Exemplar-SVM: Object Detection, Cross-domain Image Matching, and Beyond

Tomasz Malisiewicz
(Massachusetts Institute of Technology)

Joint work with:
Abhinav Shrivastava, Abhinav Gupta and Alexei A. Efros
(Carnegie Mellon University)

Medici Fountain, Paris

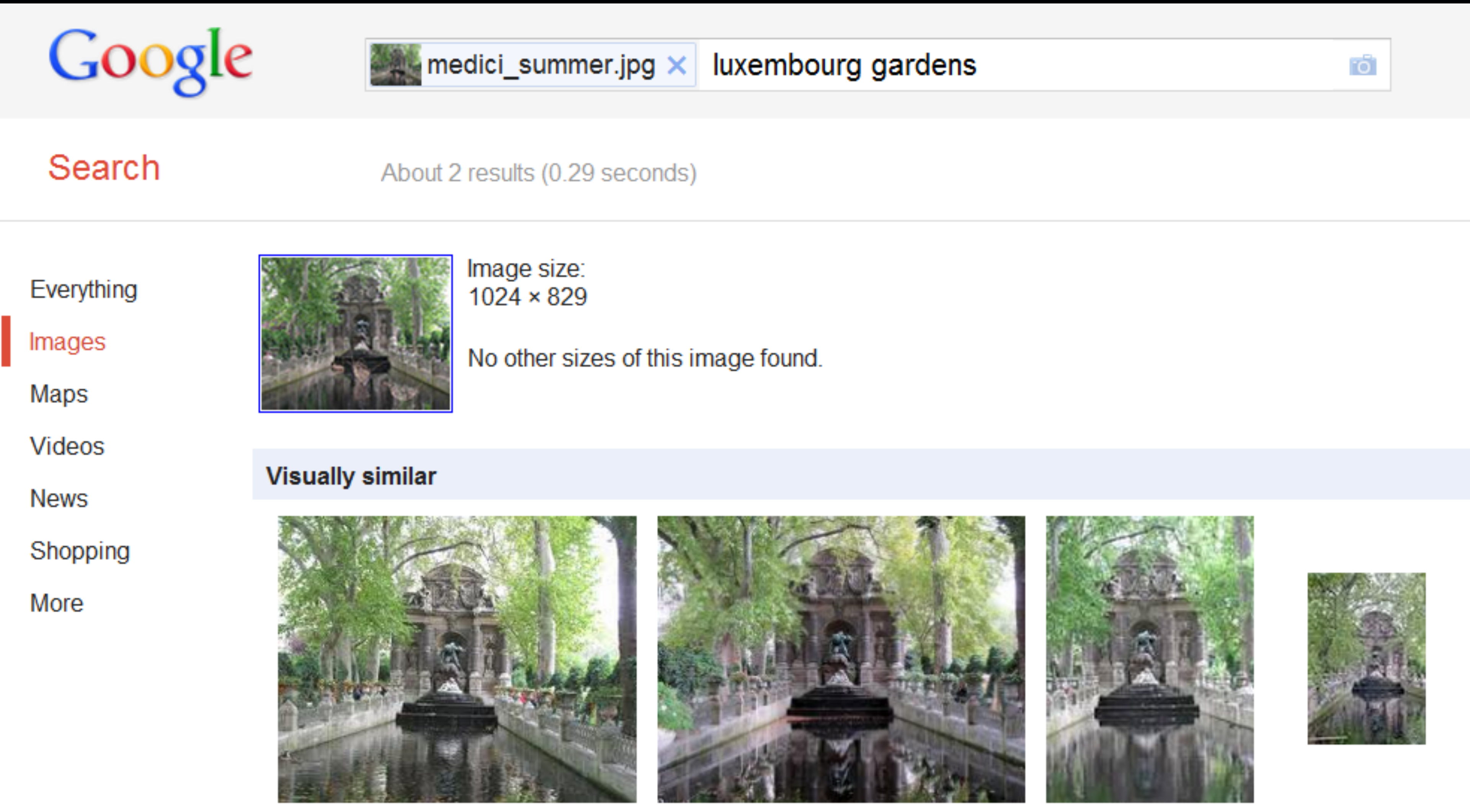
Google images

Search by image

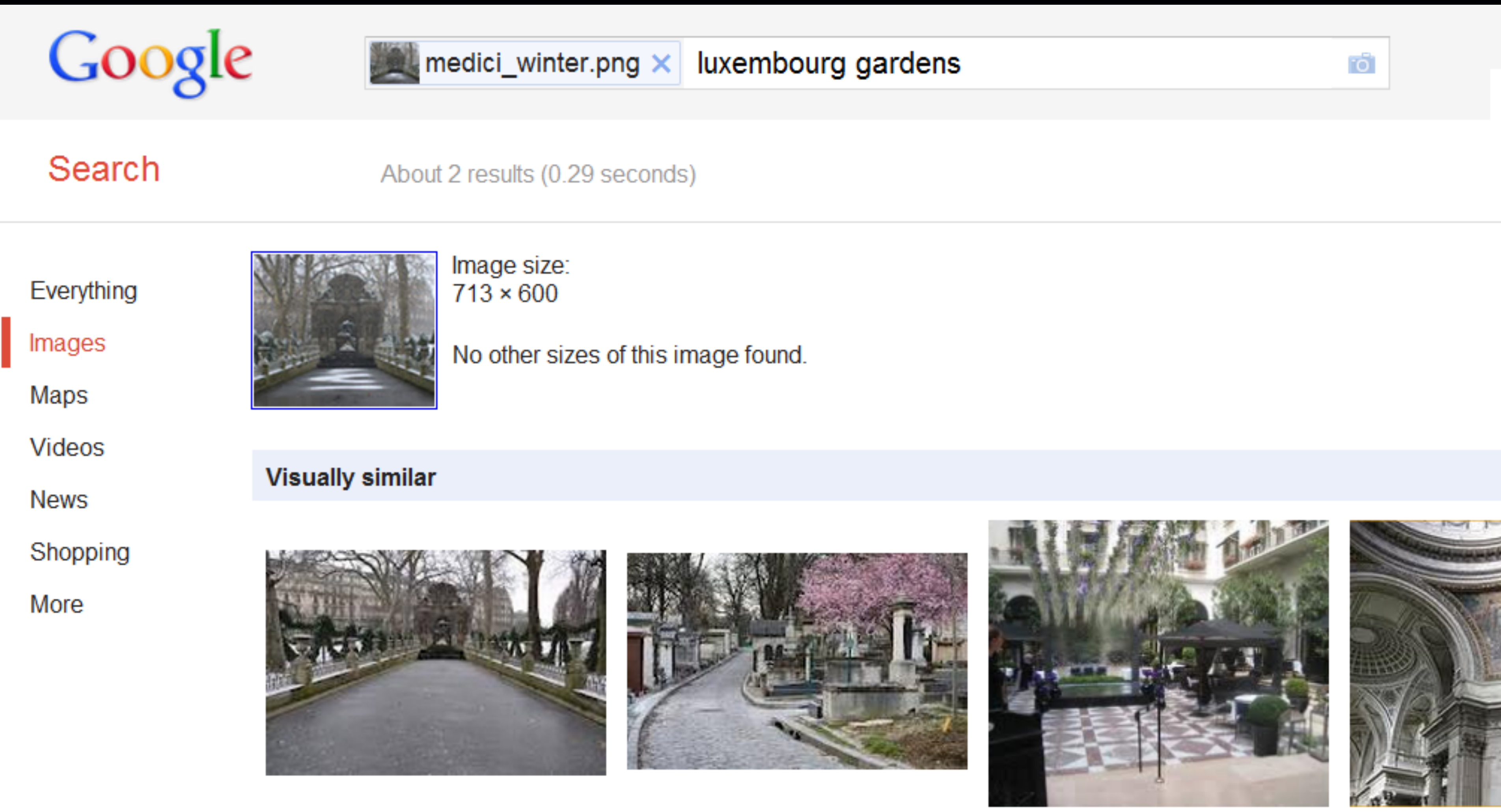
Drop image here

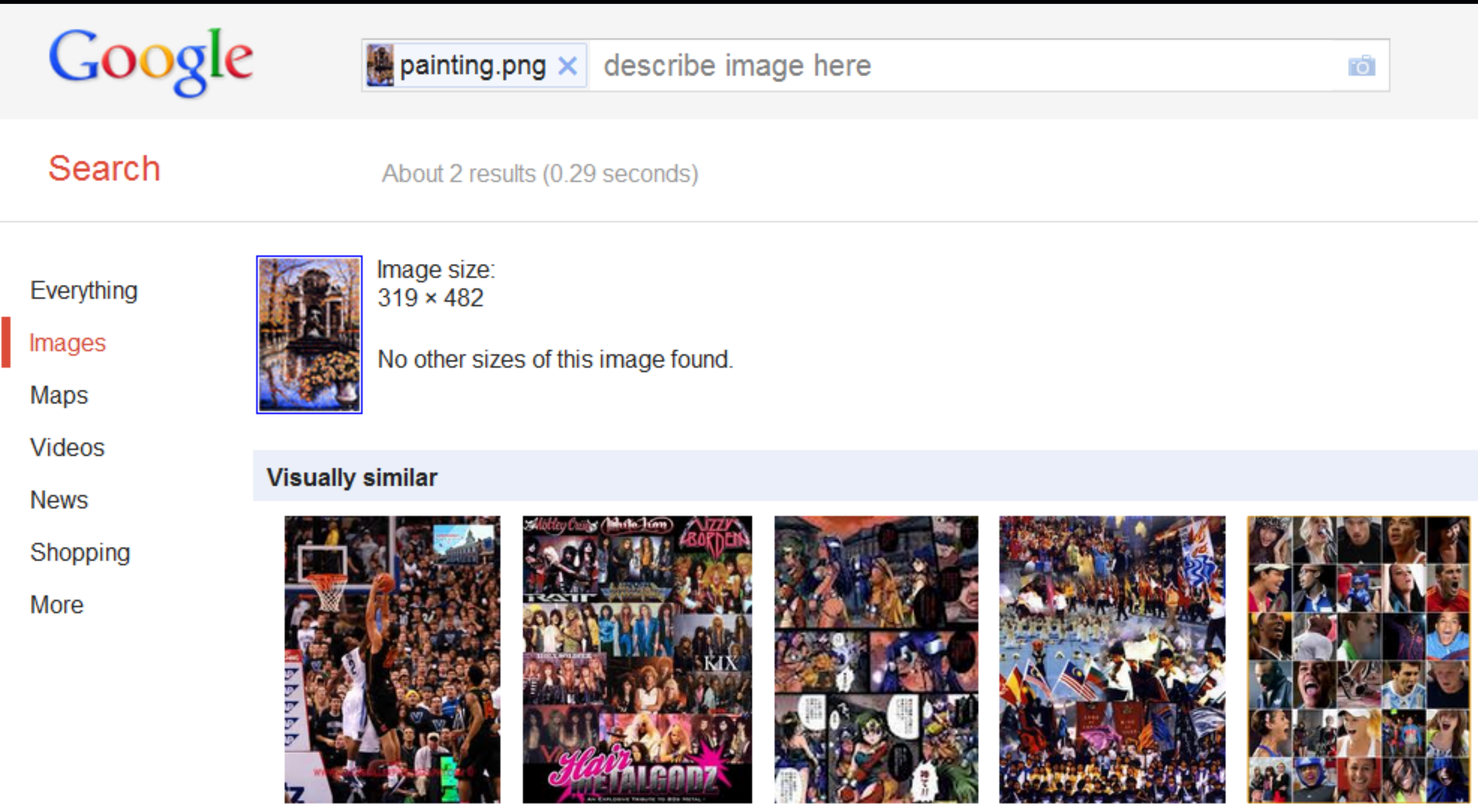
Move

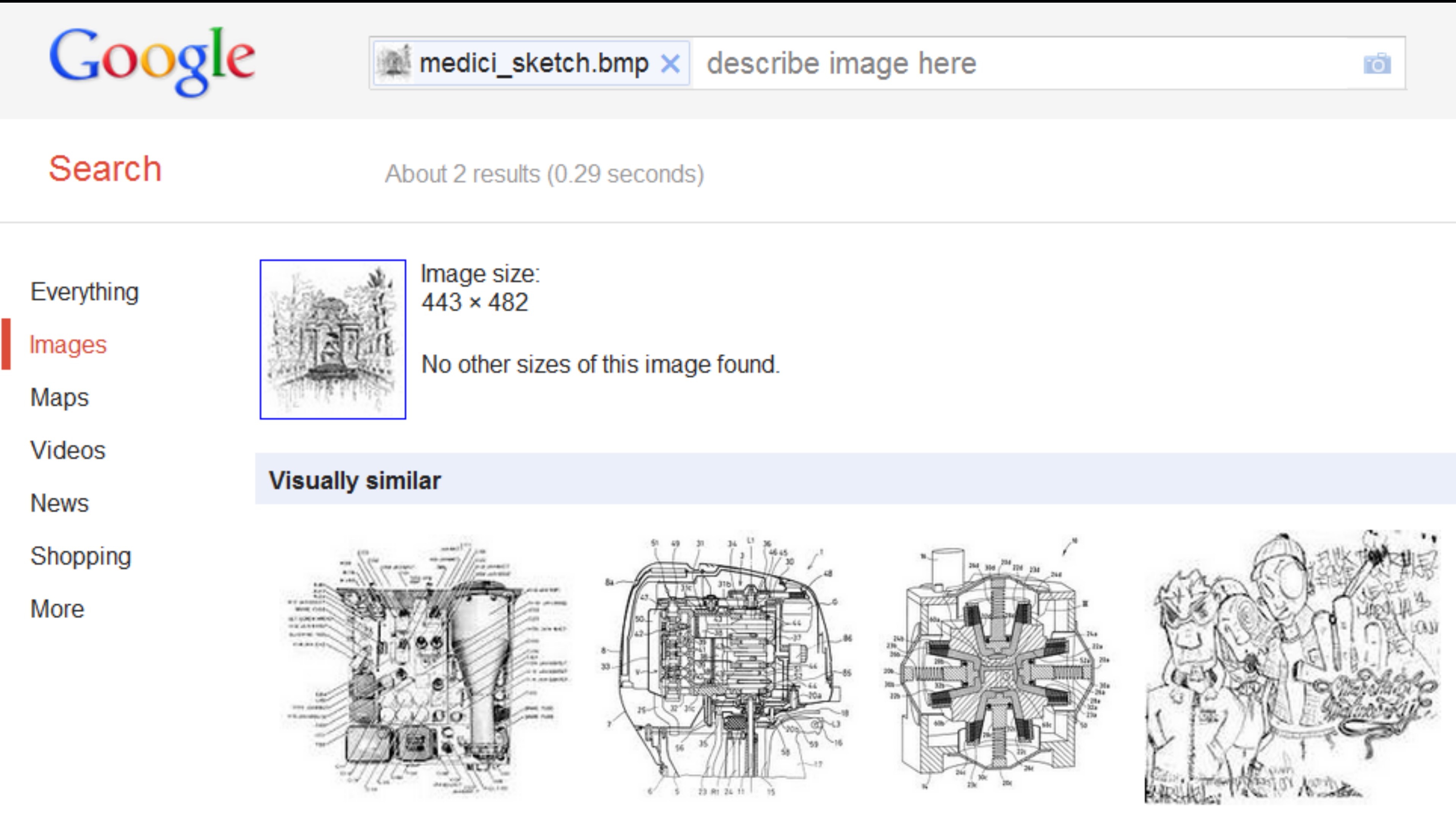
Watch a short video to learn more.


Google
medici_summer.jpg
luxembourg gardens
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
1024 × 829
No other sizes of this image found.
Visually similar

Medici Fountain, Paris (winter)


Google
medici_winter.png
luxembourg gardens
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
713 × 600
No other sizes of this image found.
Visually similar

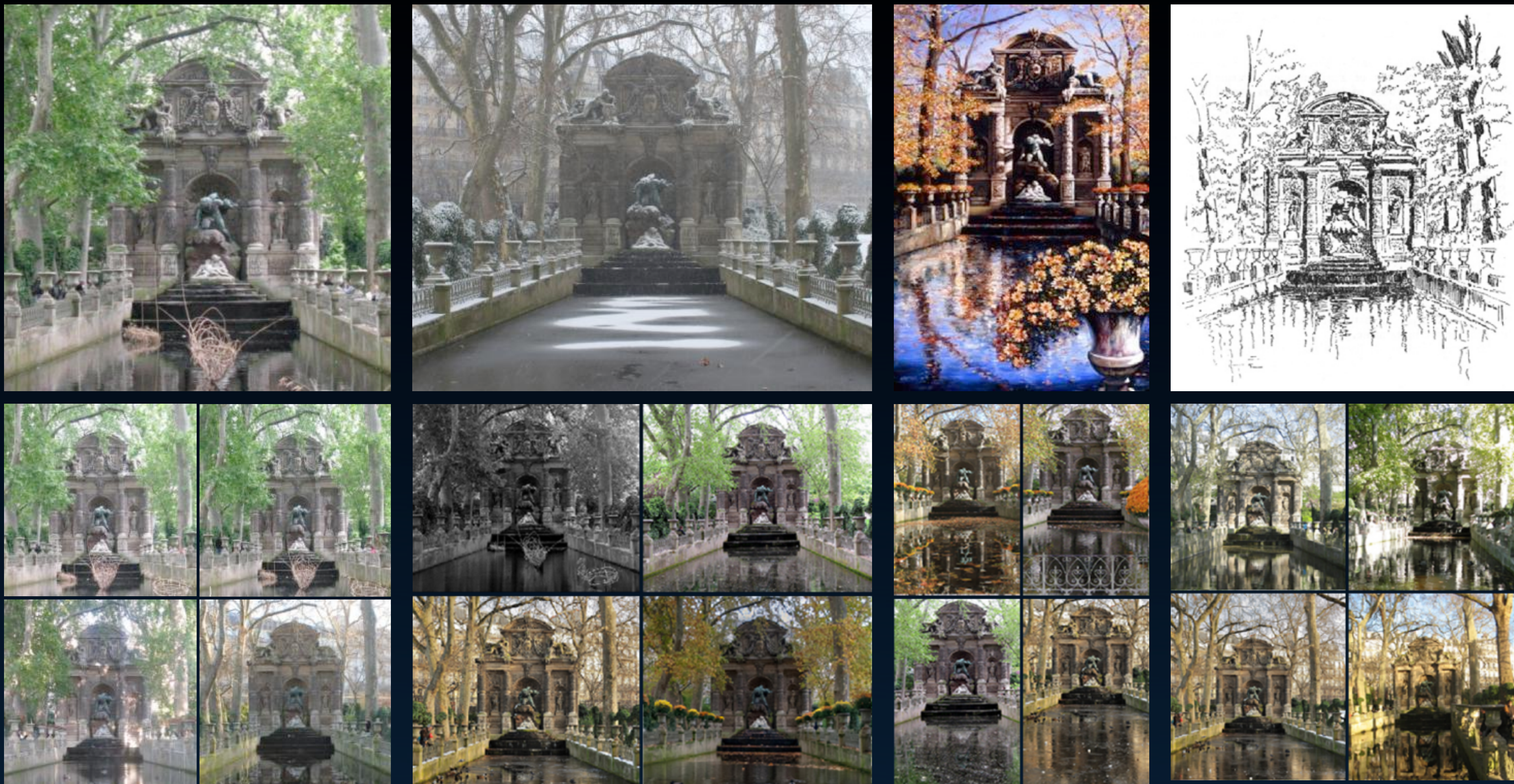

Google
painting.png
describe image here
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
319 × 482
No other sizes of this image found.
Visually similar


Google
medici_sketch.bmp
describe image here
Search
About 2 results (0.29 seconds)
Everything
Images
Maps
Videos
News
Shopping
More
Image size:
443 × 482
No other sizes of this image found.
Visually similar

OUR GOAL

Abhinav Shrivastava, Tomasz Malisiewicz, Abhinav Gupta and Alexei A. Efros.
Data-driven Visual Similarity for Cross-domain Image Matching.
In SIGGRAPH ASIA, 2011.

OUR GOAL



Abhinav Shrivastava, Tomasz Malisiewicz, Abhinav Gupta and Alexei A. Efros.
Data-driven Visual Similarity for Cross-domain Image Matching.
In SIGGRAPH ASIA, 2011.

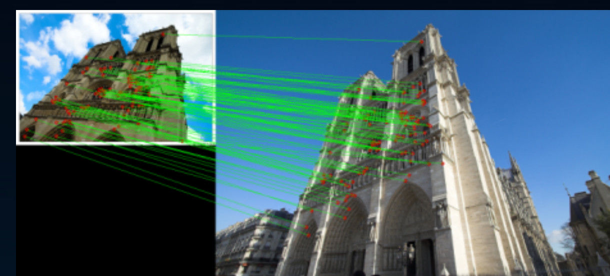
WHY IS THIS SO HARD?



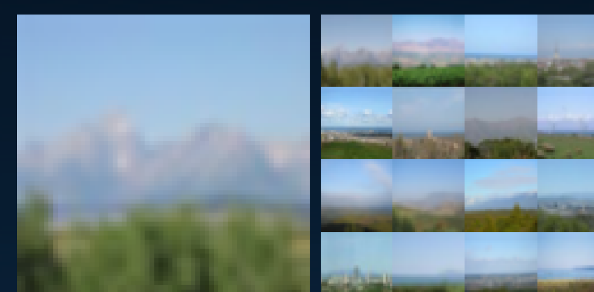

IMAGE RETRIEVAL

- Color-histograms
  - QBIC [Flickner et al., 1995]
  - Pentland et al., 1996
  - …

- SIFT-based approaches
  - Lowe, 1999, 2004
  - Sivic and Zisserman, 2003
  - Chum et al., 2007-10
  - Jegou et al., 2008-10
  - Lazebnik et al., 2009
  - …



- Gist-based “data-driven” approaches
  - Oliva and Torralba, 2006
  - Hays and Efros, 2007
  - Weiss et al., 2007
  - Torralba et al., 2008
  - …



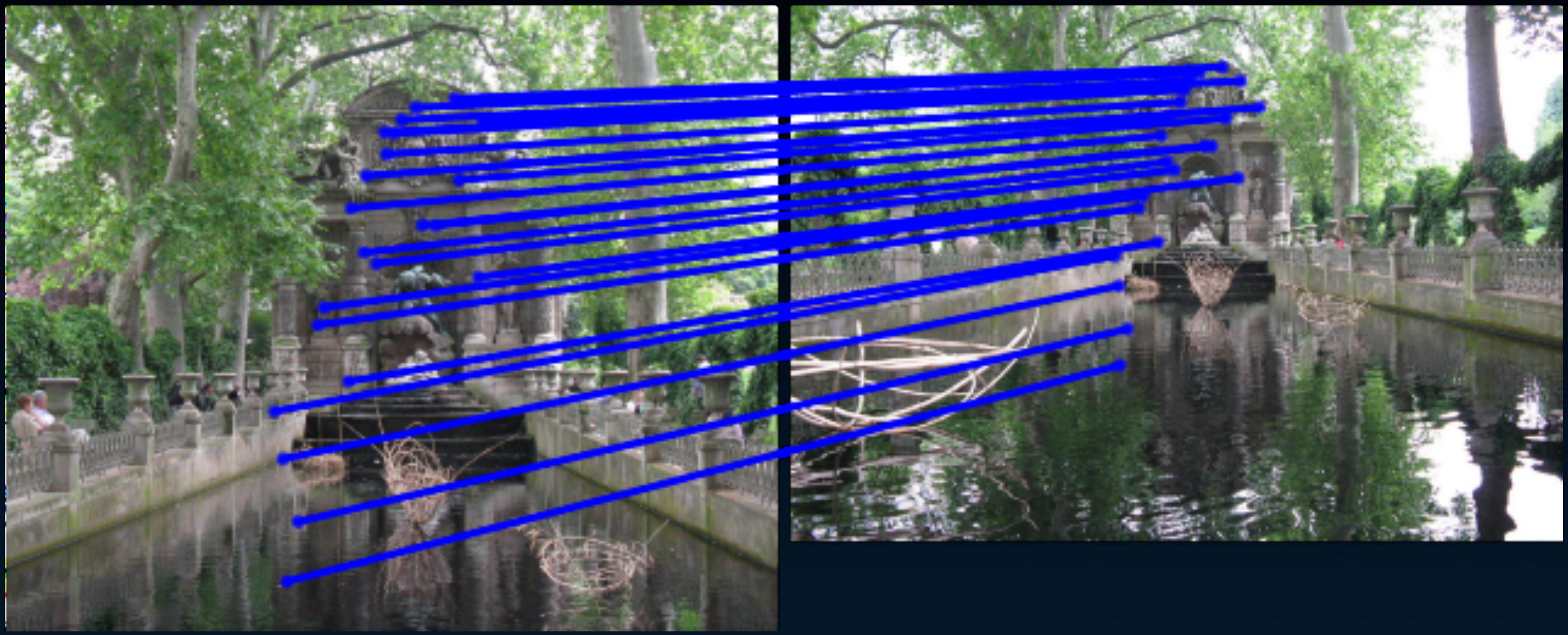
EXAMPLE: SIFT MATCHING

[SIFT: Lowe, 2004]

EXAMPLE: SIFT MATCHING

[SIFT: Lowe, 2004]

EXAMPLE: SIFT MATCHING



[SIFT: Lowe, 2004]

