What makes
Big Visual Data hard?

© Quint Buchholz

Jun-Yan Zhu *Alexei A. Efros*

UC Berkeley

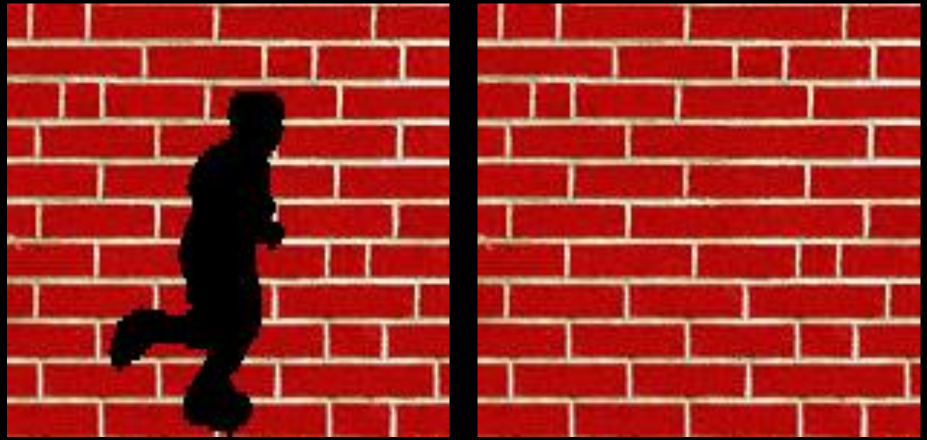
Our Goals

1. To make you fall in love with **Big <u>Visual</u> Data**
   - Very difficult to handle.
   - but holds the key to achieving real visual understanding

2. To discuss the challenges and ask for help in tackling this **Big Data Problem**

Driven by Visual Data

Texture Synthesis



Dating Historical Images

Seeing Through Water

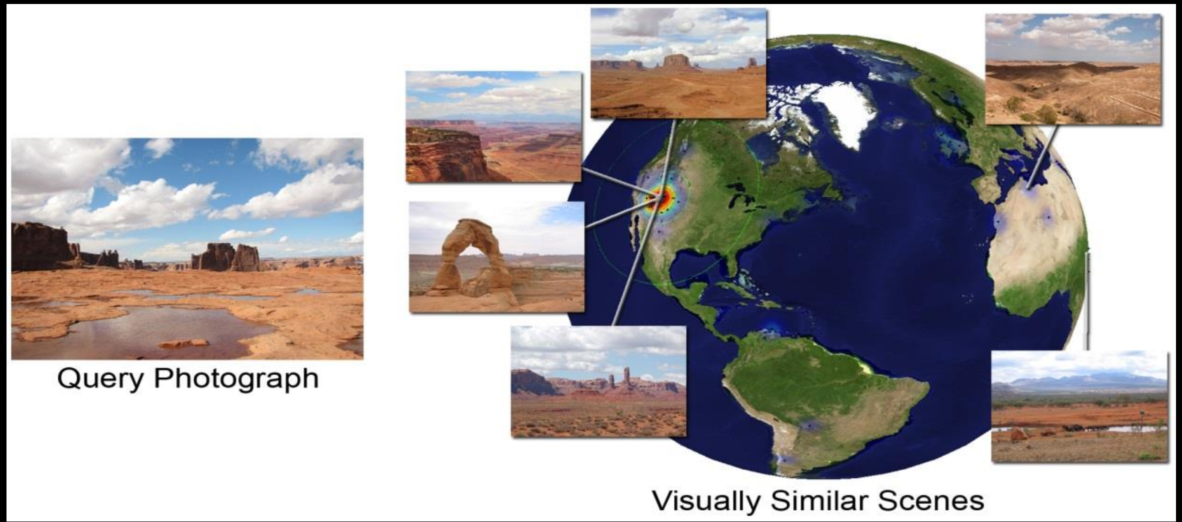
Unsupervised Object Discovery

Action Recognition

Illumination Estimation

Inferring 3D from 2D

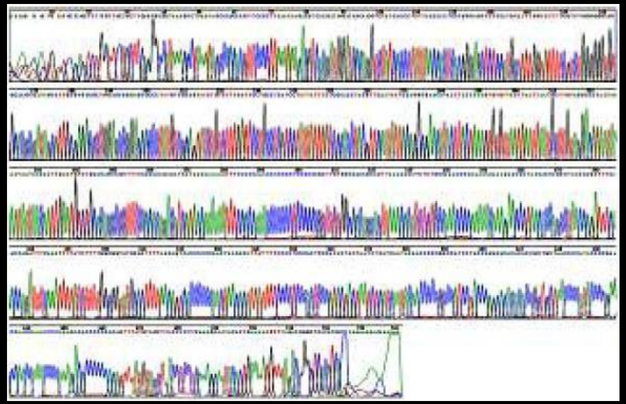
Geo-location



Two Kinds of Things in the World

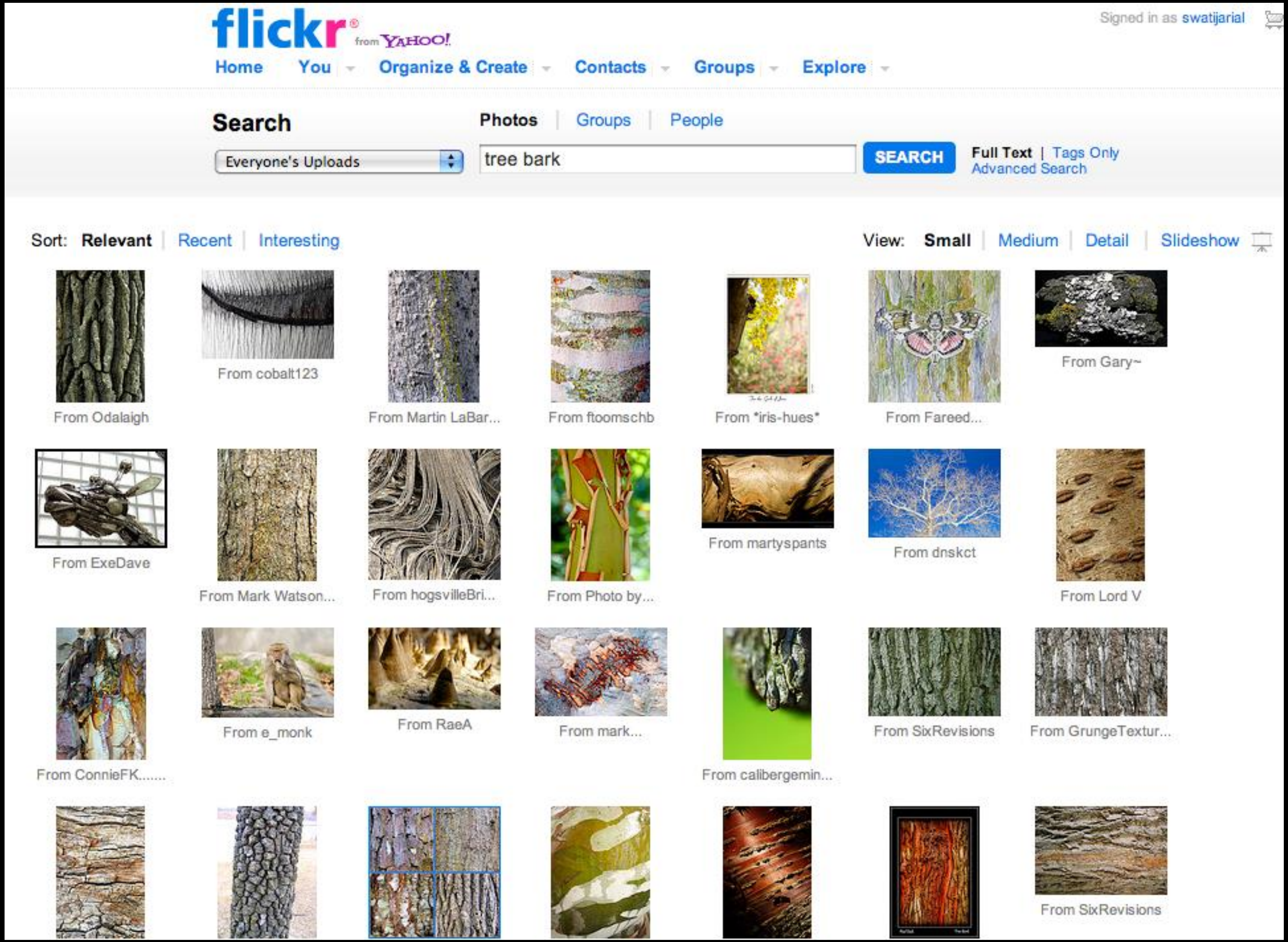
Navier-Stokes Equation

$$\frac{\partial \mathbf{u}}{\partial t} = -(\mathbf{u} \cdot \nabla)\mathbf{u} + v\nabla^2\mathbf{u} - \frac{1}{d}\nabla p + \mathbf{f}$$



+ weather
+ location
+ …

Lots of data available



“Unreasonable Effectiveness of Data”

[Halevy, Norvig, Pereira 2009]

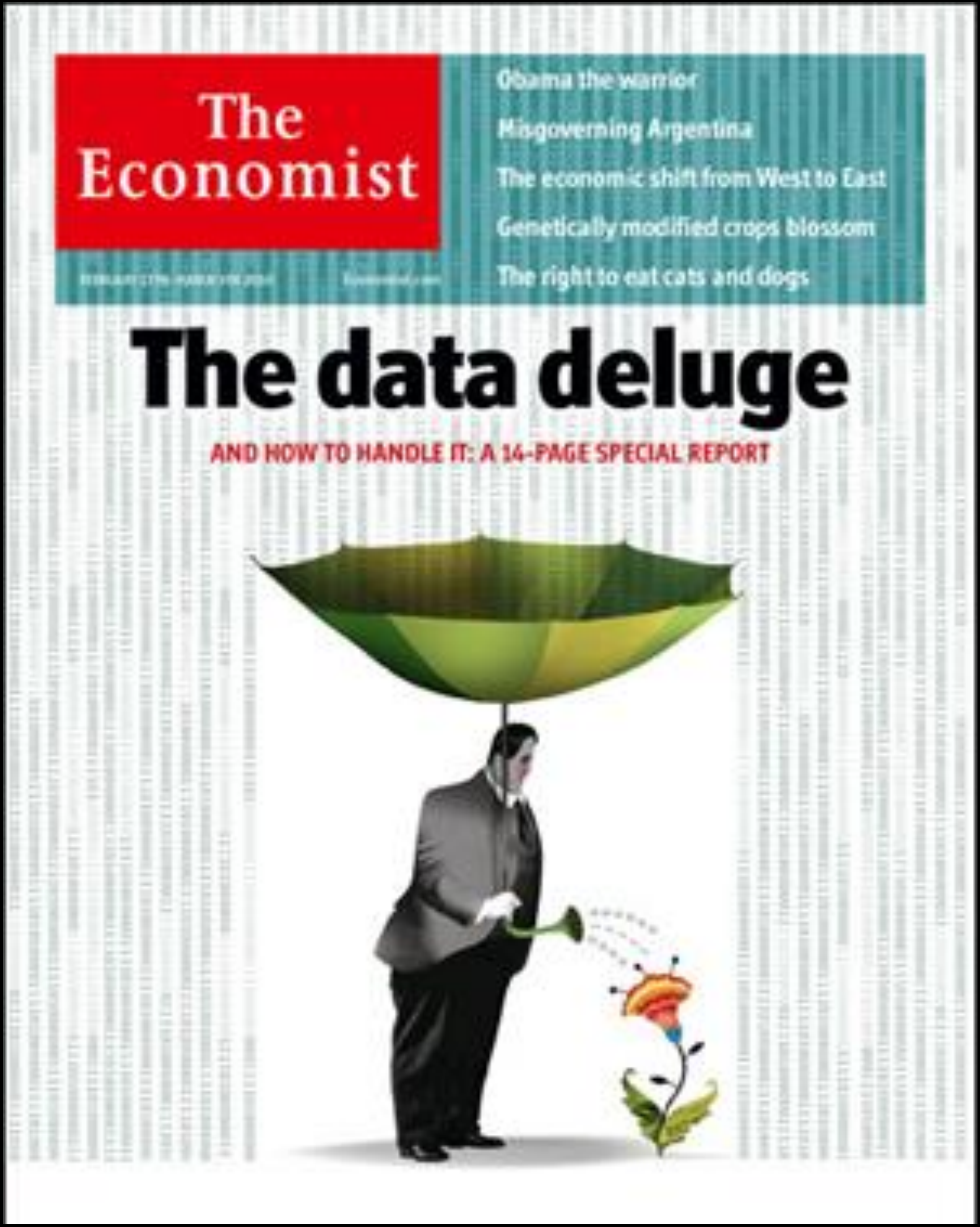
- Parts of our world can be explained by elegant mathematics:
  - physics, chemistry, astronomy, etc.
- But much cannot:
  - psychology, genetics, economics,… visual understanding?
- Enter: The **Magic of Data**
  - Great advances in several fields:
    - e.g. speech recognition, machine translation, Google

The A.I. for the postmodern world

The Good News

Really stupid algorithms + Lots of Data

= “Unreasonable Effectiveness”


The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Big Visual Data

flickr

6 billion images

the simple image sharer
imgur

1 billion images served daily

YouTube

100 hours uploaded per minute

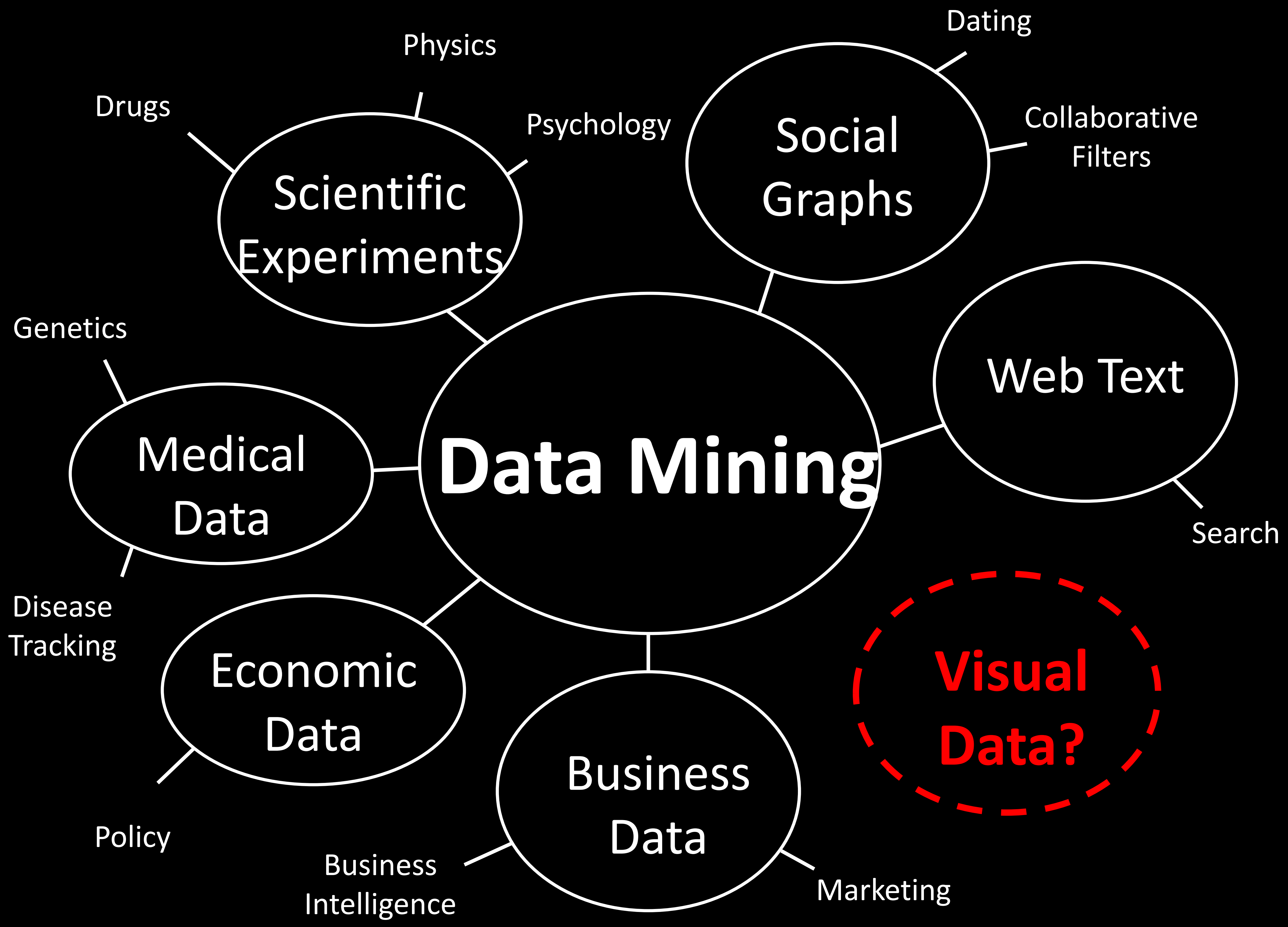
3.5 trillion photographs

facebook

70 billion images

cisco

Almost 90% of web traffic is visual!


Data Mining
Scientific Experiments
Drugs
Physics
Psychology
Social Graphs
Dating
Collaborative Filters
Web Text
Search
Medical Data
Genetics
Disease Tracking
Economic Data
Policy
Business Data
Business Intelligence
Marketing
Visual Data?

The Bad News

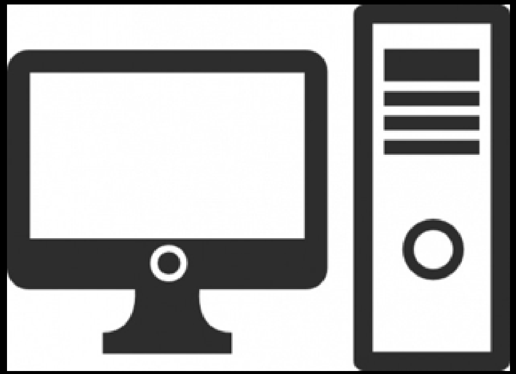
Visual Data is difficult to handle

- text:
  - clean, segmented, compact, 1D, indexable
- Visual data:
  - Noisy, unsegmented, high entropy, 2D/3D

What makes Big Visual Data hard?

for Computers

1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data

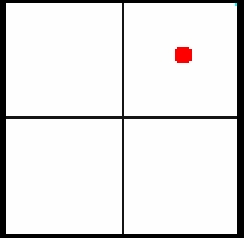
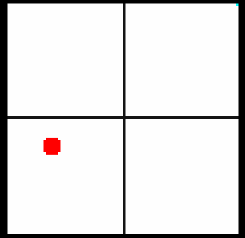


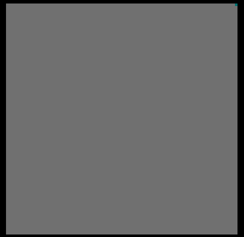
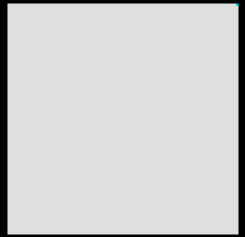
for Human Beings

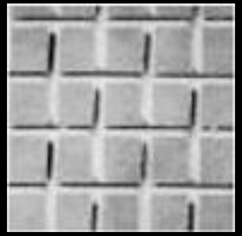
1. Visualizing Visual Data
2. Visual Communication

Computing distances is hard

CLIME - CRIME = hamming distance of 1 letter

 -  = Euclidian distance of 5 units

 -  = Grayvalue distance of 50 values

 -  = ?

How similar are two pictures?

$\stackrel{?}{=}$

Visual "Garbage Heap"

"It irritated him that the "dog" of 3:14 in the afternoon, seen in profile, should be indicated by the same noun as the dog of 3:15, seen frontally…"

"My memory, sir, is like a garbage heap."

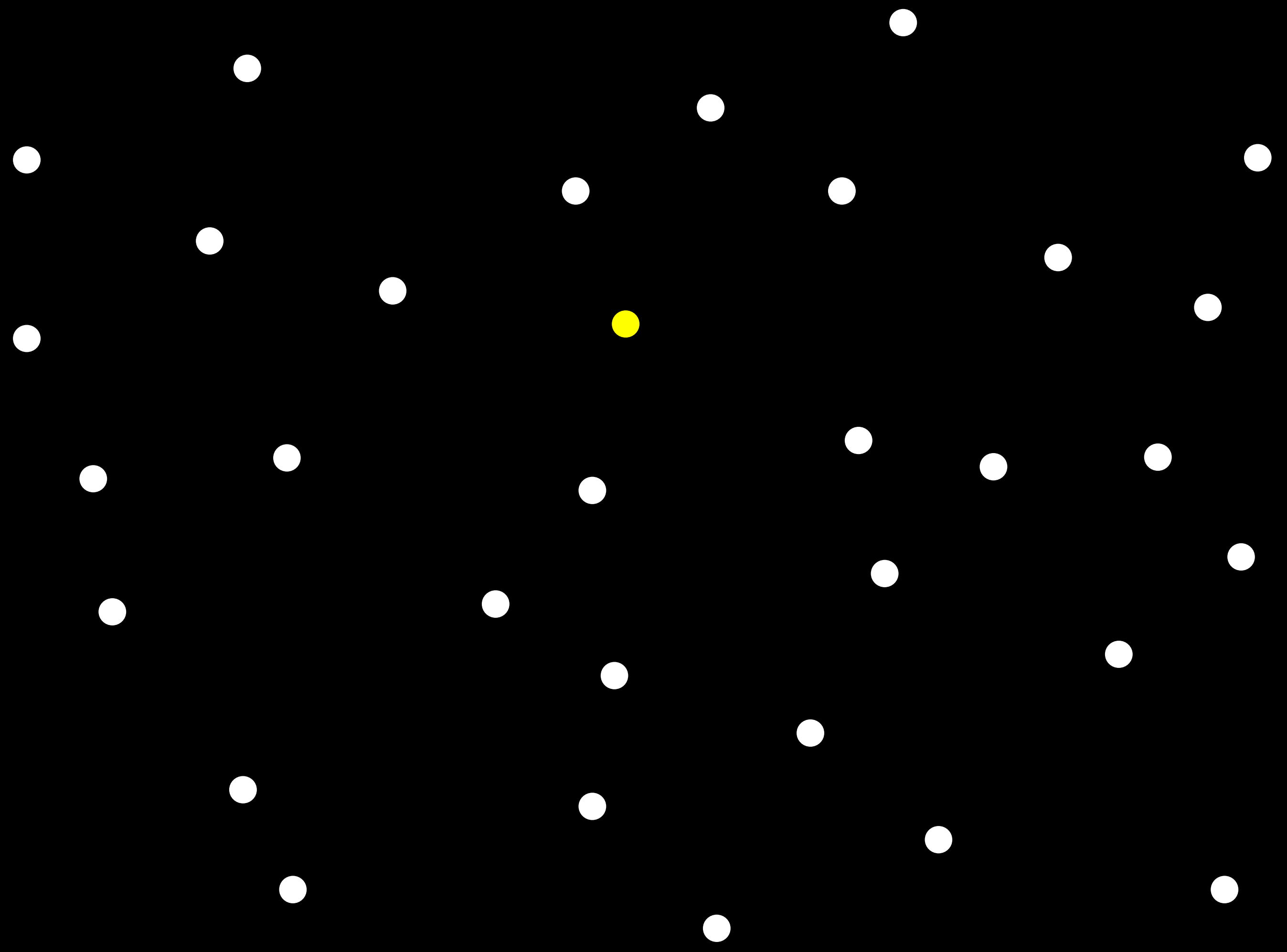
-- from *Funes the Memorious*

Jorge Luis Borges

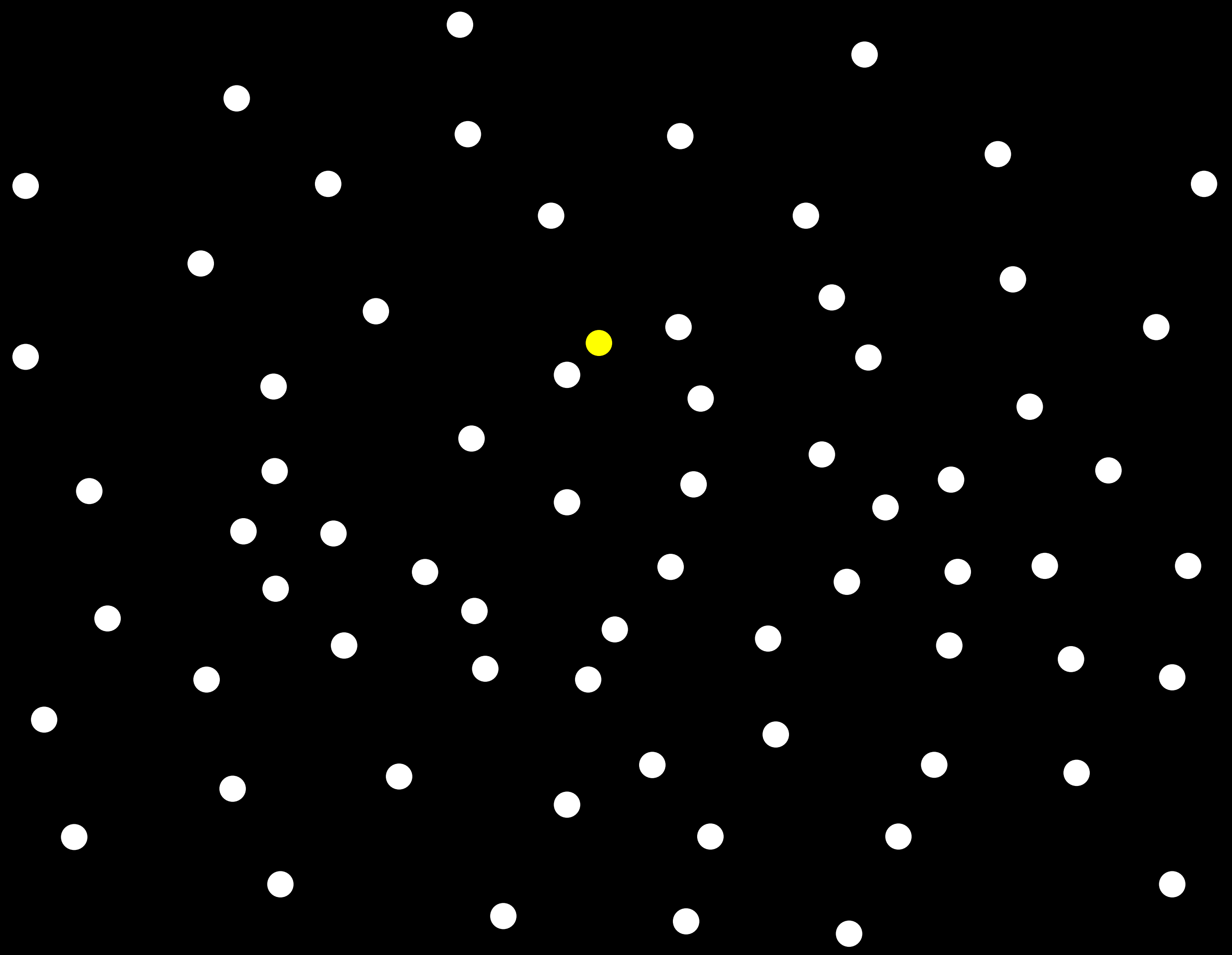
Organizing the "Garbage Heap":

- Finding visual correspondences across data
- Mining Visual Data
- Connecting visual data to enable understanding (Visual Memex)

