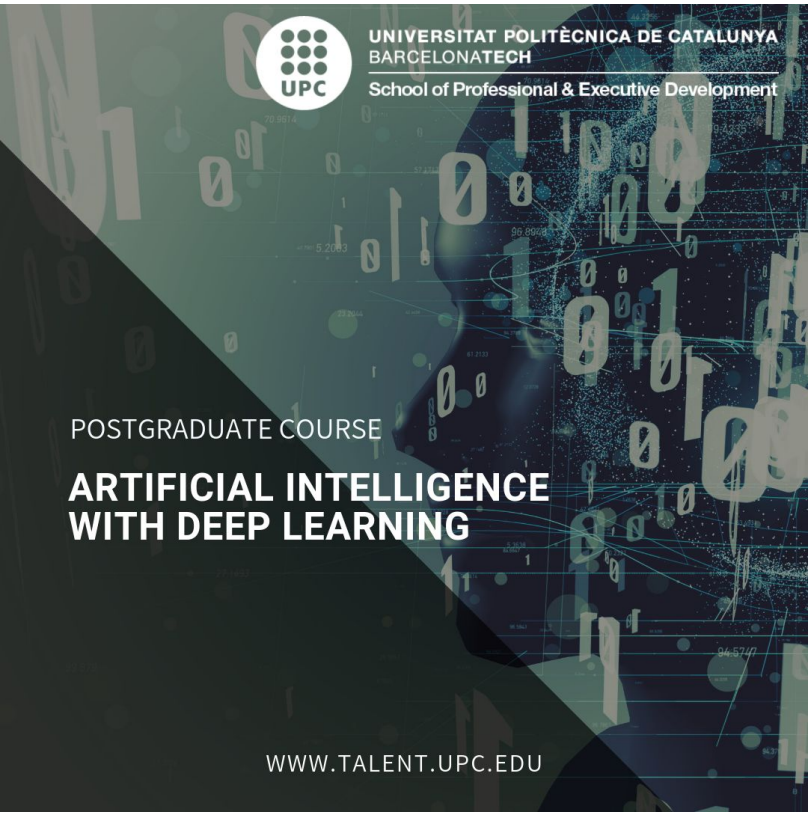


#DLUPC

Image Retrieval

Insight
Centre for Data Analytics

Eva Mohedano
eva.mohedano@satellogic.com

Data Scientist
Satellogic Solutions

SATELLOGIC

This lecture objective

Content Based Image Retrieval (CBIR)

Deep Learning approaches in CBIR

Practical session

More material online

DLCV2018

This lecture objective

Content Based Image Retrieval (CBIR)

Deep Learning approaches in CBIR

Practical session

The Problem: Query by Example

Given:

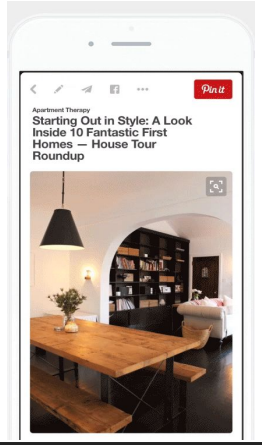
- An example query image that illustrates the user's information need
- A very large dataset of images

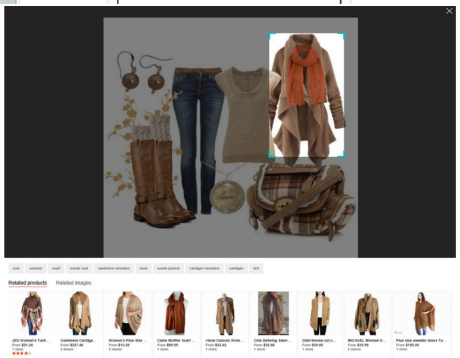
Task:

- Rank all images in the dataset according to how likely they are to fulfil the user's information need

