

# Clasificación de imágenes y asignación de textos mediante redes neuronales convolucionales y transformers multimodales

Trabajo Fin de Grado

Pablo García García

Grado en Ingeniería Informática

Universidad de Alcalá  
Escuela Politécnica Superior

16 de julio de 2024

# Índice general

- 
- 1 Introducción
  - 2 Redes neuronales convolucionales
  - 3 Transformers multimodales
  - 4 Conclusiones y trabajo futuro

# Introducción

## Niantic Wayfarer



**Información de la propuesta**  
Consulta a continuación la información de la propuesta

**Título y descripción**  
El título y la descripción no deben contener texto inapropiado. La descripción no es obligatoria.

**Estatua Víctimas del terrorismo**

Estatua del escultor Juan Carlos Díaz Durán en honor a las Víctimas del Terrorismo.



# Introducción

## Objetivos

- Clasificación de una **imagen** en un **conjunto cerrado** de clases
- Clasificación de una **imagen** en un **conjunto abierto** de clases
- **Búsqueda** de imágenes mediante una descripción **textual**
- Selección del **título** o **descripción** más adecuado para una **imagen**
- Aproximación a la detección de **imágenes** que contienen un **mismo objeto**

# Proceso ETL

## Carga y transformación

### Algunos problemas...

- No existe API
- Extracción y etiquetado manual
- Pocos datos
- Clases desbalanceadas