Improving Visual Correspondence

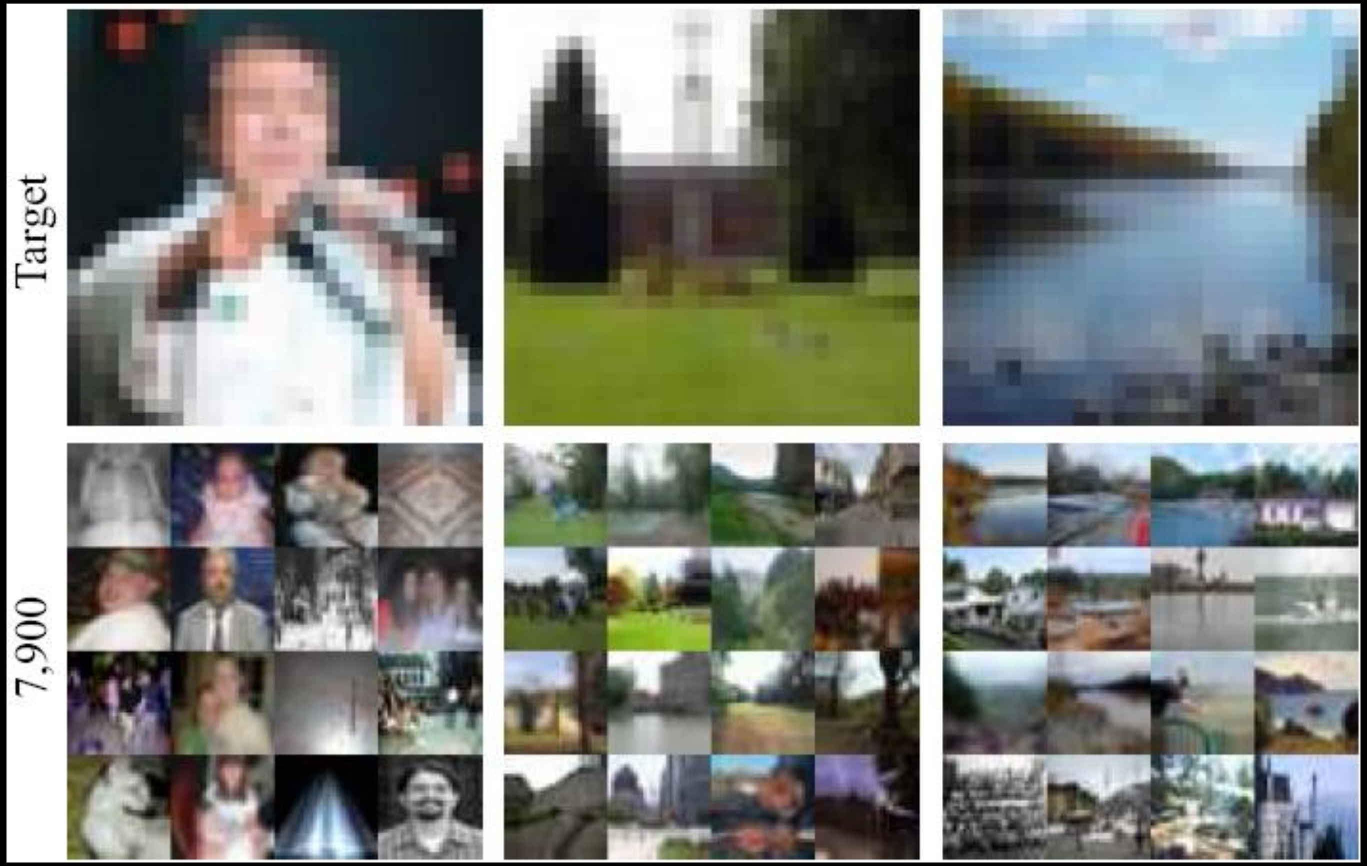


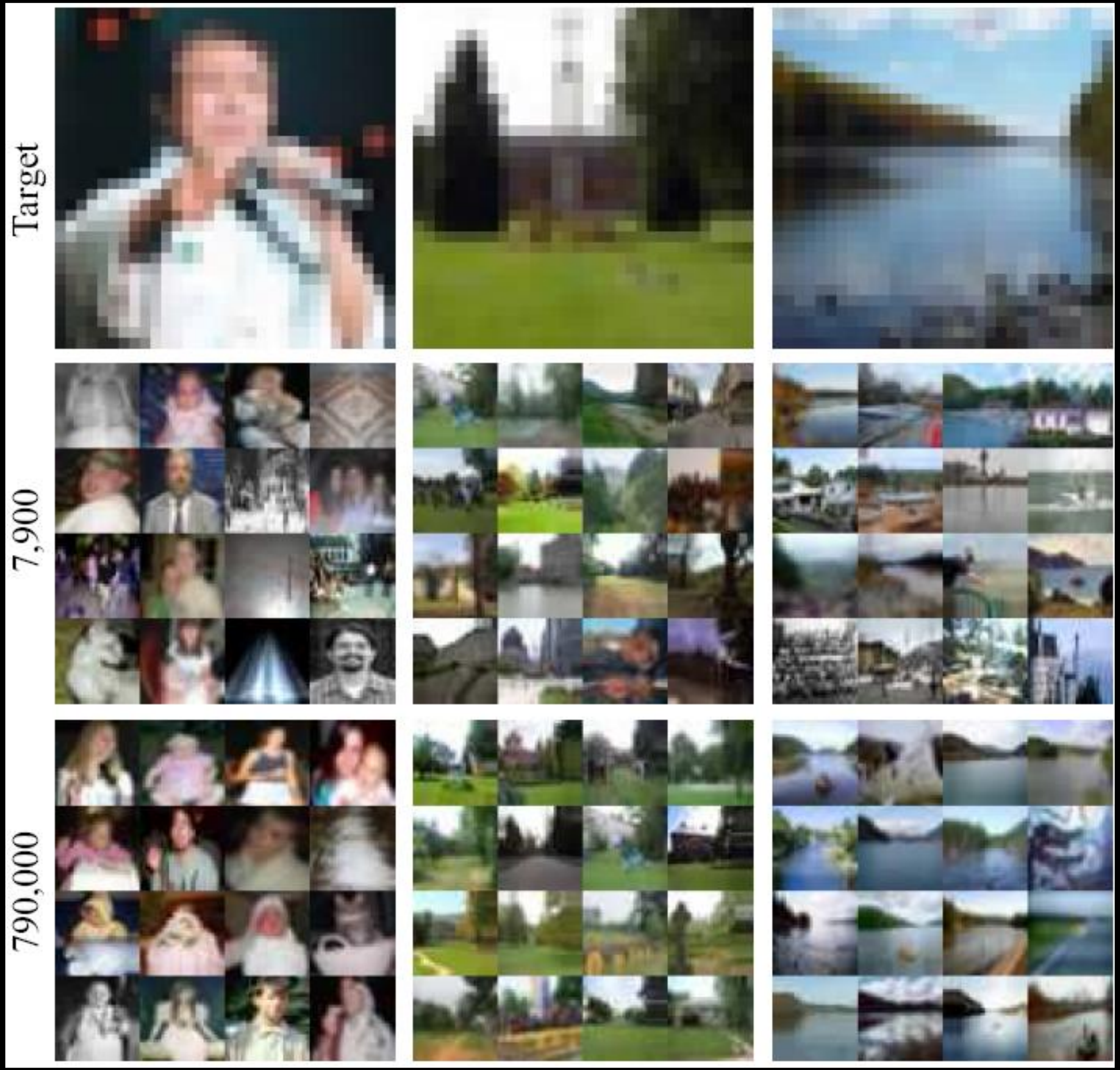
Improving Visual Correspondence

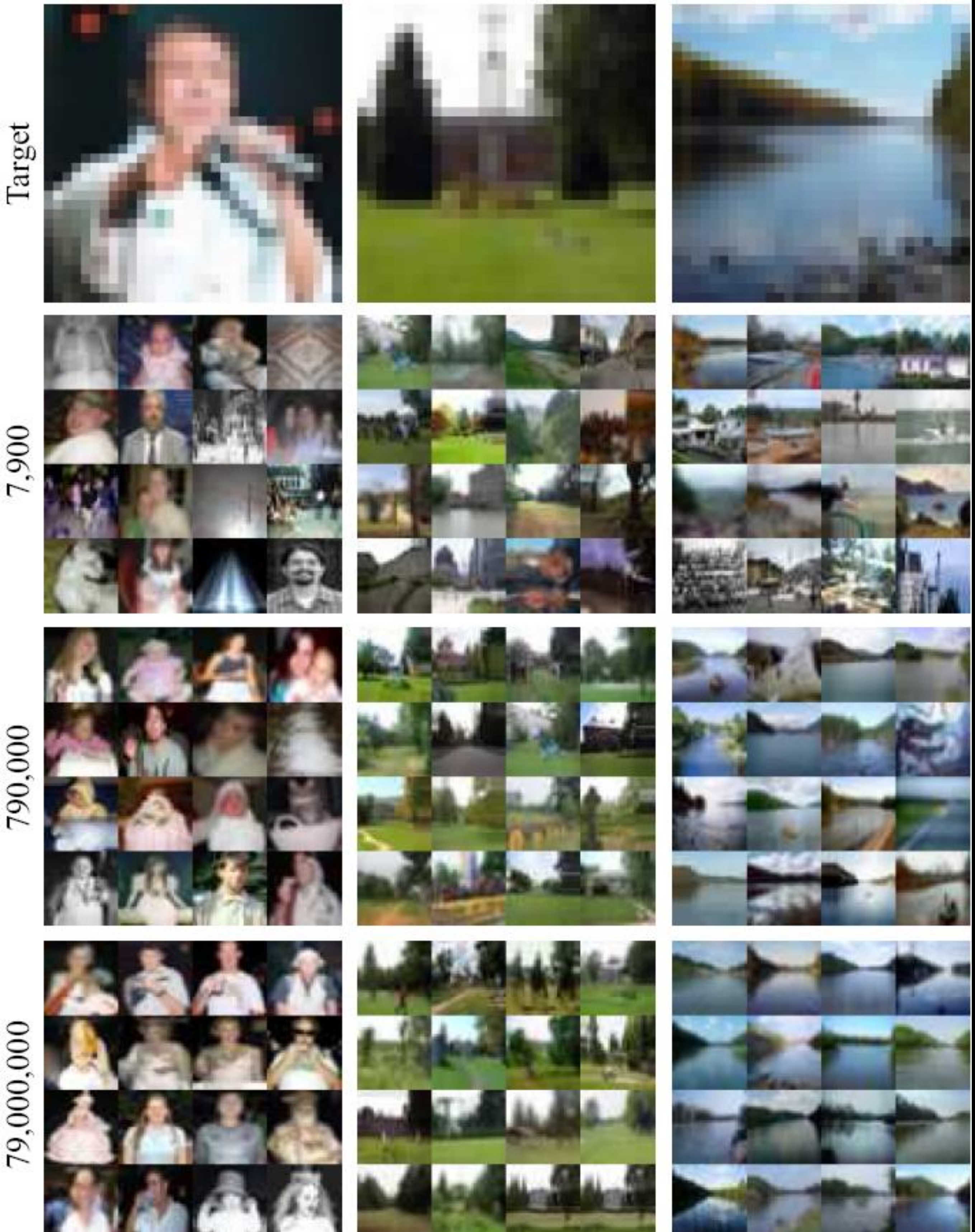


Lots of Tiny Images

- 80 million tiny images: a large dataset for non-parametric object and scene recognition **Antonio Torralba, Rob Fergus and William T. Freeman**. PAMI 2008.


Target
7,900


Target
7,900
790,000

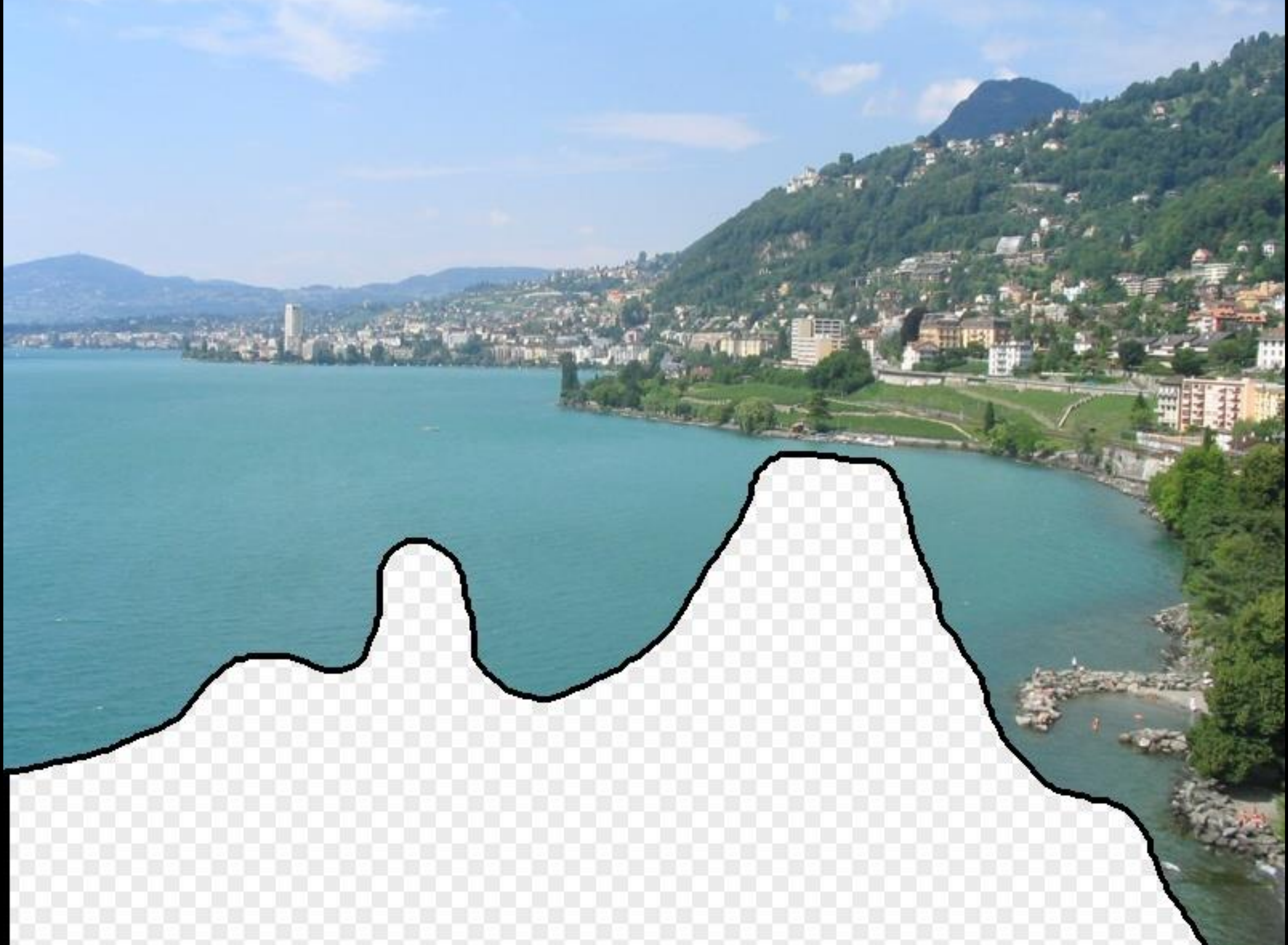

Target
7,900
790,000
79,000,000

Automatic Colorization

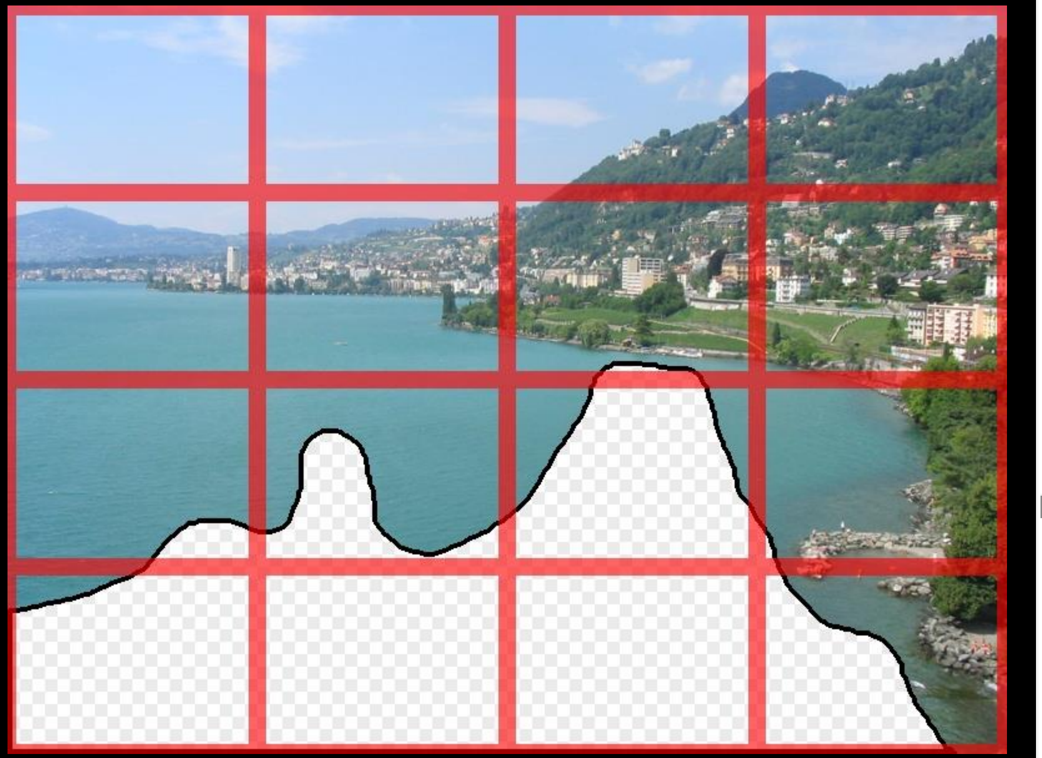
Simple Distance Metric + More Data

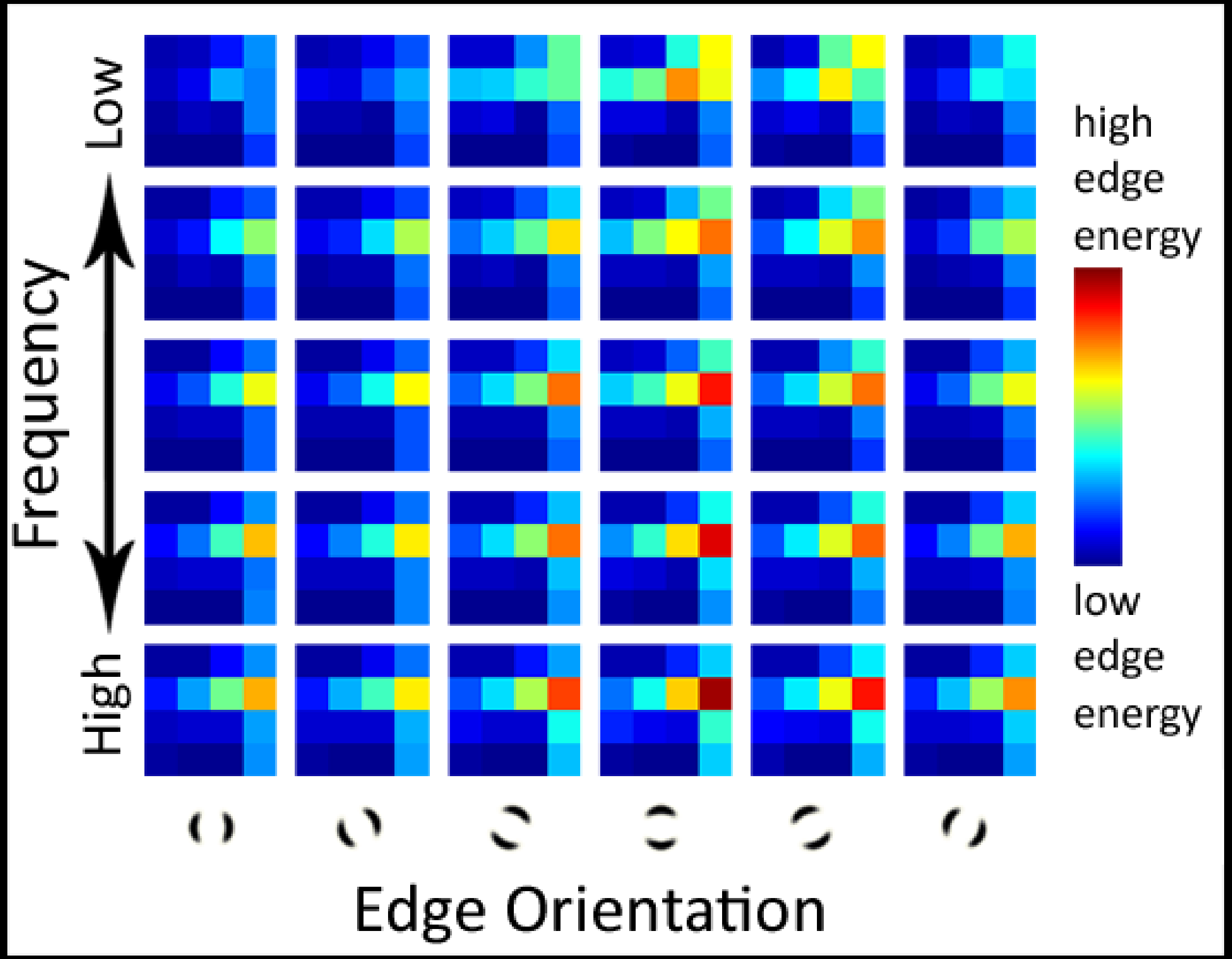
James Hays, Alexei A. Efros. *Scene Completion Using Millions of Photographs*. **SIGGRAPH** 2007


[Hays & Efros, SIGGRAPH’07]



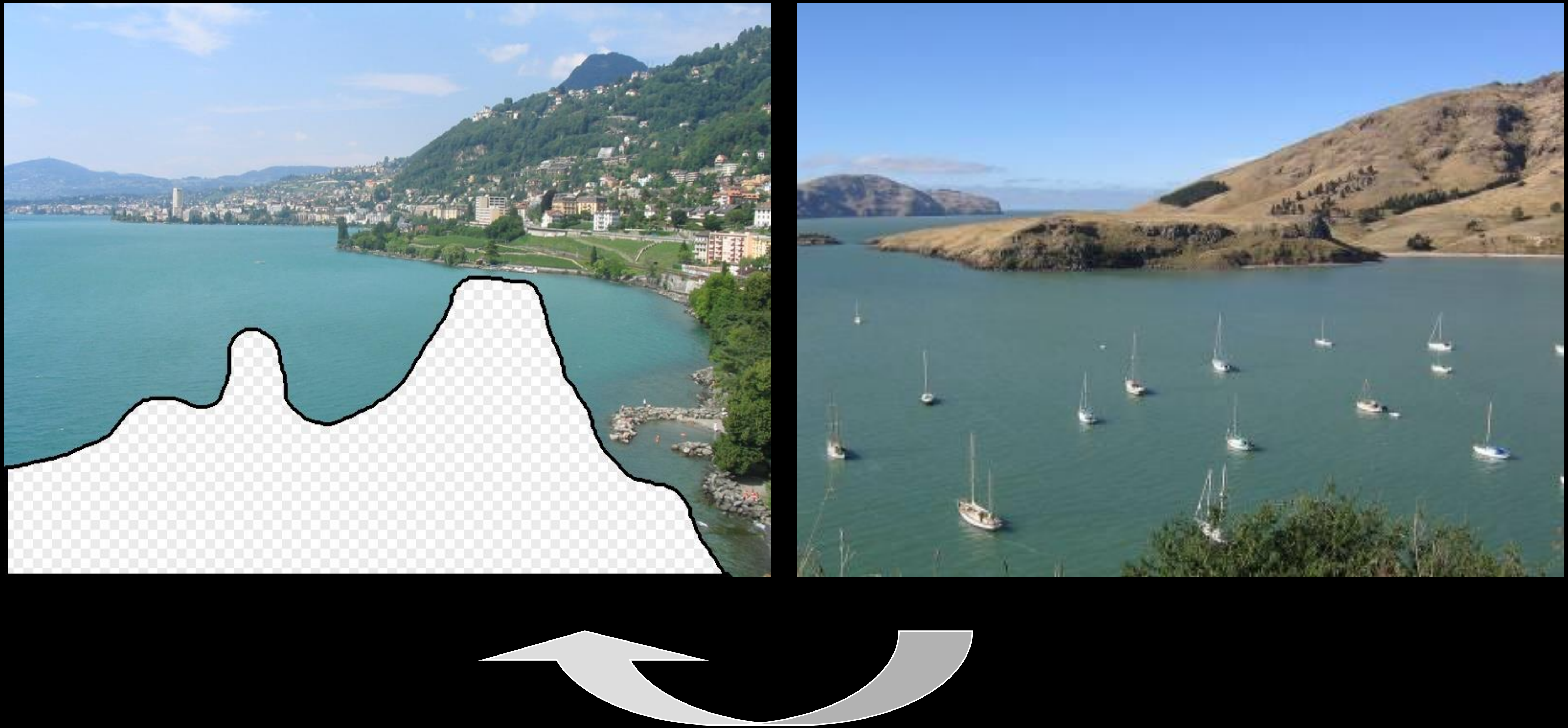
Scene Descriptor

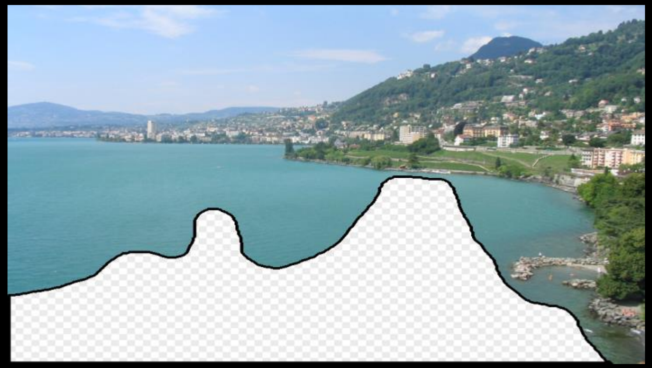




[Oliva & Torralba 01']

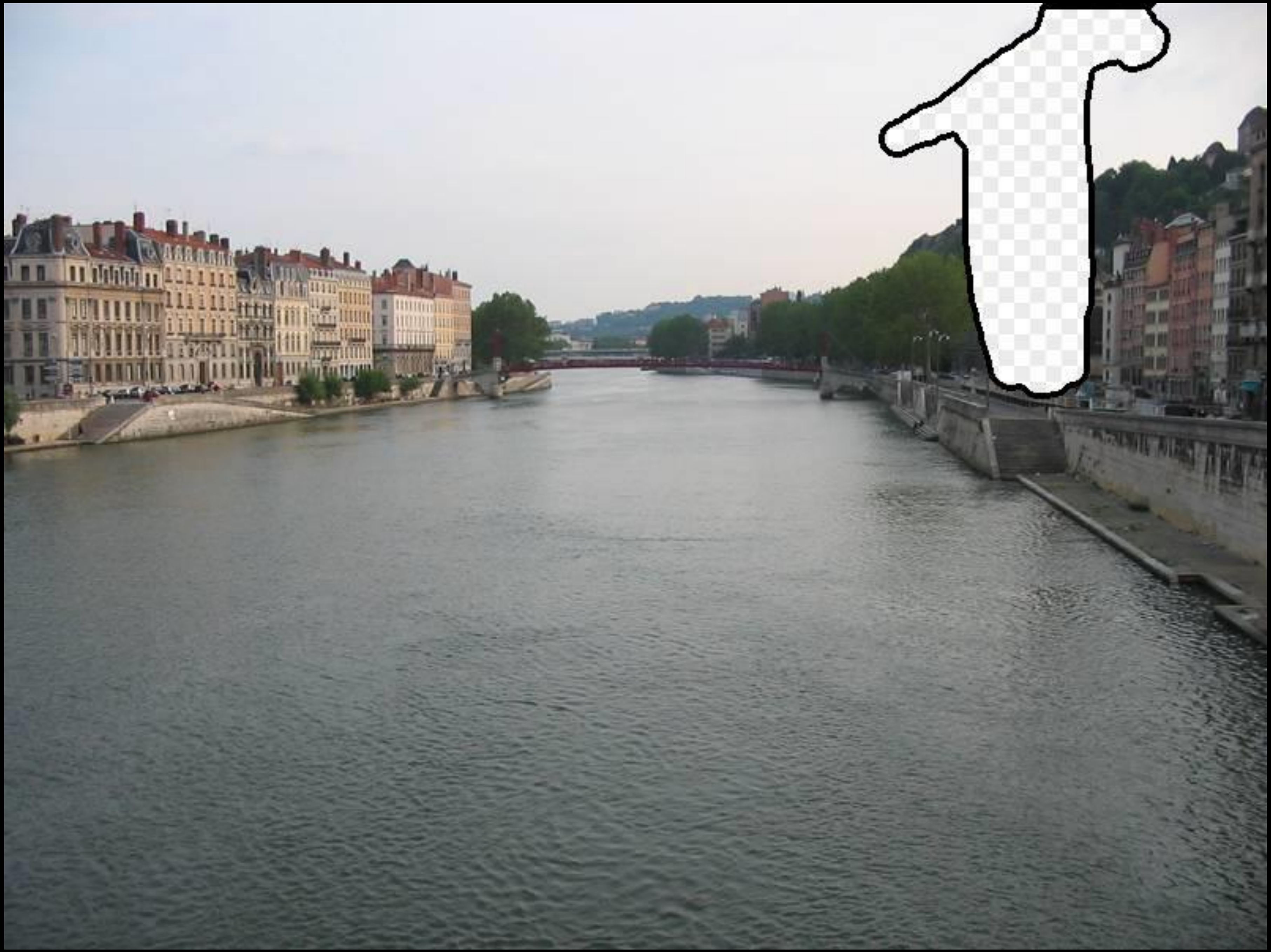
2 Million Flickr Images

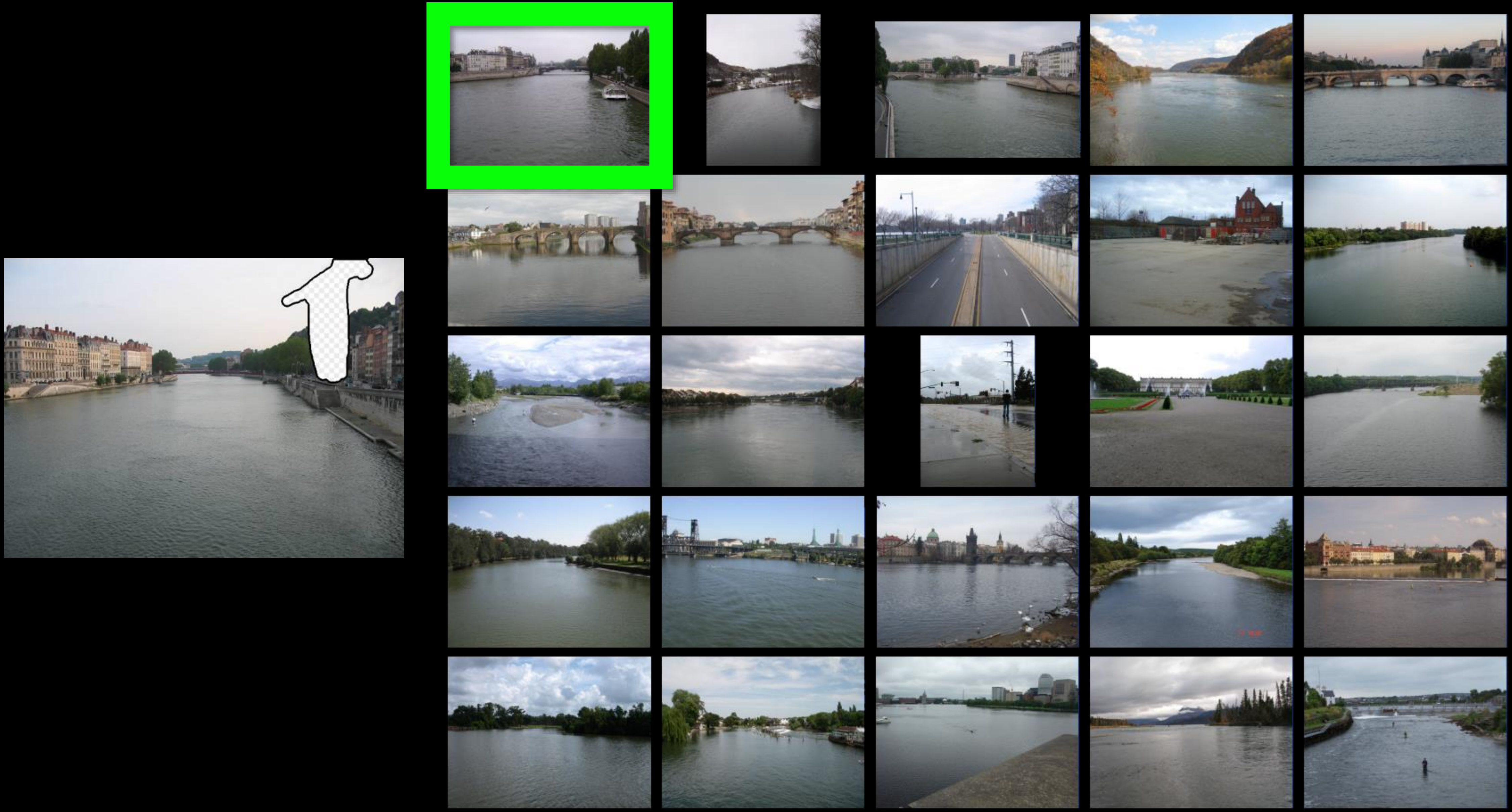




10 nearest neighbors from a collection of 20,000 images

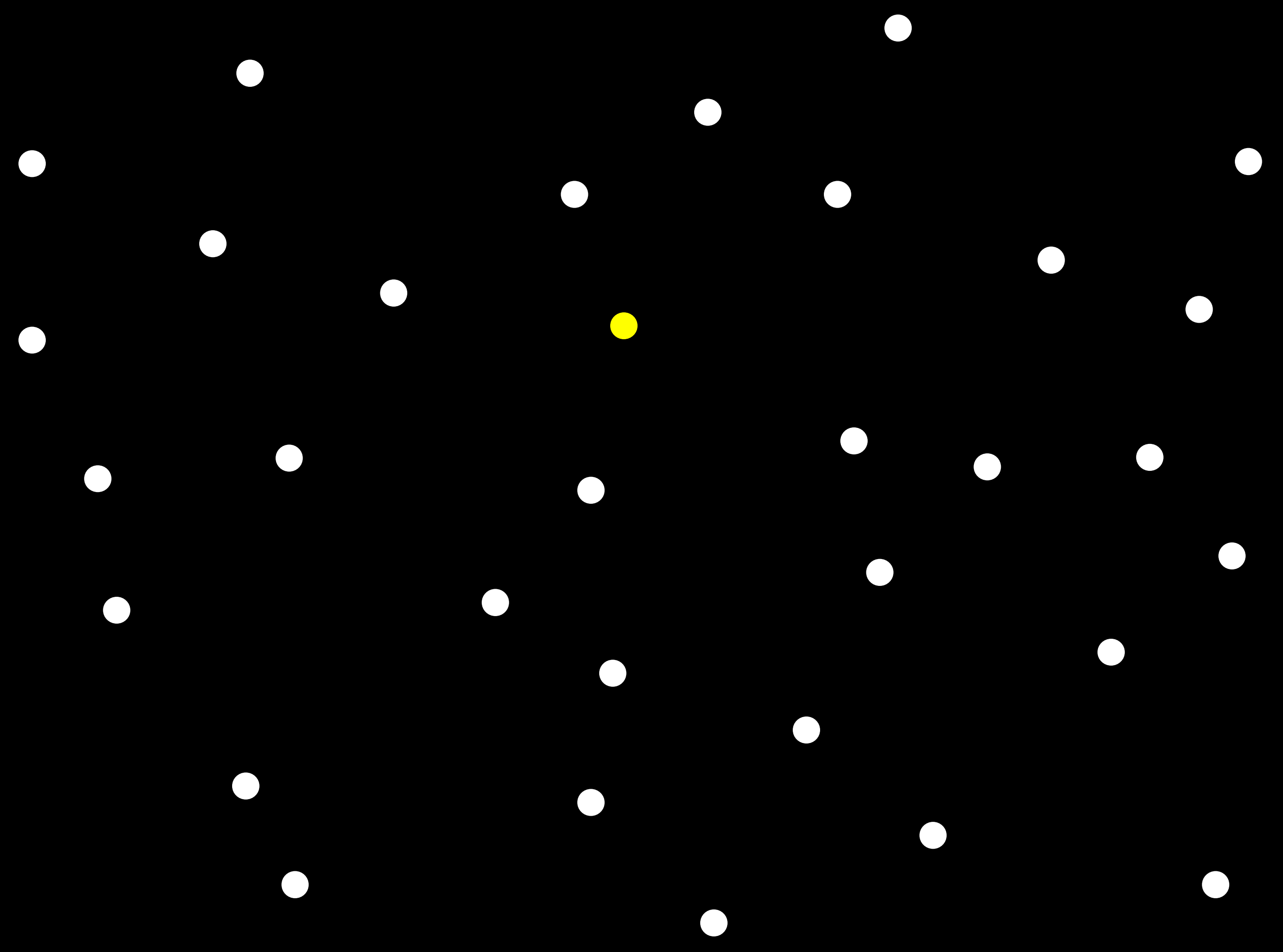
10 nearest neighbors from a
collection of 2 million images



