Modeling the shape of the scene

Amirshahed Mehrtash




Outline

- A quick review
- Descriptors
- Examples of categories
- Outliners and insiders
- Simulation
- A philosophical note
- Discussion

A QUICK REVIEW

What is a scene?

- Images
  - “objects”: 1-2 meters
  - “environments”: > 5 meters
- This paper
  - Scene representation
  - Scene statistics

Slide by Eamon Caddigan

Scene recognition

- Scenes vs. collections of objects
- Object information may be ignored
  - Fast categorization
  - Low spatial frequencies
  - Change blindness, inattention blindness
- Scene category provides context for images
- Need holistic representation of a scene

Slide by Eamon Caddigan

Holistic scene representation

- Finding a low-dimensional "scene space"
- Clustering by humans
  - Split images into groups
  - ignore objects, categories

Table 1. Spatial envelope properties of environmental scenes.

Property	S1	S2	S3	Total
Naturalness	65	12	0	77
Openness	6	53	24	83
Perspective	6	18	29	53
Size	0	0	47	47
Diagonal plane	0	12	29	41
Depth	18	12	29	59
Symmetry	0	0	29	29
Contrast	0	0	18	18

Results are in %, for each of the three experimental steps. The total represents the percent of times the attribute has been used regardless of the stage of the experiment.

Slide by Eamon Caddigan

Spatial envelope properties

- Naturalness
  - natural vs. man-made environments

Slide by Eamon Caddigan

Spatial envelope properties

- Openness
  - decreases as number of boundary elements increases


Slide by Eamon Caddigan

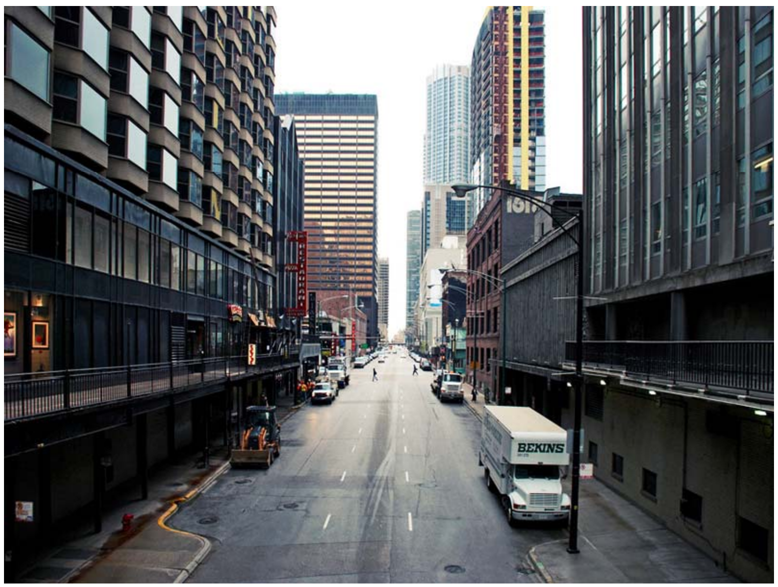
Spatial envelope properties

- Roughness
  - size of elements at each spatial scale, related to fractal dimension

Spatial envelope properties

- Expansion (man-made environments)
  - depth gradient of the space



Slide by Eamon Caddigan

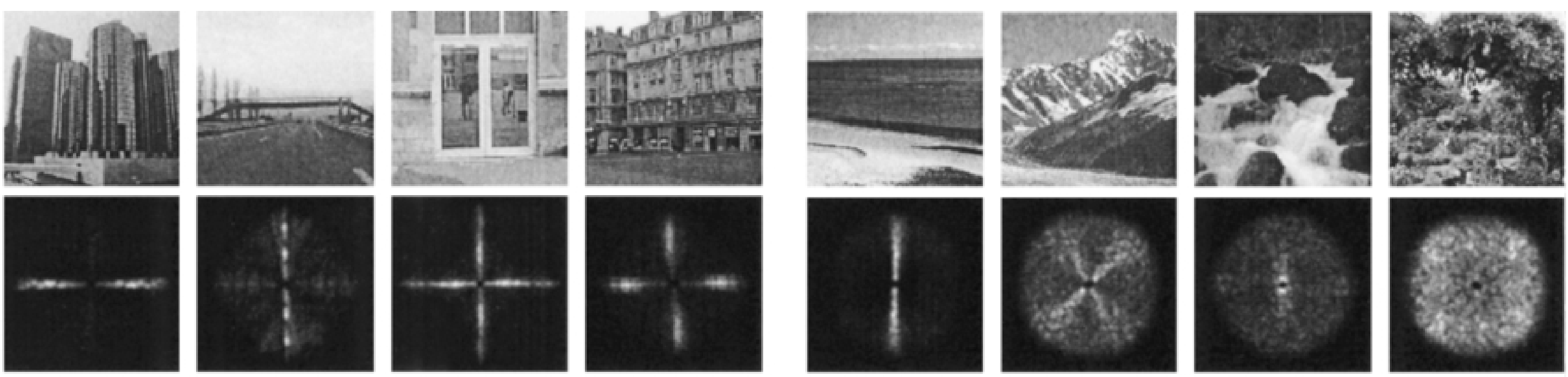
Spatial envelope properties

- Ruggedness (natural environments)
  - deviation of ground relative to horizon


Slide by Eamon Caddigan

Scene classification from statistics

- Different scene categories have different spectral signatures
  - Amplitude captures roughness
  - Orientation captures dominant edges



Slide by Eamon Caddigan

Gist of a scene

DESCRIPTORS

Pre-filtering

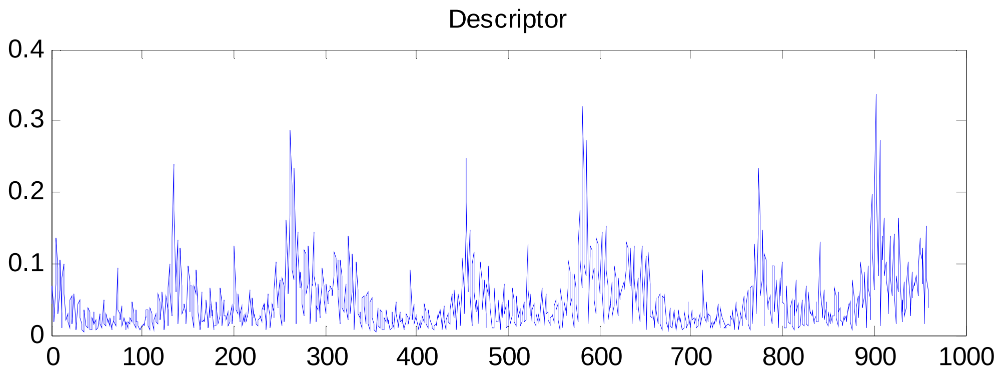
We apply a pre-filtering to the input image i (x, y) that reduces illumination effects and prevents some local image regions to dominate the energy spectrum.

$$i'(x, y) = \frac{i(x, y) * h(x, y)}{\epsilon + \sqrt{[i(x, y) * h(x, y)]^2 * g(x, y)}}$$

The pre-filtering consists in a local normalization of the intensity variance.

g(x, y) is an isotropic low-pass Gaussian spatial filter with a radial cut-off frequency at 0.015 cycles/pixel, and h(x, y) = 1 − g(x, y).

Gist descriptor



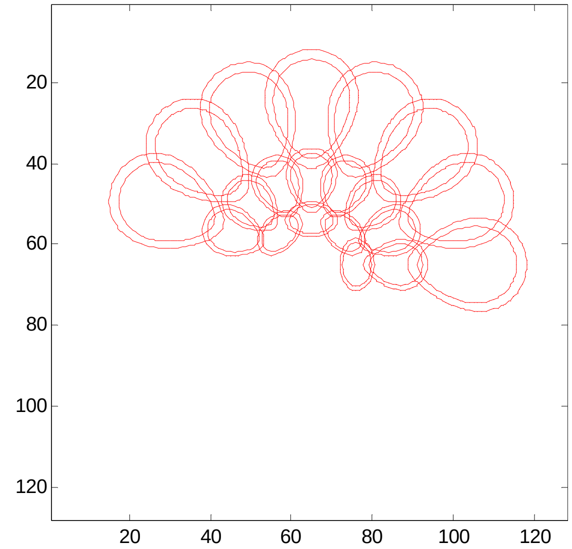
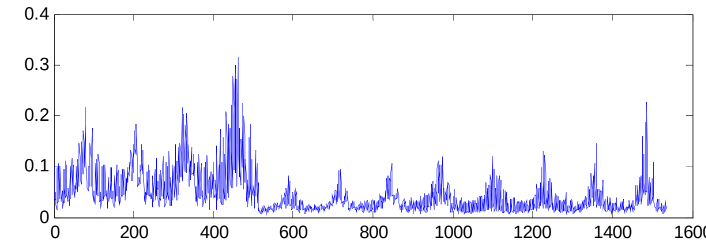
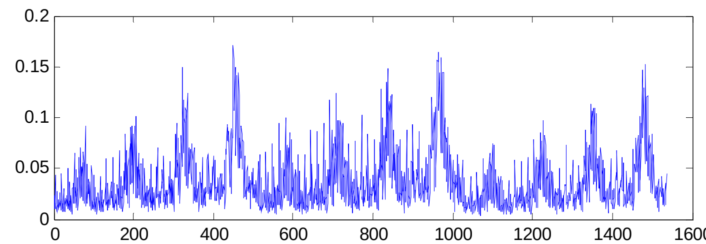
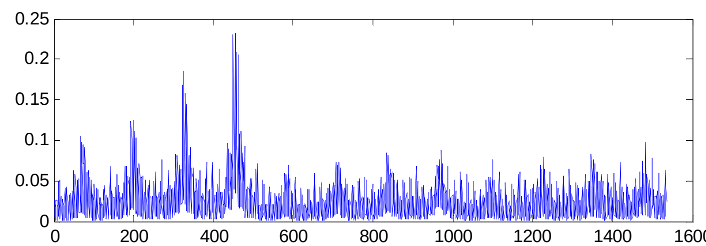


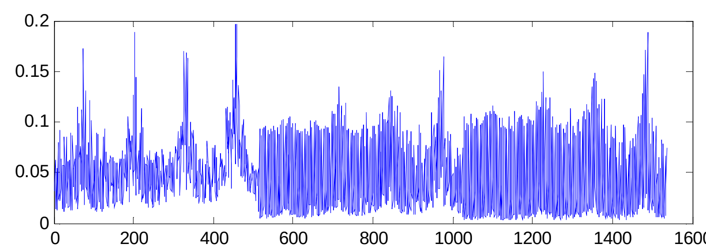
Image size = 128;
Orientations per scale = [8 8 4];
Number of Blocks = 4;

