

# Práctica 2: Trabajo de Investigación

## Procesamiento de Imagen, Vídeo y Audio

Grado en Ciencia e Ingeniería de Datos

Curso 2024/2025

Esta práctica consiste en aplicar los conocimientos adquiridos sobre las técnicas de procesado de imagen para desarrollar una metodología que permita abordar 2 problemas planteados. Para ello, se hará uso de técnicas vistas en la asignatura. Las soluciones propuestas se codificarán en **Python 3.x** con las bibliotecas de procesamiento **Numpy**, **SciPy**, **Scikit-image** y **Scikit-learn**.

- La práctica deberá realizarse de forma **individual**.
- No está permitido el uso de versiones de Python 2.x.
- Está permitido el uso de cualquier técnica y librería de procesado de imagen disponible que permita abordar el problema, siempre que se demuestre el conocimiento de la técnica utilizada y se justifique el motivo por el que se usa.
- Los sistemas desarrollados deben incluir planteamientos basados en extracción de características y aprendizaje automático, supervisado y/o no supervisado.

## Entrega y Evaluación

- Para poder superar esta práctica es necesario **entregar**:
  1. Software de procesado de imágenes desarrollado que permita ver resultados parciales.
  2. Ejemplos de resultados precalculados (solo si el software es muy lento).
  3. Pequeño documento de memoria (1-2 folios), que indique lo siguiente, a modo de guion:
    - Esquema global del método, incluyendo fases de preparación de datos, ajuste (entrenamiento), inferencia y evaluación.
    - Las soluciones planteadas para cada fase, tanto descartadas como seleccionadas para la solución final, así como los criterios cuantitativos y/o cualitativos que han llevado al planteamiento de estos métodos y su selección. Esto puede incluir, por ejemplo:
      - Métodos de preprocesamiento y postprocesamiento
      - Características visuales utilizadas
      - Metodología de entrenamiento y evaluación, incluyendo preparación de conjuntos de entrenamiento, validación y test
      - Métodos de aprendizaje automático utilizados

- Metodología de inferencia, que permita obtener el resultado esperado, usando los módulos entrenados, a partir de de la entrada en crudo.
- La fecha límite de entrega será anunciada y notificada en el Campus Virtual.
- El boletín consta de 2 ejercicios, cada uno de los cuales supone una nota máxima de 5 puntos sobre un total de 10 puntos.
- Todos los entregables anteriormente mencionados deben subirse al moodle. No se aceptarán entregas a través de enlaces a otros servicios de almacenamiento en la nube. No entregar las imágenes de entrenamiento facilitadas con este enunciado.
- Se concertará una cita para la defensa oral de la práctica en la que se mostrará el software de la práctica funcionando en un portátil, junto con el documento de memoria.
- El alumnado que presente códigos y/o respuestas con indicios de plagio (en cualquiera de las respuestas y con respecto a otras prácticas presentadas este curso o en cursos anteriores, o recursos online) obtendrá una calificación de Suspenso en la parte práctica (calificado con 0.0), independientemente de la nota que pudiera merecer la calidad de la práctica, e independientemente de si las personas implicadas resultan ser plagiadas o plagiadoras.

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

En la actualidad existen multitud de aplicaciones que hacen uso de imágenes aéreas y por satélite, entre las que se incluye la generación y actualización automática de mapas de carreteras para navegadores GPS y sistemas de mapas online. La segmentación automática de estas imágenes es un paso necesario para la automatización de dichos procesos.

## 1.1. Materiales

Para este ejercicio se usará un subconjunto de imágenes provenientes del Massachusetts Roads Dataset [1,2]. El subconjunto de imágenes seleccionado está disponible en el Moodle, aunque adicionalmente se pueden descargar más imágenes de la fuente original si se considera conveniente.

- Las imágenes se corresponden con imágenes aéreas de alta resolución de áreas urbanas y suburbanas.
- Para cada imagen se proporciona una máscara de ground truth (salida esperada) con las carreteras segmentadas.



## 1.2. Objetivos

Desarrollar un método computacional que permita:

- Segmentar las carreteras existentes en una imagen de entrada. Es decir, dada una imagen aérea, proporcionar una máscara binaria con los píxeles de carretera a 1 y el resto a 0<sup>1</sup>.
- Para un conjunto de imágenes de prueba, proporcionar un conjunto de métricas de evaluación cuantitativa de la segmentación apropiadas para la aplicación objetivo, usando el ground truth correspondiente, .