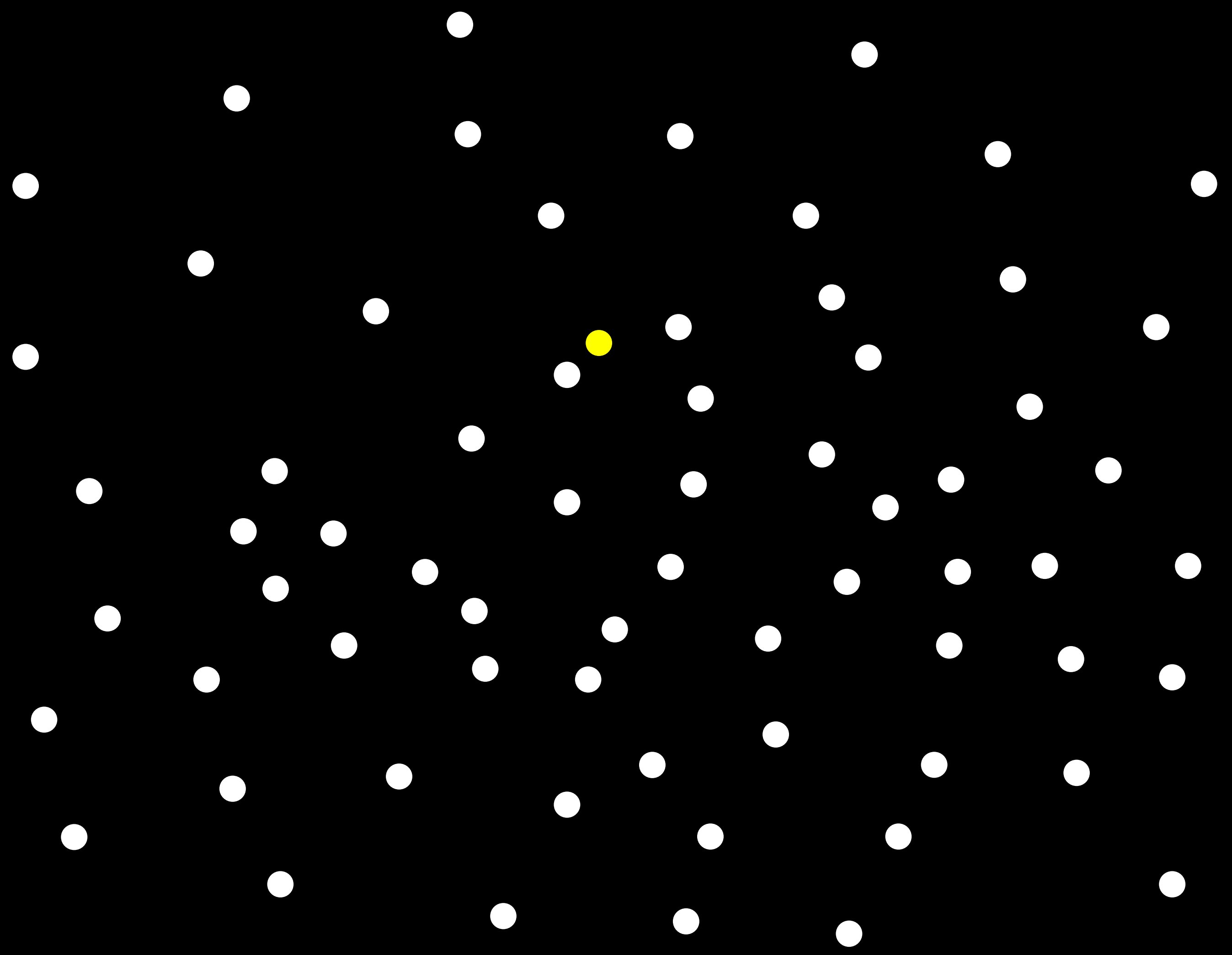


… 200 scene matches

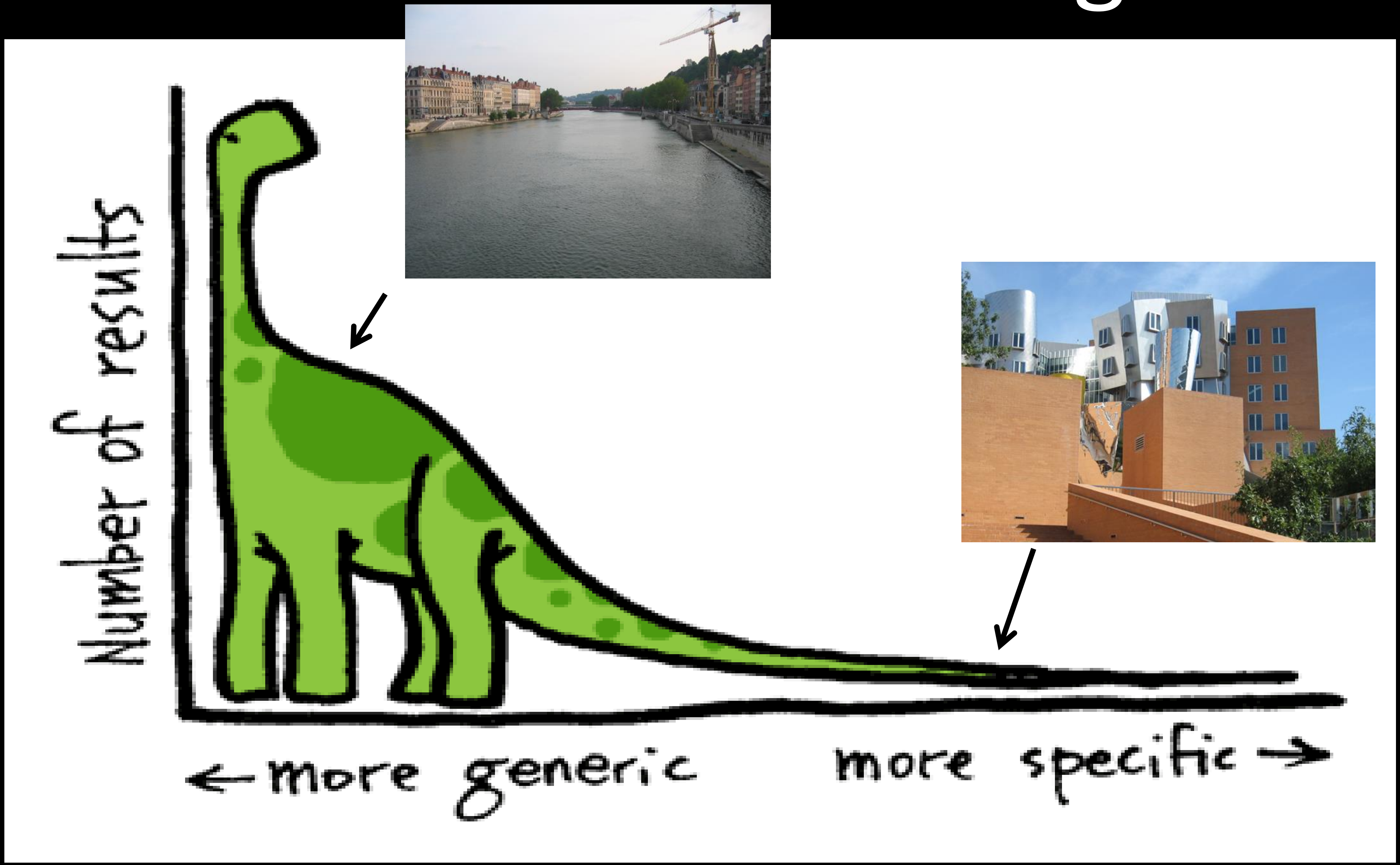

Improving Visual Correspondence



Improving Visual Correspondence



Visual Data has a Long Tail



The rare is common!

VISUAL DATA MINING

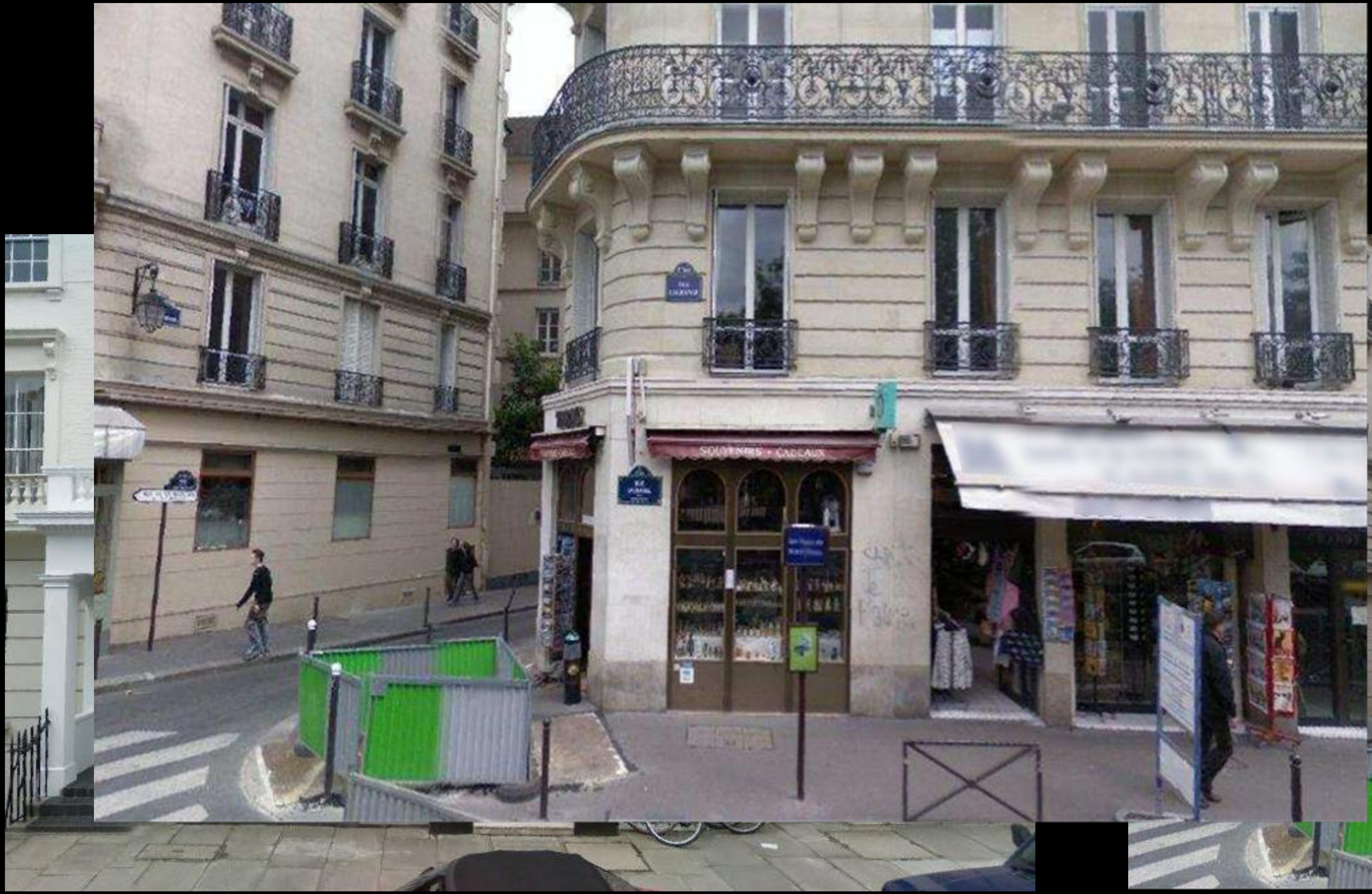
SIGGRAPH 2012

Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. *What Makes Paris Look like Paris?* SIGGRAPH 2012.

One of these is from Paris

...this is Paris

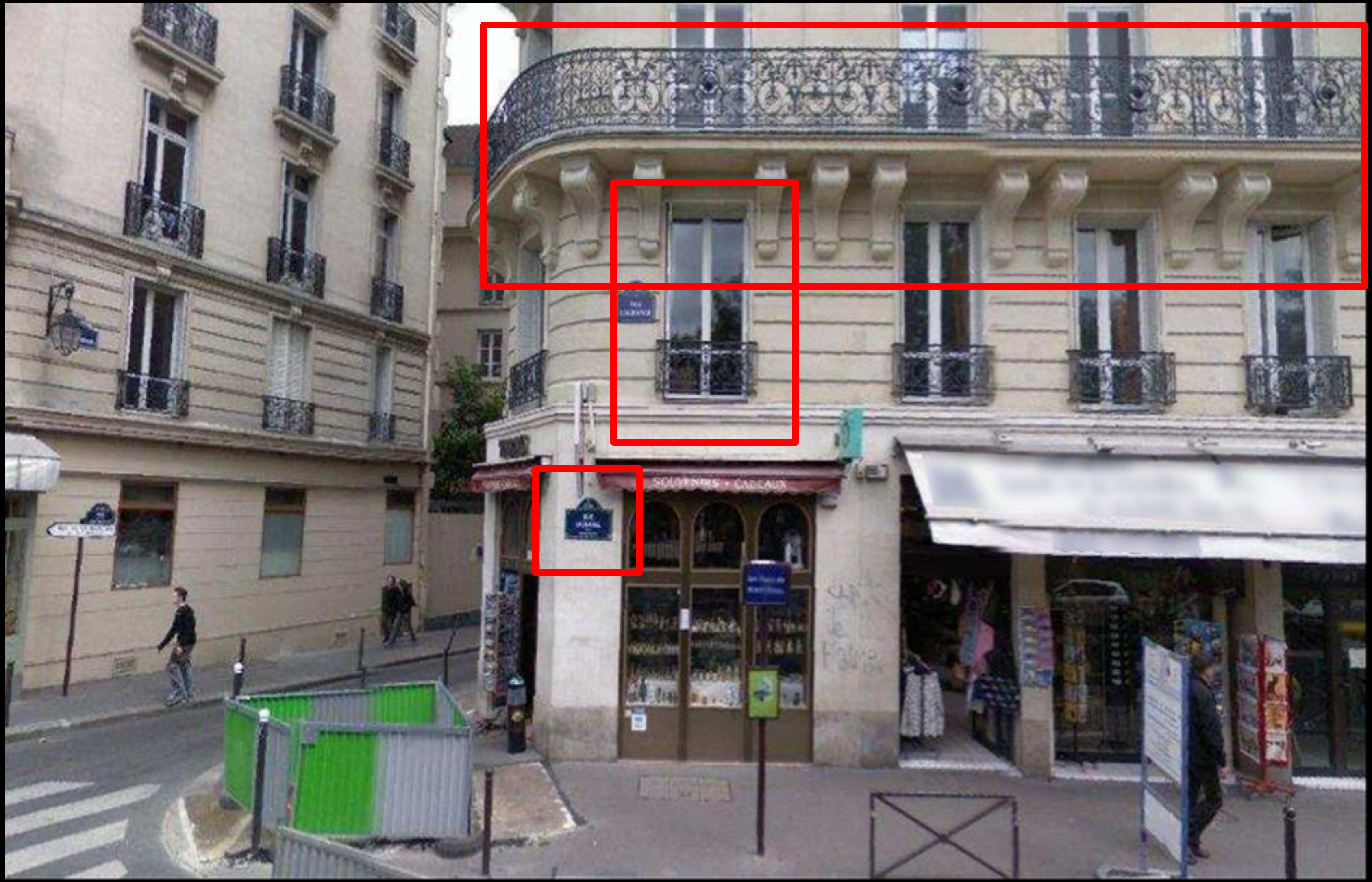
Clap if…



We showed 20 subjects:

- 100 Random Street View Images
- 50 from Paris
- They classified Paris non-Paris
- Accuracy: 79%

How do they know?



We showed 20 subjects:

- 100 Random Street View Images
- 50 from Paris
- They classified Paris non-Paris
- Accuracy: 79%

How do they know?

Our Goal:

Given a large geo-tagged image dataset,
*we automatically discover **visual elements***
that characterize a geographic location

Our Hypothesis

- The visual elements that capture Paris:
  - Frequent: Occur often in Paris
  - Discriminative: Are not found outside Paris

Note: same idea as TF-IDF if we knew the elements.

Need Both Conditions

- Discriminative only:

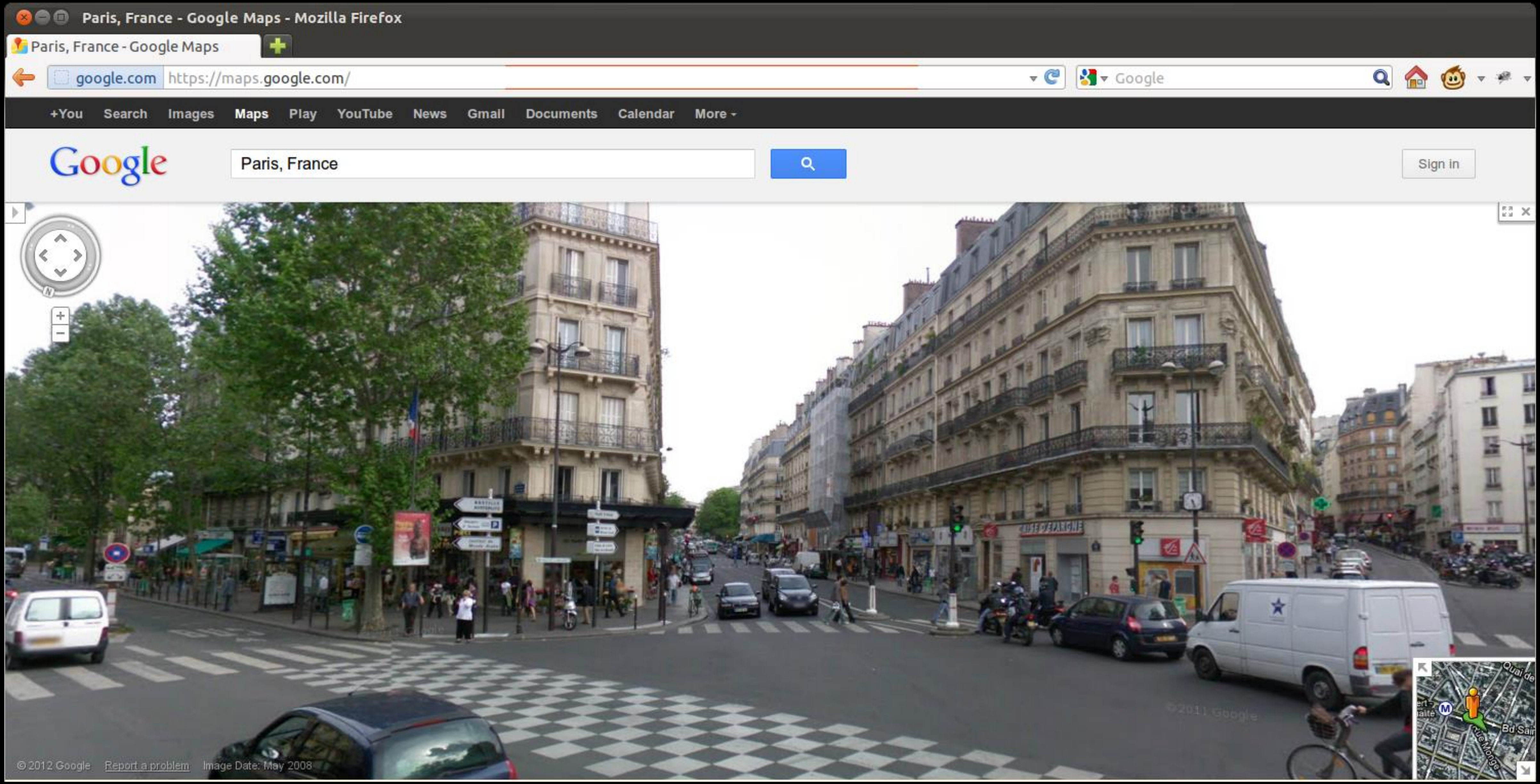
Need Both Conditions

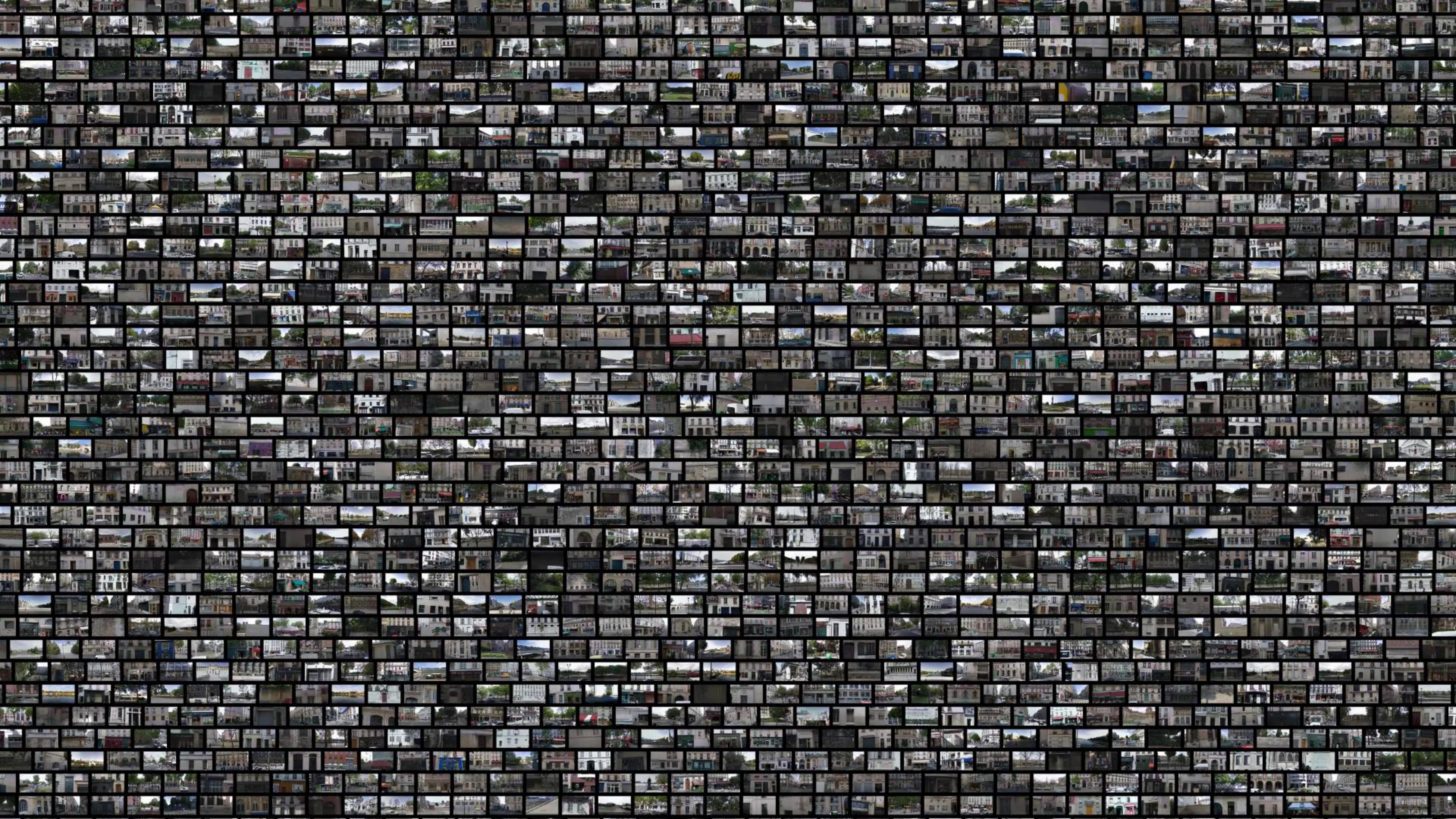
- Frequently occurring only:

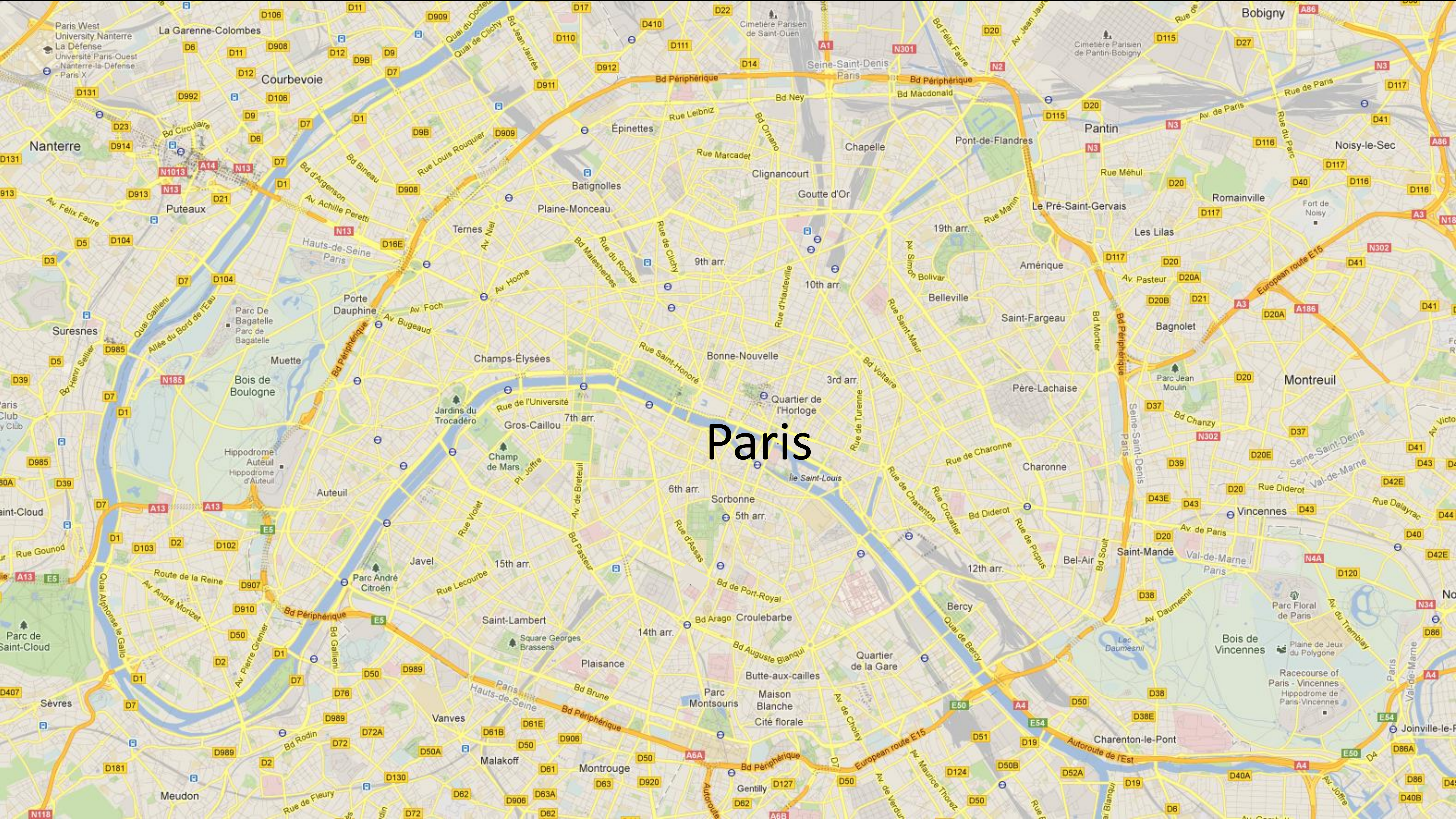


London
Prague
Paris
New York
Boston
Milan
Barcelona
Tokyo
San Francisco
Philadelphia
Mexico City
Sao Paulo
Positive Set
Negative Set