Descriptors for coast images

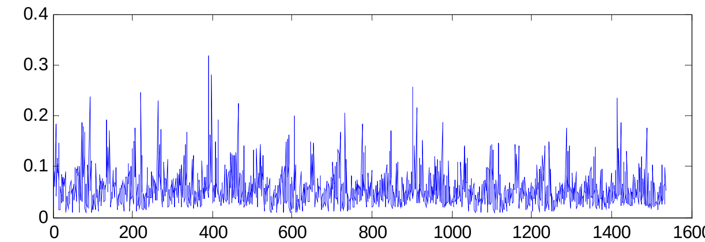


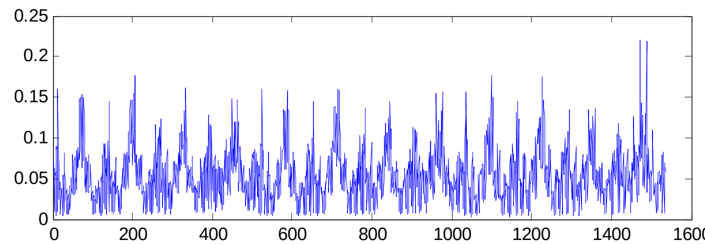


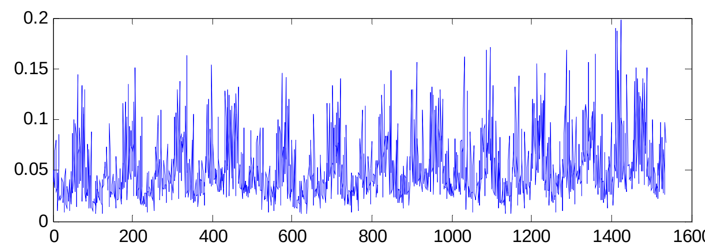


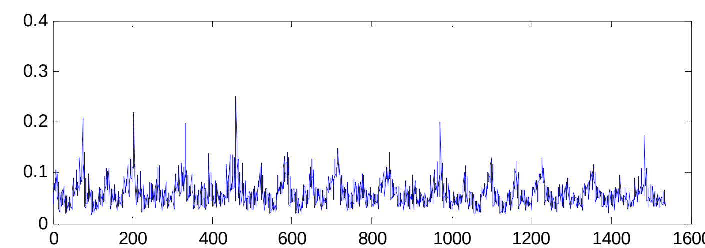


Descriptors for inside city images

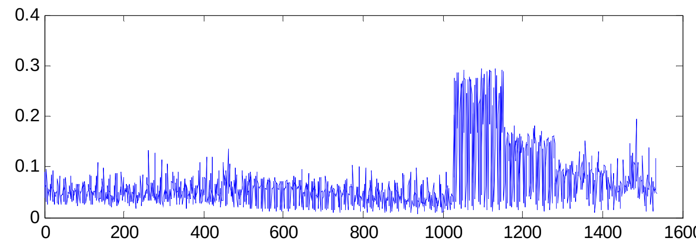


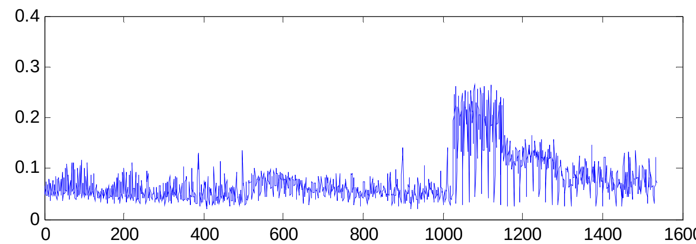


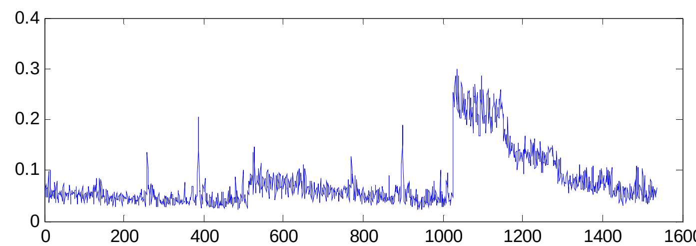


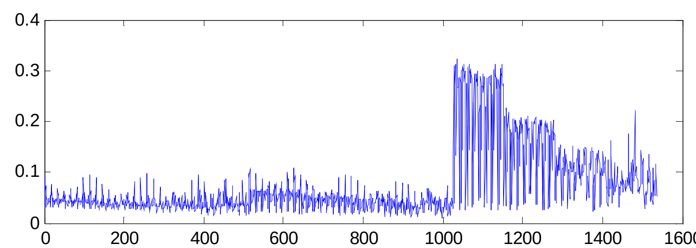


Descriptors for forest images

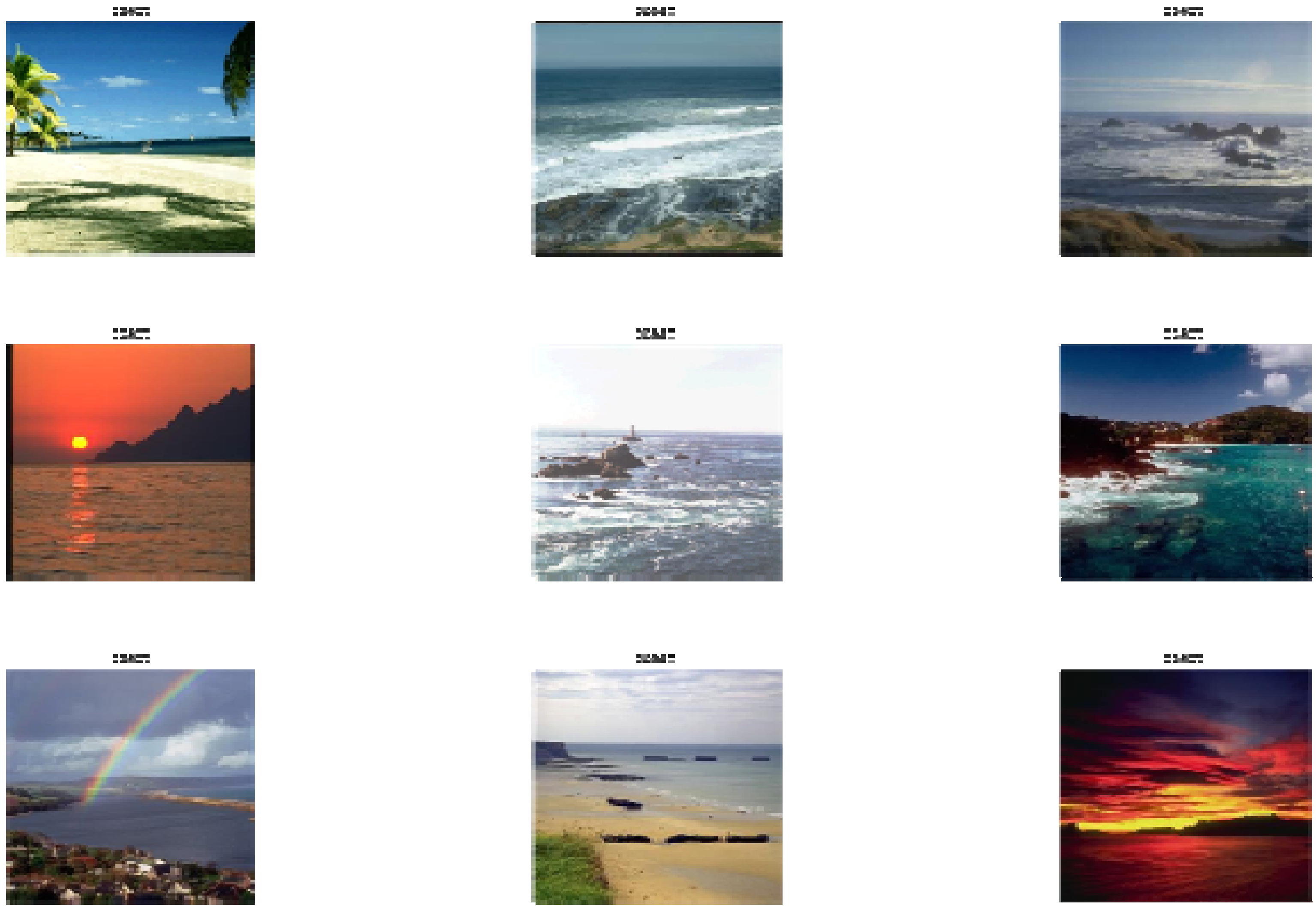


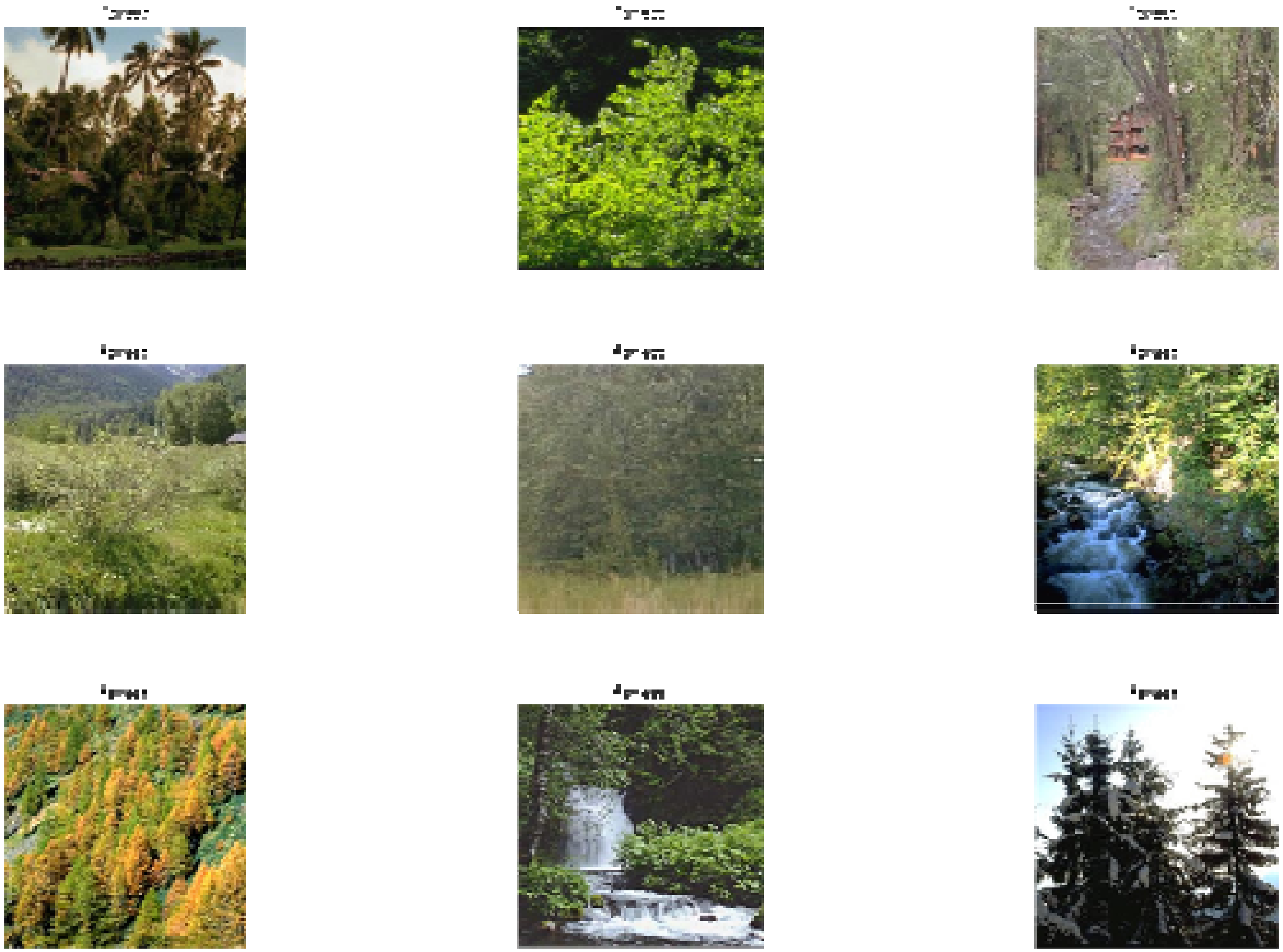


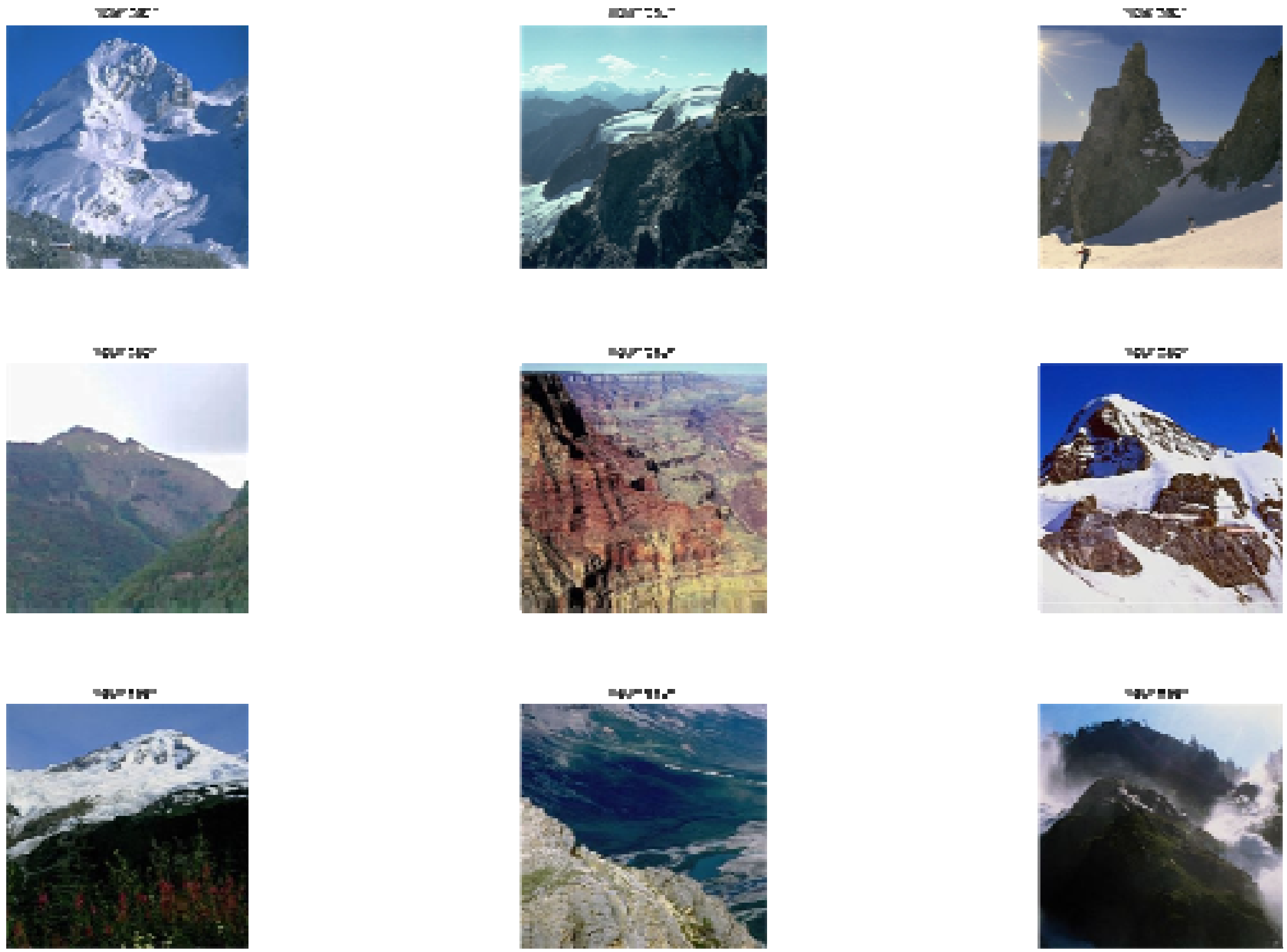