camino\_santiago



parque



parque

parque



cartel



cartel

cartel



marcador

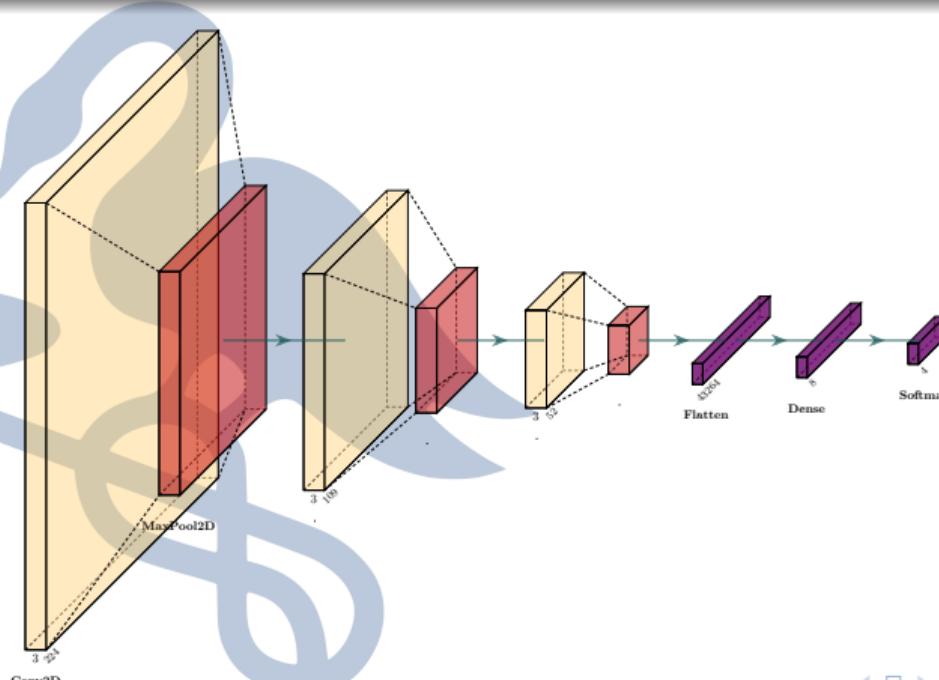


marcador



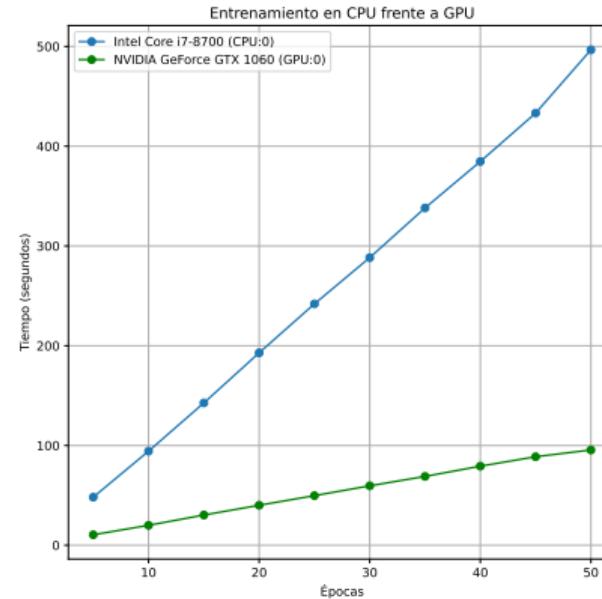
# Red neuronal convolucional

## Arquitectura



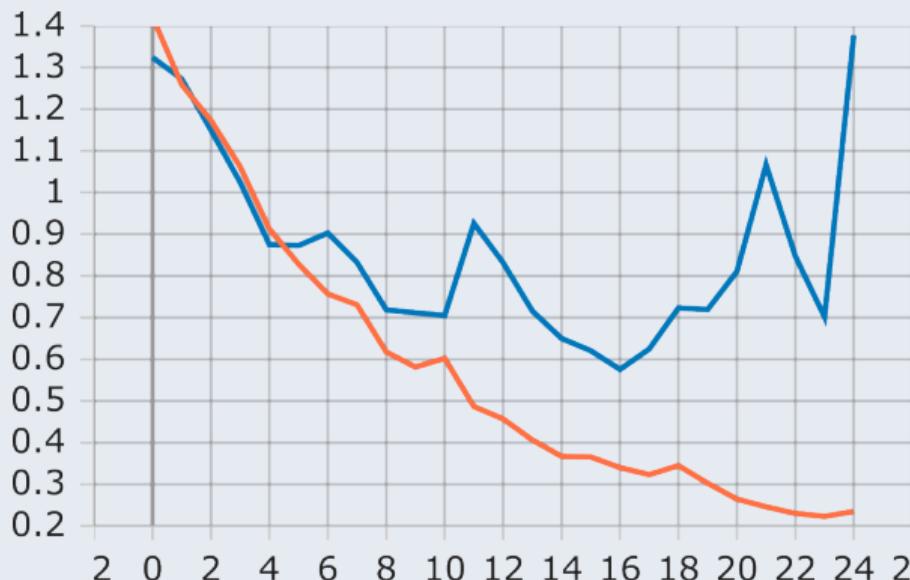
# Tecnologías

## Python y CUDA



# Entrenamiento

## Valores de la función de pérdida



# Medidas contra el sobreajuste

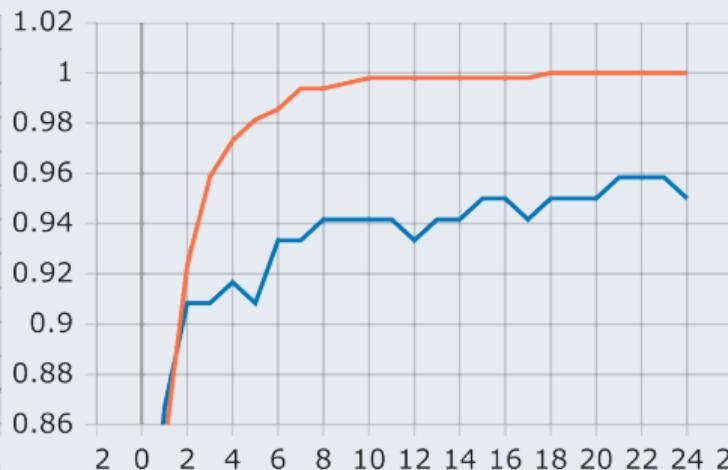
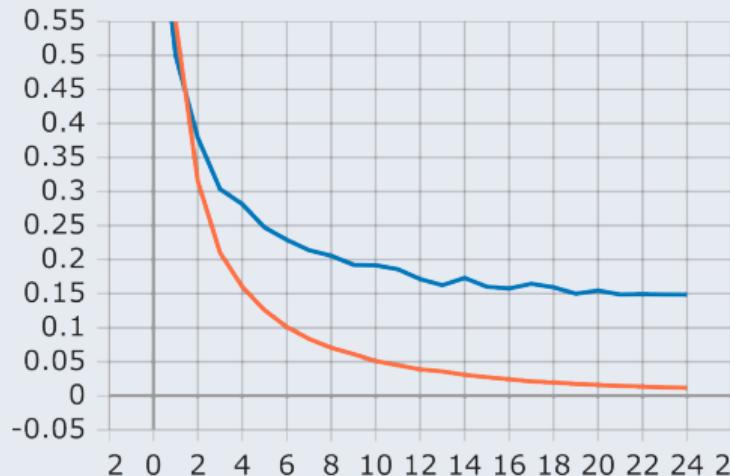
## Aumento de datos



