

# Análisis de Imágenes Satelitales por técnicas de Inteligencia Artificial

**Autor:** Muñoz Jiménez, José Antonio

**Director/es:** Vales Alonso, Javier y Troncoso Pastoriza, Francisco.

Contacto: jmunjim@mde.es

---

## Resumen:

El objetivo principal de este Trabajo Fin de Máster es el desarrollo de un proyecto para automatizar el procesado de imágenes de satélite, ofreciendo un sistema capaz de obtener información de diferentes fuentes (tanto privadas como públicas) y procesar automáticamente dicha información para garantizar la detección temprana y eficaz, y posterior identificación de elementos relevantes para la Defensa.

Entre los elementos relevantes a detectar se encuentran los barcos de un tamaño suficiente, las estructuras, la propia línea de costa y los emplazamientos de artillería de costa. El hecho de proponer una lista corta de elementos relevantes, no obsta que esta lista se pueda ampliar a otro tipo de elementos técnicamente detectables, pero esta inclusión en la lista llevaría a unas configuraciones de estructuras de detección potencialmente nuevas (adaptadas para la detección e identificación del elemento concreto) y al necesario entrenamiento dedicado en el procesado de estos nuevos elementos.

El objetivo del proyecto se plasmará en el desarrollo e implementación de una plataforma web de inteligencia geoespacial para Defensa. Esta plataforma estará basada en un conjunto de motores de inteligencia artificial que permitirán, por un lado, la colección, fusión y análisis de datos satelitales (imágenes ópticas y radar) procedentes de canales privados y fuentes abiertas y, por otro, la detección automática en las imágenes derivadas de los mismos, de “anomalías” u objetos cuya identificación resulte clave a la hora de tomar decisiones que puedan comprometer la seguridad tanto a nivel nacional como internacional.

**Palabras clave:** Entrenamiento, Detección, Identificación, Automatización.

---

## 1. Introducción y Objetivo

### 1.1. Introducción

Una de las fuentes de datos más importantes de información al mando son las fuentes satelitales., por lo que es de particular importancia la automatización de su explotación.. Esta necesidad de

procesar mucha información de manera rápida y automática, justifica el objetivo último de este estudio.

La automatización de las estructuras de detección, requiere el entrenamiento de éstas mediante miles o millones de imágenes preprocesadas, para que los sistemas aprendan a detectar e identificar correctamente. Conseguir los datos de entrenamiento y procesarlos (etiquetado) es muy laborioso, requiriendo un esfuerzo humano y técnico grande. El éxito de la tarea depende en gran medida de la cantidad y calidad del entrenamiento realizado, por eso el entrenamiento de las estructuras de detección es tan crítico. Contar previamente con estructuras pre-entrenadas para elementos similares, acorta drásticamente los tiempos de entrenamiento. Esta técnica se denomina “Transfer Learning” o transferencia del aprendizaje.

### *1.2. Objetivo*

El objetivo principal de este Trabajo Fin de Máster es el análisis y desarrollo de un sistema para automatizar el procesado de información satelital, ofreciendo un sistema capaz de obtener información de diferentes fuentes de datos (tanto privadas como públicas) y procesar automáticamente la información para garantizar una eficaz detección temprana y posterior identificación de elementos relevantes para la Defensa. Entre éstos se encuentran: barcos de tamaño suficiente, estructuras, línea de costa y emplazamientos de artillería de costa y antiaérea.

El proyecto conlleva el desarrollo e implementación de una plataforma web de inteligencia geoespacial para Defensa, con un conjunto de motores de inteligencia artificial que permitan, la colección, fusión y análisis de datos satelitales (ópticos y radar) y la detección automática, en las imágenes derivadas de los mismos, de “anomalías” u objetos relevantes para la seguridad tanto nacional como internacional.