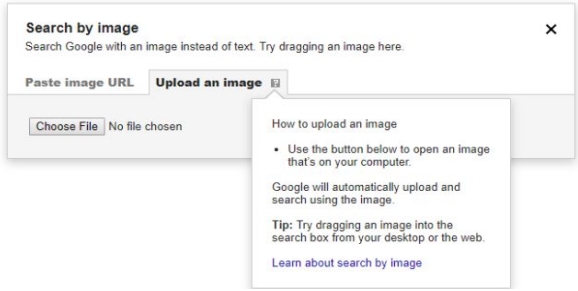
Applications

e-commerce



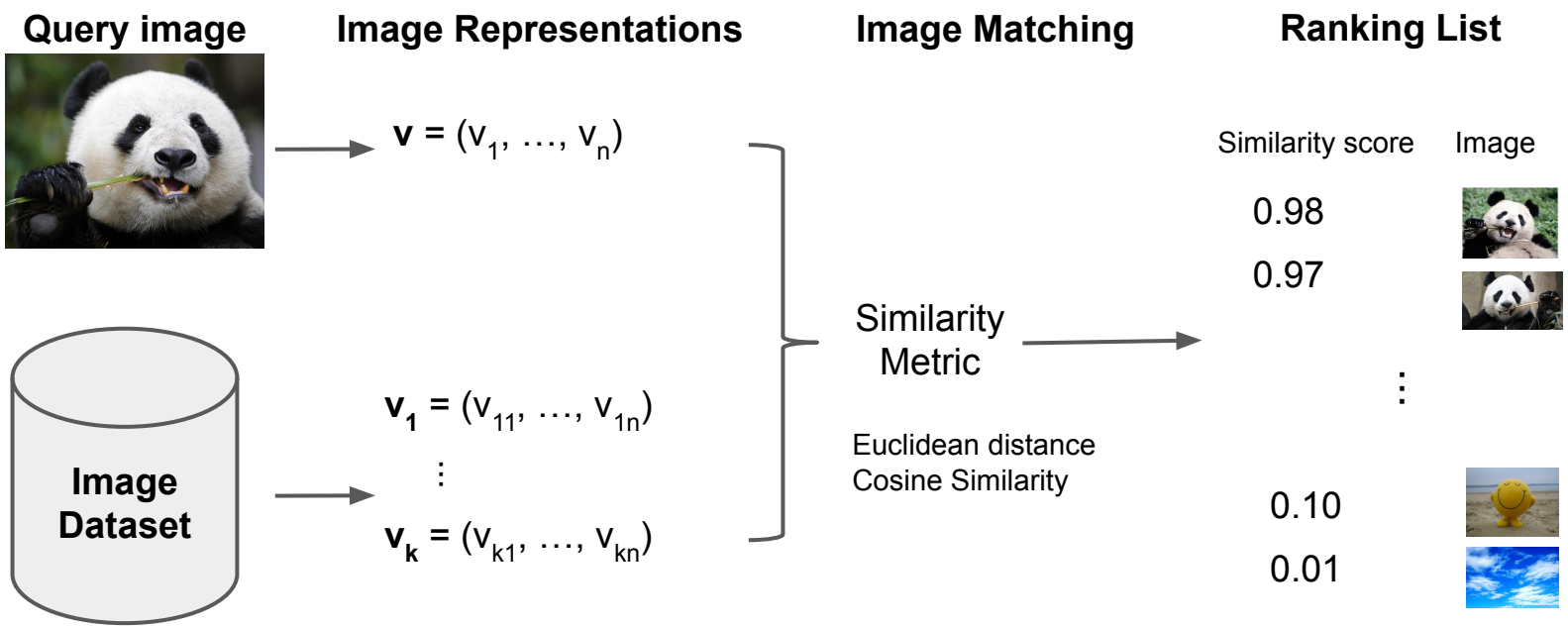


seeking information

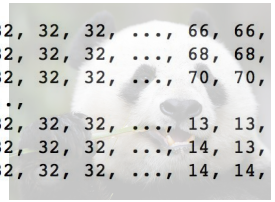


personal photo organization

CBIR Pipeline



CBIR Pipeline



```
array([[32, 32, 32, ..., 66, 66, 66],
       [32, 32, 32, ..., 68, 68, 68],
       [32, 32, 32, ..., 70, 70, 70],
       ...,
       [32, 32, 32, ..., 13, 13, 12],
       [32, 32, 32, ..., 14, 13, 12],
       [32, 32, 32, ..., 14, 14, 14]], dtype=uint8)
```