# Medidas contra el sobreajuste

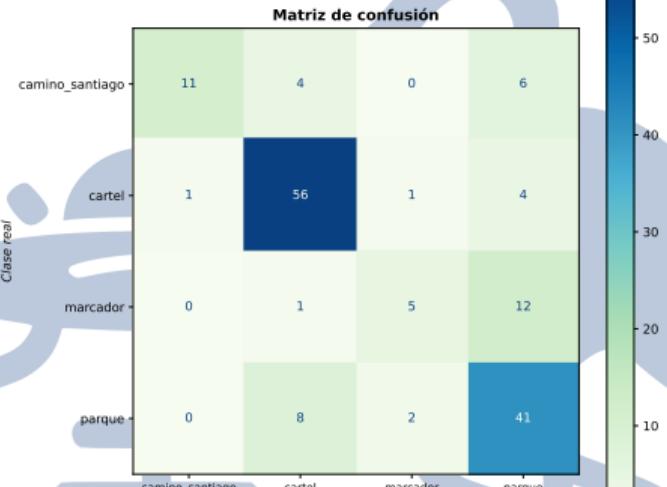
Transfer learning

## Con algunas capas de MobileNet

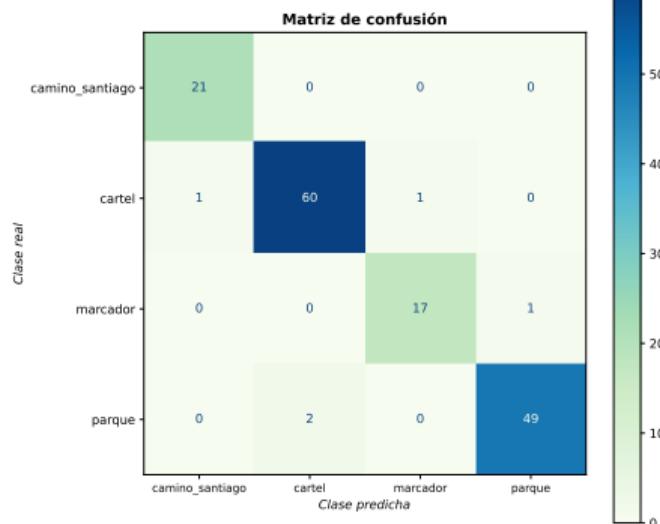


# Métricas

## Matriz de confusión



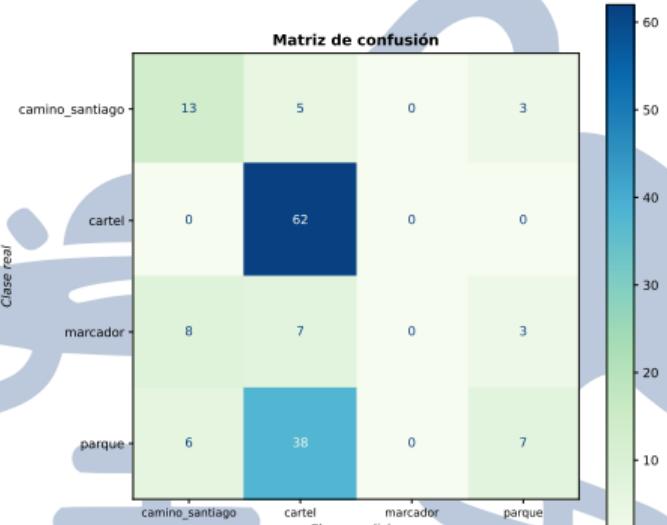
(a) Normal



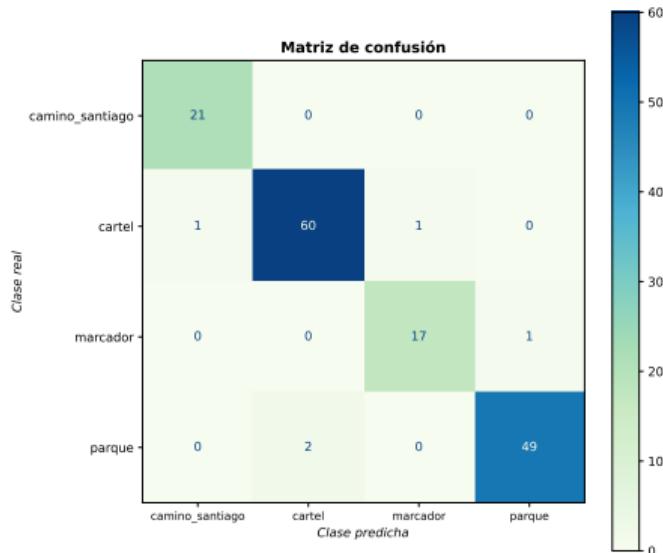
(b) Aplicando transfer learning

# Métricas

## Matriz de confusión



(c) Aplicando aumento de datos



(d) Aplicando ambas

# Métricas

Precisión, recuerdo, y  $F_1$ –score

- Original
  - $\mathcal{P} = 0,75$
  - $\mathcal{R} = 0,743$
  - $F_1 = 0,728$
- Aumento de datos
  - $\mathcal{P} = 0,473$
  - $\mathcal{R} = 0,539$
  - $F_1 = 0,439$
- Transfer learning
  - $\mathcal{P} = 0,967$
  - $\mathcal{R} = 0,967$
  - $F_1 = 0,967$

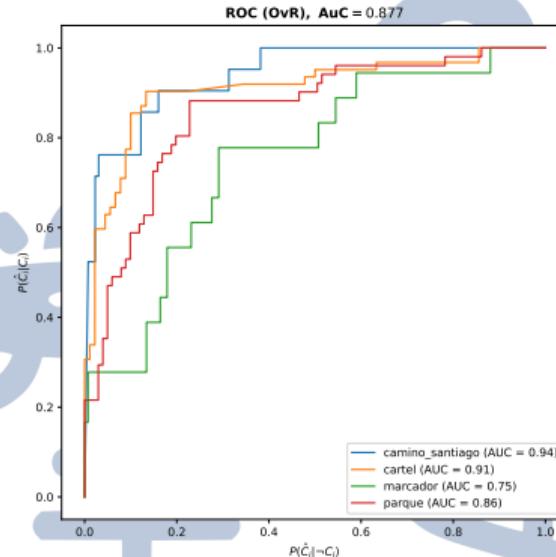
$$\mathcal{P} = P(C_i | \hat{C}_i)$$

$$\mathcal{R} = P(\hat{C}_i | C_i)$$

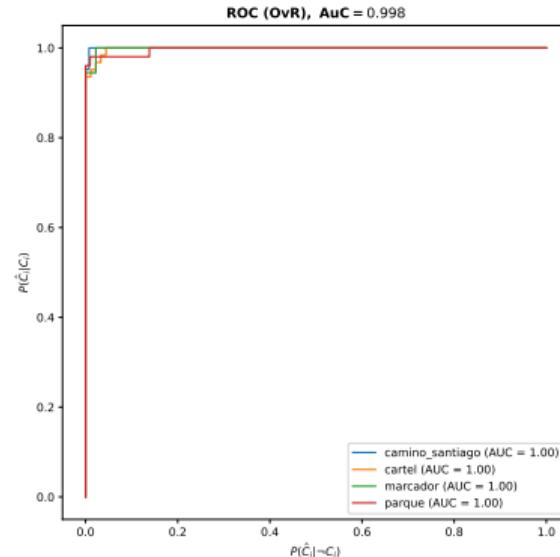
$$F_1 = \frac{2}{\mathcal{P}^{-1} + \mathcal{R}^{-1}}$$

# Métricas

## ROC (OvR) y AUC



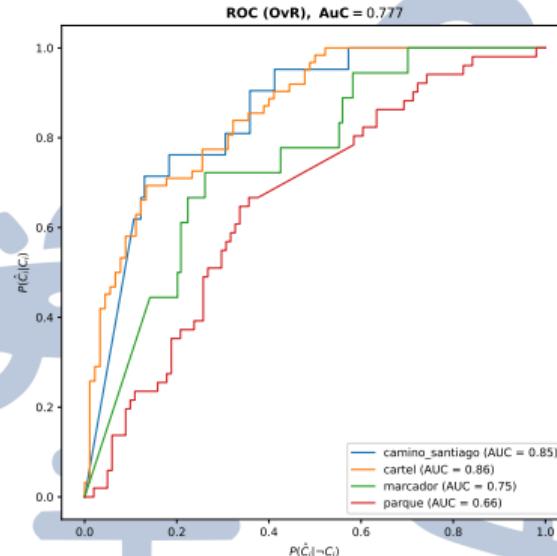
(a) Normal



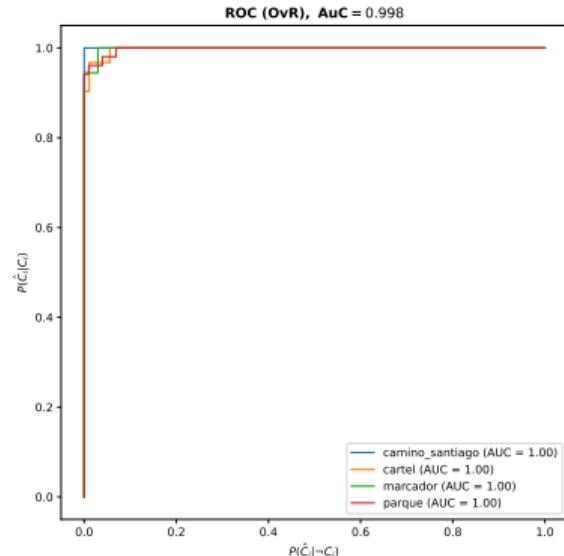
(b) Aplicando transfer learning

# Métricas

## ROC (OvR) y AUC



(c) Aplicando aumento de datos



(d) Aplicando ambas

# Capacidad de generalización

camino\_santiago (0.99)