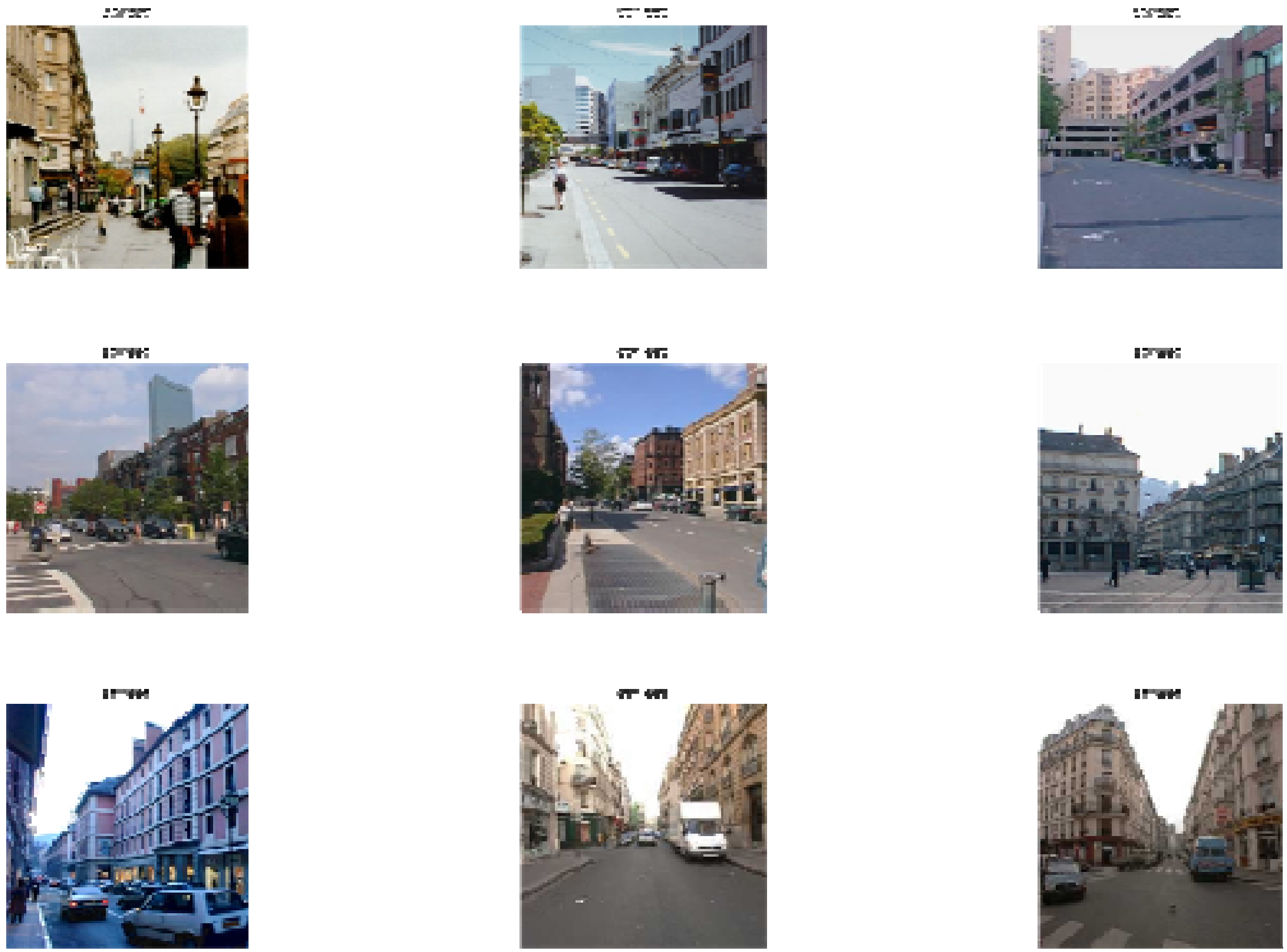


SOME EXAMPLES OF THE CATEGORIES









Adding the rows of the RBF matrix

OUTLINERS AND INSIDERS

Least and most highway-like highways

Least and most open country-like open countries

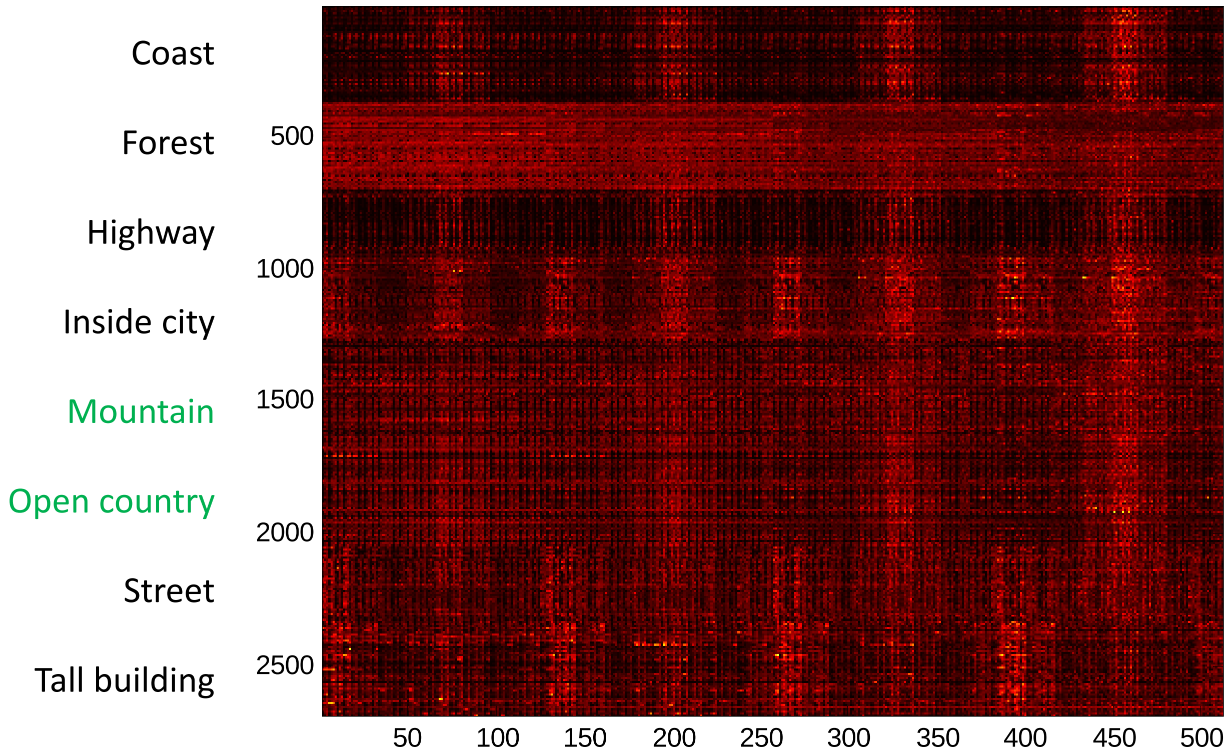
Least and most coast-like coasts

SIMULATION: SCENE RECOGNITION

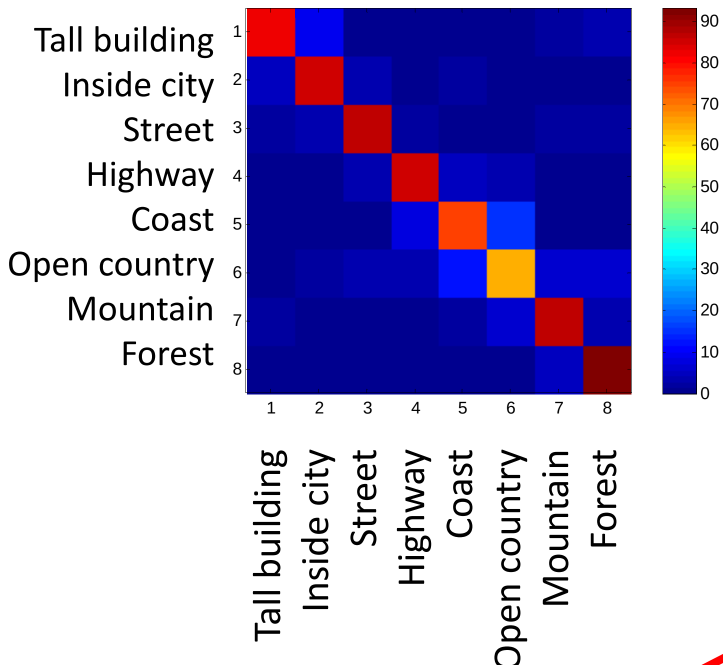
Parameters

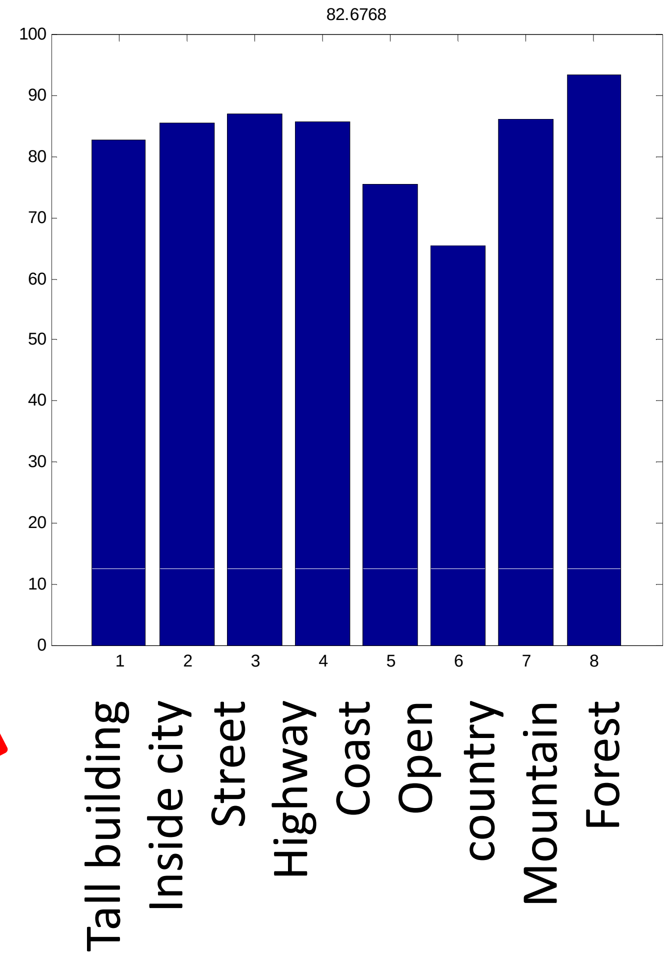
- 8 categories.
- 336 image in each category.
- 100 image from each category for training.
- The rest of the images are test images.

