

Content Based Image Retrieval

Proyecto final del curso

Integrantes:
Pablo Yañez, Magdalena De La Fuente

Agenda

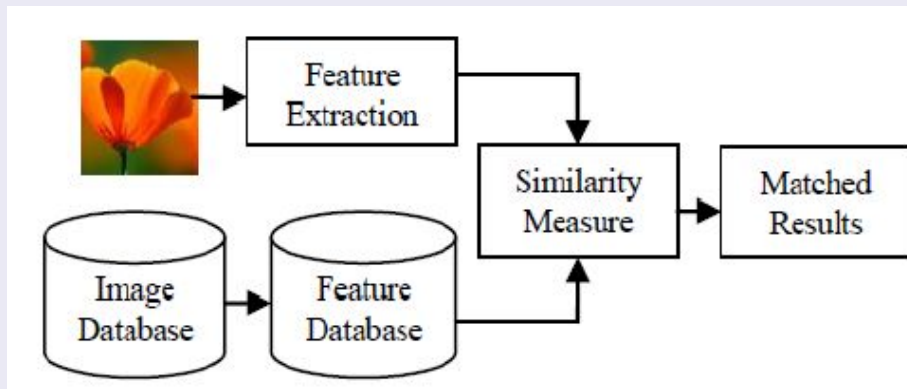
1. Contexto
2. Métodos implementados
 - a. Bag of Visual Words
 - b. CNN
 - c. LDA
 - d. Similitud Coseno
3. Resultados
 - a. Ranking
 - b. Ranking Normalizado
 - c. Curvas Precision-Recall
 - d. Tiempos de ejecución
 - e. Ejemplos
4. Conclusiones
5. Bibliografía

Que es el Content Based Image Retrieval?

El objetivo de los métodos CBIR es obtener imágenes similares a una provista por el usuario. Es en esencia, un buscador de imágenes.

El problema CBIR necesita una imagen provista por el usuario, y una base de datos. En esta base de datos se encuentran almacenados los vectores característicos de un gran número de imágenes.

Para lograr obtener imágenes similares a partir de una imagen *query*, es necesario extraer las características que identifican a una imagen. Esto para todas las imágenes de la base de datos en la que se consulta.

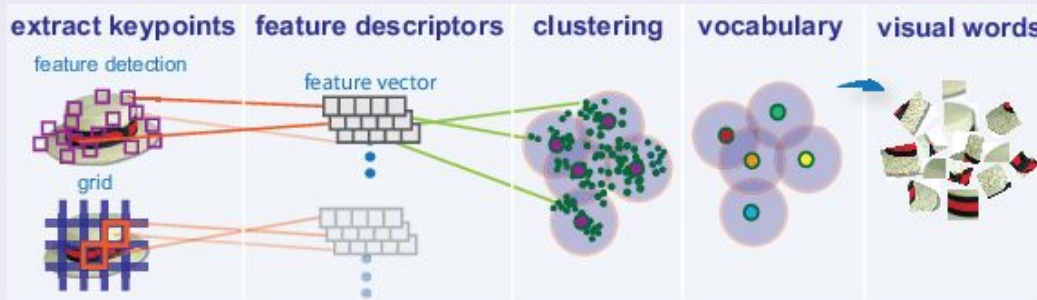


Métodos implementados: Bag of Visual Words



Se construye un modelo Bag of Visual Words (BoVW). Este modelo consta de cuatro pasos:

- Las características de las imágenes se extraen por medio de SIFT. De esta operación se obtienen descriptores para cada imagen de la base de datos.
- Se hace clustering de los descriptores.
- Se genera un vocabulario en base a los centroides de los cluster obtenidos.
- Se generan histogramas para cada imagen.



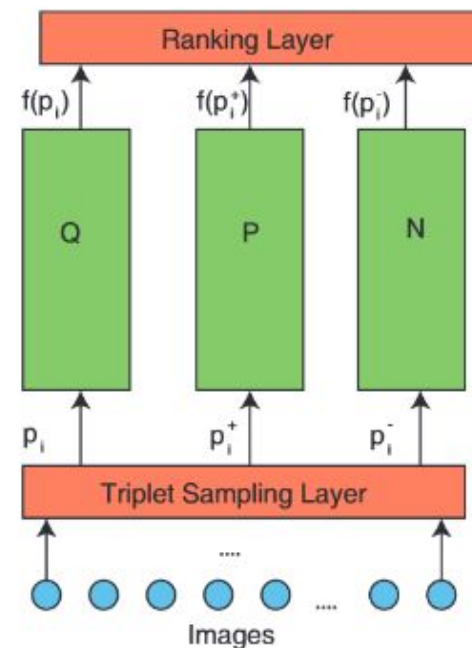
Métodos implementados: CNN

Se utiliza una red VGG16 ya entrenada, **Learning Fine-grained Image Similarity with Deep Ranking [1]**. El propósito original de la red era encontrar fine-grained similarities entre imágenes.