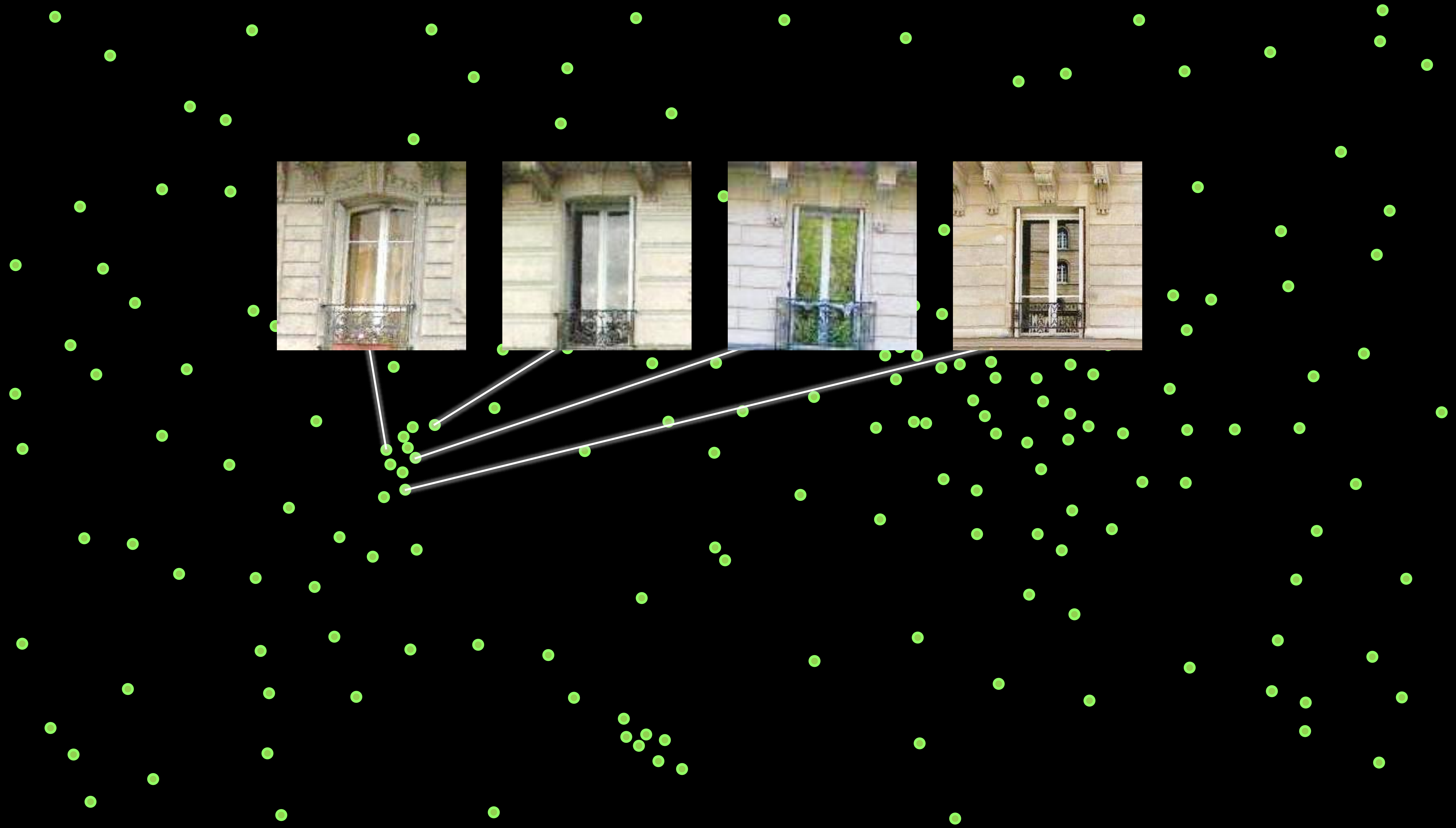
The Data: Google Street View



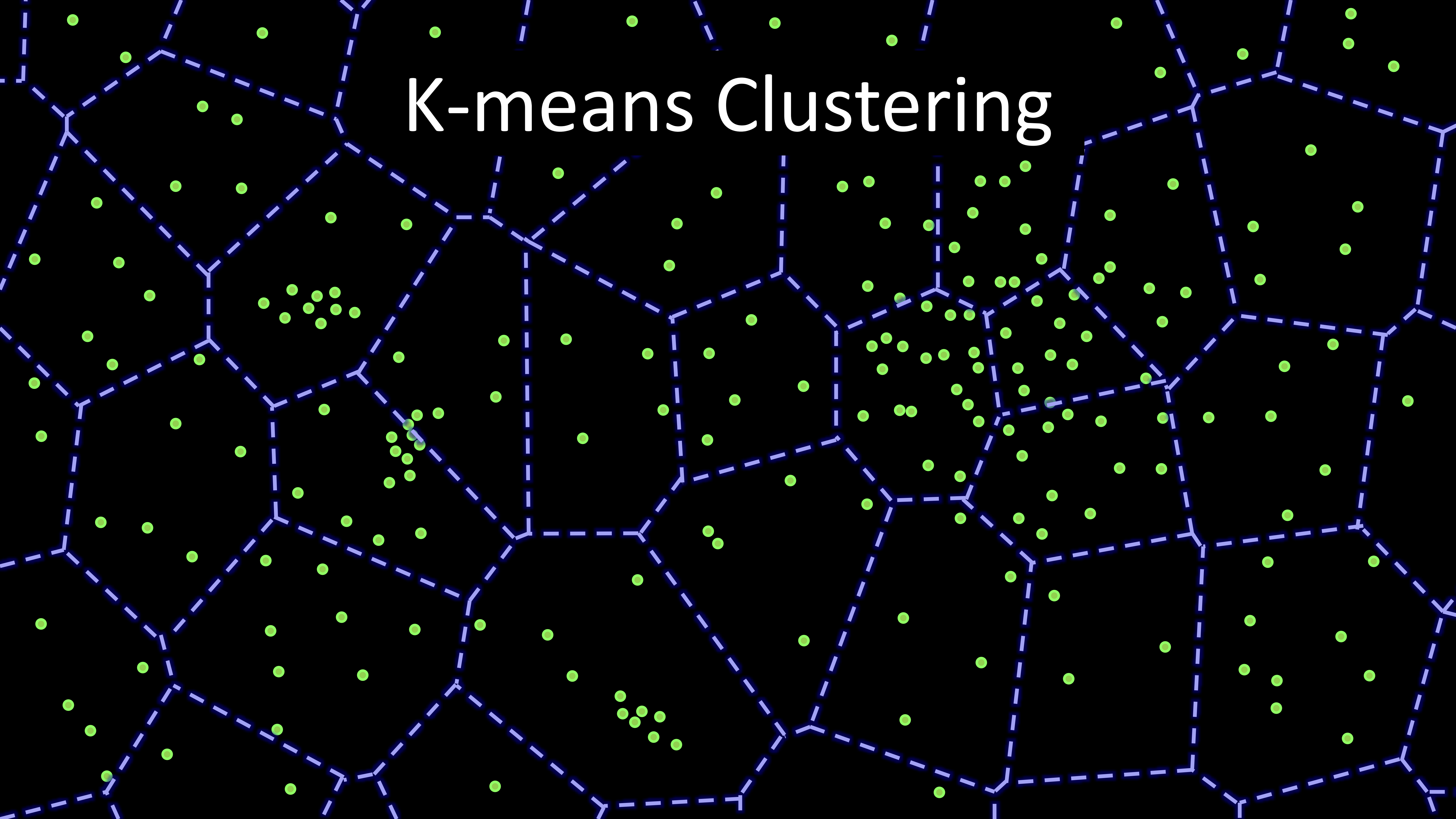


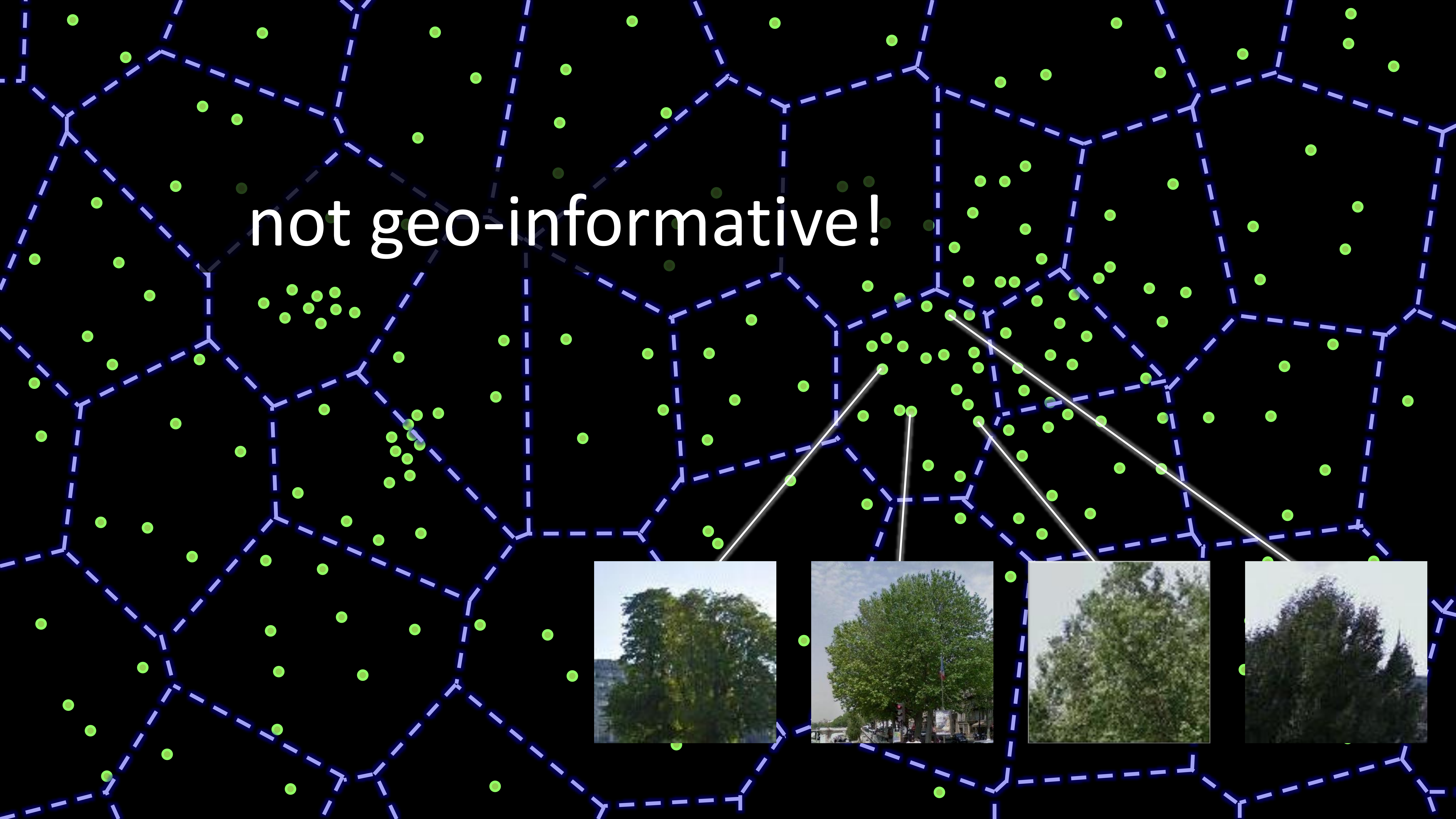

Paris
Nanterre
Courbevoie
La Garenne-Colombes
Puteaux
Suresnes
Bois de Boulogne
Porte Dauphine
Muette
Auteuil
Javel
15th arr.
Saint-Lambert
Plaisance
14th arr.
Vanves
Malakoff
Montrouge
Meudon
Sèvres
Ternes
Batignolles
Plaine-Monceau
Épinettes
Clignancourt
Goutte d'Or
Chapelle
9th arr.
10th arr.
Champs-Élysées
Gros-Caillou
7th arr.
6th arr.
Sorbonne
5th arr.
3rd arr.
Bonne-Nouvelle
Quartier de l'Horloge
Île Saint-Louis
Croulebarbe
Quartier de la Gare
Butte-aux-cailles
Maison Blanche
Cité florale
Parc Montsouris
Gentilly
Bercy
12th arr.
Bel-Air
Saint-Mandé
Vincennes
Bois de Vincennes
Charenton-le-Pont
Charonne
Père-Lachaise
Saint-Fargeau
Belleville
Amérique
19th arr.
Pont-de-Flandres
Pantin
Le Pré-Saint-Gervais
Les Lilas
Romainville
Noisy-le-Sec
Bagnolet
Montreuil
Bobigny
Bd Périphérique


4me Arr
RUE DU PETIT-MUSC
6e Arr
RUE HAUTEFEUILLE
6e Arr
RUE GARANCIÈRE
4e Arr
RUE BUDÉ
Nanterre
Courbevoie
La Garenne-Colombes
Puteaux
Suresnes
Bois de Boulogne
Porte Dauphine
Sèvres
Meudon
Malakoff
Montrouge
Gentilly
Bobigny
Noisy-le-Sec
Romainville
Les Lilas
Bagnolet
Montreuil
Vincennes
Bois de Vincennes
Charenton-le-Pont
Bd Périphérique



K-means Clustering



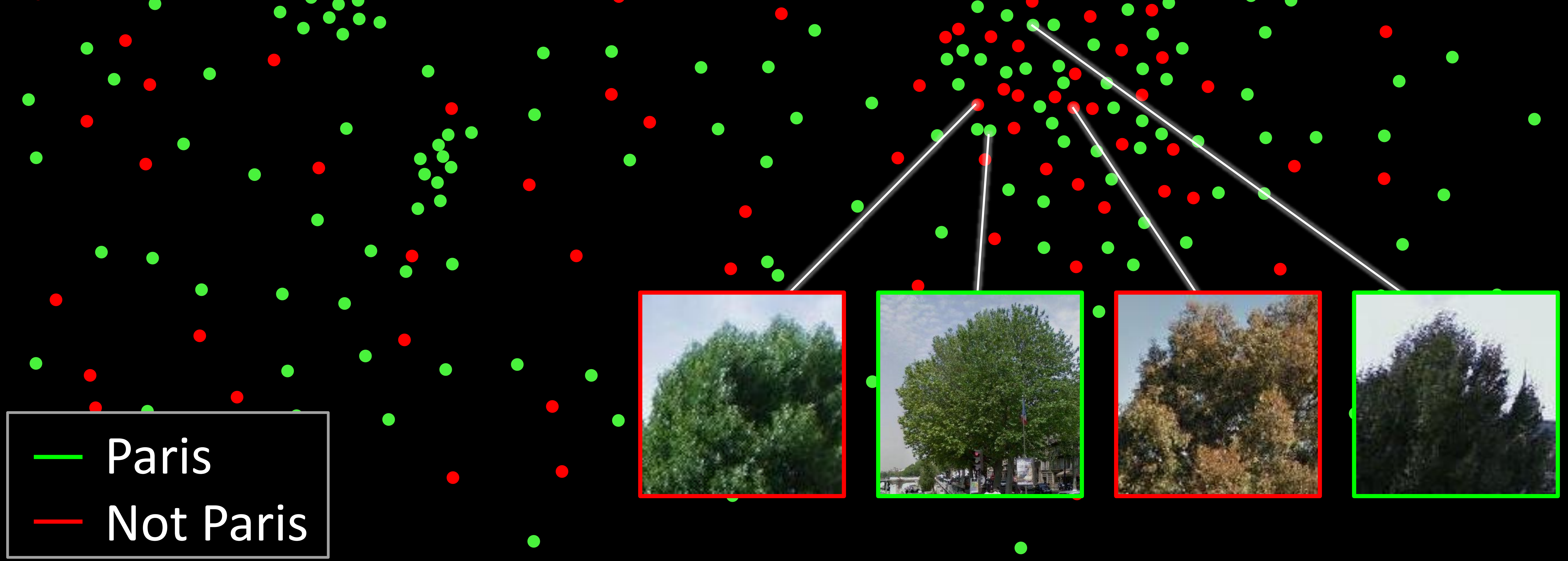

not geo-informative!


visually incoherent!

Our Approach

Our Approach

I. **Use geo-supervision**



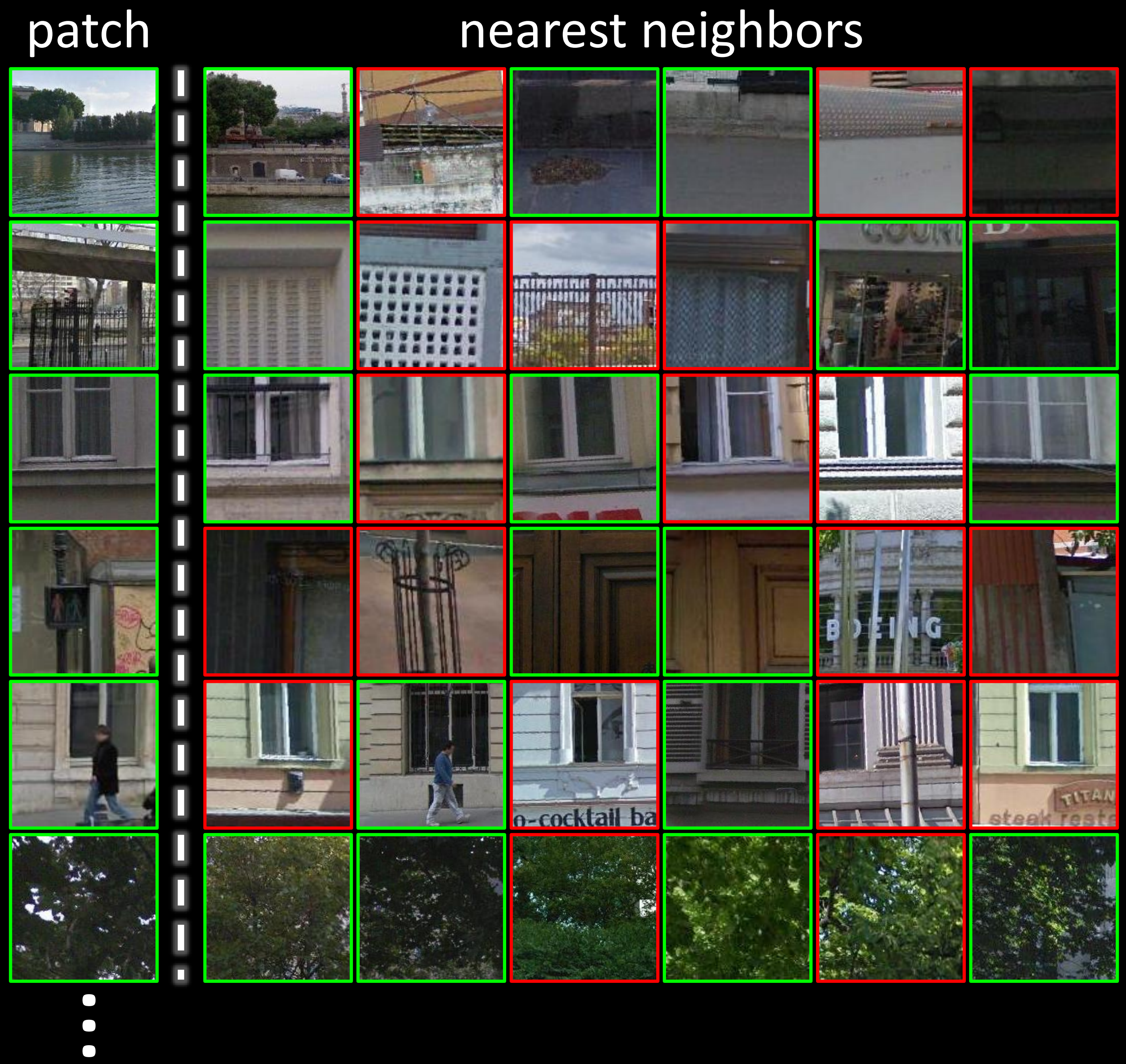
Our Approach

I. Use geo-supervision

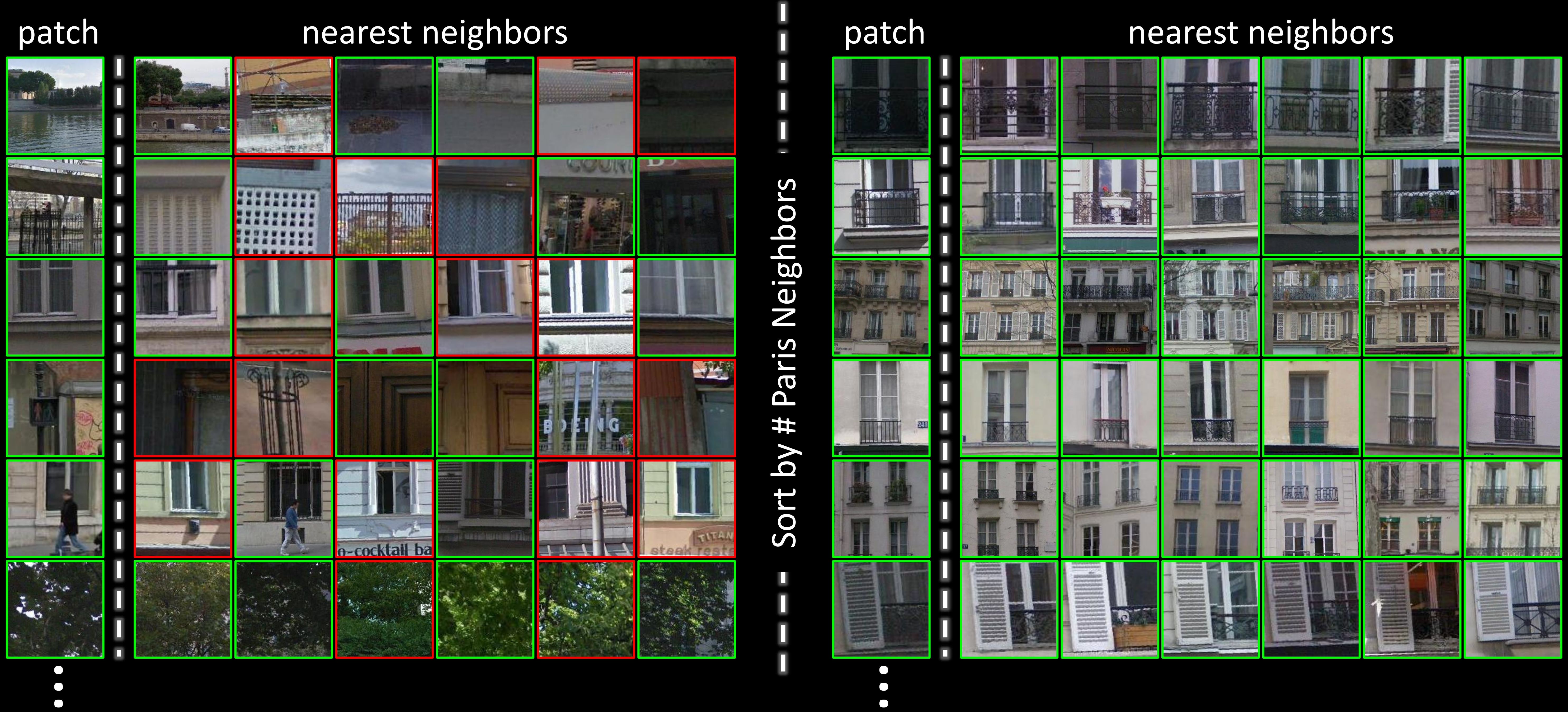
II. Don't partition the space top-down; build clusters bottom-up

Step 1: Nearest Neighbors for Every Patch

Using normalized correlation of HOG features as a distance metric



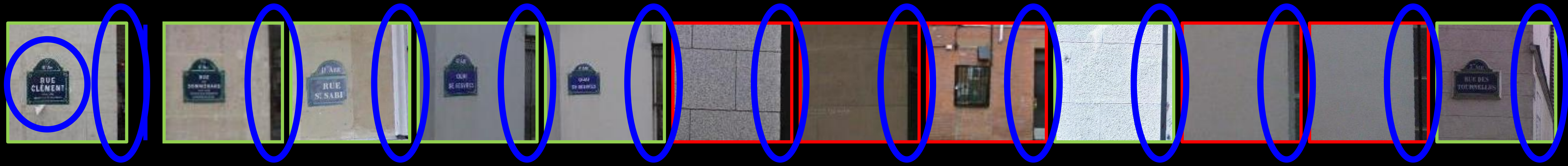
Step 2: Find the Parisian Clusters by Sorting



Rank: 1146

Good Patches may have Bad Neighbors!

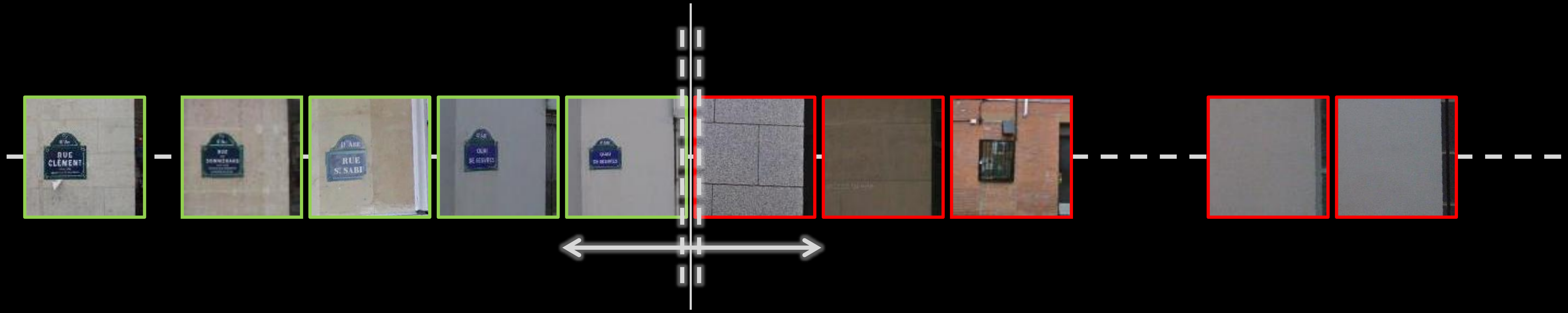
patch matches



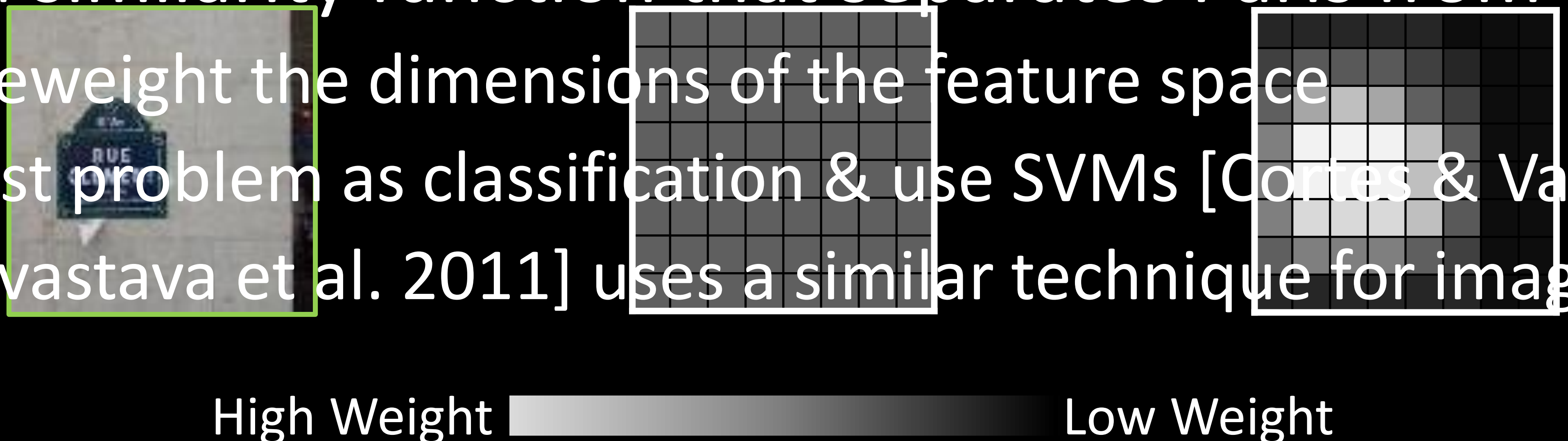
- The naïve distance metric gives equal weight to the vertical bar and the sign.

— Paris
— Not Paris

Step 3: Updating the Similarity Function

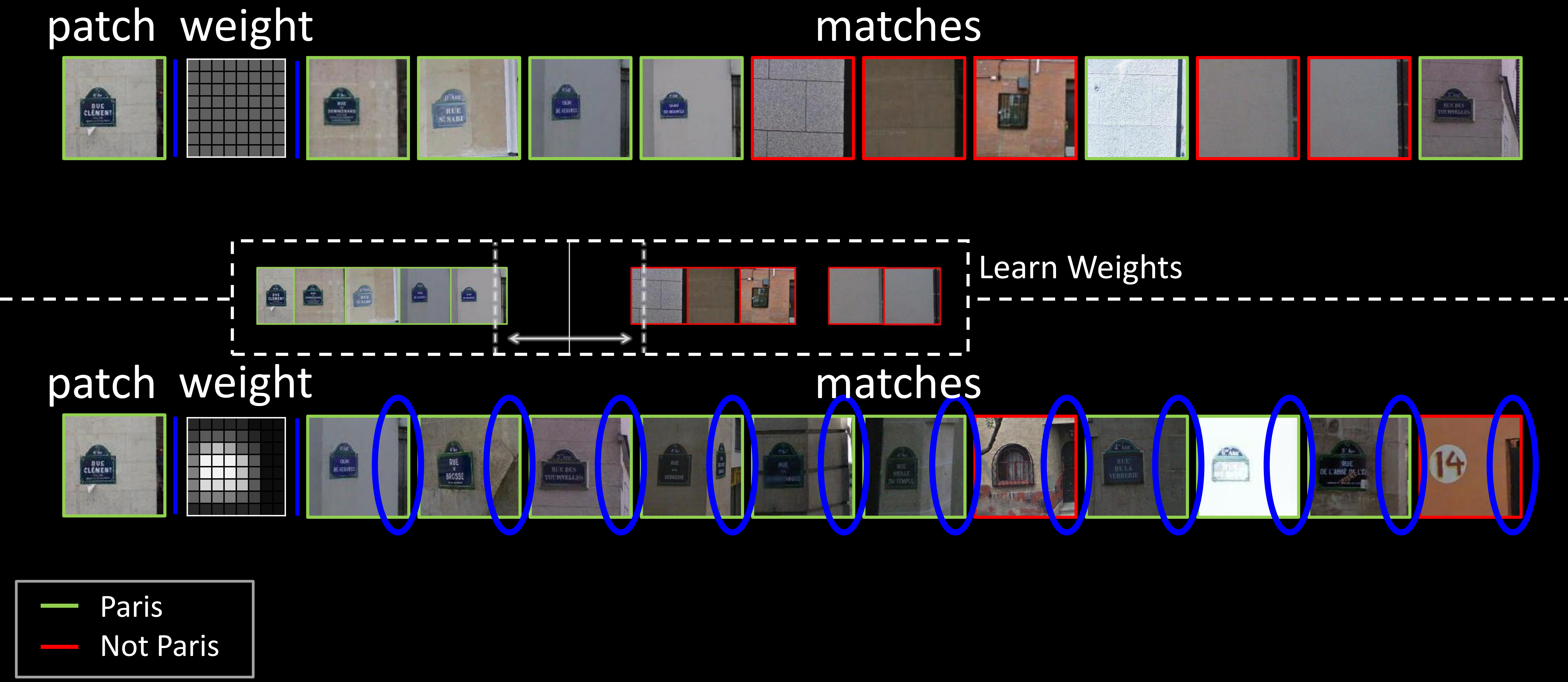


- Learn a similarity function that separates Paris from not-Paris
  - I.e. reweight the dimensions of the feature space
  - Recast problem as classification & use SVMs [Cortes & Vapnik 1995]
  - [Shrivastava et al. 2011] uses a similar technique for image retrieval