Map of all the gist descriptors



Performance





Average: 82.6768%

Confusion matrix

	Tallbuilding	In the cty	Street	Highway	Coast	Open country	Mountain	Forest
Tall building	82.8125	10.1563	0.3906	0.3906	0.3906	0	2.7344	3.125
In the city	5.2885	85.5769	4.3269	0.4808	1.9231	0.9615	0	1.4423
Street	2.0833	4.1667	86.9792	2.0833	0	0.5208	2.0833	2.0833
Highway	0	1.25	3.125	85.625	5	3.75	1.25	0
Coast	0	0	0	7.3077	75.3846	15.7692	1.1538	0.3846
Open country	0	1.9355	3.871	3.5484	11.9355	65.4839	6.129	7.0968
Mountain	1.8248	0.365	0.7299	0.365	1.4599	5.8394	86.1314	3.2847
Forest	0.4386	0	0.8772	0	0.8772	0	4.386	93.4211

Wrong classified images

insidecity

tallbuilding

insidecity

insidecity

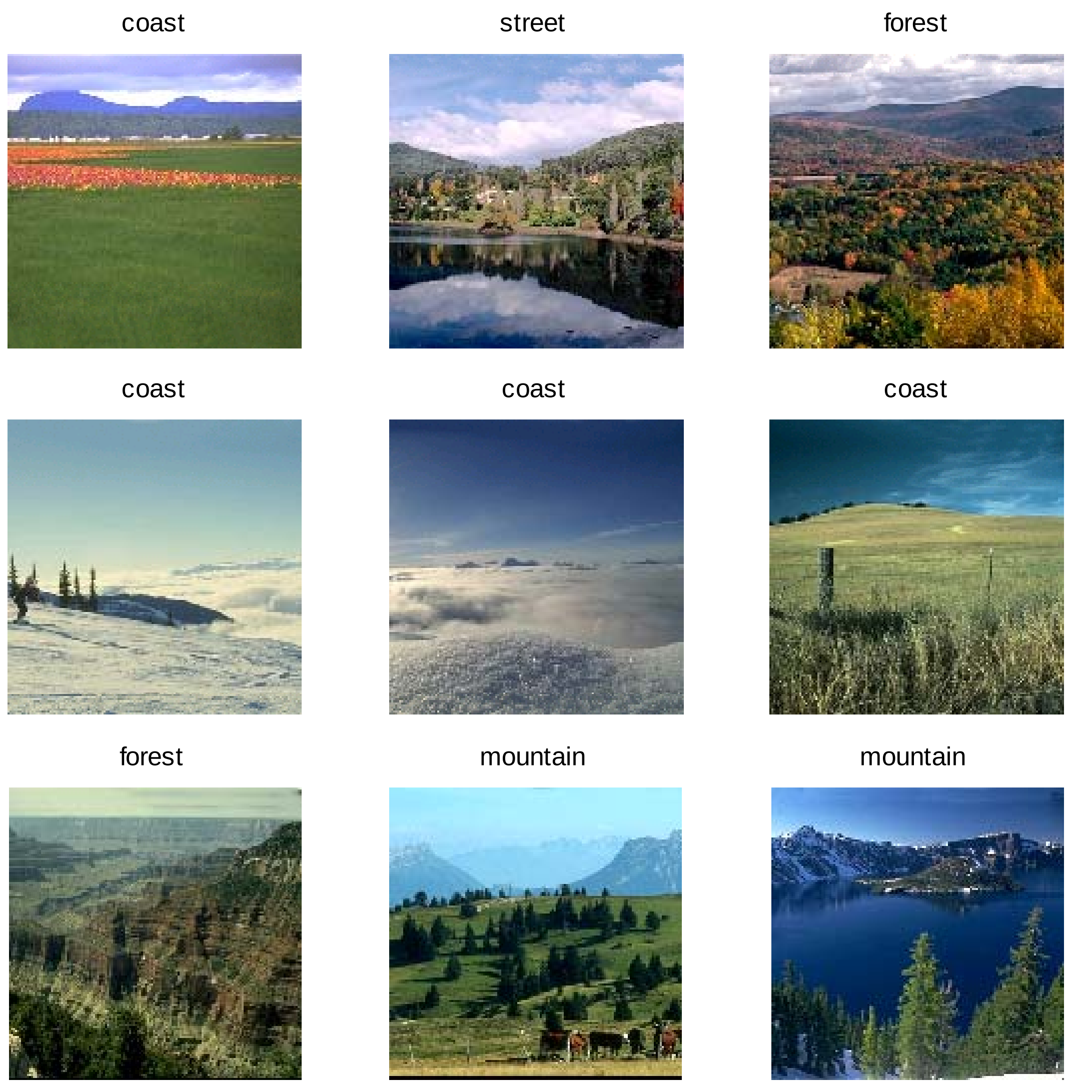
insidecity

mountain

insidecity

insidecity

forest


coast
street
forest
coast
coast
coast
forest
mountain
mountain

What is wrong with open country ?






Coast




How to fix for that

- Using color
- Fuzzy clustering (%40 open country, %30 mountain, …)

Color?

In this code:

Img = mean(img,3);

Rgb2gray:

% Calculate transformation matrix

T = inv([1.0 0.956 0.621; 1.0 -0.272 -0.647; 1.0 -1.106 1.703]);

coef = T(1,:)‘;

L = 0.2989 *R + 0.5870*G+ 0.1140*B

The color of truth is gray. -Andre Gide

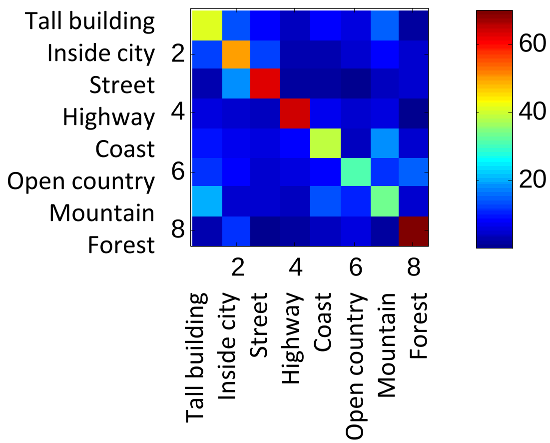
COLOR!

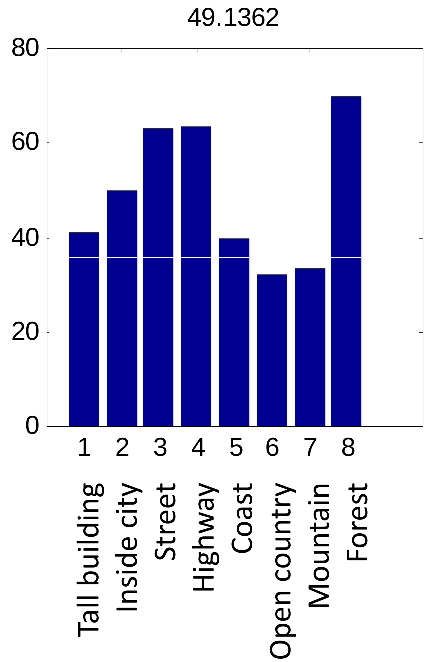
Color Descriptors

- Simple color histogram.
- RGB.
- 10 bins for each color red, green and blues

G=[.2 … .3 .4 … .6 .1 … .2]

Performance





Average is 49.1363%
Compare to chance
which is 12.5% ☺

Confusion Matrix

	Tallbuilding	In the cty	Street	Highway	Coast	Open country	Mountain	Forest
Tall building	41.0156	13.6719	8.2031	4.2969	8.2031	6.6406	15.625	2.3438
In the city	13.4615	50	12.5	2.8846	2.8846	4.8077	8.6538	4.8077
Street	3.6458	18.75	63.0208	2.0833	2.6042	0.5208	4.1667	5.2083
Highway	6.25	5.625	3.75	63.75	7.5	5	6.875	1.25
Coast	9.2308	7.3077	6.9231	8.0769	39.6154	4.6154	18.8462	5.3846
Open country	11.9355	9.0323	5.1613	6.7742	8.3871	31.9355	11.6129	15.1613
Mountain	20.073	5.8394	5.1095	4.3796	14.5985	10.5839	33.5766	5.8394
Forest	3.5088	11.4035	0.4386	1.7544	4.386	6.5789	1.7544	70.1754

Adding color

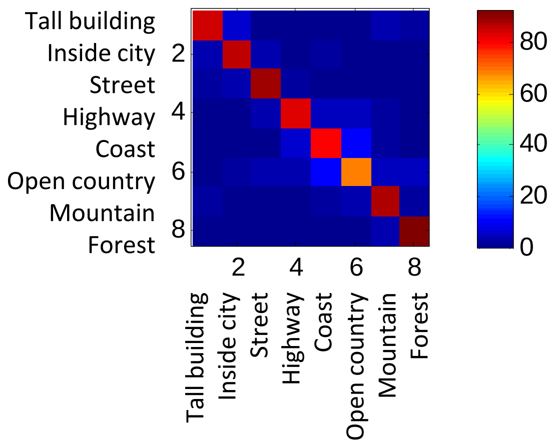
Simple color histogram

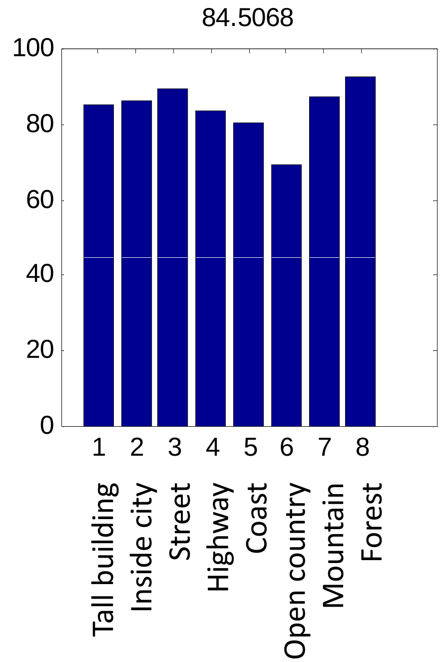
10 bins for each color Red, Green and Blue.

G = [.45 .34 …… .12 .2 … .3 .4 … .6 .1 … .2]

(using HSV was not much better)

Performance





Average is 84.5068
Compare to 82.6768

Almost 2% increase

YESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS!

Confusion matrices

Black & white

	Tallbuilding	In the cty	Street	Highway	Coast	Open country	Mountain	Forest
Tall building	82.8125	10.1563	0.3906	0.3906	0.3906	0	2.7344	3.125
In the city	5.2885	85.5769	4.3269	0.4808	1.9231	0.9615	0	1.4423
Street	2.0833	4.1667	86.9792	2.0833	0	0.5208	2.0833	2.0833
Highway	0	1.25	3.125	85.625	5	3.75	1.25	0
Coast	0	0	0	7.3077	75.3846	15.7692	1.1538	0.3846
Open country	0	1.9355	3.871	3.5484	11.9355	65.4839	6.129	7.0968
Mountain	1.8248	0.365	0.7299	0.365	1.4599	5.8394	86.1314	3.2847
Forest	0.4386	0	0.8772	0	0.8772	0	4.386	93.4211

Colorful

	Tallbuilding	In the cty	Street	Highway	Coast	Open country	Mountain	Forest
Tall building	85.1562	8.2031	0.7812	0.3906	0.7812	0	3.125	1.5625
In the city	4.3269	87.5	2.4038	1.4423	2.4038	0.9615	0	0.9615
Street	3.125	4.1667	89.0625	1.0417	0	0	1.5625	1.0417
Highway	0	0	3.75	84.375	7.5	3.75	0.625	0
Coast	0	0	0	7.6923	79.2308	11.1538	1.5385	0.3846
Open country	0	1.6129	3.5484	3.5484	12.2581	67.7419	6.129	5.1613
Mountain	1.0949	0	0.7299	0	2.5547	2.5547	90.5109	2.5547
Forest	0.4386	0	0.4386	0	0.8772	2.193	3.5088	92.5439