El input de la red consiste en tres imágenes. La consultada, una imagen considerada 'positiva' y una imagen considerada negativa.

Durante el entrenamiento, una capa de ranking evalúa (hinge loss) en cuanto se ha alterado el posicionamiento de las imágenes en el ranking e inicia el backpropagation del error.

De esta manera, la red entrenada es capaz de generar vectores de similitud para cada imagen.



Métodos implementados: Fusión de métodos

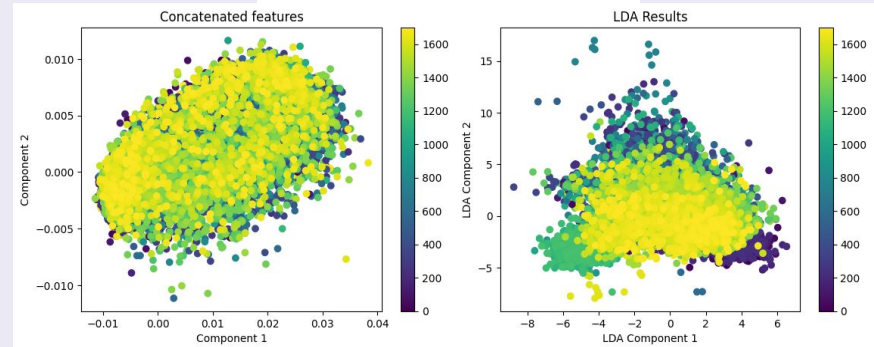
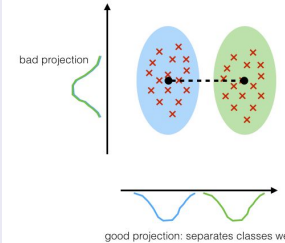
Con el objetivo de disminuir la dimensionalidad de los vectores característicos y reducir los tiempos de consulta se implementó un tercer método.

Este consiste en la concatenación de los dos vectores generados anteriormente y una transformación de la base de datos por medio de LDA.

Se obtiene un nuevo set de vectores característicos proyectado para maximizar la separación entre clases.

LDA:

maximizing the component axes for class-separation



Vectores característicos obtenidos

Método Handcrafted:

- Largo del vector: 120

Método CNN:

- Largo del vector: 4096

Método de Fusión:

- Largo del vector: 1698

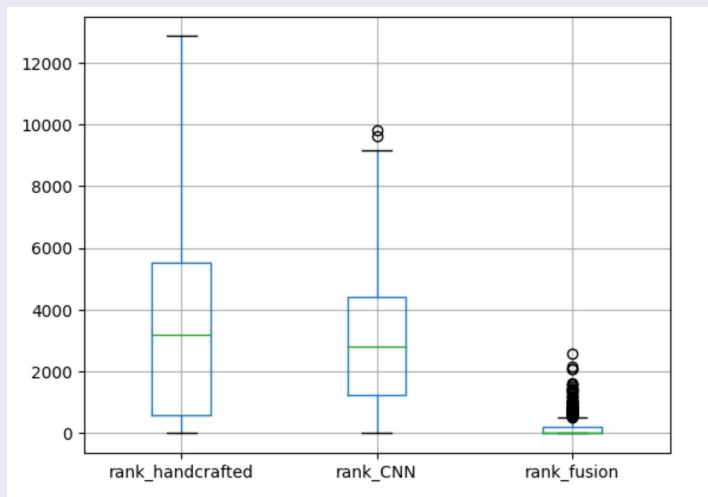
Métrica de similitud

- La Métrica de similitud de distancia coseno fue seleccionada debido a su superior rendimiento, según resultados en [2].

$$\text{Coseno: } \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \cdot \sum_{i=1}^n B_i^2}}$$

