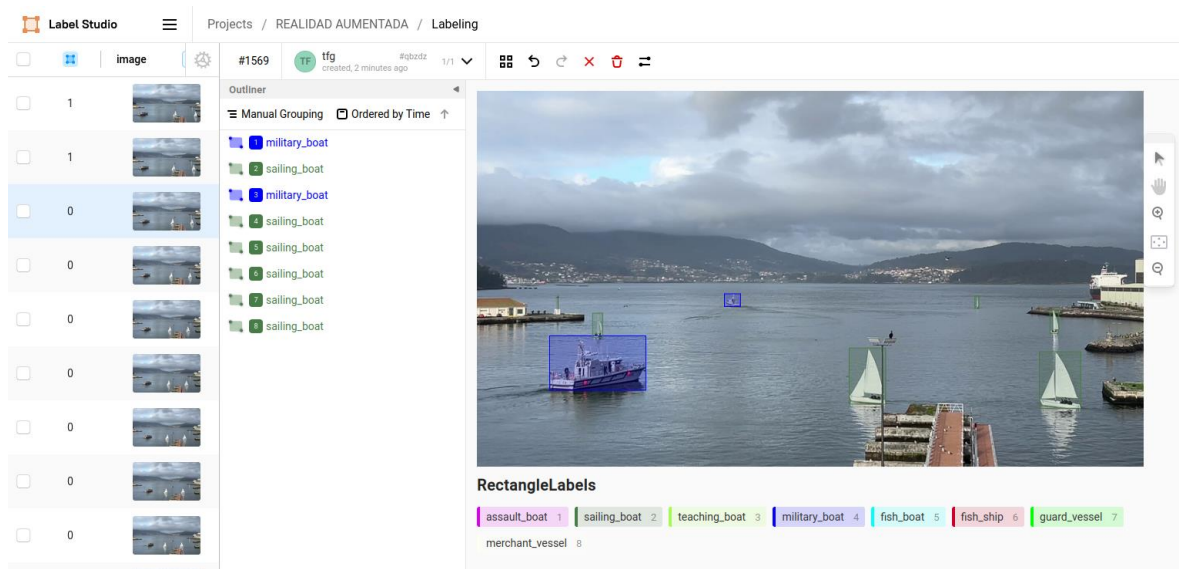


Figura 3-4 Interfaz de etiquetado de tareas en *Label-Studio* [imagen importada de *Label Studio* de autoría propia]

3.4.3 Ultralytics HUB.

Se trata de una herramienta integral de análisis de datos y entrenamiento de modelos basados en el algoritmo YOLO, que permite importar *datasets* previamente etiquetados y entrenarlos en los sistemas de altas capacidades de *Google Colab*. Como resultado se obtienen diferentes gráficas para analizar el

comportamiento de los parámetros del entrenamiento del modelo. Es importante destacar que brinda la capacidad de entrenar la detección de objetos en YOLOv5 y YOLOv8 (durante el periodo de desarrollo de este TFG se publicó el algoritmo YOLOv9, pero aún no está integrado en esta plataforma) y sus derivados como se muestra en la Figura 3-5, ya que el desarrollador del software es el mismo responsable de la creación de este algoritmo. Al igual que *Label-Studio*, su interfaz es notablemente intuitiva, lo cual simplifica el proceso de entrenamiento del modelo y visualización y análisis de los resultados de entrenamiento y validación del modelo obtenido. En esta plataforma se encuentran los modelos con los cuales es posible realizar entrenamiento mediante este software, tal y como se muestra en la Figura 3-5.

3.4.4 VS Codium

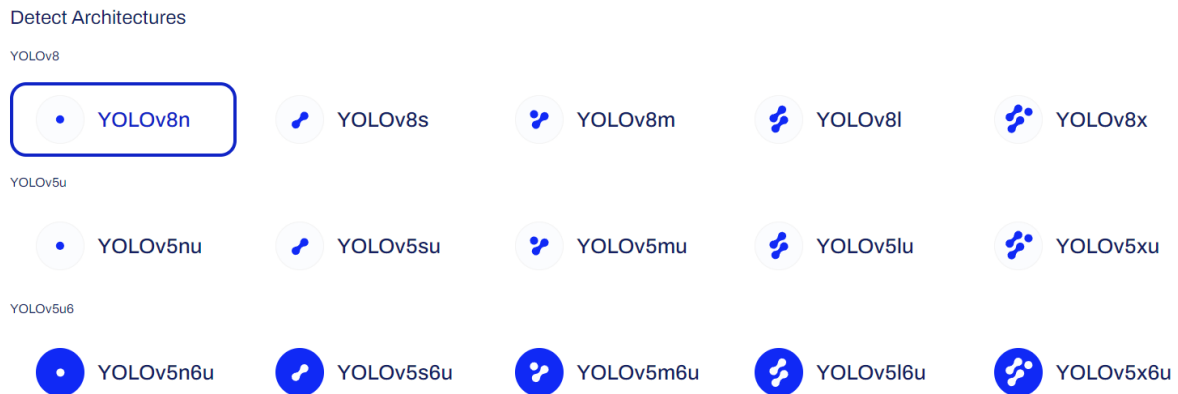


Figura 3-5 Modelos con los que permite entrenar la herramienta software *Ultralytics HUB* [imagen importada de *Ultralytics*]

VS Codium [60] es una distribución binaria de licencia gratuita, impulsada por la comunidad, del editor *VS Code* de *Microsoft* [61], lo que facilita su instalación en cualquier sistema operativo. Su interfaz intuitiva permite al usuario programar con una clara distinción de colores en el código para resaltar la sintaxis, característica estándar en los editores de código modernos que ayuda a mejorar la legibilidad al asignar diferentes colores y estilos a diferentes elementos del lenguaje de programación.

3.4.5 Entorno de trabajo VIDEO-IA-CUD

Para poder cumplir con los objetivos del proyecto ha sido necesario el empleo de una serie de *scripts* en código *Python* que se encargan de realizar diversas tareas necesarias, como la captura sincronizada de vídeos, la extracción de imágenes para crear el *dataset*, el entrenamiento del modelo basado en YOLO, etc. Esta colección de *scripts* pertenece al entorno VIDEO-IA-CUD, que está siendo desarrollado en el CUD-ENM para brindar soporte a proyectos similares relacionados con la aplicación de la visión artificial al control del tráfico marítimo. Entre los *scripts* utilizados se encuentran los siguientes:

- ***test_camera***: Se utiliza antes de tomar cualquier vídeo, para comprobar si la cámara asignada para la tarea tiene los parámetros correctos y la entrada *USB* es la indicada.
- ***video_capture_tool***: Una vez que la cámara está correctamente ajustada, este *script* les asigna el nombre deseado a los vídeos y permite elegir la ubicación de almacenamiento, además de establecer la duración de la grabación, los fotogramas por segundo (FPS) y la resolución en píxeles.
- ***video_extract_tool***: A partir de un vídeo previamente capturado y almacenado en un fichero, se extraen los *frames* que componen el vídeo y aplicando un modelo entrenado (basado en YOLO), genera las etiquetas de los objetos localizados para su posterior validación en la herramienta de etiquetado.
- ***model_train_tool***: Este *script* es empleado para entrenar un modelo a partir de un *dataset* de imágenes y las etiquetas asignadas a ellas.

- **video_track_tool:** Este *script* se emplea para verificar que el modelo entrenado funciona correctamente y sea capaz de identificar a las embarcaciones de un vídeo seleccionado, mostrando las cajas detectadas en los vídeos obtenidos.
- **video_track3d_tool:** Este script se emplea para visualizar y validar los resultados del interfaz de RA mostrando los datos de posicionamiento y cinemática de las embarcaciones en 3 dimensiones a partir de los vídeos obtenidos.
- **ais_capture_tool:** Se trata de una utilidad creada para poder leer los datos del receptor AIS. A partir de los datos obtenidos del receptor AIS, muestra por pantalla todos los mensajes recibidos en tiempo real.

3.4.6 Google Colaboratory

Se trata de un software gratuito que ofrece una plataforma basada en la nube, donde los usuarios pueden escribir y ejecutar código *Python* dentro de un entorno *Jupyter Notebook* sin la necesidad de configurar una máquina individual. La plataforma también cuenta con la posibilidad de usar GPU y TPU de alto rendimiento, por lo que se ha utilizado en las etapas finales de entrenamiento de los modelos del trabajo [62].

3.5 Descripción de tareas

Tras la introducción del desarrollo del proyecto, explicar la fase de formación inicial, y presentar de los distintos componentes, como *hardware*, software y entornos de trabajo, es necesario destacar cuales son principales desafíos encontrados para alcanzar los objetivos de este TFG. Si bien algunos de estos obstáculos ya contaban con una línea de acción bastante clara puesto que se contaban con *scripts* previamente desarrollados, otros, como los relacionados con los datos AIS y la creación del interfaz de RA son completamente nuevos.

- **Detección y clasificación de objetos en imágenes:** Este proceso implica entrenar un modelo con la capacidad de detectar objetos con precisión, centrándose específicamente en identificar embarcaciones en entornos marítimos.
- **Seguimiento de objetos en vídeo:** Después de detectar las diferentes embarcaciones, el modelo debe ser capaz de hacer su seguimiento, conocido como *tracking*, algoritmo que será explicando posteriormente. El modelo debe ser capaz de correlacionar cada detección y mantener el seguimiento del objeto durante su aparición completa en un vídeo.
- **Recepción de mensajes AIS:** A través del receptor, poder decodificar los mensajes recibidos en tiempo real para identificar su tipo, y que información relevante del buque nos ofrece.
- **Elaboración de una interfaz de RA:** Crear una ventana gráfica simple y atractiva para que el usuario que utilice el sistema de RA pueda visualizar información relevante de las embarcaciones en tiempo real que dispone a simple vista.
- **Integración de datos AIS con frames de vídeo:** Para facilitar la identificación precisa y la correlación de datos analizados por el usuario responsable de la vigilancia marítima, es necesario sincronizar el tiempo de recolección datos, tanto AIS como del modelo entrenado. En este trabajo se emplearán datos AIS que contienen la posición en coordenadas GPS de las embarcaciones para validar la precisión en la visualización de su posición.

3.6 Justificación de Algoritmos

A continuación, se exponen cuáles son los algoritmos utilizados durante el desarrollo del trabajo para resolver los problemas anteriormente mencionados.

3.6.1 Algoritmo para la detección de objetos.

En el estado del arte, se realiza una breve revisión de las capacidades del algoritmo YOLO, destacando en primer lugar, su capacidad de procesamiento *one-stage* para las imágenes de entrada, que no requiere del uso de fases de preprocesado de las imágenes y las etiquetas empleadas durante el entrenamiento, para lo cual emplea la detección de áreas de interés con la posible futura ubicación de los objetos. Además, su estructura de capas convolucionales facilita la capacidad de utilizar redes previamente entrenadas en tareas similares lo que permite extrapolar los modelos a diferentes tareas de manera sencilla, mejorando así el proceso de creación de soluciones de detección de objetos en imágenes.

De todas las versiones existentes, en el TFG se ha trabajado con la versión YOLOv8, la versión más reciente disponible a principios del año 2023. Se eligió esta versión por su capacidad para funcionar en una amplia gama de dispositivos CPU y GPU. Su compatibilidad con *Ultralytic HUB* fue el motivo principal para seccionar este modelo, ya que permite generar gráficas que permiten una fácil interpretación de los resultados.

3.6.2 Algoritmo para el seguimiento de objetos.

Tras exponer el algoritmo utilizado para la primera tarea de reconocimiento de objetos en imágenes, se presentan los algoritmos disponibles para realizar seguimiento de objetos en vídeo. Ambos están integrados en el *framework* de YOLOv8 y son interoperables entre sí, ofreciendo un alto grado de confianza en detección de objetos. Los algoritmos son:

- **ByteTrack:** El algoritmo *ByteTrack* aborda el desafío del seguimiento simultáneo de múltiples objetos en secuencias de vídeo mediante la asignación de identidades únicas a cada objeto detectado. Funciona mediante la detección de objetos en cada *frame*, seguida de la correlación de las *bounding boxes* de cada objeto, facilitando así la asociación de una identidad a lo largo de la secuencia. Cabe destacar que *ByteTrack* emplea específicamente el *IoU Tracker* para implementar este proceso. Sin embargo, uno de los problemas que presenta el algoritmo es la pérdida de *bounding boxes* cuando se produce una oclusión, es decir, cuando dos objetos se superponen. Para abordar este problema el algoritmo asocia un nivel de confianza alto a las detecciones de los *frames* anteriores, lo que permite mantener la continuidad en el seguimiento de los objetos detectados [63].
- **BoT-SORT:** Se trata de otro algoritmo empleado en la tarea de *Multi-Tracking*, que integra tanto *IoU* como *ReID* (*Re-identification tracker*). Este último, se emplea para la identificación de objetos mediante la estimación de su posición en el futuro, en contraposición al enfoque de *ByteTrack*, que compara los *IoU* con otras detecciones previas con un nivel de confianza alto. Gracias al elemento denominado *Camera Motion Compensation*, que compensa el movimiento de la cara junto con el *ReID* mencionado anteriormente y componentes del *IoU Tracker*, se logra el objetivo deseado [64].

3.7 Primera tarea: Conocer los mensajes AIS y sus tipos

Para el desarrollo de una interfaz de vigilancia marítima en puerto, es necesario conocer previamente los distintos tipos de mensajes AIS disponibles. Además del receptor AIS abordado y utilizado en el presente trabajo, existen también los transpondedores, los cuales no solo reciben los datos de otras embarcaciones y estaciones cercanas, sino que también transmiten información propia. Es esencial familiarizarse con dichos datos, ya que proporcionan información sobre algunas de las embarcaciones que serán objeto de grabación, lo que ayuda a validar los resultados finales.

Existen diferentes categorías de transceptores disponibles como se aprecia en la Figura 3-6, adaptados según el contexto de aplicación. Dichas categorías encuentran su principal aplicación en dos campos distintos: la navegación marítima de carácter profesional, encuadrada en el grupo A, y la navegación recreativa, que se divide en los grupos B y B+ [65].

Considerando que los sistemas operan en la banda de frecuencia VHF (concretamente en 161.975 MHz y 162.025 MHz), están sujetos a las limitaciones propias de la comunicación por radio en esta frecuencia. El alcance de la señal se ve limitado por la altura de la antena, siendo aproximadamente de 20 millas náuticas entre barcos en condiciones óptimas, y de 30 a 60 millas náuticas para estaciones costeras [66]. Además, cabe destacar que este alcance puede verse afectado por diversos factores como la interferencia electromagnética, las condiciones atmosféricas y la topografía del terreno, en caso de encontrarse cerca de la costa.

Un transpondedor debe tener una posición GPS, mientras que un receptor no tiene por qué tenerla ya que no transmite su posición.



Figura 3-6 Ejemplos de tipos de transpondedores y receptores [65]

Para la interpretación de los datos y su enlace con los vídeos capturados, se requiere únicamente la adquisición del primer dispositivo mencionado anteriormente. Adicionalmente es fundamental disponer de una pantalla (en este caso un ordenador) para llevar a cabo un análisis efectivo y aprovechar al máximo estos datos. Mediante una decodificación con el código adecuado es posible interpretar y obtener los diferentes tipos de mensajes AIS.

En la transmisión y recepción de datos AIS, existen tres categorías que son enviadas de manera continua por cada unidad a través de los canales de VHF mencionados anteriormente. Para la elaboración de la interfaz virtual, se resaltarán en negrita aquellos datos de mayor relevancia:

- Datos Dinámicos:
 - **Posición del buque (LAT, LON).**
 - **Velocidad sobre fondo (SOG).**
 - **Rumbo sobre fondo (COG).**
 - Hora de la posición del barco en segundos (UTC).
 - **Estado de navegación:** a motor, a vela, anclado, fondeado, sin poder maniobrar, etc.
 - **Dirección del rumbo:** equivalente al *heading* en inglés, se refiere a la orientación de la proa de la embarcación en una dirección determinada.
 - Tasa de cambio de rumbo (ROT).
- Datos estáticos:
 - Nombre del barco.
 - Distintivo de llamada.
 - **Número MMSI e indicativo internacional de radio:** es único para cada barco.
 - Número OMI.

- Tipo de embarcación: recreo, remolcador, carguero, petrolero, de pasajeros, de salvamento SA.
- Dimensiones del barco.
- Datos del viaje/itinerario:
 - Calado estático máximo actual en decímetros
 - Puerto de destino (UN/LOCODE)
 - Hora de llegada prevista (ETA)
 - Si procede, información sobre la categoría de la carga (clase de mercancía peligrosa).

Condiciones Dinámicas	Clase A	Clase B	Clase B+
El barco está fondeado o amarrado	3 minutos	3 minutos	3 minutos
SOG 0-2 nudos	10 segundos	3 minutos	3 minutos
SOG 2-14 nudos	10 segundos	30 segundos	30 segundos
SOG 2-14 nudos y cambio de rumbo	3.3 segundos	30 segundos	30 segundos
SOG 14-23 nudos	6 segundos	15 segundos	30 segundos
SOG 14-23 nudos y cambio de rumbo	2 segundos	15 segundos	30 segundos
SOG > 23 nudos	2 segundos	5 segundos	30 segundos
Información Estática	6 minutos	6 minutos	6 minutos

Tabla 2 Tasas de actualización en función de la velocidad sobre fondo [65]

Es crucial comprender la frecuencia con la que las embarcaciones transmiten sus datos, ya que esta tasa de actualización varía según el tipo de buque o el fabricante del transpondedor. En la Tabla 2 Tasas de actualización en función de la velocidad sobre fondo se muestran los intervalos de emisión estándar para un determinado tipo de embarcaciones como ejemplo.

Al igual que ocurre con la Tabla 2, en la Tabla 3 se presenta un criterio para determinar qué tipo de datos transmite cada tipo de transpondedor AIS. Es importante destacar que los datos transmitidos para el tipo B y B+, son los mismos.