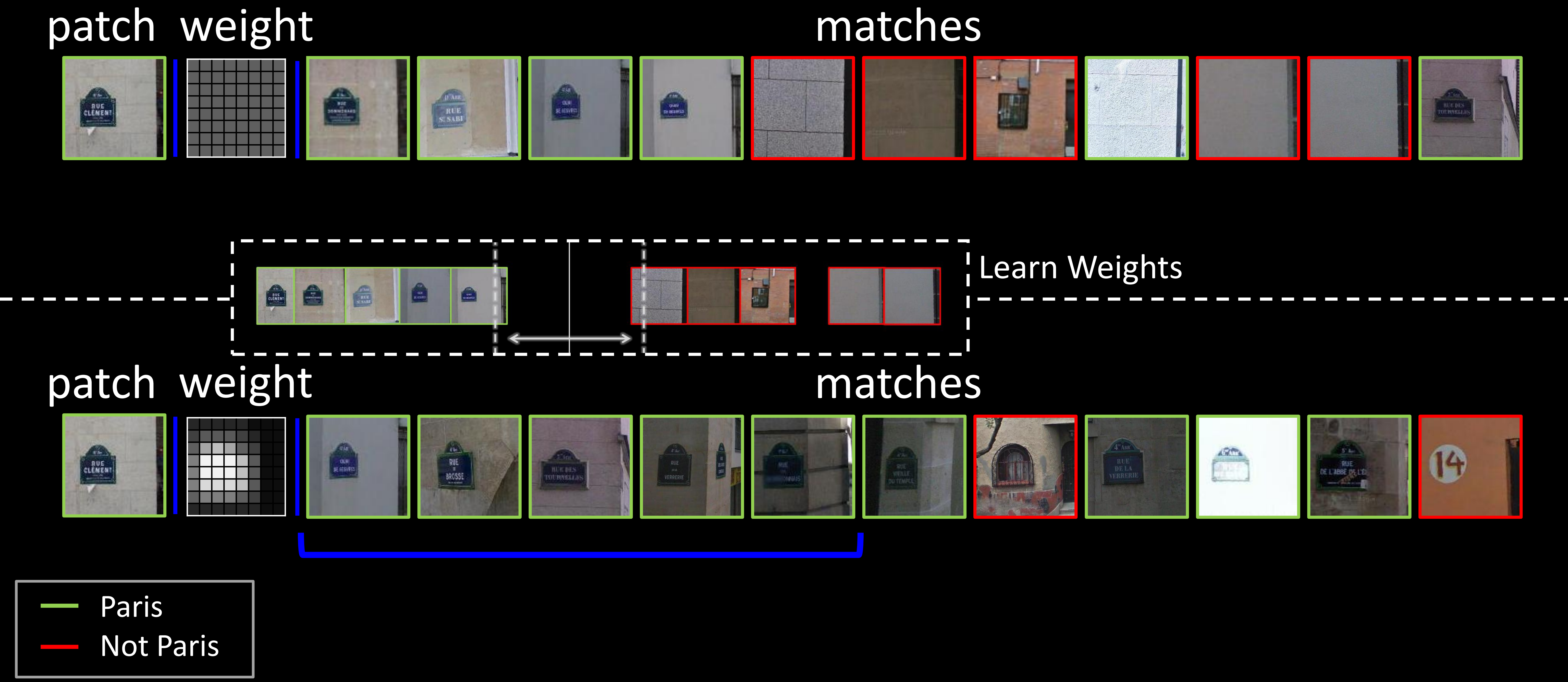


Paris
Not Paris

Resulting Matches



Resulting Matches

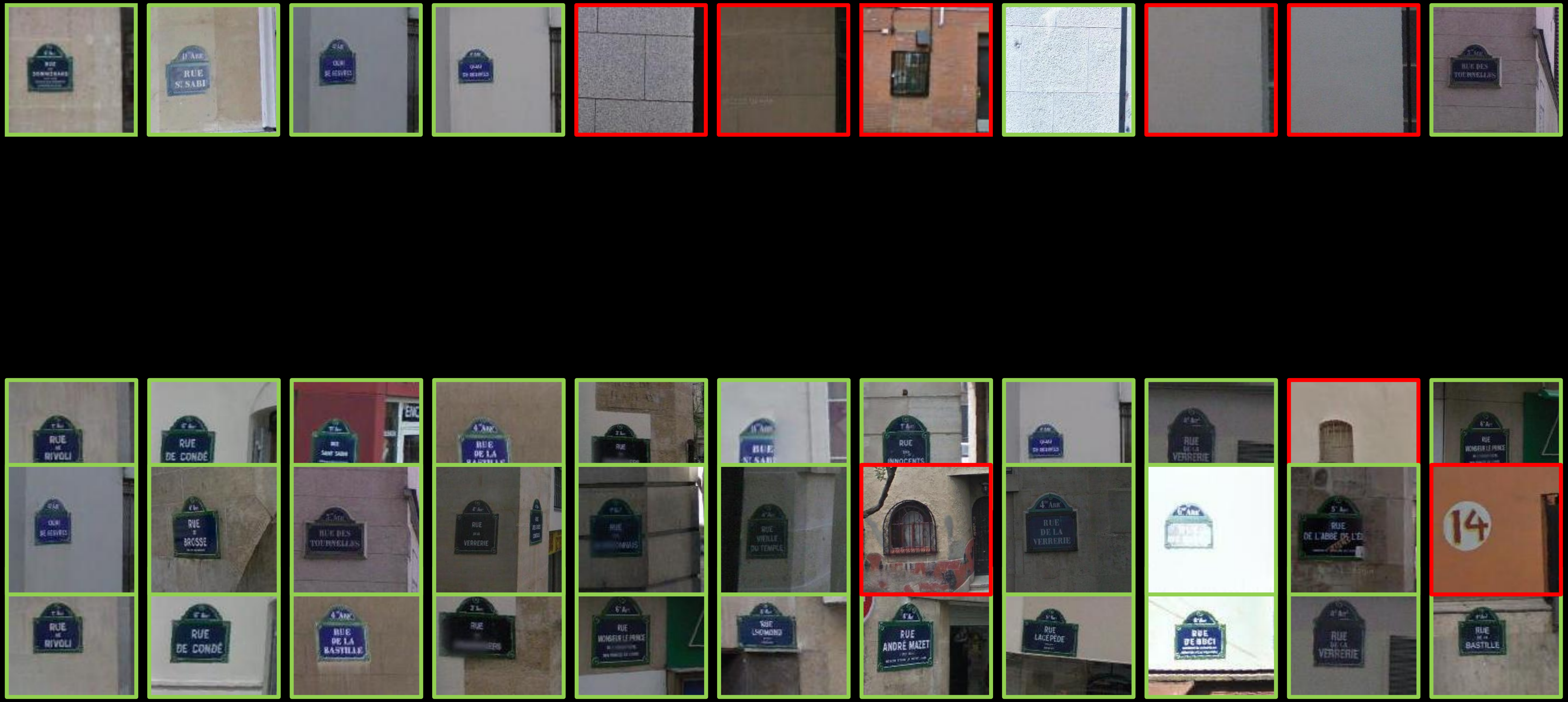


Step 4: Iterate using the new matches

patch

matches

Orig.

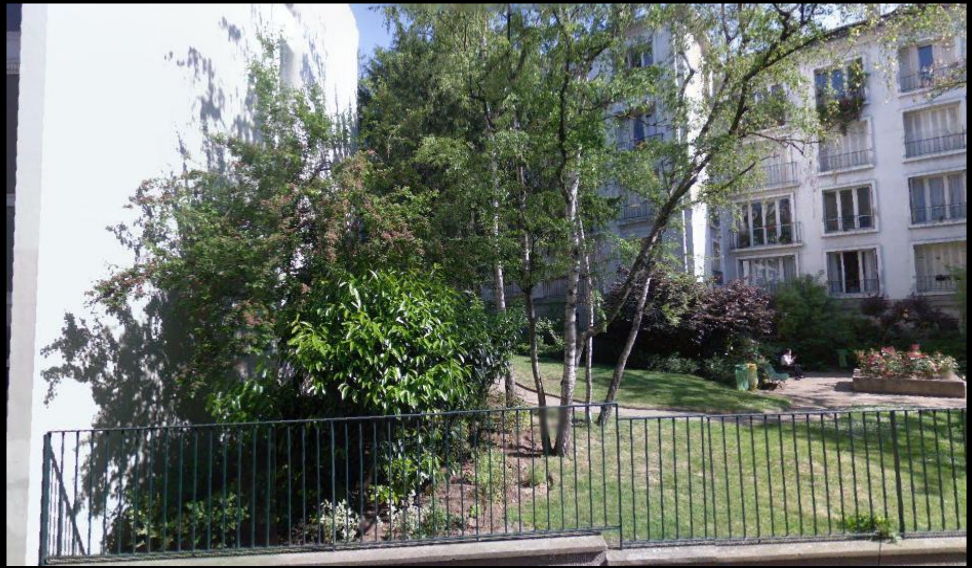


Iteration 1

Iteration 2

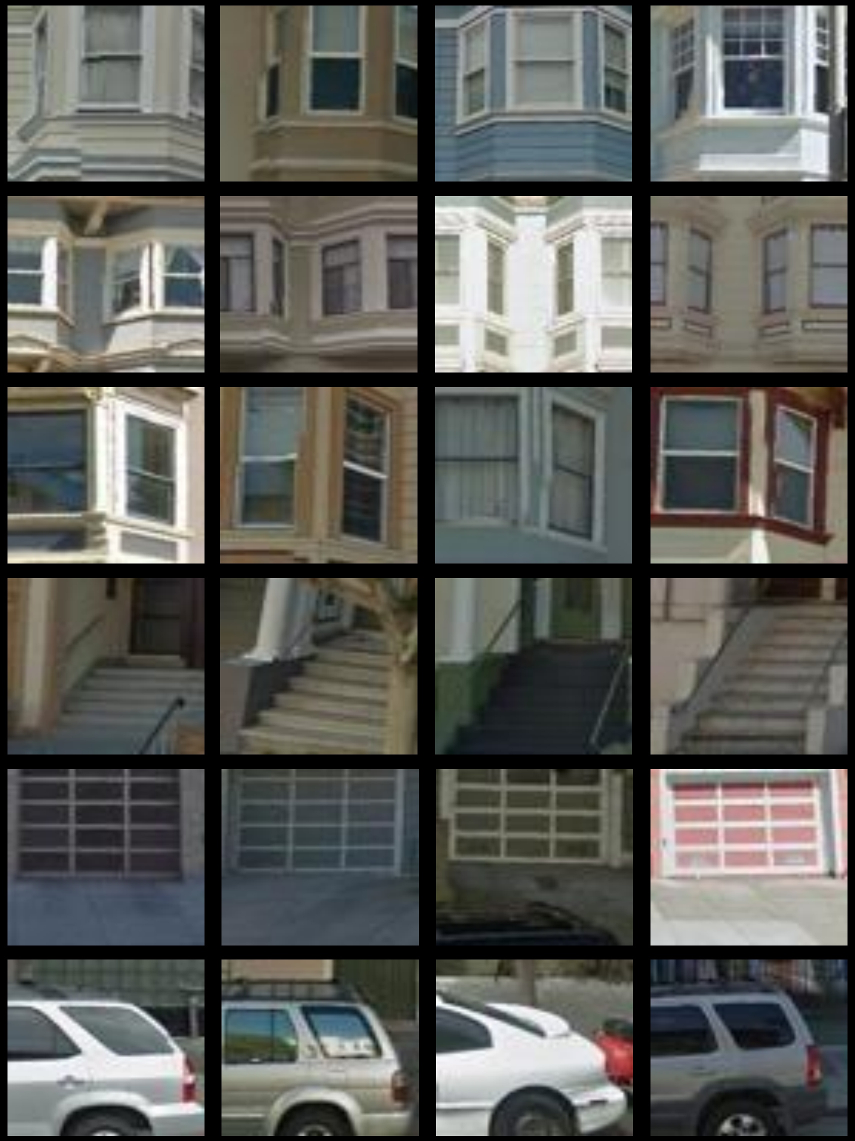
Iteration 3

Random Paris

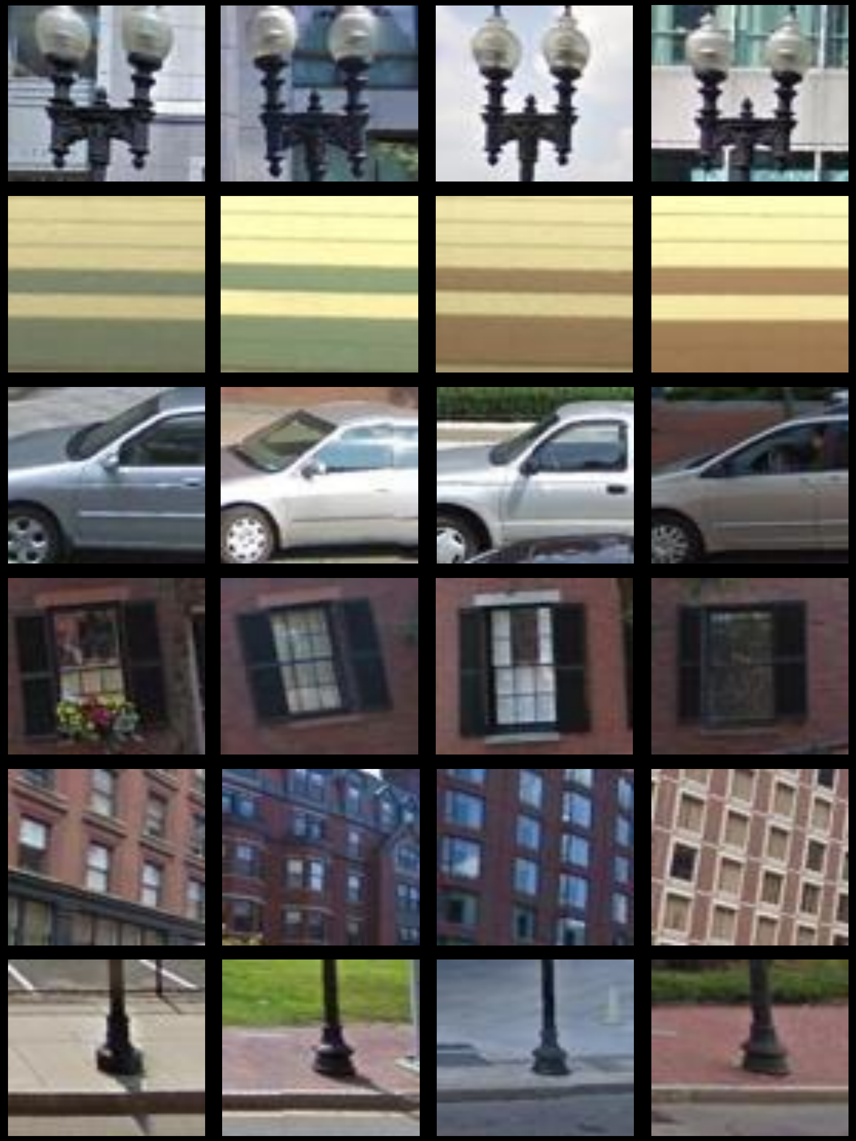


Paris: A Few Top Elements

In the U.S.



Elements from San Francisco



Elements from Boston

Elements from Prague

Elements from London

Elements from Barcelona


Google earth
Image © 2012 IGN France
Imagery Date: 12/31/2007
48°51'06.93" N 2°20'37.42" E elev 35 m
Eye alt 3.38 km

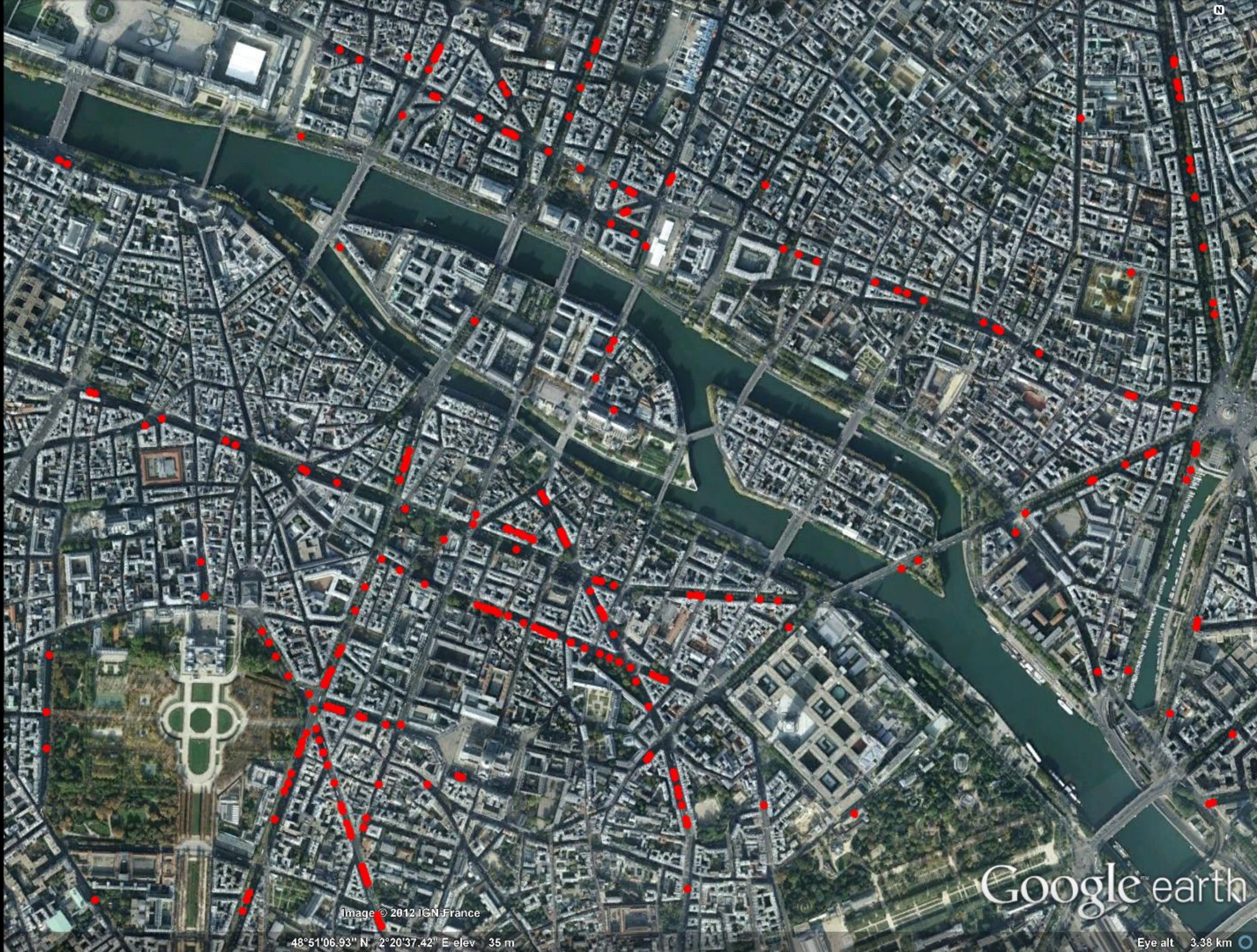
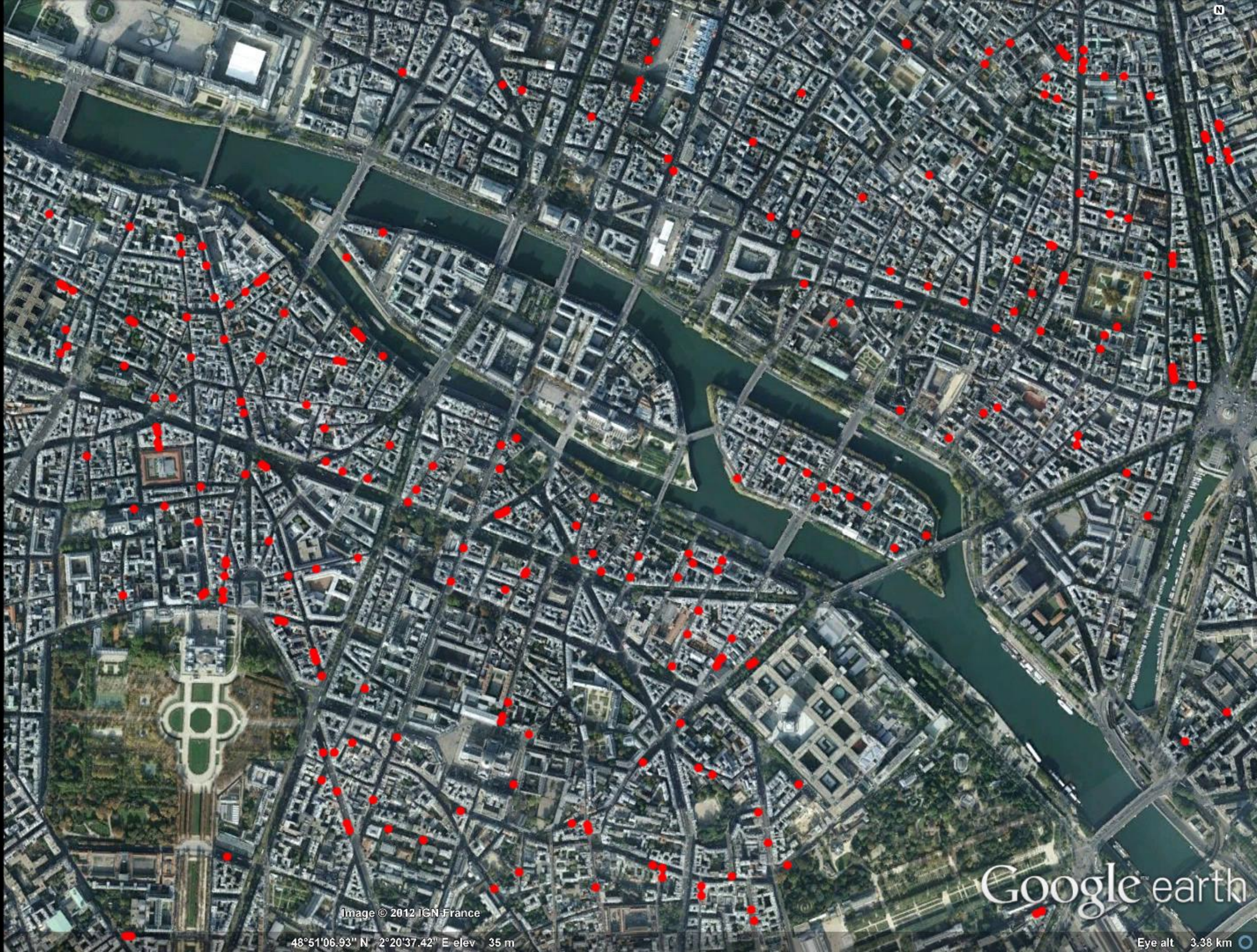


Image © 2012 IGN-France
48°51'06.93" N 2°20'37.42" E elev 35 m
Google earth
Eye alt 3.38 km





N
Image © 2012 IGN-France
48°51'06.93" N 2°20'37.42" E elev 35 m
Google earth
Eye alt 3.38 km


Louvre
/Opera
Marais
Latin Quarter
Google earth
Eye alt 3.38 km


London
Prague
RESTAURACE
U JINDŘIŠSKÉ VĚŽE
www.ujindrisskeveze.cz
Paris
Milan
Barcelona
Image © 2012 TerraMetrics
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image © 2012 GeoContent
© 2012 Cnes/Spot Image
Google earth
Imagery Date: 4/14/2004
45°33'14.44" N 5°16'25.01" E elev 568 m
Eye alt 2995.11 km


London
Prague
Paris
Milan
Barcelona
Image © 2012 TerraMetrics
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image © 2012 GeoContent
© 2012 Cnes/Spot Image
Google earth
Imagery Date: 4/14/2004
45°33'14.44" N 5°16'25.01" E elev 568 m
Eye alt 2995.11 km


London
Prague
Paris
Milan
Barcelona
Image © 2012 TerraMetrics
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image © 2012 GeoContent
© 2012 Cnes/Spot Image
Google earth
Imagery Date: 4/14/2004
45°33'14.44" N 5°16'25.01" E elev 568 m
Eye alt 2995.11 km

Paris, France

Paris, France

What if we **force** the algorithm to match elements in Prague and London?

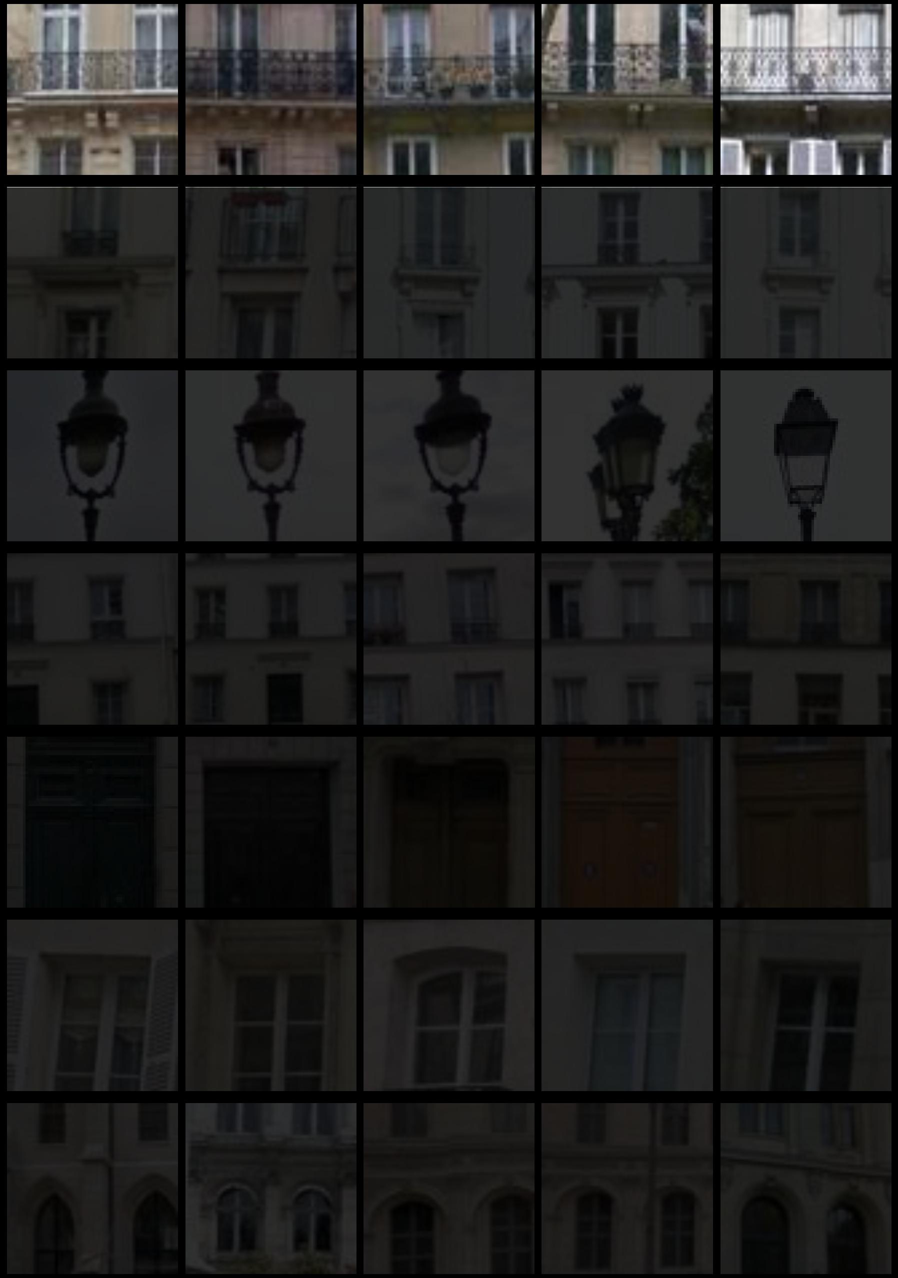
Prague, Czech Republic

London, England

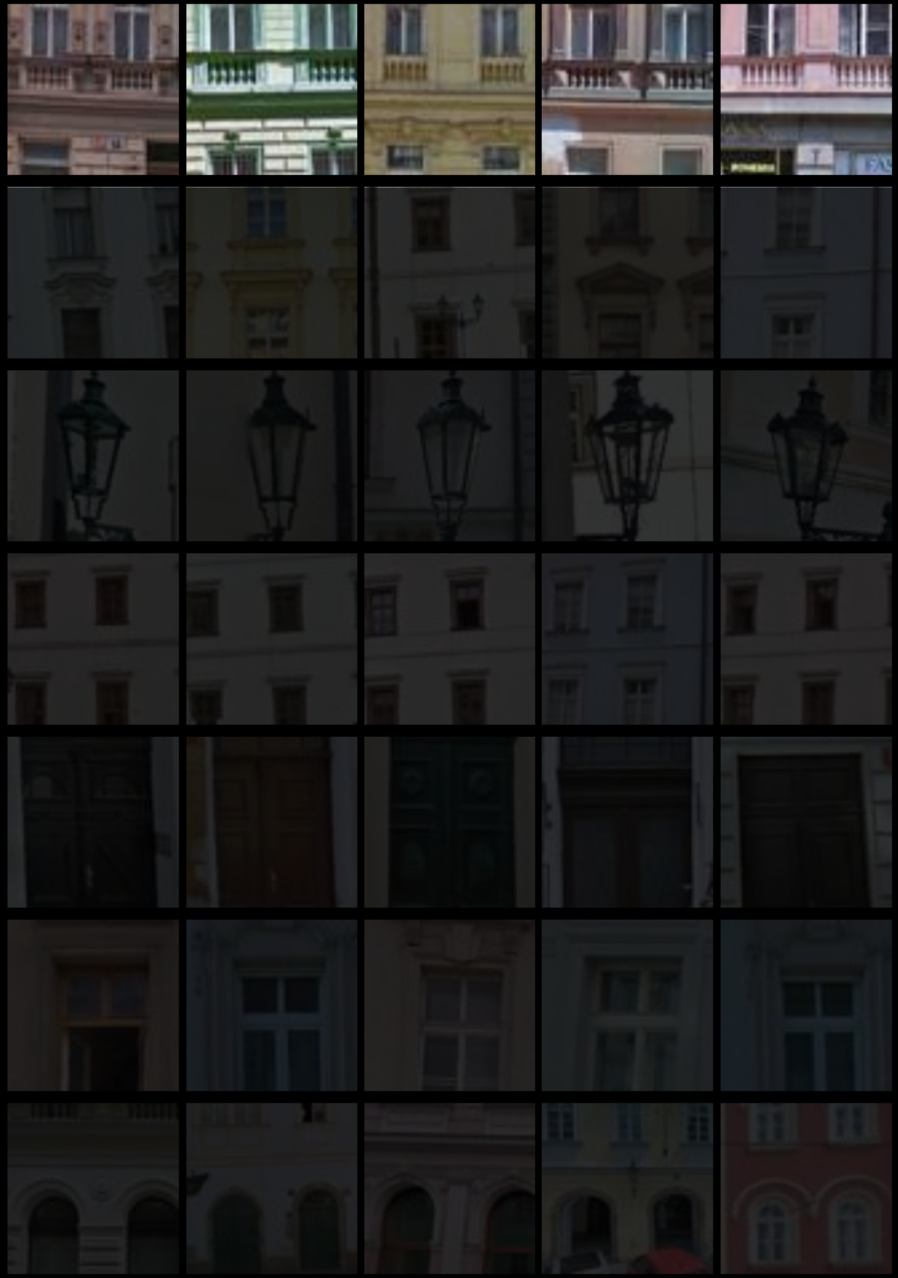
Paris, France

Prague, Czech Republic

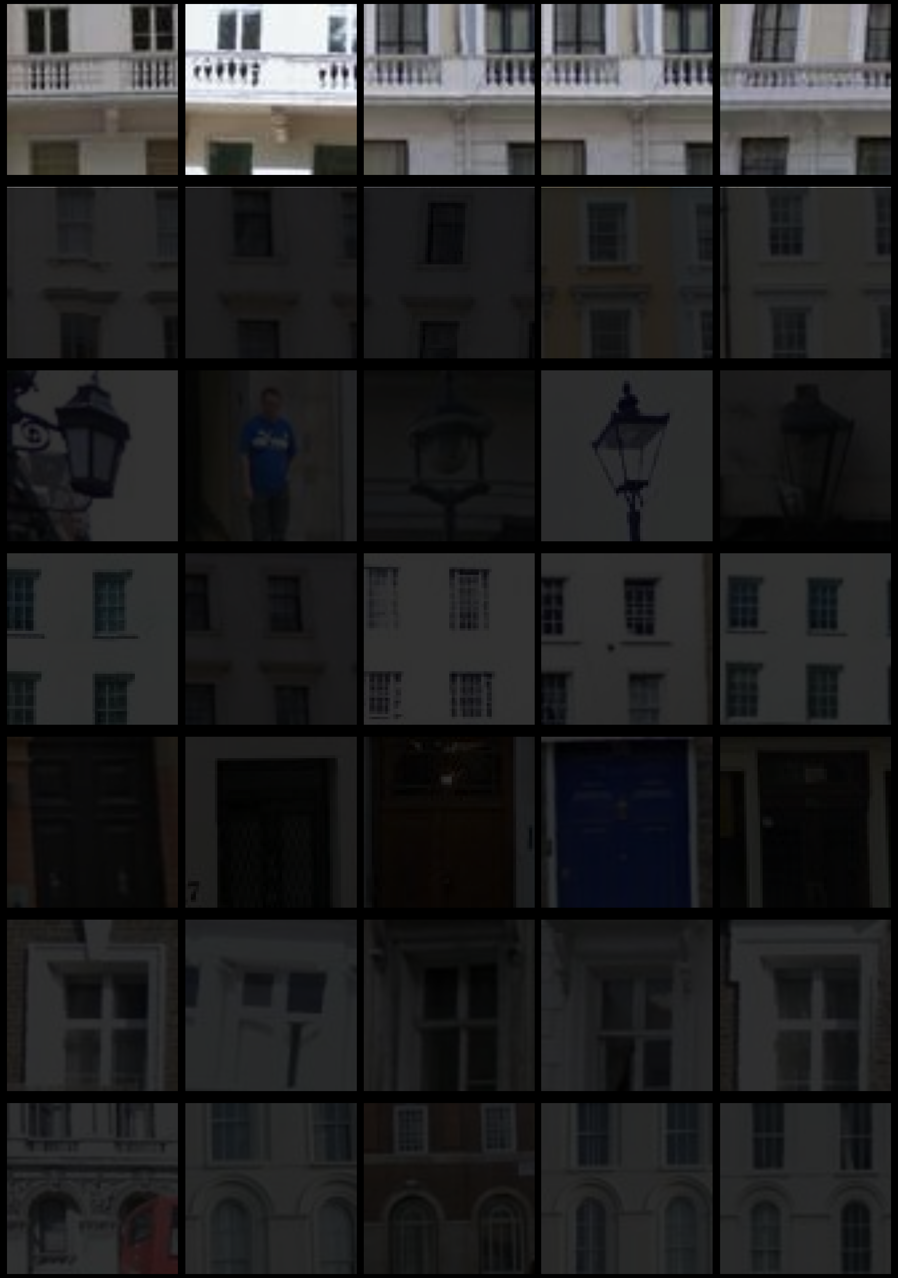
London, England



Paris, France



Prague, Czech Republic



London, England

So, what makes Paris look like Paris?