Query image

Image Representations

$\mathbf{v} = (v_1, \ldots, v_n)$

Image Dataset

$\mathbf{v_1} = (v_{11}, \ldots, v_{1n})$

$\vdots$

$\mathbf{v_k} = (v_{k1}, \ldots, v_{kn})$

Image Matching

Similarity Metric

Euclidean distance
Cosine Similarity

Ranking List

Similarity score	Image
0.98	
0.97	
$\vdots$	
0.10	
0.01	

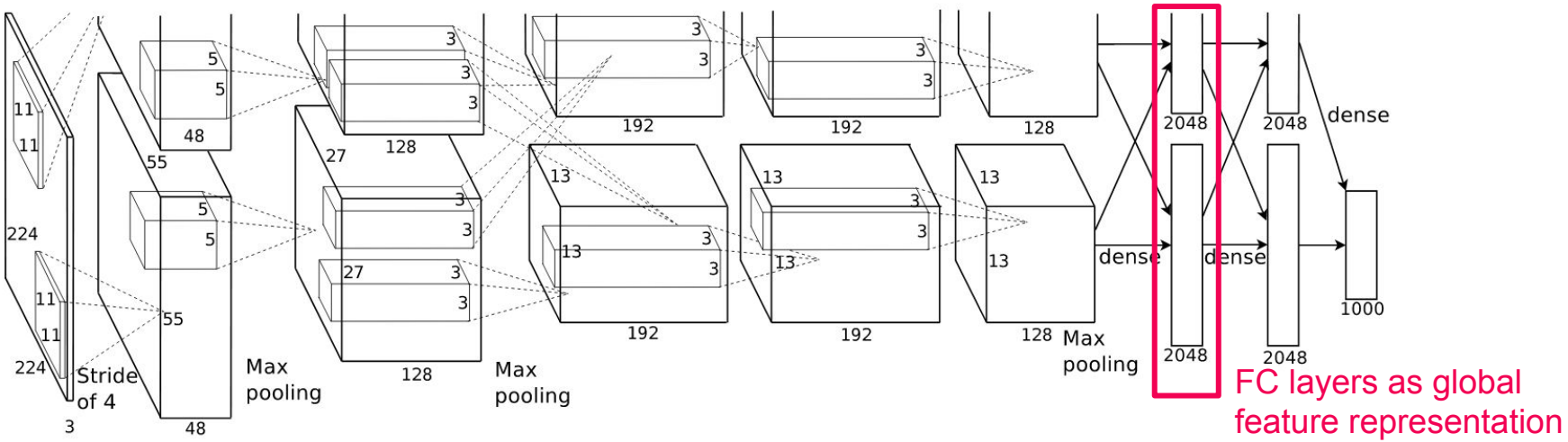
This lecture objective

Content Based Image Retrieval (CBIR)

Deep Learning approaches in CBIR

Practical session

“Off-the-shelf” classification models

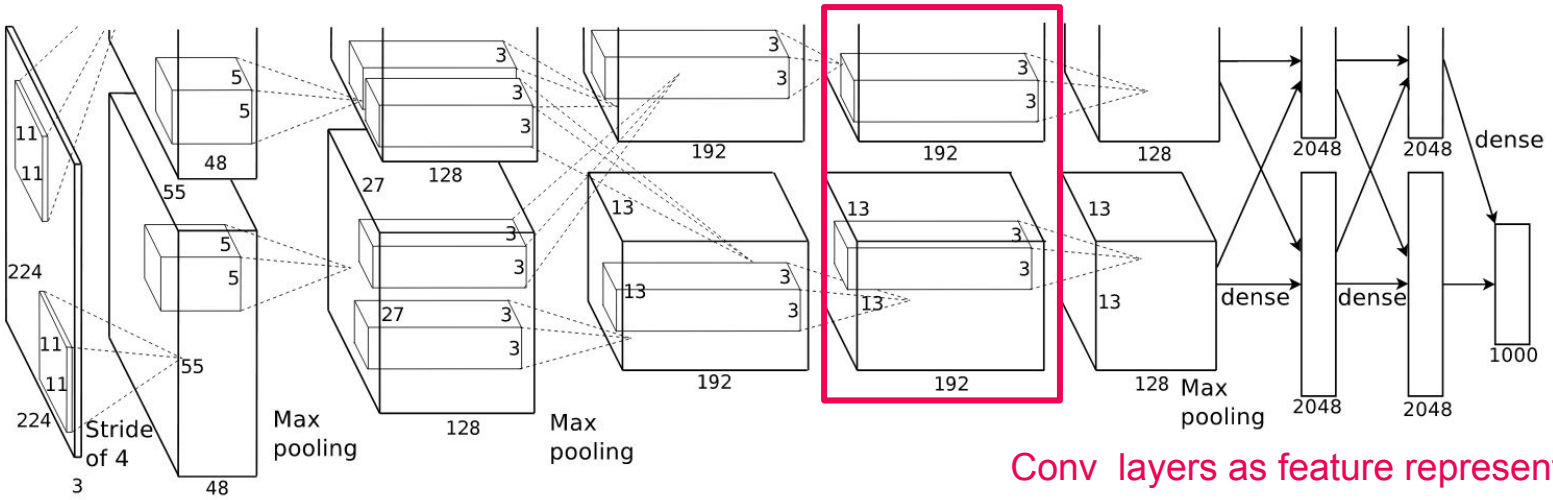


FC layers as global feature representation

- Descriptors too biased to the final classification task
- No spatial information
- Performance no better than traditional hand-crafted approaches!