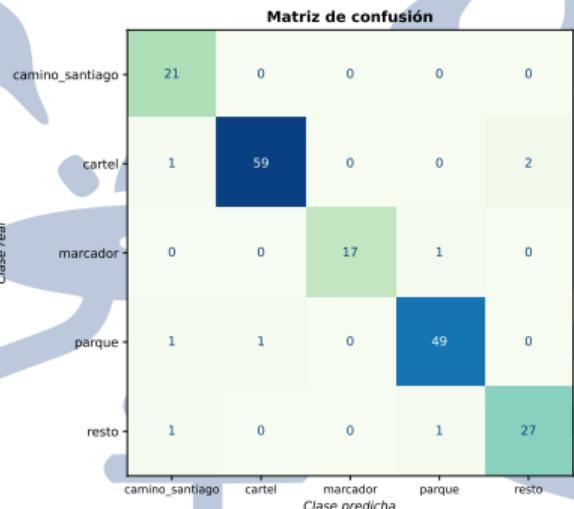


camino\_santiago (0.96)

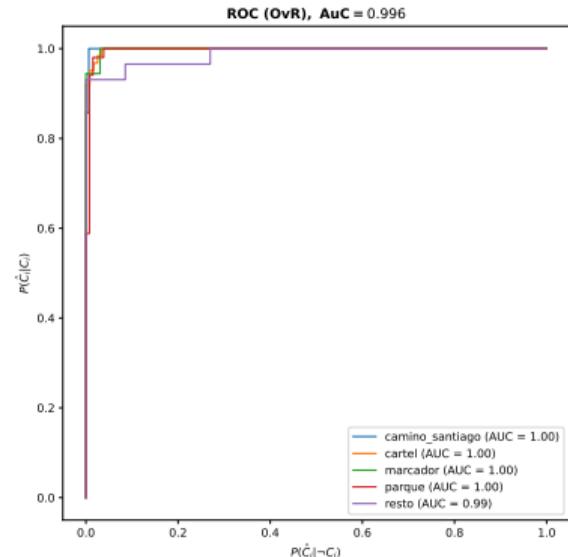


# Conjunto abierto de clases

Clase resto

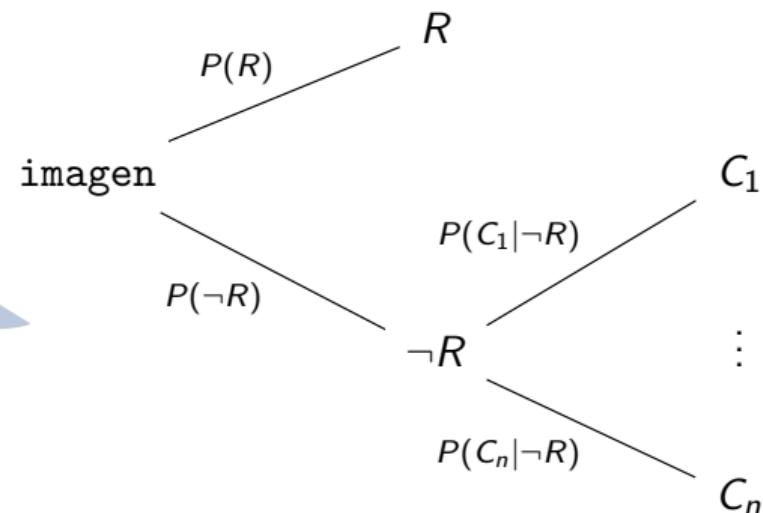
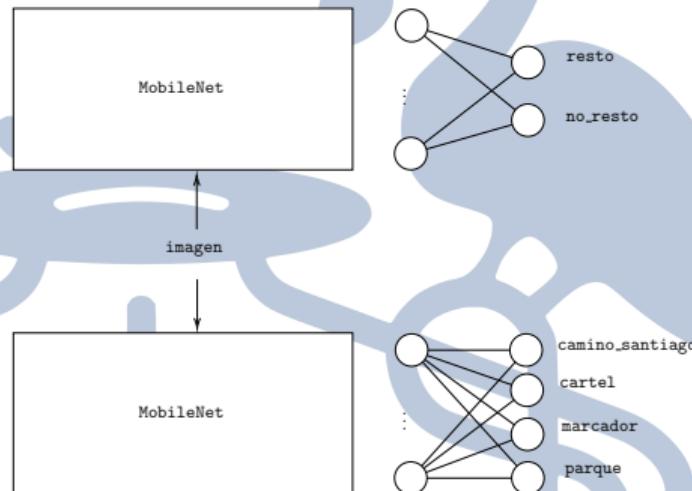