- The proposed algorithm finds visual elements that appear frequently in Paris, and not elsewhere.

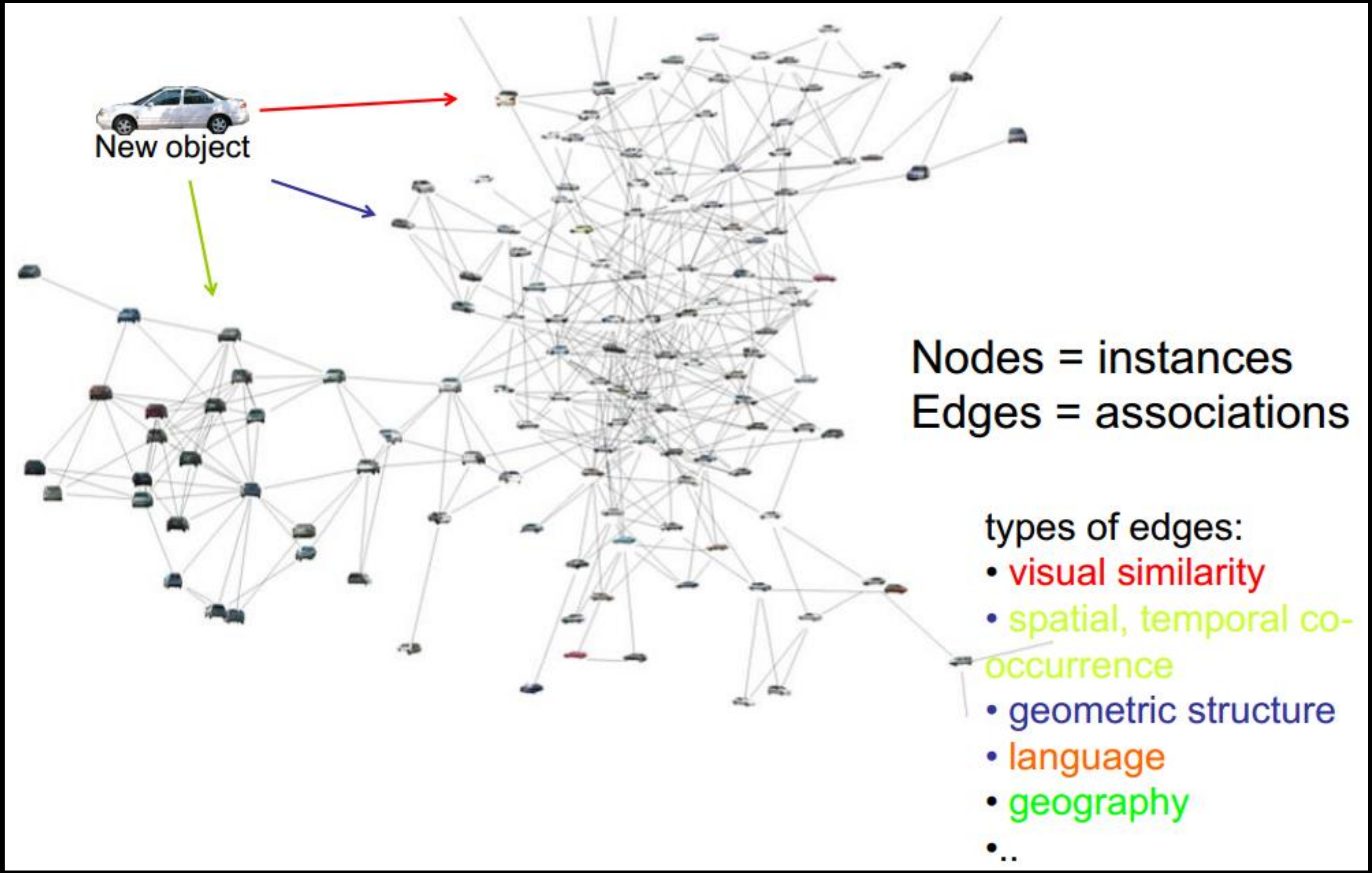



- What makes X look like X?
  - What makes a bathroom?
  - What makes a ‘50’s car?
  - What makes an Apple product?

Organizing the "Garbage Heap"

- Finding visual correspondences across data
- Mining Visual Data
- Connecting visual data to enable understanding (Visual Memex)

How to connect visual data to enable understanding (Visual Memex)



[Malisiewicz and Efros 09’]

How to build a Visual Memex with rich and dense relationships?

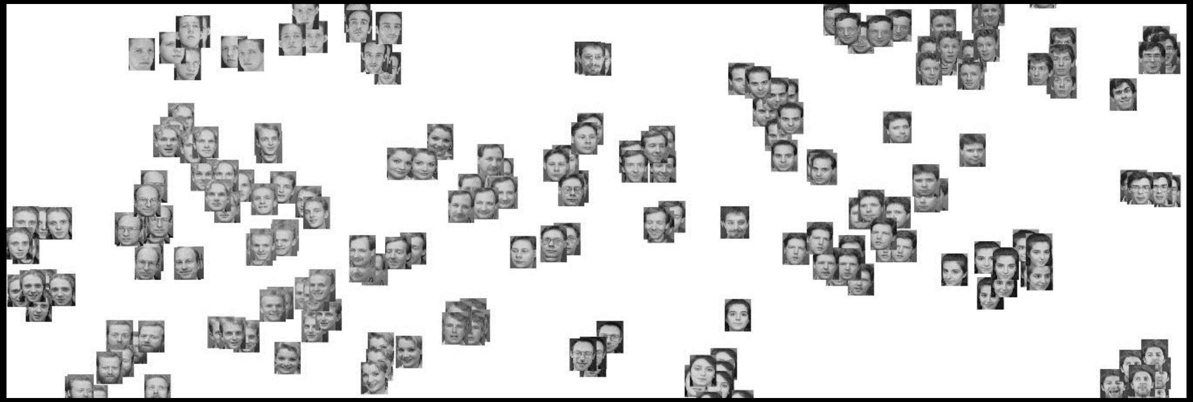
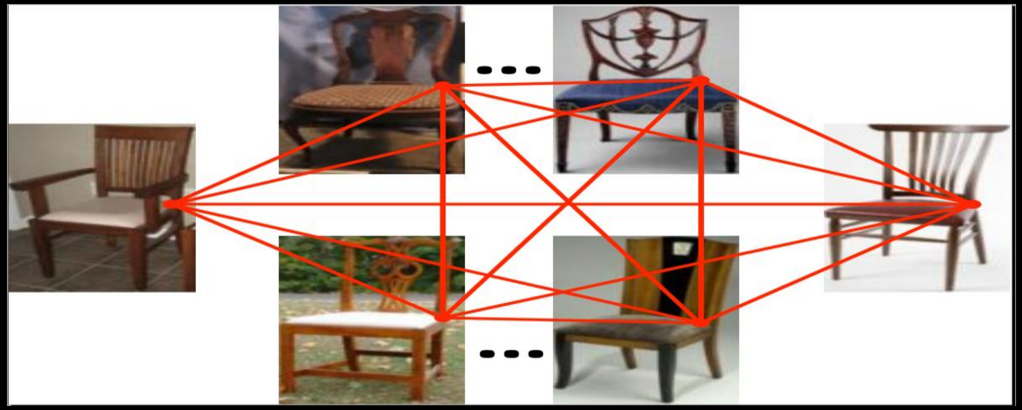
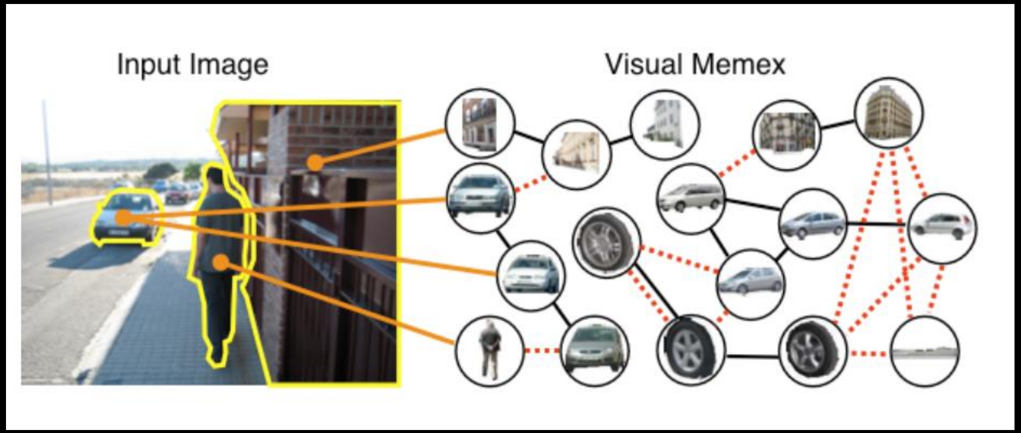


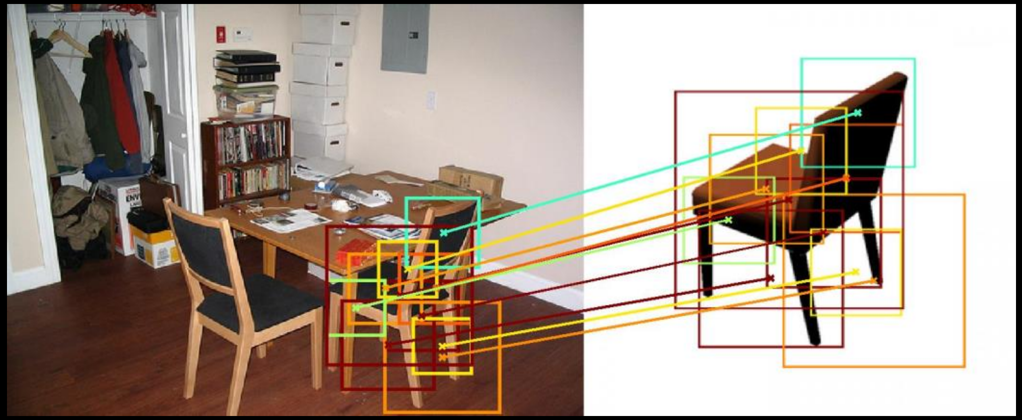
Image-Level Embedding
[van der Maaten and Hinton 2008]



Pixel-Level Graph
[Zhou et al 2014]



Object Graph
[Malisiewicz and Efros 2009]

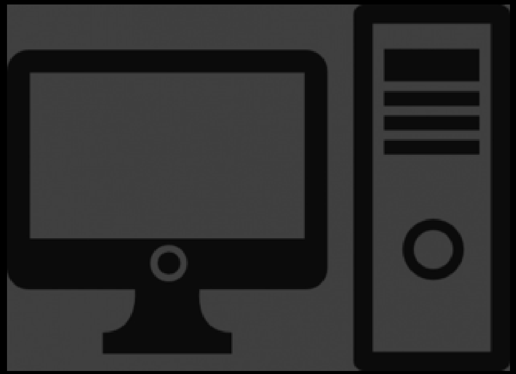


2D Image to 3D shape
[Aubry et al 2014]

What makes Big Visual Data hard?

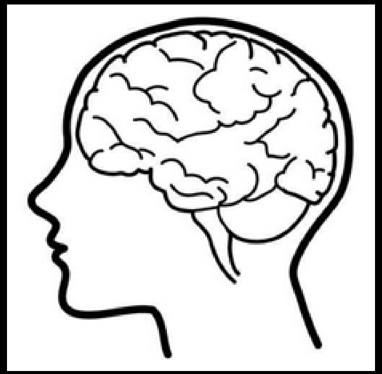
for Computers

1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data



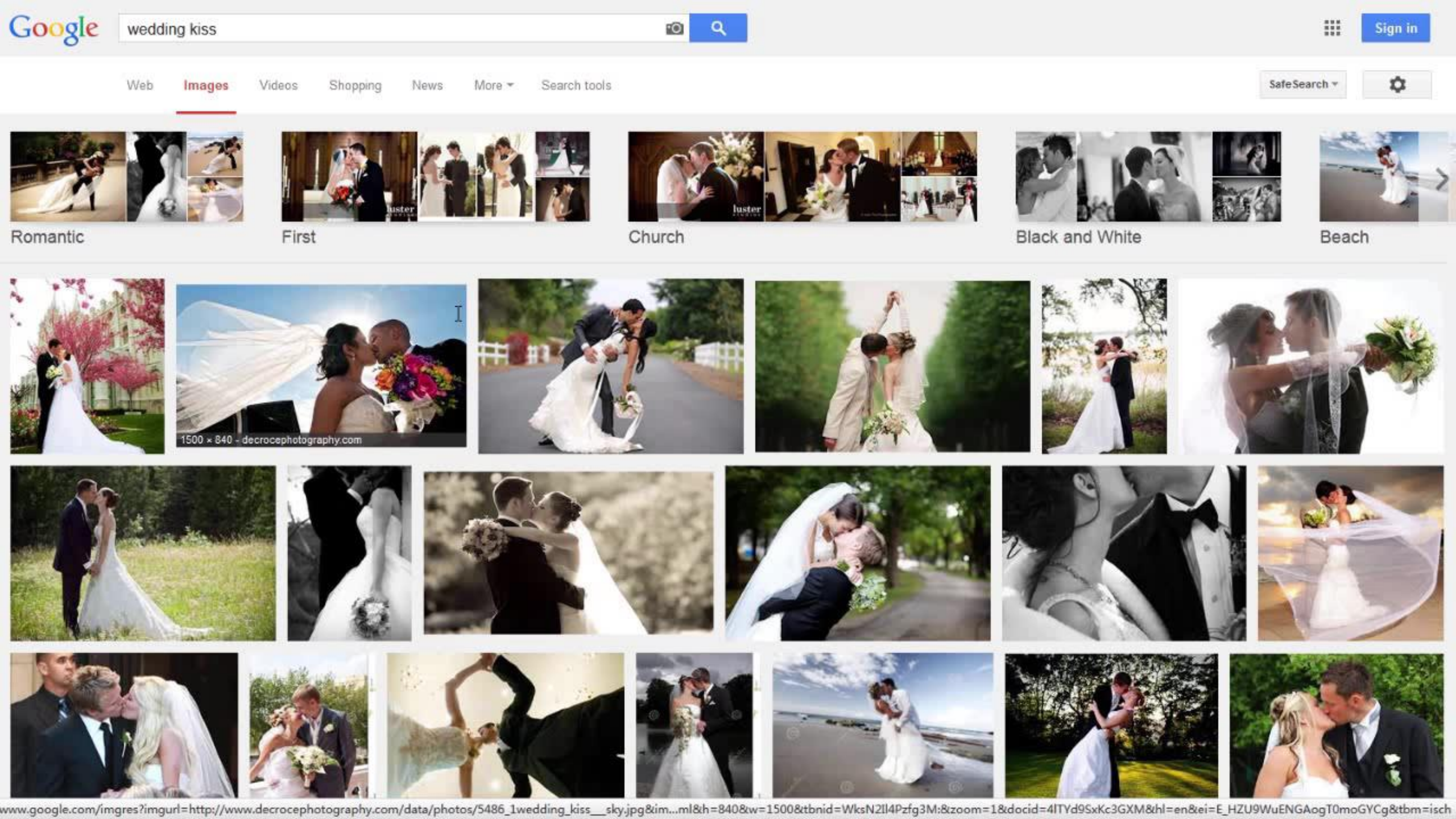
for Human Beings

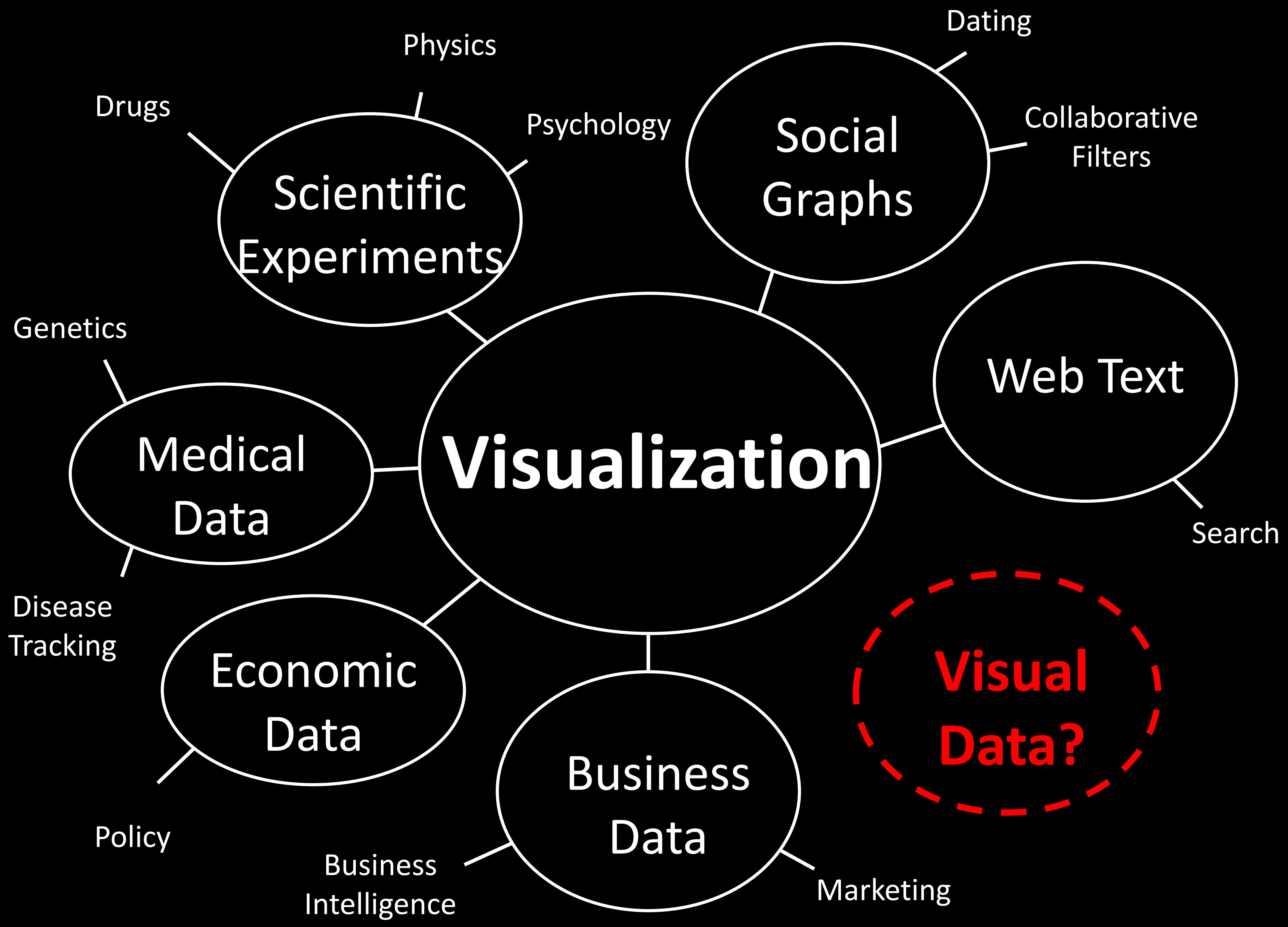
1. Visualizing Visual Data
2. Visual Communication



Too Big for Humans

[Perona 2010]


Google
wedding kiss
Sign in
Web
Images
Videos
Shopping
News
More
Search tools
SafeSearch
Romantic
First
Church
Black and White
Beach
1500 × 840 - decrocephotography.com
www.google.com/imgres?imgurl=http://www.decrocephotography.com/data/photos/5486_1wedding_kiss__sky.jpg&im...ml&h=840&w=1500&tbnid=WksN2Il4Pzfg3M:&zoom=1&docid=4lTYd9SxKc3GXM&hl=en&ei=E_HZU9WuENGAogT0moGYCg&tbm=isch


Physics
Dating
Drugs
Psychology
Social Graphs
Collaborative Filters
Scientific Experiments
Genetics
Web Text
Medical Data
Visualization
Search
Disease Tracking
Economic Data
Visual Data?
Business Data
Policy
Business Intelligence
Marketing

Data Visualization: the First Step

Data: Siggraph paper scores

Average score

4.5 4.0 3.5 2.5 3.5

3.6

Data: a collection of photos

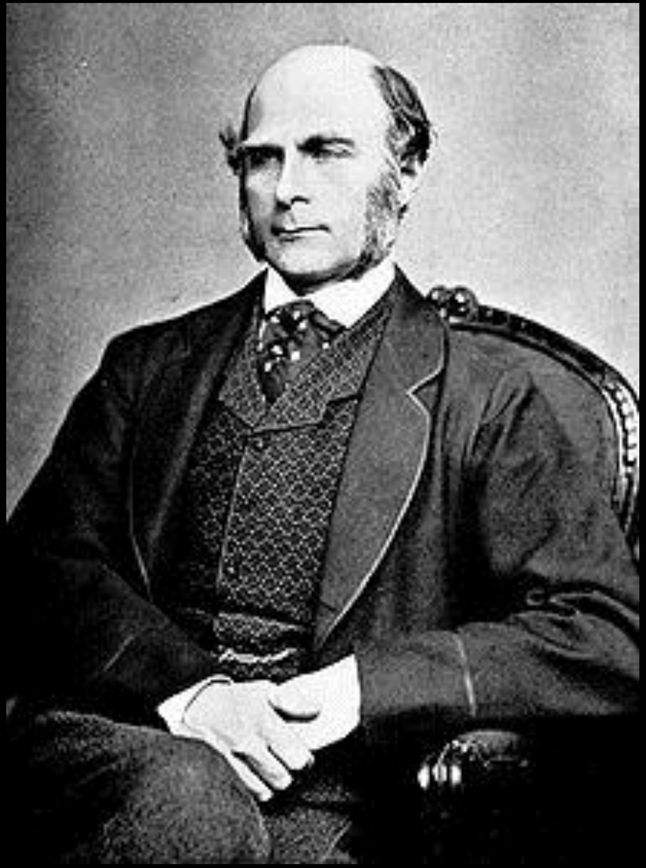
Average image

…

Visualizing Big Visual Data

Jun-Yan Zhu, Yong Jae Lee and Alexei A. Efros. *AverageExplorer: Interactive Exploration and Alignment of Visual Data Collections*. SIGGRAPH 2014.

Image Averaging



Sir Francis Galton

1822-1911

Multiple Individuals

Composite

[Galton, “Composite Portraits”, Nature, 1878]

Average Images in Art

"60 passagers de 2e classe du metro, entre 9h et 11h"

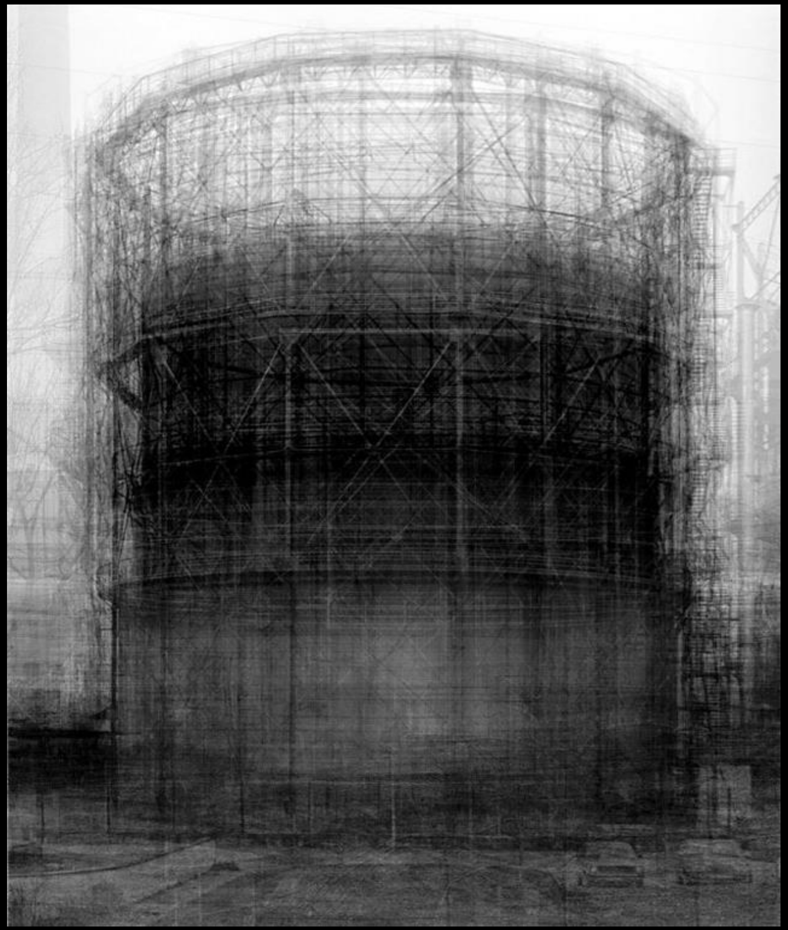
(1985)

Krzysztof Pruszkowski

"Dynamism of a cyclist"

(2001)

James Campbell



"Spherical type gasholders"

(2004)

Idris Khan

"100 Special Moments" (2004) by Jason Salavon

Newlyweds

Little Leaguer

Kids with Santa

Not so simple…

Jason Salavon

"Kids with Santa"

Google query result:

"**`kids with Santa`**"

Automatic Average

Why Difficult?

