

FACULTAD DE INFORMÁTICA

# PRÁCTICA 2

---

## Procesamiento de Imagen, Vídeo y Audio

**Carlos Hermida**

**3º GCID – Curso 2023/2024**



**UNIVERSIDADE DA CORUÑA**

# EJERCICIO 1

Para el problema de segmentación de carreteras se entrena un modelo xgboost con las imágenes satélite preprocesadas con varios métodos diferentes.

El modelo estima la probabilidad de que un píxel pertenezca a una carretera, y se aplica un umbral para determinar el resultado. Por ejemplo, un umbral de 0.6 significa que serán carretera aquellos píxeles con un 60% de probabilidad o más.

Además, en algunas aproximaciones, se postprocesa la salida del modelo para mejorar el resultado.

Se evalúa cualitativamente mediante la visualización, y cuantitativamente con las métricas:

- Accuracy: porcentaje de píxeles iguales en el ground truth y predicción.
- Precision: porcentaje de píxeles no carretera clasificados correctamente.
- Recall: porcentaje de píxeles carretera clasificados correctamente.
- f1-score: media armónica que combina los valores de precisión y recall.
- Jaccard: grado de similitud entre el ground truth y predicción.

## 1. Baseline

Para la aproximación base se entrena el modelo con las imágenes en escala de gris. Se aplica un umbral de 0.6 al resultado del modelo, porque un umbral más alto no mostraría ninguna carretera. Se obtiene un mal resultado porque detecta como carretera demasiados elementos que no son carretera. Además, en la gráfica de entrenamiento se observa que el modelo no generaliza bien.

## 2. Aproximación 1

En las imágenes satélite se aprecia que las carreteras generalmente son de color gris. Por ello, se entrena al modelo con las imágenes a color y se aplica un umbral de 0.85 al resultado del modelo, porque es el valor con el que se consigue más píxeles clasificados correctamente. Se mejora el resultado anterior.

## 3. Aproximación 2

Dado que RGB funcionó bien, se exploran otros espacios de color. Entrenar al modelo con HSV o Lab completos ofrecería resultados muy similares a la aproximación anterior, ya que estos espacios de color son transformaciones no lineales de RGB. Por ello, se buscan aquellos canales que segmenten mejor.

Tras probar varios, se observa que el canal de saturación de HSV diferencia bien las carreteras, debido a su color gris (baja saturación). Además, la componente b de Lab también funciona bien, por lo que se entrena al modelo pasándole estos dos canales (S y b). Sin embargo, los resultados empeoran.

## 4. Aproximación 3

El mejor resultado hasta el momento se obtuvo con la aproximación 1, por lo que se utilizará como base para añadir complejidad al preprocesado. En este caso, a parte de los canales RGB, se calcula el

laplaciano de la imagen en escala de gris para obtener los bordes, pero se observa que los bordes de las carreteras se aprecian mejor si se aplica el laplaciano sobre el canal de saturación de HSV utilizado anteriormente. Se mejora el resultado.

## 5. Aproximación 4

El resultado anterior es bueno, pero visualmente se aprecia ruido en la imagen de salida. Por lo que en esta aproximación se mantiene el preprocesado con los canales RGB y el laplaciano del canal saturación, y se añade postprocesado a la salida del modelo. Se utiliza un filtro de mediana, que es efectivo contra ruido de tipo sal y pimienta. Se prueban diferentes tamaños hasta obtener el mejor equilibrio entre precisión y recall. Se mejora el resultado.

## 6. Aproximación 5

A pesar de que el resultado anterior es bueno, se prueba otro método diferente de postprocesado. En este caso, se refina la salida predicha con aperturas con operadores morfológicos rectangulares.

En primer lugar, se aplica una erosión para quedarnos con los elementos rectangulares (las carreteras son rectangulares), y a través de una dilatación se recupera su tamaño. Se utiliza la apertura tanto horizontal como verticalmente, y se combinan ambas salidas para el resultado final.

En este caso, bajar el umbral de predicción a 0.6 posibilita que la operación de apertura mantenga el mayor número de carreteras. Se mejora ligeramente el resultado anterior.

## 7. Combinación

En esta última aproximación, se combinan las entradas de todas las aproximaciones (escala de gris, canales RGB, canal S de HSV, canal b de Lab, laplaciano del canal S de HSV) y se postprocesa la salida con operadores morfológicos y un filtro de mediana para eliminar imperfecciones.