EXAMPLE: SIFT MATCHING

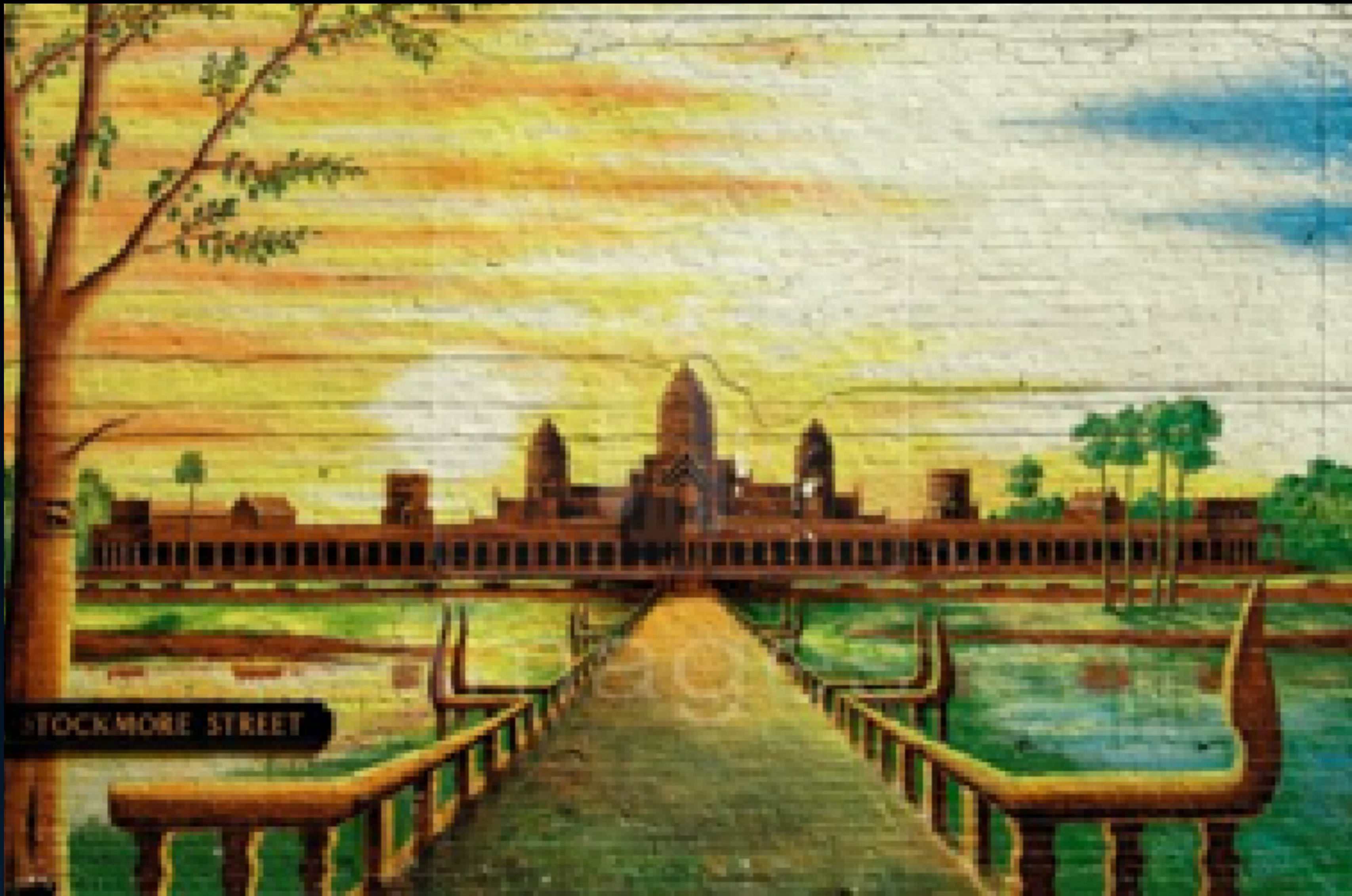
[SIFT: Lowe, 2004]

EXAMPLE: SIFT MATCHING

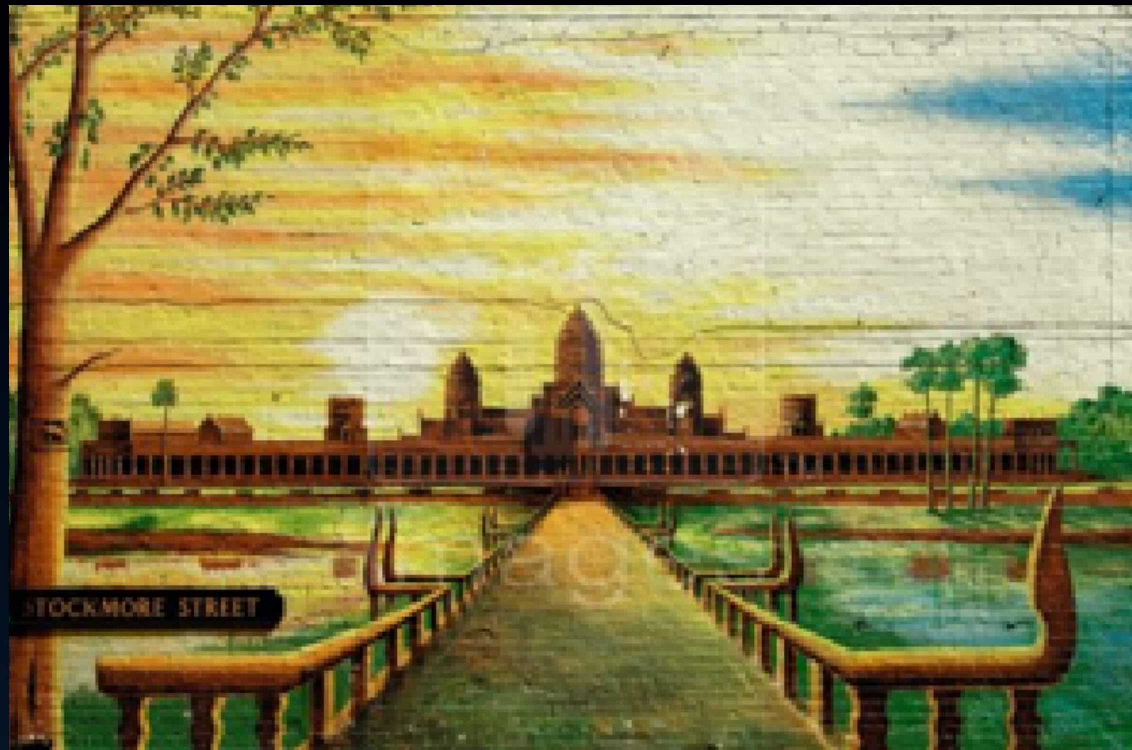
[SIFT: Lowe, 2004]

EXAMPLE: SIFT MATCHING

[SIFT: Lowe, 2004]

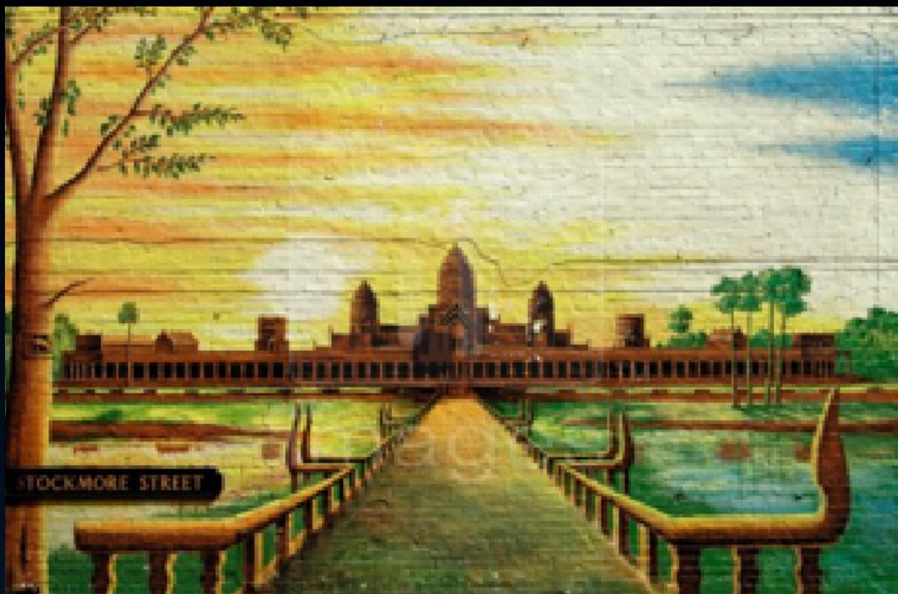

STOCKMORE STREET

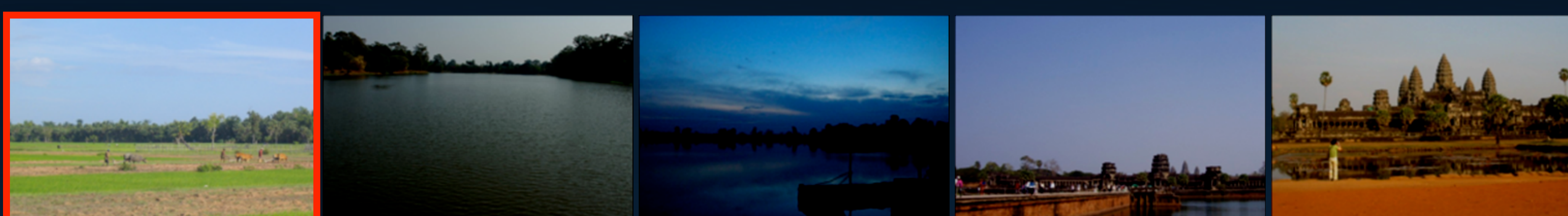
Input Query



Top GIST Matches

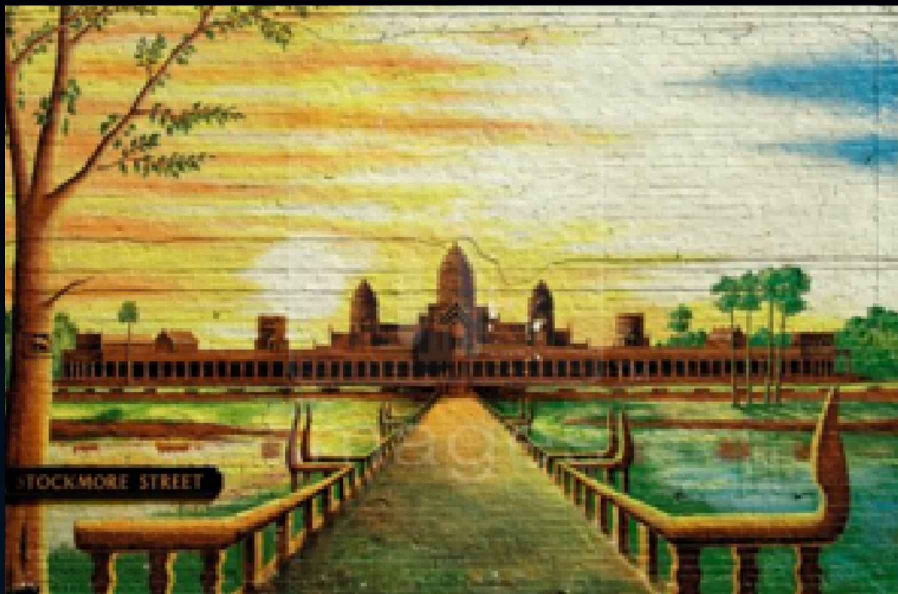
Input Query

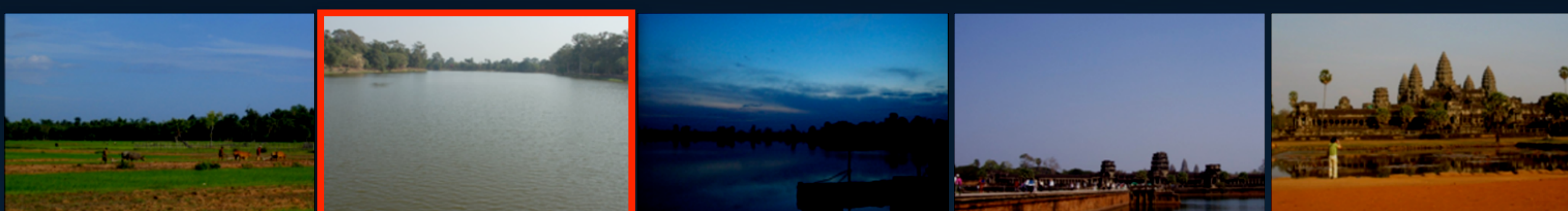




Top GIST Matches

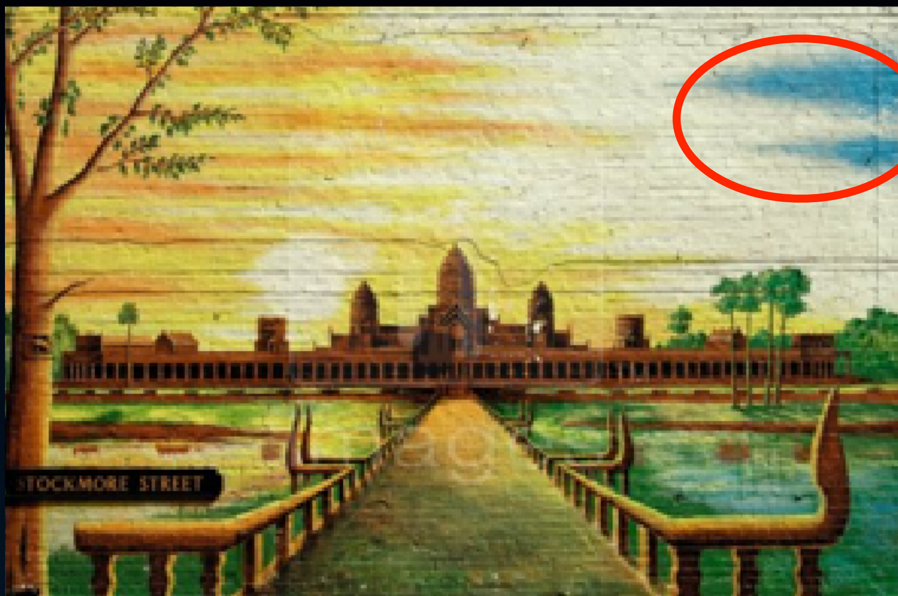
Input Query

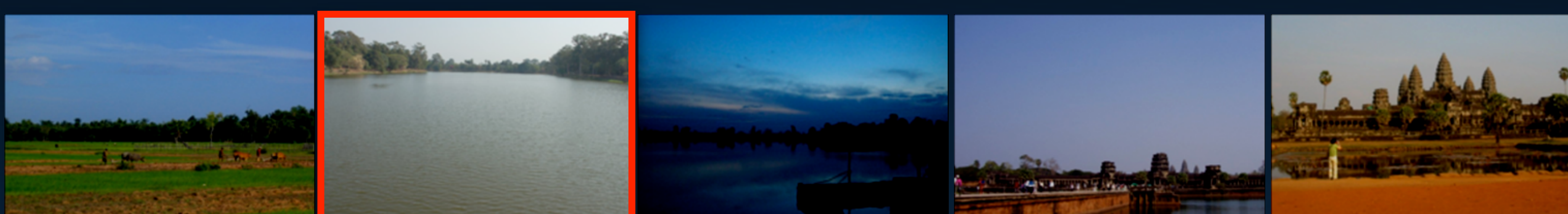




Top GIST Matches

Input Query

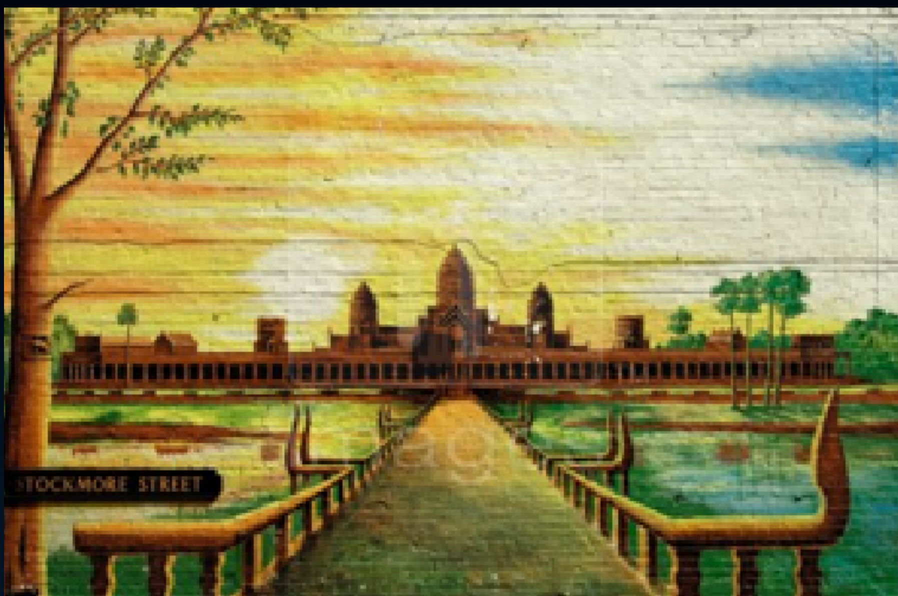




Top GIST Matches

IMPORTANT PARTS?

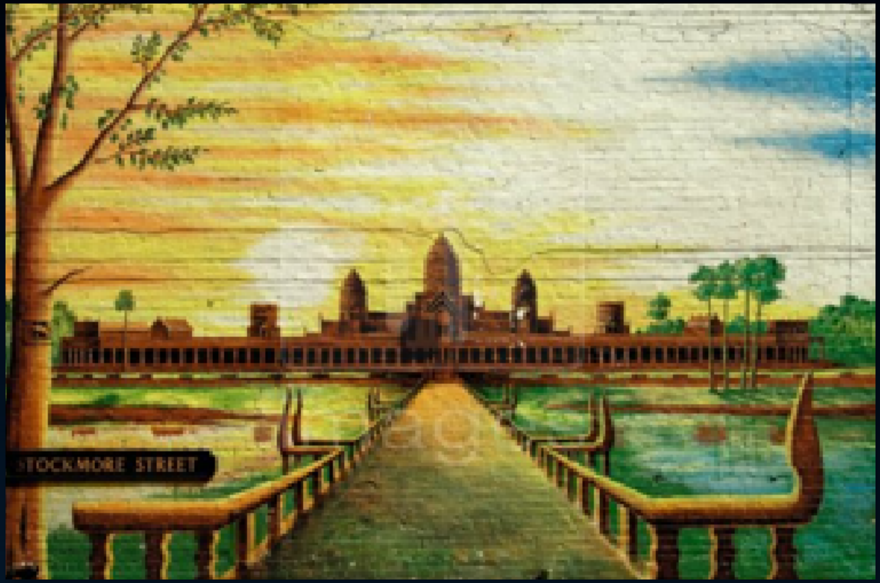
Input Query

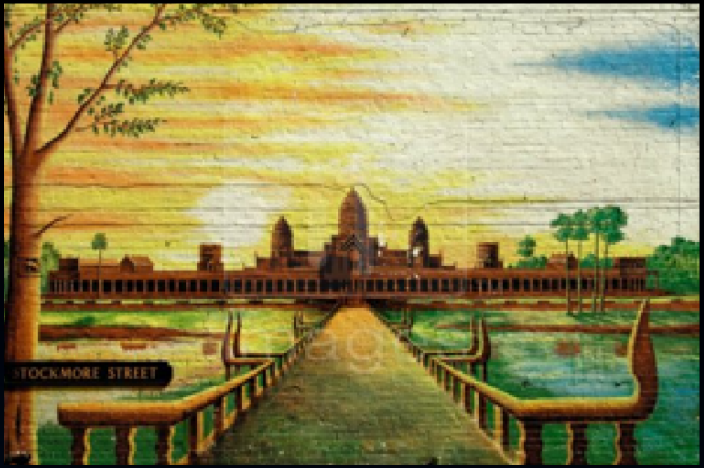


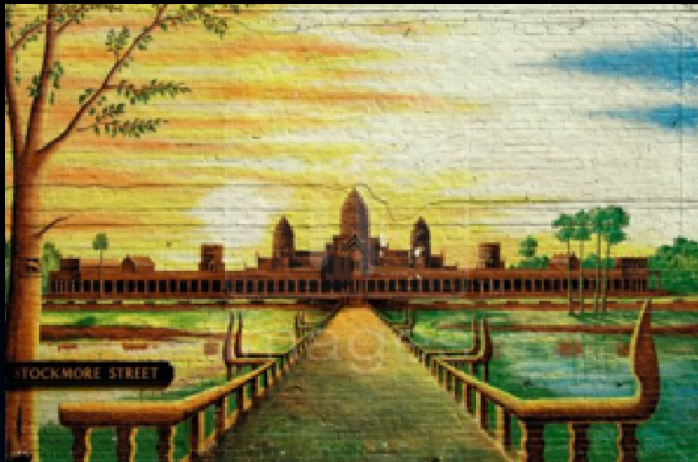
Important Parts

Our Top Matches

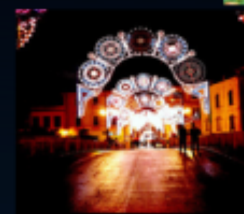
Input Query




STOCKMORE STREET

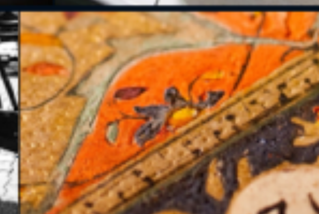

STOCKMORE STREET





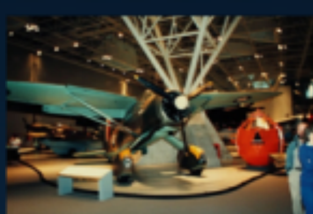


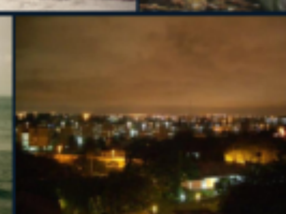




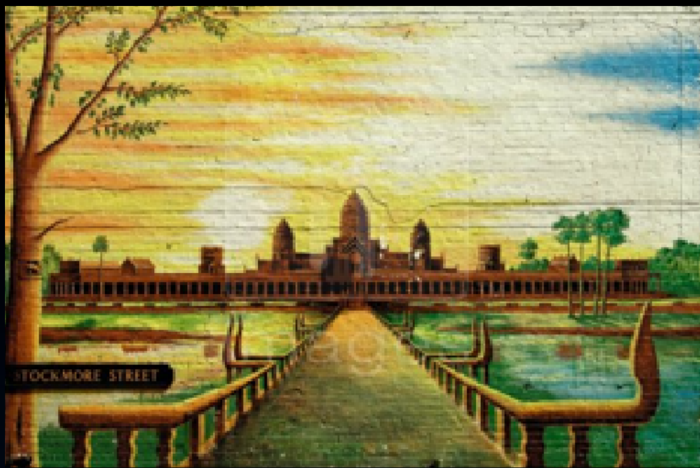


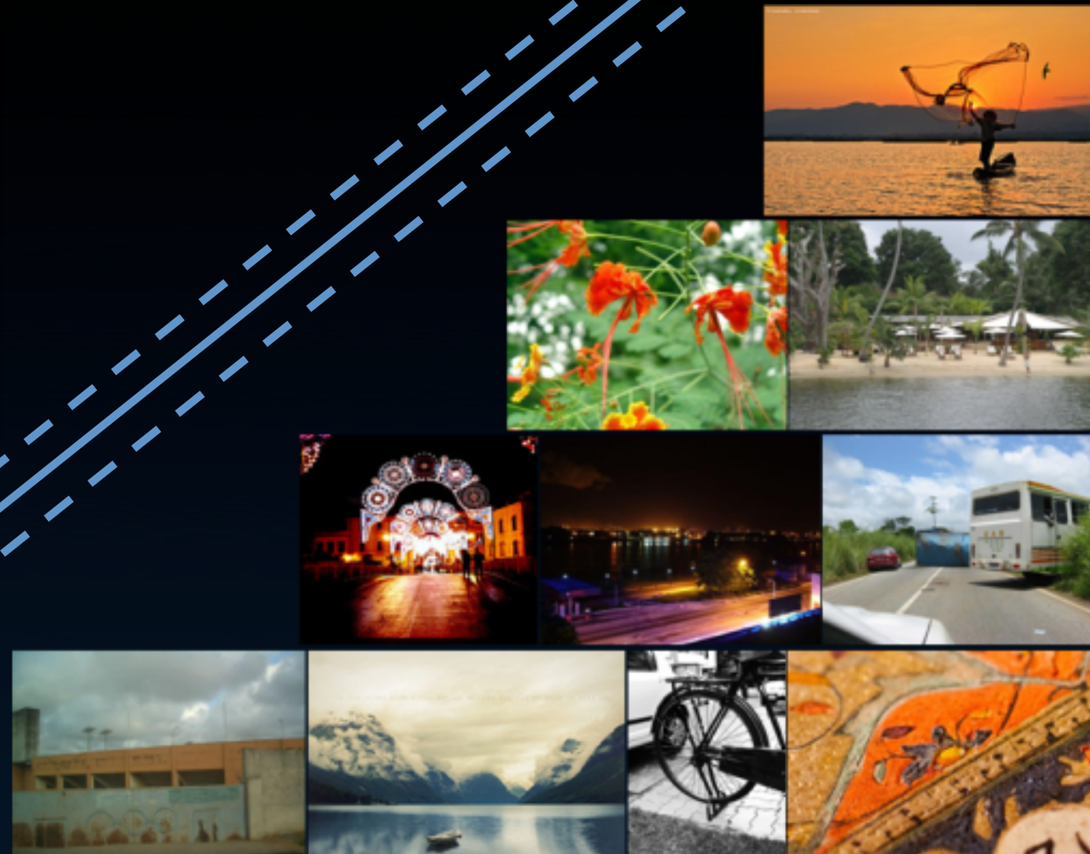


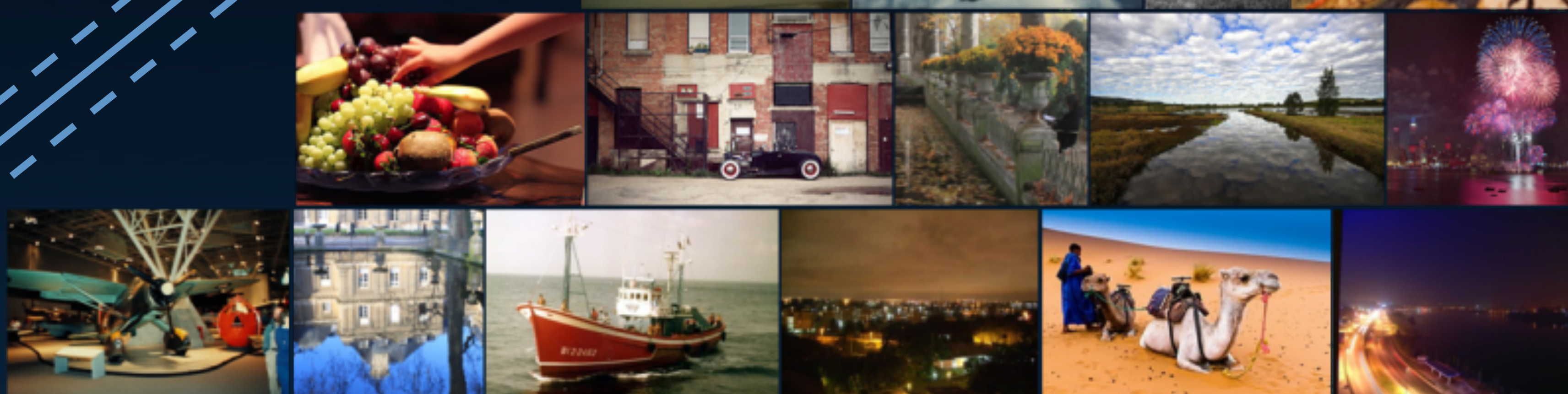


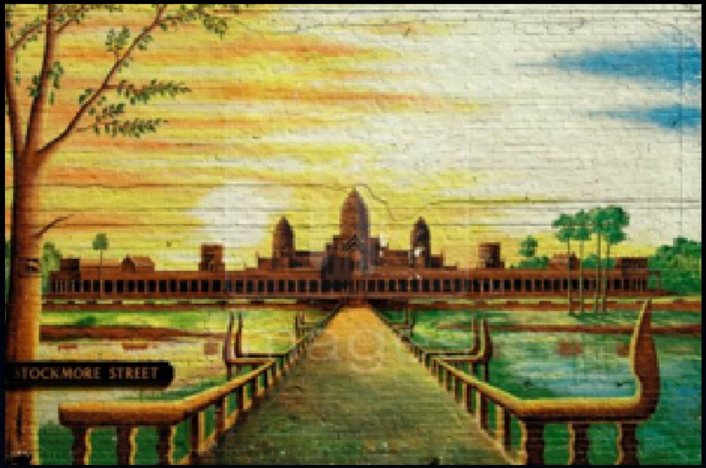



“Data-driven Uniqueness”