- $\mathcal{P} = 0,958$
- $\mathcal{R} = 0,956$
- $F_1 = 0,956$



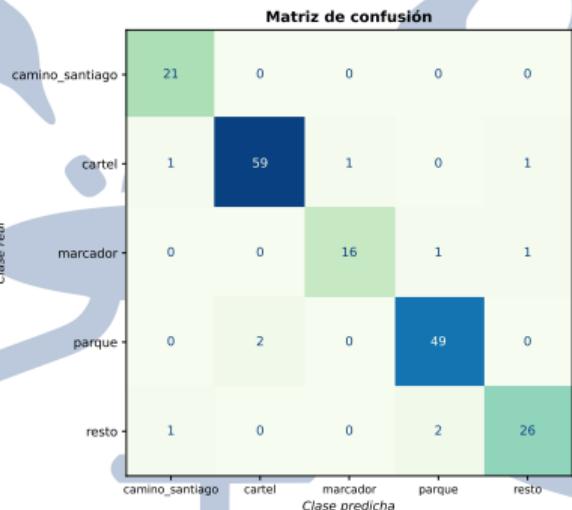
# Conjunto abierto de clases

Red neuronal convolucional doble

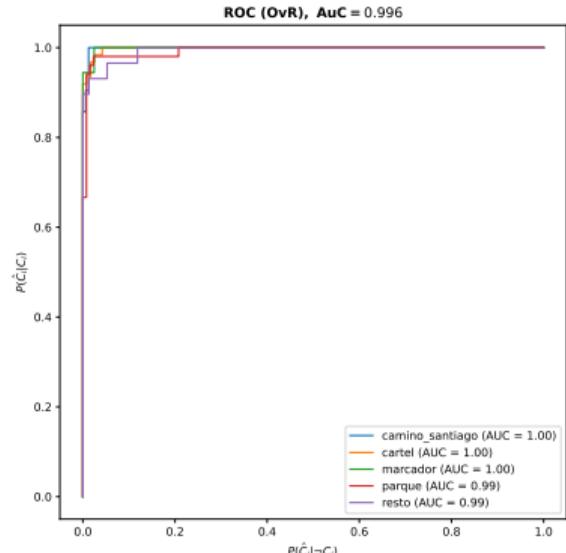


# Conjunto abierto de clases

## Red neuronal convolucional doble

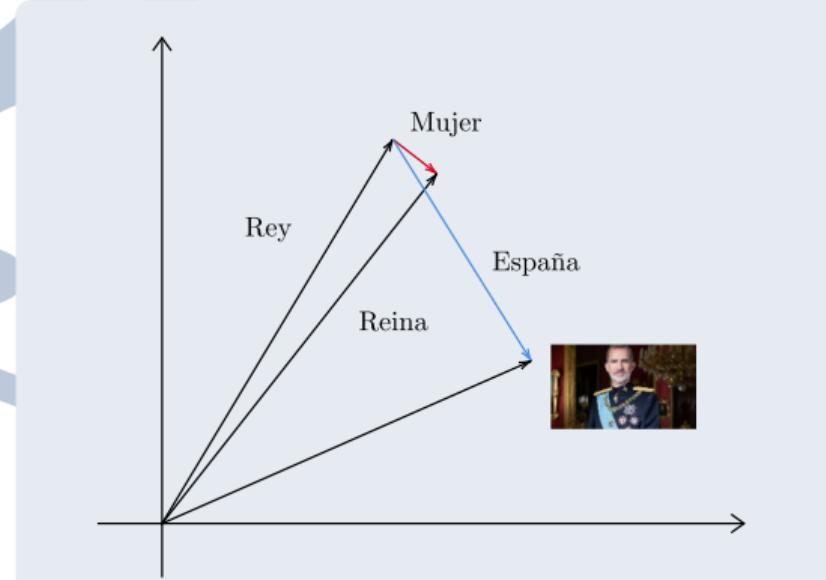


- $\mathcal{P} = 0,945$
- $\mathcal{R} = 0,945$
- $F_1 = 0,945$

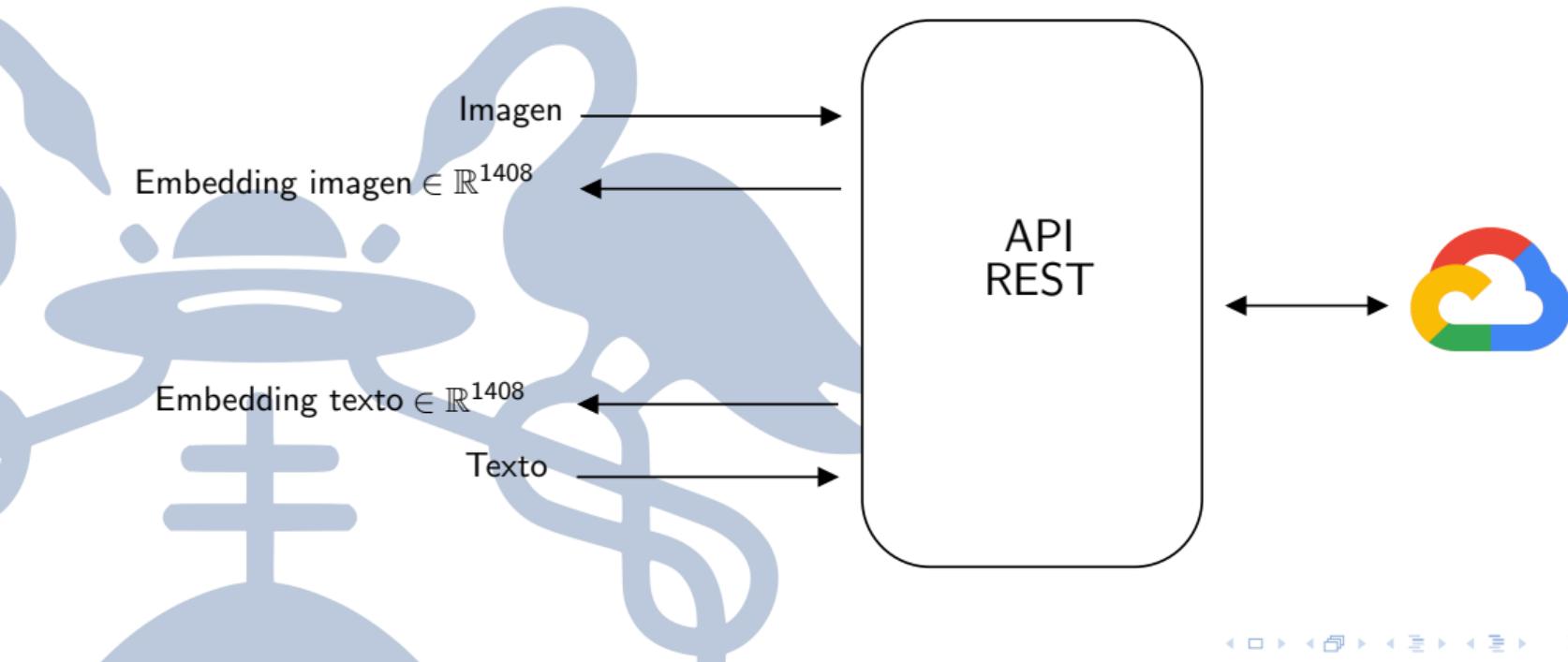