## 1.3. Referencias

- [1 ] V. Mnih, Massachusetts Roads Dataset,  
<https://www.cs.toronto.edu/~vmnih/data/>

---

<sup>1</sup>Es decir, clasificar cada pixel, incluso considerando características que usan una vecindad local entorno a este, entre carretera o fondo

- [2 ] V. Mnih, “Machine Learning for Aerial Image Labeling”, Ph.D. Thesis, University of Toronto, 2013.

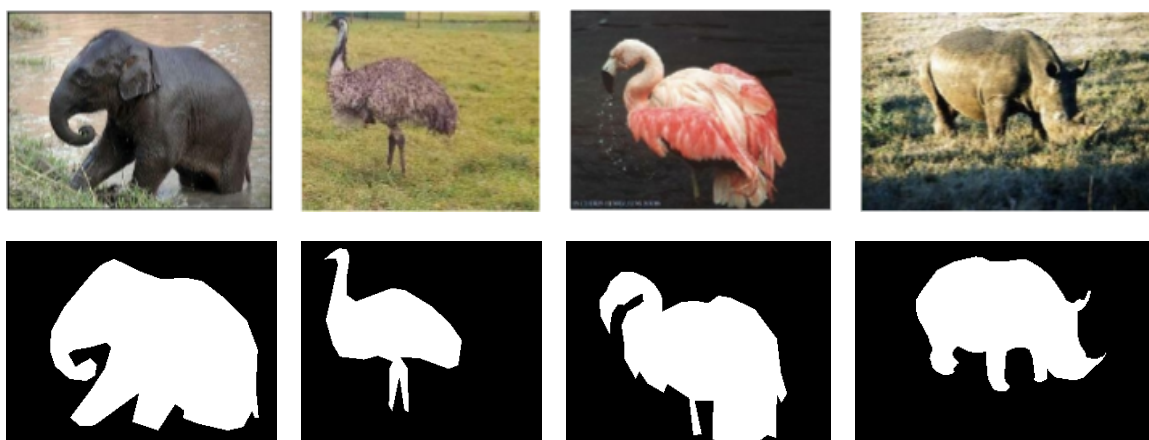
## 2. Ejercicio 2: Reconocimiento de objetos

El reconocimiento de objetos es un subproblema típico de la mayoría de sistemas de visión por computador, con aplicaciones en multitud de ámbitos. Normalmente, el reconocimiento se combina con fases previas, o simultáneas, de detección y segmentación, dando lugar a sistemas muy complejos. Sin embargo, típicamente, en todos estos sistemas, una vez localizados los candidatos, será necesario reconocer la clase a la que pertenece el objeto frente a otras categorías planteadas o el fondo (ninguna categoría). En este ejercicio nos centramos en este subproblema, mediante un planteamiento simplificado; lo que no deja de suponer un reto significativo.

### 2.1. Materiales

Para este ejercicio se usará un subconjunto de imágenes provenientes del Caltech-101 Dataset [1]. El conjunto de datos Caltech-101 contiene imágenes de 101 categorías de objetos y una categoría de fondo que contiene las imágenes que no pertenecen ninguna de las categorías de objetos. Para cada categoría de objetos, hay entre 40 y 800 imágenes, mientras que la mayoría de las clases tienen unas 50 imágenes. La resolución de la imagen es de unos  $300 \times 200$  píxeles. El subconjunto de imágenes seleccionado, disponible en el Moodle, contiene una selección de 4 clases (“elephant”, “rhino”, “emu” y “flamingo”), aunque adicionalmente se pueden descargar más imágenes de la fuente original si se considera conveniente. Para cada imagen, se proporciona:

- Su clase, correspondiendo a la carpeta donde se encuentra la imagen (salida esperada).
- Una máscara con el recorte del objeto de interés con respecto al fondo, de uso opcional como entrada adicional o alternativa.



### 2.2. Objetivos

Desarrollar un método computacional que permita:

- A partir de una imagen (y/o, opcionalmente, su máscara de recorte), identificar si pertenece a la categoría<sup>2</sup> “elephant”, o a la categoría “rhino”

---

<sup>2</sup>Es decir, clasificar cada imagen, considerando características integradas para toda la imagen, región recortada y/o contorno, entre, por ejemplo, clase objetivo y resto

- Proporcionar una evaluación cuantitativa de los objetivos del sistema, usando un conjunto de datos de prueba y una metodología de validación adecuada.

### **2.3. Referencias**

- [1 ] Fei-Fei Li, Marco Andreetto, and Marc Aurelio Ranzato, Caltech-101 Dataset,  
[http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)