Datos transmitidos	Clase A	Clase B y B+
MMSI + Nombre de la embarcación + Indicativo de llamada	SI	SI
Posición + COG + SOG	SI	SI
Verdadero heading	SI	SI
Ritmo de giro	SI	NO
Estado de Navegación	SI	NO
Numero de la OMI	SI	NO
Tipo de nave	SI	SI

Dimensiones de la nave	SI	SI
ETA + Destino + Calado	SI	NO

Tabla 3: Tipos de datos en función de clase de transpondedor A o B/B+ [65]

En la actualidad, existen 26 tipos de mensajes AIS, algunos de los cuales son específicos del transpondedor de clase A o B. Esto explica el hecho de que, en la Escuela Naval militar, se reciban mensajes de ambos tipos, desde embarcaciones de recreo en las proximidades, hasta los mercantes que se encuentran en el puerto de Marín. En la Figura 3-7, se muestra una captura de pantalla del *script ais_capture_tool* ejecutándose, leyendo datos AIS.

```
ais data msg= MessageType21(msg_type=21, repeat=1, mmsi=992246004, aid_type=<NavAid.SAFE WATER: 29>, name='MORRAZAN MIDCHANNEL', accuracy=True, lon=-8.783332, lat=42.366667, to_bow=0, to_stern=0, to_port=0, to_starboard=0, epfd=<EpfdType.Surveyed: 7>, second=0, off_position=False, reserved_1=224, raim=True, virtual_aid=True, assigned=True, spare_1=b'\x00', name_ext=None)
-----
ais data msg= MessageType1(msg_type=1, repeat=0, mmsi=224872340, status=<NavigationStatus.UnderWayUsingEngine: 0>, turn=<TurnRate.NO_TI_DEFAULT: -128.0>, speed=0.0, accuracy=True, lon=-8.70619, lat=42.396533, course=360.0, heading=511, second=37, maneuver=<ManeuverIndicator.NotAvailable: 0>, spare_1=b'\x00', raim=True, radio=34188)
-----
ais data msg= MessageType1(msg_type=1, repeat=0, mmsi=225997218, status=<NavigationStatus.UnderWaySailing: 8>, turn=-16.0, speed=0.1, accuracy=False, lon=-8.706265, lat=42.397178, course=168.8, heading=511, second=37, maneuver=<ManeuverIndicator.NotAvailable: 0>, spare_1=b'\x00', raim=False, radio=49167)
-----
ais data msg= MessageType18(msg_type=18, repeat=0, mmsi=224048250, reserved_1=0, speed=0.2, accuracy=False, lon=-8.706252, lat=42.397588, course=336.3, heading=511, second=38, reserved_2=0, cs=True, display=False, dsc=True, band=True, msg22=True, assigned=False, raim=True, radio=917510)
-----
ais data msg= MessageType4(msg_type=4, repeat=0, mmsi=2242120, year=2024, month=3, day=4, hour=8, minute=56, second=39, accuracy=True, lon=-8.690327, lat=42.40046, epfd=<EpfdType.GPS: 1>, spare_1=b'\x00\x00', raim=True, radio=2250)
-----
ais data msg= MessageType1(msg_type=1, repeat=0, mmsi=224430000, status=<NavigationStatus.UnderWayUsingEngine: 0>, turn=0.0, speed=0.0, accuracy=False, lon=-8.695467, lat=42.3976, course=351.0, heading=54, second=40, maneuver=<ManeuverIndicator.NotAvailable: 0>, spare_1=b'\x00', raim=False, radio=49207)
-----
ais data msg= MessageType1(msg_type=1, repeat=0, mmsi=224348220, status=<NavigationStatus.Undefined: 15>, turn=0.0, speed=0.0, accuracy=True, lon=-8.695658, lat=42.395073, course=360.0, heading=149, second=39, maneuver=<ManeuverIndicator.NotAvailable: 0>, spare_1=b'\x00', raim=True, radio=2231)
-----
ais data msg= MessageType20(msg_type=20, repeat=0, mmsi=2242120, spare_1=b'\x00', offset1=1208, number1=1, timeout1=0, increment1=250, offset2=2250, number2=1, timeout2=0, increment2=0, offset3=2180, number3=4, timeout3=0, increment3=1125, offset4=0, number4=None, timeout4=None, increment4=None)
-----
ais data msg= MessageType21(msg_type=21, repeat=1, mmsi=992246002, aid_type=<NavAid.SAFE WATER: 29>, name='RECALADA VIGO N/MARI', accuracy=True, lon=-8.947452, lat=42.297808, to_bow=0, to_stern=0, to_port=0, to_starboard=0, epfd=<EpfdType.Surveyed: 7>, second=0, off_position=False, reserved_1=224, raim=True, virtual_aid=True, assigned=True, spare_1=b'\x00', name_ext='N ENTR')
-----
ais data msg= MessageType4(msg_type=4, repeat=0, mmsi=2242120, year=2024, month=3, day=4, hour=8, minute=56, second=42, accuracy=True, lon=-8.690327, lat=42.40046, epfd=<EpfdType.GPS: 1>, spare_1=b'\x00\x00', raim=True, radio=2250)
-----
ais data msg= MessageType1(msg_type=1, repeat=0, mmsi=228378700, status=<NavigationStatus.UnderWayUsingEngine: 0>, turn=<TurnRate.NO_TI_DEFAULT: -128.0>, speed=0.0, accuracy=True, lon=-8.69084, lat=42.396918, course=212.0, heading=511, second=42, maneuver=<ManeuverIndicator.NotAvailable: 0>, spare_1=b'\x00', raim=True, radio=114726)
-----
ais data msg= MessageType24PartB(msg_type=24, repeat=0, mmsi=224048250, partno=1, ship_type=36, vendorid='SRT', model=1, serial=853802, callsign='EBA0', to_bow=18, to_stern=10, to_port=4, to_starboard=3, spare_1=b'\x00')
```

Figura 3-7: Lectura de datos AIS a través del *SCRIPT ais_capture_tool* [autoría propia]

Mayoritariamente, los datos recopilados corresponden a los tipos 1, 4, 18, 21 y 24, siendo estos los más representativos según lo reflejado en la Tabla 4, acompañados de otros también frecuentes. Los mencionados anteriormente aparecen se resaltan negrita:

Tipo de mensaje	Nombre	Descripción
1	Informe de posición	Informe de posición programado; Equipo móvil de a bordo de clase A.
2	Informe de posición	Informe de posición programado asignado; Equipo móvil de a bordo de clase A.
3	Informe de posición	Informe de posición especial, respuesta al interrogatorio; Equipo móvil de a bordo de clase A.
4	Informe de la estación base	Posición, UTC, fecha y número de ranura actual de la estación base.
5	Datos estáticos y relacionados con el viaje.	Informe de datos estáticos programados y relacionados con el viaje del buque, equipo móvil a bordo de Clase A
18	Informe de posición del equipo estándar Clase B	Informe de posición estándar para equipos móviles de a bordo de barcos de Clase B que se utilizará en lugar de los Mensajes 1, 2, 3

21	Informe de ayudas a la navegación	Informe de posición y estado de las ayudas a la navegación
24	Informe de datos estáticos	Datos adicionales asignados a un MMSI Parte A: Nombre Parte B: Datos estáticos

Tabla 4: Tipos de Mensajes AIS más relevantes para el trabajo [67]

Para desarrollar la interfaz de RA, como se muestra en el prototipo digital inicial en la Figura 3-8, se consideran datos clave como las coordenadas, la velocidad, el número MMSI, y el rumbo. Es especialmente importante conocer el tiempo transcurrido desde la última transmisión AIS para determinar si una embarcación ha vuelto a emitir desde la última vez que lo hizo. Por lo tanto, los datos mencionados anteriormente son válidos para el presente estudio y serán empleados en este análisis.

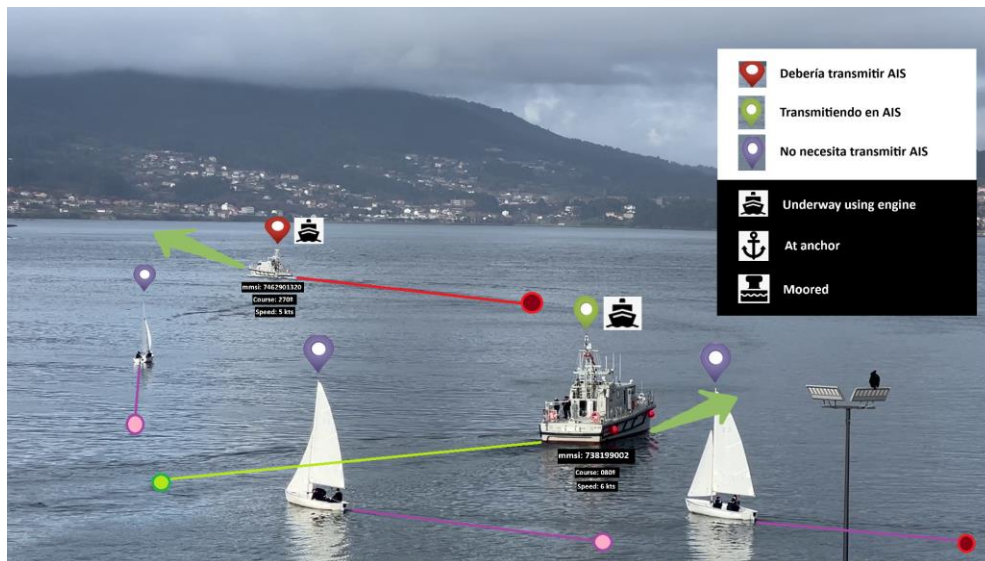


Figura 3-8 Primer prototipo digital de interfaz de RA con datos relevantes [autoría propia]

Al abordar la vigilancia marítima, y más concretamente en entornos portuarios, es fundamental tener en cuenta que únicamente algunos tipos de embarcaciones deben estar transmitiendo datos AIS. Esto se debe, en primer lugar, a los requisitos legales establecidos por la normativa de la Organización Marítima Internacional (IMO). Además, en caso de que estén navegando dentro de un puerto, los buques deben cumplir con la velocidad máxima permitida y mantenerse dentro de los límites autorizados para su navegación.

Cabe destacar que en caso de producirse una colisión entre dos embarcaciones debido al incorrecto cumplimiento del Reglamento Internacional para Prevenir Abordajes en la Mar (RIPAM) [68], un interfaz de vigilancia debería ser capaz de saber cuál de las dos ha infringido dicha normativa. Esto se lograría mediante el análisis de los datos cinemáticos proporcionados por el transpondedor AIS, en combinación con la trayectoria efectuada por cada buque.

3.7.1 Casos Particulares

Es importante considerar ciertos casos particulares que se reflejan en los transpondedores mediante emisiones específicas, es decir códigos particulares que indican situaciones relevantes para tener en cuenta. A continuación, se muestran los siguientes casos:

- **Heading:** este valor varía entre 0 y 359 grados, sin embargo, si devuelve el valor 511, indica que la información no se encuentra disponible.
- **COG:** Se espera que esté en el rango de 0 a 359 grados, pero una lectura de 511, indica que la información no se encuentra disponible.

- **SOG:** Este valor debe oscilar entre 0 y 102.2 nudos, pero si registra el valor 102.3, indica que la información no se encuentra disponible.
- **Número MMSI:** En ciertas instancias, el número MMSI puede devolver de manera extraordinaria el valor 7777777. Esto sugiere que puede tratarse de un nuevo transpondedor sin un número MMSI válido asignado, o en su defecto, tratarse de un error.

3.8 Segunda tarea: grabación de vídeo y entrenamiento modelo inicial

En cuanto al aspecto práctico del TFG, la segunda tarea consiste en encontrar una ubicación adecuada para la captura de vídeos, que permita obtener imágenes, con altura y distancia necesarias para simular la perspectiva de una cámara real de vigilancia marítima en puerto. Además, se detallarán algunas observaciones particulares surgidas durante la fase de etiquetado, y se proporcionarán análisis preliminares de los parámetros obtenidos tras el entrenamiento del modelo con *Ultralytics HUB*.

Posteriormente, se procede al entrenamiento local con modelos basados en YOLOv8 utilizando *scripts* de *Python*, junto con una biblioteca adecuada que permite cargar y entrenar los modelos para esta versión del algoritmo. Estos *scripts*, previamente mencionados, fueron extraídos del entorno VIDEO-IA-CUD y utilizados por el autor del presente TFG para llevar a cabo el proyecto.

Es relevante señalar que, para este modelo inicial, no se consideran datos AIS, ya que es prioritario abordar la problemática de la detección y seguimiento de objetos en vídeos.

3.8.1 Elección de la ubicación de grabación

La elección de un lugar idóneo para la captura de vídeos y la creación de un *dataset* de alta calidad es fundamental para poder garantizar el entrenamiento de un modelo en condiciones óptimas. Para lograr el cometido, se han evaluado diversas localizaciones dentro de la Escuela Naval Militar de Marín, Pontevedra, como la residencia de suboficiales, la terraza del edificio de investigación y el balcón de la segunda planta del edificio Isaac Peral. Tras considerar dichos espacios, la torre del patio de aulas cumplía con los requisitos necesarios, proporcionando una vista amplia del puerto, como se aprecia en la Figura 3-9 y Figura 3-10.

La implementación de esta tecnología brinda la capacidad de capturar simultáneamente varios tipos de embarcaciones, lo cual enriquecerá la diversidad de categorías del modelo, incluyendo buques que emiten señales AIS como aquellos que, según la normativa no están obligados a hacerlo. Además, se asegura la continuidad de la captura de vídeos, incluso con condiciones meteorológicas adversas.

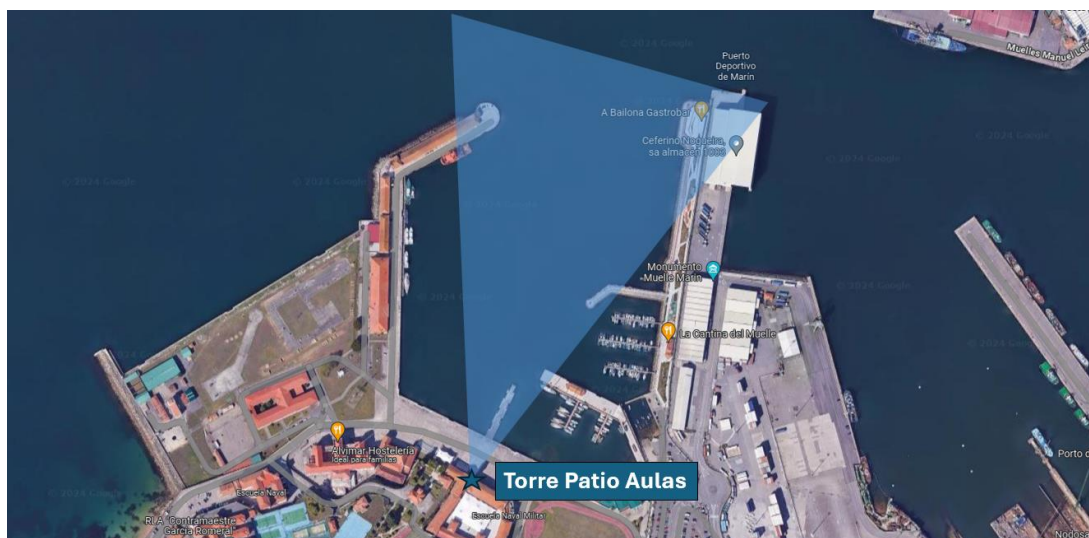


Figura 3-9 Captura satélite de *Google Maps* para la generación del *dataset* [autoría propia]



Figura 3-10 A la izquierda vista desde la ventana la Torre del Patio de Aulas, a la derecha con un círculo rojo la ventana [autoría propia]

3.8.2 Etiquetado de vídeos

Previamente a proceder con el entrenamiento del modelo, es necesario realizar una selección del conjunto de datos a ser utilizados. Se ha optado por el *dataset* denominado *MARIN_PORT*, que consta de 100 imágenes obtenidas de ocho vídeos diferentes capturados desde la torre, es decir, un total de 800 imágenes. La abundancia de *frames*, contribuirá a un entrenamiento del modelo significativamente más preciso y de mejor calidad, para lo cual es indispensable etiquetar cada embarcación de interés presente en cada imagen de este *dataset*.

Todos los buques han sido previamente etiquetados en *Label-Studio*, tal y como se ha indicado anteriormente, dando lugar a diversas imágenes junto con sus respectivas etiquetas, lo que constituye una tarea. En la Figura 3-11 se muestra un ejemplo de las embarcaciones habituales que se pueden encontrar en la dársena de la Escuela Naval Militar, que cuenta con las siguientes peculiaridades:

- En el entorno marítimo se pueden observar elementos propios de la infraestructura de un puerto, como luces, espigones y zona literal, como la isla de Tambo.
- La cámara se mantiene en una posición fija, sin variaciones en la orientación, profundidad o elevación.
- Las condiciones meteorológicas y de luminosidad pueden variar entre diferentes tomas de vídeo.
- Durante actividades de formación que tienen lugar por la tarde, tanto los veleros, como las lanchas de instrucción salen al mismo tiempo, lo que puede ocasionar una diferencia de 20 minutos entre la toma del primer y último vídeo.
- Dado que el puerto de Marín se encuentra muy próximo, es posible que pasen buques mercantes o se encuentren atracados en las proximidades del entorno de grabación.

En el proceso de etiquetado de clases para el entrenamiento del modelo, es crucial establecer un patrón coherente en los nombres, particularmente en relación con el tamaño de las embarcaciones. Por ejemplo, tratándose de un pesquero, se distingue entre aquellos de eslora³ pequeña (*fish_boat*), y los de mayor porte (*fish_ship*), e incluso se consideran categorías de mayores dimensiones como los buques mercantes (*merchant_vessel*). Como resultado la dificultad del entrenamiento del modelo es mayor, a pesar de ello, garantiza una mejor generalización y capacidad de discriminación en la clasificación de los buques detectados. Las etiquetas diseñadas siguen el siguiente patrón:

³ El término eslora es la longitud total de un barco, medida desde la proa hasta la popa, utilizada en la industria naval y la navegación para determinar el tamaño y capacidad de las embarcaciones.

- **assault_boat:** Embarcaciones motorizadas que utilizan los infantes de marina de la Escuela Naval Militar para realizar sus actividades de formación específicas, con una tripulación de 5 a 6 integrantes cada una.
- **sailing_boat:** Veleros de pequeño tamaño utilizados por los alumnos de cuerpo general para complementar su formación a vela, con una tripulación de 2 a 6 personas.
- **teaching_boat:** Embarcaciones similares los *assault_boat* empleadas por los instructores para supervisar las actividades de instrucción marinera, con 1 o 2 tripulantes.
- **military_boat:** Lanchas de instrucción que utilizan los alumnos de cuerpo general para practicar diversos tipos de maniobras en la mar, con una dotación de entre 15 a 20 miembros.
- **fish_boat/fish_ship:** Pesqueros comunes en la ría de Pontevedra que suelen pasar cerca mientras se dirigen a faenar, tripulados por 2 o 3 personas.
- **row_boat:** Pequeñas embarcaciones de remo con capacidad para 4 o 5 remeros.
- **guard_vessel:** Algunas embarcaciones de la Guardia Civil que prestan Servicio Móvil Marítimo tienen su puerto en la Escuela Naval Militar, que cuentan con una eslora considerable.
- **merchant_vessel:** Buques de gran eslora, de entre 150 a 300 metros, reconocibles por transportar grandes cantidades de mercancía.



Figura 3-11: Etiquetado primer modelo Escuela Naval Militar, donde se pueden apreciar *military_boat*, *sailing_boat*, *teaching_boat* y *row_boat* [imagen extraída de *Label-Studio* de autoría propia]

Previamente al análisis de los resultados, es fundamental tener en cuenta diversas consideraciones a cerca del proceso de creación del *dataset*. No sólo se limita al recuento de etiquetas o de la precisión a la hora de ajustar los cuadros que encuadran los buques en cada *frame*, sino que también otros aspectos inciden significativamente y limitan el nivel de éxito de nuestro modelo. Elementos como el brillo de la imagen, la resolución, el número de píxeles, el contraste y la superposición de buques que se cruzan, además de las variaciones de tamaño dentro de una misma clase, impidiendo la consecución de una situación idílicamente favorable.

3.8.3 Entrenamiento dataset *MARIN_PORT* en *VS CODIUM*

Para el primer entrenamiento, se utiliza un modelo en YOLOv8 que ha sido entrenado previamente con el *dataset* COCO, el cual fue discutido previamente en el segundo capítulo del presente TFG, el cual contiene numerosas categorías relativas a miles de imágenes que están integradas en el modelo YOLOv8. También cabe destacar que se requiere de un archivo *YAML* para definir la configuración del conjunto de datos, el cual integra información acerca del número de clases presentes.

Al proceder con el entrenamiento, se ejecuta el *script model_train_tool*, en el cual se especifica el nombre del proyecto, el modelo a utilizar en el entrenamiento (YOLOv8) y la posible modificación de otros parámetros que puedan influir en el proceso, como se aprecia en la Figura 3-12.

En primer lugar, es importante indicar que el término *epoch*, que en ML representa un ciclo completo de datos de entrenamiento. A medida que aumenta el número de *epochs*, se requiere un mayor tiempo de cálculos en el ordenador de forma que el tiempo necesario para el entrenamiento aumenta de manera directamente proporcional. Asimismo, también es importante destacar el parámetro *patience*, el cual establece que, si el modelo no mejora su aprendizaje durante el número preestablecido de *epochs*, el entrenamiento se detendrá automáticamente y se guardarán sus datos.

```
PROJECT_NAME = "MARIN_PORT" # Name of the project
MODEL_NAME   = 'yolov8n.pt' # Name of the model used to create annotations
TRAIN_DEVICE = 'cpu'
TRAIN_EPOCHS = 1000
TRAIN_PACIENCE = 200
SAVE_FOLDER  = 'MARIN_PORT_YOLOV8N'
RUN_BEST_REPLACE = True # Set it best model replaces original source model
```

Figura 3-12 Fragmento del script empleado para el entrenamiento del modelo [autoría propia]

Tras el primer entrenamiento con YOLOv8, se identificó una detección insuficiente de las embarcaciones, lo que motivó a un reentrenamiento del modelo. Para ello, se emplearon los resultados del mejor modelo previamente obtenido, junto con el *dataset* de *MARIN_PORT* que consta de 800 tareas previamente etiquetadas en *Label-Studio*. La elaboración de estas tareas para el entrenamiento resultó en una mejora significativa en el rendimiento del modelo. La Figura 3-13 muestra una matriz que refleja el porcentaje de *precision* entre las asignaciones y la realidad, comparando los niveles de confusión en cada categoría proporcionada para el entrenamiento, así como en el *background* de las imágenes.