Google results

Visual Modes

Misaligned

“Object-Centric Averages” (2001) by Antonio Torralba

…

Manual Annotation and Alignment

Average Image

With Alignment

Our Goal:

An interactive system to rapidly explore and align a large image collection using ***image averaging***

Weighted Averages + Alignment

Overview

Image Collection $\{I_1 \cdots I_N\}$ (e.g. “Kids with Santa” images)

Average I_{avg}

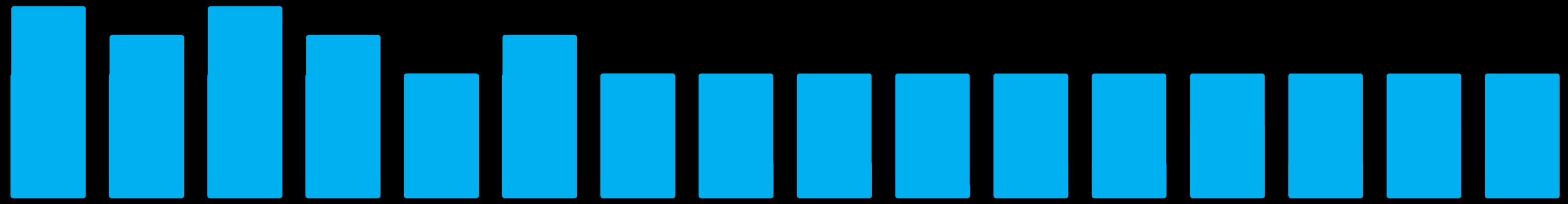
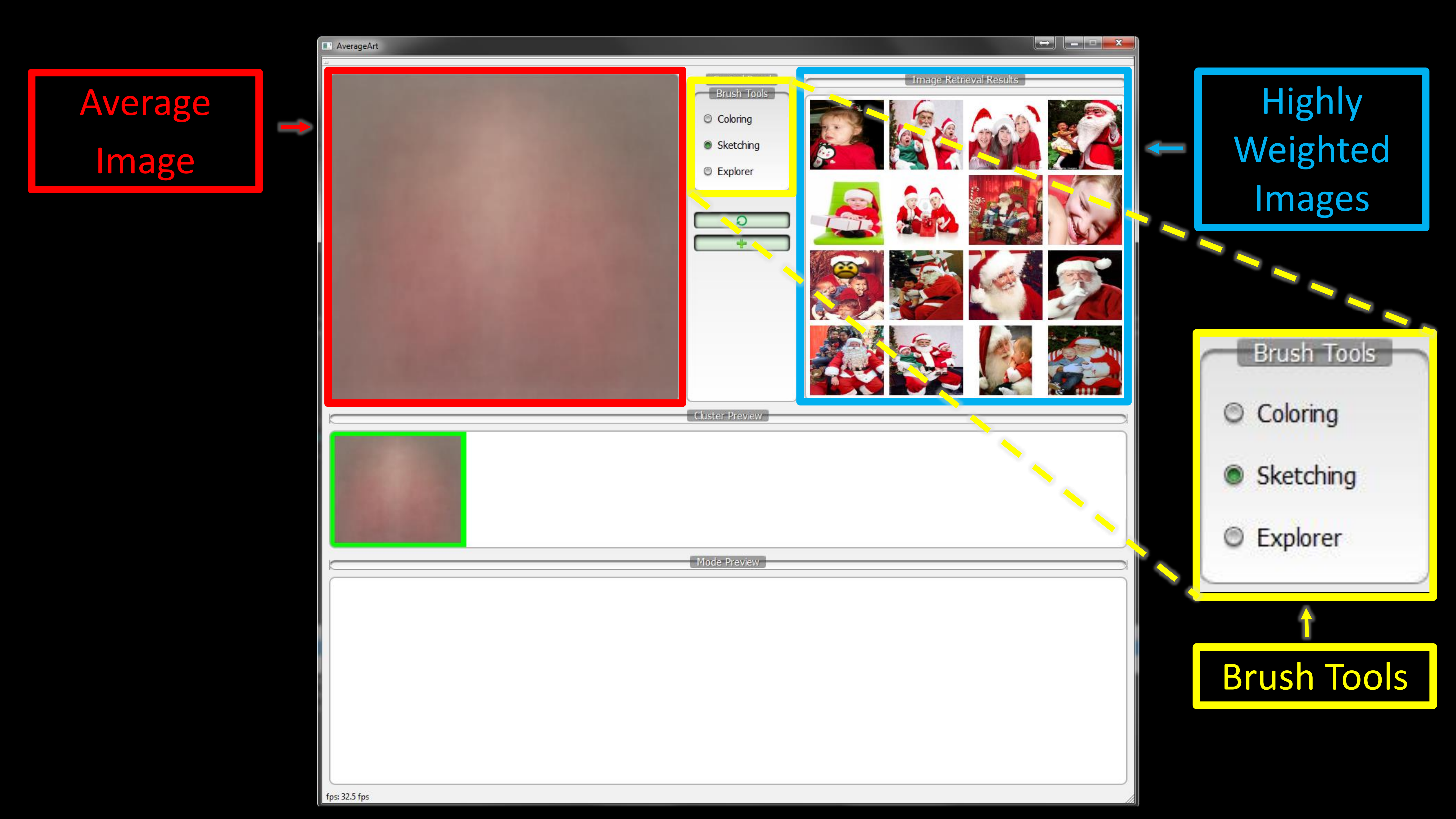


Image Weights $\{s_1 \cdots s_N\}$

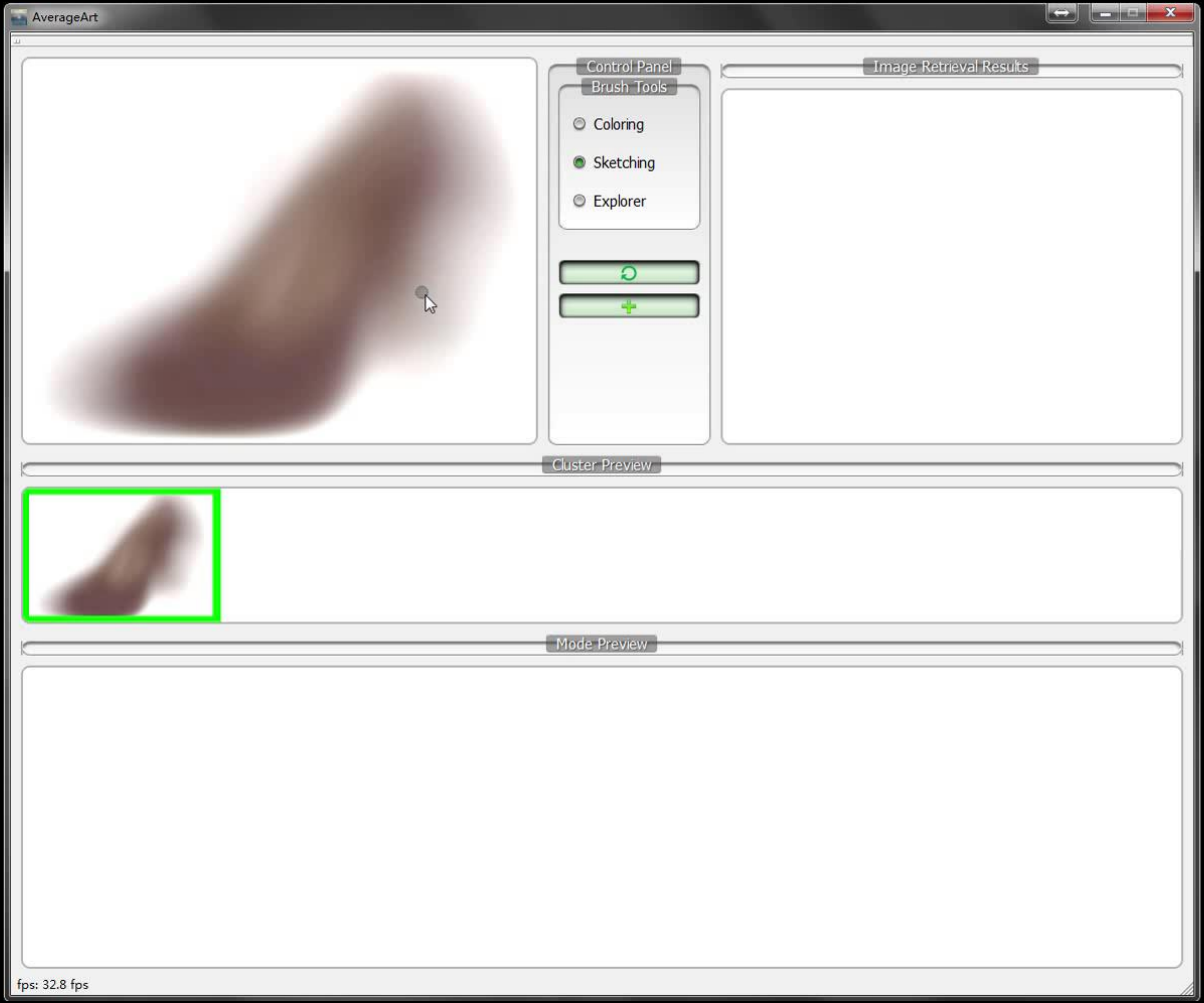
$$I_{avg} = \frac{1}{N}\sum_{i=1}^{N} I_i$$

$$I_{avg} = \sum_{i=1}^{N} s_i I_i$$



Average Image
AverageArt
Brush Tools
Coloring
Sketching
Explorer
Image Retrieval Results
Cluster Preview
Mode Preview
fps: 32.5 fps
Highly Weighted Images
Brush Tools
Coloring
Sketching
Explorer
Brush Tools

Zappos “Shoes” (5, 703 Images)

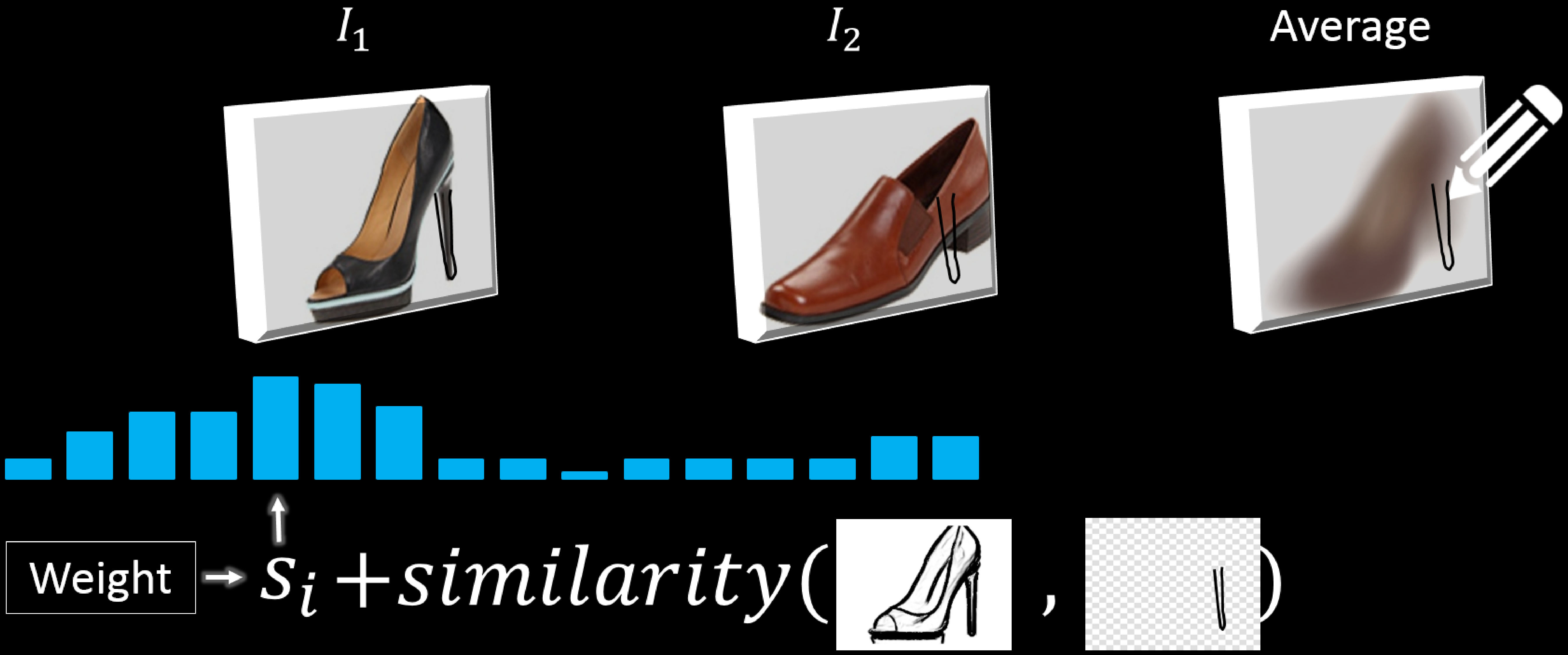


Sketching Brush

ShadowDraw

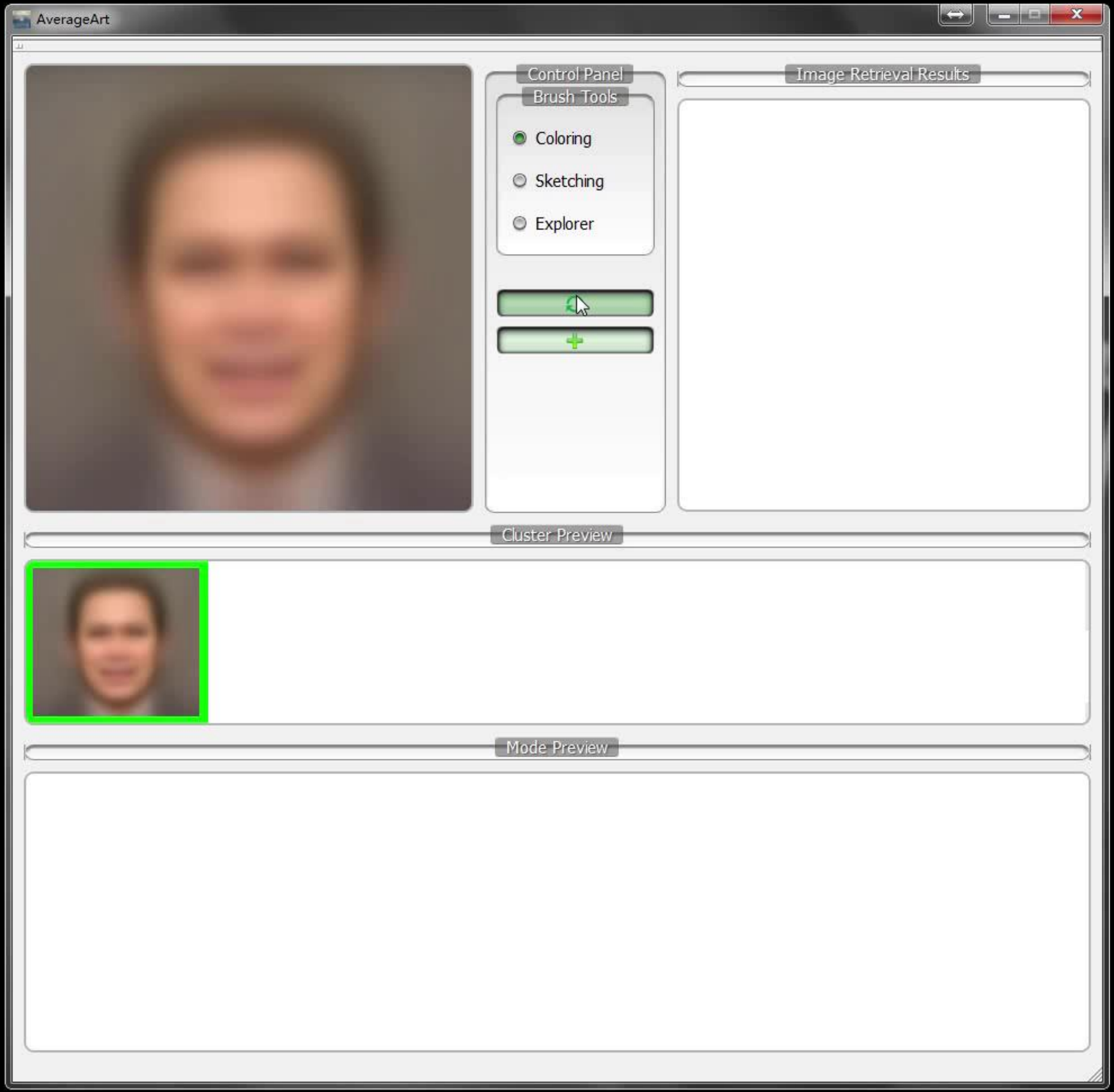
[Lee et al. 2011]

Sketching Brush



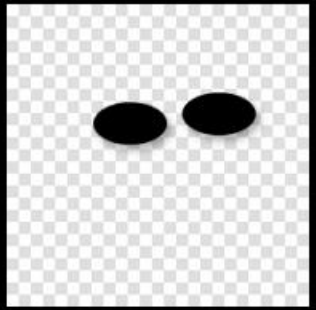
“Face” Dataset (13,233 Images)

Coloring Brush



Coloring Brush

I_1 I_2 Average

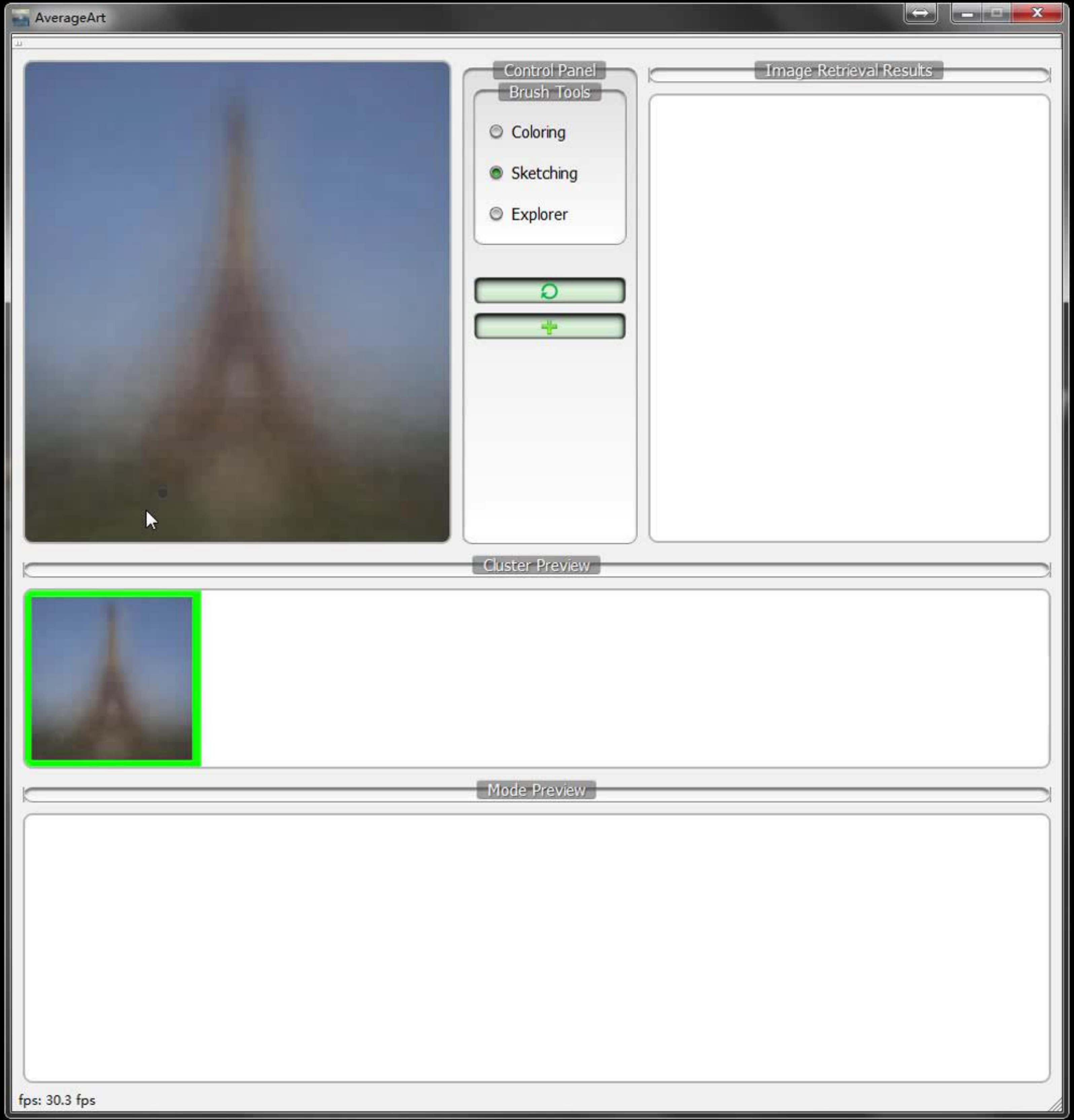
Weight → $s_i + similarity($  ,  $)$

Flickr + Google Query:
'Eiffel Tower'
(412 Images)

Sketching Brush

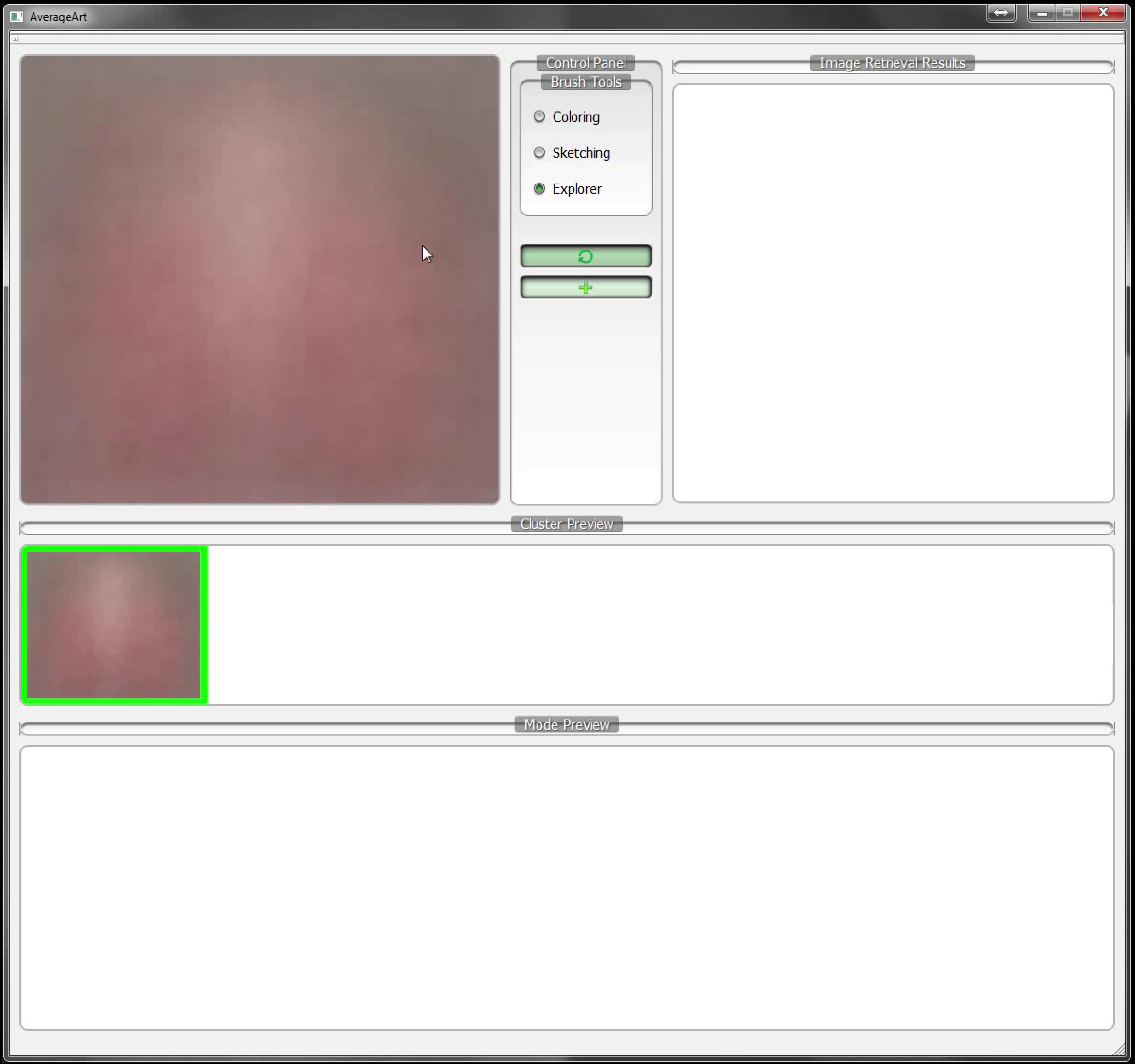
+

Coloring Brush



How to Start?

Blurry Average



Explorer Brush

Explorer Brush: Select a Local Mode

N Local Patches

…

Visual Mode Discovery

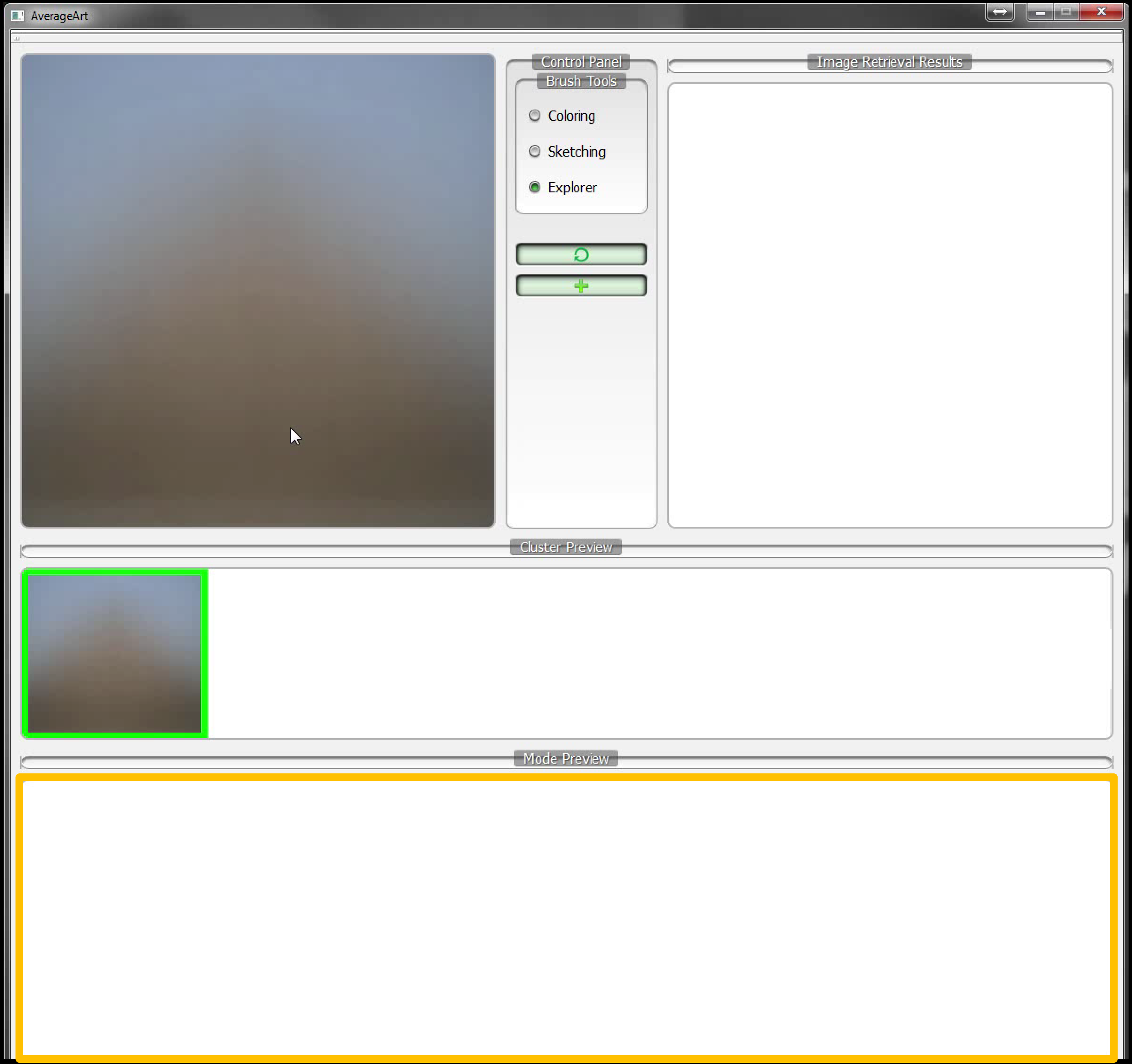
Local Visual Modes

Average

$s_i = s_i + similarity($  ,  $)$

Mid-level Discriminative Patch Discovery [Doersch et al. 2012]

Google Query
‘Church’
(11,007 Images)



Select different modes at the same location

Local visual modes

Weighted Averages + Alignment

Image Alignment

Flickr + Google Query
'Bridge of Sighs'
(829 Images)

Bridge of Sighs
Oxford

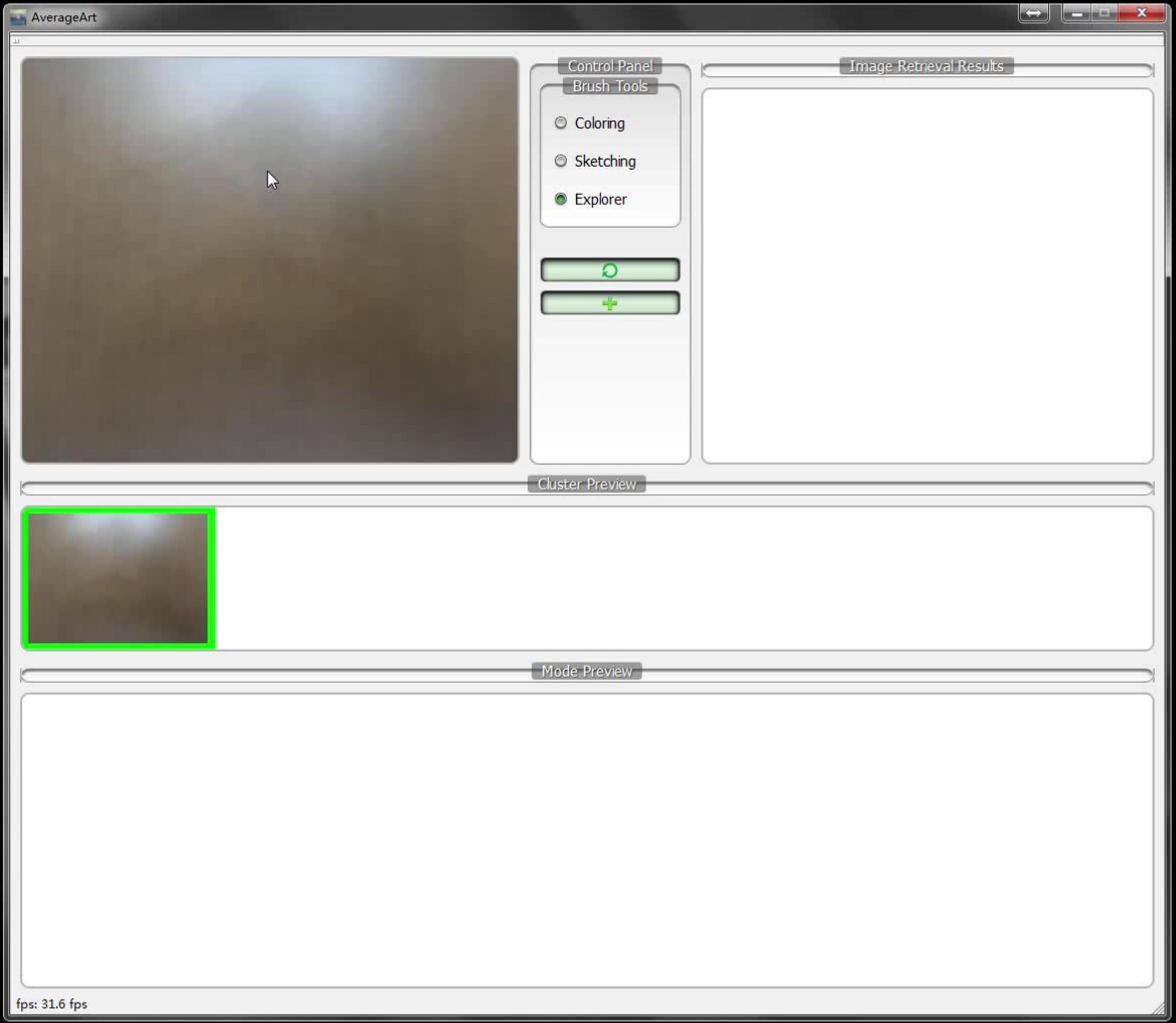
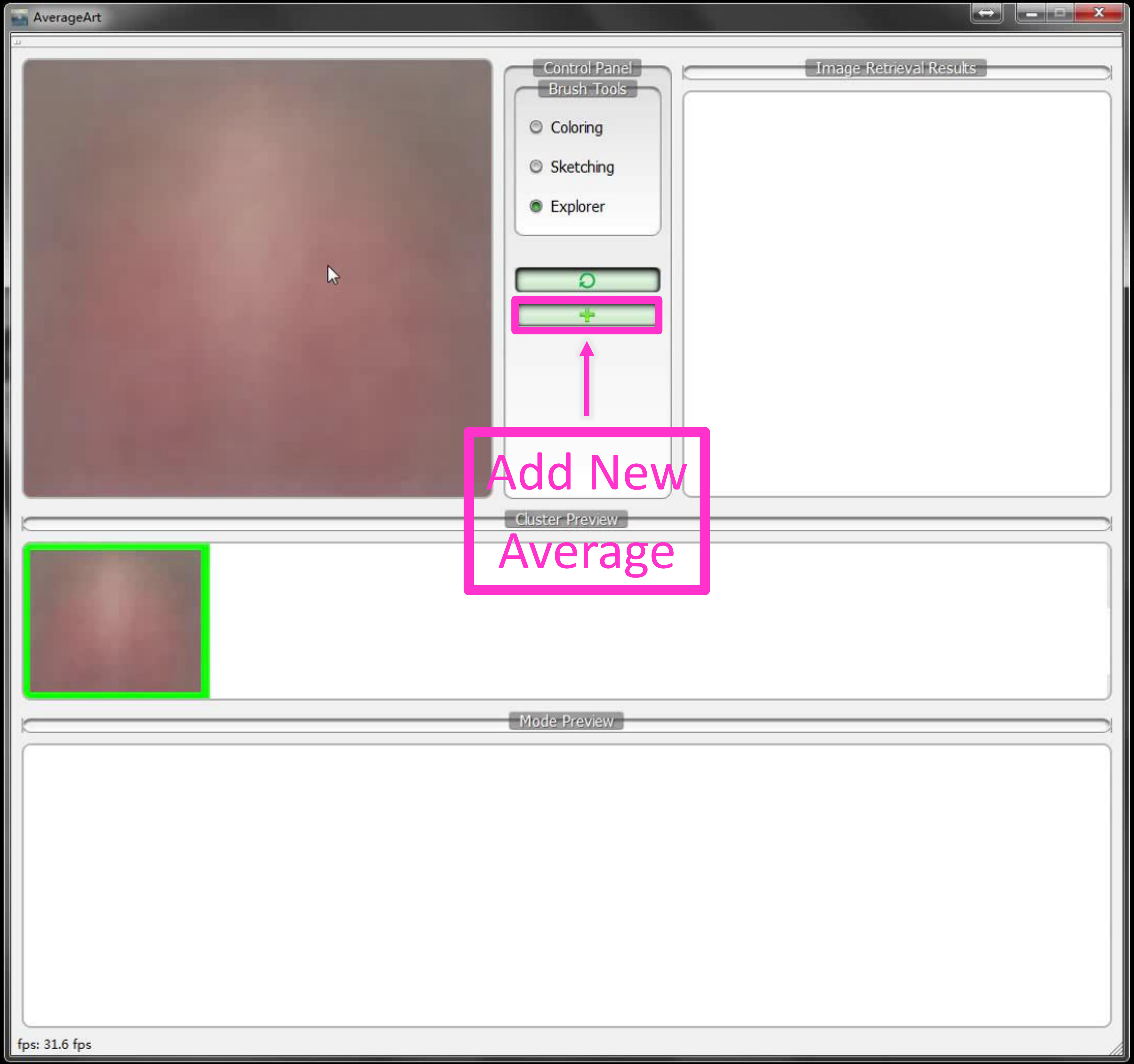