Una vez indicados los parámetros de búsqueda en la plataforma (zona a inspeccionar, tipo de acción u objeto a buscar), el sistema buscará en ese área por los canales disponibles. Con esa información, el sistema fusionará y analizará los datos, extrayendo aquellos realmente relevantes. Estos datos se procesarán mediante un motor de inteligencia artificial basado en redes neuronales profundas (Deep Learning), que buscará el objeto, elemento o “anomalía” que el usuario indicó inicialmente en su patrón de búsqueda. Si el sistema identifica el objeto buscado, lanzará una alerta al usuario. El alcance final es disponer de un sistema operativo, entrenado para los casos de uso que se determinen de interés para el MINISDEF y preparado para su potencial despliegue en un entorno seguro, de manera que el software resultante esté preparado para ser instalado en la unidad operativa del MINISDEF que la Administración determine.

## **2. Desarrollo**

Se parte de un estudio de la arquitectura general para posteriormente detallar los modelos de Deep Learning y se detalla cómo estos módulos se estructuran y organizan para resolver los casos de uso del proyecto.

### **2.1 Análisis y diseño**

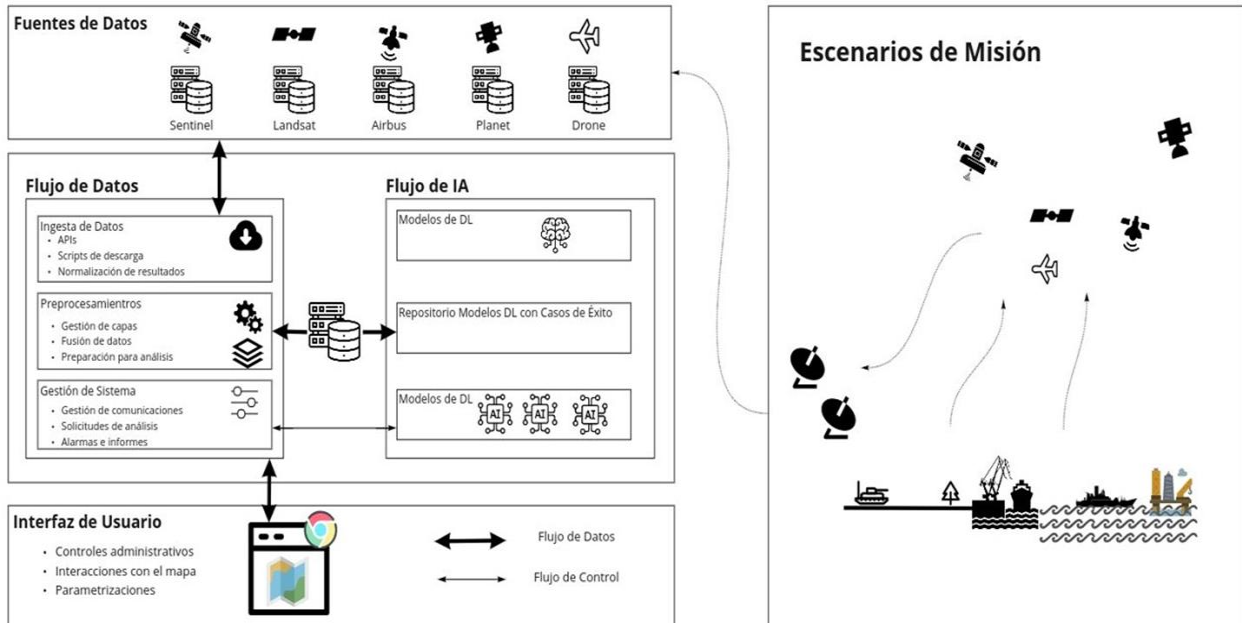
La arquitectura lógica del sistema se basa en tres dispositivos con sus respectivas tareas asociadas: “Flujo de procesamiento de Datos”, “Sistema de almacenamiento” y “Flujo de Deep Learning”. También se incluyen partes ajenas al sistema como: “PC del usuario”, protocolo de comunicaciones y fuentes de datos satelitales incluidas en entorno “Cloud” a las que se accede para obtener las imágenes.

### 2.1.1 Sistema “Flujo del procesamiento de datos”

Servidor que ejecuta los procesos de control principal del sistema así como los de comunicación con las fuentes de datos satelitales, el cliente y el resto de sistemas.