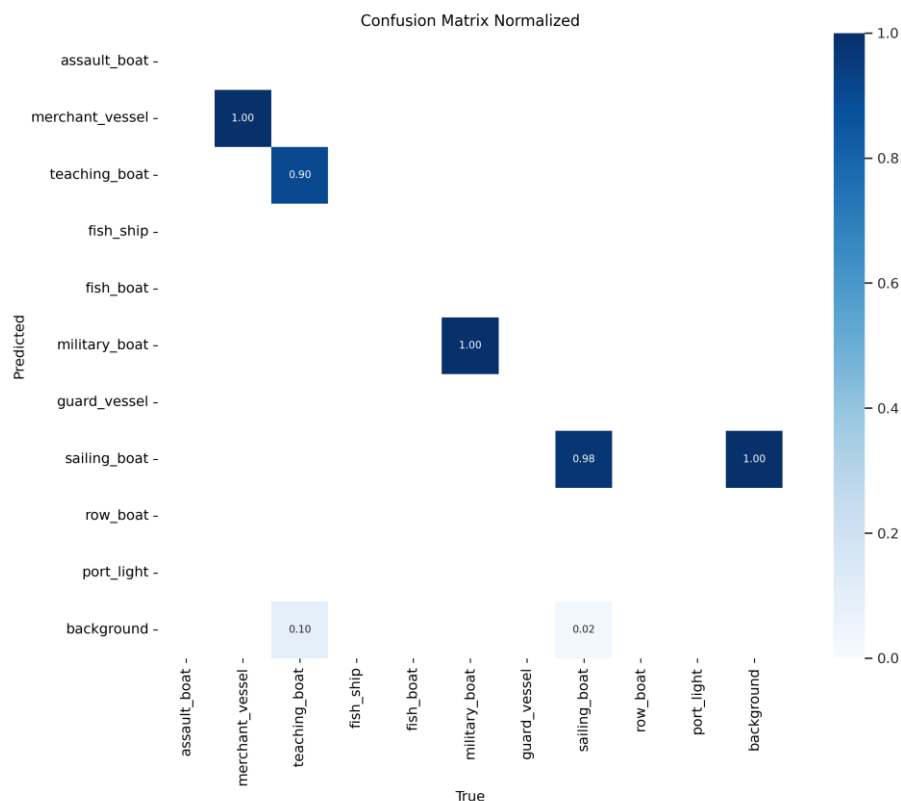


Figura 3-13 Matriz de confusión generada durante el entrenamiento del modelo *MARIN_PORT* [autoría propia]

Como parte del análisis paramétrico, el proceso entrenamiento no solo genera una matriz de confusión, sino también una serie de gráficas que facilitan la evaluación de los parámetros *precision* y *recall*. A continuación, se presentarán dichas gráficas de manera individual, para posteriormente relacionar ambos parámetros en un único gráfico.

La Figura 3-14 presenta la gráfica obtenida para el parámetro de medida *precision*. La precisión se refiere a la proporción de ejemplos clasificados correctamente como positivos (verdaderos positivos) entre todos los ejemplos clasificados como positivos, es decir, la proporción de verdaderos positivos respecto al total de instancias clasificadas como positivas. La precisión mide la exactitud de las predicciones positivas del modelo.

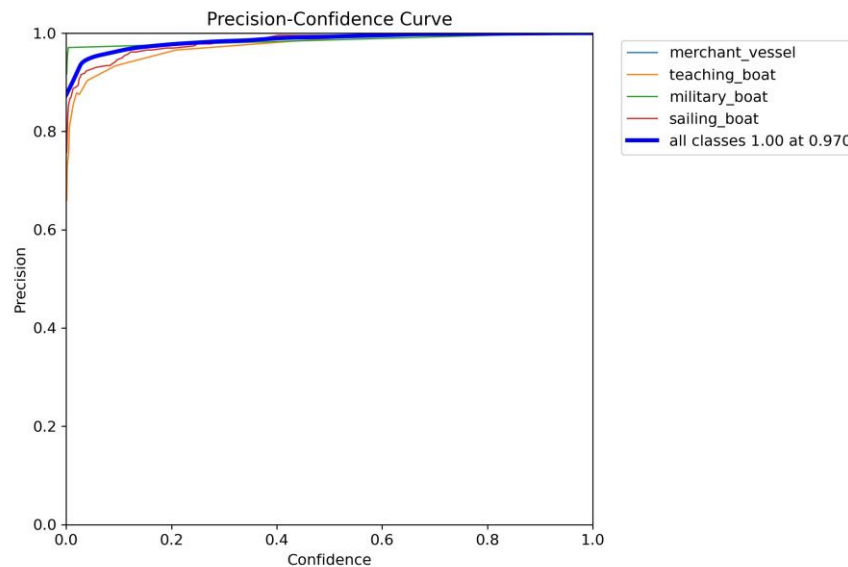


Figura 3-14 Gráfica que muestra *precision* para cada categoría de objeto detectado [autoría propia]

Asimismo, el parámetro de medida *recall* representa la proporción de ejemplos clasificados correctamente como positivos (verdaderos positivos) entre todos los ejemplos que realmente son positivos, es decir, la proporción de verdaderos positivos respecto al total de instancias positivas en el conjunto de datos. El *recall* mide la capacidad del modelo para encontrar todos los casos positivos en el conjunto de datos. La representación gráfica queda representada en la Figura 3-15.

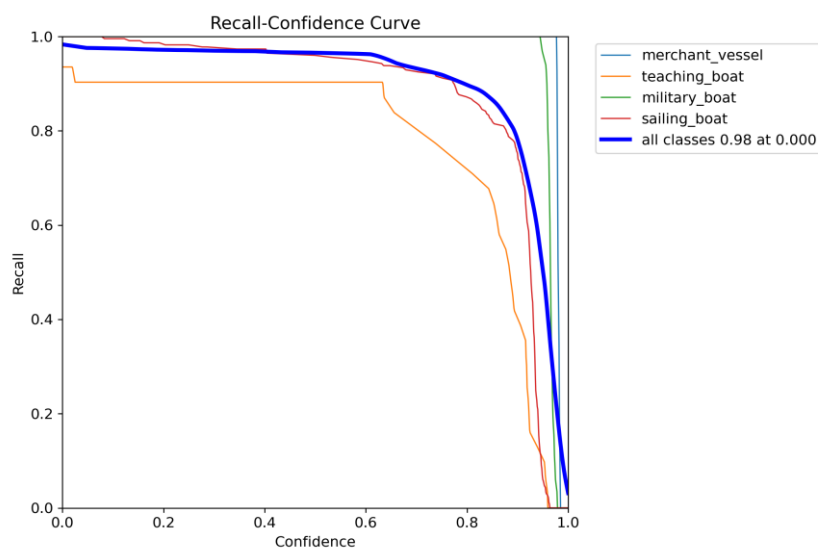


Figura 3-15 Gráfica que muestra *recall* para cada categoría de objeto detectado [autoría propia]

A partir del *script*, se genera como resultado una gráfica que correlaciona el *precision* y el *recall*, los mismos datos que se representan en la Figura 3-14 y Figura 3-15. Se podría esperar que la gráfica generada por un modelo perfecto tenga una forma similar a la de un cuadrado, dado que alcanzaría el valor de la unidad en ambos parámetros.

Sin embargo, lograr este nivel de perfección representa un desafío importante, ya que cualquier intento de optimizar los resultados de uno de los parámetros inevitablemente resultará en la disminución de los valores del otro. Con base en los resultados positivos observados en la Figura 3-16, se puede inferir que la curva efectivamente posee una forma cuadrada, corroborando lo mencionado anteriormente.

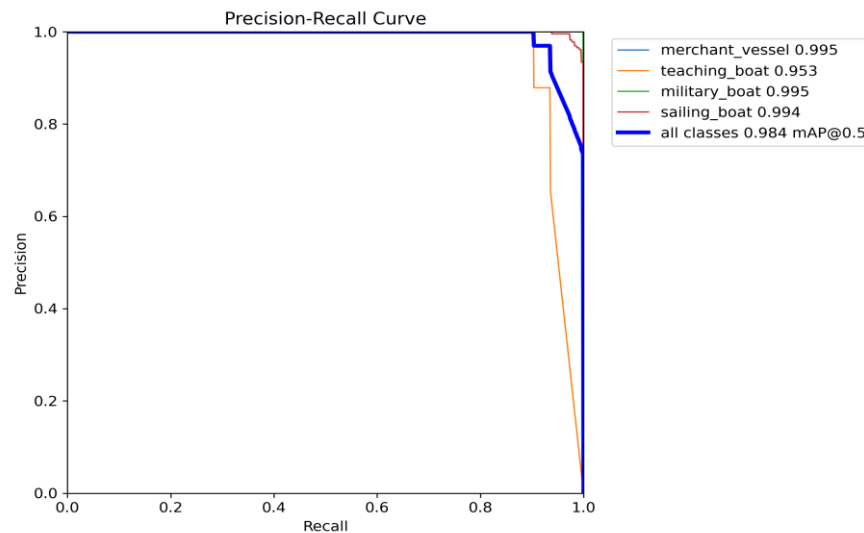


Figura 3-16 Gráfica que muestra la relación entre *recall* y *precision* [autoría propia]

Otro elemento de gran relevancia que emerge del proceso de entrenamiento, son imágenes de prueba, las cuales se utilizan para evaluar la capacidad es el modelo después de la fase *train*. En la Figura 3-17 se muestra un ejemplo de algunas de estas imágenes generadas por el *script*, capturadas en diversos entornos.



Figura 3-17 Imágenes generadas por el modelo YOLOv8 que ha sido entrenado [autoría propia]

3.8.4 Obtención de vídeos con seguimiento de objetos

Una vez concluida la etapa de entrenamiento, es fundamental llevar a cabo una evaluación para determinar si el modelo posee la capacidad de hacer seguimiento de objetos y abordar una de las tareas planteadas al comienzo del desarrollo del TFG. Este proceso de seguimiento se refiere a la capacidad del modelo para identificar un objeto y establecer una correlación coherente en los *frames* consecutivos del vídeo, garantizando que mantenga la misma identidad en todo momento, es decir, que estén correlacionados. Uno de los aspectos fundamentales de este tipo de soluciones radia en cómo evitar la pérdida del objeto en cuestión, ya que, si ocurriese, el modelo volvería a identificarlo como uno nuevo, en lugar del mismo. Es importante destacar que YOLOv8 ya cuenta con los algoritmos de *tracking* que facilitan la ejecución de esta tarea.

Para lograr el objetivo, se emplea el *script video_track_tool* en VS Codium generando como resultado un vídeo que muestra el seguimiento de las embarcaciones, distinguiendo entre las diferentes categorías como se aprecia en la Figura 3-18. A continuación, se adjunta uno de los vídeos obtenidos tras la ejecución del *script*. Es importante mencionar que el acceso a dicho vídeo podría implicar la descarga por parte del lector dependiendo de factores como el dispositivo, el sistema antivirus, el sistema operativo, etc. Se recomienda la descarga para evitar una pérdida notable en la imagen e interrupciones durante la reproducción debido a la conexión a internet.

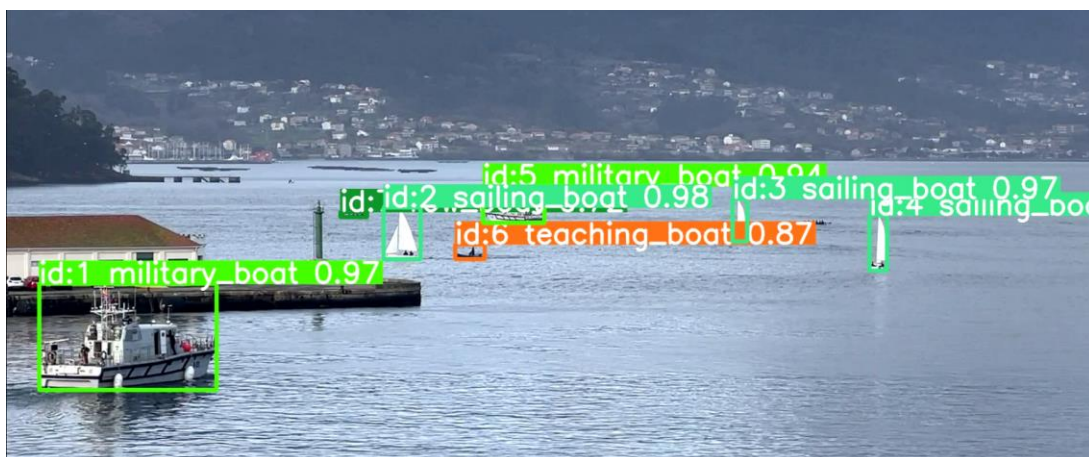


Figura 3-18 Captura de un *frame* de los vídeos obtenidos con *tracking* [autoría propia]

<https://drive.google.com/file/d/1bg2iaUq6dWvSHwpDggew7iIGlmAMmo9-/view?usp=sharing>

Es importante recalcar que, tras la obtención de los vídeos, se identifican ciertas irregularidades que serán examinadas en el capítulo de Resultados / Validación / Prueba. Estas anomalías deben considerarse al seleccionar nuevas imágenes para el *dataset*, con el fin de evitar repeticiones al reentrenar el modelo.

3.8.5 Entrenamiento *dataset MARIN_PORT* en *Ultralytics HUB*

Para reforzar el entrenamiento del *dataset*, es importante realizar un entrenamiento en *Ultralytics HUB* para garantizar que el modelo pueda identificar los objetos con mejor precisión.

En la fase de procesamiento del programa se utilizan modelos previamente entrenados existentes en la plataforma *Ultralytics-HUB* con el *dataset* COCO, siendo el algoritmo YOLOv8n escogido para este propósito. Dicho programa divide automáticamente el *dataset* en tres grupos, como se ilustra en la Figura 3-19. Cada grupo tiene un propósito único y desempeña una función particular para mejorar el proceso de aprendizaje de la red.

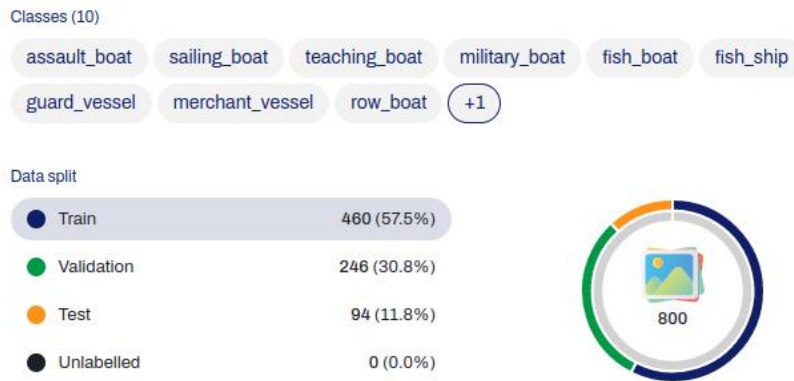


Figura 3-19 Distribución de las imágenes que componen el dataset "MARIN_PORT" [gráfica de Ultralytics HUB de autoría propia]

En primer lugar, el conjunto de datos denominado *train* contiene el mayor número de imágenes por lo que es elegido para el entrenamiento de la red. Es fundamental asegurarse de que esta selección abarque una variedad de fotogramas que representen diferentes aspectos del conjunto total de los datos, evitando que la red no se entrene únicamente con imágenes de momento específico del día. El objetivo es proporcionar a la red de una experiencia de aprendizaje variada al exponerla a una variedad de escenarios durante su fase de validación.

Por otro lado, el segundo grupo se constituye de imágenes destinadas a *validation*, las cuales, durante el entrenamiento de la red, son utilizadas para verificar su capacidad de generalizar los resultados a partir de nuevos datos que no formaron parte del conjunto *train*. Como resultado, este grupo desempeña un papel fundamental en la optimización del modelo durante su fase de aprendizaje. Además, aunque la red accede constantemente a este grupo, se asegura de no aprender nunca de él en ningún momento.

En tercer lugar, se reserva un pequeño porcentaje de las imágenes para la última fase, denominada *test*. Esta fase se caracteriza por no ser utilizada por la red durante el proceso de aprendizaje hasta la realización de la prueba final de validación, en la que la red se somete a una serie de pruebas para evaluar su fiabilidad y precisión en la entrega de resultados.

El programa genera una serie de gráficas que se presentarán a continuación. Haciendo referencia a la primera de ellas, como se muestra en la Figura 3-20, esta ilustra la relación existente entre los valores del *recall* y *precision*, como se observó en el entrenamiento local a través de *VS Codium*, junto con los resultados del *mAP*. Realizando un análisis superficial, se aprecia que el *mAP* para un *IoU* 50 es satisfactorio, pero disminuye notablemente para un *IoU* entre 50 y 95. Sin embargo, la regularidad en los valores de *precision* y *recall*, sugiere fiabilidad en el modelo para la detección y clasificación de objetos, tras el proceso de aprendizaje. Un aspecto importante considerar es el número de *epochs*, lo que indica que sería necesario aumentar su cantidad para obtener resultados más confiables.

Continuando el análisis de las gráficas proporcionadas por *Utralytics HUB* para evaluar el modelo obtenido, se dispone del *box loss* (traducido como pérdida de cajas en castellano), que devuelve un valor para cada entrada y permite observar como el modelo mejora a medida que procesa un mayor número de imágenes. Particularmente, en la Figura 3-21 se observa como el valor disminuye tanto para el conjunto de entrenamiento como para la validación. Además, cabe destacar que como se mencionó anteriormente, el modelo no tiene acceso a las imágenes reservadas para la prueba, por lo que no se reflejan en la Figura 3-21.

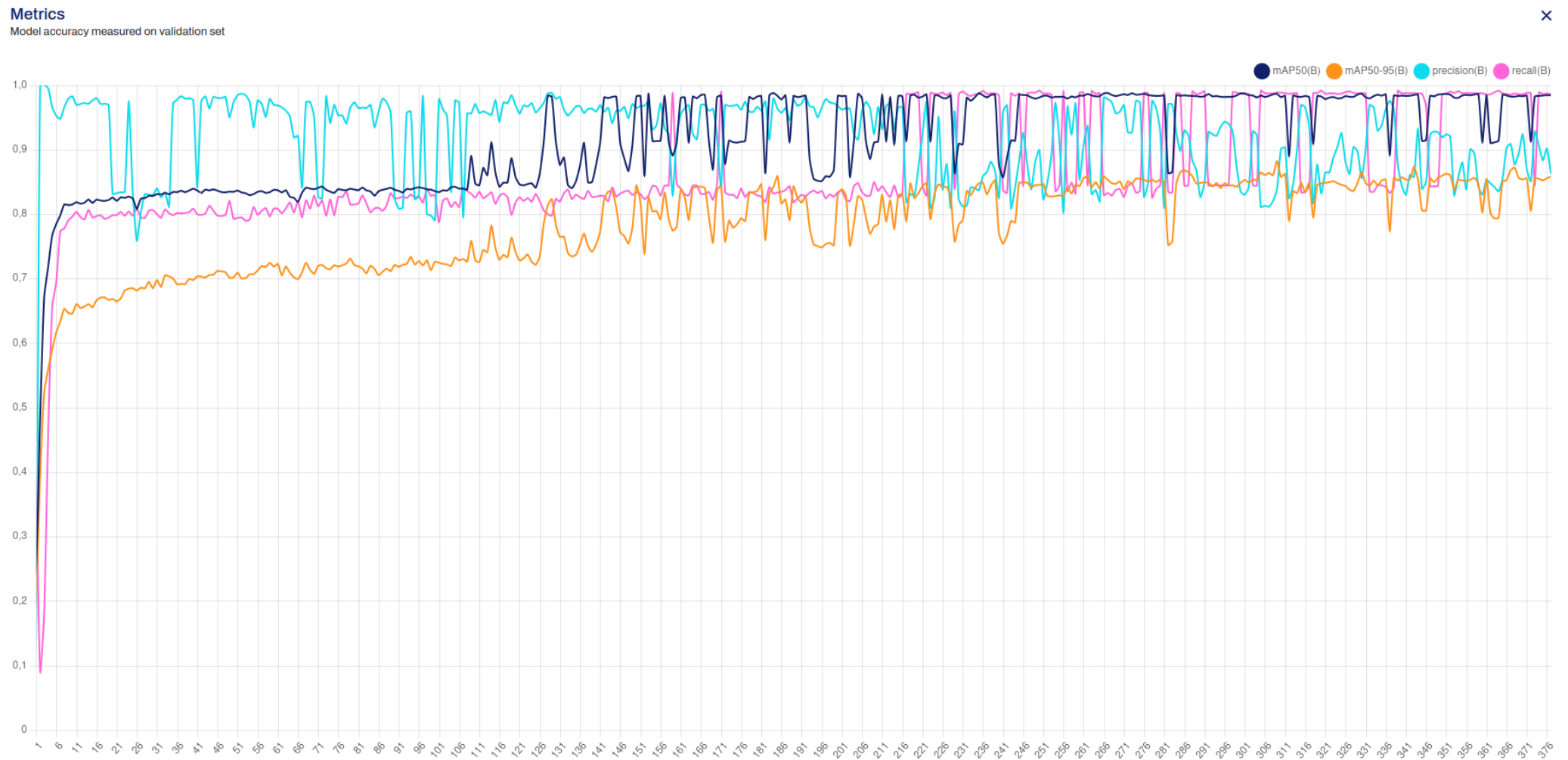


Figura 3-20 Representación visual proporcionada por Ultralytics HUB que muestra los resultados del mAP, recall y precision para cada uno de los epochs analizados [gráfica generada por Ultralytics HUB de autoría propia]

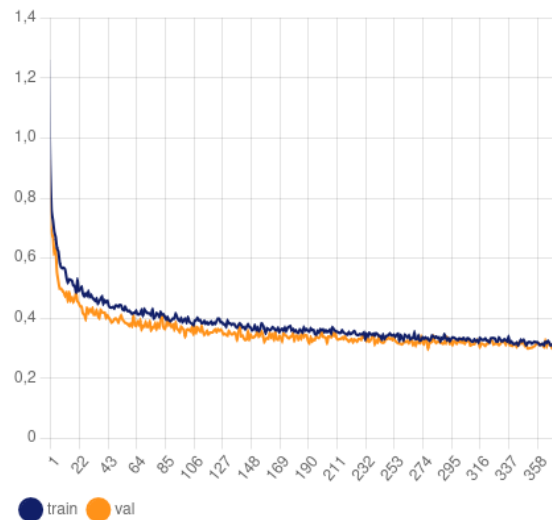


Figura 3-21 Gráfica del *box loss* para cada entrada evaluada [gráfica generada por *Ultralytics HUB* de autoría propia]

Se obtiene otra gráfica importante que muestra el parámetro conocido como *class loss* (pérdida de clase en castellano), representado en la Figura 3-22. Este cuantifica la capacidad del modelo para clasificar un objeto detectado correctamente. Al igual que en la gráfica anterior que mostraba el *box loss*, se observa una disminución en el número de errores conforme se incrementa el número de entradas.