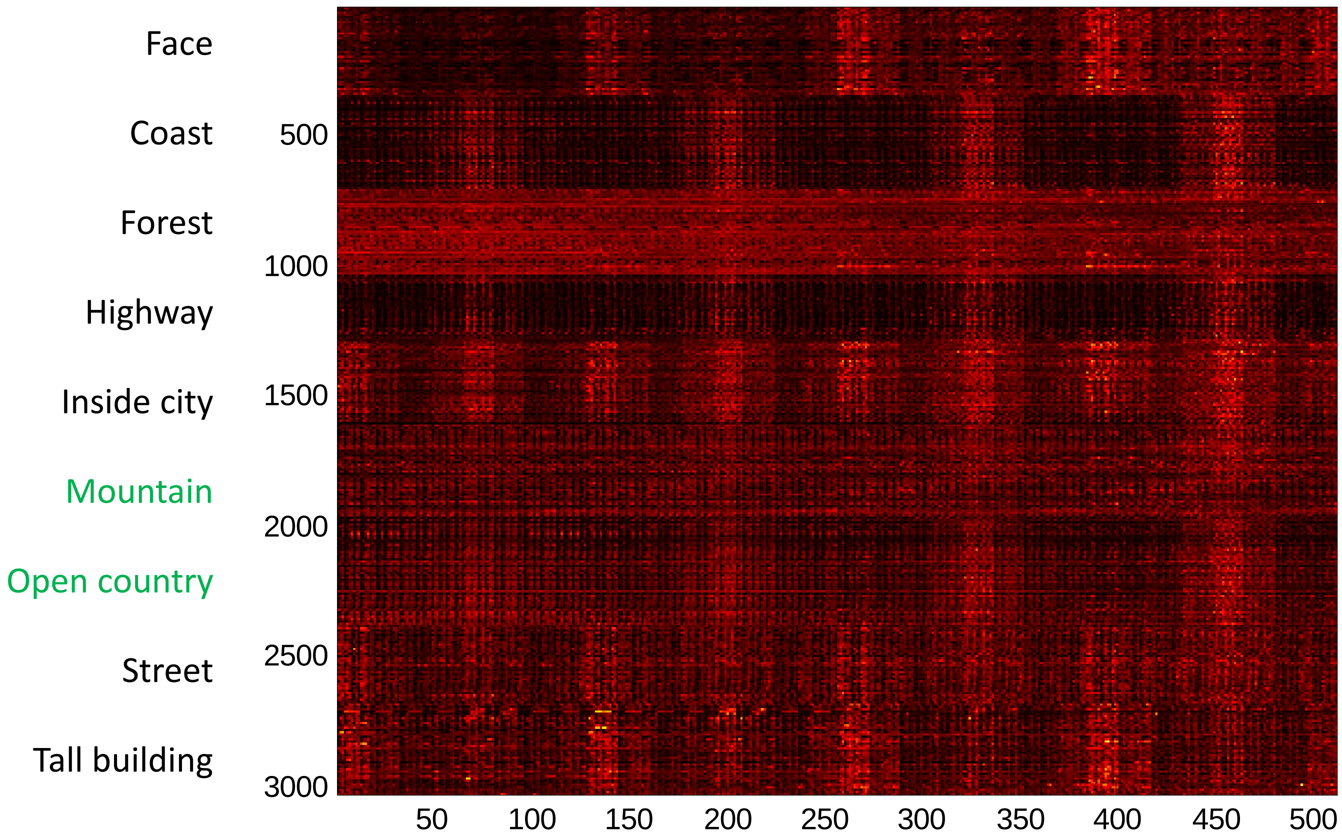
Why God? why?

WHAT HAPPENS IF WE INTRODUCE A RANDOM CATEGORY?

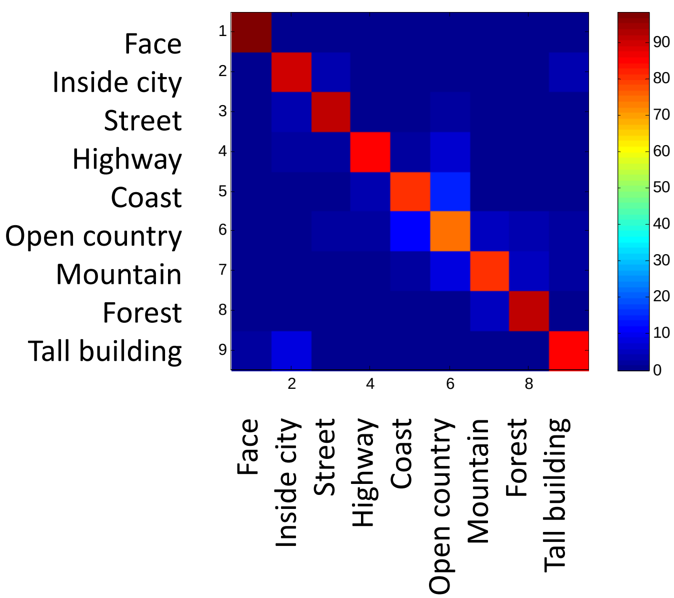
Parameters

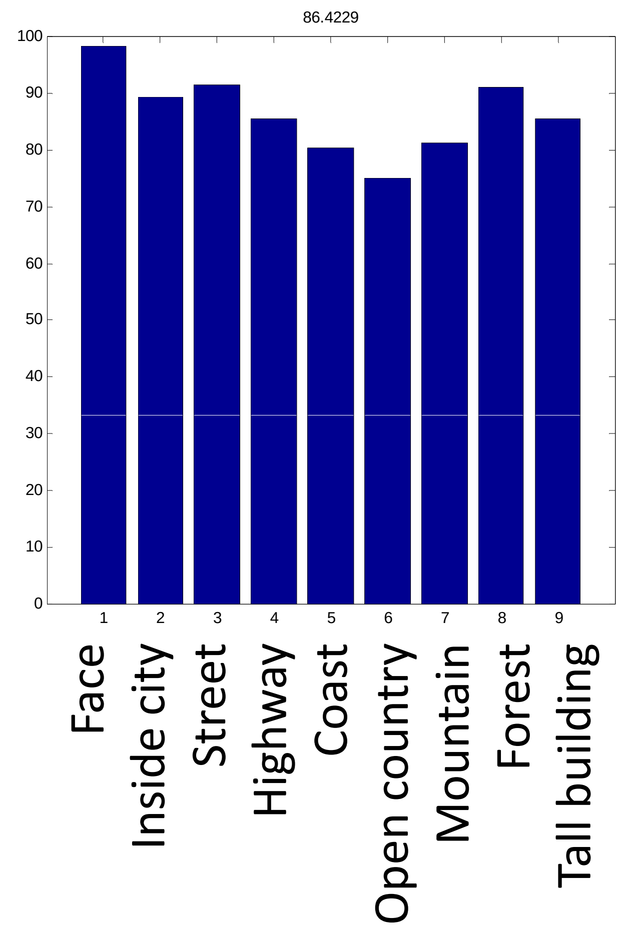
- 9 categories (including faces).
- 336 image in each category.
- 150 image from each category for training.
- The rest of the images are test images.

Map of all the gist descriptors



Performance





Confusion matrix

	Face	In the cty	Street	Highway	Coast	Open country	Mountain	Forest	Tallbuilding
Face	98.3871	1.0753	0	0	0	0	0.5376	0	0
In the city	0	89.2405	3.7975	0.6329	0.6329	1.2658	0	0	4.4304
Street	0.7042	3.5211	91.5493	1.4085	0	2.1127	0	0	0.7042
Highway	0.9091	1.8182	1.8182	85.4545	1.8182	6.3636	0.9091	0.9091	0
Coast	0	0	0	3.8095	80.4762	15.2381	0.4762	0	0
Open country	0	0.7692	2.3077	1.5385	11.1538	75	4.6154	3.0769	1.5385
Mountain	0	0	0	0.4464	2.6786	8.0357	81.25	5.3571	2.2321
Forest	0.5618	0	1.1236	0	0	1.1236	5.0562	91.0112	1.1236
Tall building	1.9417	8.7379	1.4563	0	0.4854	0.4854	1.4563	0	85.4369

Performance discussion

- Why does it perform this good for faces?
  - Could be because faces are so different from the rest.

Now, is this good or bad?

The method is specifically designed for scenes while the best performance is achieved with faces.

More is more.

YET ANOTHER CATEGORY

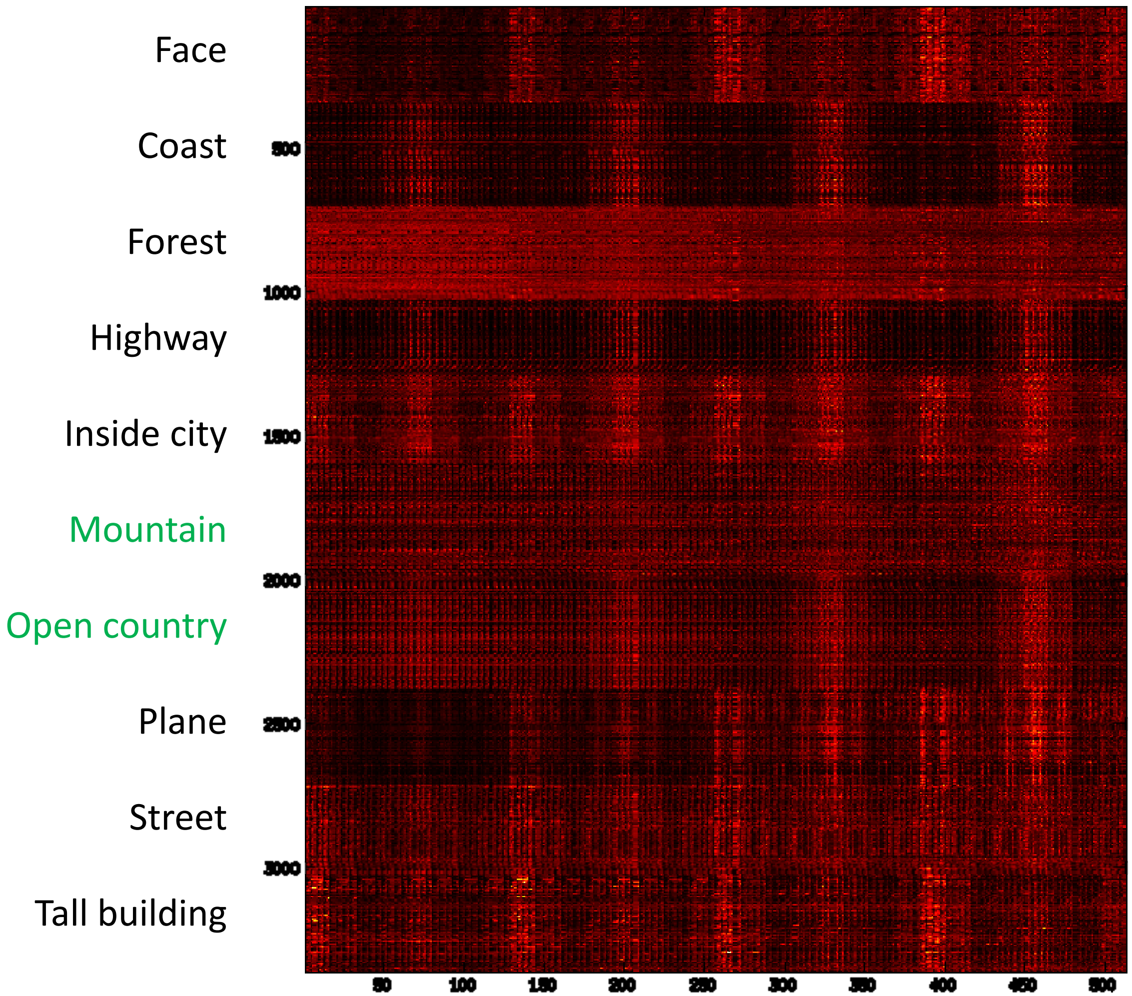
A more similar category

- What if we add another category which is more similar to the rest of the scenes.
- Plane images are similar to coast and open country images in that there is a strong horizontal element.
- Their background is often similar to other scenes.

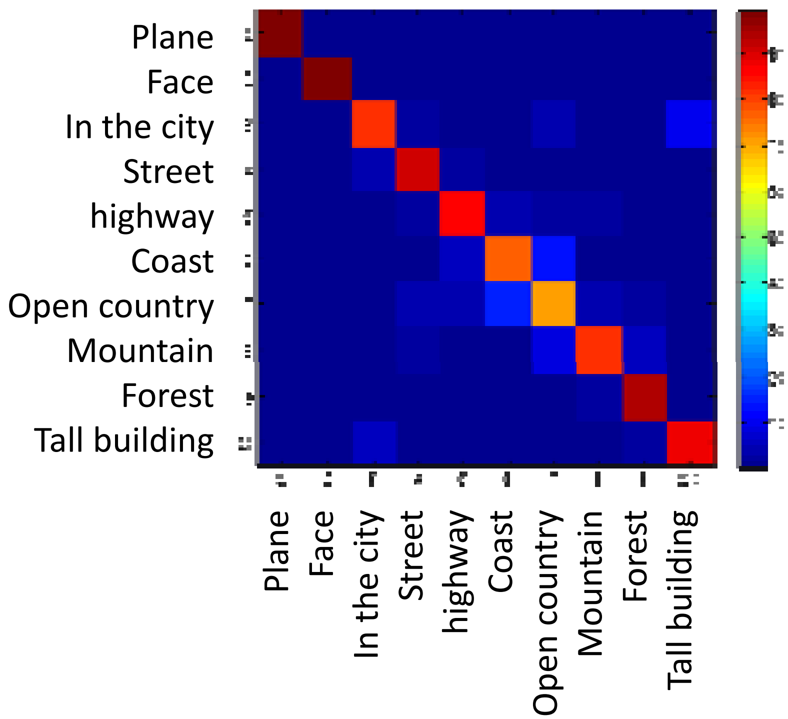
Parameters

- 10 categories (including planes and faces).
- 336 image in each category.
- 150 image from each category for training.
- The rest are used for testing.