### 2.1.1 Sistema “Flujo de Deep Learning”

Servidor que ejecuta los procesos de análisis y entrenamiento para resolver con éxito los casos de uso expuestos.



Esquema arquitectura lógica del sistema 1

### 2.1.2 Sistema “Almacenamiento”

Servidor NAS para almacenamiento y protección de los datos satelitales.

## 2.2 Modularización de los modelos

En esta sección se describen todos los modelos que se desplegarán en el sistema, especificando para cada uno, el tipo de datos de entrada y salida, y una o varias arquitecturas de modelos Deep Learning posibles, para ir conformando un modelo listo para ser desplegado.

En los diferentes modelos, usamos arquitecturas que comparten partes entre los modelos. Estas arquitecturas se conocen como de aprendizaje multi-tarea.

### 2.3 Detección y segmentación de objetos y estructuras

El primer paso es la detección de los elementos de interés en las imágenes. Por tanto, es necesario entrenar un modelo para cada uno de los casos de uso.

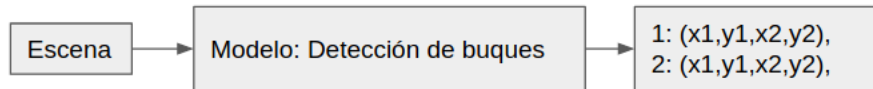
Estos modelos tienen una primera parte (backbone) que toma las imágenes de entrada y los metadatos extrayendo un mapa de características de las mismas, que luego es utilizado por cada una de las partes finales de los modelos correspondientes a los casos de uso.

Los casos de uso que requieran de una clasificación de barcos/estructuras, toman el resultado de éstas detecciones como input para clasificar los elementos de las imágenes

### 2.3.1 Modelo de Detección de barcos (MM-1.1)

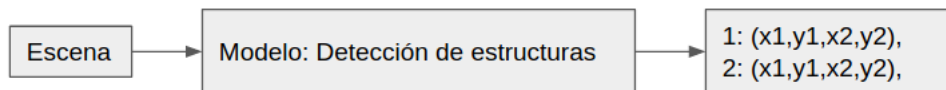
Este modelo compartirá el backbone común y detectará los buques que aparezcan en la imagen satelital. Se trata de un modelo básico para la resolución de los casos de uso y es en gran parte el primer paso necesario para analizar las imágenes.

Los datos de entrada a este modelo son imágenes satelitales + metadatos y los datos de salida son regiones de interés o bounding boxes que sitúan los barcos en la imagen. En definitiva el backbone común para extrae las características, y un módulo llamado “cabeza” propone las regiones de interés y devuelve la detección de los barcos (coordenadas de los bounding boxes).



### 2.3.2 Modelo de Detección de estructuras (MM-1.2)

Similar al anterior, usa backbone común, el objetivo de este modelo es la detección de estructuras en una imagen. Los datos de salida corresponden a la localización en la imagen de las estructuras detectadas. Este modelo tiene arquitectura similar al modelo de detección de buques.



### 2.3.3 Modelo de Segmentación de barcos (MM-1.3)

Para mayor resolución en las detecciones y mejor detección de cambios (salidas de puertos, construcciones, etc), usamos modelos de segmentación semántica para complementar el análisis de las imágenes. Los datos de entrada del modelo son imágenes satelitales + metadatos y la salida son predicciones para los diferentes píxeles.

Al tener un modelo de segmentación de estructuras entrenado hacemos un seguimiento de determinadas zonas en busca de nuevos elementos, cambios graduales como la construcción de nuevas estructuras, etc.

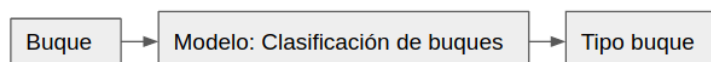
## 2.4 Clasificación de barcos y estructuras

Los modelos de clasificación, con un backbone común, establecen la categoría de las imágenes generadas por los modelos de detección. Gran parte de los casos de uso implicará la categorización de los elementos de las imágenes.

