

#DLUPC

# Image Retrieval



Eva Mohedano  
eva.mohedano@satellogic.com

Data Scientist  
Satellogic Solutions



Centre for Data Analytics



SATELLOGIC

# This lecture objective

Content Based Image Retrieval (CBIR)

Deep Learning approaches in CBIR

Practical session

# More material online

The screenshot shows a presentation slide for a Deep Learning for Computer Vision course. The title 'DEEP LEARNING FOR COMPUTER VISION' is at the top, followed by the subtitle 'Summer School at UPC Telecom-BCN Barcelona, June 29-July 6, 2018'. Below the title is a photograph of a city skyline at night. The word 'Instructors' is above a row of nine small profile pictures. Logos for 'Organized by' (UPC, Universitat Politècnica de Catalunya) and 'Supported by' (vilynx, Google Cloud Platform, Grads Education) are present. A link '+ info: <http://bit.ly/dlcv2018>' is at the bottom. To the right of the slide, there is a Twitter icon and a blue box containing the hashtag '#DLUPC'. The main content of the slide is 'Day 1 Lecture 5 Content-based Image Retrieval'. Below this, there is a portrait of a woman named Eva Mohedano, described as a Postdoctoral Researcher at the Insight Centre for Data Analytics, Dublin City University. Logos for DCU and Insight are also shown.

DEEP LEARNING  
FOR COMPUTER VISION

Summer School at UPC Telecom-BCN Barcelona, June 29-July 6, 2018

Instructors

Organized by

Supported by

+ info: <http://bit.ly/dlcv2018>

Day 1 Lecture 5

**Content-based Image Retrieval**

Eva Mohedano

Postdoctoral Researcher

Insight Centre for Data Analytics

Dublin City University

DCU

Insight

[DLCV2018](#)

# This lecture objective

**Content Based Image Retrieval (CBIR)**

Deep Learning approaches in CBIR

Practical session

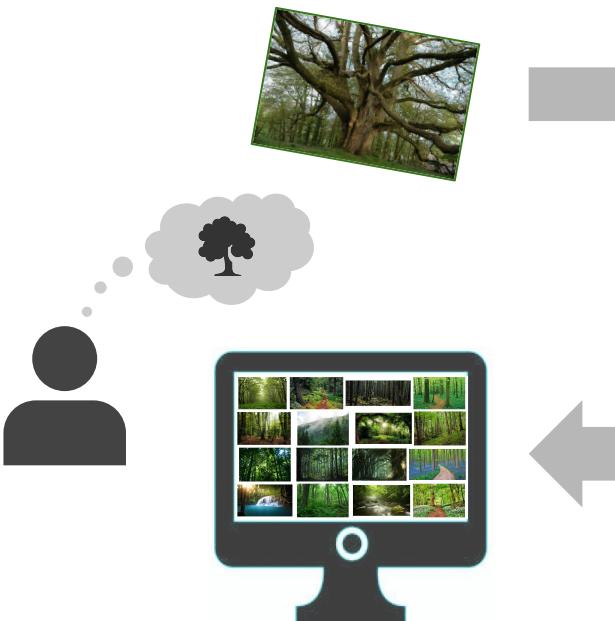
# The Problem: Query by Example

**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



Planters At Screwfix - Here When...  
Great Range Of Planters At Trade Prices. Buy  
Online, Collect In Store.  
[www.screwfix.com](http://www.screwfix.com)

Related Images [See All >](#)



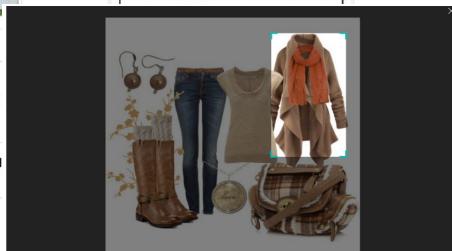
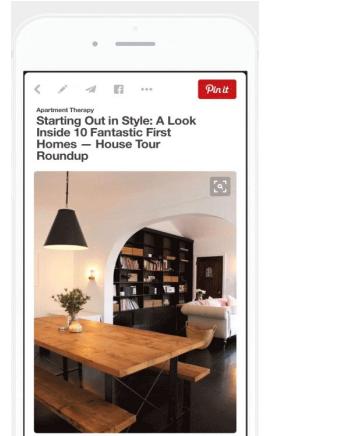
Web Results

**white wooden planters | eBay**

Find great deals on eBay for white wooden planters and white square wooden planters. Shop with confidence.  
[www.ebay.co.uk](http://www.ebay.co.uk)

**Amazon.co.uk: Window Boxes: Garden & Outd**

Online shopping for Window Boxes from a great selection at  
[Garden & Outdoors Store](#)



seeking  
information

Google  
Images

Search by image  
Search Google with an image instead of text. Try dragging an image here.