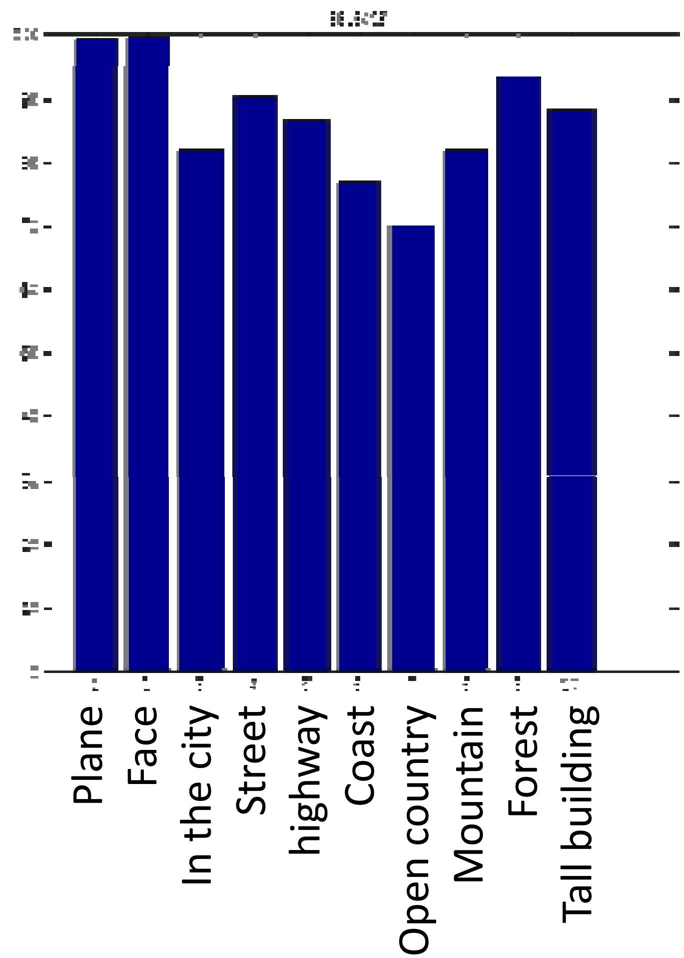
All the gist representations



Results



Plane
Face
In the city
Street
highway
Coast
Open country
Mountain
Forest
Tall building



Plane
Face
In the city
Street
highway
Coast
Open country
Mountain
Forest
Tall building

Confusion matrix

	Plane	Face	In the cty	Street	Highway	Coast	Open cou	Mountai	Forest	Tallbuildi
Plane	98.9247	0	0	0	0	0	0	1.0753	0	0
Face	0	99.4624	0	0	0	0	0	0	0.5376	0
In the cit	0	0	81.6456	2.5316	0.6329	0.6329	3.7975	1.2658	0	9.4937
Street	0	0	3.5211	90.1408	2.8169	0	0.7042	1.4085	1.4085	0
Highway	0	0.9091	0.9091	1.8182	86.3636	4.5455	2.7273	1.8182	0	0.9091
Coast	0	0.4762	0.4762	0	6.1905	76.6667	13.8095	1.4286	0.9524	0
Open cou	0.3846	0	0.3846	3.4615	4.2308	14.2308	70	4.6154	1.9231	0.7692
Mountai	0.4464	0	0	1.7857	0.4464	1.3393	8.0357	81.6964	4.9107	1.3393
Forest	0	0	0	1.1236	0.5618	0	1.1236	2.809	93.2584	1.1236
Tall build	0	1.4563	5.3398	1.4563	0	0	0	1.4563	2.4272	87.8641

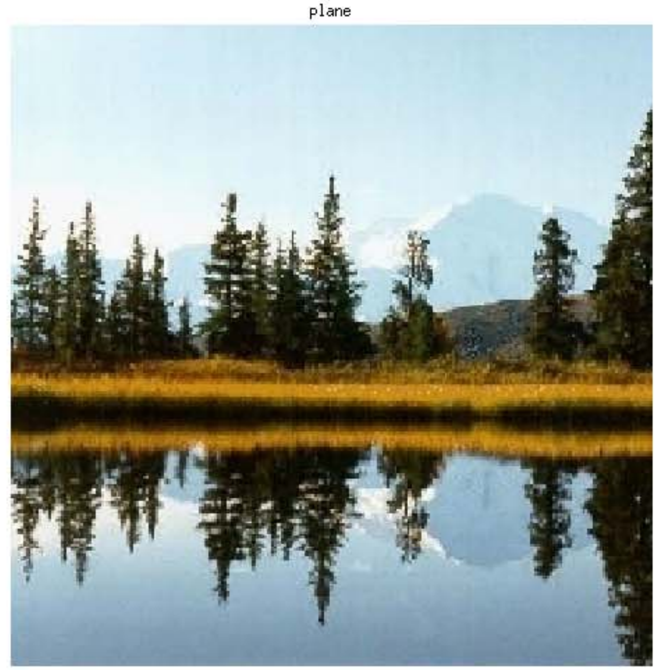
Wrong examples are cooler!

THE WRONG EXAMPLES IN EACH CATEGORY

Plane!

plane

plane



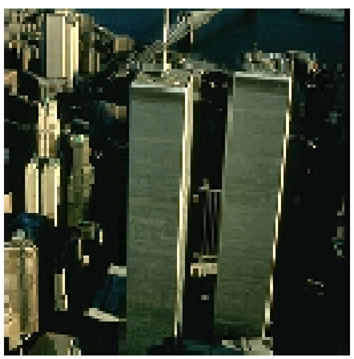
Face!