Razavian et al. CNN features off-the-shelf: an astounding baseline for recognition. CVPR 2014
Babenko et al. Neural codes for image retrieval. ECCV 2014

“Off-the-shelf” classification models

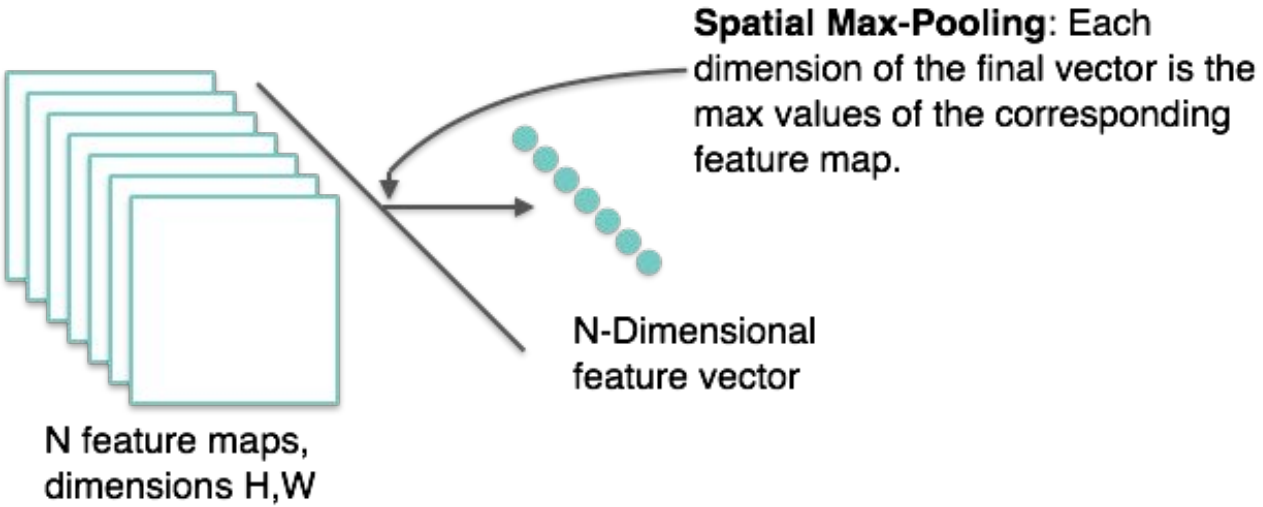


Conv layers as feature representation

- Spatial information is encoded
- Need to aggregate local information into a single vector
- Better performance than traditional approaches

Babenko and Lempitsky, Aggregating local deep features for image retrieval. ICCV 2015
Tolias et al. Particular object retrieval with integral max-pooling of CNN activations. *arXiv 2015*.
Kalantidis et al. Cross-dimensional Weighting for Aggregated Deep Convolutional Features. ECCV 2016 Workshops.