Paste Image URL  Upload an Image

Choose File  No file chosen

How to upload an image  
• Use the button below to open an image that's on your computer.  
Google will automatically upload and search using the image.

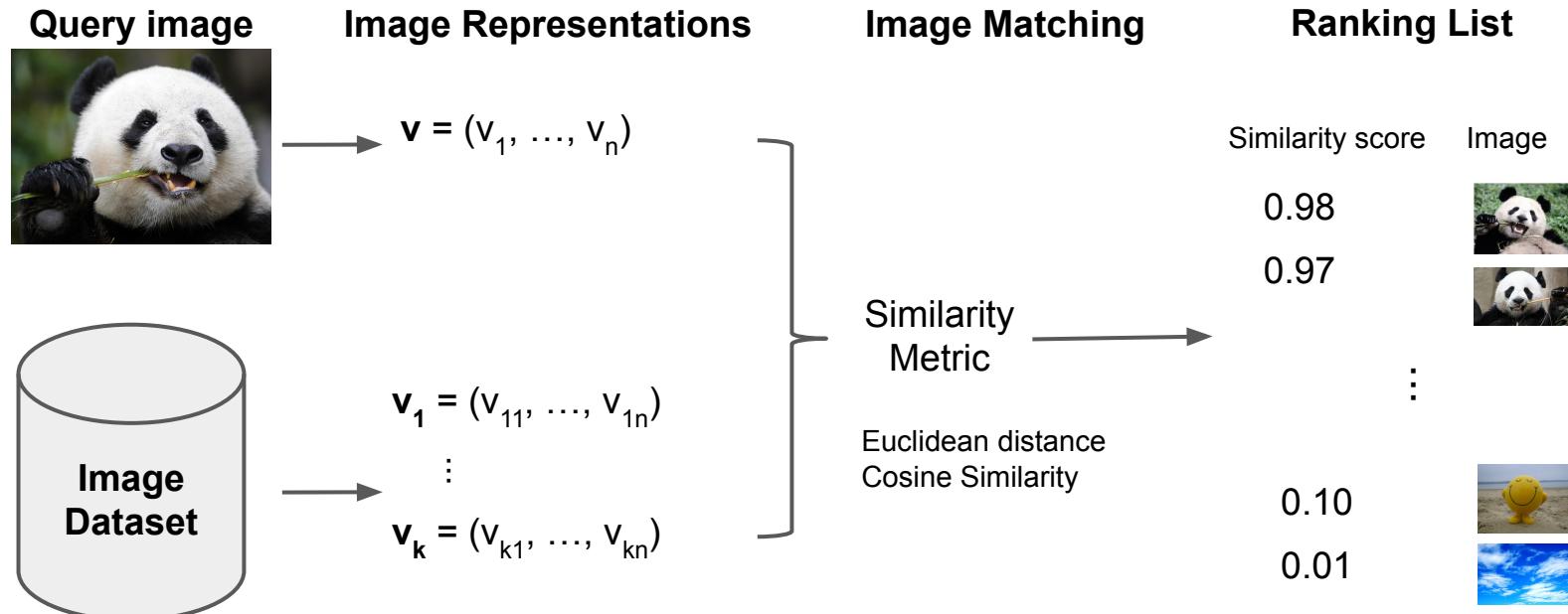
Tip: Try dragging an image into the search box from your desktop or the web.