face

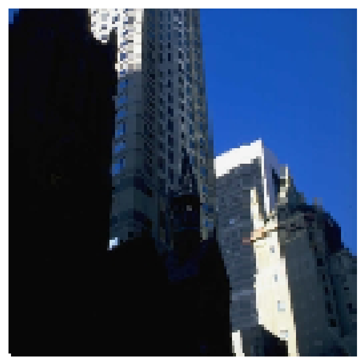
face

face

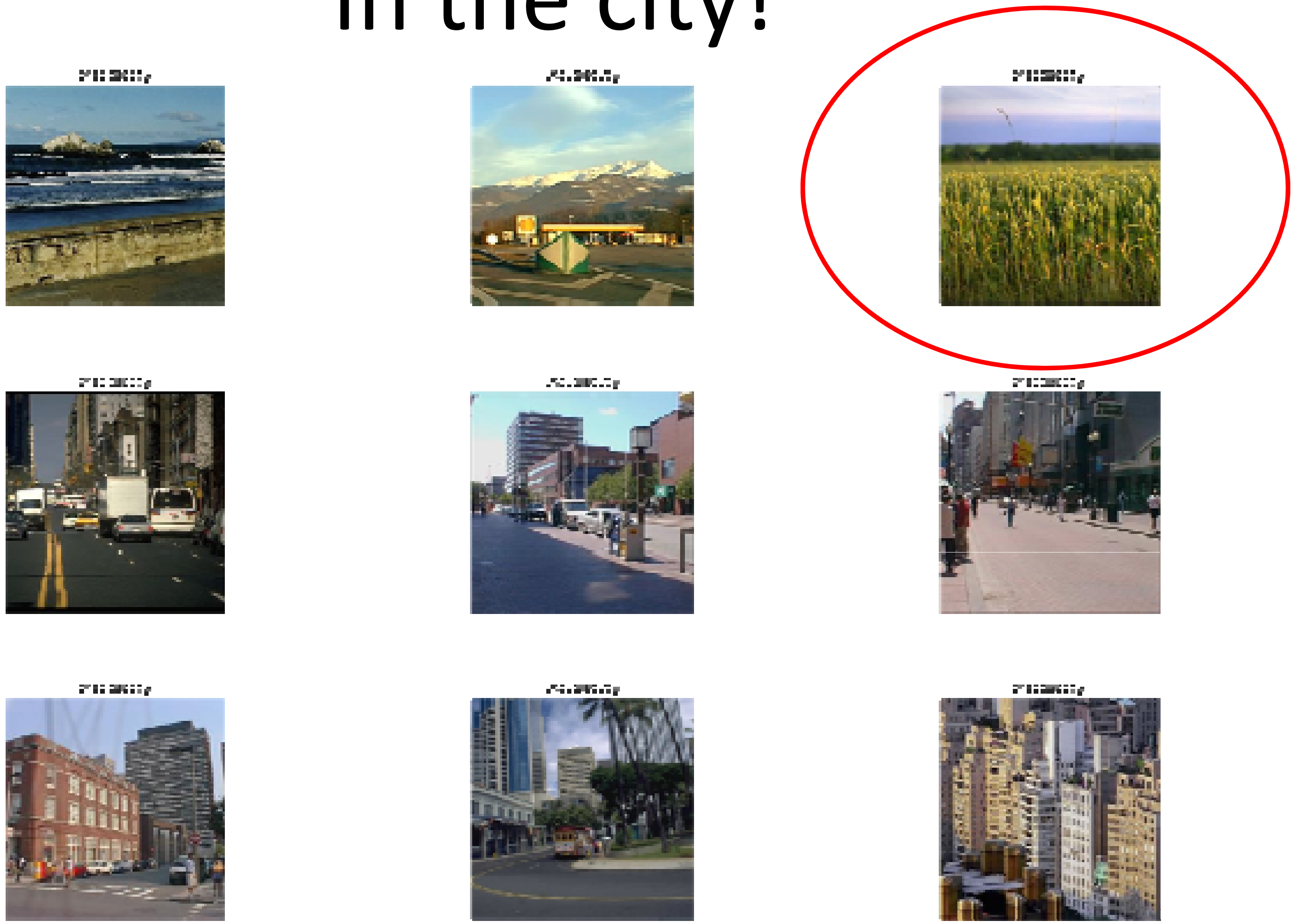
face



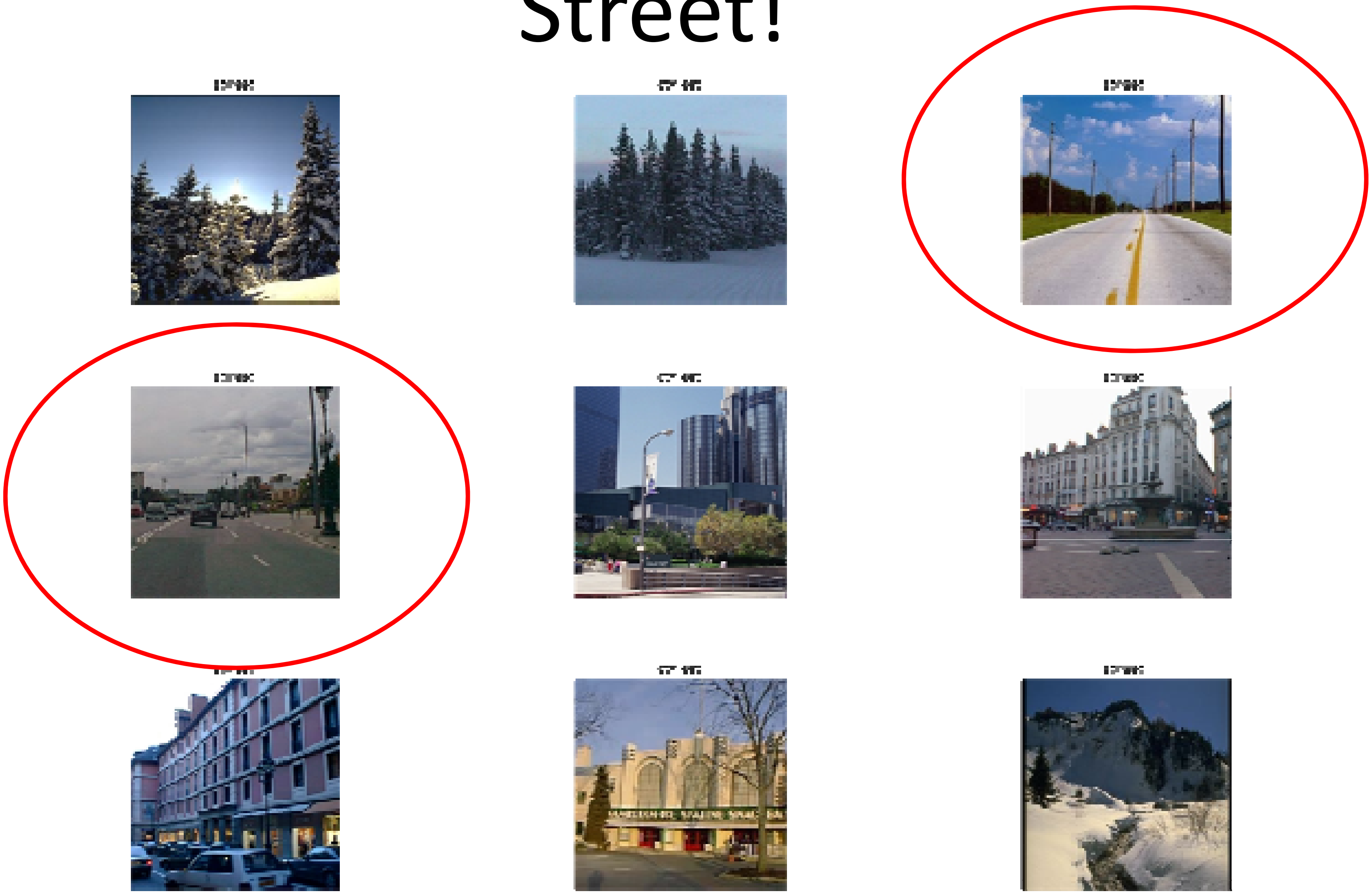
face



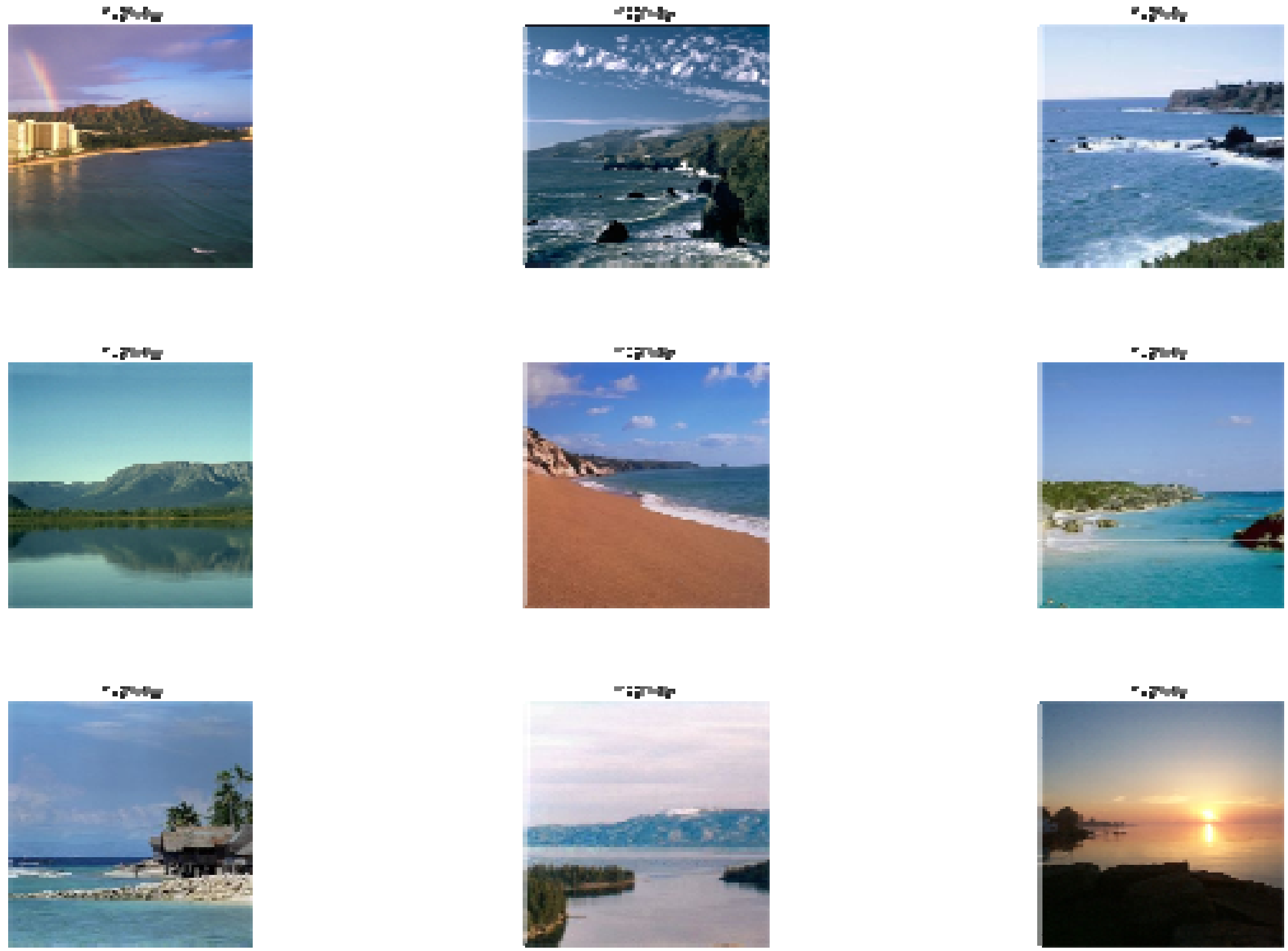
In the city!



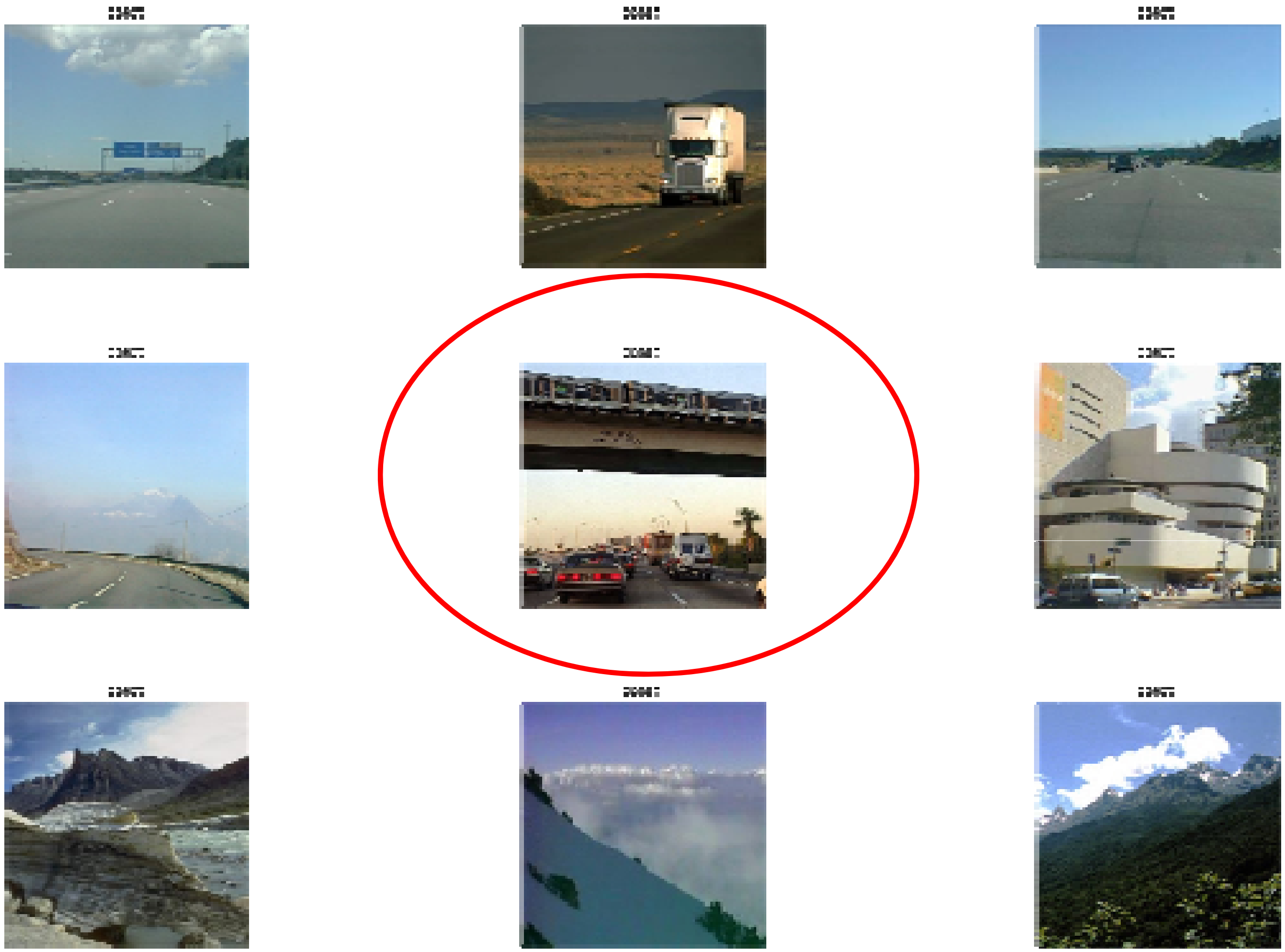
Street!



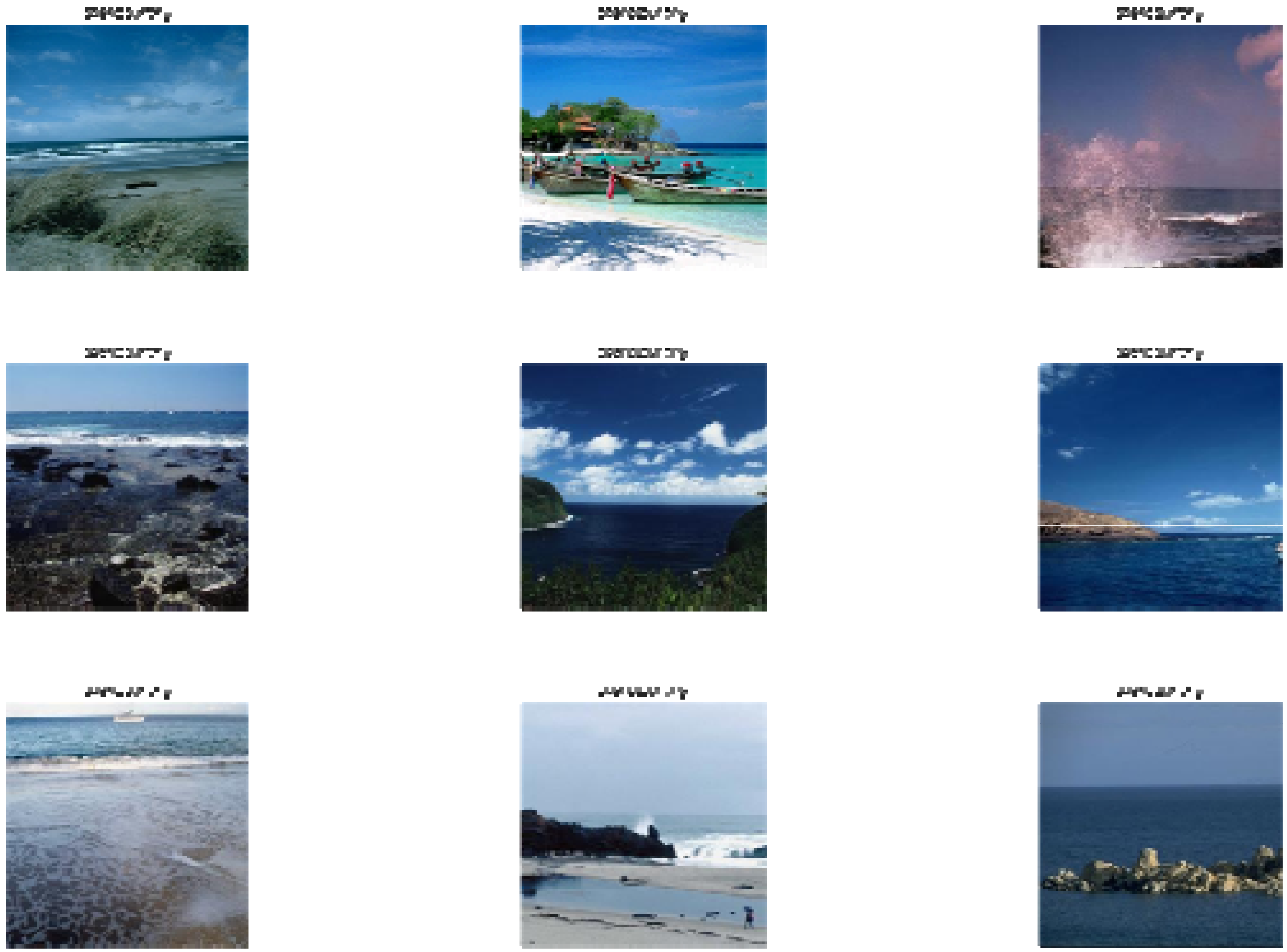
Highway!