Representation from Conv layers

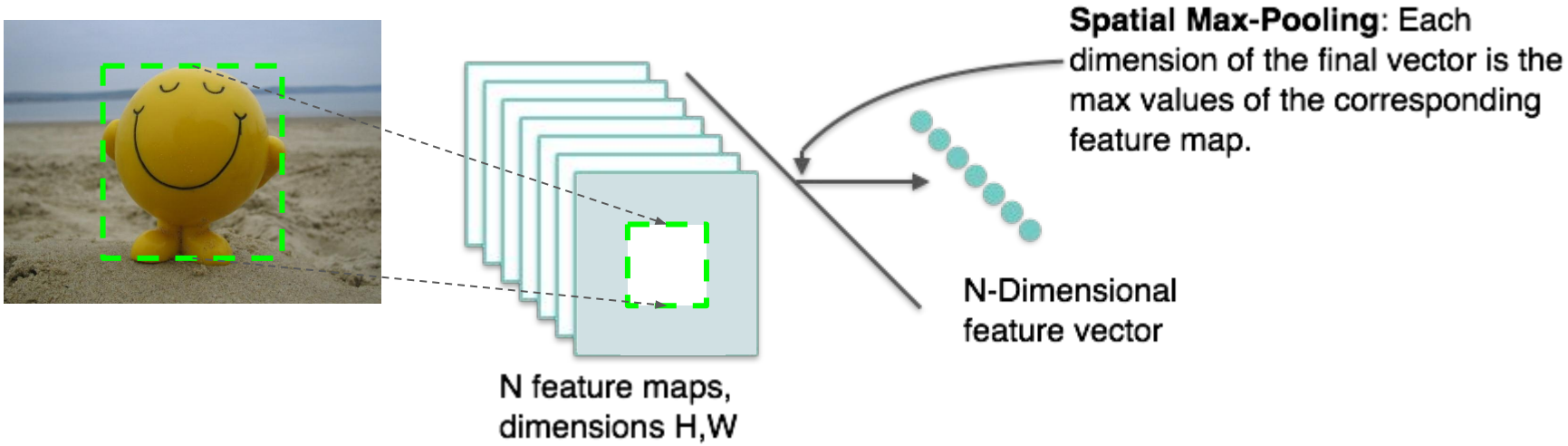


Lower level of abstraction (off-the-self model)

Allow processing images at higher resolution

Representation from Conv layers

Aggregate information over specific regions



“Off-the-shelf” network limitations

Classification

Query: This chair

Results from dataset classified as “chair”

“Off-the-shelf” network limitations

Retrieval

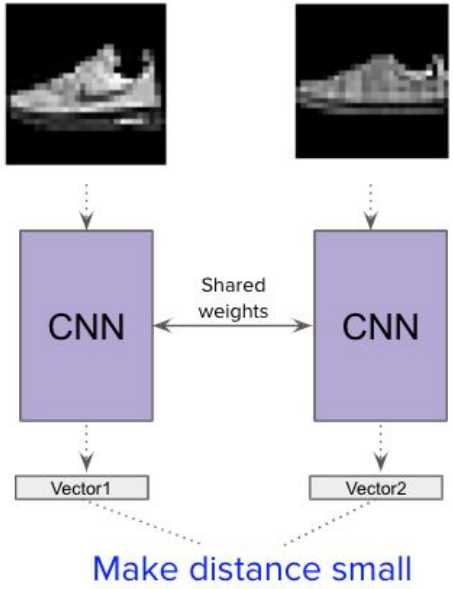
Query: This chair

Results from dataset **ranked** by similarity to the query

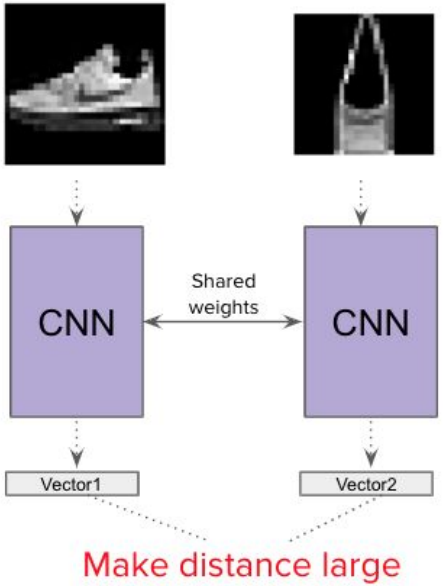
Siamese Networks for retrieval

$$l(\mathbf{x}_1, \mathbf{x}_2, \delta) = \delta \cdot l_P(d_D(\mathbf{x}_1, \mathbf{x}_2)) + (1 - \delta) \cdot l_N(d_D(\mathbf{x}_1, \mathbf{x}_2))$$