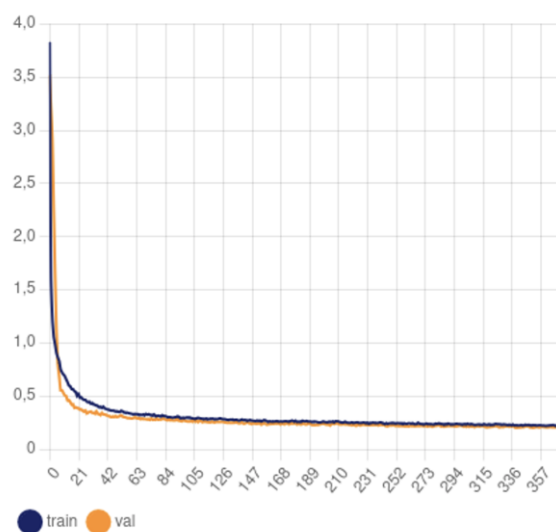


Figura 3-22 Gráfica del *class loss* para cada entrada evaluada [gráfica generada por *Ultralytics HUB* de autoría propia]

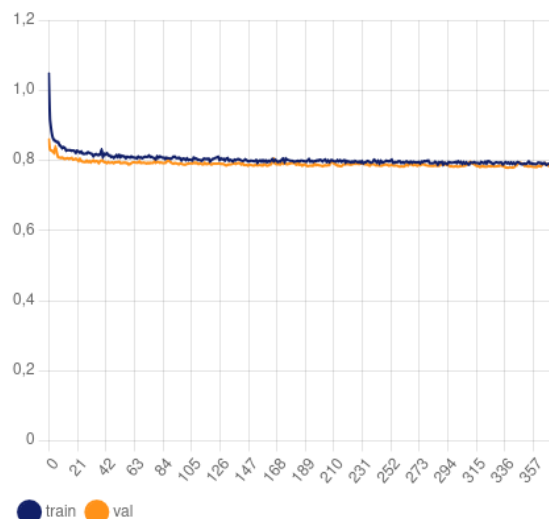


Figura 3-23 Gráfica del *object loss* para cada entrada evaluada [gráfica generada por *Ultralytics HUB*]

En último lugar, tras analizar *class loss* y *box loss*, se muestra en la Figura 3-23 la gráfica *object loss* refiriéndose a la cantidad de objetos que es capaz de detectar el modelo en un *frame*. Como se observa, al igual que el resto de las gráficas anteriores, este valor decrece con el número de épocas de entrenamiento ejecutadas. Además, en la Figura 3-24 se muestra una de las imágenes procesadas por el modelo durante la prueba *test*, en donde se muestran las detecciones que ha generado.

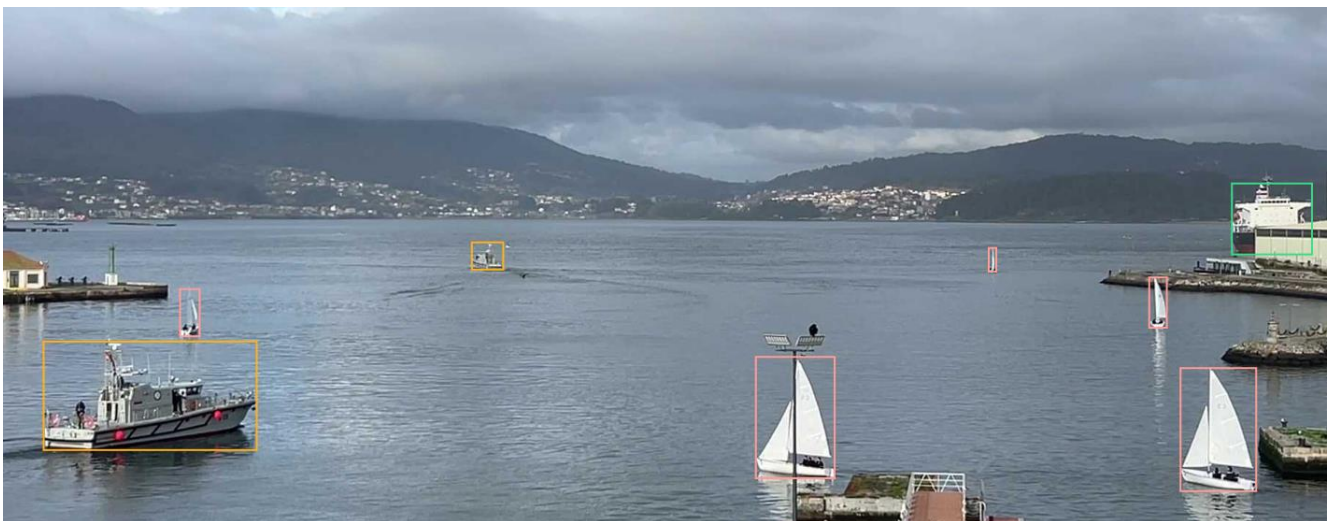


Figura 3-24 Imagen sometida a *test* tras entrenamiento del modelo [imagen generada por *Ultralytics HUB* de autoría propia]

3.9 Tercera tarea: integración de datos AIS y vídeo

En esta tercera fase, se procese al desarrollo necesario para poder integrar los vídeos capturados con los datos AIS recolectados simultáneamente. Posteriormente, se pretende llevar dicha información a la interfaz de RA con el fin de proporcionar la información necesaria al usuario.

En los siguientes apartados se expone el proceso de integración de datos utilizando dos cámaras, tal y como se requirió en el TFG de Don Víctor Recio Rico, el cual necesitaba realizar el *tracking* de embarcaciones y medir las distancias utilizando puntos ya conocidos en los vídeos. De esta manera, se verificará la coherencia entre las coordenadas GPS incluidas en los datos AIS capturados junto con los calculados empleando una solución de visión estereoscópica propuesta en el otro TFG. Es crucial asegurar la sincronización temporal entre ambos conjuntos de datos, es decir, entre los datos capturados

por dos cámaras de vídeo y por el receptor de datos AIS, ya que, de lo contrario, no sería posible integrar los *frames* del vídeo con los datos AIS de manera precisa.

3.9.1 Elección de ubicación de grabación.

Para lograr el objetivo de la grabación con dos cámaras, se requirió cambiar el lugar de filmación, debido a que en la ubicación inicial la distancia entre las cámaras no era adecuada para conseguir una diferencia de píxeles significativa. Debido a esto, se seleccionó una nueva ubicación en la Escalera Monumental como se aprecia en la Figura 3-25, para comprobar si se pudiera cumplir con el objetivo establecido.



Figura 3-25 A la izquierda vista desde la Escalera Monumental, a la derecha con un círculo rojo la escalera [autoría propia]

3.9.2 Entrenamiento del modelo

Una vez registrados los nuevos vídeos, se necesita entrenar un nuevo modelo actualizado para el entorno renovado. Si bien es verdad que el modelo anterior era capaz de detectar las embarcaciones correctamente, su incapacidad para adaptarse a las diferentes condiciones de iluminación plantea desafíos para lograr una detección precisa y libre de errores de las embarcaciones, como se destacó anteriormente en este capítulo a cerca de la generalización.

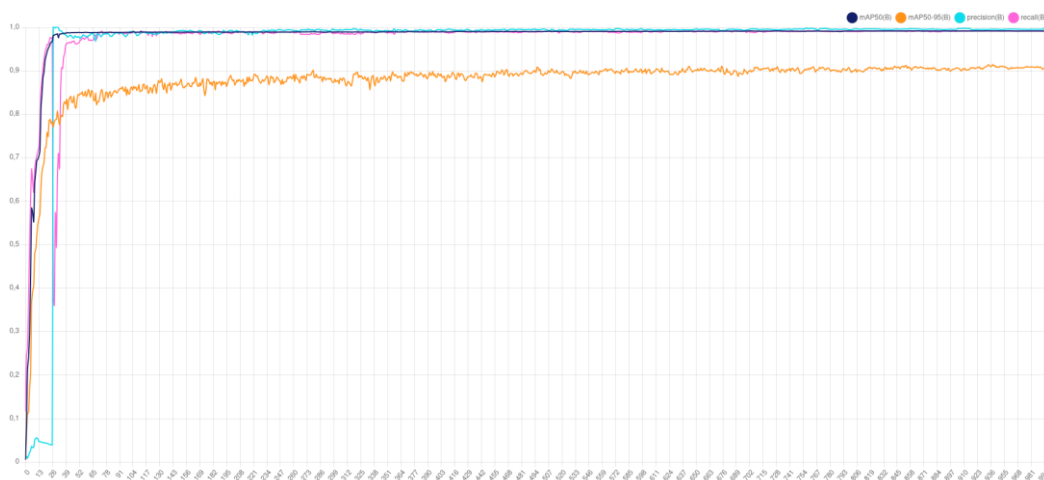


Figura 3-26 Representación visual proporcionada por *Ultralytics HUB* que muestra los resultados del *mAP*, *recall* y *precision* para cada uno de los *epochs* analizados [gráfica generada por *Ultralytics HUB* de autoría propia]

Con el fin de mejorar la eficiencia y reducir el tiempo necesario, se ha tomado la decisión de realizar un reentrenamiento del nuevo modelo en *Ultralytics HUB*, basado en el modelo base YOLOv8n y

utilizando el nuevo *dataset* bautizado como “ESCALERA MONUMENTAL”. Este abarca los *frames* que presentan una alta probabilidad de confusión, principalmente por la presencia de embarcaciones que entran o salen del entorno de grabación, así como la posible situación de cruce entre ellas. El progreso del proceso de entrenamiento se muestra en la Figura 3-26.

Se aprecia como el entrenamiento, tras someterse al proceso de aprendizaje, muestra fiabilidad en la coherencia de los valores de *precision* y *recall* que produce.

3.9.3 Información temporal

Para iniciar la sincronización de los vídeos, es necesario conocer el momento exacto de su grabación antes de comenzar el proceso de captura. Específicamente, se decide añadir la hora con precisión de segundos, para asociar un determinado *frame* (cuyo número también se muestra en el vídeo) con los datos AIS correspondientes que se adquirirán simultáneamente durante el proceso de captura de datos. Como se ha mencionado anteriormente en la tarea inicial, estos datos proporcionan información GPS en un instante del tiempo determinado de forma que pueda emplearse esa información temporal para verificar el posicionamiento generado por medio del sistema de visión artificial.

Para mostrar con precisión la fecha y la hora en un vídeo, se emplea el *framework ffmpeg* que, mediante una función de superposición de texto específica, permite la visualización precisa de la información temporal en la ubicación deseada del vídeo lo que permite el análisis temporal posterior. Para este propósito, se ha decidido crear un nuevo *script* denominado **video_capture_ffmpeg**. En la Figura 3-27, se muestra la incorporación de esta función, junto con el número de *frame*. Cabe destacar que el color del texto se ha decidido poner en negro debido al cielo parcialmente nublado, facilitando así su visualización.

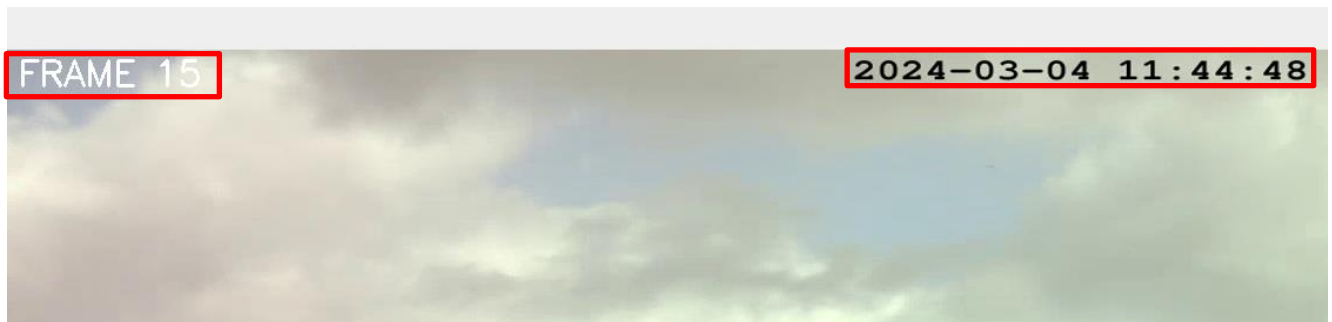


Figura 3-27 Captura de un *frame* del vídeo capturado, a la izquierda el número correspondiente y a la derecha la fecha y hora [autoría propia]

Posteriormente, durante el proceso de análisis, se concluyó que, si para mejorar la precisión habría sido necesario emplear valores temporales con precisión de décimas de segundo, ya que en cada segundo una cámara obtiene un total de 10 *frames*. Por otra parte, la importancia de este factor se atenúa en cierta medida si se considera que los objetos a seguir se mueven a bajas velocidades. No obstante, este aspecto será analizado con detalle en el capítulo de Sincronización de datos empleando dos cámaras

Tras completar el entrenamiento del modelo y mostrar los datos requeridos en la captura de vídeo, se ejecutan simultáneamente el *script* **ais_camera_capture**, junto con **video_capture_tool**, para almacenar ambos datos al mismo tiempo. Ambas capturas se guardan en un directorio específico para su posterior procesamiento con otro *script* denominado **video_track_3d_ais_tool**, el cual sincroniza ambas fuentes de datos y utiliza el sistema de 3 dimensiones desarrollado por Víctor Recio para medir distancias entre embarcaciones. Como se indicó con anterioridad, este sistema permitirá verificar las coordenadas GPS que devuelven los transpondedores AIS con las obtenidas por la aplicación tridimensional.

En la cámara de la izquierda durante el *track*, tal y como se muestra en la Figura 3-28, se presenta las detecciones de ambas cámaras, pero con las detecciones de la cámara derecha trasladadas de forma

constante para poder asociar las diferentes detecciones obtenidas en ambas cámaras. Al asociar estas cajas sobre un mismo contacto, para evitar el error y la confusión entre sí, se tienen en cuenta los siguientes factores para calcular una distancia de similitud básica que permita establecer que dos cajas se relacionan entre sí para establecer la detección de un único objeto:

- La distancia entre los centros de ambas cajas.
- La diferencia entre el tamaño de ambas cajas.

De esta manera, si dos contactos tienen tamaños notablemente diferentes, no existirá confusión al asociarles sus correspondientes boxes cuando se encuentren próximos entre sí. En el siguiente enlace se aprecia el vídeo en funcionamiento:

https://drive.google.com/file/d/1CgGk5356hqHC-WPwe-6xSPbE-CqPrZg_/view?usp=sharing



Figura 3-28 Cámara izquierda durante el track con ambas cajas en cada embarcación [autoría propia]

Sin embargo, cuando dos veleros se cruzan, el sistema puede perder el contacto, y encontrarlo de nuevo, pero identificándolo como uno nuevo objeto (con un nuevo número identificador) en lugar del anterior. Es aquí, en los momentos en que se producen oclusiones entre objetos, donde aparece uno de los problemas característicos del seguimiento de objetos.

Por otro lado, en la cámara de la derecha se muestra el resultado de la interfaz virtual, tanto como para calcular las distancias que existen entre los diferentes contactos realizada por Víctor (Figura 3-29), como para mostrar el AIS virtual y los datos cinemáticos de cada embarcación (Figura 3-30).

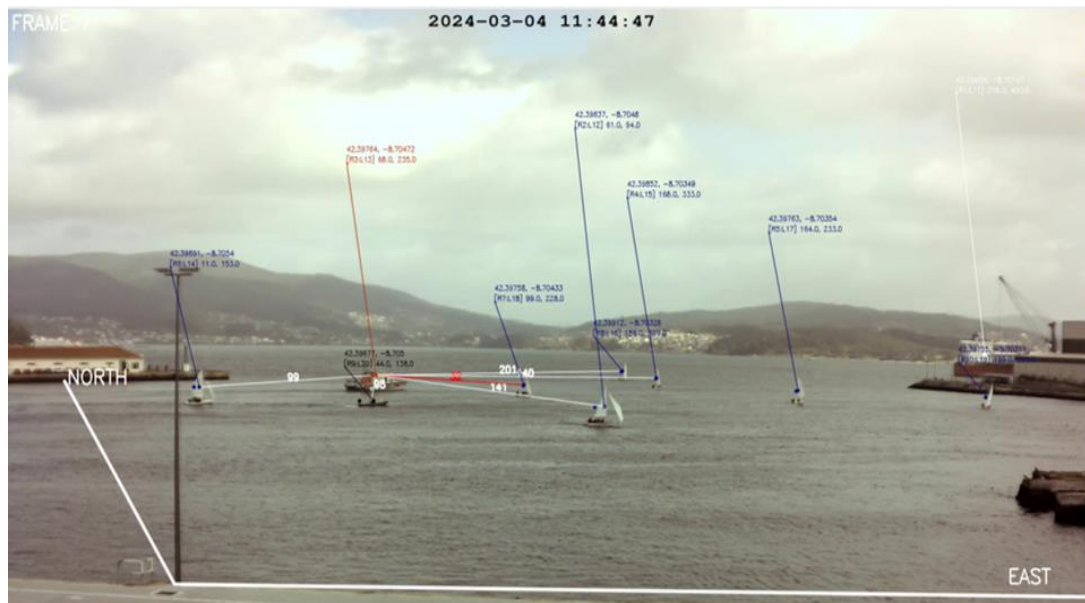


Figura 3-29 Captura del *video_track_3d_tool* en su función distancia entre embarcaciones [autoría propia]

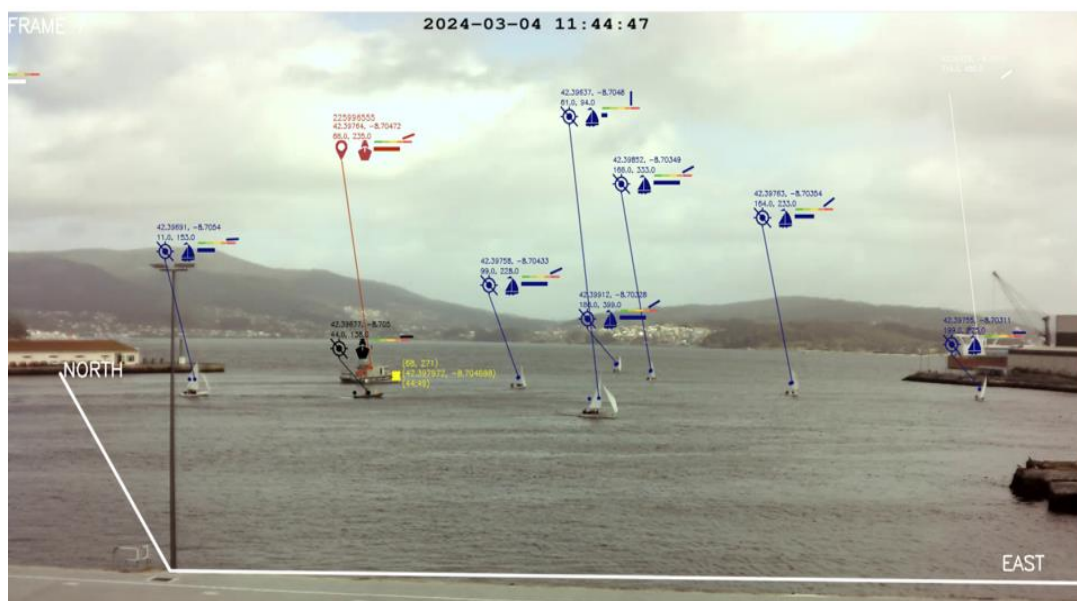


Figura 3-30 Captura del *video_track_3d_tool* en su función de interfaz virtual [autoría propia]

3.9.4 Visualización de los puntos GPS en Google Maps

Para verificar la obtención de los puntos en el plano, es necesario obtener los puntos GPS de distintos *frames* seleccionados del vídeo y emplear la herramienta *Google My Maps* para realizar su representación, es decir pasar datos 3D estimados de los vídeos a una vista 2D. Durante la ejecución del *script*, se selecciona cuáles son los *frames* deseados para que se guarde un archivo “cvs” que contenga la posición GPS de cada embarcación, facilitando su posterior importación al mapa.