### 2.4.1 Modelo de Clasificación de barcos (MM-2.1)

Input: Imagen Satelital (Output MM-1.1) + Metadatos

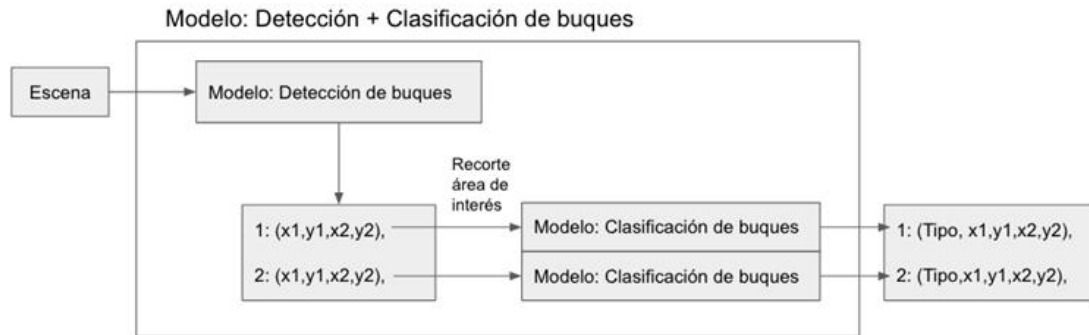
Output: Probabilidad del barco de la imagen de pertenecer a cada una de las clases.



Los tipos de barcos que se desean identificar en una primera versión son los siguientes:

- Buques mercantes: carga general, Ro-Ro (transporte de carga rodada y/o de personas), transportador o ferry o pesqueros
- Buques militares: portaaviones, submarinos, destructores, fragatas, corbetas, dragaminas, guardacostas o patrulleras

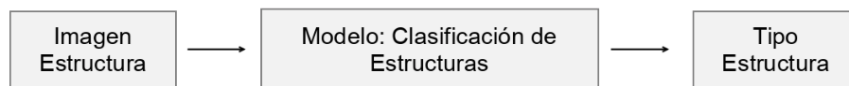
Combinando ambos modelos (Detección y Clasificación de buques), al procesar automáticamente una imagen satelital, obtenemos la ubicación de los barcos detectados en la imagen (RoIs, Region of Interest) y la clasificación de dichos barcos.



### 2.4.2 Modelo de Clasificación de estructuras (MM-2.2)

In Input: Imagen Satelital + Metadatos

Output: Probabilidad del objeto de la imagen de pertenecer a cada una de las clases de estructuras.



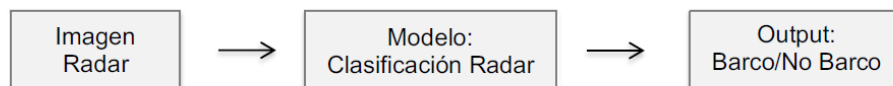
Para el modelo de clasificación de estructuras, se usa una arquitectura estándar de clasificación como EfficientNet<sup>1</sup>, ResNet<sup>2</sup>, Inception<sup>3</sup>, MobileNet<sup>4</sup>, etc. Las clases en las que incluir las estructuras, podrán ser boyas, plataformas petroleras y/o aerogeneradores.

### 2.4.3 Modelo de Clasificación de Áreas de Interés en Imágenes Radar (MM-2.3)

Input: Imagen Satelital de tipo radar + Metadatos

Output: Probabilidad de pertenecer a una clase predefinida

Para las imágenes radar elaboramos un modelo de clasificación de barco/no barco.



El modelo toma como entrada una imagen radar + metadatos y como salida predice la probabilidad de que exista un barco en la imagen

<sup>1</sup> Mingxing Tan, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

<sup>2</sup> Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition

<sup>3</sup> <https://doi.org/10.48550/arXiv.1512.00567>

<sup>4</sup> <https://doi.org/10.48550/arXiv.1704.04861>

Debido a la gran diferencia entre las imágenes radar y RGB, este modelo usa una red neuronal aparte de la de los modelos de clasificación para imágenes RGB, de arquitectura parecida, basada en EfficientNet pero comparten el mismo *backbone*.