Positive Pairs



Negative Pairs



$$l_P(d_D(\mathbf{x}_1, \mathbf{x}_2)) = d_D(\mathbf{x}_1, \mathbf{x}_2)$$

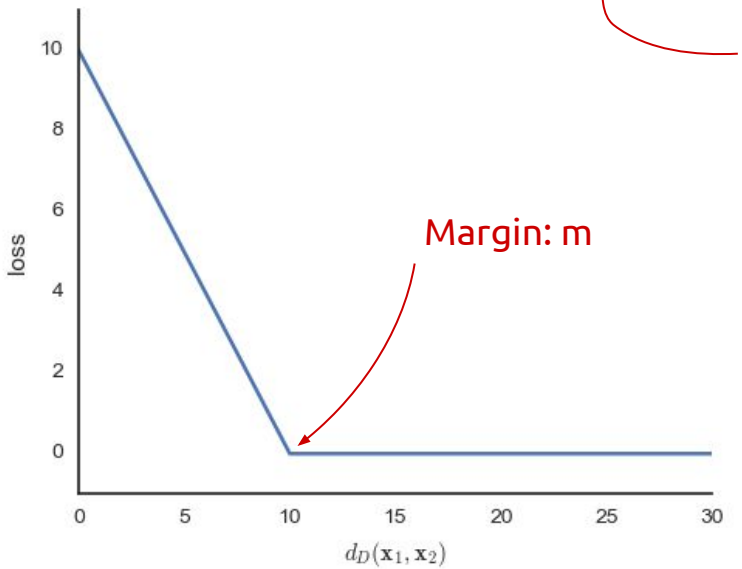
$$l_N(d_D(\mathbf{x}_1, \mathbf{x}_2)) = \max(0, m - d_D(\mathbf{x}_1, \mathbf{x}_2))$$

Hinge loss

Negative pairs: if nearer than the margin, pay a linear penalty

$$l_N(d_D(\mathbf{x}_1, \mathbf{x}_2)) = \max(0, m - d_D(\mathbf{x}_1, \mathbf{x}_2))$$

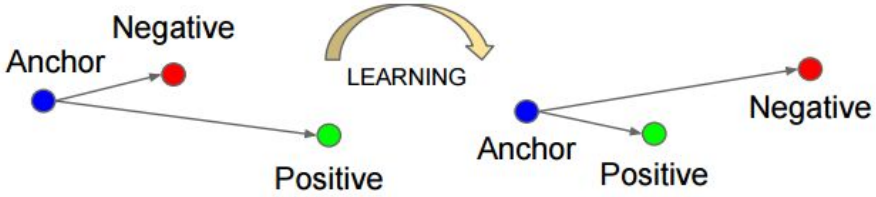
Hinge loss



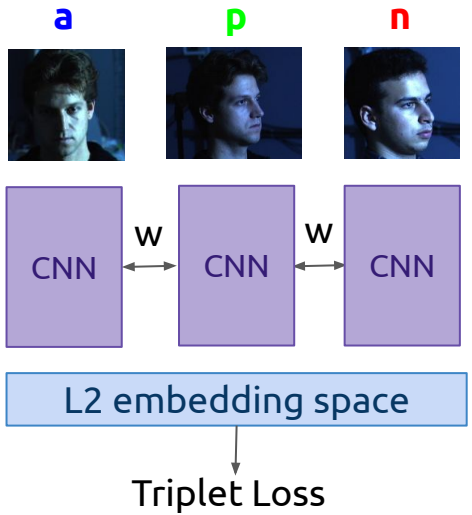
The margin is a **hyperparameter**

Triplet Networks

Siamese network with triplet loss: loss function minimizes distance between query and positive and maximizes distance between query and negative