STOCKMORE STREET

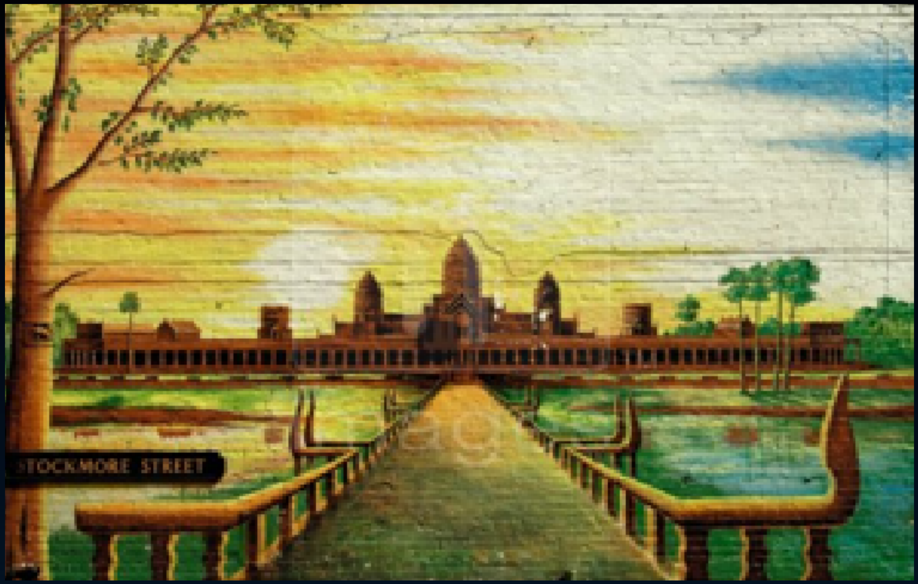





STOCKMORE STREET

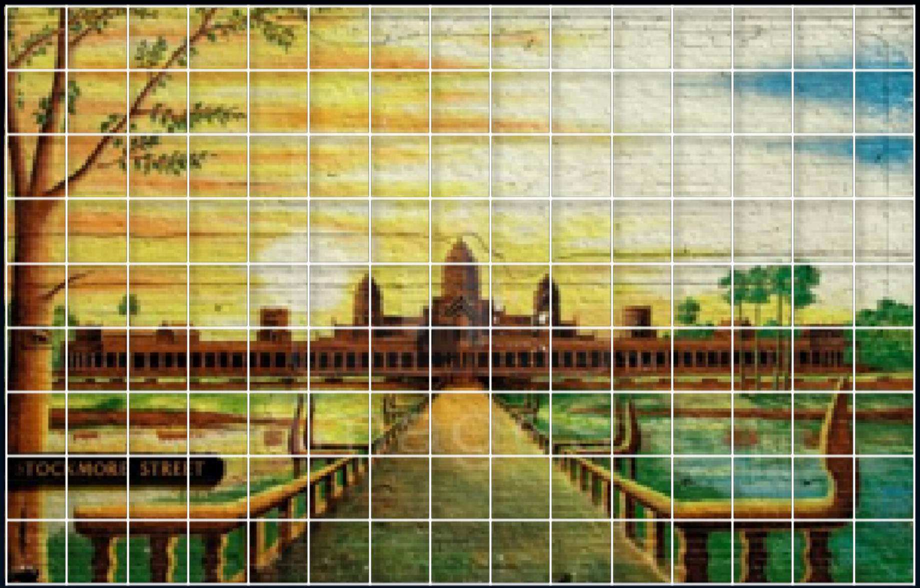

STOCKMORE STREET

FEATURE REPRESENTATION
HISTOGRAM OF ORIENTED GRADIENTS (HOG)



[Dalal and Triggs, CVPR, 2005]

FEATURE REPRESENTATION
HISTOGRAM OF ORIENTED GRADIENTS (HOG)



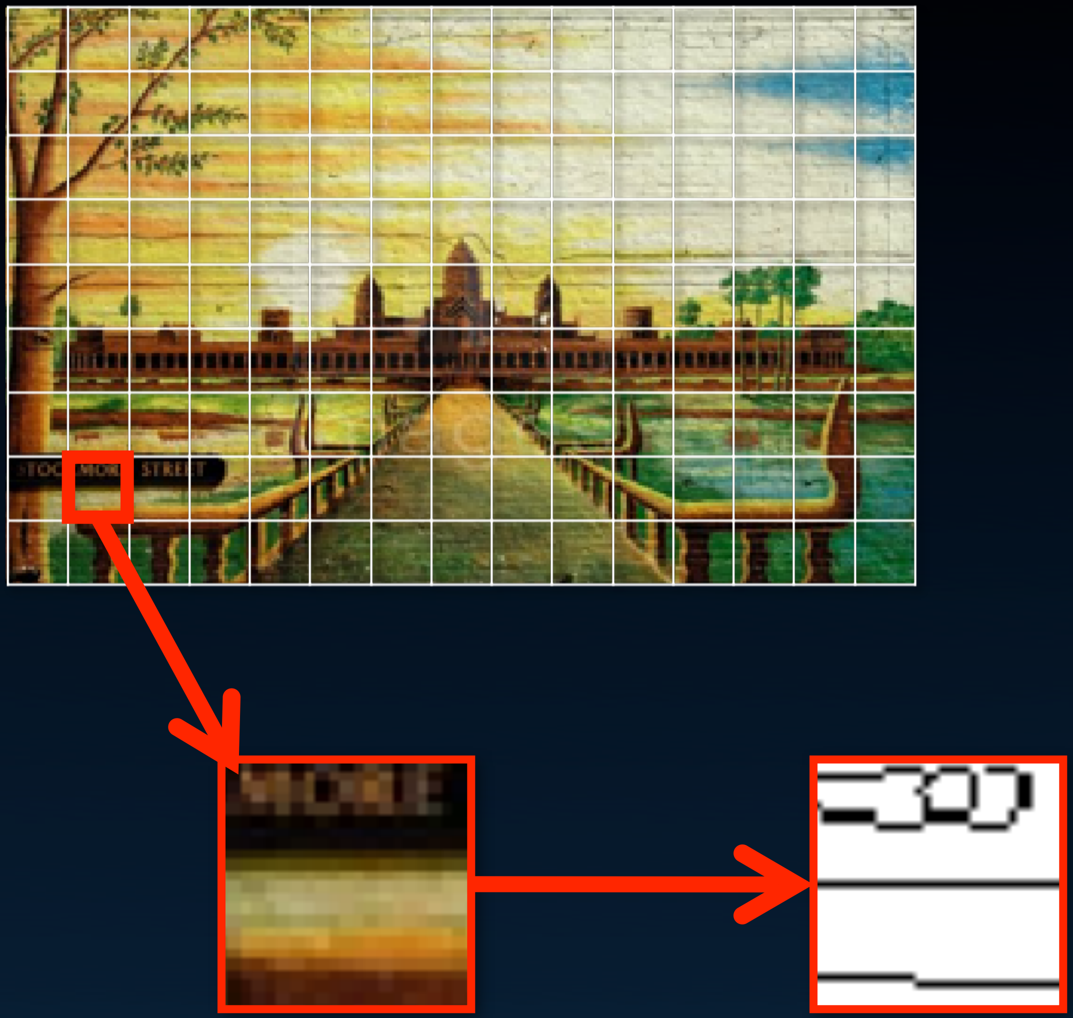
[Dalal and Triggs, CVPR, 2005]

FEATURE REPRESENTATION
HISTOGRAM OF ORIENTED GRADIENTS (HOG)

[Dalal and Triggs, CVPR, 2005]

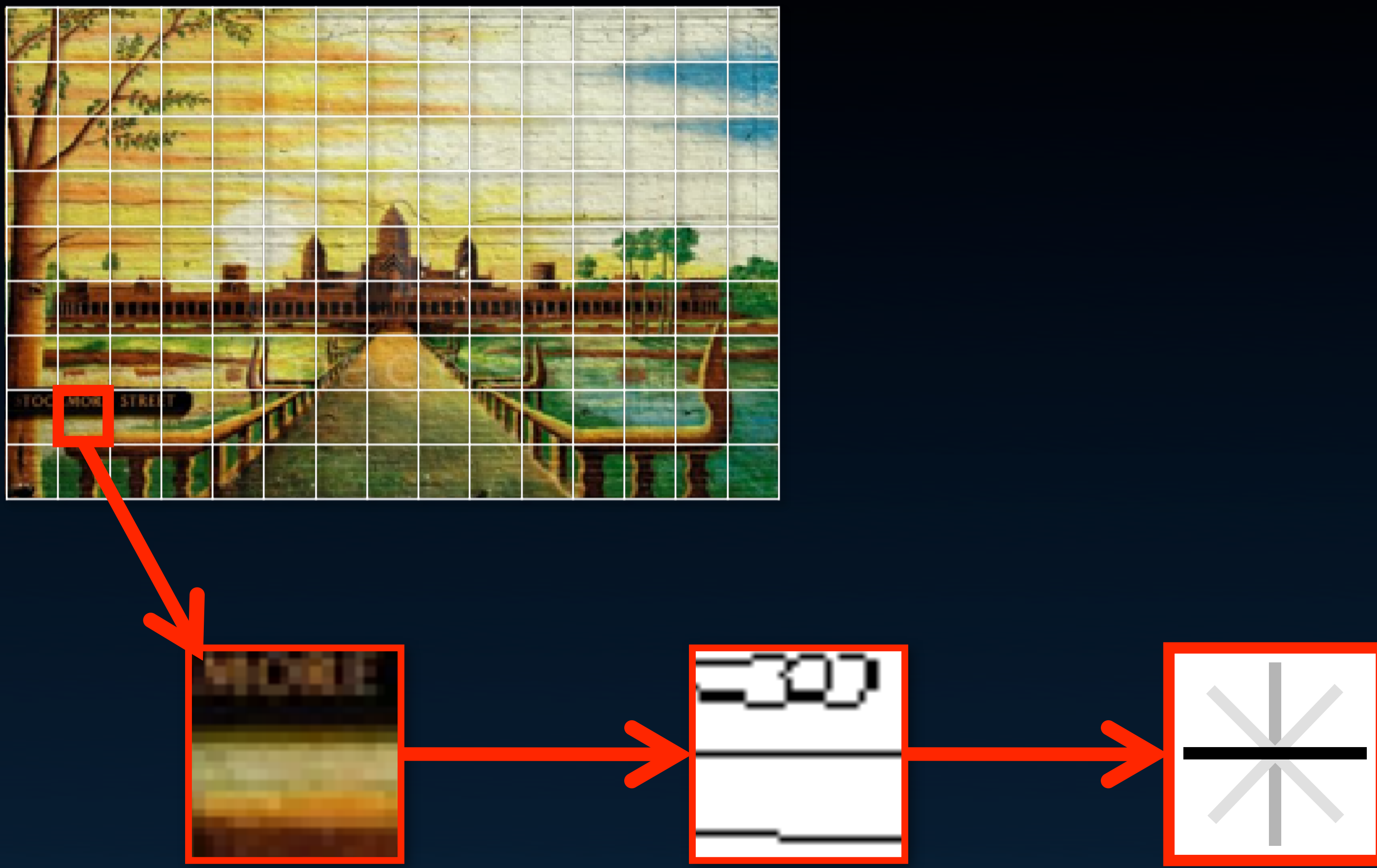
FEATURE REPRESENTATION

HISTOGRAM OF ORIENTED GRADIENTS (HOG)



[Dalal and Triggs, CVPR, 2005]

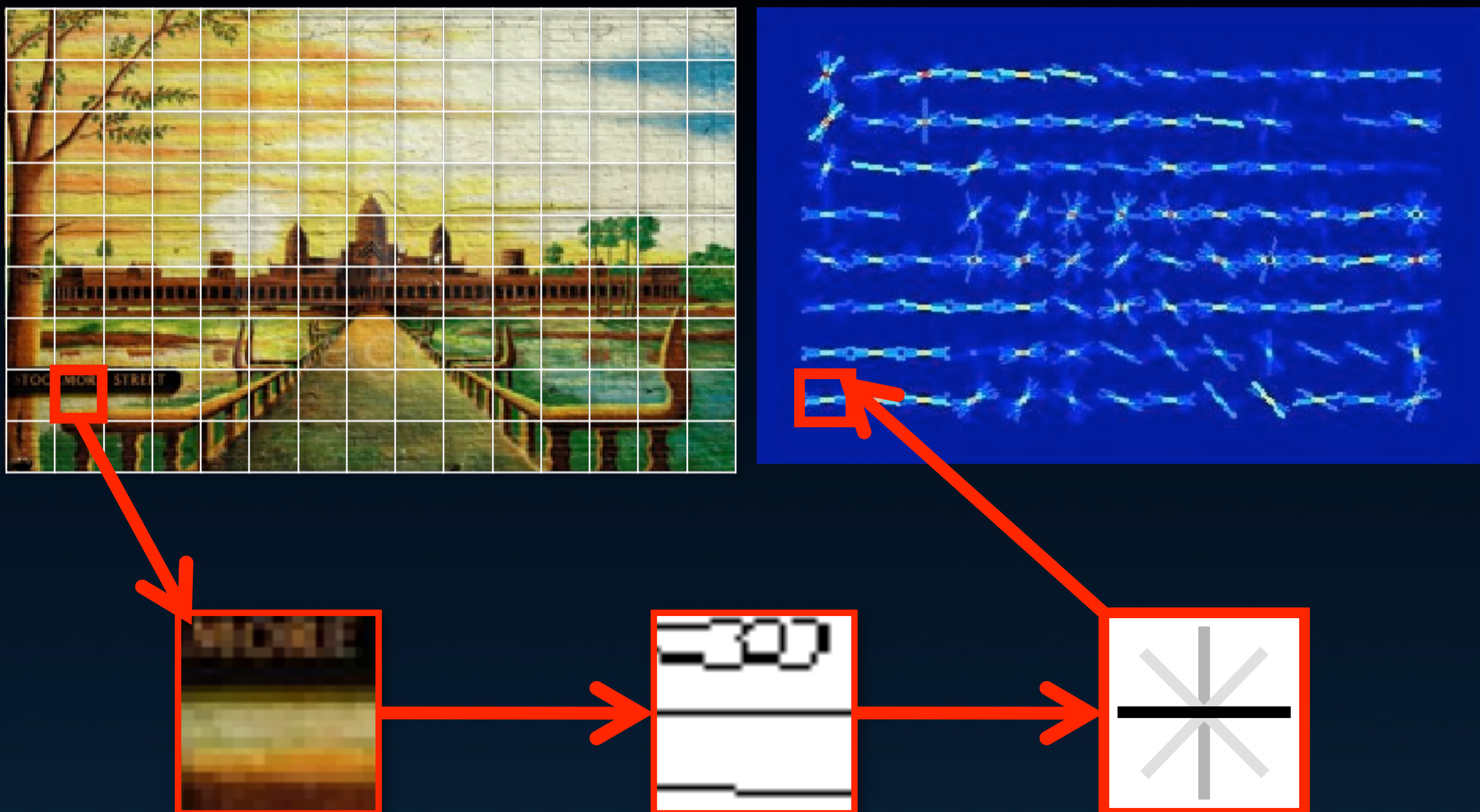
FEATURE REPRESENTATION
HISTOGRAM OF ORIENTED GRADIENTS (HOG)



[Dalal and Triggs, CVPR, 2005]

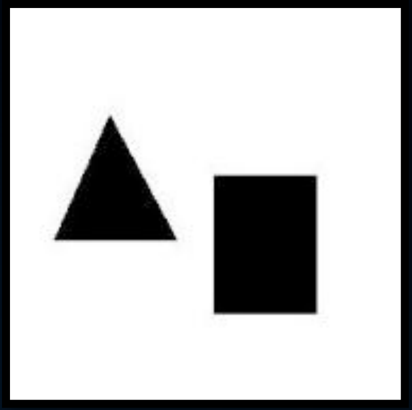
FEATURE REPRESENTATION

HISTOGRAM OF ORIENTED GRADIENTS (HOG)

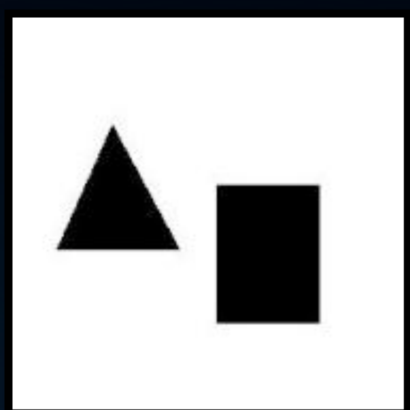


[Dalal and Triggs, CVPR, 2005]

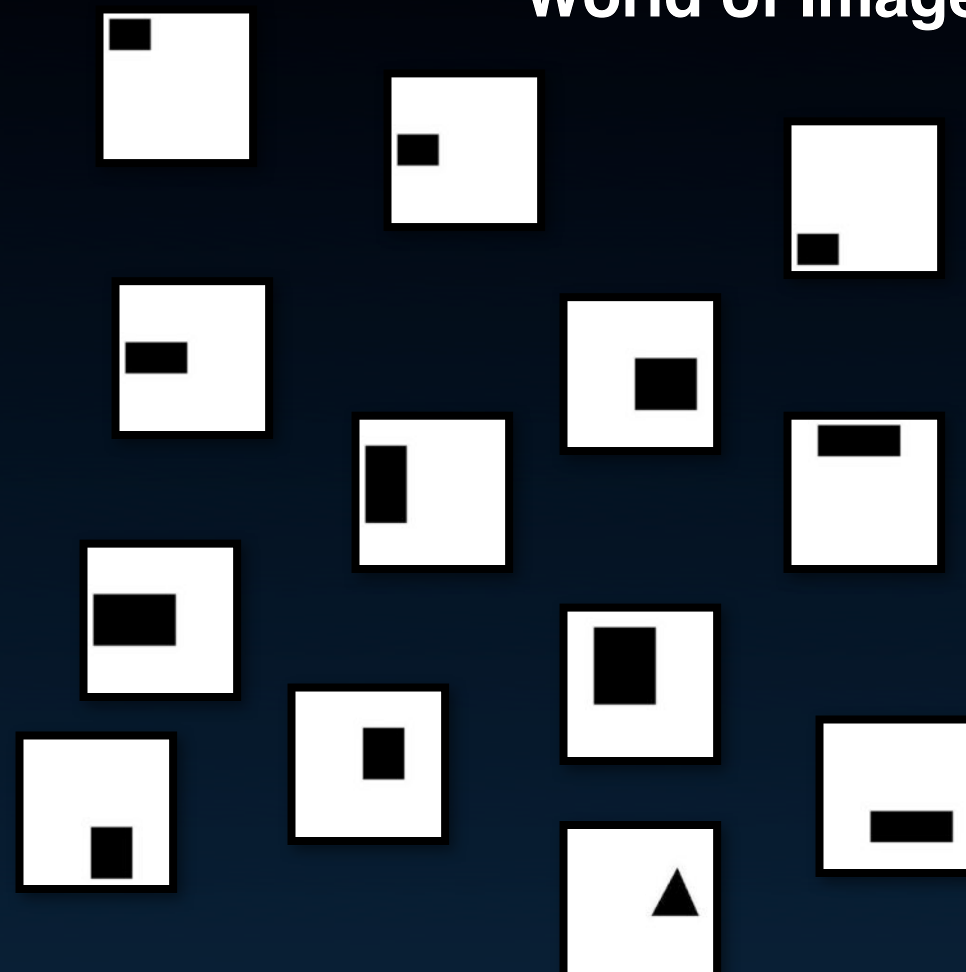
WHAT IS UNIQUE?



WHAT IS UNIQUE GIVEN THIS WORLD?



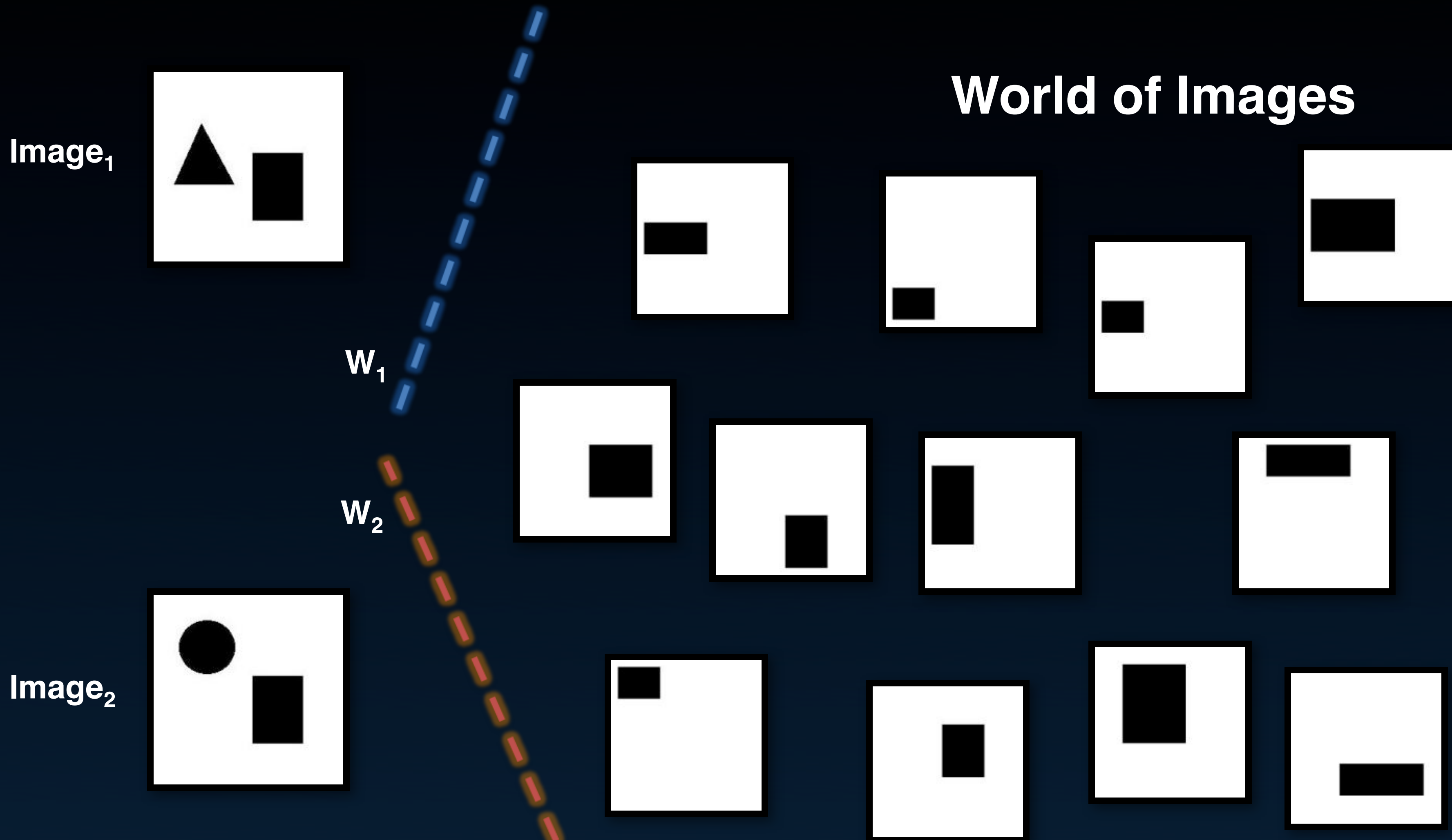
World of Images



EXEMPLAR-SVM

Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros.
Ensemble of Exemplar-SVMs for Object Detection and Beyond. In ICCV, 2011.