[Learn about search by image](#)

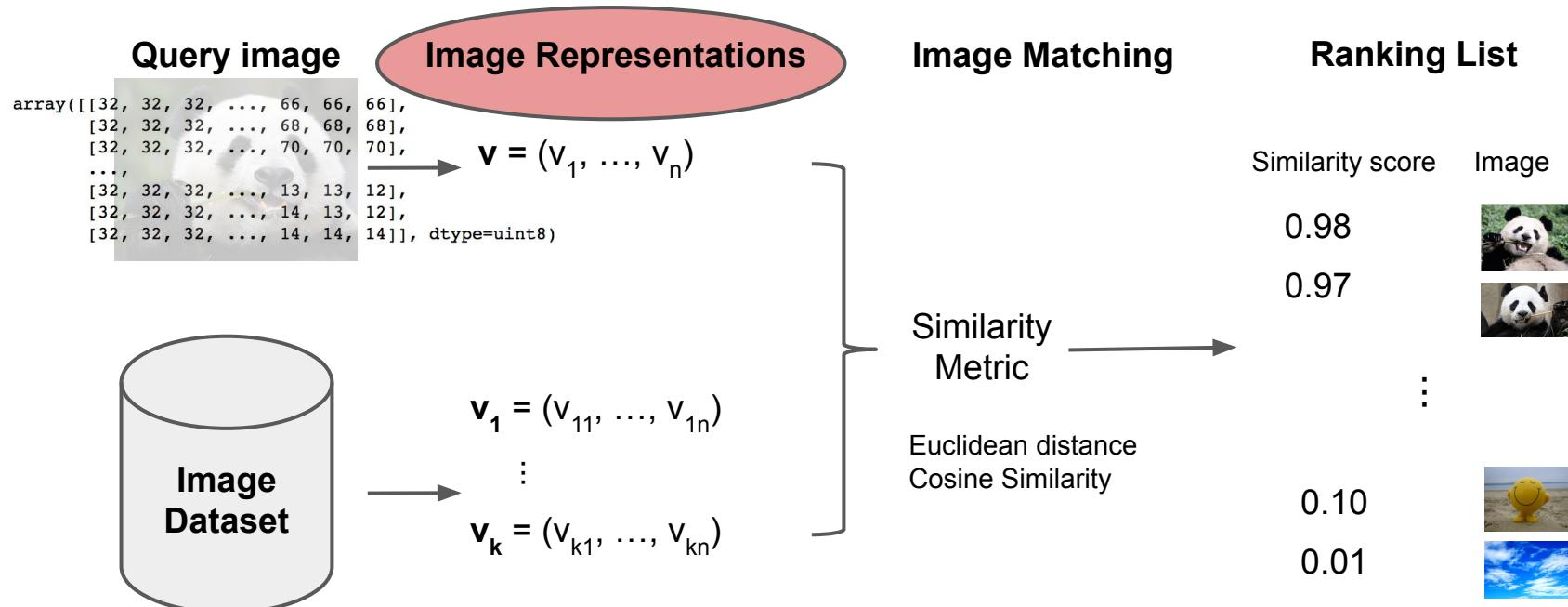
personal photo organization



# CBIR Pipeline



# CBIR Pipeline



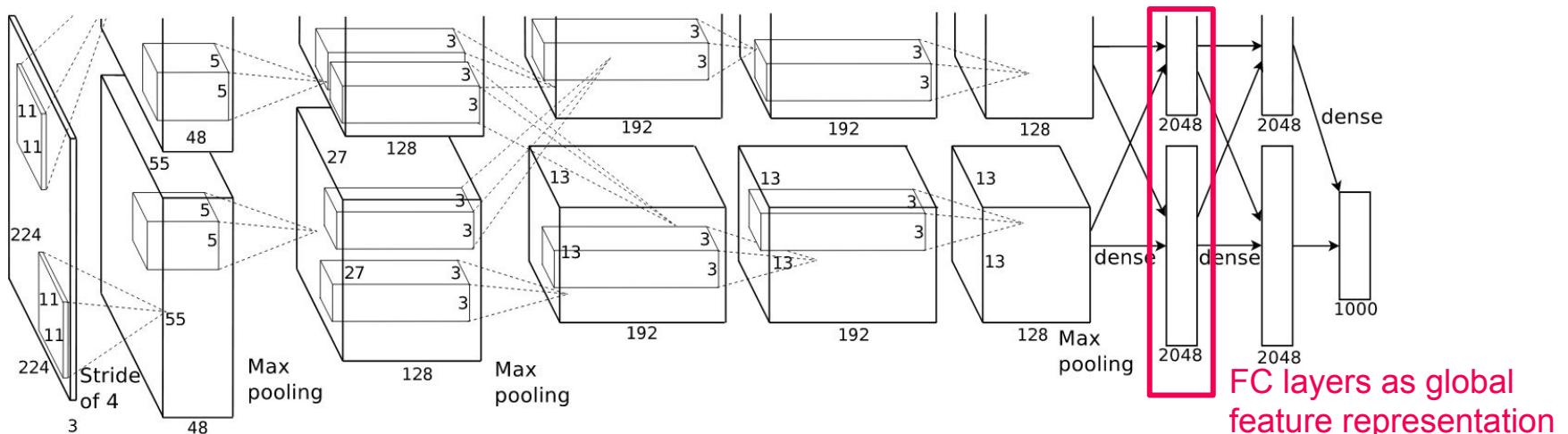
# This lecture objective

Content Based Image Retrieval (CBIR)