# Transformers multimodales

## Embeddings multimodales



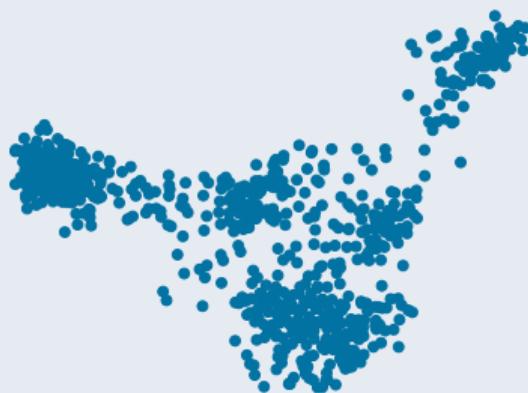
# Generación de embeddings multimodales en VertexAI



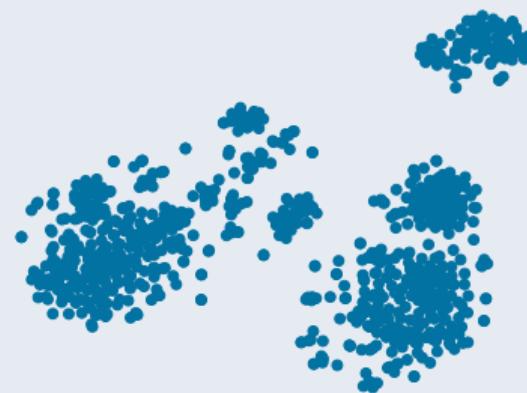
# Embeddings en $\mathbb{R}^2$

PCA y t-SNE

Embeddings (PCA)



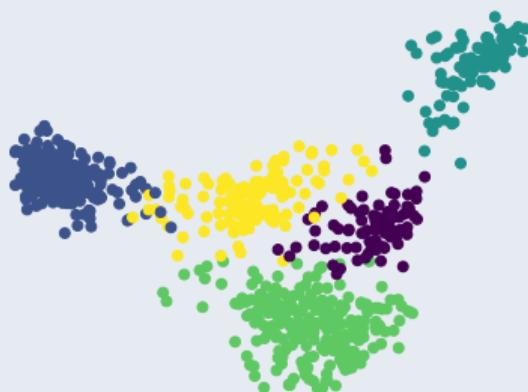
Embeddings (t-SNE)



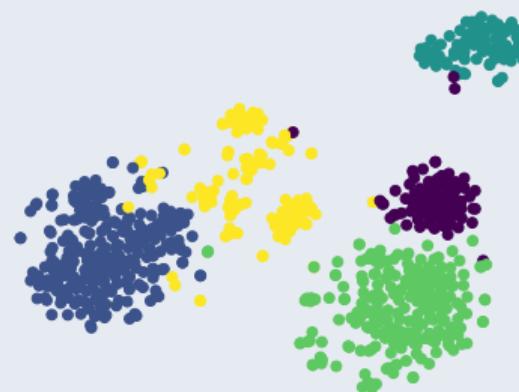
# Clasificación no supervisada

## *k*-means

Embeddings (k-means + PCA)



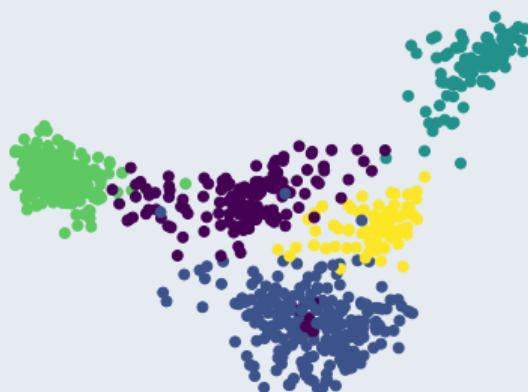
Embeddings (k-means + t-SNE)