Posteriormente, en la plataforma *Google My Maps*, se introducen dos capas diferentes, una primera en la que se muestran los puntos GPS base utilizados para crear el sistema de cálculo de distancias, como se aprecia en la Figura 3-31. La segunda capa incluye los puntos GPS devueltos por el sistema para un *frame* de vídeo en concreto.

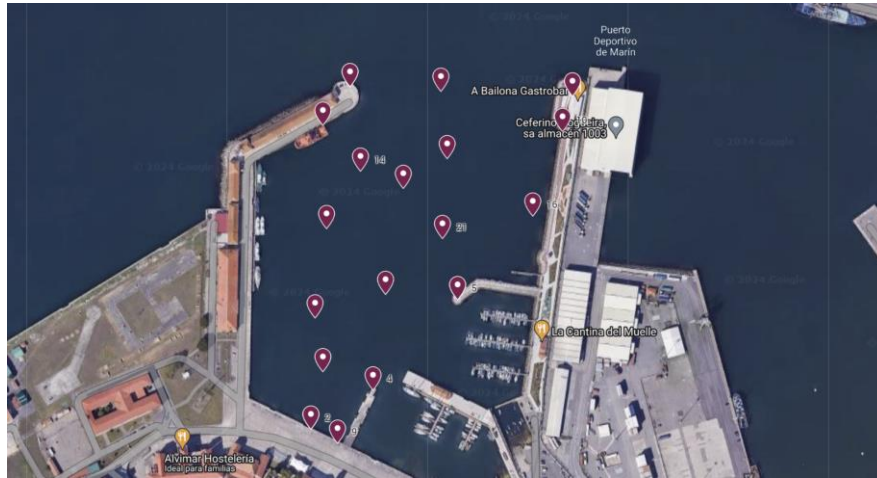


Figura 3-31 Imagen extraída de *Google Maps* donde se ven los puntos GPS base utilizados para crear el sistema de medición de distancias [Alfárez de fragata Don Víctor Recio]

De esta manera, al agregar las coordenadas de un dato AIS devuelto por la lancha de instrucción, es posible visualizar el error que presenta, es decir, la distancia que existe entre la posición estimada por el sistema y la que realmente devuelve el transpondedor de la embarcación.

Para ilustrar el funcionamiento de dicho sistema, se ha decidido extraer el *frame* número 350, donde se observa la posición de la lancha de instrucción con una notable precisión respecto de los datos AIS, es decir, en la Figura 3-32 rodeado por un cuadrado rojo, se puede apreciar el instante en el que ambos datos coinciden en dos puntos que se encuentran muy cercanos. Sin embargo, al observar la desde *Google Maps*, se percibe que la distancia en profundidad, aunque parezca más precisa, en realidad es mayor. Esto se debe a que la estimación de la profundidad en la dirección norte (eje Y) representa una mayor dificultad, con respecto a los desplazamientos de la dirección del este (eje X).

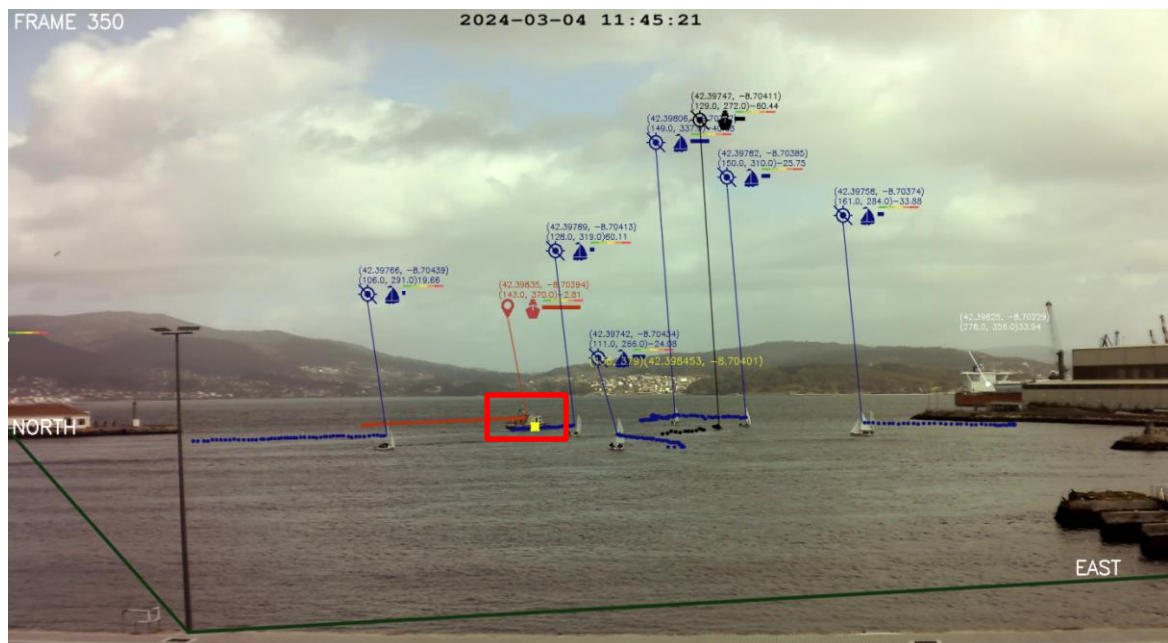


Figura 3-32 Captura del *frame* número 350 en donde se puede apreciar como el dato AIS coincide en la posición de la lancha de instrucción [autoría propia]

En la captura de dicho *frame* mencionado anteriormente, se pueden apreciar elementos adicionales de la interfaz, pero todos serán explicados con detenimiento en “Cuarta tarea: elaboración interfaz Realidad Aumentada”. Para verificar la posición de los datos AIS, como ya se ha mencionado, es necesario contrastar dicha información en 2D utilizando *Google Maps*. En la Figura 3-33, se pueden ver las diferentes embarcaciones de color rojo, destacando a la lancha de instrucción en verde y sus datos AIS de un tono más claro.

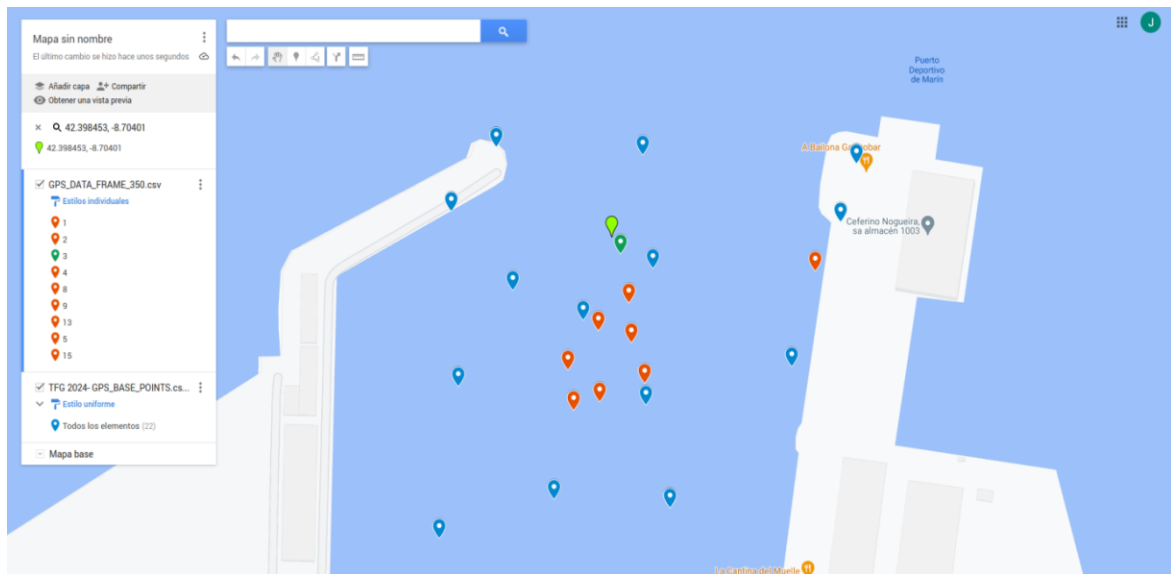


Figura 3-33 Captura de las embarcaciones en el *frame* 350 en un plano 2 dimensiones [Plano de *Google My Maps* de autoría propia]

En la Figura 3-34, se verifica la discrepancia entre la posición real de la lancha respecto de sus datos AIS, confirmando un error mayor en la profundidad del eje Y que en el eje X, tal y como se había indicado anteriormente. A pesar de ello, es importante señalar que la lancha se encuentra bastante lejos de la posición desde donde se tomaron los vídeos, resultando en un margen aproximado de 12 metros



Figura 3-34 Distancia entre la lancha de instrucción y la coordenada GPS que devuelve su transpondedor AIS [Plano de *Google My Maps* de autoría]

3.10 Cuarta tarea: elaboración interfaz Realidad Aumentada

A continuación, en este último capítulo del desarrollo se va a justificar como se ha elaborado la interfaz, desde los símbolos elegidos para cada embarcación y situación, hasta los datos que se muestran para cada una de ellas.

3.10.1 Justificar los símbolos

En la Tabla 5 Justificación de los símbolos escogidos para la interfaz virtual [autoría propia] siguiente aparecen cuales son los símbolos elegidos para cada embarcación y situación:

 military_boat/teaching_boat	• Cuando se trate de un <i>teaching_boat</i>, o de un <i>military_boat</i>, se mostrará este símbolo tras ser reconocido por el modelo de entrenamiento. • El color cambiará en función del número de barcos diferentes que haya en el entorno, o si hay varias embarcaciones emitiendo datos AIS.
 sailing_boat	• Cuando se trate de un <i>sailing_boat</i>, se mostrará este símbolo tras ser reconocido por el modelo de entrenamiento. • En este caso serán del mismo color si ninguna emite AIS, si hubiera alguna transmitiendo se le cambiaría el color.
 gps_globe	• Este icono, será del mismo color que el elegido para el icono de la embarcación. • Aparecerá cuando un buque este transmitiendo datos AIS.
 no_gps_globe	• Este icono, será del mismo color que el elegido para el icono de la embarcación. • Aparecerá cuando un buque no tenga la necesidad de transmitir datos AIS, como ocurre con los veleros.
 speedometer	• Esta barra aparecerá para todos los buques, acompañado de una barra justo debajo indicando la velocidad que lleva la embarcación en cuestión.

Tabla 5 Justificación de los símbolos escogidos para la interfaz virtual [autoría propia]

3.10.2 Disposición de la ventana

A continuación, en la Figura 3-35 Datos que ofrece la interfaz virtual para vigilancia portuaria [autoría propia] se presenta la distribución de los datos correspondientes a cada embarcación, incluyendo su número MMSI, coordenadas GPS, datos cinemáticos y emisión de datos AIS (en caso de encontrarse emitiendo), junto con la clasificación del tipo de embarcación. Para la construcción de dicha interfaz, se ha utilizado un *script*, cuya función principal es determinar qué elementos se mostrarán en la interfaz y cómo se distribuirán.

Es importante señalar que cada buque traza una línea de puntos discontinua que representa su trayectoria durante la reproducción del vídeo, teniendo está el mismo color que los datos asociados a la embarcación.

Para la calcular la velocidad, ha sido necesario la extracción de coordenadas de las embarcaciones a lo largo del tiempo, a partir de lo cual se estima la velocidad mediante una suma ponderada de su histórico de velocidades expresada de la siguiente forma:

$$v = \sum_{i=1}^n (v_i * p_i)$$

Siendo v_i , es el valor del dato i-ésimo en el histórico de velocidades, y p_i es el porcentaje asociado al dato i-ésimo, siendo n es el número total de datos empleado para el cálculo de la velocidad.

Se ha empleado un proceso similar para determinar el rumbo de la embarcación, aunque en este caso debido a la presencia de ruido a la disminución de precisión en el posicionamiento cuando la distancia en el eje Y aumenta, se han observado ciertos errores y variaciones drásticas en el rumbo observado, lo cual lleva a pensar que se necesita mejorar el sistema de posicionamiento en 3D para poder realizar una estimación más precisa del rumbo de las embarcaciones.

Para mejorar la visualización de los resultados, se ha seleccionado un sistema de coordenadas con dos dimensiones (x, y), medidas en metros cuyo origen se sitúa en la farola que se encuentra en la explanada y que es claramente visible en los vídeos capturados. Estas coordenadas GPS se calculan también con el mismo sistema, aprovechando que los puntos utilizados para su creación ya disponen de sus propias coordenadas GPS como referencia.

Para facilitar que la presentación de la interfaz sea lo más clara posible y las ventanas de información añadida de cada embarcación no se mezclen entre sí, se ha optado por una función que escalona cada ventana a una altura dejando cierto espacio de separación entre ellas. De esta manera, el usuario encargado puede supervisar la información que se muestra por pantalla sin que esta se mezcle entre sí. Por otra parte, esta solución puede no ser suficiente para otro tipo de vídeos, por lo que el desarrollo de un algoritmo de posicionamiento dinámico para visualización de las ventanas de información de cada embarcación puede ser una tarea para desarrollar en futuros trabajos sobre este tema.

En último lugar, como ocurre con la lancha de instrucción, aquellas embarcaciones que se encuentren emitiendo datos AIS, mostrarán en la parte superior de la ventana, su número MMSI asociado.

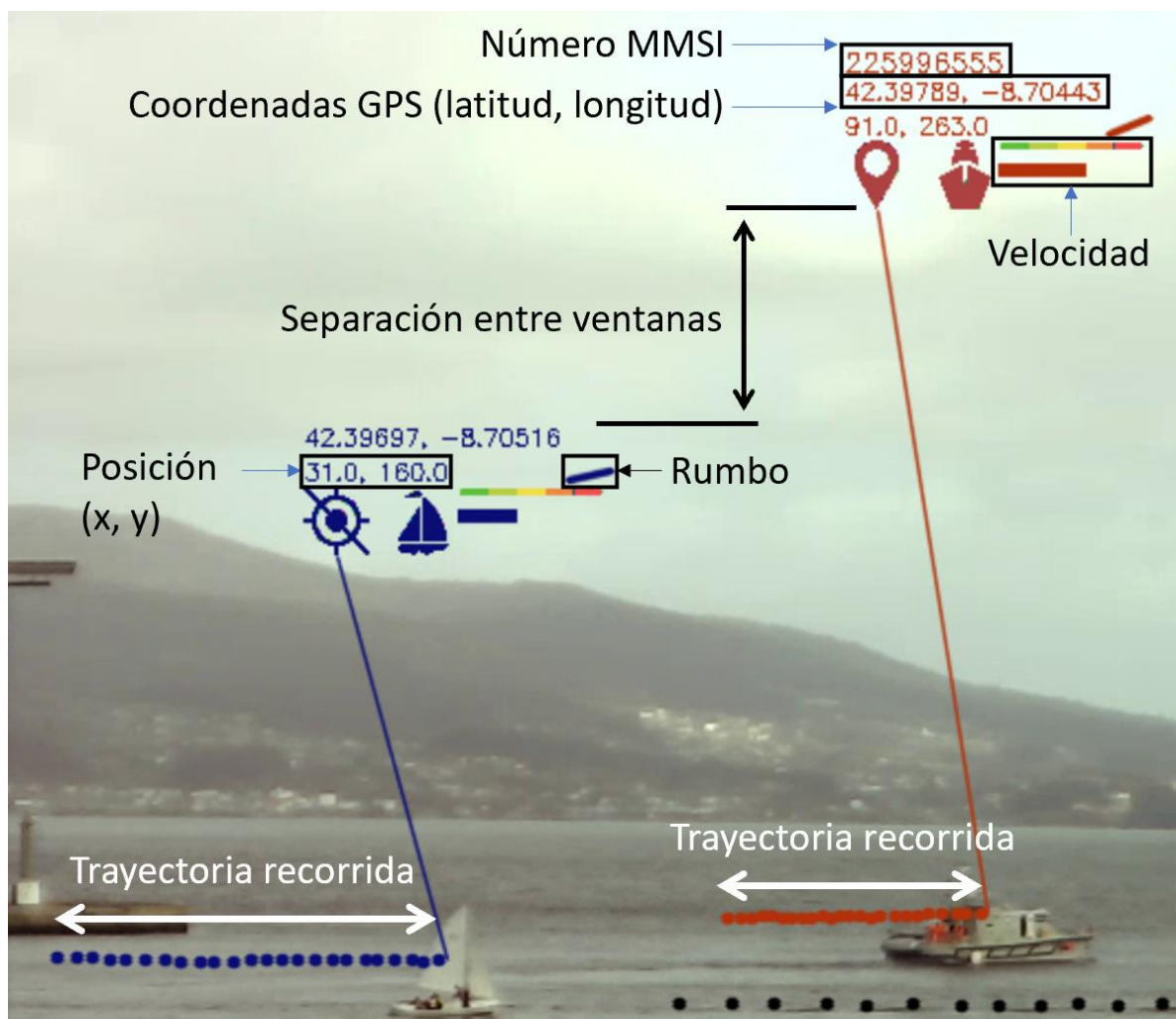


Figura 3-35 Datos que ofrece la interfaz virtual para vigilancia portuaria [autoría propia]

4 RESULTADOS / VALIDACIÓN / PRUEBA

4.1 Introducción

En este capítulo, se evalúan detalladamente los resultados obtenidos y recopilados en cada etapa del desarrollo, es decir, se trata de extraer conclusiones justificadas y válidas, poniendo especial énfasis en aquellos hallazgos que merecen un reconocimiento especial por su singularidad y contexto del presente trabajo.

En primer lugar, se justifica el modelo final entrenado con *Ultralytics HUB*, el cual se ha utilizado posteriormente en la interfaz, y se discuten los problemas y fallos más significativos que se han presentado. Posteriormente, se examinan los datos AIS en *Google Maps* junto con las trayectorias obtenidas a través de las coordenadas GPS, para, en último lugar, evaluar el rendimiento de la interfaz virtual, y verificar si la información que muestra cumple con los objetivos establecidos.

4.2 Resultados de la prueba realizada con el *dataset* de Escalera Monumental en *Ultralytics HUB*

Como se puede apreciar en la Figura 4-1, que muestra la gráfica del entrenamiento en *Ultralytics HUB*, el modelo entrenado exhibe un alto nivel de precisión para 1000 *epochs*, demostrado por el buen comportamiento que muestran los parámetros *precision*, *mAP*, y *recall*. Además, esta afirmación se puede respaldar con las demás gráficas de *box_loss*, *object_loss* y *class_loss*, las cuales confirman el que el entrenamiento del modelo ha sido adecuado para el volumen de datos existente.

Como ya se había mencionado en Entrenamiento dataset MARIN_PORT en *Ultralytics HUB*, se ha decidido entrenar el modelo final en *Ultralytics HUB* para ahorrar tiempo y por ser más rápido gracias a la GPU online de *Google Colab* con la que trabaja, además de devolver gráficas que ayudan a dar una mejor explicación de los resultados como se ha visto anteriormente. A continuación, en los siguiente enlaces se pueden ver los resultados del modelo haciendo *tracking* de uno de los vídeos que se va ha utilizado para la creación de la interfaz virtual.

<https://drive.google.com/file/d/11zUG6R9DaKod-hWQV0sj5y1GScDvgdh4/view?usp=sharing>

https://drive.google.com/file/d/1_Sf6vt_qWFuHfWzz54jCHp-imCbU1dac/view?usp=sharing

Además, como se puede ver en la Figura 4-2, **video_track_tool** hace uso del modelo entrenado para hacer *tracking* de las embarcaciones. En el siguiente enlace, se adjunta un vídeo del mismo *script*, donde cada clase del modelo es identificada. Muestra cómo un barco deja una línea de puntos sobre su trayectoria a medida que avanza en una dirección específica.