Resultados de Rank

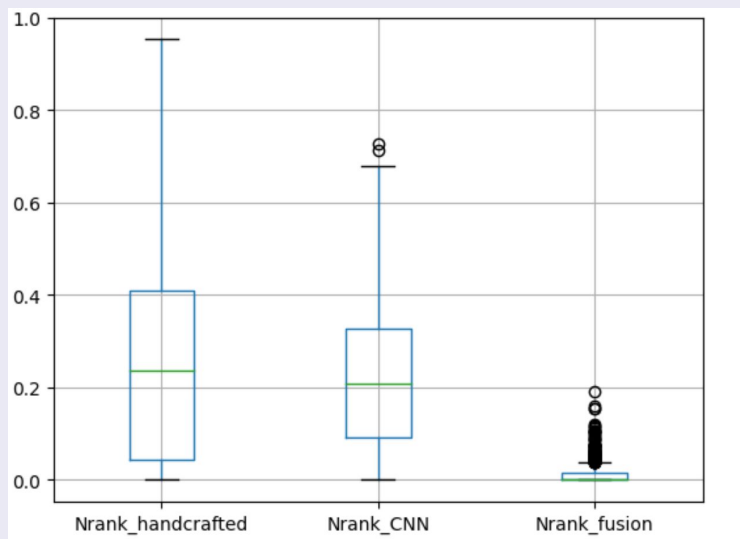
$$Rank = \frac{1}{N_{rel}} \sum_{i=1}^{N_{rel}} R_i$$



	rank_handcrafted	rank_CNN	rank_fusion
count	1700.000000	1700.000000	1700.000000
mean	3381.544337	2960.011163	145.996183
std	2794.454267	1965.402397	253.381632
min	1.000000	1.000000	1.000000
25%	578.444444	1255.861111	2.500000
50%	3190.727778	2813.494444	14.505556
75%	5526.461111	4429.769444	212.938889
max	12855.500000	9808.333333	2562.677778

Resultados de Rank Normalizado

$$\widetilde{\text{Rank}} = \frac{1}{N \times N_{rel}} \left(\sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel} (N_{rel} + 1)}{2} \right)$$

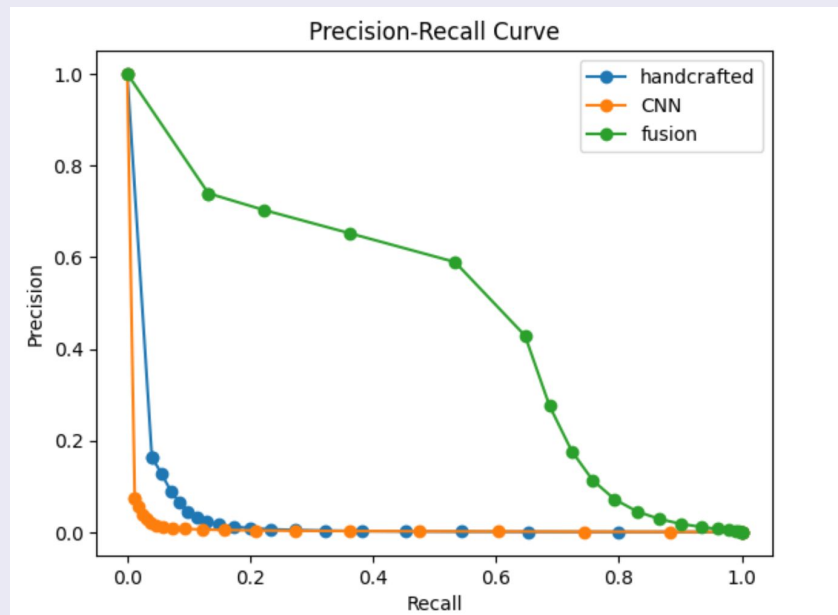


	Nrank_handcrafted	Nrank_CNN	Nrank_fusion
count	1700.000000	1700.000000	1700.000000
mean	0.250358	0.219112	0.010528
std	0.207071	0.145630	0.018739
min	0.000000	0.000000	0.000000
25%	0.042617	0.092718	0.000000
50%	0.236284	0.208175	0.000719
75%	0.409270	0.327979	0.015413
max	0.952783	0.726657	0.189584

Curva PR

$$\text{Precisión} = \frac{TP}{TP+FP} = \frac{TP}{\text{Total retrieved}}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{TP}{\text{Total relevant}}$$



Tiempos de ejecución

- A pesar de las diferencias en precisión y recuperación, el tiempo de búsqueda es similar en todos los métodos.
- Esto sugiere que el costo computacional de la búsqueda es aproximadamente constante, sin importar la dimensión del vector característico.