Esta combinación obtiene los mejores resultados tanto cuantitativa como cualitativamente.

	aproximación	ciclos	duración (s)	threshold	accuracy	precision	recall	f1	jaccard
0	baseline	12	14.39	0.60	0.8123	0.1316	0.4690	0.2055	0.1145
1	aproach_1 (RGB)	100	78.45	0.85	0.9241	0.3521	0.5544	0.4306	0.2744
2	aproach_2 (H[S]V + La[b])	100	74.51	0.85	0.9161	0.3015	0.4720	0.3680	0.2255
3	aproach_3 (RGB + laplace(H[S]V))	100	85.55	0.85	0.9379	0.4259	0.5712	0.4880	0.3227
4	aproach_4 (RGB + laplace(H[S]V) + median_filter)	100	85.02	0.85	0.9504	0.5197	0.5493	0.5341	0.3643
5	aproach_5 (RGB + laplace(H[S]V) + morph)	100	86.12	0.60	0.9426	0.4613	0.6530	0.5407	0.3705
6	combination	100	115.49	0.60	0.9505	0.5181	0.6341	0.5702	0.3988



## EJERCICIO 2

Para el problema de clasificación de animales se entrena un modelo xgboost con las imágenes de los animales preprocesadas con varios métodos diferentes. El modelo asigna la etiqueta con mayor probabilidad, siendo: 0 --> elefante / 1 --> rinoceronte / 2 --> otros.

Se evalúa cuantitativamente atendiendo a la matriz de confusión y con las métricas:

- Accuracy: porcentaje individuos clasificados correctamente.
- Precision:  $VP / (VP + FP)$
- Recall:  $VP / (VP + FN)$
- f1-score: media armónica que combina los valores de precisión y recall.

### 0. Baseline

Para la aproximación base se entrena el modelo con las imágenes en escala de gris. Se obtiene un accuracy de 48%.

### 1. Aproximación 1

Se elimina el fondo de las imágenes en escala de gris utilizando la máscara dada, pues el fondo no es relevante para el problema de clasificación. Se obtiene un accuracy de 76%.

### 2. Aproximación 2

Dado que los rinocerontes y los elefantes tienen un color diferente a los otros dos animales, se utilizan los canales RGB sin el fondo. Sin embargo, el resultado empeora.

### 3. Aproximación 3

El mejor resultado hasta el momento se obtuvo con la aproximación 1, por lo que se utilizará como base para añadir complejidad al preprocesado. Se le resaltan los bordes, detectados con el algoritmo de Canny. De esta manera, se utilizan las formas de cada animal para distinguirlos mejor. Por ejemplo, las orejas de los elefantes o las alas de los flamencos. El resultado empeora.

### 4. Aproximación 4

Con la operación morfológica skeletonize se obtiene un esquema de la estructura del animal, para que el modelo aprenda la forma genérica de cada uno. Por ejemplo, el cuello en forma de 'S' de los flamencos, las patas de los rinocerontes y elefantes, o la trompa de este último. Se anhea este esqueleto con una operación de dilatación para que sea más fácil de identificar. El resultado no mejora.

## 5. Aproximación 5

A la imagen en escala de gris sin fondo, se le aplica cuatro umbrales con Otsu para diferenciar tres regiones diferentes en cada animal, y de esta forma destacar elementos clave como los colmillos de los elefantes o los cuernos de los rinocerontes. El resultado empeora.

## 6. Combinación

En esta última aproximación, se combinan las entradas de todas las aproximaciones (escala de gris, escala de gris sin fondo, canales RGB sin fondo, bordes, esqueleto y regiones con multi-umbralización). Esta combinación no mejora los resultados de la primera aproximación.

	aproximación	ciclos	duración (s)	accuracy	precision (mean)	recall (mean)	f1 (mean)
0	baseline	31	36.24	0.48	0.40	0.40	0.39
1	aproach_1 (grayscale + mask)	18	14.04	0.76	0.76	0.74	0.75
2	aproach_2 (RGB + mask)	17	39.65	0.56	0.53	0.51	0.51
3	aproach_3 (grayscale + mask + Canny)	13	10.72	0.64	0.62	0.64	0.62
4	aproach_4 (grayscale + mask + skeleton)	20	11.46	0.72	0.79	0.70	0.70
5	aproach_5 (grayscale + mask + otsu)	17	6.10	0.64	0.60	0.60	0.58
6	combination	18	246.33	0.60	0.55	0.54	0.53
7	aproach_(?)	35	14.43	1.00	1.00	1.00	1.00

Sin embargo, probando diferentes métodos para aplicar en las aproximaciones, al utilizar un filtro de medias con un sigma bastante alto ( $\sigma=10$ ) sobre las imágenes en escala de gris sin fondo, se obtuvo un accuracy del 100%. Pudiendo entender este hecho como fruto de la casualidad, se probó con diferentes semillas a la hora de hacer el split en train, val y test; y curiosamente estas imágenes extremadamente suavizadas continuaban dando buenos resultados. Además, su gráfica de entrenamiento es la única donde se observa que el modelo aprende y no sobreentrena como en el resto de aproximaciones.