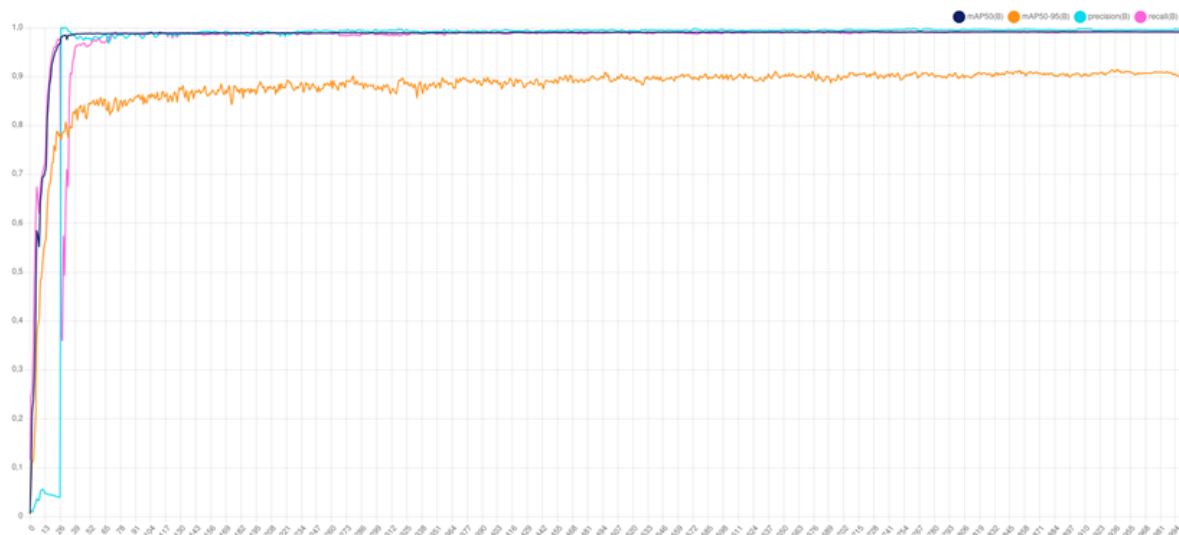


Figura 4-1 Representación visual proporcionada por *Ultralytics HUB* que muestra los resultados del *mAP*, *recall* y *precision* para cada uno de los *epochs* analizados [gráfica generada por *Ultralytics HUB* de autoría propia]



Figura 4-2 Captura del script *video_track_tool* [autoría propia]

4.2.1 Fallos comunes encontrados

Tal y como es natural esperar, el modelo generado no es perfecto, y hay ocasiones en las que le resulta difícil hacer seguimiento de algunas embarcaciones durante breves periodos de tiempo. Es el caso que ocurre cuando una de las lanchas no ha atravesado aún la farola que se encuentra en la explanada, como se aprecia en la Figura 4-3 y el primer enlace. Para solucionarlo, se extraen más *frames* del momento exacto en el que el modelo falla, para hacer un reentrenamiento de este y solucionar el error en cuestión. Otro fallo común es cuando se da una situación de cruce entre varias embarcaciones, es decir, se produce una oclusión entre objetos, de forma que el sistema las pierde y las vuelve a identificar como nuevos elementos, no como el anterior, como se aprecia en la Figura 4-4 y el segundo enlace.

https://drive.google.com/file/d/1UFGQMzMkv3UgRkc_YHyW7l0Eu2vaSEMF/view?usp=sharing

https://drive.google.com/file/d/1nmnK0Y1JROZR_qN26vzbLDc9rnsiJcCd/view?usp=sharing



Figura 4-3 Lancha de instrucción que el sistema no detecta correctamente [autoría propia]



Figura 4-4 Pérdida de embarcaciones cuando se produce una situación de cruce [autoría propia]

A pesar de la corrección realizada, se observa que en ocasiones sigue perdiendo la lancha en ese momento, por lo que haría falta tomar nuevos vídeos y extraer cientos de *frames* para reentrenar el modelo hasta que este fallo se solucione. Es aquí donde vuelve a aparecer el problema que se expuso desde momentos iniciales en el presente trabajo, la generalización.

Es por ello, que un modelo, cuando se extrapola del entorno en el que ha sido entrenado ya sea porque se trata de otro puerto o incluso porque las condiciones meteorológicas hayan modificado la luminosidad, tiene dificultades para reconocer embarcaciones que ya era capaz de reconocer en el entorno en el que se había sometido al proceso de aprendizaje.

Para evitar en mayor medida dicho contratiempo, se ha escogido una cámara cuyo *zoom*, enfoque e ISO pueden ajustarse manualmente, de forma que se garantice la consistencia de los vídeos capturados para el entrenamiento del modelo.

Como en los vídeos utilizados para el primer modelo algunos de ellos fueron tomados con un teléfono *Iphone*, que cuenta con la función de autoenfoco, se apreciaba que vídeo contaba con una condición de luminosidad diferente, dificultando notablemente la generalización del modelo. Como se aprecia en la Figura 4-5 con tan solo una diferencia de 15 minutos, la luminosidad cambia considerablemente debido a las condiciones meteorológicas.



Figura 4-5 Diferencia de luminosidad debido a las condiciones meteorológicas [autoría propia]

4.2.2 Verificación de datos AIS en Google Maps

Tras la creación del sistema tridimensional capaz de medir distancias a las embarcaciones de la dársena de la Escuela Naval Militar, como se aprecia en la Figura 4-6, y el siguiente enlace, se va a poder verificar que tan preciso es el mismo, respecto de las coordenadas GPS que nos devuelve el transpondedor a bordo de la lancha de instrucción.

<https://drive.google.com/file/d/1LYdWK-d-wk1fVDV9euXUrvzB0vNcDHS/view?usp=sharing>

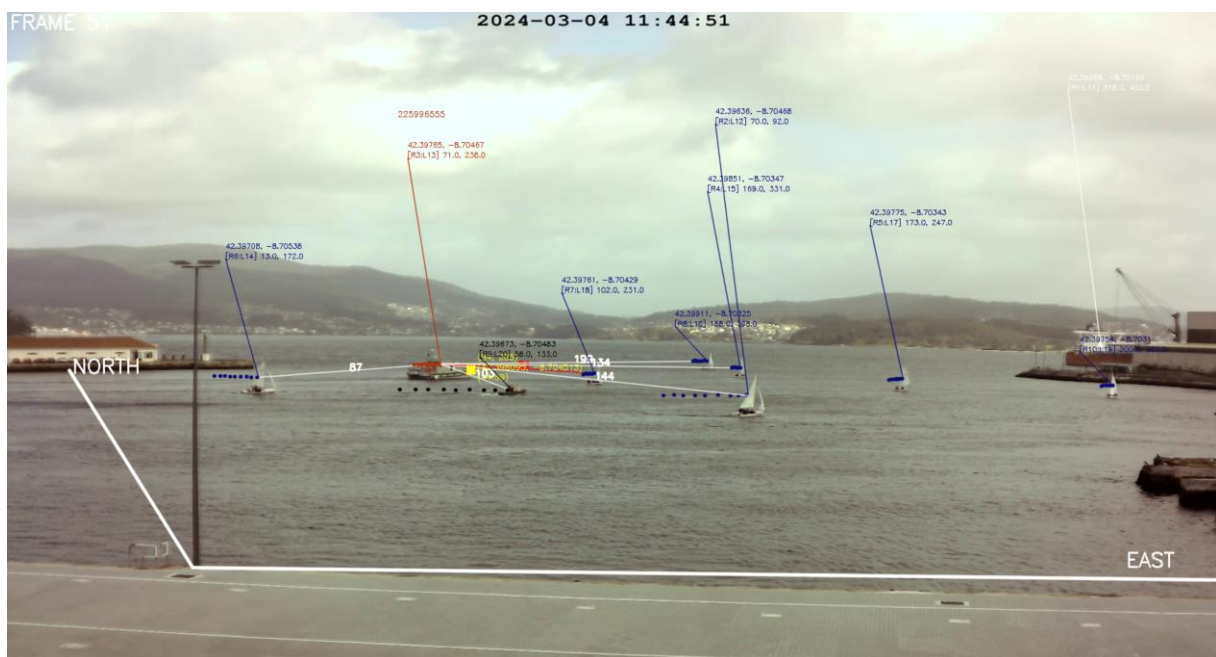


Figura 4-6 Captura del *video_track_3d_tool* en su función distancia entre embarcaciones [autoría propia]

Como ya se explicaba en el capítulo de desarrollo, fueron necesarias dos cámaras separadas para ser capaces de medir las distancias a las que se encuentran las embarcaciones, gracias a la diferencia de píxeles que existe entre ambas cámaras, y el conocimiento de coordenadas GPS de puntos reconocibles en los vídeos capturados.

En primer lugar, en la Figura 4-7 se puede apreciar como el sistema es capaz de devolver las coordenadas GPS de *frames* deseados, para de esta manera, pasarlos a una visión en 2 dimensiones y analizar la trayectoria que siguen los buques.



Figura 4-7 Vista en 2 dimensiones de la trayectoria de las embarcaciones en *Google Maps* [autoría propia]

Además, en la Figura 4-8 se verifica que a pesar de que, en el vídeo, dos embarcaciones que parece que se encuentran muy próximas entre sí, al pasar las coordenadas al plano en 2 dimensiones, se averigua que se encuentran más alejadas en profundidad de lo que parece en un primer instante.

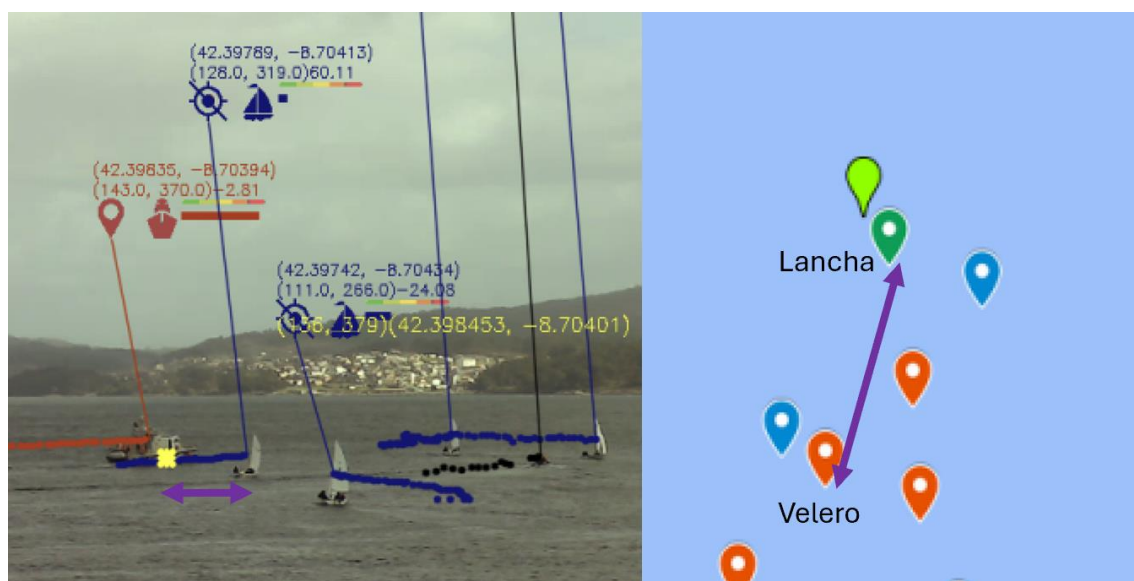


Figura 4-8 Diferencia de vista en 2 dimensiones en *Google Maps* respecto de la interfaz virtual [autoría propia]

Para la coordenada GPS (en color verde claro en la Figura 4-8Figura 4-9) que devuelve la lancha de instrucción en el *frame* 350 de la Figura 3-34, se puede apreciar que ocurre lo mismo, en un primer instante parece que coincide en la misma posición de la lancha, pero en realidad se encuentra a una distancia superior en profundidad.

Una posible solución a este problema consistiría en colocar una de las cámaras en un ángulo de 90°, pudiendo situarla a la altura del edificio de investigación, de forma que cada cámara tuviese una buena estimación en cada uno de los ejes X e Y. Sin embargo, haría falta contar con otro ordenador, y la toma de datos de manera simultánea incluyendo la sincronización de los vídeos podría llegar a verse comprometida sino se realiza con cuidado.

Otro resultado apreciado es que el buque mercante que se encuentra atracado en el puerto de Marín y que aparece en la parte derecha de los vídeos, parece estar dentro de la dársena con los puntos GPS utilizados para el sistema de distancias tridimensional, pero en realidad se encuentra en otro muelle fuera de la dársena. También se aprecia un problema similar con el velero que aparece en el extremo más a la derecha de los vídeos, ya que las distancias que devuelve son muy poco precisas.

Esto se debe a que, en los extremos, por más puntos que se coloquen, el sistema encuentra dificultades para calcular distancias, por lo que, colocando otra cámara con una orientación perpendicular, se podría ayudar a ampliar tanto la precisión del sistema, como el rango dentro del que mide distancias.

4.3 Validación del Interfaz Virtual de RA

A continuación, se emplea el script *video_track_3d_tool* para mostrar la información del interfaz de RA integrada en el vídeo, de forma que se presentan los datos de posiciones, velocidad y rumbo estimados, el tipo de embarcación detectada, así como su MMSI si se trata de una embarcación que emite datos AIS, así como la trayectoria que ha recorrido. En el siguiente enlace se aprecia un vídeo del script en funcionamiento:

<https://drive.google.com/file/d/149VDFQsxkB9OunBf8mEA1jNAO2CZuGUbx/view?usp=sharing>

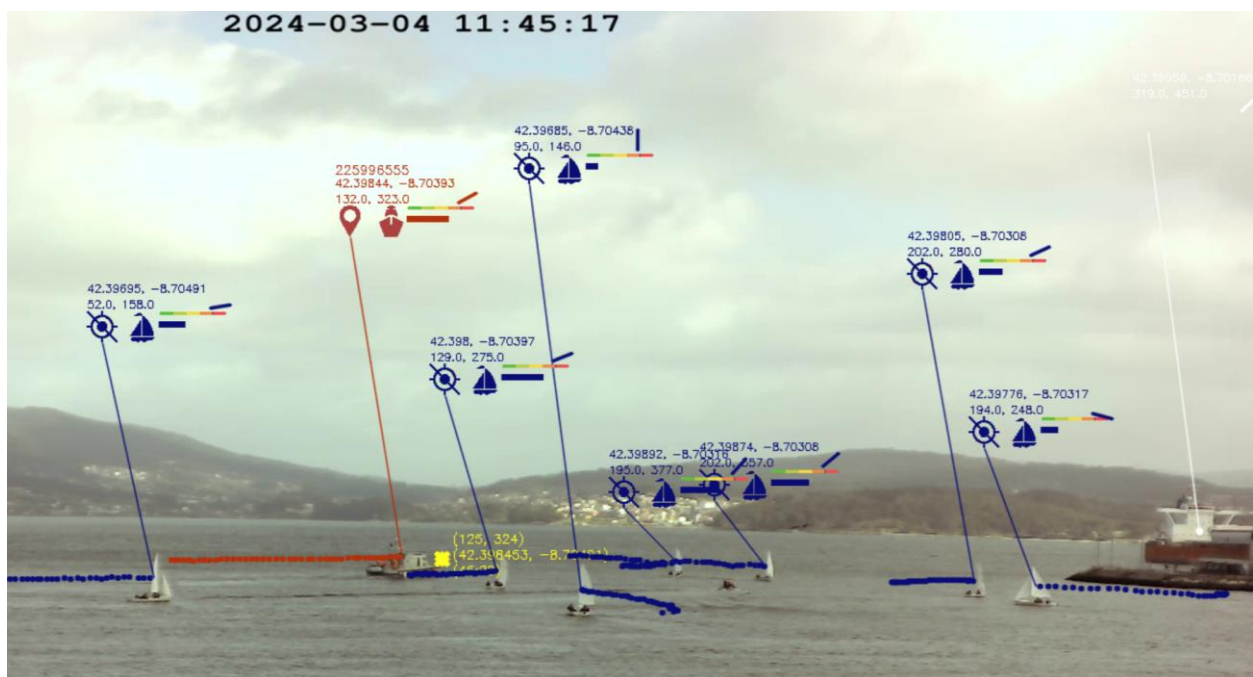


Figura 4-9 Captura del *video_track_3d_tool* en su función de interfaz virtual [autoría propia]

Como se puede apreciar en la Figura 4-9, donde se presenta el resultado obtenido (el cual aparece ampliado para facilitar la visualización de la información de cada buque), el interfaz generado permite:

- Estimar la futura trayectoria de una embarcación empleando la visualización de la trayectoria que ha recorrido.
- Comprobar si la velocidad y el rumbo con el que navegan son correctos y se encuentran dentro de unos valores razonables.
- Mejorar el seguimiento de las embarcaciones en la visualización empleando un tipo de color diferente para cada tipo de embarcación.
- Los datos AIS, mostrados en amarillo, permiten evaluar visualmente la precisión del sistema de posicionamiento empleado en el interfaz de RA
- La ventana de información de cada embarcación se sitúa a diferentes alturas para que, en los momentos de cruces de embarcaciones, sus datos puedan seguir visualizándose sin que se mezclen en la pantalla, aunque en ocasiones debido a la falta de espacio esto no es posible.

4.4 Posibles mejoras de los resultados

Una vez analizados los resultados, además de proponer que en el futuro los vídeos se graben con dos cámaras en posiciones perpendiculares para igualar la precisión en ambos ejes de coordenadas, se han observado otra serie de cuestiones que sería importante tener en cuenta para generar mejoras en el sistema de localización visual desarrollado.

Por una parte, al analizar en los vídeos los resultados obtenidos, se ha observado que, habría sido mejor haber integrado la hora en los vídeos incluyendo las décimas de segundo, , ya que cada segundo de vídeo se compone de un total de 10 *frames*.

Por otra parte, cabe destacar que también sería posible generar modelos con mejores prestaciones empleando versiones de YOLOv8 con un mayor número de parámetros, pero debido a los límites en los equipos de trabajo no sería posible realizar su entrenamiento de forma local. Por otra parte, aunque se entrenase un modelo de mayor tamaño de forma *online*, a posteriori su carga y ejecución sobre un vídeo requeriría de ordenadores con la suficiente potencia para poder ser ejecutados en tiempo real, o su adaptación algún tipo de dispositivo especializado en la ejecución de modelos de redes neuronales, tales como procesadores GPU o TPU portátiles que se conectan por USB a un PC.

Por último, con el objetivo de mejorar las prestaciones del sistema, sería necesario obtener nuevos vídeos con diferentes condiciones meteorológicas y que incluyan todos los tipos de embarcaciones que se pretende que el sistema pueda identificar, de forma que es necesario aumentar el tamaño del *dataset* para volver a entrenar el modelo hasta dar soporte a las nuevas condiciones de detección y seguimiento de embarcaciones.

5 CONCLUSIONES Y LÍNEAS FUTURAS

5.1 Introducción

En el presente capítulo se dispone a extraer las conclusiones extraídas tras la realización del mismo, independientemente de la calidad de las mismas. Además, se indican cuáles han sido los objetivos alcanzados, y si se corresponden con los planteados previamente del desarrollo del trabajo. En caso de no haber sido posible el cumplimiento de estos, justificar el motivo y que se alternativas hubieran sido necesarias para conseguirlo.

En último lugar se exponen las líneas futuras sobre las que arroja luz la realización de este TFG, para poder continuar con proyectos de carácter similar, o su continuación.

5.2 Conclusiones extraídas

El campo de la IA en la vigilancia marítima ha sido el que ha protagonizado el proyecto desde momentos iniciales. Se trata, como ya se indicó en el Estado del arte, de tecnologías cuya aplicación en la vida cotidiana, se encuentra en auge, especialmente las técnicas de DL, integrada con la RA en la visión artificial. Esta revolución ha dado lugar a la sustitución de numerosas tareas realizadas por personas, dando lugar a la economización de recursos y tiempo.

Por lo tanto, la primera conclusión extraída de este TFG es la importancia que podría tener la implementación de estas tecnologías en el entorno marítimo, ya que serían capaces de procesar y asimilar mucha más información en tiempo real de la que es capaz de procesar un individuo que cuente, únicamente, con una cámara de vigilancia estándar y su percepción del escenario que está vigilando.

Esta idea se ve reforzada por los acontecimientos sucedidos las últimas semanas, en donde se ha podido comprobar que los accidentes en puerto, y sus proximidades, son más frecuentes de lo que realmente se piensa. De esta forma, el pasado, 26 de Marzo en Baltimore, un buque portacontenedores impactó contra uno de los pilares del puente Francis Scott Key, provocando su derrumbe. Pocos días más tarde, otra embarcación choca con un puente en Oklahoma, pero esta vez sin haber heridos o daños estructurales.

Ante situaciones de este tipo, es lógico pensar que la continua vigilancia a través de una interfaz virtual como la desarrollada en este trabajo ayudaría a prevenir este tipo de accidentes a tiempo, siendo importante destacar el elevado impacto económico que causa un accidente de dicho calibre frente al coste que puede suponer un sistema de vigilancia automático como el que se propone. Por otra parte, este tipo de sistemas podrían ser empleados también para realizar el posterior análisis de accidentes que ya han ocurrido, al poder contar con el seguimiento de los datos cinemáticos del mismo sumado a los datos AIS que emiten la mayoría de las embarcaciones, de forma que ayude a esclarecer las causas de dichos eventos.