Ejemplos de consultas (1/3)

E1

Handcrafted Features



CNN Features

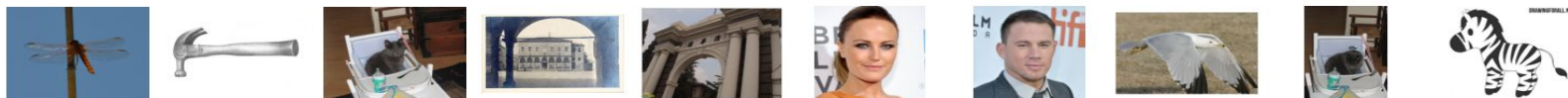


Fused Features



E2

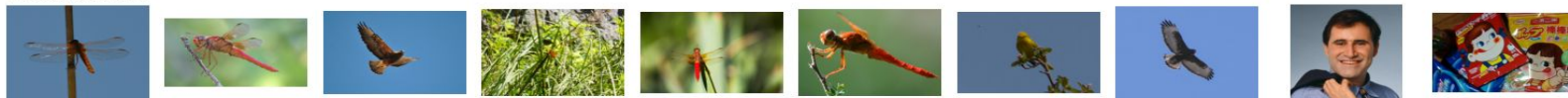
Handcrafted Features



CNN Features



Fused Features



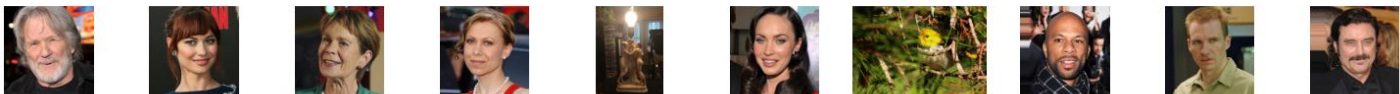
Ejemplos de consultas (2/3)

E3

Handcrafted Features



CNN Features

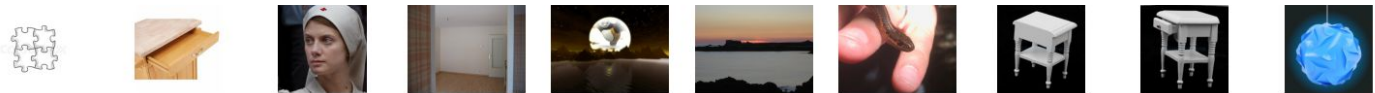


Fused Features



E4

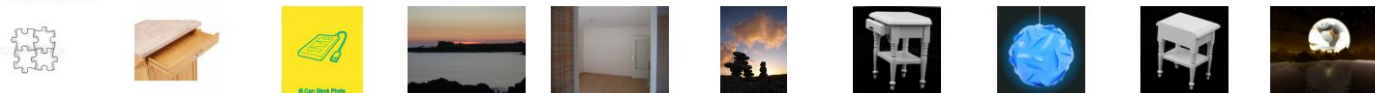
Handcrafted Features



CNN Features



Fused Features



Ejemplos de consultas (3/3)

E5

Handcrafted Features



CNN Features



Fused Features



E6

Handcrafted Features



CNN Features



Fused Features



Conclusiones

- Algoritmo CBIR implementado con éxito usando técnicas "handcrafted", CNN, y combinadas.
- Método "Handcrafted" útil y pero menos eficiente en promedio.
- Método CNN más sensible a patrones y texturas.
- Combinación de técnicas mejora precisión y recall. Fundamental fué el uso de LDA para reducir dimensionalidad.
- Tiempos de búsqueda similares en todos los casos.

Bibliografía

[1] Learning Fine-grained Image Similarity with Deep Ranking (<https://arxiv.org/abs/1404.4661>)

[2] A Large-scale Dataset and Benchmark for Similar Trademark Retrieval (<https://arxiv.org/pdf/1701.05766.pdf>)

Content Based Image Retrieval

Proyecto Final del curso

Integrantes:
Pablo Yañez, Magdalena De La Fuente