EXEMPLAR-SVM



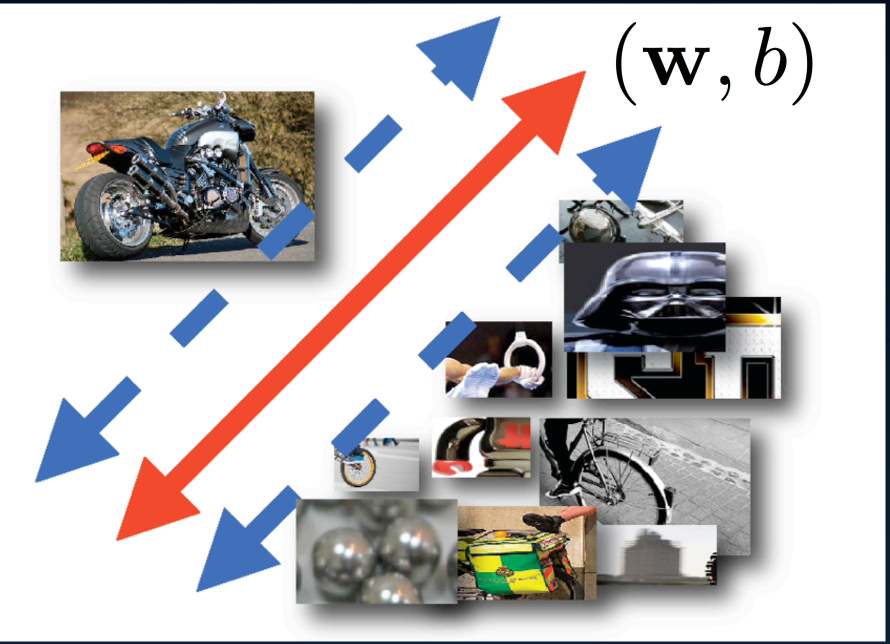
Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros.
Ensemble of Exemplar-SVMs for Object Detection and Beyond. In ICCV, 2011.

EXEMPLAR-SVM

Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0)

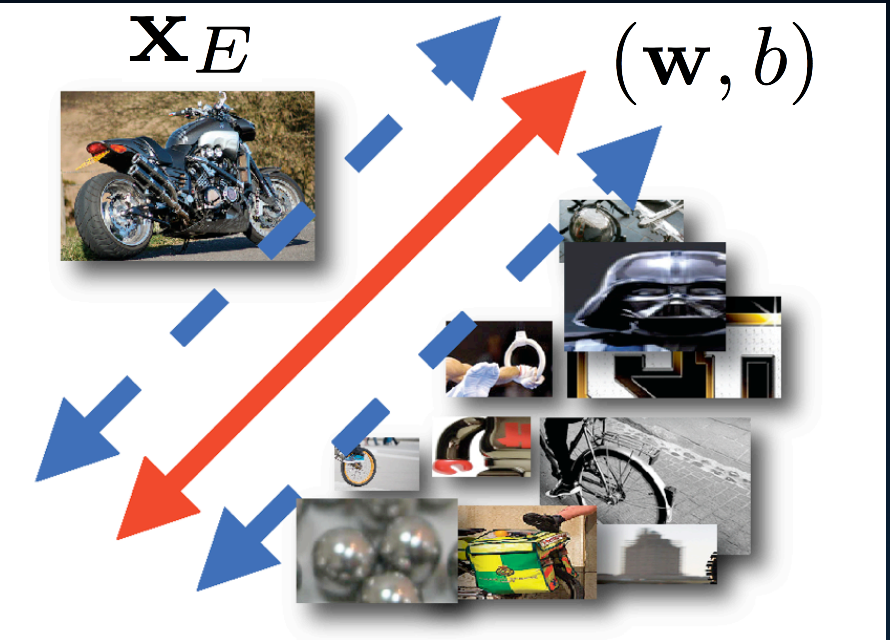


EXEMPLAR-SVM

Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0)



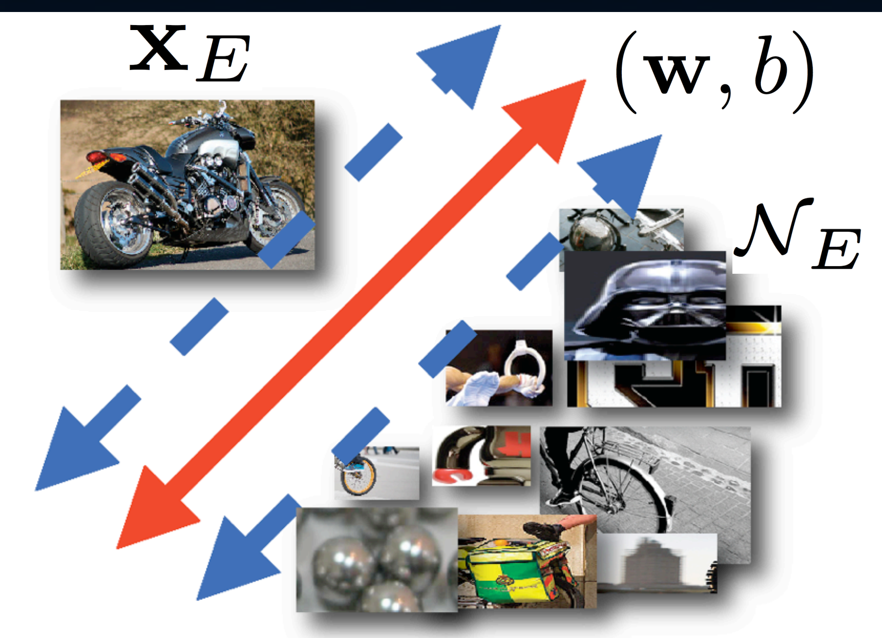
$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,100 features)

EXEMPLAR-SVM

Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0)

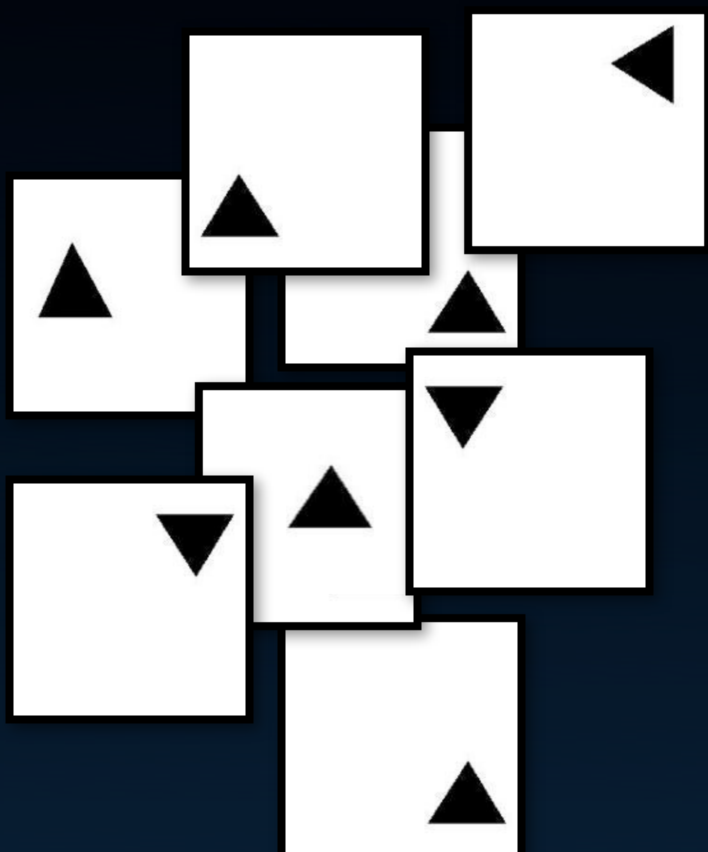


$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,100 features)

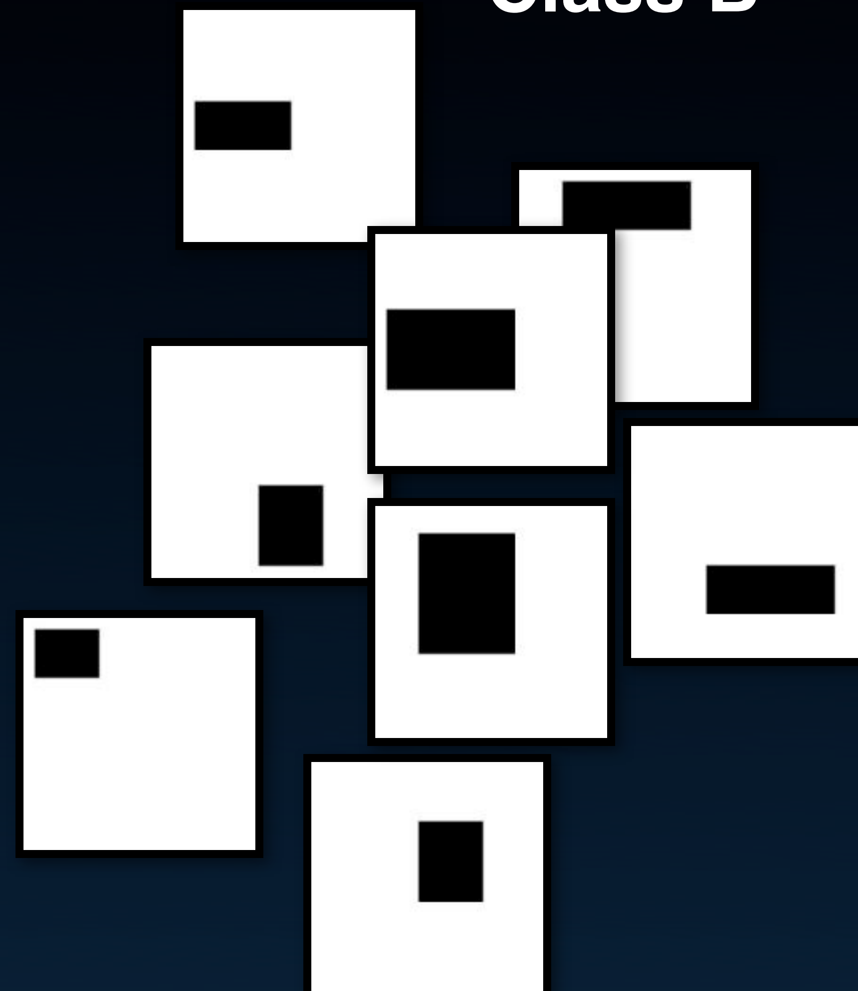
$\mathcal{N}_E$ Image windows from negative images (~2,000 images x ~10,000 windows/image =~20M negatives)

SUPPORT VECTOR MACHINE (SVM)

Class A

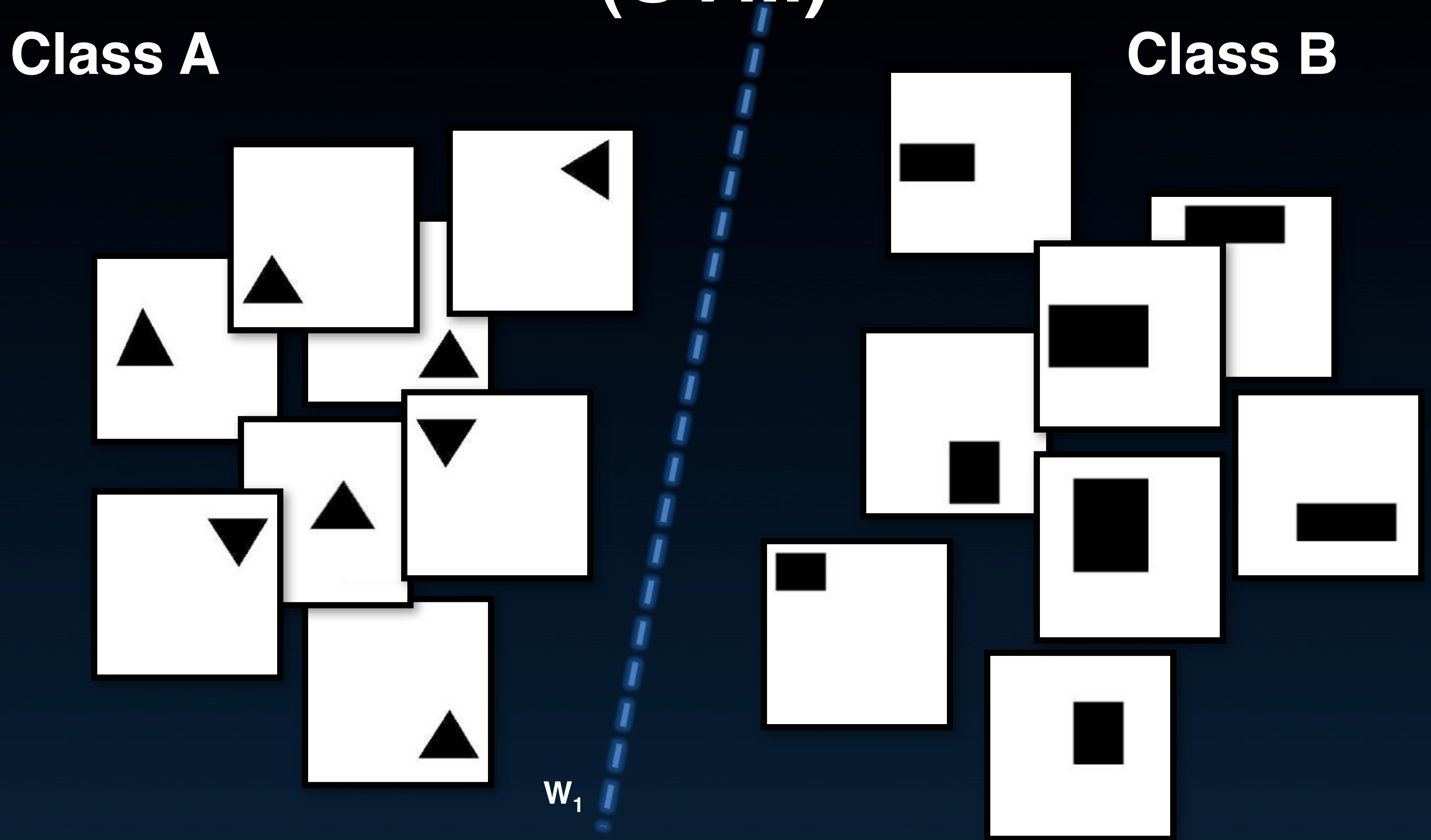


Class B



[Cortes and Vapnik, Machine Learning, 1995]

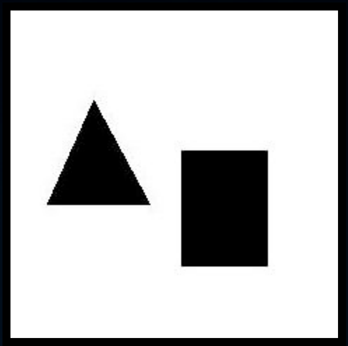
SUPPORT VECTOR MACHINE (SVM)

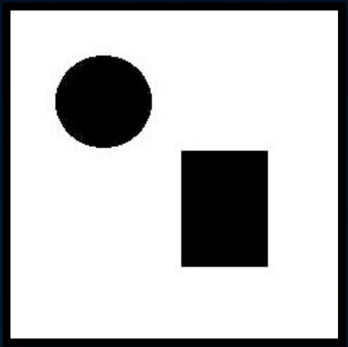


[Cortes and Vapnik, Machine Learning, 1995]

VISUALIZING UNIQUENESS

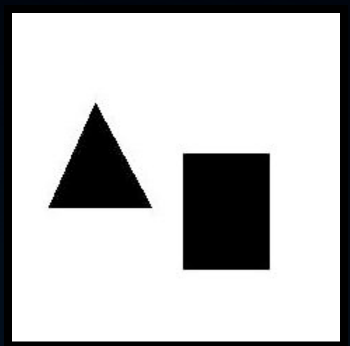
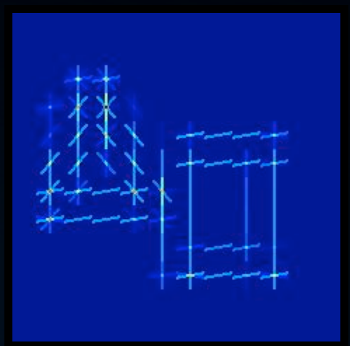
Query

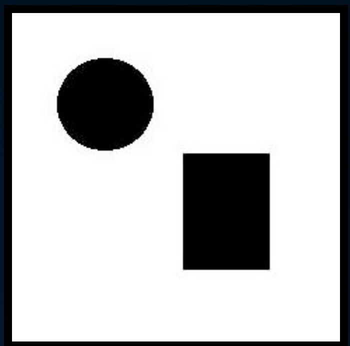
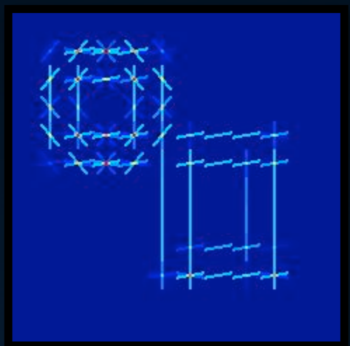