**Deep Learning approaches in CBIR**

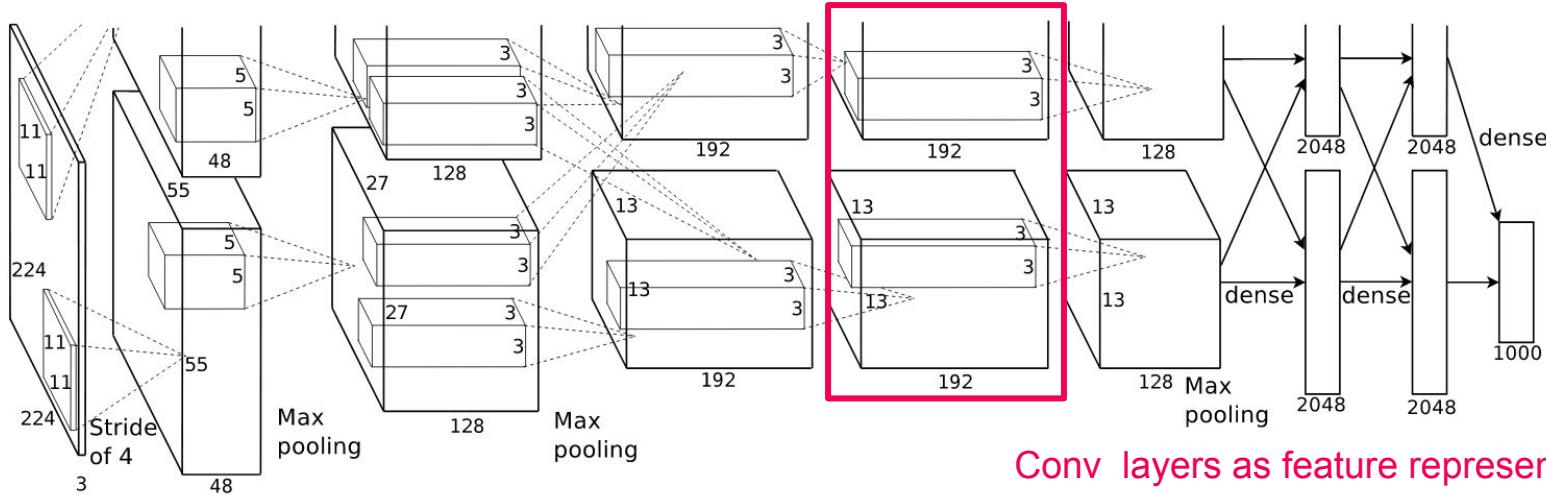
Practical session

# “Off-the-shelf” classification models



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

# “Off-the-shelf” classification models



Conv layers as feature representation

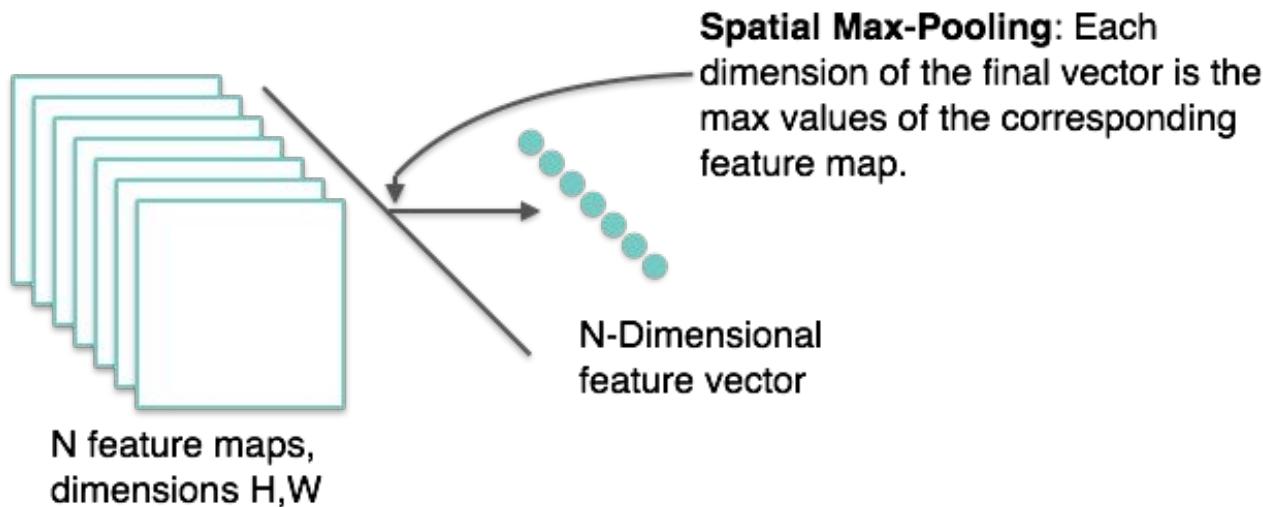
- Spatial information is encoded
- Need to aggregate local information into a single vector
- Better performance than traditional approaches

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

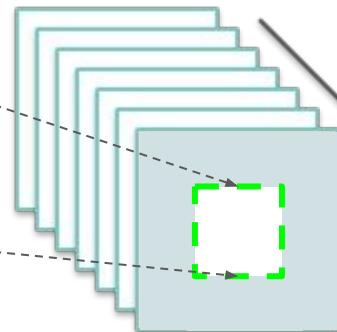


Lower level of abstraction (off-the-self model)

Allow processing images at higher resolution 12