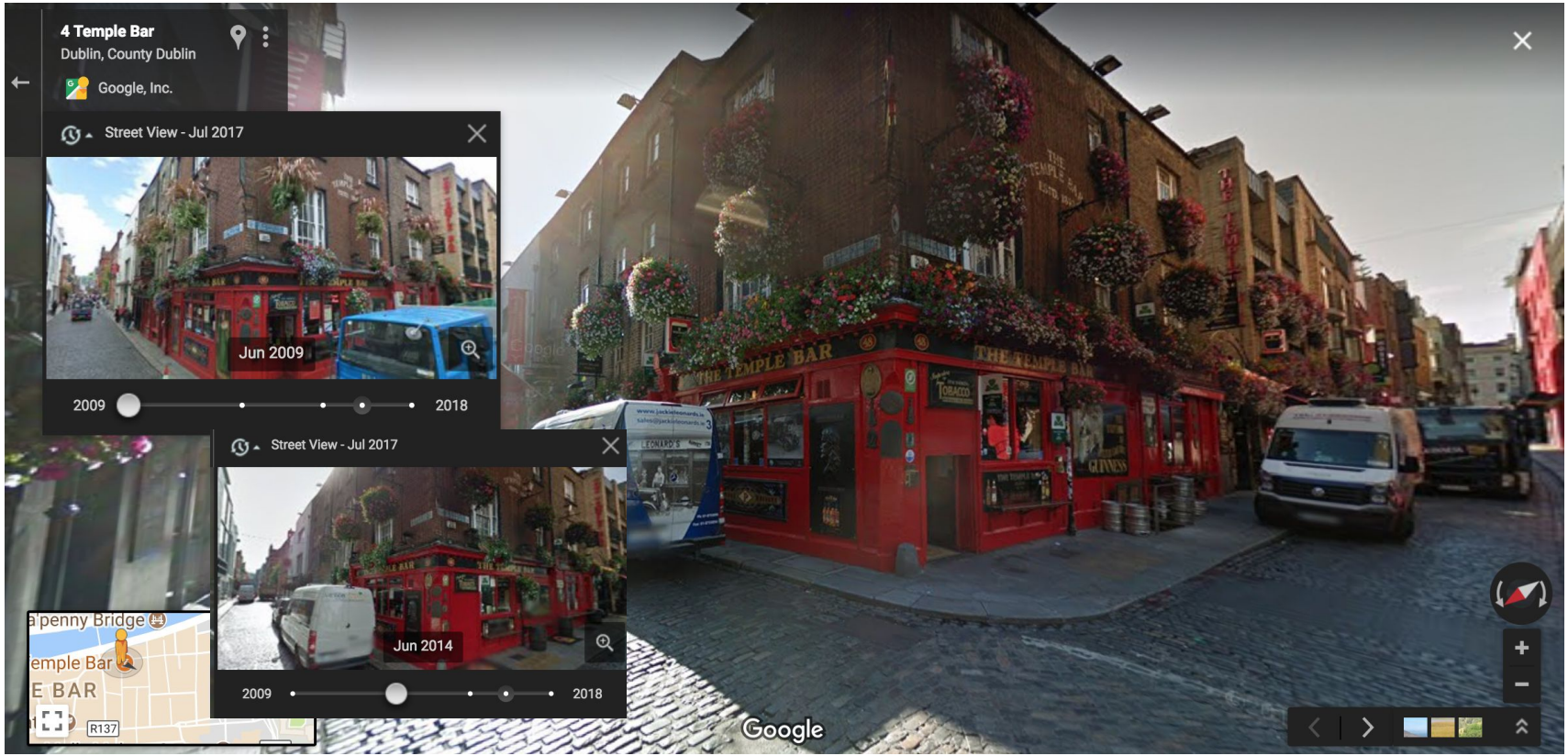
$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$