VISUALIZING UNIQUENESS

Query **Before**

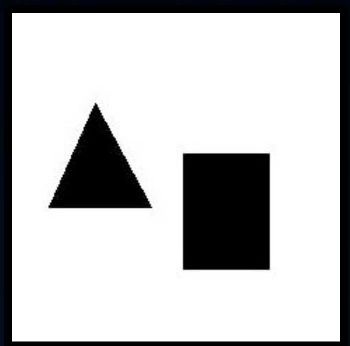



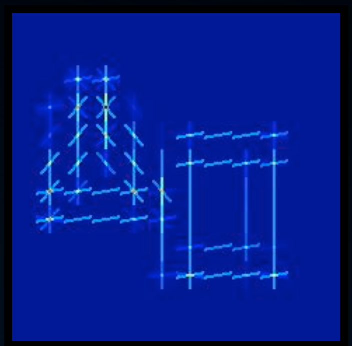



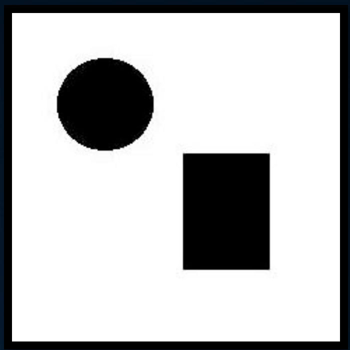
VISUALIZING UNIQUENESS

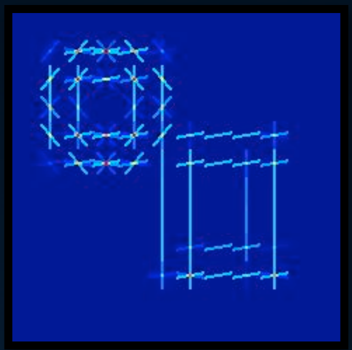
Query

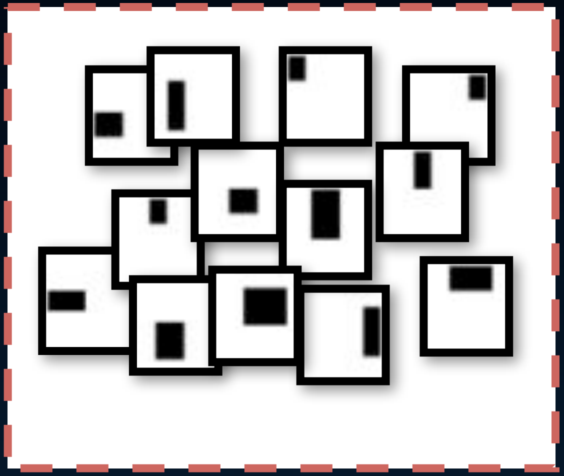
Before





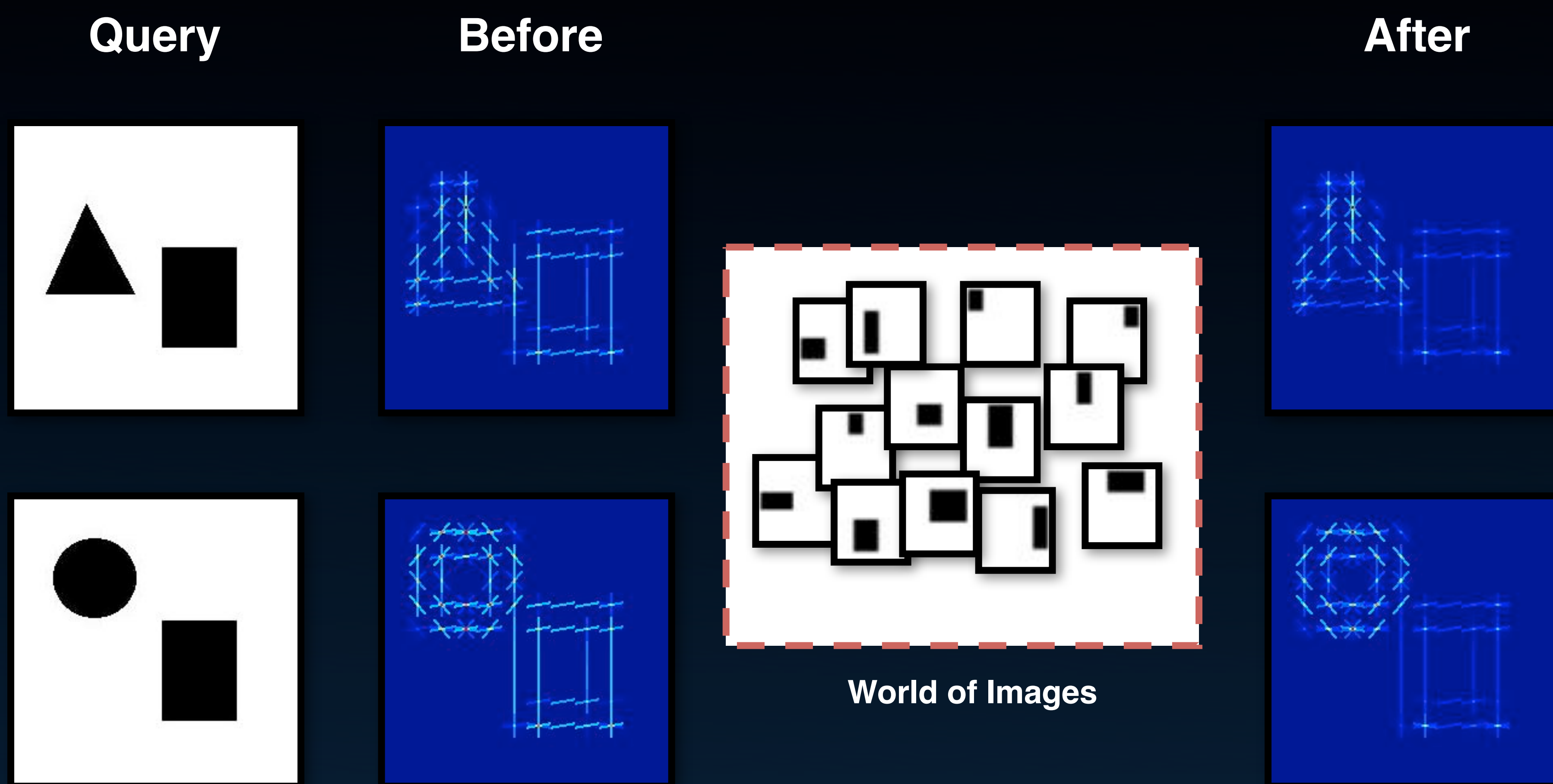




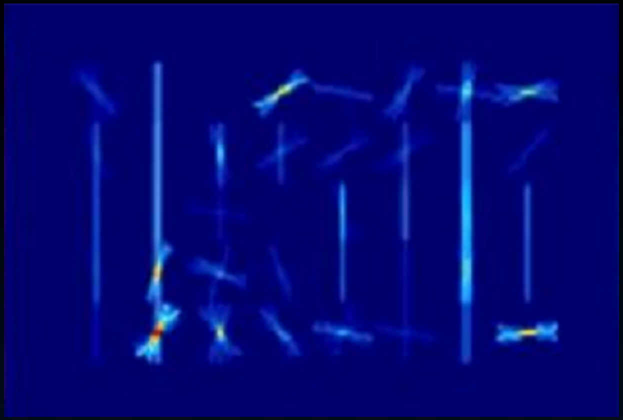


World of Images

VISUALIZING UNIQUENESS



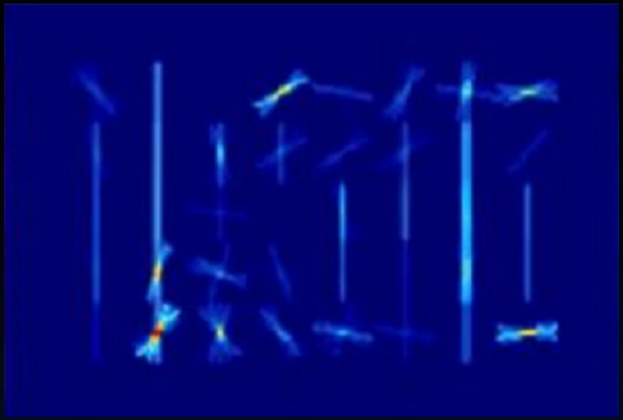
Input Query



HOG

Input Query

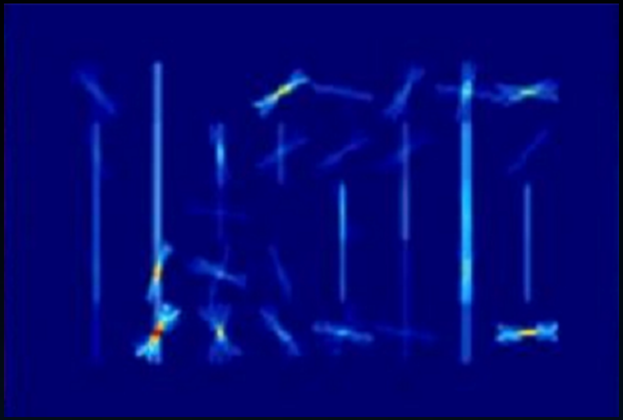
Input Query



HOG

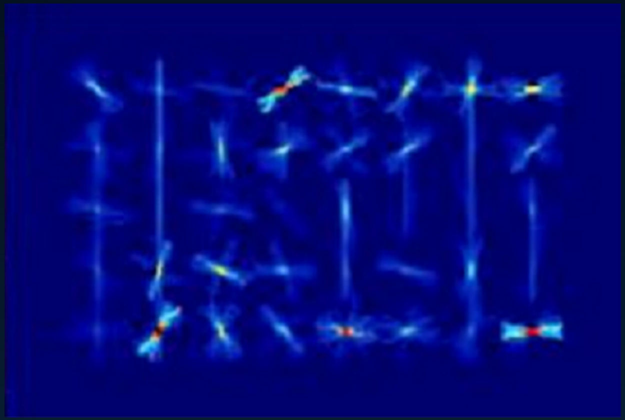
Top Match

Input Query



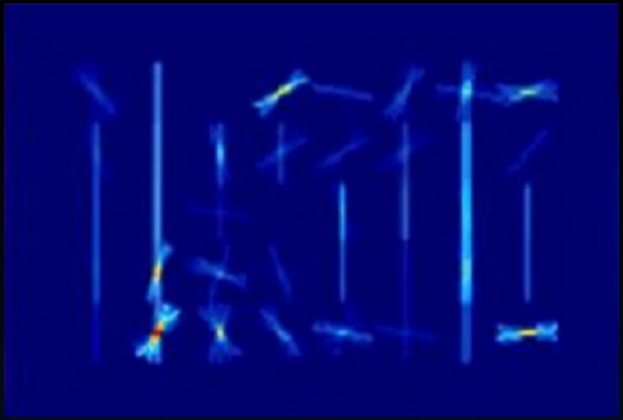
HOG

Top Match



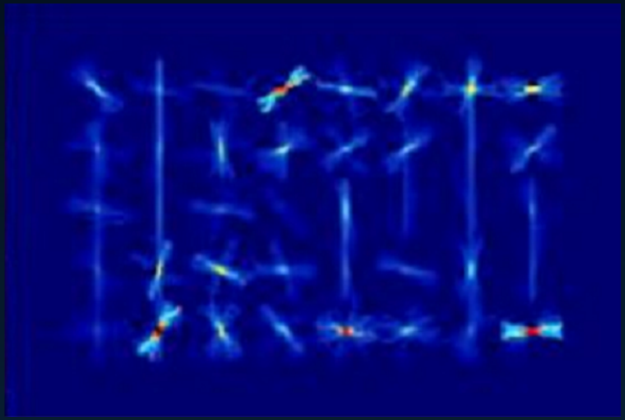
Learnt Weights

Input Query



HOG

Top Match



Learnt Weights

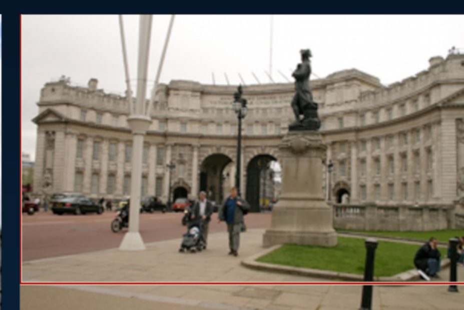
Top Match

SEARCH USING IMAGES

Input Query

SEARCH USING IMAGES

Input Query



Our Top Matches

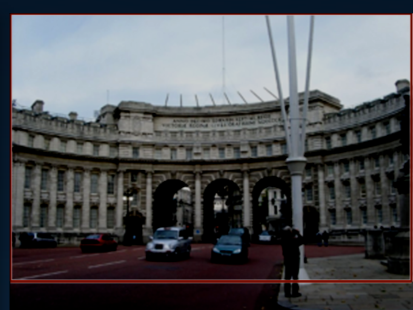
SEARCH USING IMAGES

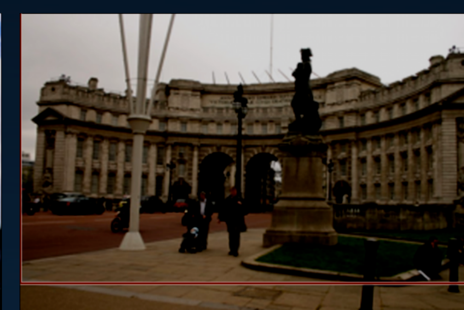
Input Query

Our Top Matches

SEARCH USING IMAGES

Input Query





Our Top Matches

SEARCH USING PAINTINGS

Input Painting

SEARCH USING PAINTINGS

Input Painting

GIST

SEARCH USING PAINTINGS

Input Painting

GIST

Bag-of-Words

SEARCH USING PAINTINGS

Input Painting

GIST

Bag-of-Words

Tiny Images

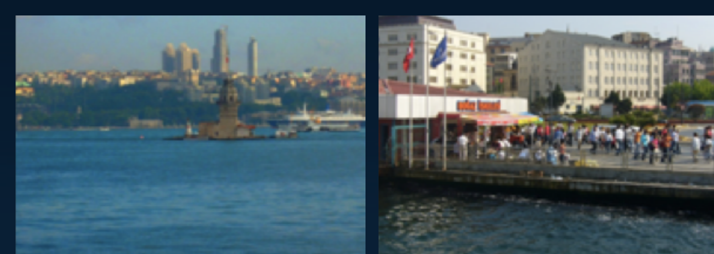
SEARCH USING PAINTINGS

Input Painting

GIST

Bag-of-Words

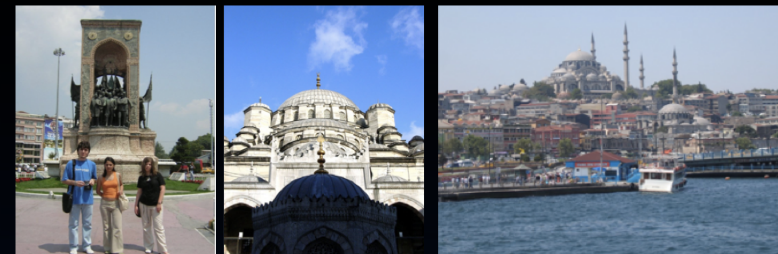
Tiny Images



HOG

SEARCH USING PAINTINGS

Input Painting

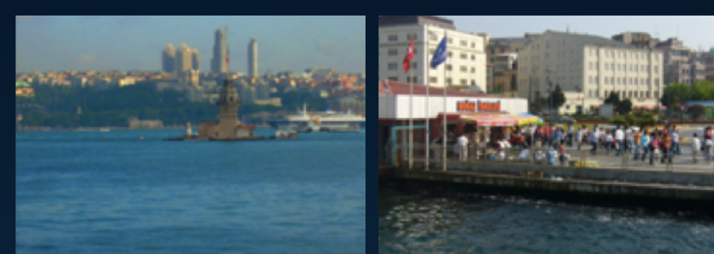


GIST

Bag-of-Words

Tiny Images

Our Approach



HOG

SEARCH USING PAINTINGS

Input Painting

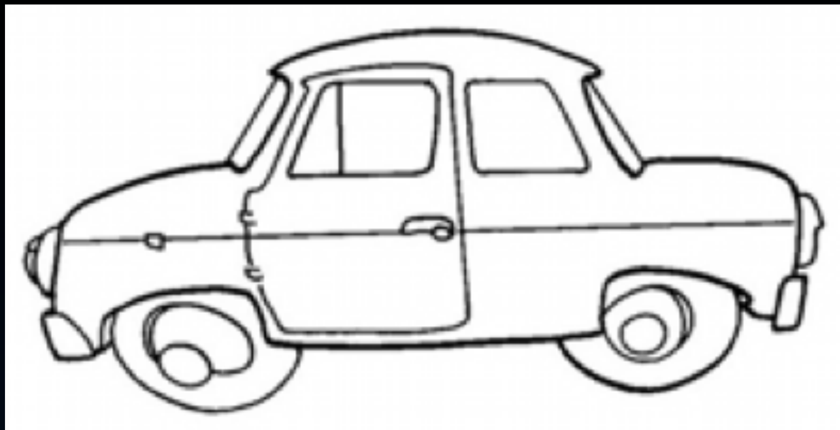
Our Top Matches

SEARCH USING PAINTINGS

Input Painting

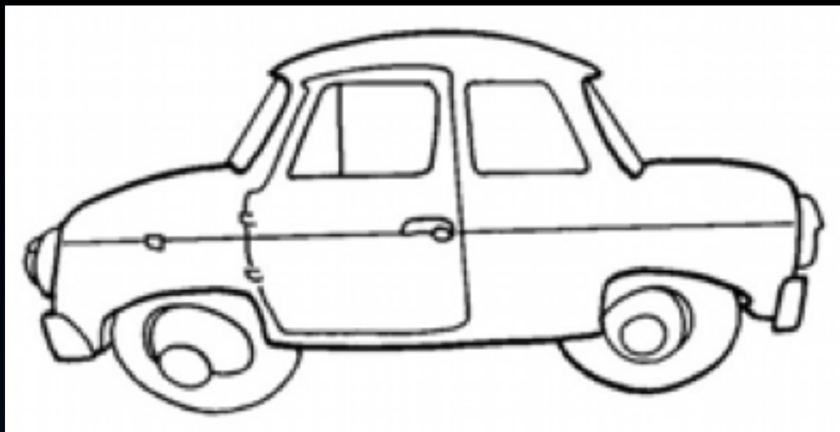
Our Top Matches

SEARCH USING SKETCHES



Input Sketch

SEARCH USING SKETCHES



Input Sketch

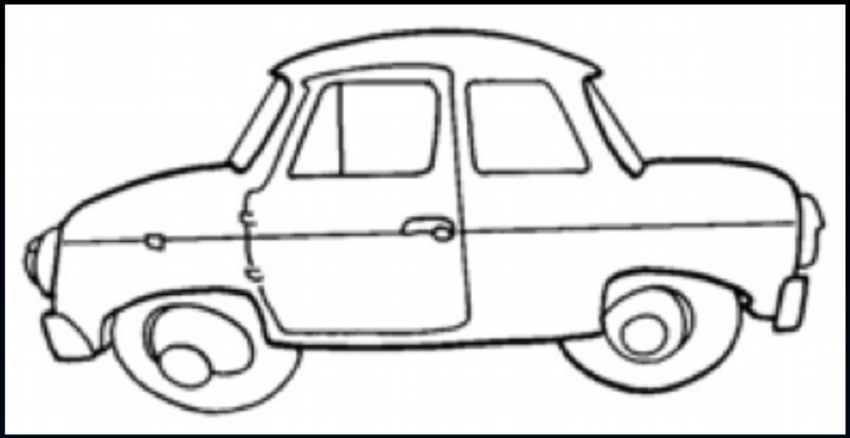
Tiny Images

GIST

Bag-of-Words

HOG

SEARCH USING SKETCHES



Input Sketch

Our Approach

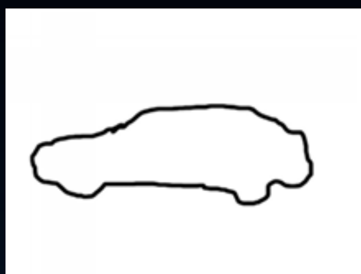
Tiny Images

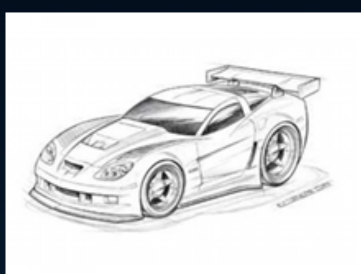
GIST

Bag-of-Words

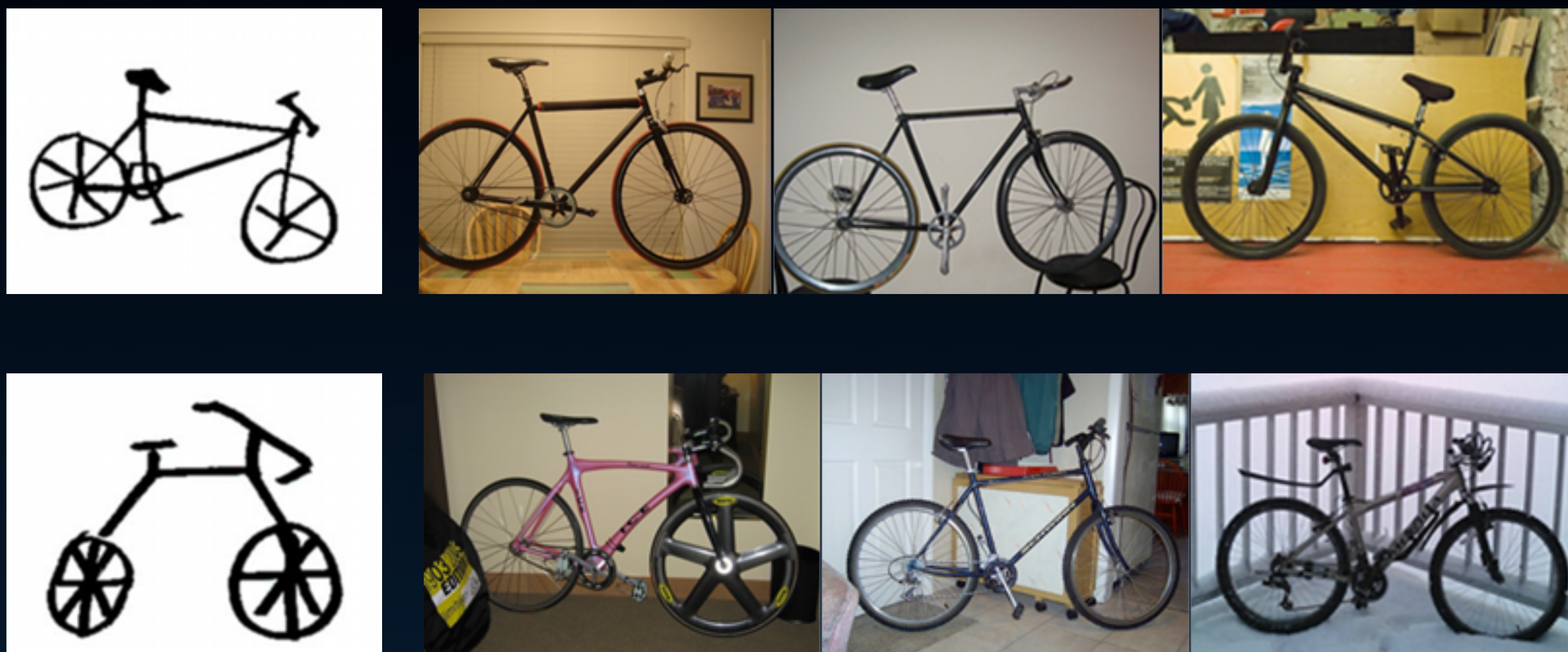
HOG

SEARCH USING SKETCHES

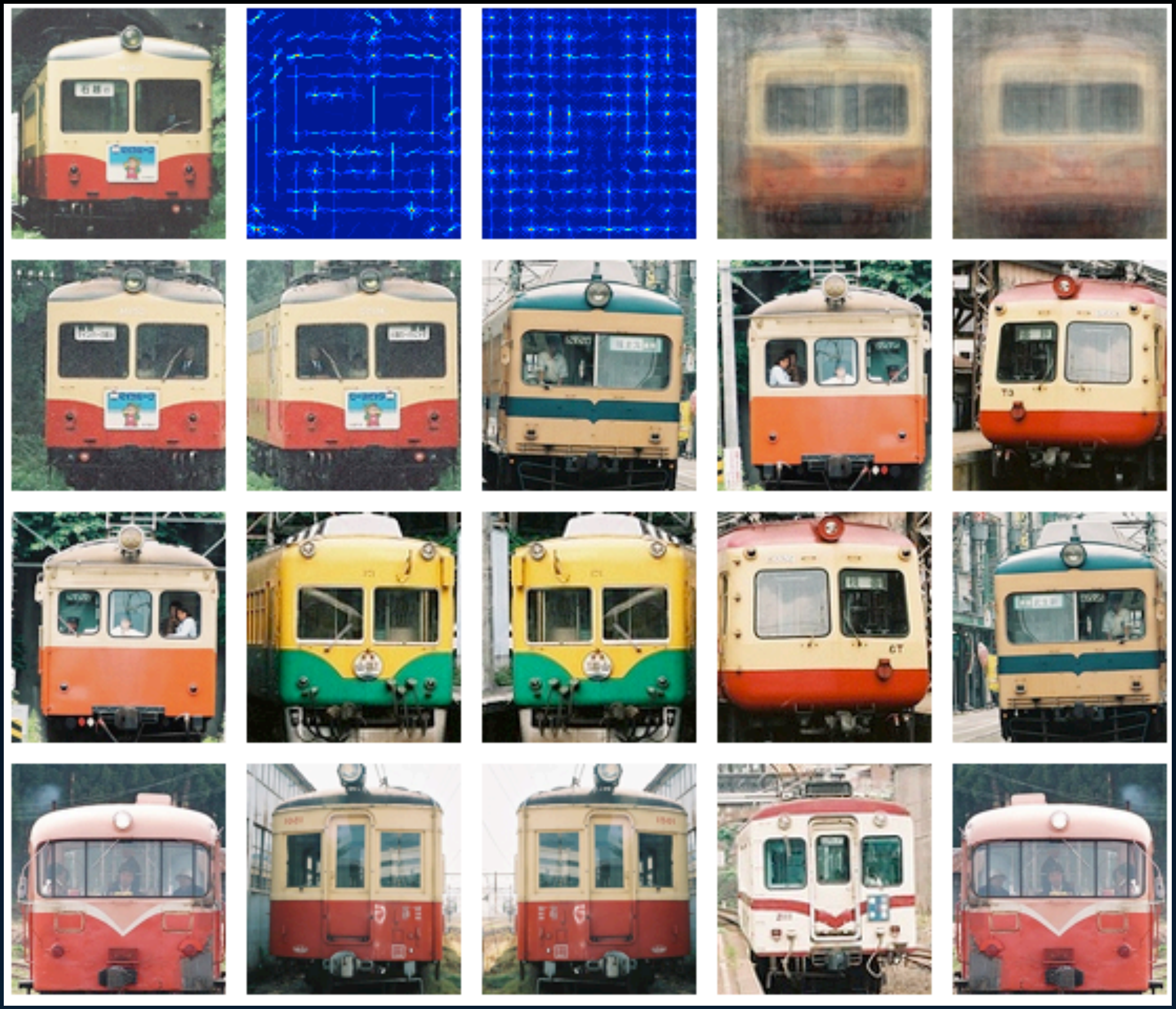




SEARCH USING SKETCHES

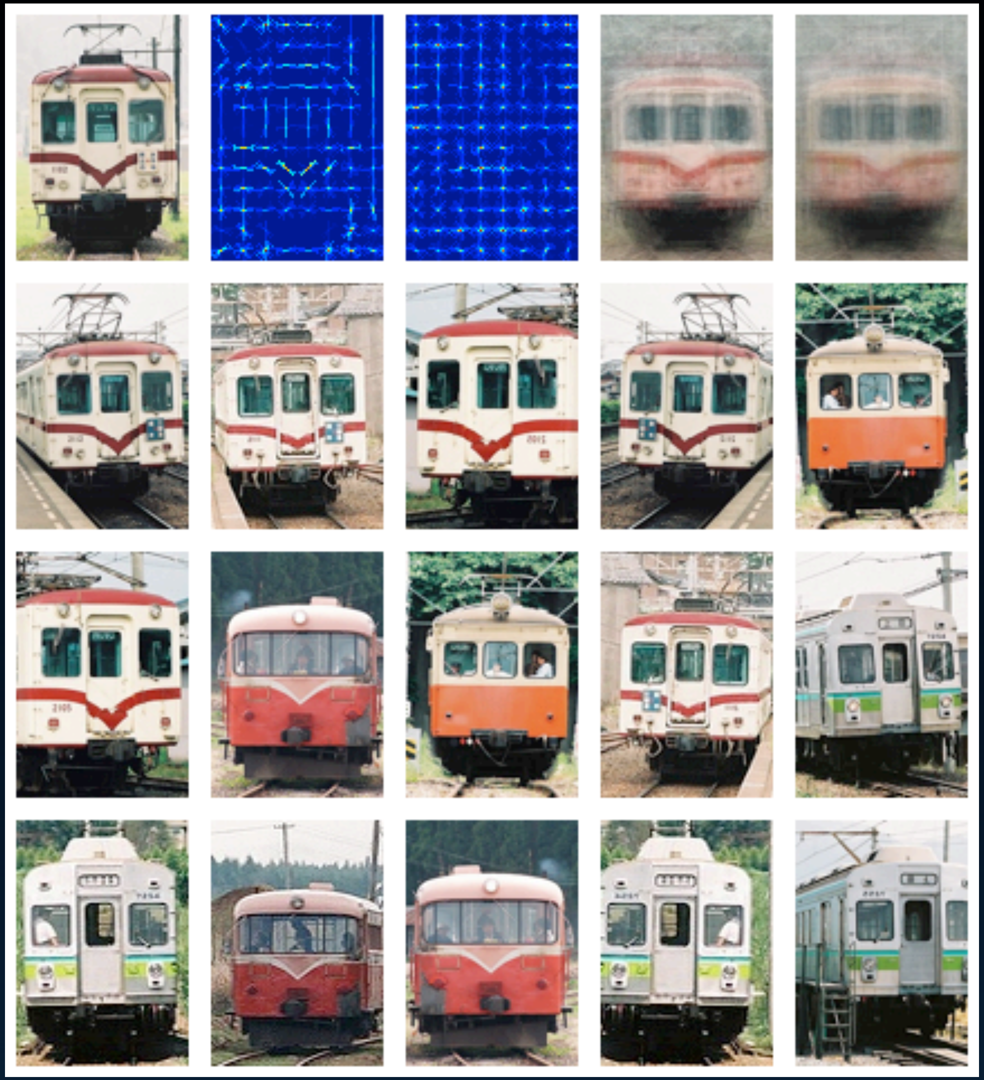


SEARCH USING OBJECTS



Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros.
Ensemble of Exemplar-SVMs for Object Detection and Beyond. In ICCV, 2011.

SEARCH USING OBJECTS



Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros.
Ensemble of Exemplar-SVMs for Object Detection and Beyond. In ICCV, 2011.

QUANTITATIVE EVALUATIONS

SKETCH-BASED IMAGE RETRIEVAL

SKETCH-BASED IMAGE RETRIEVAL

Query Sketches:

25 Car & 25 Bicycle Sketches

SKETCH-BASED IMAGE RETRIEVAL

Query Sketches:

25 Car & 25 Bicycle Sketches

Retrieval Set:

SKETCH-BASED IMAGE RETRIEVAL

Query Sketches:

25 Car & 25 Bicycle Sketches

Retrieval Set:

10,000 Annotated Images

SKETCH-BASED IMAGE RETRIEVAL

Query Sketches:

25 Car & 25 Bicycle Sketches

Retrieval Set:

10,000 Annotated Images

Pascal VOC 2007 Dataset

SKETCH-BASED IMAGE RETRIEVAL

Query Sketches:

25 Car & 25 Bicycle Sketches

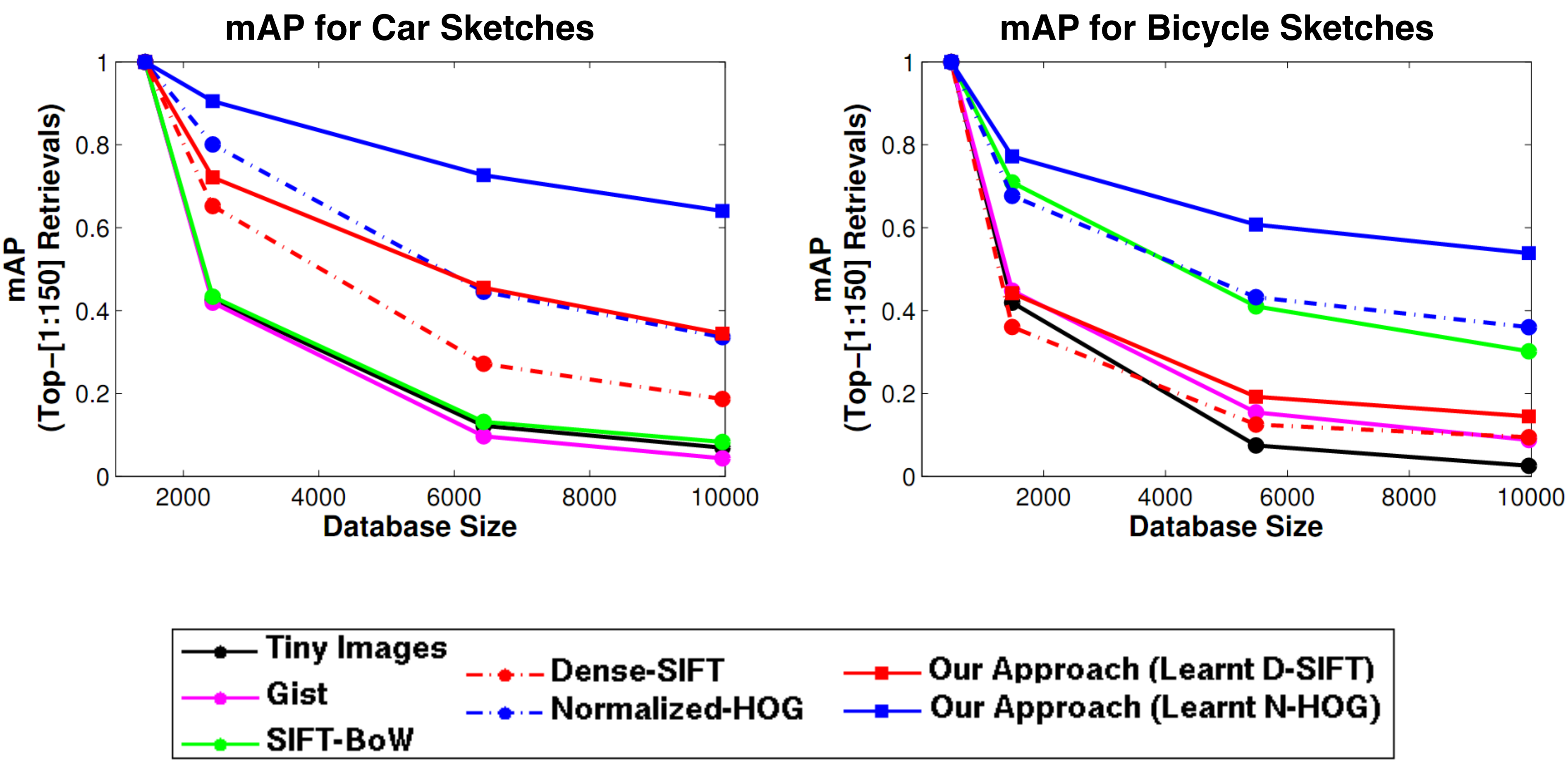
Retrieval Set:

10,000 Annotated Images

Pascal VOC 2007 Dataset

[Everingham et al., 2008]

SKETCH-BASED IMAGE RETRIEVAL



PASCAL VOC Object Detection

Approach	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	diningtable	dog	horse	motorbike	person	pottedplant	sheep	sofa	train	tvmonitor	mAP
NN	.006	.094	.000	.005	.000	.006	.010	.092	.001	.092	.001	.004	.096	.094	.005	.018	.009	.008	.096	.144	.039
NN+Cal	.056	.293	.012	.034	.009	.207	.261	.017	.094	.111	.004	.033	.243	.188	.114	.020	.129	.003	.183	.195	.110
DFUN+Cal	.162	.364	.008	.096	.097	.316	.366	.092	.098	.107	.002	.093	.234	.223	.109	.037	.117	.016	.271	.293	.155
E-SVM+Cal	.204	.407	.093	.100	.103	.310	.401	.096	.104	.147	.023	.097	.384	.320	.192	.096	.167	.110	.291	.315	.198
E-SVM+Co-occ	.208	.480	.077	.143	.131	.397	.411	.052	.116	.186	.111	.031	.447	.394	.169	.112	.226	.170	.369	.300	.227
CZ [6]	.262	.409	–	–	–	.393	.432	–	–	–	–	–	–	.375	–	–	–	–	.334	–	–
DT [7]	.127	.253	.005	.015	.107	.205	.230	.005	.021	.128	.014	.004	.122	.103	.101	.022	.056	.050	.120	.248	.097
LDPM [9]	.287	.510	.006	.145	.265	.397	.502	.163	.165	.166	.245	.050	.452	.383	.362	.090	.174	.228	.341	.384	.266

Table 1. **PASCAL VOC 2007 object detection results.** We compare our full system (ESVM+Co-occ) to four different exemplar based baselines including NN (Nearest Neighbor), NN+Cal (Nearest Neighbor with calibration), DFUN+Cal (learned distance function with calibration) and ESVM+Cal (Exemplar-SVM with calibration). We also compare our approach against global methods including our implementation of Dalal-Triggs (learning a single global template), LDPM [9] (Latent deformable part model), and Chum et al. [6]'s exemplar-based method. [The NN, NN+Cal and DFUN+Cal results for person category are obtained using 1250 exemplars]

PASCAL VOC Object Detection

Approach	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	diningtable	dog	horse	motorbike	person	pottedplan	sheep	sofa	train	tvmonitor	mAP
NN	.006	.094	.000	.005	.000	.006	.010	.092	.001	.092	.001	.004	.096	.094	.005	.018	.009	.008	.096	.144	.039
NN+Cal	.056	.293	.012	.034	.009	.207	.261	.017	.094	.111	.004	.033	.243	.188	.114	.020	.129	.003	.183	.195	.110
DFUN+Cal	.162	.364	.008	.096	.097	.316	.366	.092	.098	.107	.002	.093	.234	.223	.109	.037	.117	.016	.271	.293	.155
E-SVM+Cal	.204	.407	.093	.100	.103	.310	.401	.096	.104	.147	.023	.097	.384	.320	.192	.096	.167	.110	.291	.315	.198
E-SVM+Co-occ	.208	.480	.077	.143	.131	.397	.411	.052	.116	.186	.111	.031	.447	.394	.169	.112	.226	.170	.369	.300	.227
CZ [6]	.262	.409	–	–	–	.393	.432	–	–	–	–	–	–	.375	–	–	–	–	.334	–	–
DT [7]	.127	.253	.005	.015	.107	.205	.230	.005	.021	.128	.014	.004	.122	.103	.101	.022	.056	.050	.120	.248	.097
LDPM [9]	.287	.510	.006	.145	.265	.397	.502	.163	.165	.166	.245	.050	.452	.383	.362	.090	.174	.228	.341	.384	.266

Table 1. **PASCAL VOC 2007 object detection results.** We compare our full system (ESVM+Co-occ) to four different exemplar based baselines including NN (Nearest Neighbor), NN+Cal (Nearest Neighbor with calibration), DFUN+Cal (learned distance function with calibration) and ESVM+Cal (Exemplar-SVM with calibration). We also compare our approach against global methods including our implementation of Dalal-Triggs (learning a single global template), LDPM [9] (Latent deformable part model), and Chum et al. [6]'s exemplar-based method. [The NN, NN+Cal and DFUN+Cal results for person category are obtained using 1250 exemplars]

Equal or better in performance than Felzenszwalb et al's Deformable Part-based Model in 7 PASCAL VOC 2007 categories.

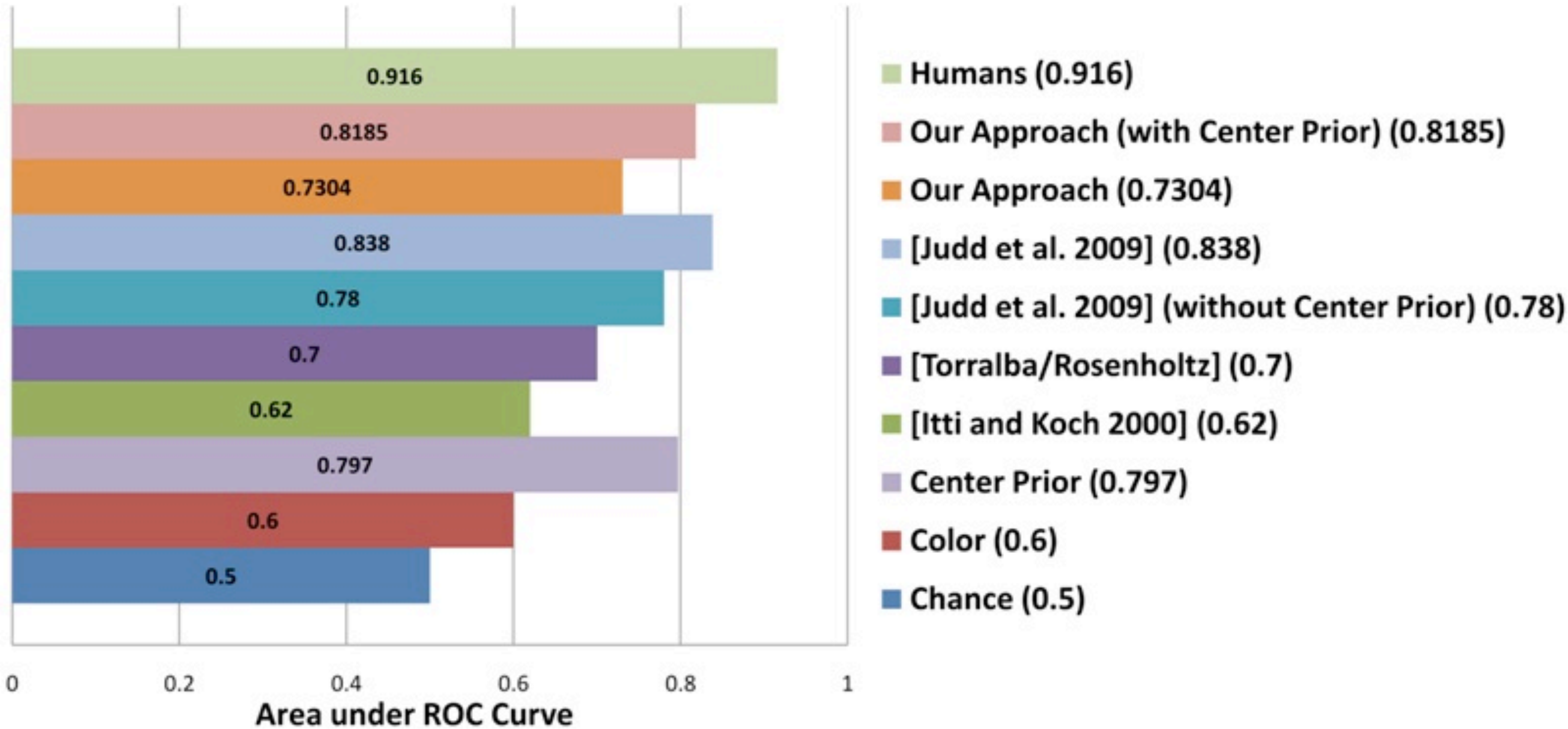
SALIENCY

[Yarbus, 1962]

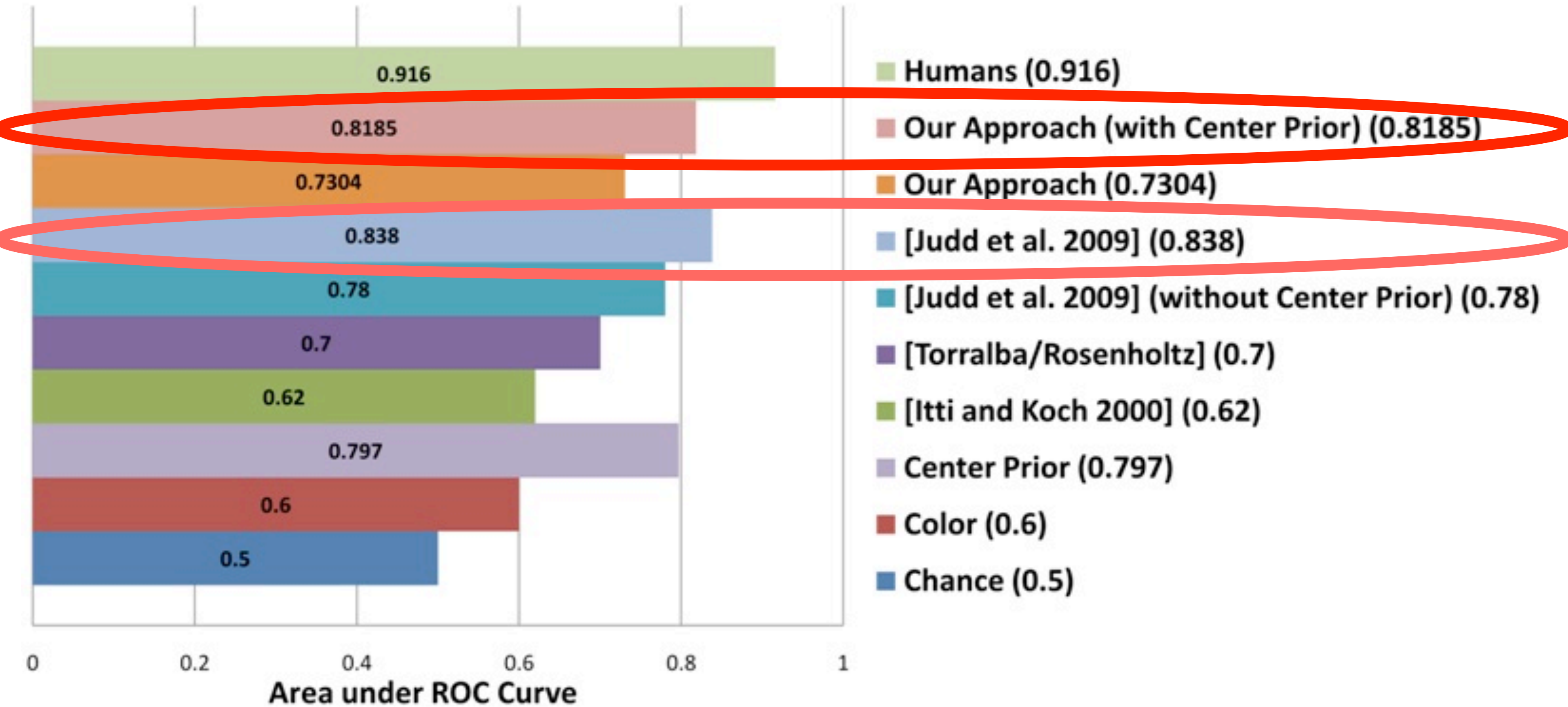
PROXY FOR SALIENCY

PROXY FOR SALIENCY

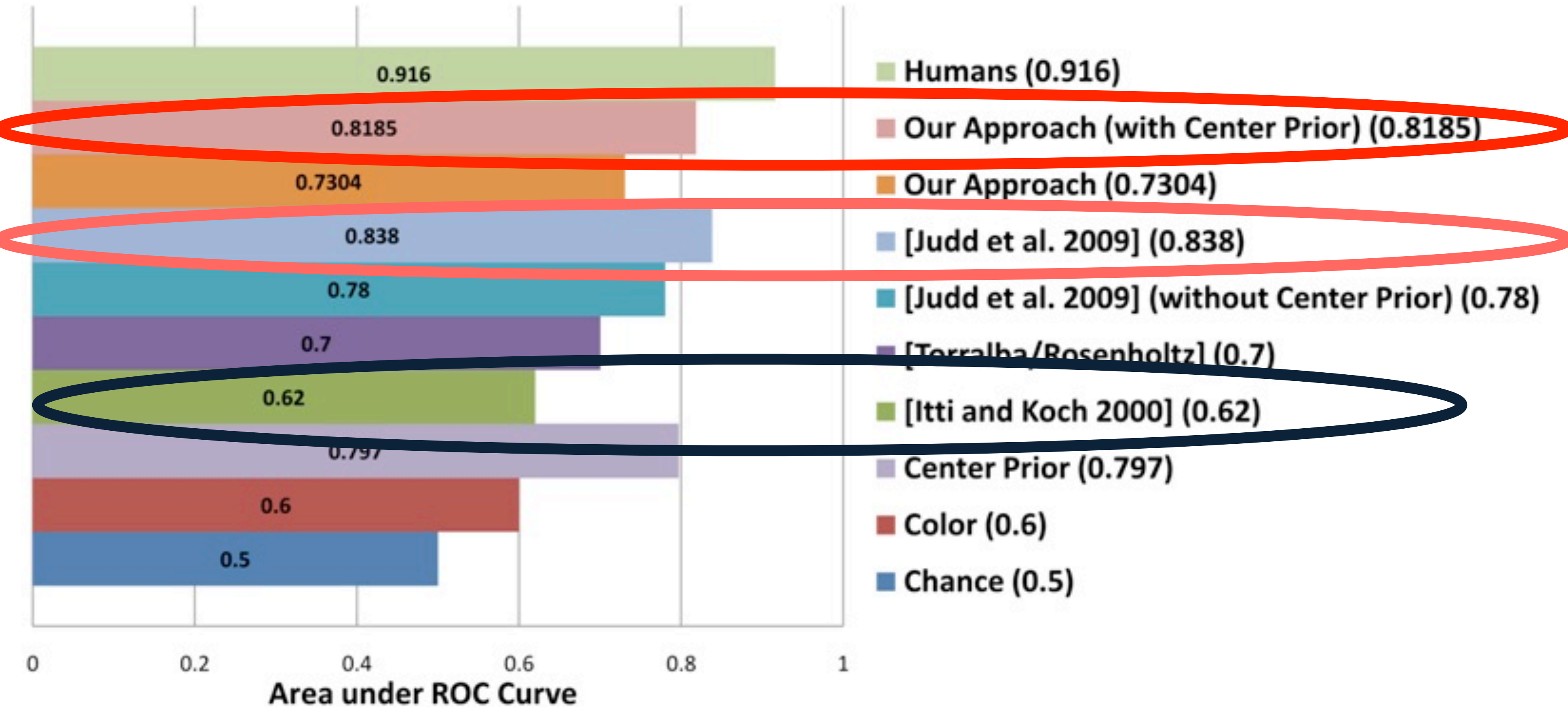
PREDICTING SALIENCY
SALIENCY DATASET [Judd et al., 2009]



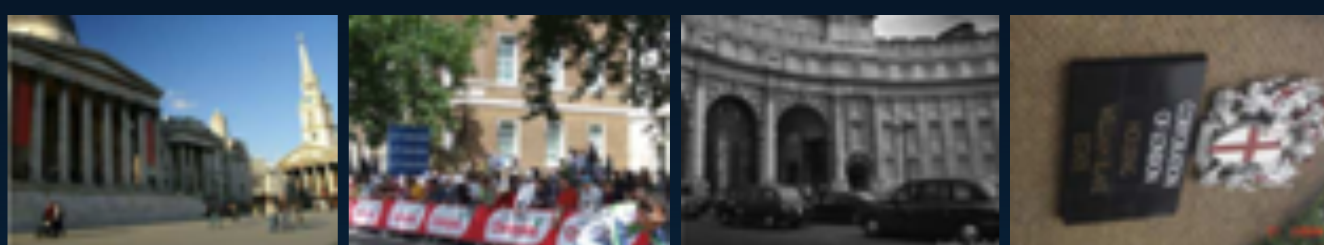
PREDICTING SALIENCY
SALIENCY DATASET [Judd et al., 2009]



PREDICTING SALIENCY
SALIENCY DATASET [Judd et al., 2009]



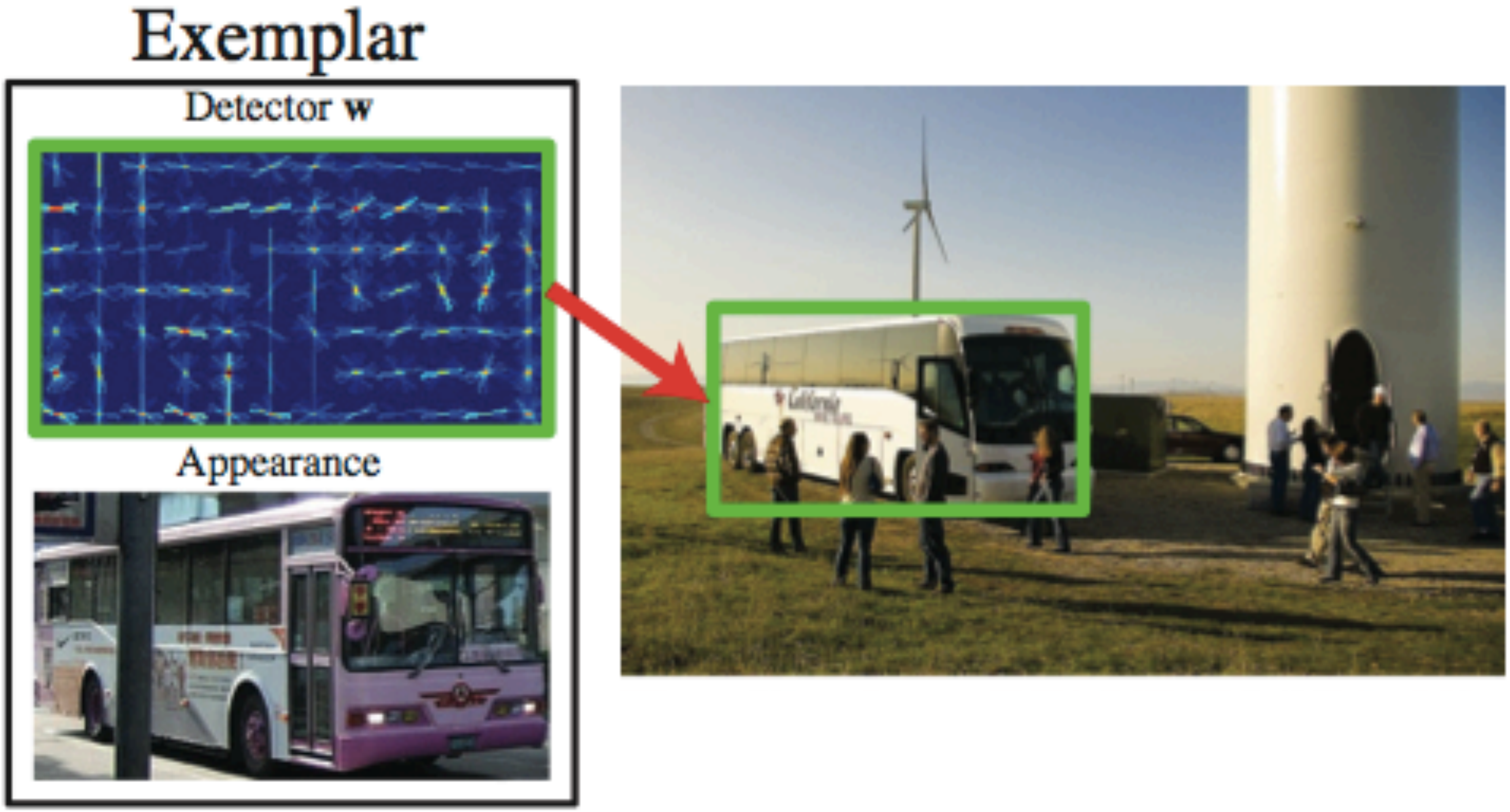
WHERE DOES IT FAIL?