# Representation from Conv layers

Aggregate information over specific regions



$N$  feature maps,  
dimensions  $H, W$

**Spatial Max-Pooling:** Each dimension of the final vector is the max values of the corresponding feature map.



$N$ -Dimensional  
feature vector

# “Off-the-shelf” network limitations

Query: This chair



Classification



Results from dataset classified as “chair”

# “Off-the-shelf” network limitations

Query: This chair



Retrieval



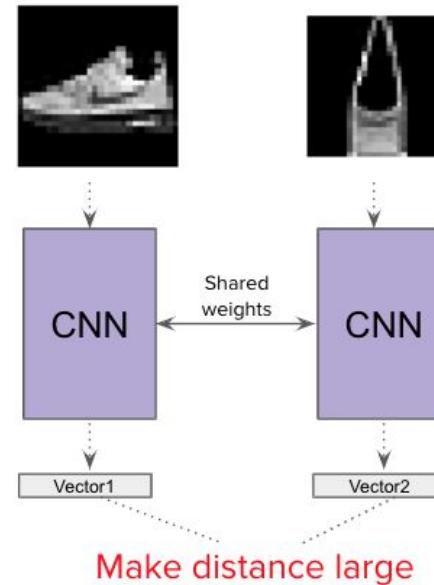
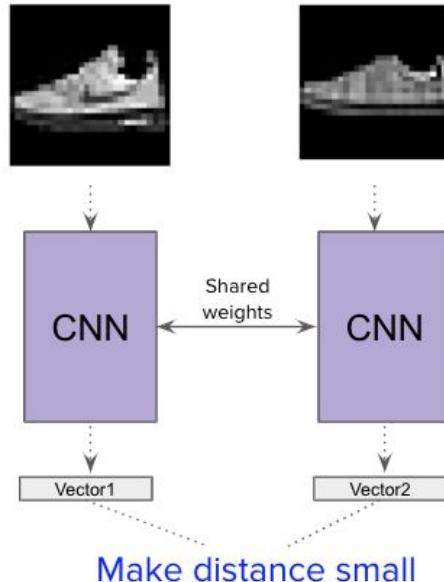
Results from dataset **ranked** by similarity to the query