Image Warping

Moving Least Square
[Schaefer et al. 2006]

Google Query
'Kids with Santa'
(1,640 Images)

Creating Multiple Averages



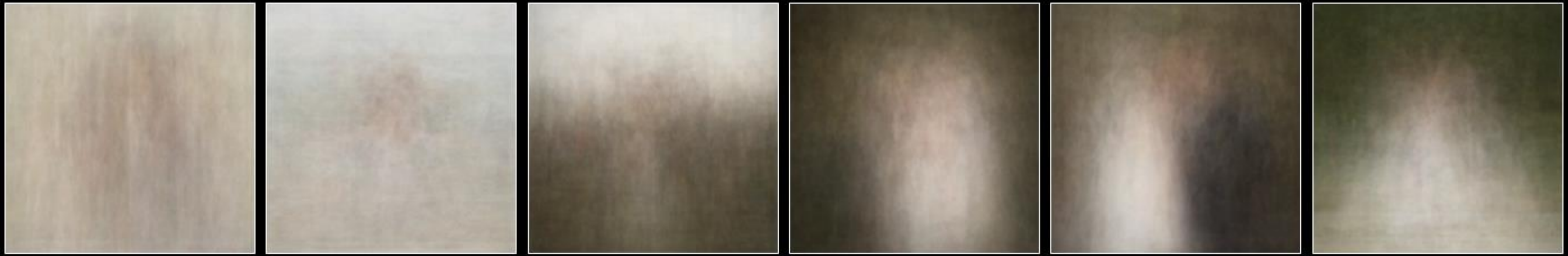
Automatic Clustering

- K-means, GMM
- Spectral Clustering
  - e.g. [Shi and Malik 2000]
- Discriminative Clustering
  - e.g. [Hoai and Zisserman 2013]

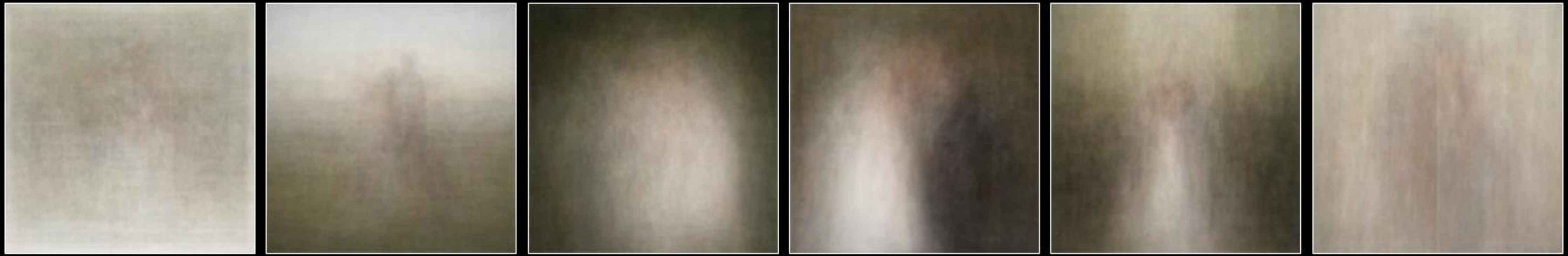
Automatic Clustering

Google Query
'Wedding Kiss'
(16, 868 Images)

Average Image



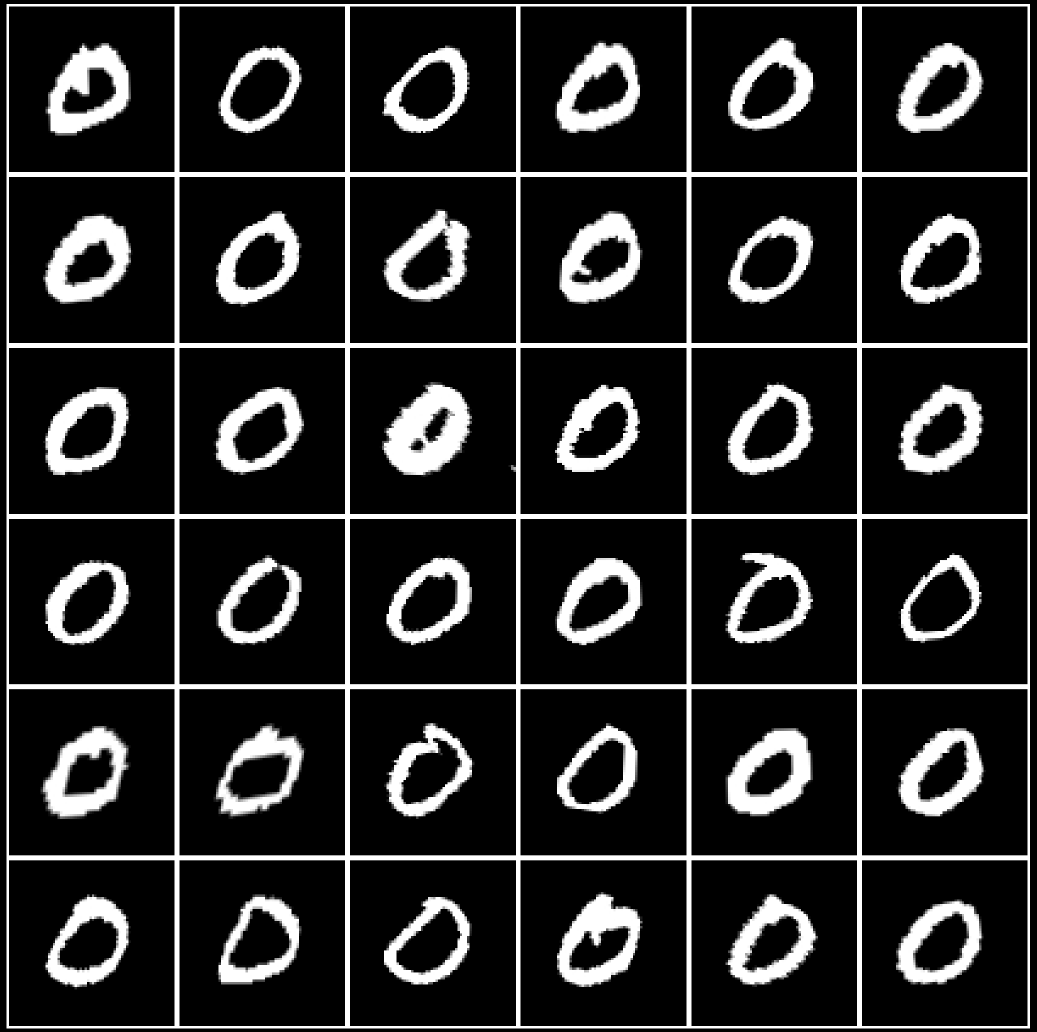
K-means



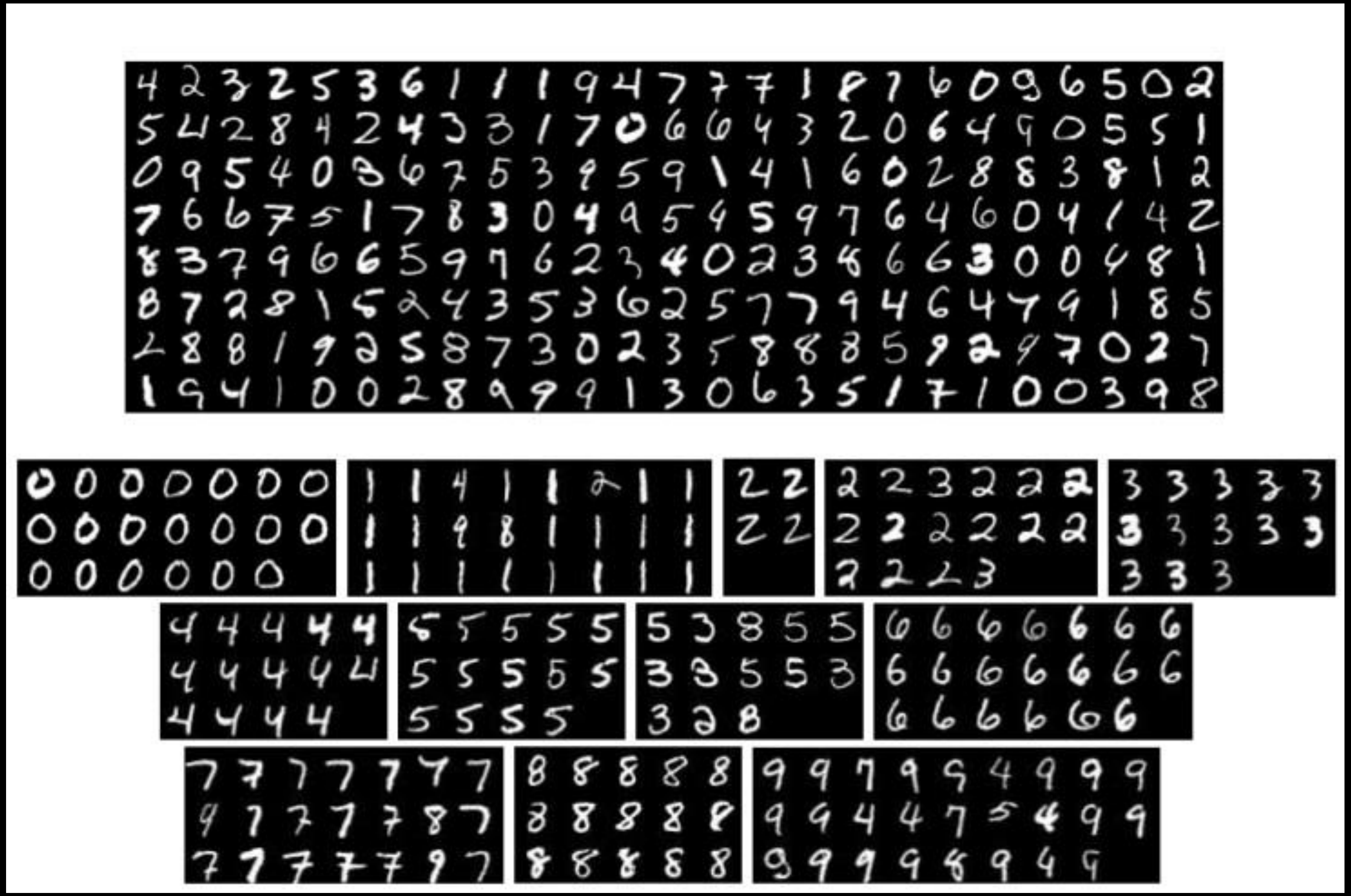
Spectral Clustering [Shi and Malik 2000]

Discriminative Clustering [Hoai and Zisserman 2013]

Automatic Alignment



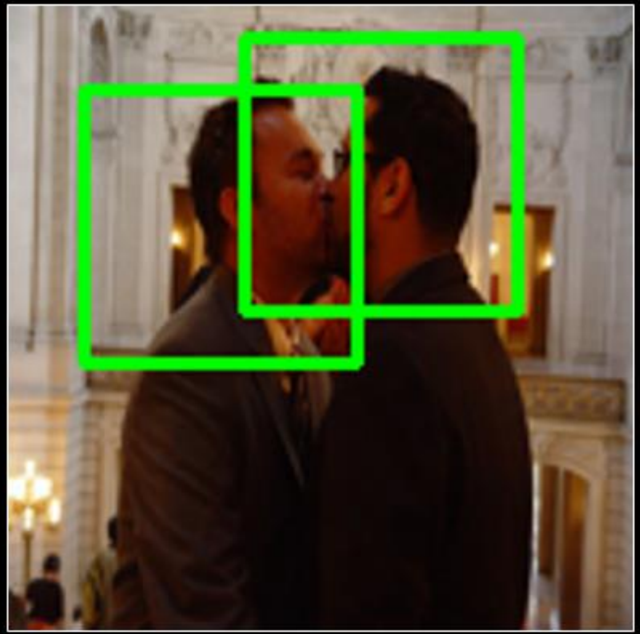
[Learned-Miller 06]
[Huang et al. 07]

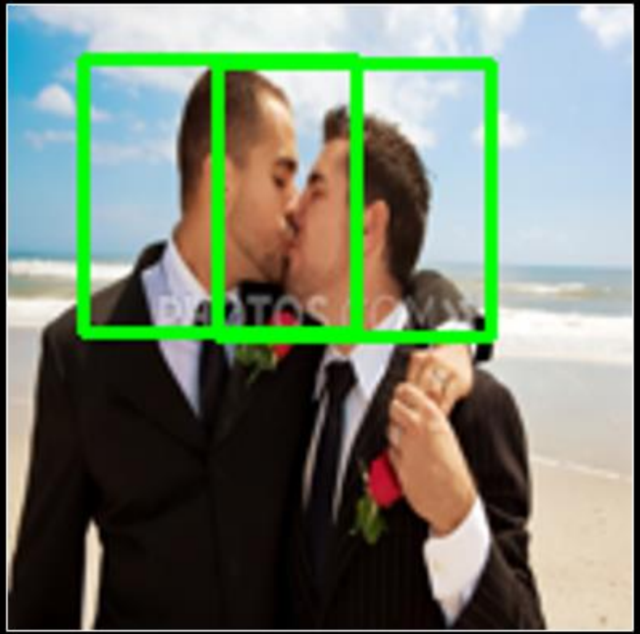


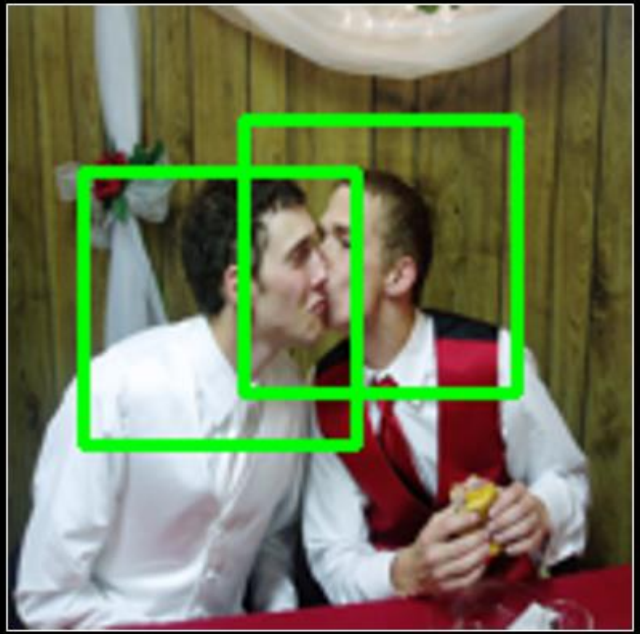
[Mattar et al. 12]

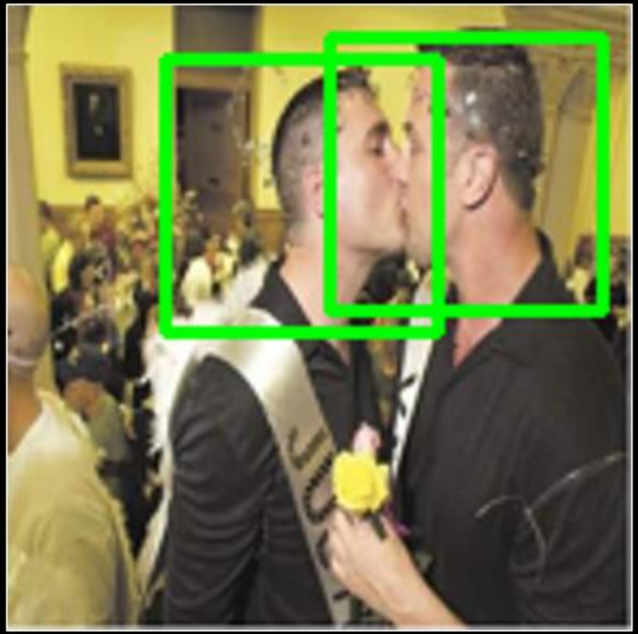
Interactive Clustering and Alignment

Average image









Our Contribution:

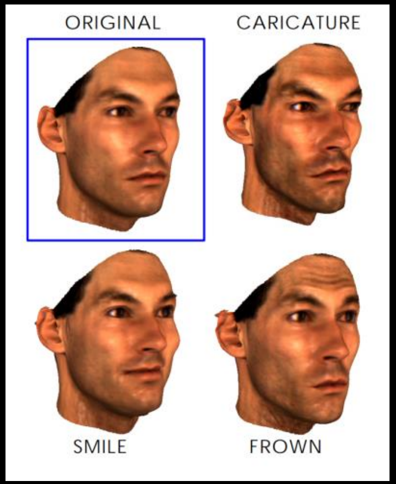
User-Guided Clustering

+

User-Guided Alignment

Face Keypoint Alignment

[Cootes et al. 1998]



[Blinz & Vetter, 1999]

Africa American | Afghan | Central African | Burmese | Cambodian | English

French | German | Greek | Indian | Iranian | Irish

“Average Face by Country”

using FaceResearch.org

Different Cat Breeds (Simple Average)

Abyssinian

Sphynx

Birman

Bombay

Egyptian Mau

Ragdoll

British Shorthair

Persian

Maine Coon

Russian Blue

Siamese

Bengal

Different Cat Breeds (Our Result)


Abyssinian


Sphynx


Birman


Bombay


Egyptian Mau


Ragdoll


British Shorthair


Persian

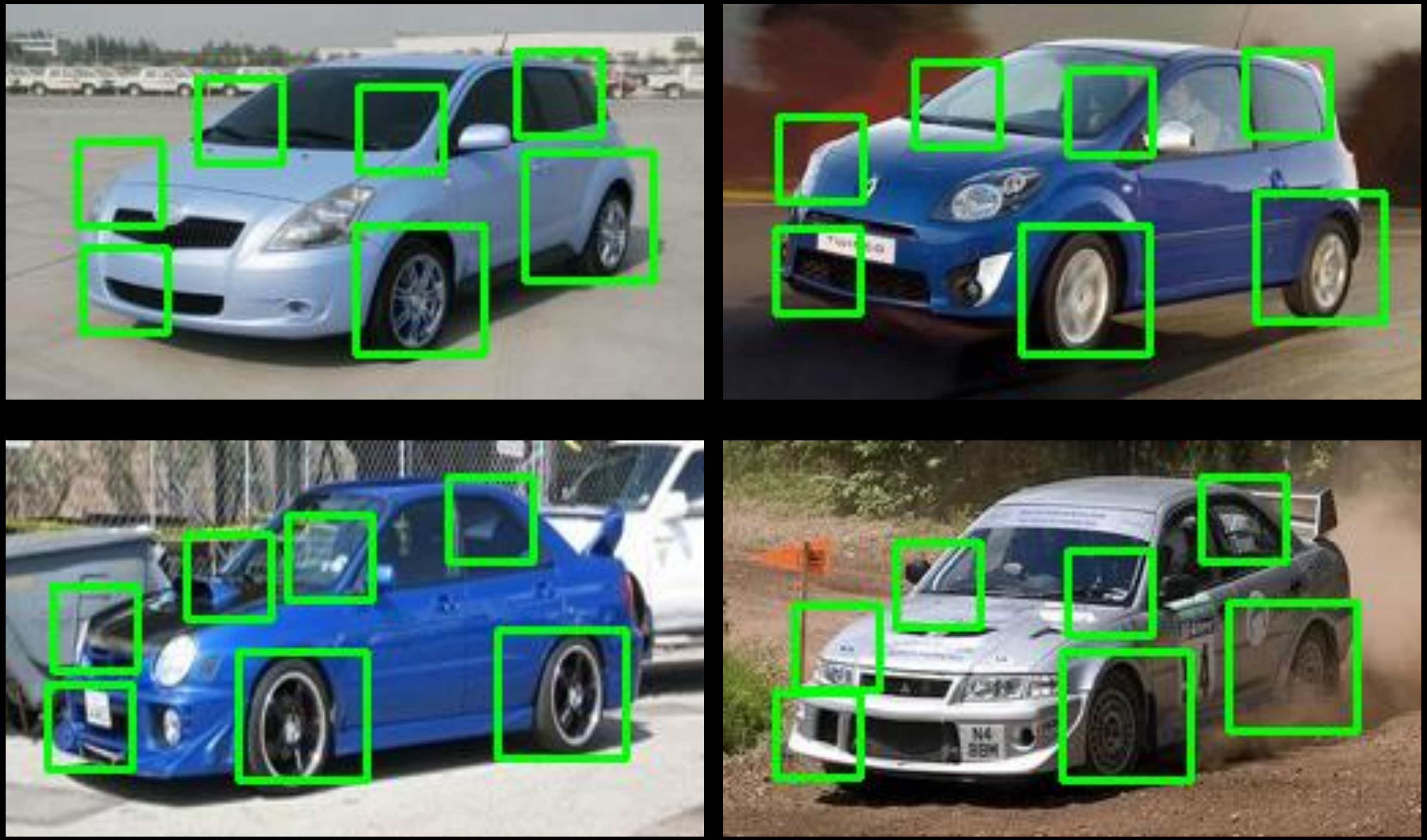

Maine Coon


Russian Blue

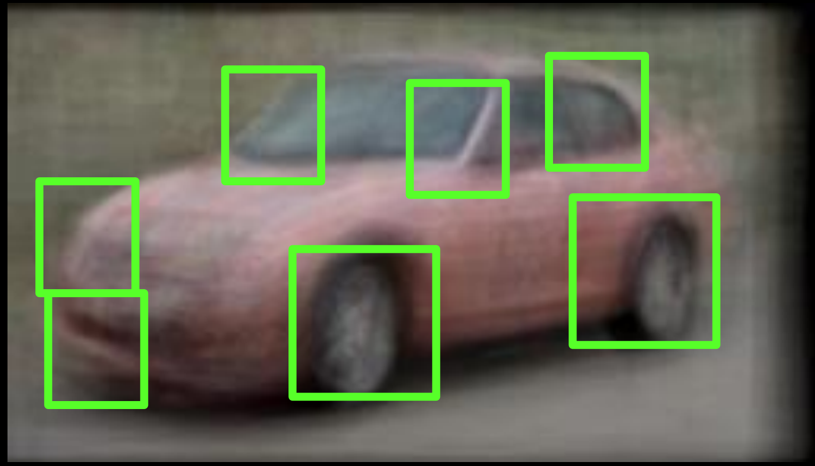

Siamese


Bengal

Application: Keypoint Annotation

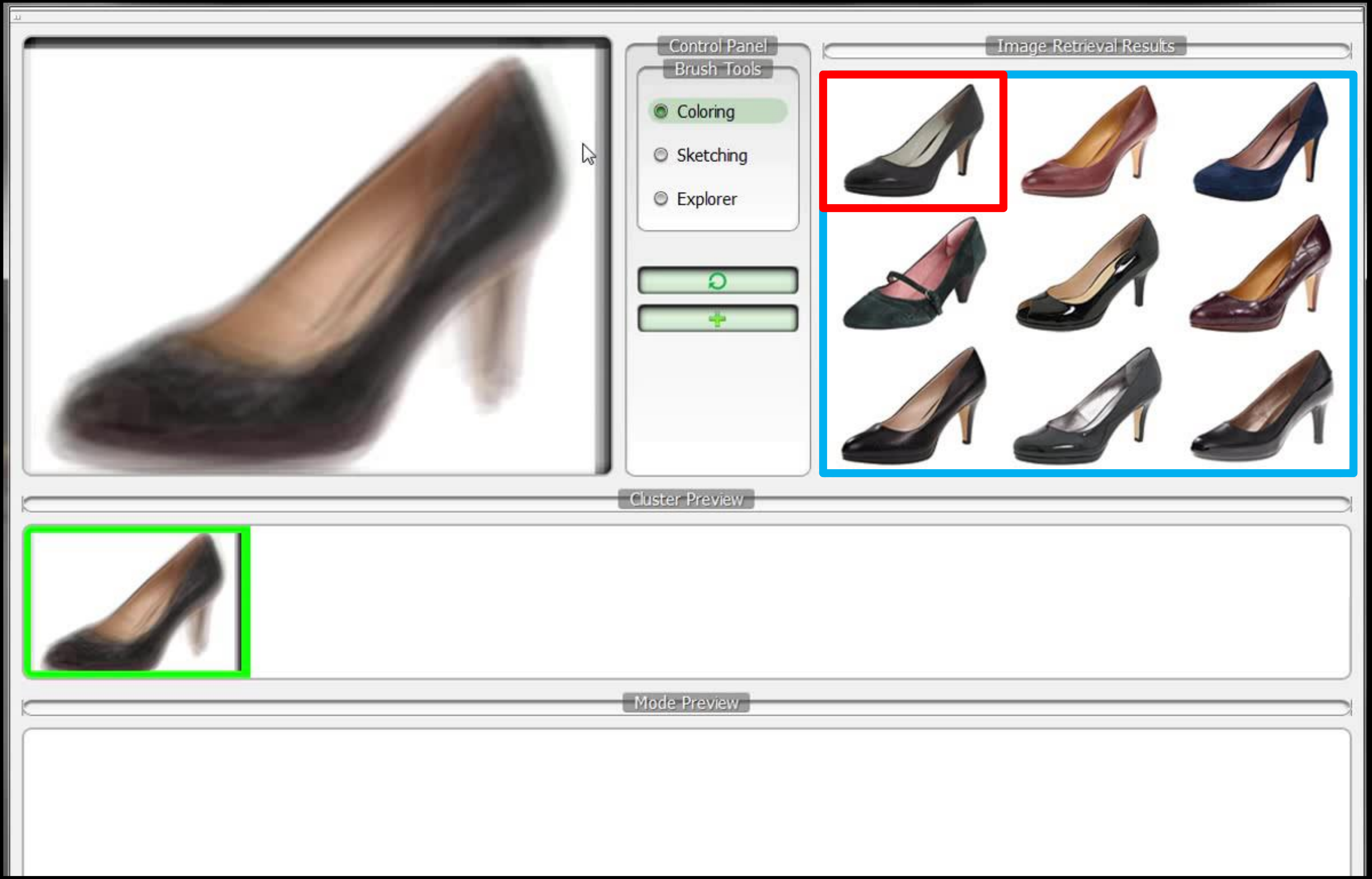


Car Parts Annotation



Average Image

Application: Online Shopping

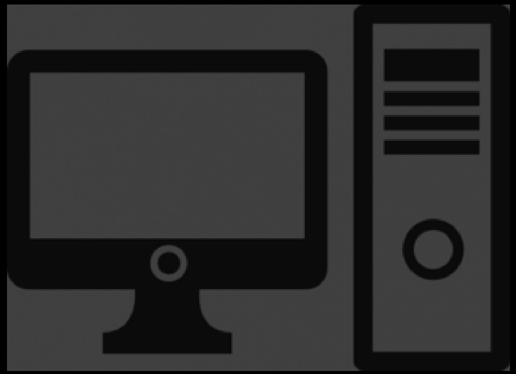


Recommended Products

What makes Big Visual Data hard?

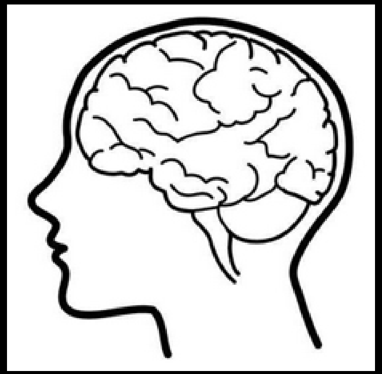
for Computers

1. Finding Correspondences
2. Mining Visual Data
3. Connecting Visual Data



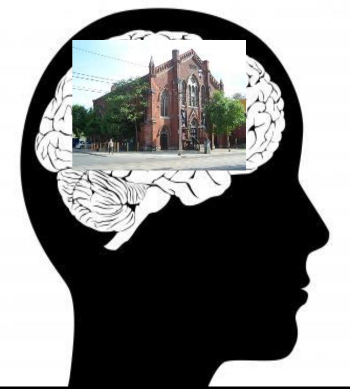
for Human Beings

1. Visualizing Visual Data
2. Visual Communication

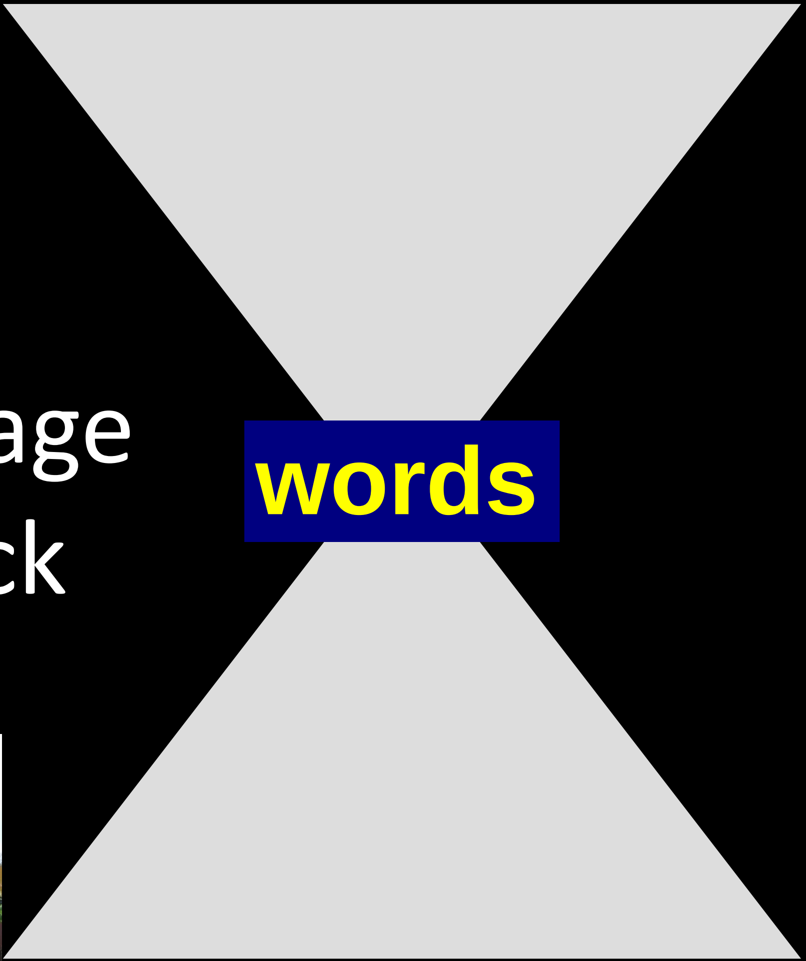


How to connect Humans’ Mental Picture to Big Visual Data?

Mental Picture



The Language Bottleneck

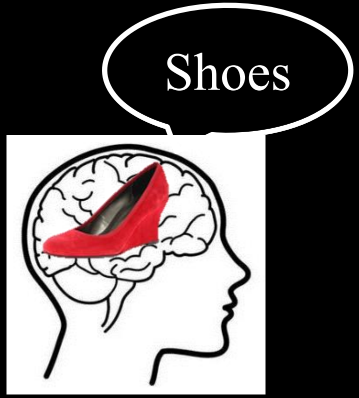


words

Image

Forensic Sketch

The Identi-Kit System








THANK YOU!