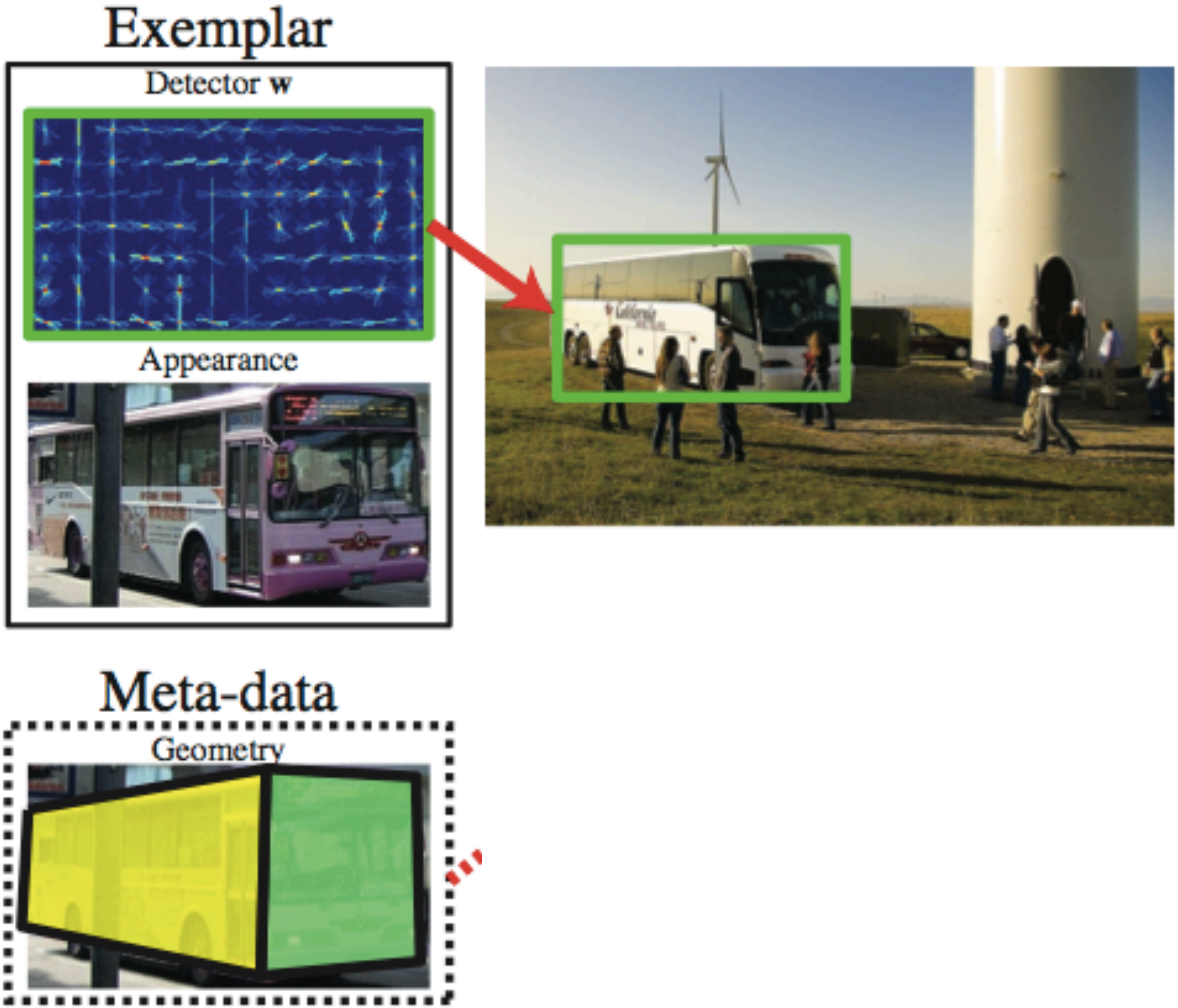


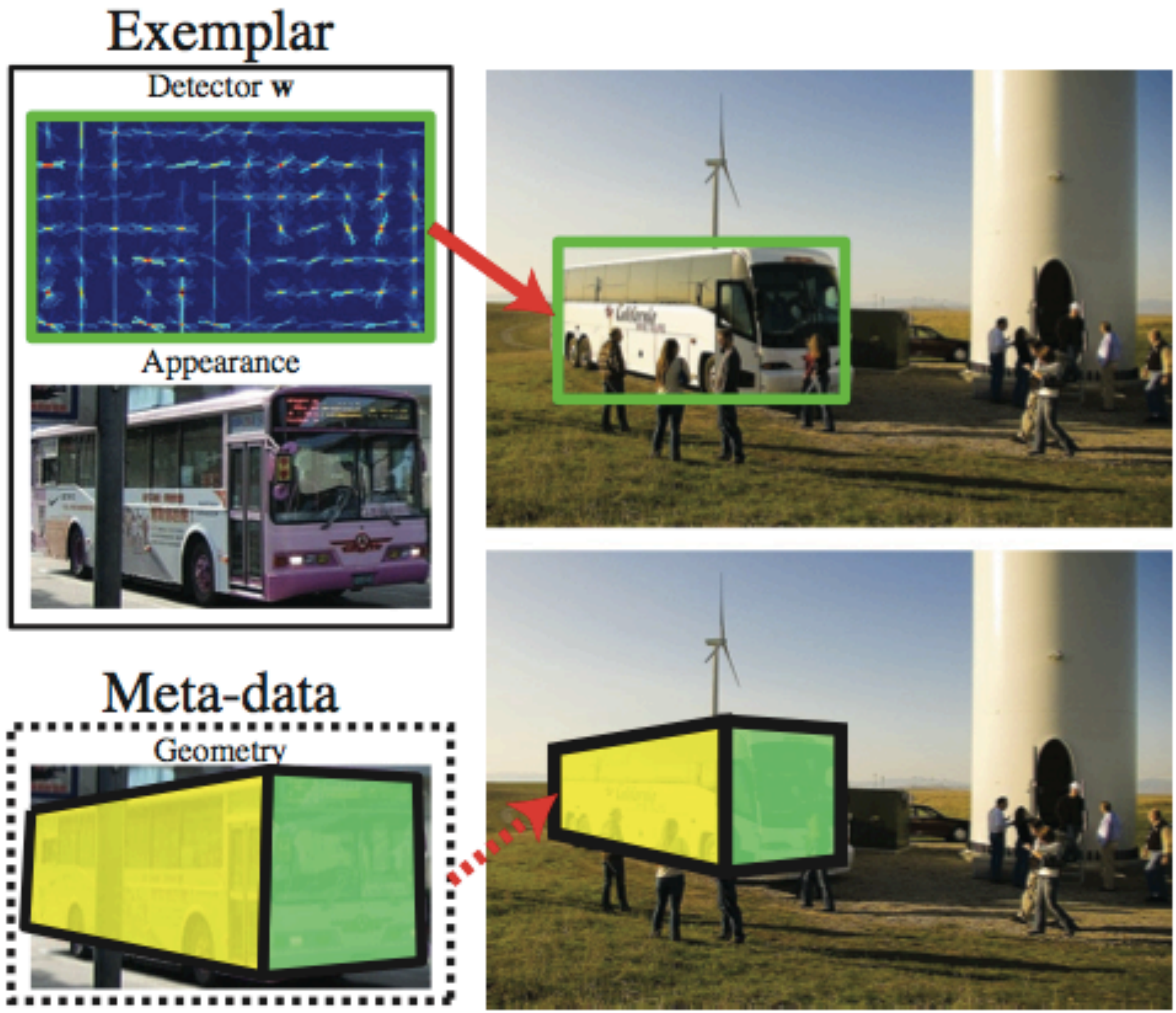
Top Matches

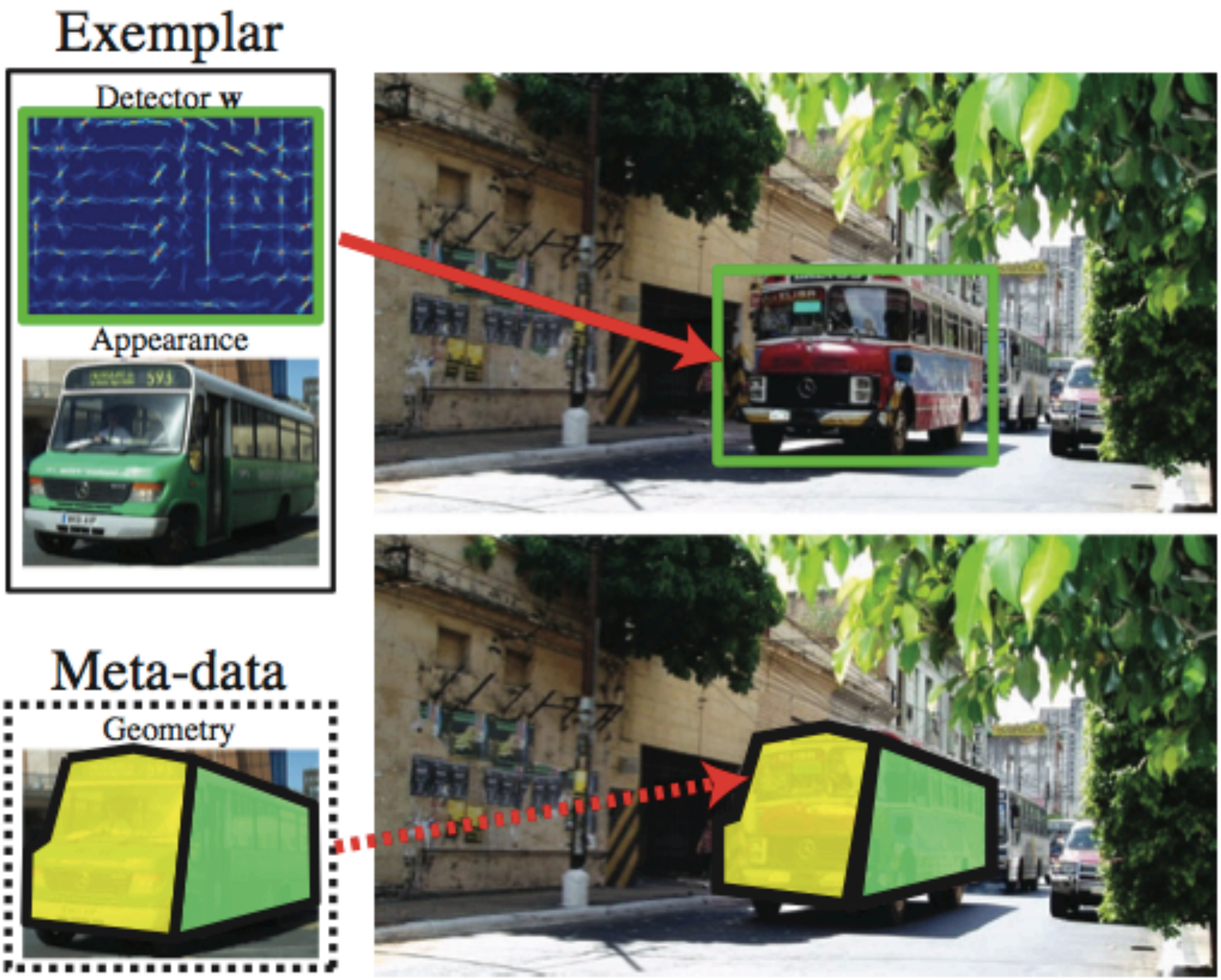
APPLICATIONS

Label Transfer

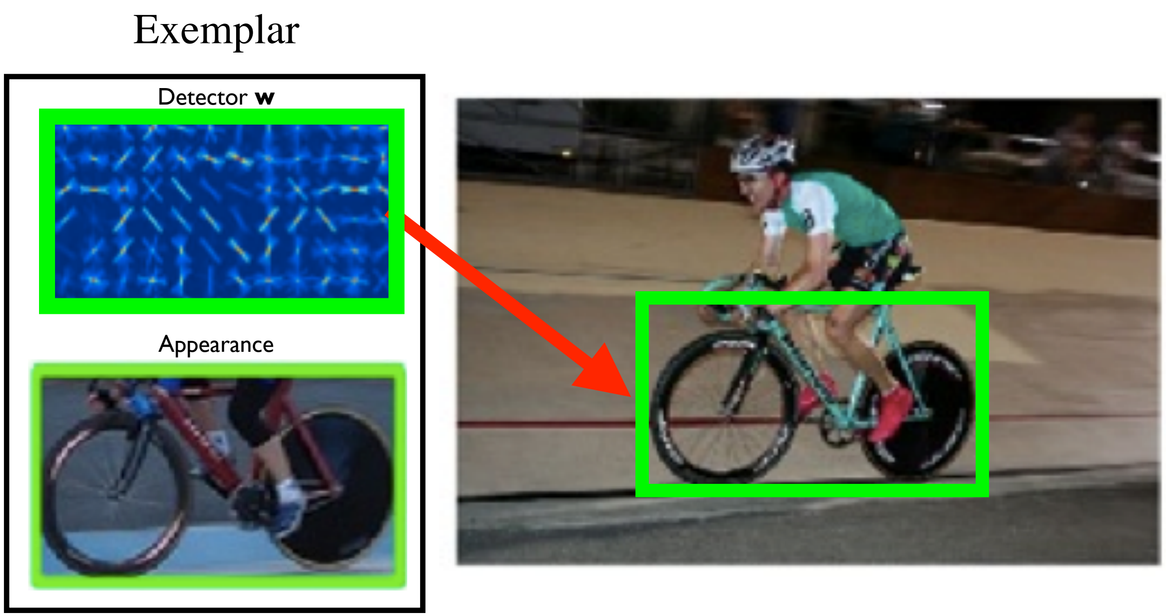

Exemplar
Detector w
Appearance

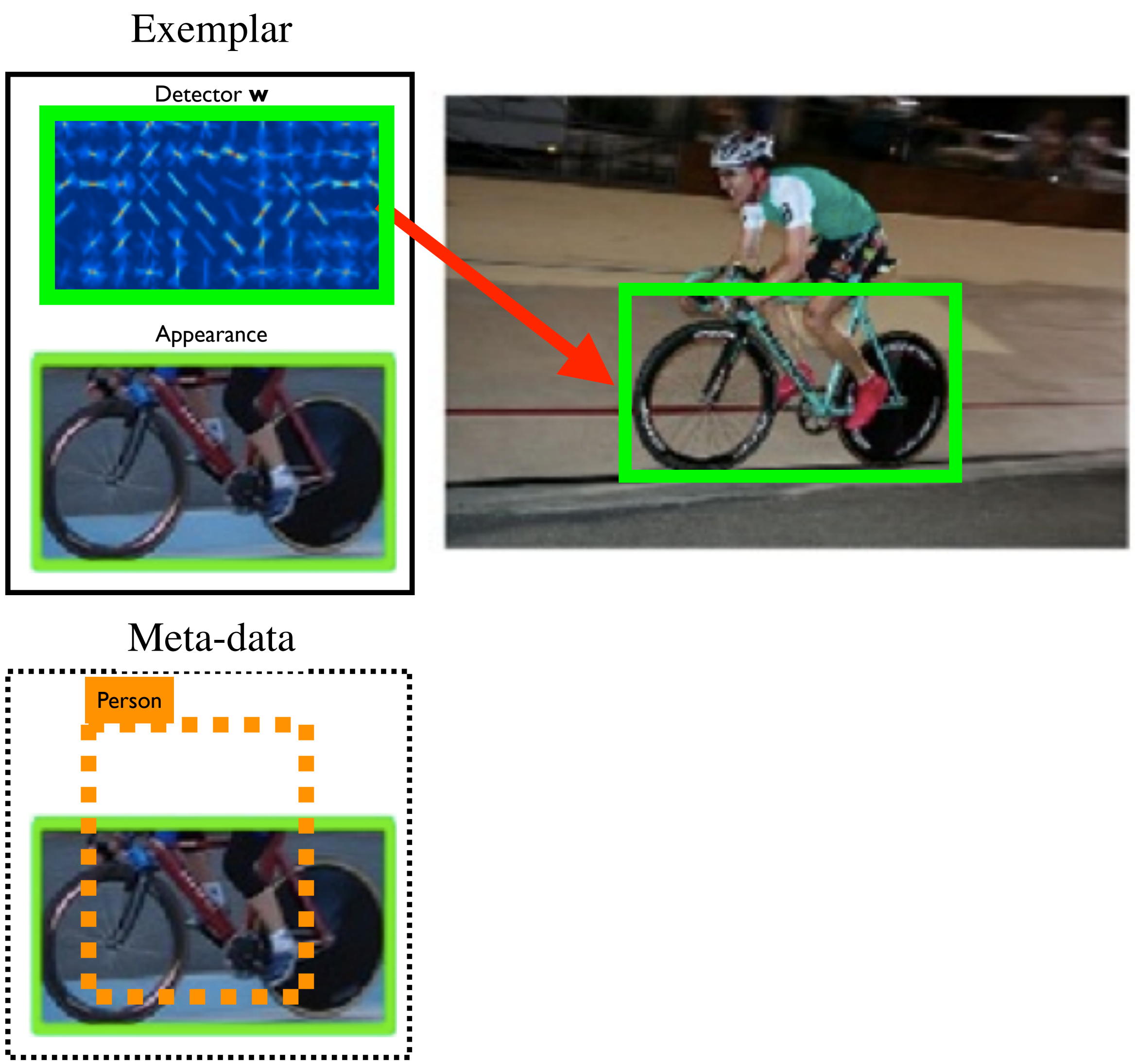

Exemplar
Detector w
Appearance
Meta-data
Geometry

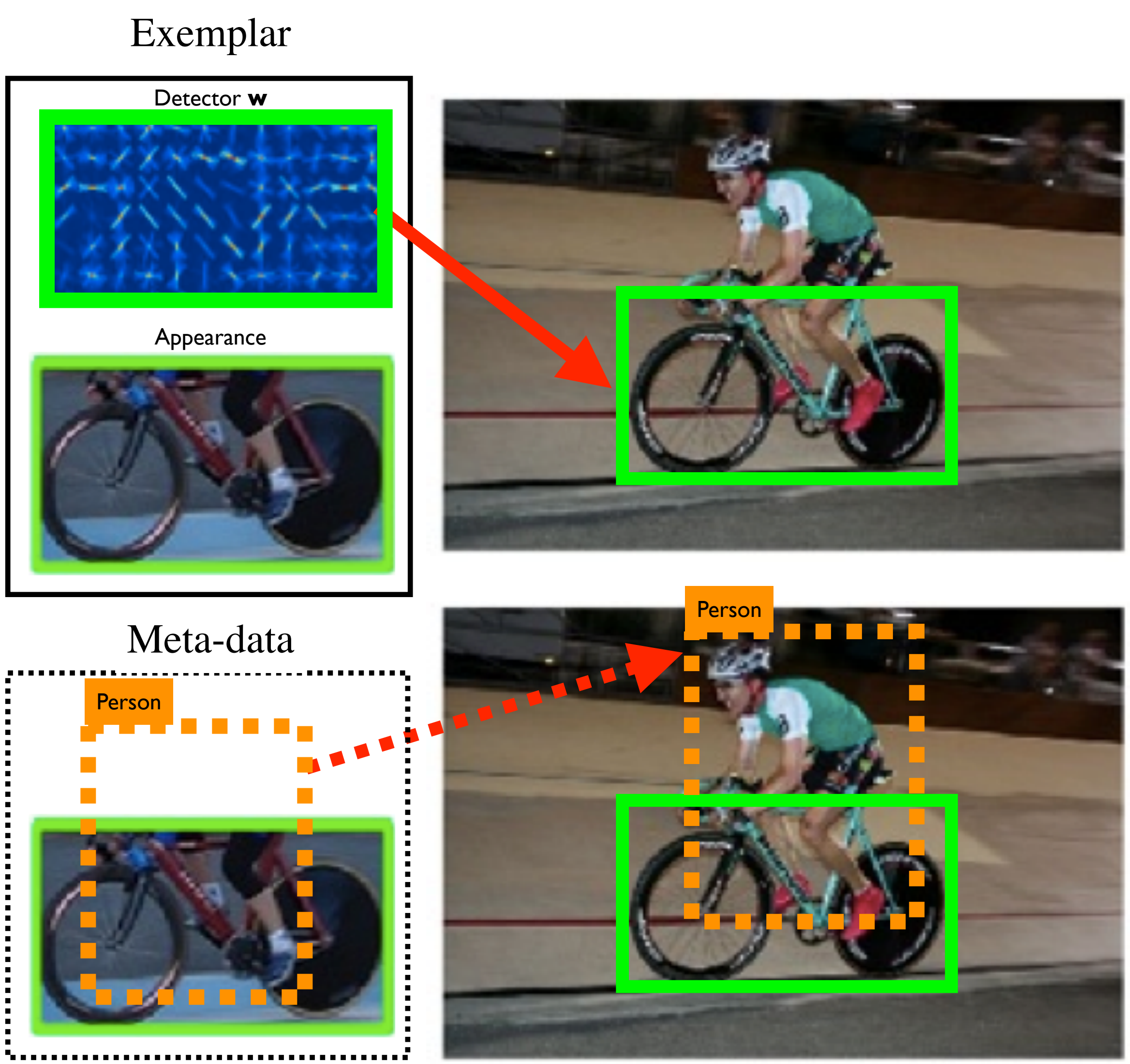

Exemplar
Detector w
Appearance
Meta-data
Geometry

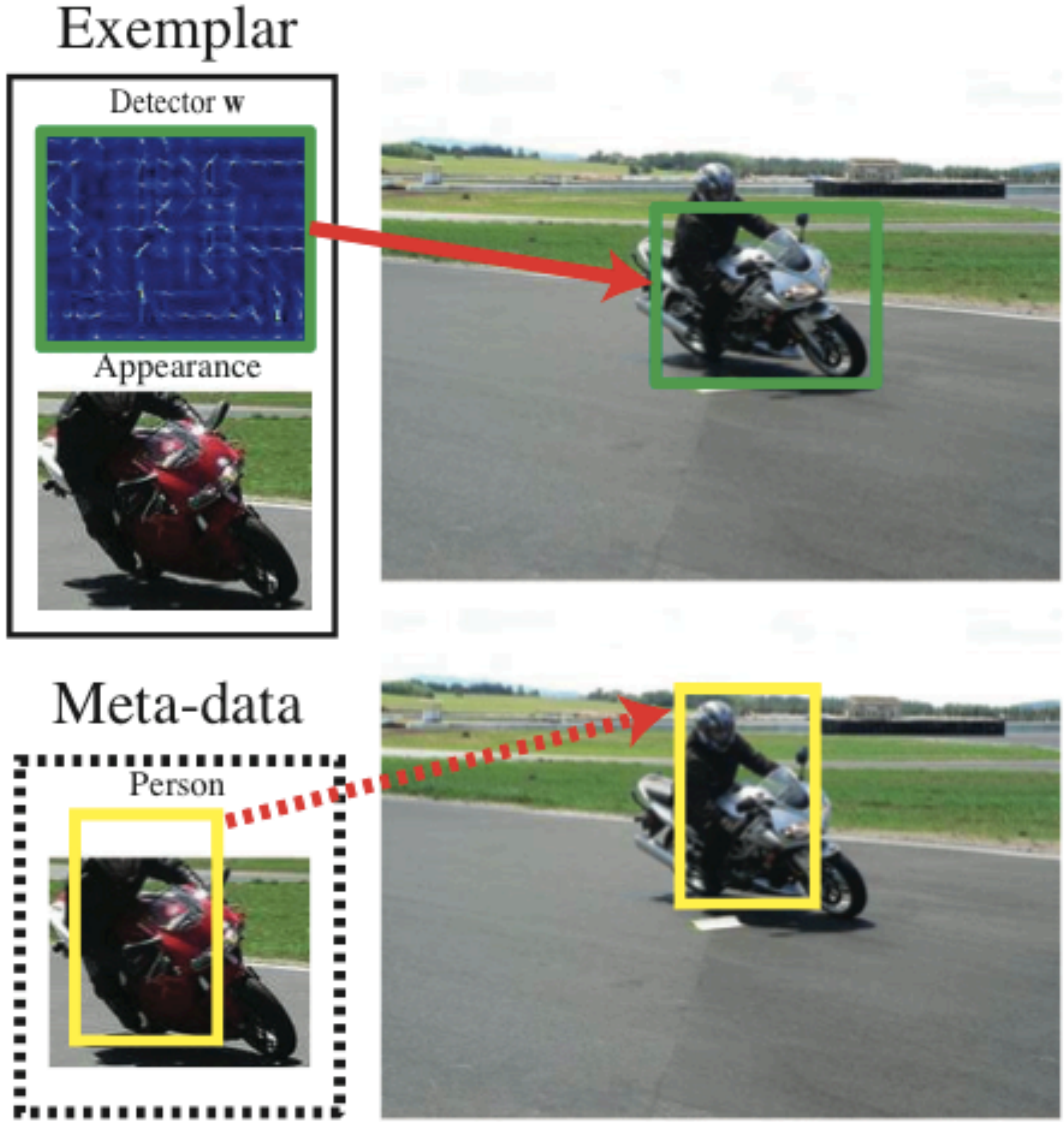

Exemplar
Detector w
Appearance
Meta-data
Geometry

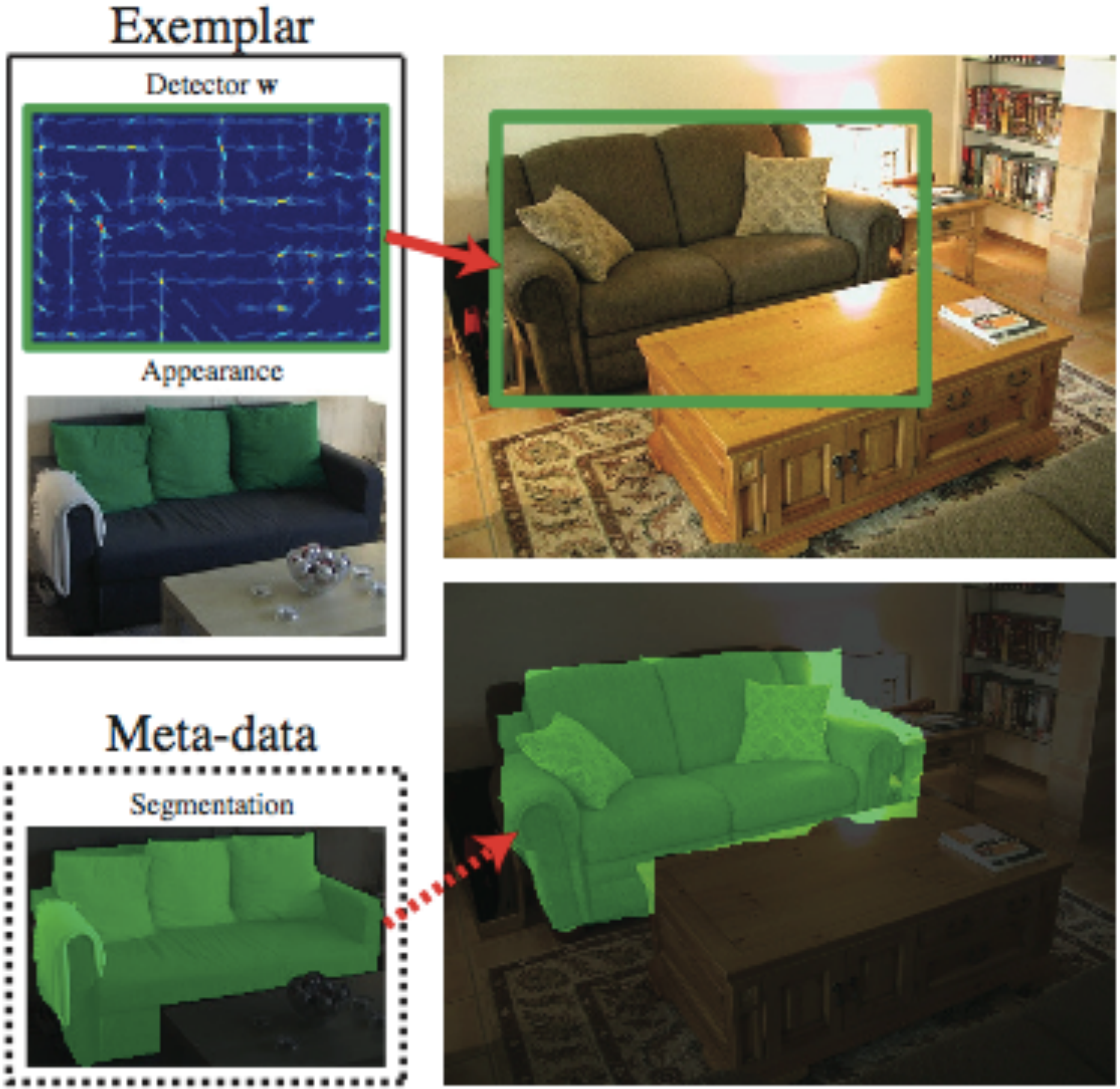

Exemplar
Detector w
Appearance
Meta-data
Geometry

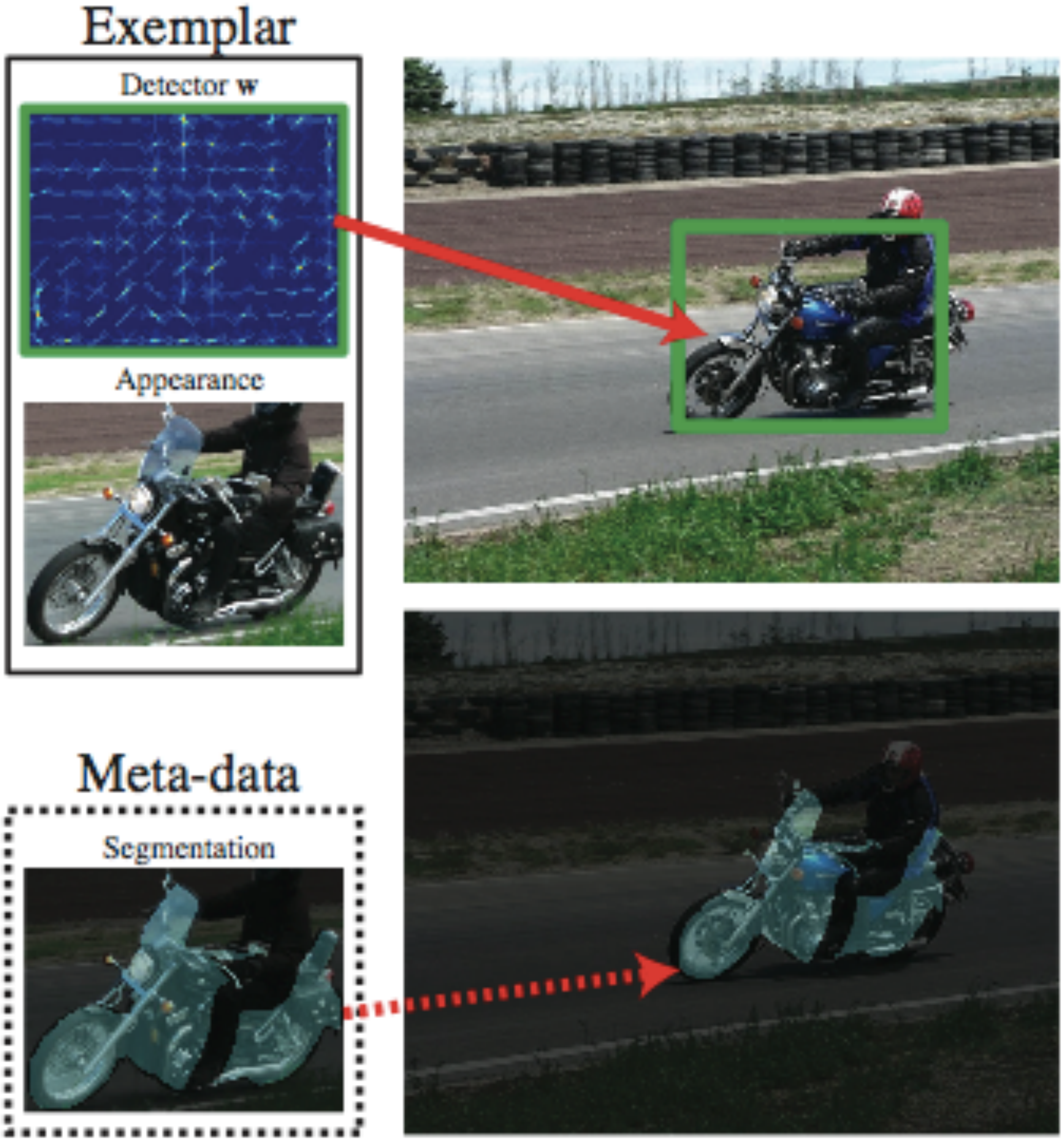

Exemplar
Detector **w**
Appearance


Exemplar
Detector w
Appearance
Meta-data
Person

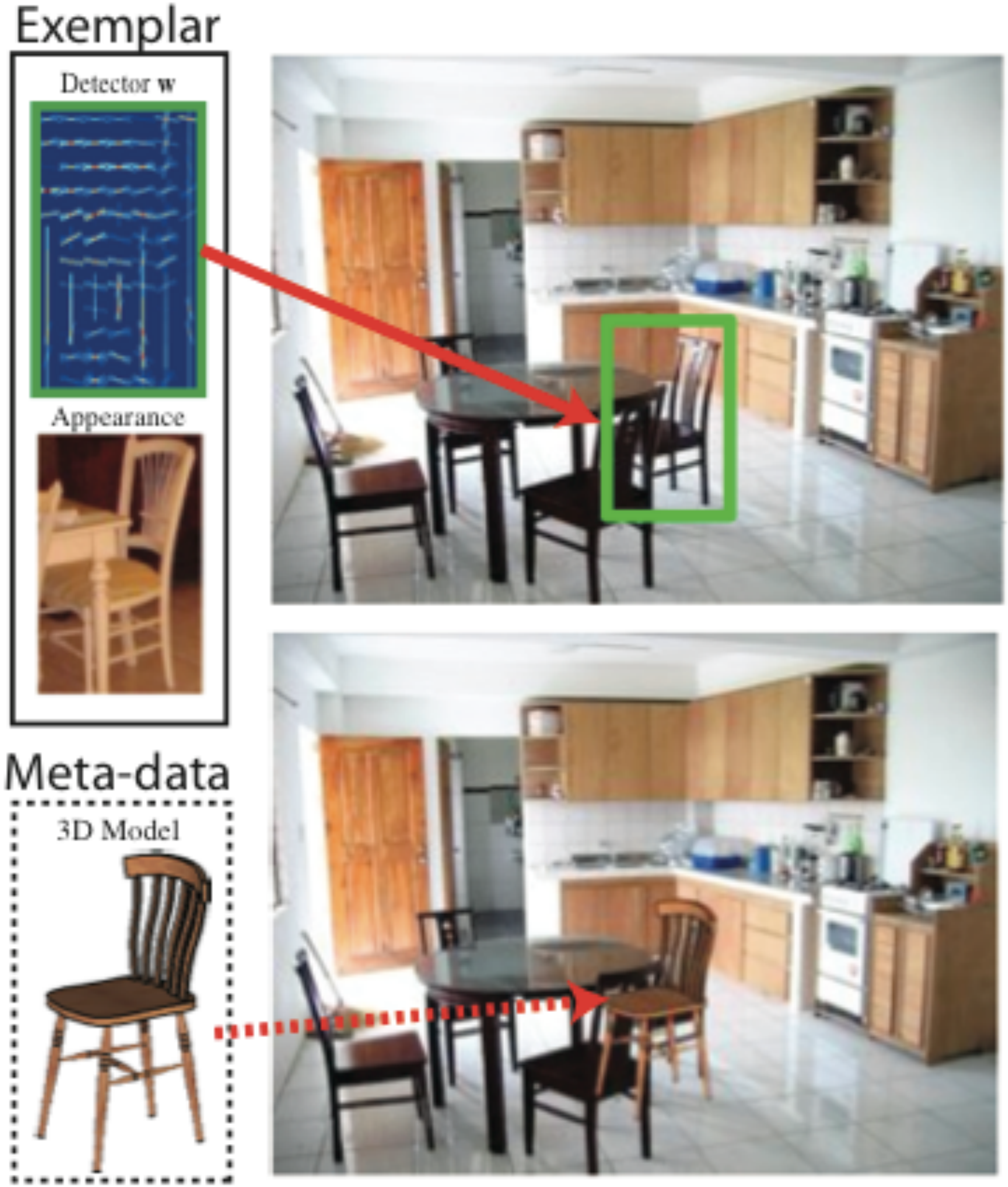

Exemplar
Detector w
Appearance
Meta-data
Person
Person


Exemplar
Detector w
Appearance
Meta-data
Person


Exemplar
Detector w
Appearance
Meta-data
Segmentation


Exemplar
Detector w
Appearance
Meta-data
Segmentation


Exemplar
Detector w
Appearance
Meta-data
3D Model


Exemplar
Detector w
Appearance
Meta-data
3D Model

RE-PHOTOGRAPHY

Historical Image of Boston Station

RE-PHOTOGRAPHY

Historical Image of Boston Station

Re-photographed Image

RE-PHOTOGRAPHY

Computational Re-photography
(Bae et al., 2010)

Historical Image of
Boston Station

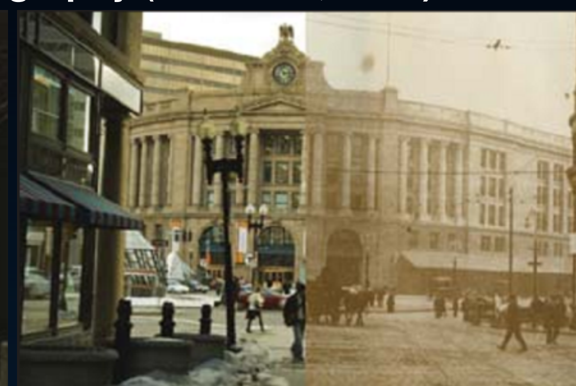
Re-photographed Image

RE-PHOTOGRAPHY

Historical Image of
Boston Station

Computational Re-photography (Bae et al., 2010)

Re-photographed Image



Then & Now View

INTERNET RE-PHOTOGRAPHY

Computational Re-photography (Bae et al., 2010)

Historical Image of Boston Station

Re-photographed Image

Then & Now View

Historical Image of Boston Station

INTERNET RE-PHOTOGRAPHY

Computational Re-photography (Bae et al., 2010)

Historical Image of
Boston Station

Re-photographed Image

Then & Now View