# Siamese Networks for retrieval

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

Positive Pairs

Negative Pairs



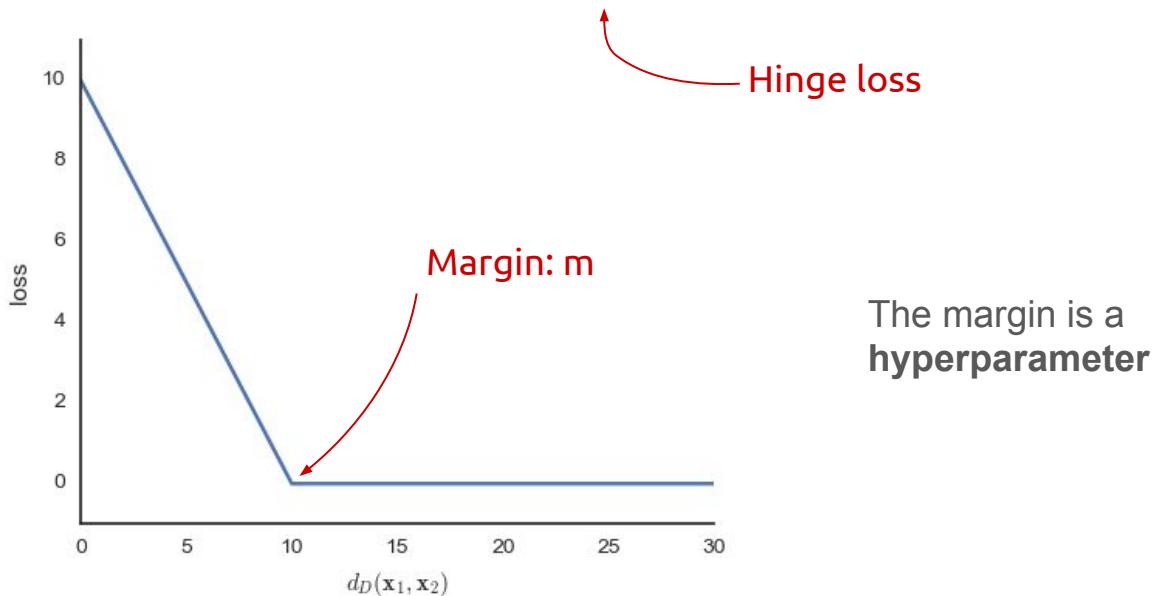
$$l_P(d_D(\mathbf{x}_1, \mathbf{x}_2)) = d_D(\mathbf{x}_1, \mathbf{x}_2)$$

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

# Hinge loss

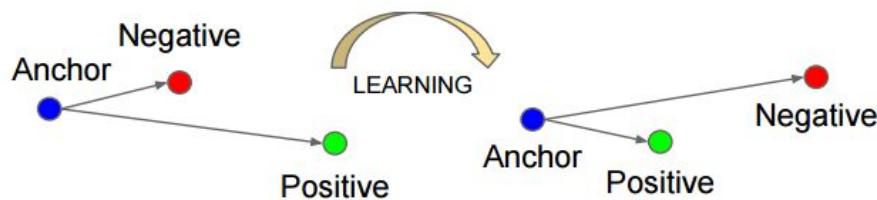
$$l_N(d_D(\mathbf{x}_1, \mathbf{x}_2)) = \max(0, m - d_D(\mathbf{x}_1, \mathbf{x}_2))$$