#### 2.4.4 Modelo de Detección de Cambios (MM-3)

Input: par de imágenes satelitales.

Output: máscara de los píxeles donde se considera que existen cambios sustanciales en la imagen.

Se realiza un análisis de la escena mediante un modelo de segmentación semántica de entornos costeros, pudiendo ser distinto a los modelos de segmentación MM-1.3 y MM-1.4 con el objetivo de separar las clases a predecir y ganar precisión. Una vez segmentado el par de imágenes, se utiliza un modelo capaz de detectar cambios entre las dos máscaras de segmentación mediante análisis de imágenes del mismo lugar en distintos momentos en el tiempo.

Para segmentar las imágenes tomadas en distintos momentos, podemos utilizar los modelos de segmentación de barcos y estructuras definidos anteriormente.

#### 2.4.5 Modelo de comparación de buques (MM-4)

Input: par de imágenes RGB

Output: probabilidad de que el buque sea el mismo

Este modelo se encarga de decir si un buque es similar a otro. Así, teniendo una base de datos de buques objetivos a buscar, se puede comparar y detectar si son parecidos o no. Para buques parecidos devuelve un valor cercano a 1, mientras que para buques diferentes devuelve valores cercanos a 0.

Este modelo se basa en arquitecturas de redes siamesas, donde se codifican los buques con la misma red y se comparan sus mapas de características para ver lo diferentes que son estos buques.

Tabla resumen de modelos de red neuronal

Identificador de modelo	Descripción
MM-1.1	Modelo de detección de barcos.
MM-1.2	Modelo de detección de estructuras.
MM-1.3	Modelo de segmentación de barcos.
MM-1.4	Modelo de segmentación de estructuras.
MM-2.1	Modelo de clasificación de barcos.
MM-2.2	Modelo de clasificación de estructuras
MM-2.3	Modelo de clasificación de áreas de interés en imágenes radar
MM-3	Modelo de segmentación de puertos y zonas costeras
MM-4	Modelo de comparación de buques

Resumen modelos.

### 2.5 Casos de uso

Aquí se detalla qué modelos de red neurona se usa en cada caso de uso y la denominación y explicación intuitiva de cada caso de uso.

La siguiente tabla sirve a modo de resumen.

Caso de uso	Denominación del caso de uso	Modelo usado
CU-01	Reconocimiento de tipo de buque	MM-1.1, MM-2.1
CU-02	Identificación de buque específico	MM-1.1, MM-2.1

CU-03	Detección zonal de buques y alertado	MM-1.1, MM-2.3
CU-04	Detección de buques abarloados en la mar	MM-1.1
CU-05	Detección de estructuras	MM-1.2, MM-2.2
CU-06	Seguimiento de trazas y obtención de histórico	MM-1.1, MM-2.1
CU-07	Obtención de mapas de calor	MM-1.1
CU-08	Alerta por salidas de puerto y fondeos	MM-1.1, M-2.1
CU-09	Detección de cambios en puertos o línea de costa	MM-3
CU-10	Detección emplazamientos artillería de costa o A/A	MM-1.4, MM-2.2

Casos de uso y modelos que intervienen en cada caso de uso.

### 3. Resultados y discusión

Pese al tremendo progreso en el campo de la detección de objetos, la tecnología sigue siendo significativamente más primitiva que la visión humana y aún no puede abordar satisfactoriamente los desafíos del mundo real. Los progresos en este campo son vertiginosos y mientras se escribe este artículo acontecen progresos que determinan futuras líneas de investigación, no tenidas aquí en cuenta, que acercan aún más la visión artificial a la humana.

#### 3.1 Conclusiones

Tras evaluar las diferentes estructuras de detección, se enumeran los factores clave que han surgido de los diferentes estudios y desarrollos dentro de la detección genérica de objetos basada en el aprendizaje profundo.