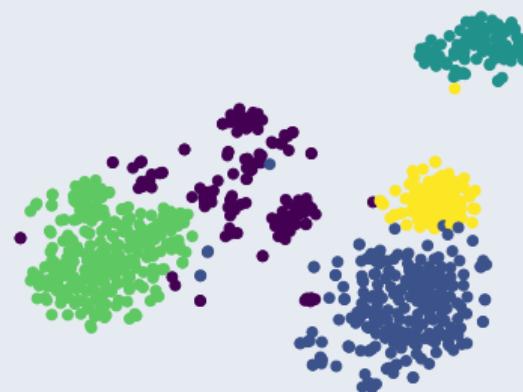
# Clasificación no supervisada

## Clustering jerárquico aglomerativo

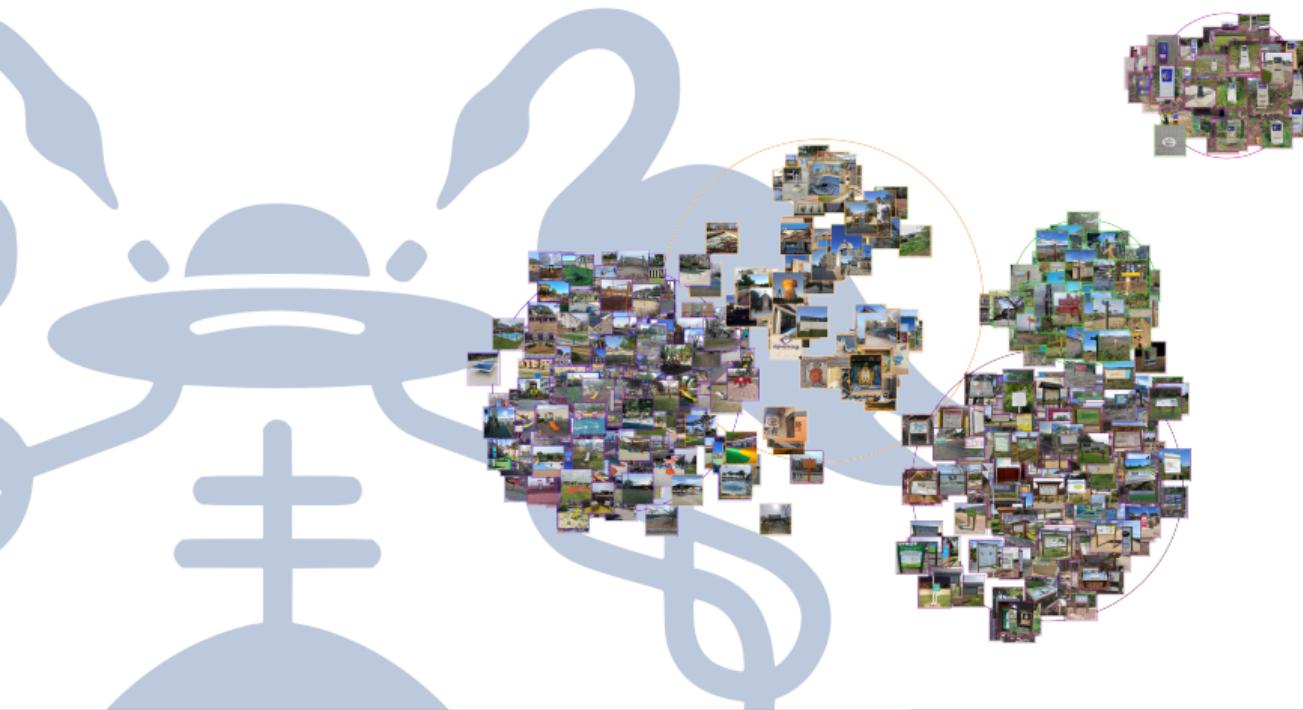
Embeddings (Clustering jerárquico aglomerativo + PCA)



Embeddings (Clustering jerárquico aglomerativo + t-SNE)

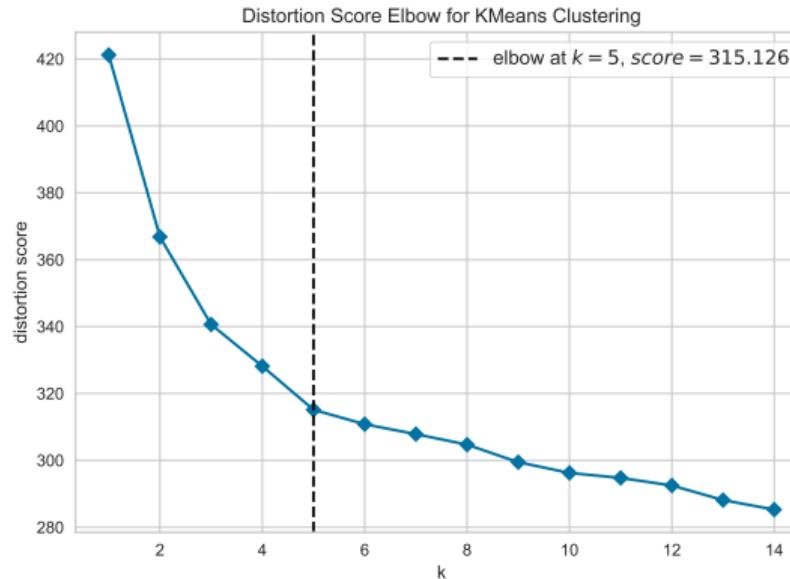


# Clasificación no supervisada



# ¿Y si no se conoce el número de clusters?

Método del codo y kneedle



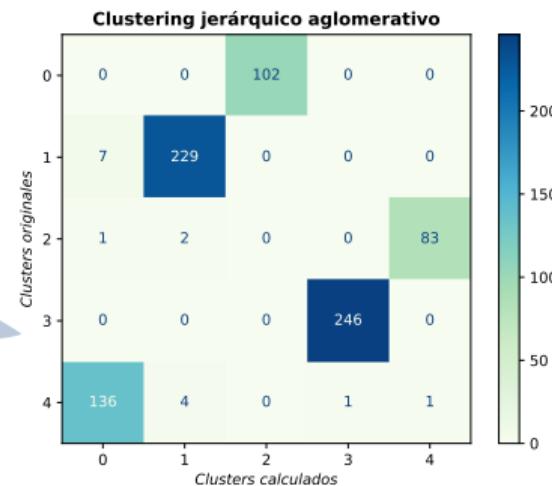
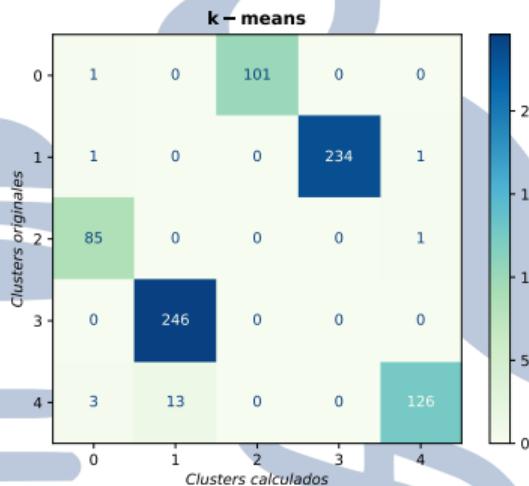
# Métricas

