

# Práctica 2: Trabajo de Investigación

Procesamiento de Imagen, Vídeo y Audio

Grado en Ciencia e Ingeniería de Datos

Curso 2022/2023

*Borja Souto Prego*

## Esquema global del método

El esquema global del método consiste en un modelo xgboost que, primeramente, como baseline, utiliza únicamente los canales RGB propios de las imágenes.

Tras realizar un preprocesado de las imágenes y adaptar los conjuntos de datos para que tengan la forma que espera xgboost, a partir de este modelo se irán añadiendo características, en forma de nuevos canales, de forma que aumente su complejidad, además del ajuste de los parámetros (como el número de estimadores o la profundidad máxima) con la intención de obtener mejores resultados.

Tras cada actualización del modelo se realiza una evaluación cuantitativa de la segmentación a través de diferentes métricas. Si las métricas se mejoran, las características añadidas a dicho modelo se mantienen, en caso contrario se descartan.

## Ejercicio 1: Segmentación de carreteras en imagen aérea de alta resolución

### Soluciones planteadas para cada fase

#### Preprocesamiento

Como método de preprocesamiento, se dividen las imágenes entre 255 para llevar los valores de los píxeles a un rango de 0 a 1, lo que facilita el procesamiento.

#### Primera fase

Lo primero que se trata con la intención de mejorar el baseline es de eliminar el ruido en las imágenes de entrada para añadirlo como una nueva característica. Esto se realiza a través de la utilización del filtro de medias y el filtro de la mediana. Comparando los resultados de las métricas obtenidos con ambos, nos quedamos con el filtro de medias.

#### Segunda fase

En una segunda fase se prueba a agregar los canales HSV y/o LAB para comprobar si puede mejorar la robustez frente a cambios de iluminación, lograr una mejor separación de objetos y reducir la correlación entre los canales, lo que puede conducir a una segmentación más precisa y eficaz de las carreteras en las imágenes. Estos canales se añaden sobre las imágenes anteriores, resultantes de añadir los canales rgb con eliminación de ruido.

Tras probar a añadir los canales HSV, los canales LAB y los canales HSV y LAB juntos, nos quedamos con la combinación de ambos en base a los resultados obtenidos en las métricas.

## **Tercera fase**

En esta tercera fase se realizó detección de bordes mediante el operador de Canny, el filtro de Laplaciano y el filtro de Sobel, añadiendo los bordes detectados como otro nuevo canal. Para estas tres aproximaciones los resultados obtenidos han sido altamente similares.

## **Métricas**

Las métricas utilizadas para la evaluación de las diferentes fases han sido: Accuracy, F1-score, Precision, Jaccard y Recall.

Además de las métricas, se muestran también las imágenes predichas obtenidas para cada fase, para complementar los resultados de forma visual.

# **Ejercicio 2: Reconocimiento de objetos**

## **Soluciones planteadas para cada fase**

### **Preprocesamiento**

Para el preprocesamiento de las imágenes, además de dividir las entre 255 para normalizarlas al igual que en el ejercicio anterior, también convertimos las imágenes que están en escala de grises, a diferencia de la mayoría, a rgb, para tener todas en un mismo formato de 3 canales, y las reescalamos a un tamaño fijo (al tamaño máximo de todas las imágenes, 300x300).

### **Primera fase**

En una primera fase, eliminamos el fondo de las imágenes para que el modelo se centre en los objetos que queremos detectar. Para ello, utilizaremos la máscara de la imagen, que nos indica qué píxeles pertenecen al fondo y cuáles a los objetos. En base a las métricas se puede observar que se mejoran los resultados, por lo que se conserva esta característica en nuestro modelo.

### **Segunda fase**

Eliminamos el ruido en las imágenes de entrada para añadirlo como una nueva característica, de forma similar al ejercicio anterior. Probamos con el filtro de la mediana y el filtro de medias.

Ambos mejoran los resultados obtenidos en las métricas, pero nos quedamos con el filtro de la mediana que obtiene unas mejoras mayores.

### **Tercera fase**

A continuación, al igual que en el ejercicio anterior, probamos a agregar los canales HSV y/o LAB. Los espacios de color HSV y LAB se asemejan más a cómo los humanos percibimos y describimos los colores. Al incorporar estos canales, el modelo puede capturar mejor las características visuales relevantes para la detección de animales, alineándose más con la intuición humana.

De las tres aproximaciones, nos quedamos con la que añade únicamente los canales HSV, pues es la que mejores resultados nos proporciona en las métricas.

## **Cuarta fase**

Como última fase se aplica el laplaciano, por un lado, en una de las aproximaciones, a los canales por separado (los diferentes canales pueden contener información de bordes o características distintas, al aplicar el Laplaciano a cada canal por separado, pueden resaltar diferentes tipos de bordes o características en la imagen), y por otro lado, a las imágenes completas (esto puede ser útil si los canales de color están altamente correlacionados y la información de bordes o características está distribuida en todo el espectro de color).

Ninguna de estas aproximaciones mejora los resultados obtenidos en las métricas, por lo que las descartamos.

## **Métricas**

Las métricas utilizadas para la evaluación de las diferentes fases han sido: Accuracy, F1-score, Precision y Recall.

La F1-score, Precision y Recall se muestran tanto para cada clase como para el conjunto total de todas las clases.