Otra conclusión obtenida es la consideración de las condiciones del entorno de entrenamiento, como son la meteorología, la luminosidad, el contraste, la diversidad de los elementos, el número de clases y sus posibles similitudes como la eslora o el tipo de embarcación. De esta manera se deduce que estos sistemas son sometidos a procesos de aprendizaje muy específicos y concretos, por lo que, si se extrae del entorno en el que ha sido entrenado, aun sería capaz de reconocer las embarcaciones, pero con ciertas dificultades si no se parece a la situación bajo la cual ha sido entrenado. Debido a esto, es probable que el modelo deba someterse a fases de reentrenamiento tantas veces como sea necesario ampliar su capacidad de procesamiento en nuevos entornos, lo cual implica la necesidad de ampliar los *datasets* con más imágenes, especialmente con aquellas donde se aprecie mayor dificultad para detectar a los buques.

Relacionada con la conclusión anterior, y en términos de detección de objetos, hay situaciones particulares en las que puede resultar bastante problemático conseguir un seguimiento preciso de los buques, independientemente del número de entrenamientos que se realicen. Respecto a este punto se puede afirmar que se ha alcanzado un grado de formación adecuado, ya que durante el Desarrollo del TFG la experiencia obtenida al intentar formar y modelar en los *trackings* realizados a vídeos con los que no ha trabajado éste, los objetos son fácilmente enmascarables en situaciones de oclusión. Uno de los casos más extendidos es cuando varios objetos se ocultan entre ellos, es decir, se da una situación de cruce. Otros casos en los que los objetos se pierden ocurren cuando entran o salen de la escena, ya que algunos vídeos capturados poseen los extremos borrosos o de baja calidad.

Otra conclusión que se ha destacado a través del trabajo es el correcto uso de las cámaras para medir distancias, en el sentido de que se debe prever una cuidadosa sincronización de las mismas para que el sistema pueda funcionar de manera eficaz.

Además, la presencia de datos de cronometraje (en minutos, segundos, décimas y centésimas) en el propio vídeo es un requisito fundamental para la sincronización de los datos AIS y a la hora de generar la interfaz virtual. Adicionalmente, es muy crítico tener más de una cámara no sólo como solución al error de distancia en profundidad, sino también para disminuir el error de distancia observado cuando los objetos se sitúan desde lados o periferias. Para resolver este problema se ha propuesto que en futuros trabajos la instalación de las cámaras se haga de forma que estén perpendiculares entre sí, de forma que cada una se encargue de dar una mayor resolución en cada uno de los dos ejes de coordenadas establecidos.

En lo que a dispositivos de grabación se refiere, se concluye que han sido un éxito, ya que las cámaras han sido de fácil adquisición y cómodas para transportar entre las diferentes ubicaciones de captura para cualquier usuario. Además, su capacidad de ajustar los parámetros de grabación de manera manual, han sido de gran utilidad a la hora de obtener vídeos uniformes que faciliten entrenamientos con entornos similares, que faciliten posteriormente la generalización de los modelos entrenados.

Por último, pero no menos importante, estaría la influencia que podría ejercer sobre una serie de tareas realizadas por la Armada. Se pueden ver fácilmente situaciones en las que sería muy valioso contar con ayudas para la toma de decisiones y, en este sentido, las técnicas de inteligencia artificial se configuran como el medio principal para permitir la flexibilidad del personal. Si a esto le sumamos que estos sistemas se basan en técnicas de aprendizaje profundo como las CNN con operaciones realizadas sobre funciones matemáticas, su confiabilidad sería, probablemente, significativamente mayor que la de los sistemas que utilizan cualquier otro medio.

5.3 Objetivos alcanzados

En cuanto al balance general de los objetivos alcanzados tras la realización del trabajo, si se realiza un análisis imparcial de los mismos, se puede considerar positivo.

En primer lugar, la formación por parte del alumnado en cuestiones relacionadas con la inteligencia, visión, y redes artificiales ha sido realizada, de forma que el posterior uso de los conceptos estudiados ha sido de utilidad a la hora de analizar los resultados obtenidos en este trabajo.

De esta manera, enumerando los objetivos que se han cumplido, el hecho de crear una interfaz virtual que aplique técnicas de DL y ML para detectar y hacer seguimiento de las embarcaciones que han estado presentes durante la captura de videos en el entorno portuario de la Escuela Naval Militar, ha sido conseguido satisfactoriamente. Si bien, es necesario indicar que las cámaras no ofrecían gran nitidez y calidad de la imagen para altas distancias, al ser desde un inicio elementos de bajo coste sin grandes prestaciones, por lo que se considera que el objetivo queda ha sido realizado con éxito a partir de los recursos existentes.

Además, se ha podido integrar la interfaz con datos AIS, coincidiendo las coordenadas GPS de los mismos con la trayectoria que describe la lancha en muchos de los vídeos capturados, lo cual ha sido empleado para determinar la precisión del posicionamiento del sistema planteado. Para conseguir este objetivo, se ha logrado combinar el trabajo del presente proyecto con el de Víctor Recio, cuya función es la de medir distancias empleando dos cámaras empleando un sistema de visión estereoscópica con entrenamiento de un modelo de redes neuronales. Gracias a la integración del mismo, se ha podido enriquecer la información que se presenta sobre las ventanas de la interfaz virtual con el objetivo de que pudiera servir de gran utilidad para ser implementado en un sistema de vigilancia portuaria.

Sin embargo, al comienzo del desarrollo del trabajo, el objetivo era localizar una ventana física sobre la que, con la ayuda de pantallas acrílicas, poder proyectar la interfaz creada y aplicar la RA. Esto no ha sido posible, en primer lugar, por la imposibilidad de disponer de una ventana que cumpliera con los requisitos necesarios para poder crear la ventana, y, en segundo lugar, debido a que las pantallas acrílicas son de elevado coste y la única alternativa que existía era el papel celofán, cuya calidad es inferior.

A pesar de este último objetivo, como se indicaba al comienzo del apartado, el balance es positivo por haber logrado los propuestos previamente al desarrollo del TFG.

5.4 Tendencias futuras

En resumen, después del desarrollo del presente TFG y se hayan evaluado el desarrollo y las conclusiones, esta sección pretende destacar algunas ideas de posibles líneas futuras para la investigación. Existen algunas preguntas que se plantearon, y en base a estas, se podría explorar un poco más lo que se ha hecho aquí hasta ahora, lo que con suerte dará una base sólida para elaborar un producto de alta calidad que pueda usarse por la Armada en cuestiones de defensa en un futuro más bien próximo. Las líneas sugeridas son las siguientes:

- La elaboración de interfaces similares, pero combinando aún más cámaras, para resolver el problema de la profundidad y utilizarlo en sectores de mayor tamaño. Esta ampliación permitiría mejorar la precisión y la calidad de la información recopilada en entornos marítimos para fines de vigilancia y seguridad.
- Otra línea aparecerá al examinar cómo afecta la eficacia del sistema de vigilancia portuaria, el aumento significativo del número de barcos. Sería necesario un análisis que examinase la capacidad del sistema para procesar este incremento de datos y mejorar su detección y seguimiento de embarcaciones. El objetivo final sería garantizar que el sistema pueda seguir funcionando correctamente incluso cuando exista un tráfico considerablemente denso.
- Otra implementación para futuros proyectos podría ser el uso, por ejemplo, de un dispositivo dominado *Coral USB Accelerator*, que contiene un ASIC especializado (Google Edge TPU) para acelerar los cálculos de inferencia en Machine Learning, por lo que, debido a su tamaño, podría transportarse fácilmente y hacer uso de la interfaz virtual en tiempo real.
- Mediante el uso de gafas de realidad virtual, se podría simular un puente de navegación, y utilizar una interfaz de realidad aumentada de manera virtual para poder comprobar cómo se verían dichas interfaces en realidad.
- Se considera también la implementación del *keypoint detection*, que permitiría detectar armas, personas y elementos característicos de los buques. Sin embargo, su viabilidad se ve limitada por la necesidad de imágenes de alta calidad, lo cual presenta un desafío dada la naturaleza del

presente TFG, centrado en técnicas de aprendizaje profundo con dispositivos portátiles de grabación.

En definitiva, las técnicas de ML y DL están llamadas a ser claves en el desarrollo de sistemas que automaticen muchas de las tareas repetitivas en la Armada y en el ámbito marítimo en general. Su aplicación en la interfaz virtual para la vigilancia portuaria promete mejorar la detección y seguimiento de embarcaciones y prevenir accidentes en el futuro, pero hay que seguir explotando estas líneas. Sin embargo, mejorar su rendimiento implica que se necesitan equipos más potentes, ya que como se ha indicado en el capítulo, se aspira a obtener modelos más eficientes y precisos.

6 BIBLIOGRAFÍA

En esta sección se reflejarán todas las referencias como son los recursos web, artículos, libros, etc., contando con la información de autores, año, título, edición junto con la fecha del último acceso realizado en caso de tratarse de una referencia a recursos online.

- [1] «Armada_Lineas_Generales_2.0_reducido_5sept2017.pdf». Accedido: 17 de marzo de 2024. [En línea]. Disponible en: https://www.ieee.es/Galerias/fichero/OtrasPublicaciones/Nacional/2017/Armada_Lineas_Generales_2.0_reducido_5sept2017.pdf
- [2] A. V. Sales, «ENSEÑANZA NAVAL Y ENTORNO 4.0».
- [3] «El Port de Barcelona y la startup AllRead utilizan la inteligencia artificial para identificar contenedores y vagones | Port de Barcelona». Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <http://www.portdebarcelona.cat/es/comunicacion/noticias/el-port-de-barcelona-y-la-startup-allread-utilizan-la-inteligencia-artificial>
- [4] «Vista de Ecosistema de tecnologías emergentes: realidad aumentada, virtual y mixta». Accedido: 21 de enero de 2024. [En línea]. Disponible en: <https://www.tecnologia-ciencia-educacion.com/index.php/TCE/article/view/1148/5491>
- [5] «What is Augmented Reality (AR)?», WhatIs. Accedido: 26 de diciembre de 2023. [En línea]. Disponible en: <https://www.techtarget.com/whatis/definition/augmented-reality-AR>
- [6] «A Brief History of Augmented Reality (+ Future Trends & Impact)». Accedido: 21 de enero de 2024. [En línea]. Disponible en: <https://www.g2.com/articles/history-of-augmented-reality>
- [7] C. Hoy, «Para qué podemos usar las Google Glass», Computer Hoy. Accedido: 21 de enero de 2024. [En línea]. Disponible en: <https://computerhoy.com/noticias/hardware/que-podemos-usar-google-glass-20591>
- [8] «Qué es la realidad aumentada | Xperiencia Virtual». Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://www.xperienciavirtual.es/es/que-es-la-realidad-aumentada/>
- [9] «Pokémon Go: la realidad aumentada pone a prueba la propiedad intelectual». Accedido: 22 de enero de 2024. [En línea]. Disponible en: https://www.wipo.int/wipo_magazine/es/2017/01/article_0005.html
- [10] A. Alcolea, «visionOS: el prometedor software de las Apple Vision Pro mezcla el mundo real y el virtual», Xataka. Accedido: 21 de enero de 2024. [En línea]. Disponible en: <https://www.xataka.com/aplicaciones/visionos-prometedor-software-apple-vision-pro-mezcla-mundo-real-virtual>

- [11] F. van den Oever, B. Sætrevik, M. Fjeld, y K. Nordby, «A Virtual Reality Experiment shows that Augmented Reality can improve Collaboration in Ship Navigation», ene. 2024, doi: 10.31234/osf.io/ts5wx.
- [12] «VIVE Pro | VIVE European Union». Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://www.vive.com/eu/product/vive-pro/>
- [13] E. Gernez, K. Nordby, J. O. Eikenes, y O. S. Hareide, «A review of augmented reality applications for ship bridges».
- [14] Armada., *TROZO DE SEGURIDAD INTERIOR EN EJERCICIO DE ZAFARRANCHO DE COMBATE*. 2013. Accedido: 11 de enero de 2024. [Photo]. Disponible en: <https://www.flickr.com/photos/armadamde/8379868227/>
- [15] R. D. InfoDefensa, «Industria premia al Ejército por la automatización y robotización de los almacenes del Pcami», Infodefensa - Noticias de defensa, industria, seguridad, armamento, ejércitos y tecnología de la defensa. Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://www.infodefensa.com/texto-diario/mostrar/3890738/industria-premia-ejercito-automatizacion-robotizacion-almacenes-pcami>
- [16] «Líder mundial en computación de IA», NVIDIA. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://www.nvidia.com/es-es/>
- [17] R.- ASALE y RAE, «inteligencia | Diccionario de la lengua española», «Diccionario de la lengua española» - Edición del Tricentenario. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://dle.rae.es/inteligencia>
- [18] «Marvin Minsky», *Wikipedia, la enciclopedia libre*. 30 de noviembre de 2023. Accedido: 22 de enero de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Marvin_Minsky&oldid=155715561
- [19] «GNU/Linux», *Wikipedia, la enciclopedia libre*. 21 de febrero de 2024. Accedido: 4 de marzo de 2024. [En línea]. Disponible en: <https://es.wikipedia.org/w/index.php?title=GNU/Linux&oldid=158341923>
- [20] A. G. ISABEL MARIA, C. Q. ANGEL MIGUEL, C. P. OTTO, E. R. FRANCISCO, y L. O. ANGEL MIGUEL, *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. Ediciones Paraninfo, S.A., 2003.
- [21] «01_inteligencia_artificial_juan_sossa.pdf». Accedido: 22 de enero de 2024. [En línea]. Disponible en: https://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/89/1/01_inteligencia_artificial_juan_sossa.pdf
- [22] A. M. IESC, «Artificial Intelligence, Machine Learning, and Deep Learning: Same context, Different concepts», Master Intelligence Economique et Stratégies Compétitives. Accedido: 22 de enero de 2024. [En línea]. Disponible en: <https://master-iesc-angers.com/artificial-intelligence-machine-learning-and-deep-learning-same-context-different-concepts/>
- [23] «Qué es overfitting y underfitting y cómo solucionarlo | Aprende Machine Learning». Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/>
- [24] J. L. Gonzalez, «Tipos de aprendizaje automático», SoldAI. Accedido: 23 de enero de 2024. [En línea]. Disponible en: <https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2>
- [25] L. González, *Diferencia aprendizaje supervisado y no supervisado* 3. 2019. Accedido: 23 de enero de 2024. [Photo]. Disponible en: <https://www.flickr.com/photos/ligdieli/48054954426/>
- [26] «TFG_Garcia_de_Andoin_Bolaño_Mikel.pdf». Accedido: 23 de enero de 2024. [En línea]. Disponible en: https://addi.ehu.es/bitstream/handle/10810/49049/TFG_Garcia_de_Andoin_Bola%C3%B1o_Mikel.pdf?sequence=4&isAllowed=y

- [27] C.-I. Cira, «CONTRIBUTION TO OBJECT EXTRACTION IN CARTOGRAPHY: A NOVEL DEEP LEARNING-BASED SOLUTION TO RECOGNISE, SEGMENT AND POST- PROCESS THE ROAD TRANSPORT NETWORK AS A CONTINUOUS GEOSPATIAL ELEMENT IN HIGH-RESOLUTION AERIAL ORTHOIMAGERY».
- [28] «Charles Darwin», *Wikipedia, la enciclopedia libre*. 17 de enero de 2024. Accedido: 23 de enero de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Charles_Darwin&oldid=157312852
- [29] K. Palasundram, N. Sharef, K. Kasmiran, y A. Azman, «Enhancements to the Sequence-to-Sequence Based Natural Answer Generation Models», *IEEE Access*, vol. PP, pp. 1-1, mar. 2020, doi: 10.1109/ACCESS.2020.2978551.
- [30] «Trackmania | Ubisoft». Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.ubisoft.com/es-es/game/trackmania/trackmania>
- [31] L. Neinders, «Improving Trackmania Reinforcement Learning Performance: A Comparison of Sophy and Trackmania AI».
- [32] «¿Cuál es la diferencia entre el machine learning y el deep learning?» Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://blog.bismart.com/diferencia-machine-learning-deep-learning>
- [33] Coursesteach, «Computer Vision (Part 1)-What is Computer Vision», Medium. Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://medium.com/@Coursesteach/computer-vision-part-1-ff493a713887>
- [34] I. G. S y V. C. S, «La visión artificial y los campos de aplicación», *Tierra Infin.*, vol. 1, n.º 1, Art. n.º 1, dic. 2015, doi: 10.32645/26028131.76.
- [35] Ultralytics, «Ultralytics YOLOv8 Tareas». Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://docs.ultralytics.com/es/tasks>
- [36] «Tipos de redes neuronales (Clasificación)». Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://inteligencia-artificial.dev/tipos-redes-neuronales/>
- [37] «Redes Neuronales y Deep Learning. Capítulo 1: Preludio», Future Space S.A. Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://www.futurespace.es/redes-neuronales-y-deep-learning-capitulo-1-preludio/>
- [38] D. Calvo, «Clasificación de redes neuronales artificiales», Diego Calvo. Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>
- [39] «Centeno Franco Alba TFG.pdf». Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://docs.google.com/viewerng/viewer?url=https://idus.us.es/bitstream/handle/11441/90004/Centeno+Franco+Alba+TFG.pdf?sequence%3D1&isAllowed=y>
- [40] «memoria.pdf». Accedido: 4 de abril de 2024. [En línea]. Disponible en: <https://upcommons.upc.edu/bitstream/handle/2117/129220/memoria.pdf>
- [41] «Convolutional Neural Networks: Architectures, Types & Examples». Accedido: 25 de marzo de 2024. [En línea]. Disponible en: <https://www.v7labs.com/blog/convolutional-neural-networks-guide>, <https://www.v7labs.com/blog/convolutional-neural-networks-guide>
- [42] «Reconocimiento de objetos con Deep Learning - Damavis Blog». Accedido: 24 de enero de 2024. [En línea]. Disponible en: <https://blog.damavis.com/reconocimiento-de-objetos-con-deep-learning/>
- [43] Ultralytics, «YOLO Métricas de rendimiento». Accedido: 26 de marzo de 2024. [En línea]. Disponible en: <https://docs.ultralytics.com/es/guides/yolo-performance-metrics>
- [44] «Sistema de Identificación Automática», *Wikipedia, la enciclopedia libre*. 23 de diciembre de 2022. Accedido: 4 de abril de 2024. [En línea]. Disponible en:

- https://es.wikipedia.org/w/index.php?title=Sistema_de_Identificaci%C3%B3n_Autom%C3%A1tica&oldid=148140684
- [45] «IMO_8814275.jpg (5616×3744)». Accedido: 21 de enero de 2024. [En línea]. Disponible en: https://upload.wikimedia.org/wikipedia/commons/9/97/IMO_8814275.jpg
- [46] «¿Qué Es Un Dataset? Descubre Qué Son Los Conjuntos De Datos». Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://www.pontia.tech/que-es-un-dataset-conjunto-de-datos/>
- [47] «Bienvenido a Toyota España | Vehículos Toyota», Toyota ES. Accedido: 10 de marzo de 2024. [En línea]. Disponible en: <https://www.toyota.es>
- [48] Ultralytics, «COCO». Accedido: 11 de marzo de 2024. [En línea]. Disponible en: <https://docs.ultralytics.com/es/datasets/detect/coco>
- [49] R. Zhang, S. Li, G. Ji, X. Zhao, J. Li, y M. Pan, «Survey on Deep Learning-Based Marine Object Detection», *J. Adv. Transp.*, vol. 2021, pp. 1-18, nov. 2021, doi: 10.1155/2021/5808206.
- [50] Y. Jie, L. Leonidas, F. Mumtaz, y M. Ali, «Ship Detection and Tracking in Inland Waterways Using Improved YOLOv3 and Deep SORT», *Symmetry*, vol. 13, p. 308, feb. 2021, doi: 10.3390/sym13020308.
- [51] J. Zhong, M. Li, J. Qin, Y. Cui, K. Yang, y H. Zhang, «REAL-TIME MARINE ANIMAL DETECTION USING YOLO-BASED DEEP LEARNING NETWORKS IN THE CORAL REEF ECOSYSTEM», *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLVI-3-W1-2022, pp. 301-306, abr. 2022, doi: 10.5194/isprs-archives-XLVI-3-W1-2022-301-2022.
- [52] «Cesadar – Arisnova – Ingeniería de Sistemas». Accedido: 26 de marzo de 2024. [En línea]. Disponible en: <http://www.arisnova.com/naval/cesadar/>
- [53] J. D. G.-A. Lacalle y F. L. López, «EVOLUCIÓN DEL CENTRO DE SUPERVISIÓN Y ANÁLISIS DE DATOS DE LA ARMADA (CESADAR)».
- [54] J. M. L. Ros, «INTELIGENCIA ARTIFICIAL APLICADA A LA GUERRA DE MINAS NAVALES».
- [55] «Ucrania llama a los civiles a conformar un “ejército de drones” para combatir a Rusia», ELMUNDO. Accedido: 27 de marzo de 2024. [En línea]. Disponible en: <https://www.elmundo.es/internacional/2022/07/16/62d29da3e4d4d876358b45b8.html>
- [56] «Cámara web con zoom de 5 megapíxeles con objetivo manual variable de 5 a 50 mm, UVC Plug & Play, USB para conferencia/enseñanza online USB500W05G-MFV (5-50) : Amazon.es: Informática». Accedido: 4 de abril de 2024. [En línea]. Disponible en: <https://www.amazon.es/megap%C3%ADxeles-Raspberry-conferencia-ense%C3%B1anza-USB500W05G-MFV/dp/B08X46NND9>
- [57] «iPhone 13 Pro Max de 256 GB reacondicionado - Grafito (Libre) - Apple (ES)». Accedido: 6 de marzo de 2024. [En línea]. Disponible en: <https://www.apple.com/es/shop/product/FLLA3QL/A/iphone-13-pro-max-256-gb-reacondicionado-grafito-libre>
- [58] «QK-A026 Receptor inalámbrico AIS con GPS -Vendedor del Reino Unido : Amazon.es: Deportes y aire libre». Accedido: 4 de abril de 2024. [En línea]. Disponible en: https://www.amazon.es/qk-a026-inal%C3%A1mbrico-AIS-GPS-Receptor-vendedor/dp/B074TFY95B/ref=sr_1_1?__mk_es_ES=%C3%85M%C3%85%C5%BD%C3%95%C3%91&crd=38OFAJ42QXKZS&dib=eyJ2IjojMSJ9.Y7kUIIrubulzR2Nq1mz_IWBnrIBxLB3pWHDtOJouBbPb_AvPmHcN67mHbu6RkOy9aMIPzyPAvSVVMxA2y8eSDFQBR2DsNqWRj08r5jpYuQw.CVsTxu3j871CEAOgU24L1smDw_VyCYqRgq9ZsXJ1rc&dib_tag=se&keywords=3.3.3+Receptor+AIS+inal%C3%A1mbrico+QK-A024&qid=1712228325&srefix=3.3.3+receptor+ais+inal%C3%A1mbrico+qk-a024%2Caps%2C223&sr=8-1&ufe=app_do%3Aamzn1.fos.5e544547-1f8e-4072-8c08-ed563e39fc7d
- [59] «Open Source Data Labeling», Label Studio. Accedido: 4 de marzo de 2024. [En línea]. Disponible en: <https://labelstud.io/>

- [60] «VSCodium - Open Source Binaries of VSCode». Accedido: 5 de marzo de 2024. [En línea]. Disponible en: <https://vscodium.com/>
- [61] «Microsoft: nube, ordenadores, aplicaciones y juegos». Accedido: 6 de marzo de 2024. [En línea]. Disponible en: <https://www.microsoft.com/es-es>
- [62] «Google Colaboratory». Accedido: 8 de marzo de 2024. [En línea]. Disponible en: <https://colab.research.google.com/>
- [63] R. Khandelwal, «ByteTrack: A Simple Yet Effective Multi-Object Tracking Technique», Medium. Accedido: 26 de marzo de 2024. [En línea]. Disponible en: <https://arshren.medium.com/bytetrack-a-simple-yet-effective-multi-object-tracking-technique-86f1f3632a85>
- [64] N. Aharon, R. Orfaig, y B.-Z. Bobrovsky, «BoT-SORT: Robust Associations Multi-Pedestrian Tracking». arXiv, 7 de julio de 2022. doi: 10.48550/arXiv.2206.14651.
- [65] «Guía-del-No-Idiota-para-el-AIS.pdf». Accedido: 4 de marzo de 2024. [En línea]. Disponible en: <https://digitalyacht.es/blog/wp-content/uploads/2020/06/Gu%C3%ADa-del-No-Idiota-para-el-AIS.pdf>
- [66] S. V. B. GmbH, «Todo lo que debe saber sobre AIS», SVB GmbH. Accedido: 29 de marzo de 2024. [En línea]. Disponible en: <https://www.svb-marine.es/es/guia/todo-lo-que-debe-saber-sobre-ais.html>
- [67] «AIS Messages | Navigation Center». Accedido: 5 de marzo de 2024. [En línea]. Disponible en: <https://www.navcen.uscg.gov/ais-messages>
- [68] «Reglamento internacional para prevenir los abordajes en la mar (1972). 14ª ed., 3ª reimp. = International regulations for preventing collisions at sea (1972). 14ª ed., 3ª reimp», Castellano. Accedido: 5 de marzo de 2024. [En línea]. Disponible en: <https://publicaciones.defensa.gob.es/reglamento-internacional-para-prevenir-los-abordajes-en-la-mar-1972-14-edicion-21870.html>
- [69] «La inteligencia artificial tiene un impacto ambiental mucho mayor del que se creía». Accedido: 16 de marzo de 2024. [En línea]. Disponible en: <https://www.lavanguardia.com/tecnologia/innovacion/20190617/462863973194/inteligencia-artificial-impacto-ambiental-mayor-creia.html>
- [70] «EUNAVFOR “Operación Atalanta” - Ministerio de Defensa de España». Accedido: 17 de marzo de 2024. [En línea]. Disponible en: https://www.defensa.gob.es/misiones/en_exterior/actuales/listado/atalanta.html
- [71] «BOE-A-2005-18933 Ley Orgánica 5/2005, de 17 de noviembre, de la Defensa Nacional.» Accedido: 17 de marzo de 2024. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2005-18933>
- [72] «Sea Guardian - Ministerio de Defensa de España». Accedido: 17 de marzo de 2024. [En línea]. Disponible en: https://www.defensa.gob.es/misiones/en_exterior/actuales/listado/otan-sea-guardian.html
- [73] «Estrategia de Seguridad Nacional - Ministerio de Defensa de España». Accedido: 17 de marzo de 2024. [En línea]. Disponible en: <https://www.defensa.gob.es/defensa/politicadefensa/estrategiaseguridad/>
- [74] «Coches eléctricos, energía solar y limpia», Tesla. Accedido: 17 de marzo de 2024. [En línea]. Disponible en: <https://www.tesla.com/>
- [75] J. A. P. Estapé, «Autopilot de Tesla: 736 accidentes y 17 muertos», Computer Hoy. Accedido: 17 de marzo de 2024. [En línea]. Disponible en: <https://computerhoy.com/motor/autopilot-tesla-736-accidentes-17-muertos-1258862>
- [76] «Así es como las Apple Vision Pro pueden ayudar a las personas con problemas de visión a ver mejor», El Español. Accedido: 17 de marzo de 2024. [En línea]. Disponible en:

https://www.lespanol.com/omicrono/hardware/20240215/apple-vision-pro-pueden-ayudar-personas-problemas-vision-ver-mejor/832916950_0.html

- [77] V. P. Rodríguez, «El aterrador efecto secundario de las Vision Pro», Business Insider España. Accedido: 17 de marzo de 2024. [En línea]. Disponible en: <https://www.businessinsider.es/apple-vision-pro-son-experimento-gigantesco-cerebro-humano-1364023>

ANEXO I: IMPLICACIONES SOCIALES, Y/O ECONÓMICAS, Y/O AMBIENTALES.

En este primer anexo se recogen algunas de las implicaciones sociales, económicas y medio ambientales de la visión artificial en las cámaras portuarias utilizadas para la vigilancia marítima. Desde la mejora de la seguridad portuaria y el control de las operaciones ilegales, hasta una gestión más eficaz del flujo y densidad del tráfico marítimo y la minimización de los efectos negativos sobre los ecosistemas locales, la IA junto con la RA tiene un potencial de gran alcance en diversas áreas. También se analiza el impacto que tiene la implementación de la RA y técnicas de IA como en la Armada Española y en la industria.

En el contexto de las cámaras de vigilancia portuaria, las implicaciones económicas e industriales de la aplicación de técnicas de IA y ML son notables. Contrario a ciertas percepciones, su implementación no siempre conlleva la pérdida de empleos, sino que permite capacitar sistemas y reubicar al personal en tareas más estratégicas y menos tediosas. Además, contribuyen a la seguridad marítima y a reducir positivamente el impacto medio ambiental. A continuación, se muestran algunos ejemplos importantes:

- En primer lugar, las cámaras de vigilancia portuaria, reducen el impacto ambiental en gran medida, ya que su modelo de entrenamiento puede haber sido entrenado para hacer seguimiento no solo de buques, si no de ser además capaz de hacerlo con otros objetos, como contenedores a la deriva o tanque de combustibles que puedan haberse precipitado accidentalmente de buques mercantes, o incluso ser capaces de distinguir manchas de otro color en la superficie portuaria que indiquen la presencia de combustible.
- A su vez, en el campo de la fauna marina, pueden servir de estudio para la presencia de especies que puedan encontrarse en peligro de extinción, ya que muchas veces algunos cetáceos se acercan a puerto por encontrarse desubicados por la contaminación acústica que produce el tráfico marino. Sin embargo, no todas las implicaciones ambientales son positivas porque algunos investigadores aseguran que la huella ecológica de los sistemas que emplean IA sería más grande que la que se esperaba como indican investigadores de la Universidad de Amherst en Massachusetts, Estados Unidos [69]. Los estudios aseguran que la energía consumida por las unidades que son capaces de procesar esa gran cantidad de datos puede llegar a 284 toneladas de CO₂e, es decir lo que 47 ciudadanos españoles contaminan a lo largo de un solo año.
- Económicamente, en relación con la detección de objetos flotando a la deriva en inmediaciones portuarias, suponen un grave peligro para las diferentes embarcaciones que se encuentren entrando o saliendo de puerto. Como resultado existe la posibilidad de que les produzca un impacto, o que tengan que variar su rumbo de manera repentina y tengan un abordaje contra la infraestructura del muelle. Gracias a la IA, estas podrían identificar dichos objetos y ayudar a los buques a prevenir su derrota ante estos peligros mejorando así la seguridad marítima.
- En caso de que dicha seguridad fallase y dos buques tuvieran un accidente bajo una situación de vuelta encontrada⁴, las cámaras podrían justificar sobre quien reside la responsabilidad del accidente a la hora de necesitar demostraciones visuales legales que verifiquen los hechos.
- Algunos puertos cuentan con un tráfico muy denso y activo, como es el caso del Puerto de Vigo, por lo que muchas veces se juntan mercantes con una eslora considerable junto con pequeños veleros o embarcaciones de recreo de tamaño similar. Si se dispone de una cámara

⁴ Vuelta encontrada: En el Reglamento Internacional para Prevenir Abordajes en el Mar, una situación de vuelta encontrada es cuando dos buques se aproximan entre sí con rumbos opuestos, es decir, proa con proa, y existe riesgo de abordaje si los rumbos se mantienen.

de vigilancia con un modelo que capaz de reconocer el tipo de embarcación y hacerles seguimiento además de distinguir si emite AIS o no, puede controlar el tráfico marítimo e incluso contribuir a la seguridad marítima.

- Son muchos puertos los que sufren de entradas de pequeñas embarcaciones de narcotráfico por lo que, gracias a las cámaras, si estas cuentan con modelos capaces de detectar su velocidad, pueden identificarlas más fácilmente, reduciendo el impacto del contrabando en la sociedad. Esta implementación también puede ser de utilidad para la armada a la hora de desplegarse en el cuerno de África en su lucha contra la piratería en la Operación Atalanta⁵.
- Una característica de este tipo de proyectos es la necesidad de contar con capacidad tecnológica, la cual puede requerir inversiones en *hardware* y *software* que no estén al alcance de todos los sectores. En la Armada podría verse reflejado por el ajuste del presupuesto financiero de ese momento.
- Otra consecuencia es la necesidad de contar con personal cualificado para implementar estas metodologías. Podría resultar fácil encontrar especialistas de este calibre en organizaciones privadas, pero en instituciones públicas como la Armada que cuenta con tecnologías de distintas épocas, la formación del personal deberá adaptarse a estas nuevas corrientes tecnológicas, cuyo proceso de aprendizaje también es elevado.

En conclusión, la introducción de IA en las cámaras destinadas a la vigilancia marítima y portuaria implica una variedad de impactos significativos en los ámbitos del medio ambiente y de la economía del entorno marino. El presente anexo, ha explorado los aspectos económicos y medio ambientales con mayor detenimiento que los sociales ya que los últimos serán abordados con mayor profundidad en el Anexo II: Reflexiones Éticas y Sociales para evitar en mayor medida la redundancia.

⁵ Operación Atalanta: Operación de la OTAN, que se centra en la lucha contra la piratería en el cuerno de África y en el Océano Índico [70].

ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

En este segundo anexo se reflexiona acerca de las implicaciones ético-sociales abordadas en el TFG. Para conseguirlo, iniciamos estableciendo el marco social para acotar el proyecto como un medio que trata de mejorar la seguridad dentro de un contexto específico como es el entorno marítimo. Esto se consigue implicando el compromiso que tiene la Armada como uno de los guardianes clave de la seguridad marítima a nivel nacional, además de su capacidad para adaptarse a los desafíos e incertidumbres que continuamente se presentan, como ocurre en el escenario entre Rusia y Ucrania. Por otro lado, se realiza un examen conciso para evaluar las implicaciones éticas asociadas a la introducción de sistemas con ML en la sociedad moderna, haciendo énfasis en el hecho de que el avance de las técnicas de DL no es favorable en la totalidad de los casos y como la realidad virtual y aumentada podrían influenciar en nuestra vida cotidiana.

La Armada Española siempre ha mantenido un compromiso firme y fuertemente estructurado con la defensa de la soberanía e integridad territorial como se refleja en la ley de defensa nacional [56]. A pesar de que, durante las últimas décadas, se hayan afrontado bajos presupuestos debido a la crisis económica que ha atravesado el país, la Armada siempre se ha mantenido fuerte ante dicho compromiso y ha sabido estar a la altura explotando los medios de los que dispone. Como fruto de ello afina la formación de sus miembros para exprimir al máximo la capacidad de dichos. El ejemplo más reciente de ello se vio cuando buques de la Armada como el Juan Carlos I, fueron activados durante la Operación Balmis para servir de hospital a las víctimas del coronavirus durante la pandemia.

Extrapolando las operaciones militares dentro del margen europeo, se puede apreciar como en las EUNAVFOR Sea Guardian⁶ o Atalanta, se lleva una tendencia hacia la defensa colectiva, dejando un poco más al lado la individual junto con la interoperabilidad más allá de la frontera de una nación. Esta tendencia toma fuerzas con la principal preocupación que tiene la OTAN actualmente con Rusia, obligando a Estados Unidos, junto con sus miembros europeos, fortificar la unión de fuerzas y empezar a pensar como un único ejército en caso de que el conflicto que tiene contra Ucrania escalase a niveles mayores.

Haciendo hincapié en los sistemas inteligentes abordados el presente TFG, el conflicto presenta en el Mar Negro por parte de las dos naciones mencionadas actualmente, ya ha causado la destrucción de más de un buque militar. Algunas de las técnicas empleadas están siendo USV's, pequeños drones controlados a distancia que impactan contra fragatas antes incluso de que puedan percatarse de su aproximación. Implementando estos modelos de IA, como los expuestos para las cámaras de vigilancia de puerto, se puede ayudar a prevenir este tipo de ataques.

Como ya se ha expuesto en la Introducción y objetivos del presente trabajo, la Armada en sus líneas generales de acción [OJO], es firmemente consciente de la importancia de la digitalización a todos sus niveles, tanto administrativos como a bordo de sus buques. Es por ello que además toma en gran parte importancia el concepto de ciberseguridad en cuando a nuevo espacio a defender en la estrategia de seguridad nacional [58]. Es aquí donde entran las técnicas de Inteligencia Artificial y ML de gran importancia para la actualidad, que facilitan a mantener los sistemas con una mayor seguridad y fiabilidad.

En el presente anexo se pretende exponer dichas técnicas desde un punto de vista social, ya que se promueve mejorar la situación de seguridad en la mar, como uno de los principales cometidos de la Armada. Pero, por otro lado, desde un punto de vista ético, el desarrollo de algunas de las técnicas de ML puede acarrear repercusiones que pudieran generar rechazo en determinados círculos de la sociedad. A continuación, se muestran y analizan algunas de estas técnicas:

⁶ Operación Sea Guardian: Operación de la OTAN, que trata de conocer el entorno marítimo con el fin de luchar contra el terrorismo en el Mar Mediterráneo [57].

- El primer factor a tener en cuenta es el creciente protagonismo que están teniendo las RNA como las que se han expuesto en el del presente trabajo, trasladadas a la de toma de decisiones de alto riesgo. Un ámbito donde cuestiones morales toman un grado de importancia considerable. Es decir, hasta qué punto una máquina, que independientemente del grado de exactitud y precisión que posea, sea capaz de tomar decisiones cuyo error pudiera poner en riesgo la integridad y salud de los usuarios. Bajo qué porcentaje de riesgo se estaría dispuesto a implementar dichas técnicas manteniendo una conciencia tranquila de que, en caso de accidente, dicho porcentaje sea de menor importancia que el fallo humano. Es el caso que encontramos en los vehículos desarrollados por Tesla [59], la compañía de Elon Musk, que lleva varios años apostando por coches eléctricos que cuenten con *Autopilot*⁷. Sin embargo, informes muestran que hasta la fecha han ocurrido 736 accidentes y hasta 17 muertos [60].
- Investigando a cerca de las implicaciones sociales que pueden tener herramientas como las gafas de realidad virtual, el ejemplo de las nuevas *Apple Vision Pro* de *Apple* es perfecto. Su tecnología es capaz de revolucionar la vida de las personas con discapacidad visual [61], y ofreciéndoles posibilidades con las que no podían contar antes de disponer de ellas, gracias a la superposición en tiempo real, de indicaciones de navegación o descripciones de objetos. Pero por desgracia, estas podrían tener consecuencias muy dañinas para el cerebro humano. Los Investigadores han descubierto que la inmersión durante periodos prolongados llevada al hábito podría cambiar nuestra forma de percibir el mundo y a las demás personas que nos rodean. También aparece un nuevo termino bautizado como ‘mareo del simulador’, que se manifiesta en náuseas, dolores de cabeza, etc., en usuarios que han utilizado las gafas durante dos semanas para todas sus actividades rutinarias [62]. Tras quitárselas experimentaban dificultades para relacionarse con el entorno, por alteraciones en su percepción del espacio y las distancias.

En pocas palabras, es crucial considerar las implicaciones éticas y sociales que tienen las técnicas de DL y ML, y hacer una revisión que permita un uso responsable por parte de los usuarios y una implementación más segura y fiable en el entorno civil y militar. La defensa de España puede servirse de dichas tecnologías, pero sus bases deben ser asentadas con mucha precaución para sacarles provecho. Además, las gafas de realidad virtual suponen un gran avance y son altamente explotables en numerosos entornos, pero aún están lejos de gozar de un uso a nivel cotidiano, hasta que no sean debidamente desarrolladas en el mundo profesional.

⁷ El *Autopilot* de *Tesla* es un sistema avanzado de asistencia al conductor que utiliza sensores y cámaras para realizar tareas de conducción, como mantenerse en el carril y controlar la velocidad de manera automática.

ANEXO III: GLOSARIO DE SIGLAS Y ACRÓNIMOS.

AIS: *Automatic Identification System.*

AJEMA: Almirante General Jefe de Estado Mayor de la Armada.

CESADAR: Centro de Supervisión y Análisis de Datos de la Armada.

CIA2: Centro de Inteligencia Artificial en la Armada.

CNN: *Convolutional neural networks.*

CO2e: Dióxido de Carbono equivalente.

COCO: *Common Objects in Context.*

ETA: *Estimated Time of Arrival.*

EUNAVFOR: *European Naval Force Operation.*

FPS: *Frames per second.*

GPS: *Global positioning system.*

IoU: *Intersection over Union.*

mAP: *mean Average Precision.*

ML: *Machine Learning*

MMSI: *Maritime Mobile Service Identity.*

MR: *Mixed Reality.*

NMEA: *National Marine Electronics Association.*

NMS: *Non-Maximum Suppression.*

OTAN: Organización del Tratado del Atlántico Norte.

PIP: Periodo de Inmovilización Programado.

RA: Realidad Aumentada.

RIPAM: Reglamento Internacional para Prevenir Abordajes en el Mar.

USB: *Universal Serial Bus.*

USV: *Unmanned Surface Vehicle.*

VHF: *Very High Frequencies.*

VR: *Virtual Reality.*

XR: *Extended Reality.*

YOLO: *You Only Look Once*