#### Estructuras de detección seleccionadas

- Cuando el coste computacional no es el factor limitante, los detectores de dos etapas generalmente producen precisiones de detección más altas que los de una etapa, porque su estructura es más flexible y más adecuada para la clasificación basada en regiones. Los marcos más utilizados son Faster R-CNN<sup>5</sup>, R-FCN<sup>6</sup> y Mask R-CNN<sup>7</sup>. En nuestro caso se ha elegido este tipo de detectores casi de manera exclusiva, salvo en los casos que implicaban detección de cambios donde usamos estructuras basadas en redes siamesas.
- Las configuraciones de una etapa (e.g. YOLO<sup>8</sup> y SSD<sup>9</sup>) suelen tener un rendimiento mucho más bajo en la detección de objetos pequeños que las arquitecturas de dos etapas como Faster R-CNN y R-FCN, pero son competitivos en la detección de objetos grandes.

#### Mejora de la fiabilidad de la representación de objetos frente a circunstancias adversas.

La variabilidad de las imágenes del mundo real (e.g. iluminación, posturas, deformaciones, desorden de fondo, oclusiones, desenfoque, resolución, ruido y distorsiones de cámara) son un desafío clave en el reconocimiento de objetos.

#### Procesamiento de entorno del objeto.

Queda por explorar cómo incorporar información contextual de manera eficiente.

#### Propuestas de detección.

Las propuestas de detección reducen significativamente los espacios de búsqueda en comparación con las propuestas de región.

<sup>5</sup> (Ren et al. 2015)

<sup>6</sup> (Dai et al. 2016c)

<sup>7</sup> (He et al. 2017)

<sup>8</sup> Redmon et al. 2016

<sup>9</sup> Liu et al. 2016

## Otros factores.

Por ejemplo, aumento de datos, estrategias de entrenamiento novedosas, combinaciones de modelos de redes troncales, marcos de detección múltiples, incorporación de información de otras tareas, métodos para reducir el error de localización, etc.

## 1.1.Líneas futuras de investigación

En cuanto a los problemas que quedan por resolver y que marcan las tendencias futuras se prevén las siguientes direcciones de investigación:

- Aprendizaje de mundo abierto: Capacitar a los algoritmos para reconocer categorías de objetos fuera de su conjunto de datos de entrenamiento, para mejorar comportamiento en ambientes reales.
- Conseguir estructuras de detecciones mejores y más eficientes. Los detectores basados en regiones tienen mayor precisión, los detectores de una etapa son generalmente más rápidos y simples. Los detectores de objetos dependen en gran medida de las redes troncales subyacentes, que se han optimizado para clasificación de imágenes.
- Diseñar funciones de CNN compactas y eficientes. Las CNN tienen millones de parámetros, lo que requiere datos masivos, memoria y capacidad de cálculo para el entrenamiento. Se necesitan mayores esfuerzos en compactación de las redes (e.g. mediante uso compartido de fases o módulos) para aprovechar más la capacidad de cómputo.
- Búsqueda automática de arquitecturas neuronales mediante programas de diseño automático.
- Detección débilmente supervisada. A día de hoy se usan modelos totalmente supervisados con datos etiquetados con cuadros delimitadores de objetos o máscaras de segmentación. Sería precisa diseñar CNNs que solo usen datos anotados débil o parcialmente.
- Detección de objetos con pocos o cero intentos. La capacidad de aprender de unos pocos ejemplos y la detección en pocos intentos, es otro de los campos de mejora para el futuro. Otra línea futura de investigación es la mejora en el reconocimiento de objetos nunca antes vistos, entrenados o conocidos, por tanto como intento “0”.
- Detección de objetos en otras modalidades. La mayoría de los detectores usan imágenes fijas en 2D; la detección en otras modalidades puede ser muy relevante para vehículos autónomos, vehículos aéreos no tripulados y robótica.
- Detección universal de objetos. Detección de cualquier clase de objeto, no sólo aquellos en los que se ha especializado la red.

## Referencias

1. Mingxing Tan, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks
2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition
3. <https://doi.org/10.48550/arXiv.1512.00567>
4. <https://doi.org/10.48550/arXiv.1704.04861>
5. (Ren et al. 2015)
6. (Dai et al. 2016c)
7. (He et al. 2017)
8. Redmon et al. 2016
9. Liu et al. 2016