Historical Image of
Boston Station

Search
10,000 Flickr Images
of Boston

INTERNET RE-PHOTOGRAPHY

Historical Image of
Boston Station

Computational Re-photography (Bae et al., 2010)

Re-photographed Image

Then & Now View

Historical Image of
Boston Station

Search
10,000 Flickr Images
of Boston

Our Approach

Top Match

INTERNET RE-PHOTOGRAPHY

Computational Re-photography (Bae et al., 2010)

Historical Image of Boston Station

Re-photographed Image

Then & Now View

Our Approach

Historical Image of Boston Station

Top Match
From 10,000 Flickr Images

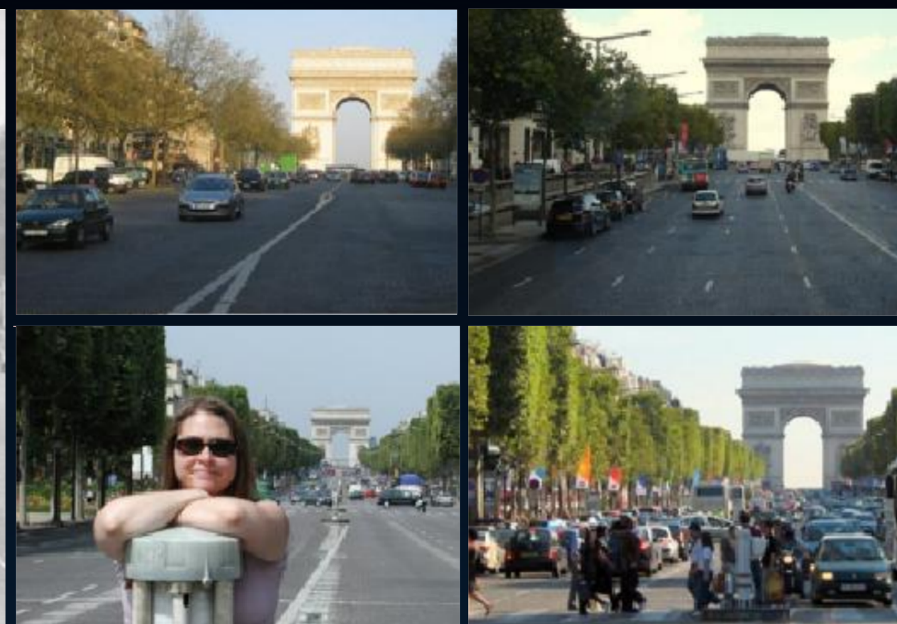
Then & Now View

INTERNET RE-PHOTOGRAPHY

Paris (1940)

INTERNET RE-PHOTOGRAPHY

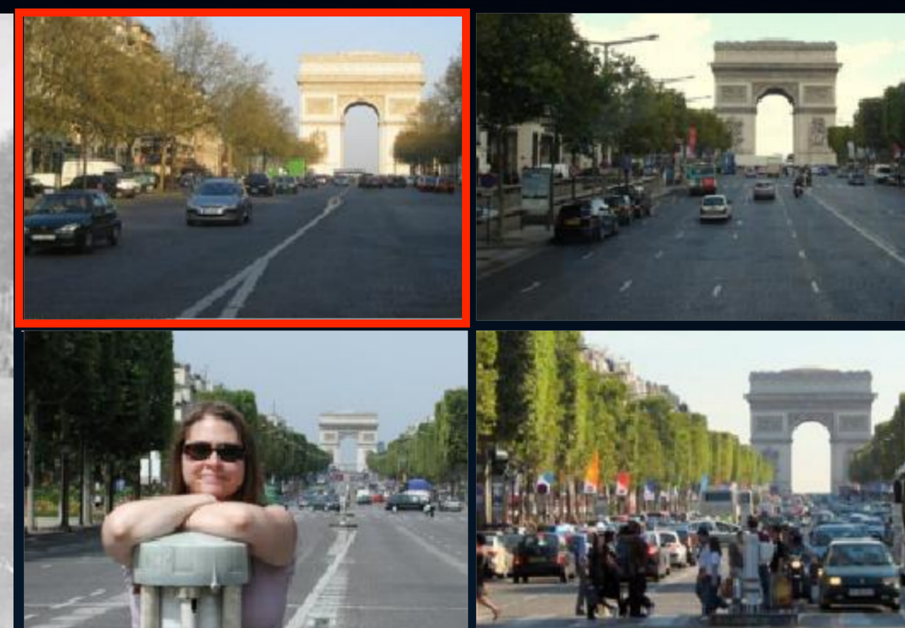
Paris (1940)



Top Matches

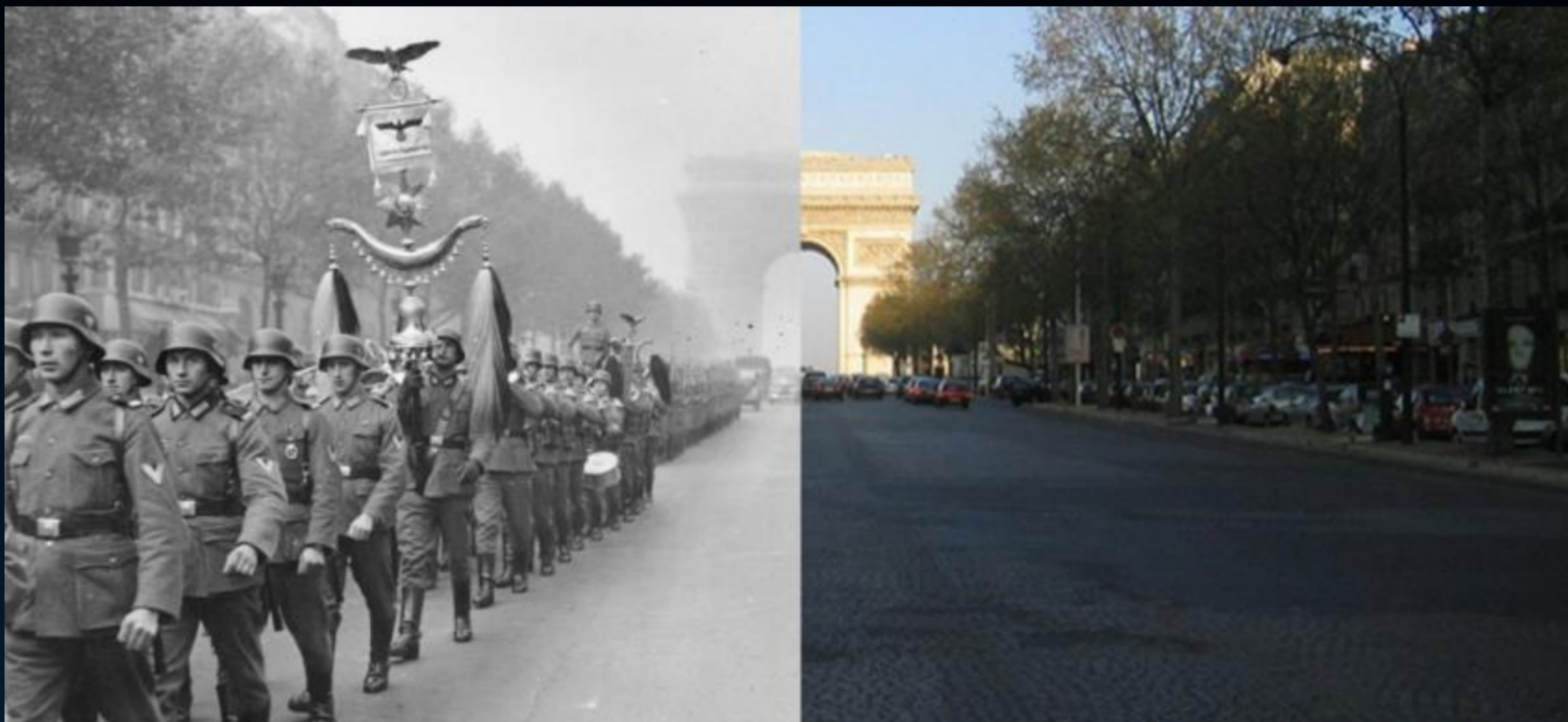
INTERNET RE-PHOTOGRAPHY

Paris (1940)



Top Matches

INTERNET RE-PHOTOGRAPHY



WHERE WAS THE PAINTER STANDING?

Input Painting

PAINTING2GPS

Input Painting

Retrieval set
10,000 Geo-tagged Flickr Images

100 top matches used to estimation

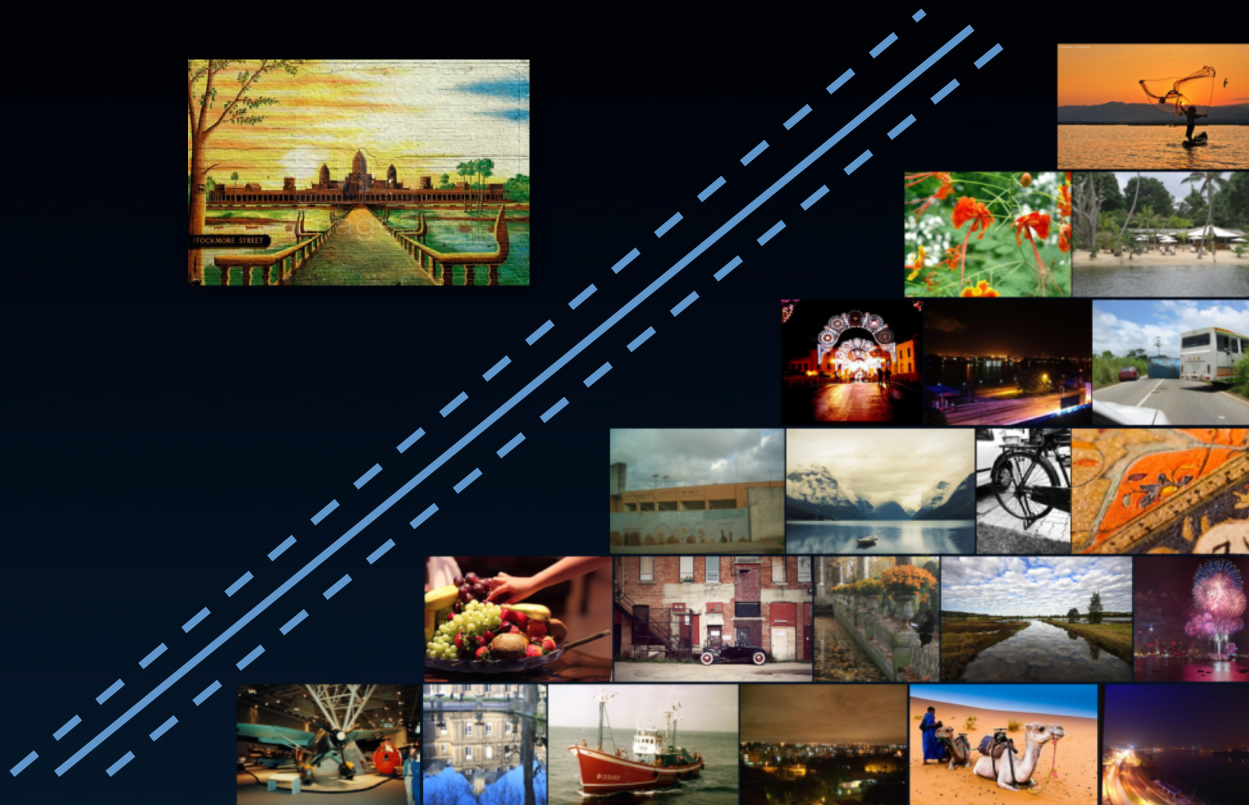
PAINTING2GPS

Input Painting

Estimated Geo-location

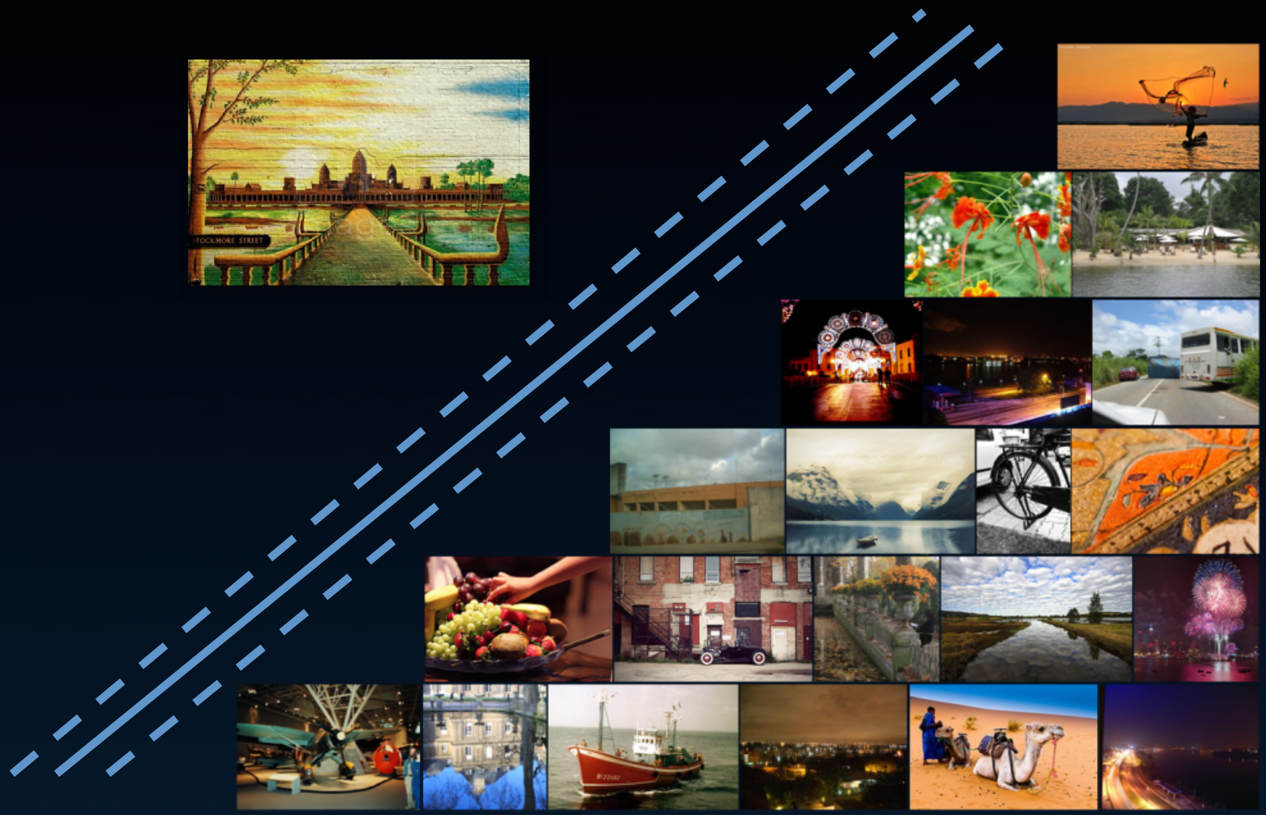
Estimated using 100 top matches

CONCLUSION



- Good News:
  - Results surprisingly nice, embarrassingly parallel learning
- Bad News:
  - Computationally expensive

CONCLUSION



Website:

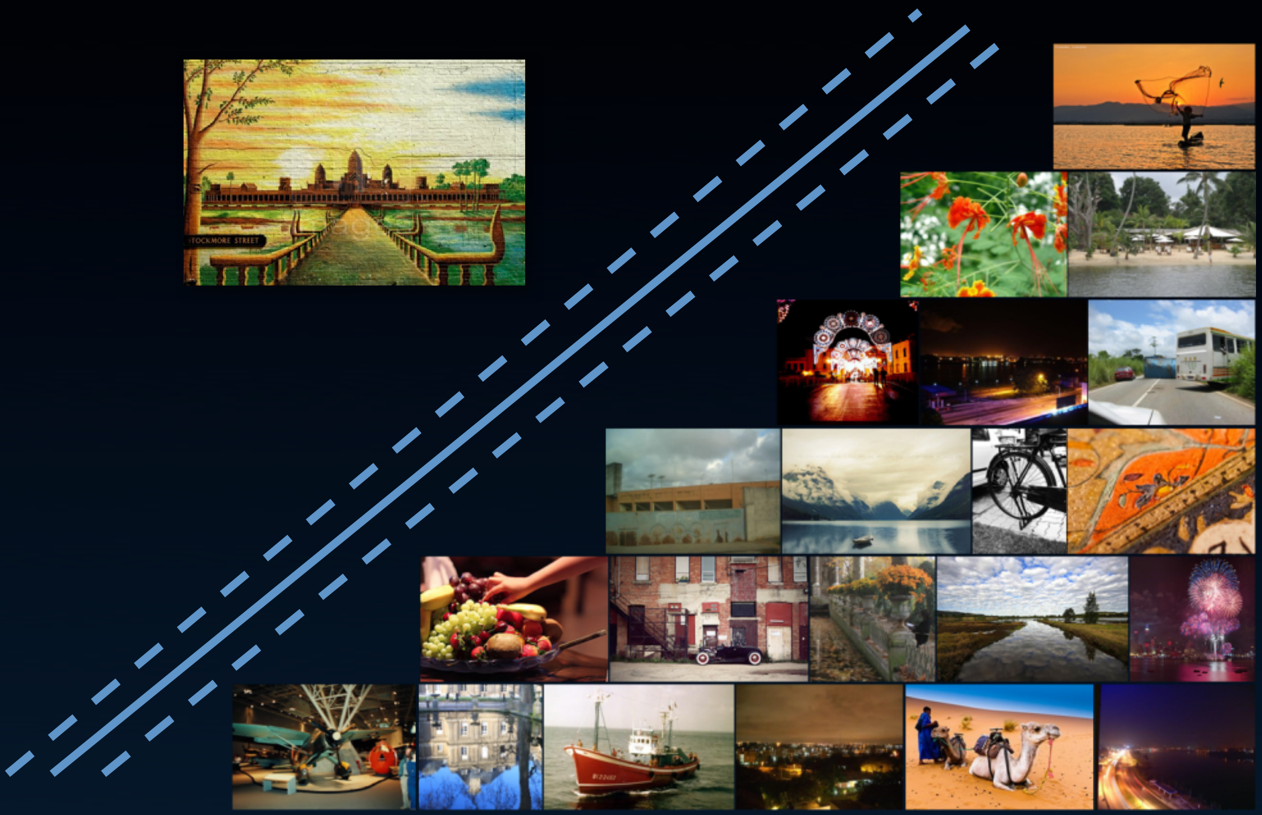
http://graphics.cs.cmu.edu/projects/crossDomainMatching/

http://www.cs.cmu.edu/~tmalisie/projects/iccv11/

Code:

https://github.com/quantombone/exemplarsvm

Thank You



Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros. **Ensemble of Exemplar-SVMs for Object Detection and Beyond.** In ICCV, 2011.

Abhinav Shrivastava, Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros. **Data-driven Visual Similarity for Cross-domain Image Matching.** In SIGGRAPH ASIA, 2011.