## Coeficiente de la silueta



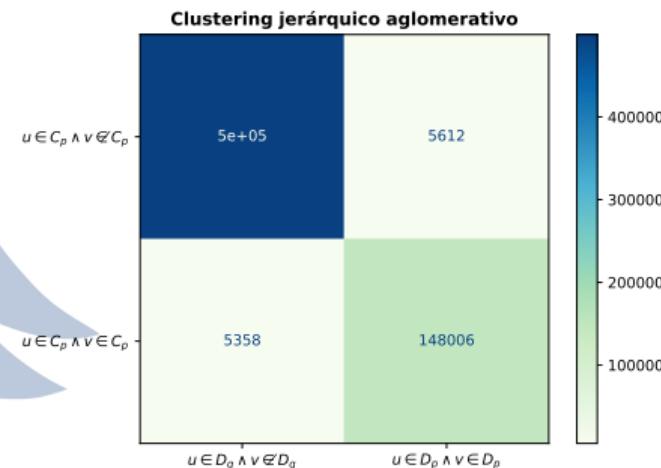
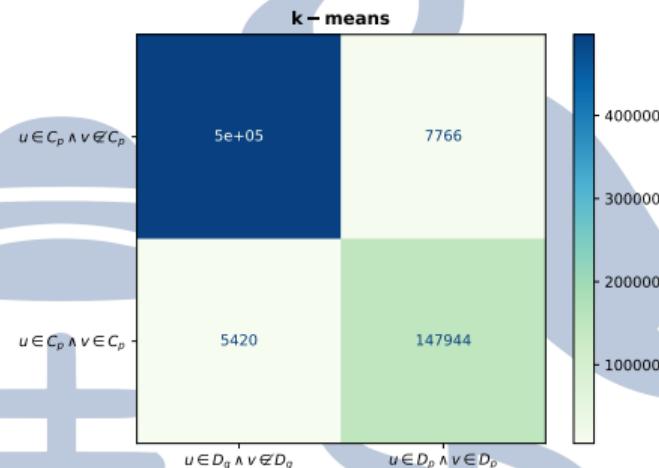
# Métricas

## Matrices de contingencia



# Métricas

## Matrices de confusión



# Métricas

ARI y  $V$ -measure

## $k$ -means

- ARI: 0,944
- $V$ -measure: 0,932

## CJA

- ARI: 0,953
- $V$ -measure: 0,938

Cuidado si hay pocas observaciones y/o muchos clusters...

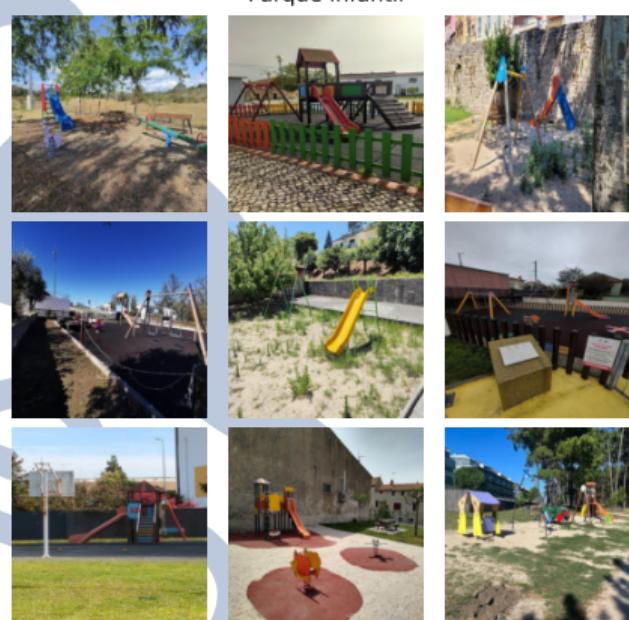
# Búsqueda semántica de imágenes

Encontrar las imágenes que mejor se ajustan a una descripción textual

$$-1 \leq \cos(\theta) \leq 1$$

- Parque infantil
- Marcador de ruta
- Cartel informativo
- Fuentes, piscinas, y elementos de agua
- Murales, grafitis, dibujos, y arte

# Búsqueda semántica de imágenes



# Búsqueda semántica de imágenes

Marcador de ruta



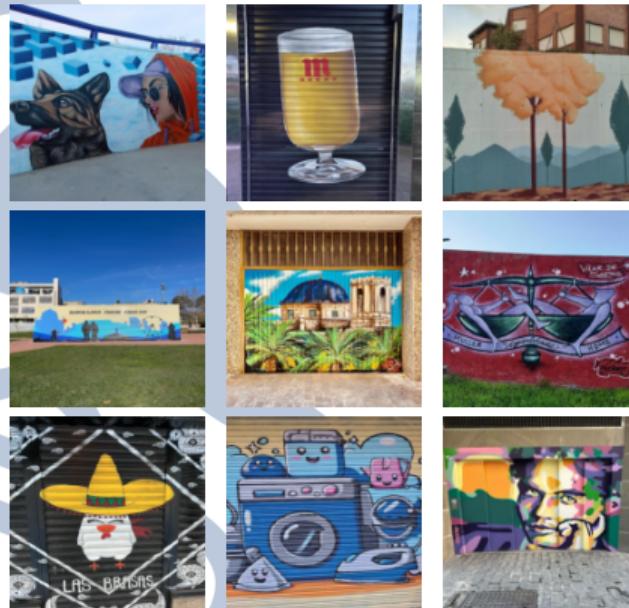
# Búsqueda semántica de imágenes

Cartel informativo



# Búsqueda semántica de imágenes

Murales, grafitis, dibujos y arte



# Búsqueda semántica de imágenes

Fuentes, piscinas, y elementos de agua



# Búsqueda semántica de imágenes



# Conclusiones y trabajo futuro

## Logros

- **Clasificación** de imágenes en **conjuntos abiertos** y **cerrados**, de manera **supervisada** y **no supervisada**
- **Selección** de **textos descriptivos** para las imágenes

# Conclusiones y trabajo futuro

Para automatizar al completo el proceso...

- Verificación de coordenadas
- Detección de elementos duplicados



0,89

0,75

0,44

0,49

# Clasificación de imágenes y asignación de textos mediante redes neuronales convolucionales y transformers multimodales

Trabajo Fin de Grado

Pablo García García

Grado en Ingeniería Informática

Universidad de Alcalá  
Escuela Politécnica Superior

16 de julio de 2024