Schroff et al. FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015

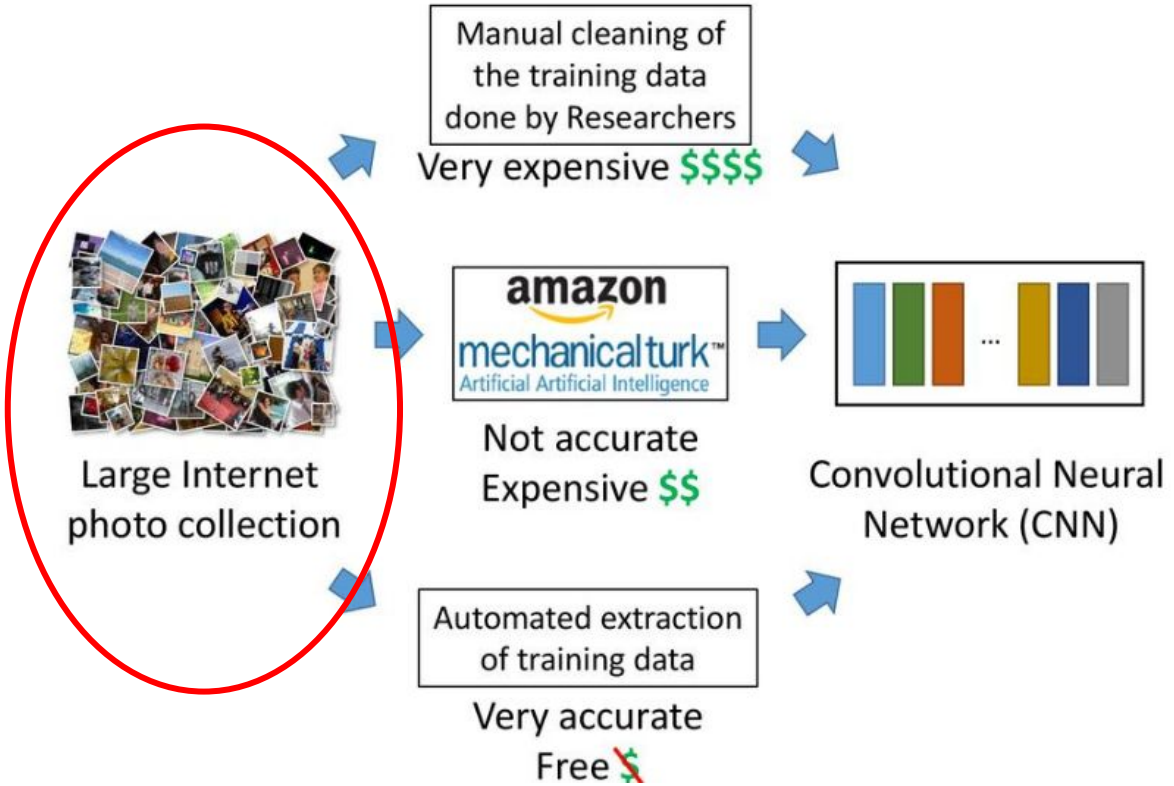
GT data generation example

Exploring Image datasets with GPS coordinates (Google Street View Time Machine)



Relja Arandjelović et al, NetVLAD: CNN architecture for weakly supervised place recognition, CVPR 2016

GT data generation example



Automatic data cleaning

Strong baseline (SIFT) + geometric verification + 3D camera estimation[1]

Further manual inspection

Further manual inspection

Radenović el at., CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples, CVPR 2016
Gordo et al., Deep Image Retrieval: Learning global representations for image search, CVPR 2016
[1] Schonberger et al. From single image query to detailed 3D reconstruction, CVPR 2015

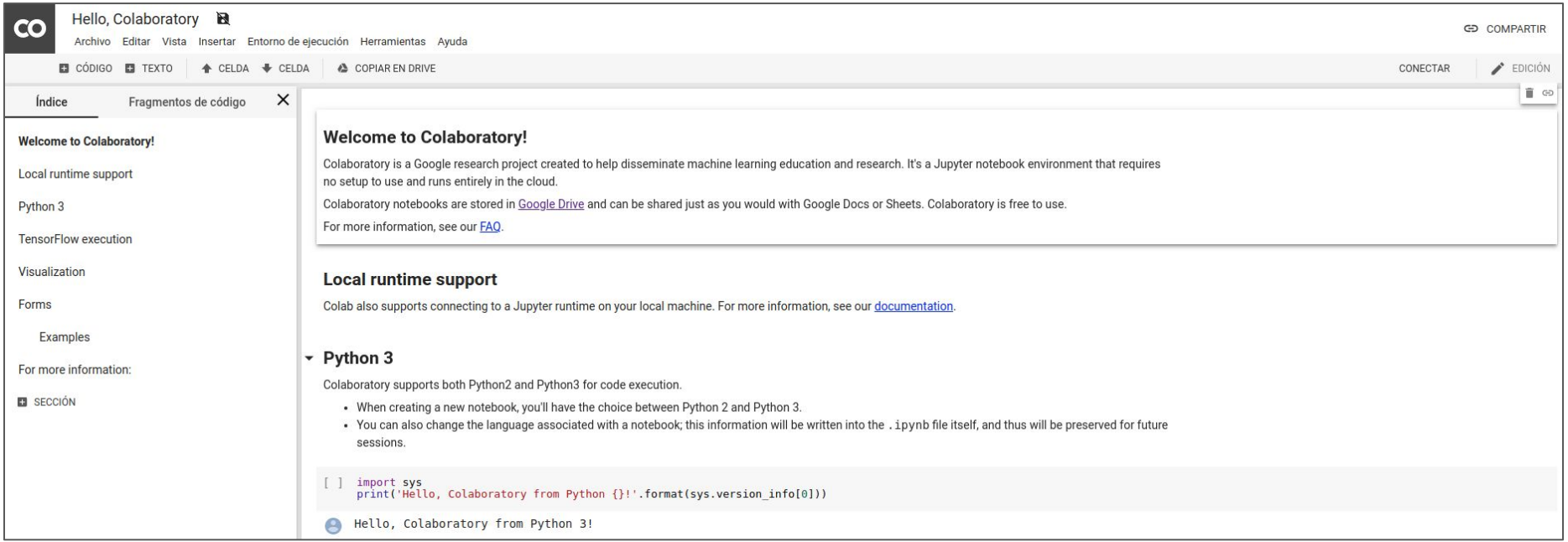
This lecture objective

Content Based Image Retrieval (CBIR)

Deep Learning approaches in CBIR

Practical session

Google Colab



https://colab.research.google.com/

This lecture objective

- Implement a retrieval pipeline
- Qualitative and quantitative evaluation of results
- Training a Siamese network

https://colab.research.google.com/drive/1nZ35NdTILmegBcOLdEXzUJkAHedfh8-K