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

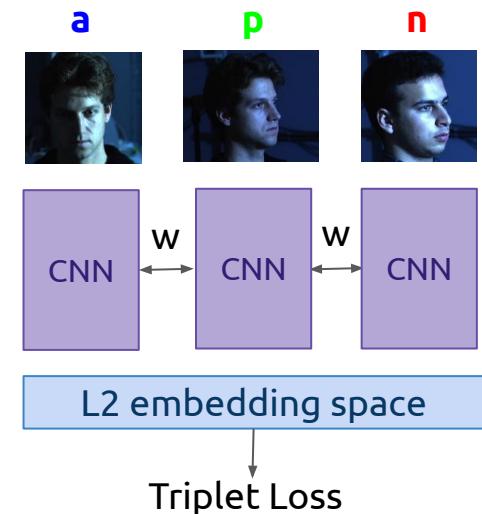


# Triplet Networks

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

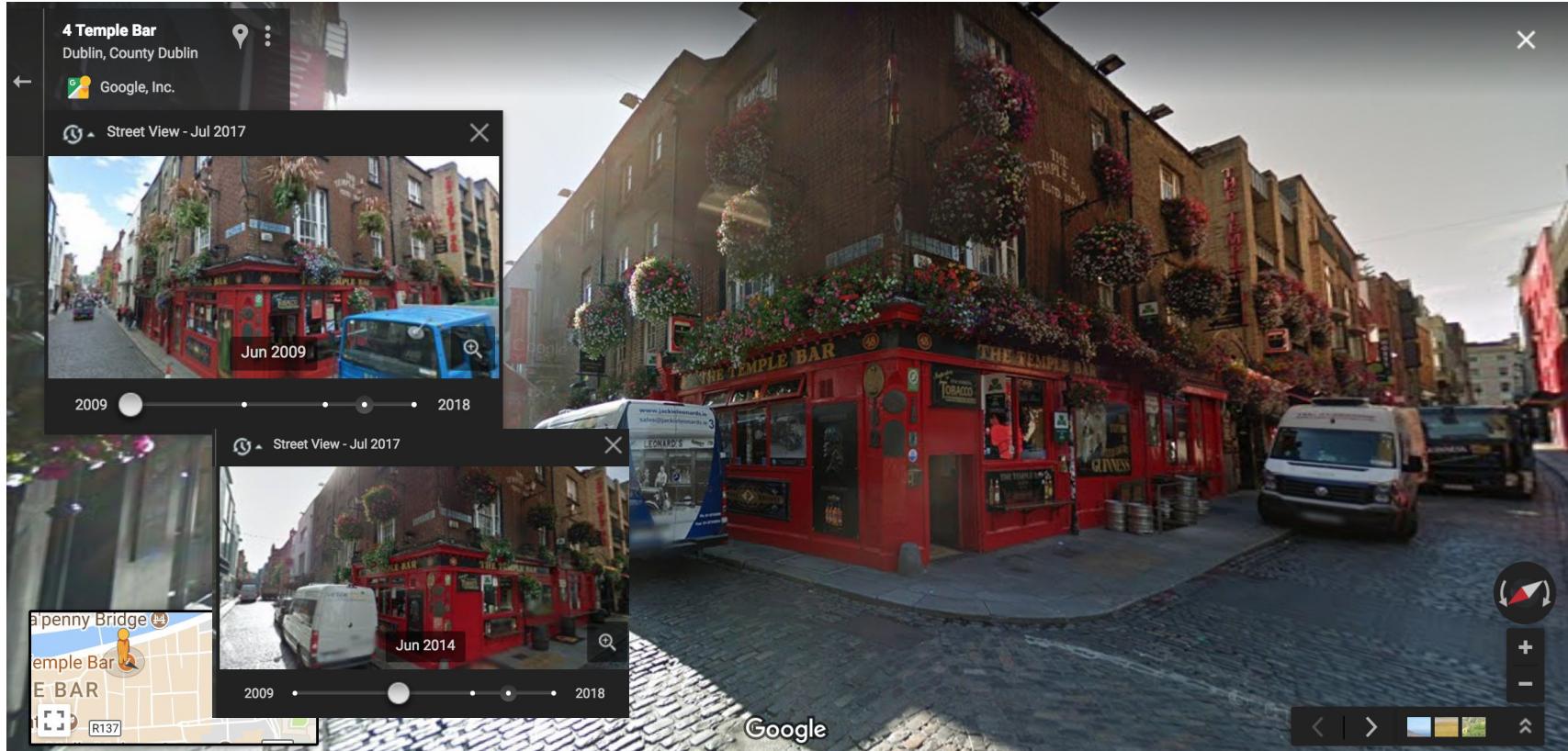


$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

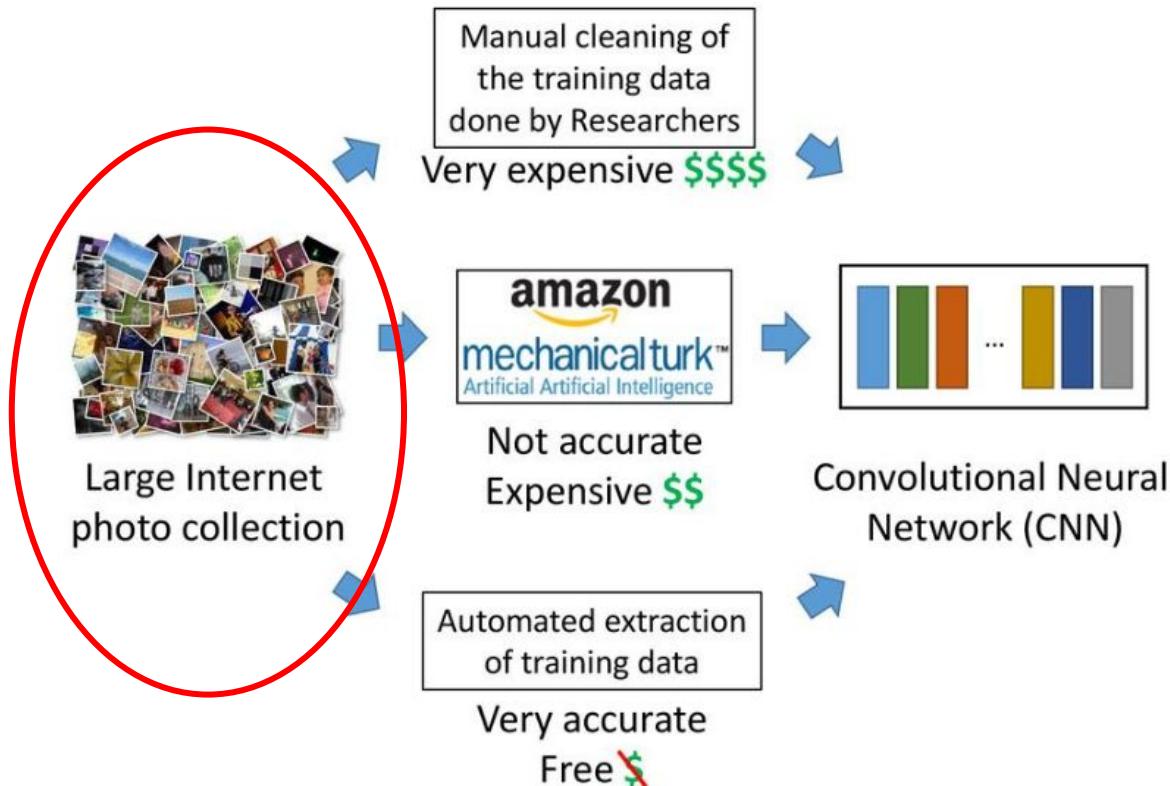


# GT data generation example

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



# GT data generation example



## Automatic data cleaning

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

## Further manual inspection

Further manual inspection

Radenović et al., [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

The screenshot shows the Google Colab interface. At the top, there's a navigation bar with options like Archivo, Editar, Vista, Insertar, Entorno de ejecución, Herramientas, Ayuda, and a CO icon. Below the bar are buttons for CÓDIGO, TEXTO, CELDA, CELDA, COPIAR EN DRIVE, CONECTAR, and EDICIÓN. On the left, a sidebar titled 'Índice' lists various sections: Welcome to Colaboratory!, Local runtime support, Python 3, TensorFlow execution, Visualization, Forms, Examples, and For more information: SECCIÓN. The main content area displays the 'Welcome to Colaboratory!' page, which includes a heading, a paragraph about Colaboratory, a note about Google Drive, and a link to the FAQ. Below this, the 'Local runtime support' section is shown, followed by a expanded 'Python 3' section with a bullet list and a code cell that outputs 'Hello, Colaboratory from Python 3!'. The URL <https://colab.research.google.com/> is visible at the bottom of the page.

<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>