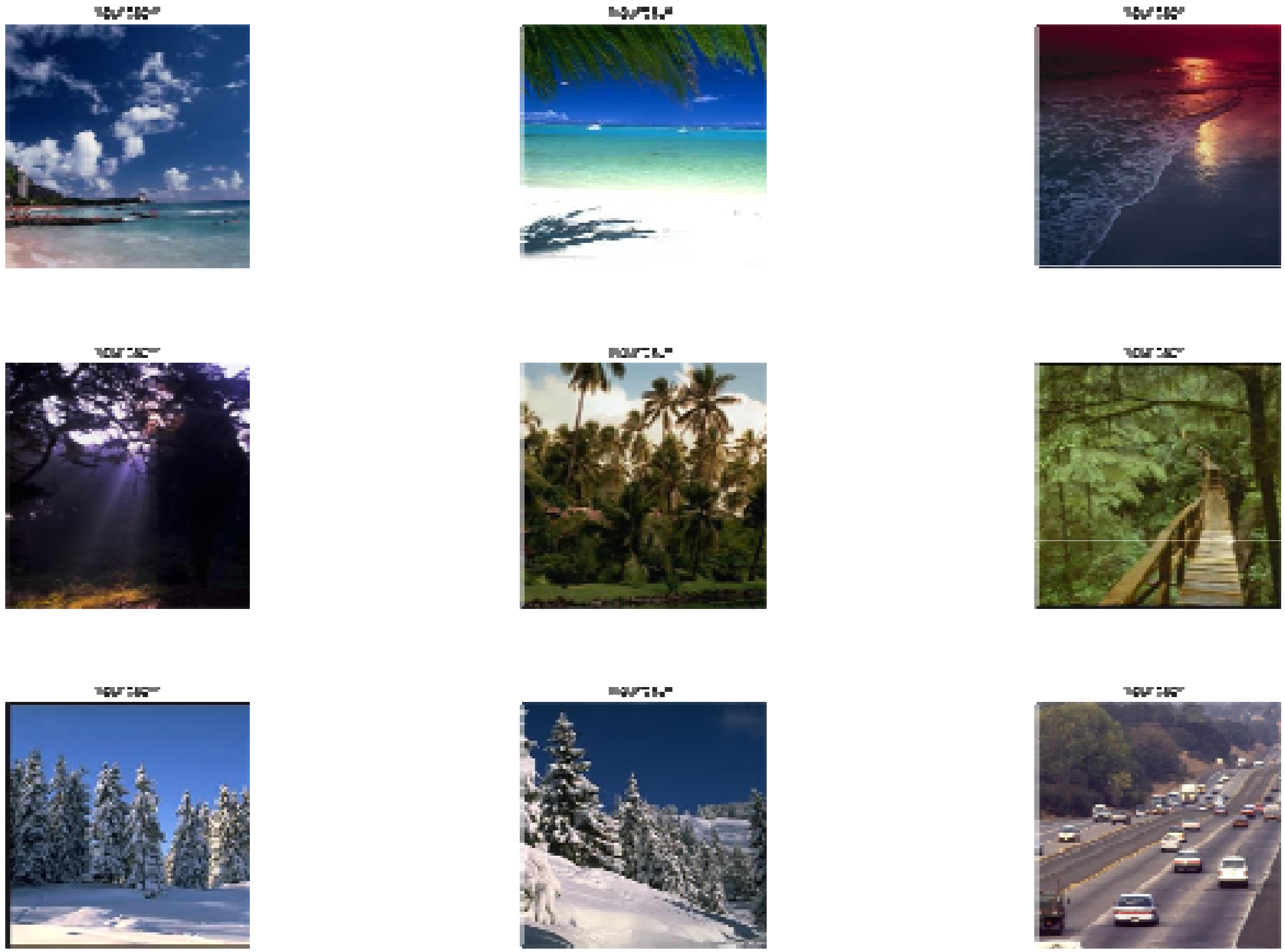
Coast!



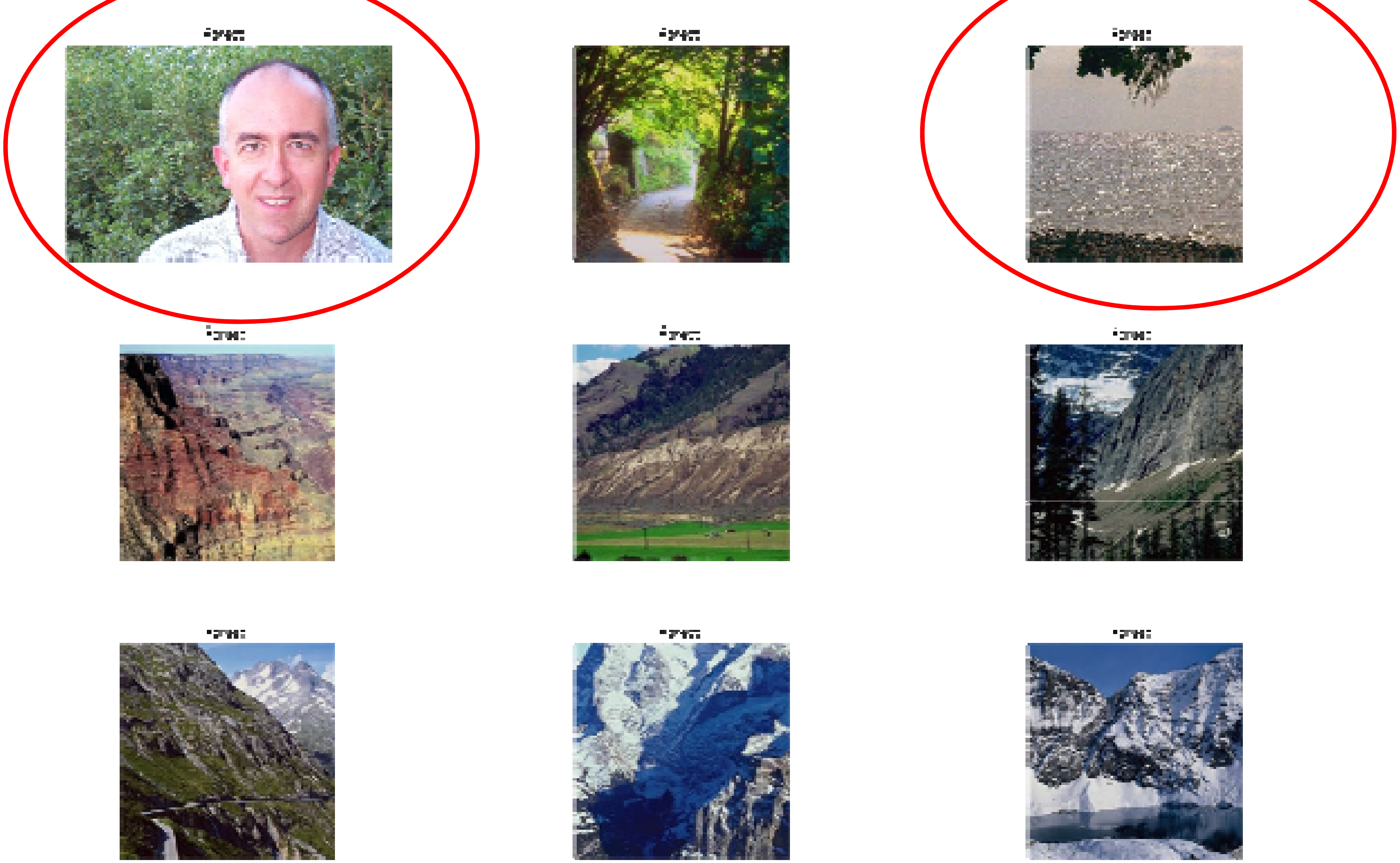
Open Country!



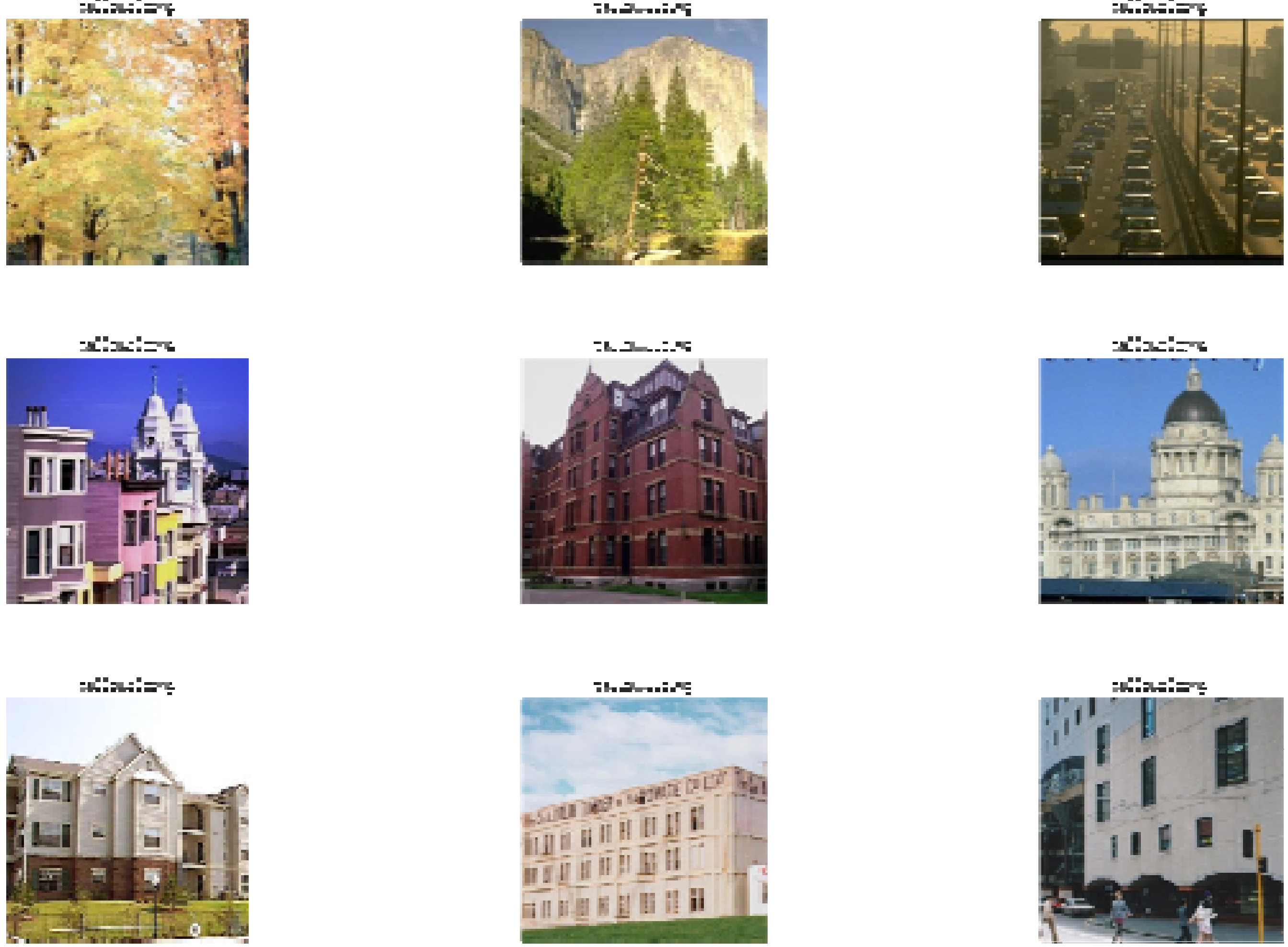
Mountain!



Forest!



Tall building!

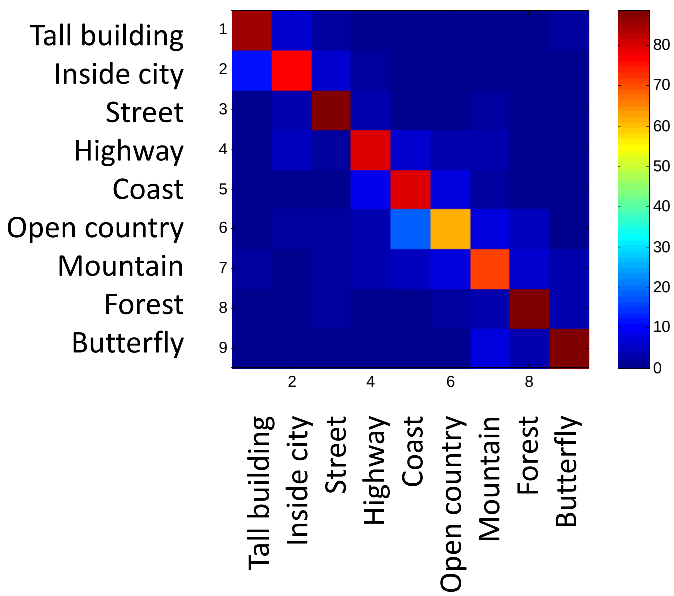


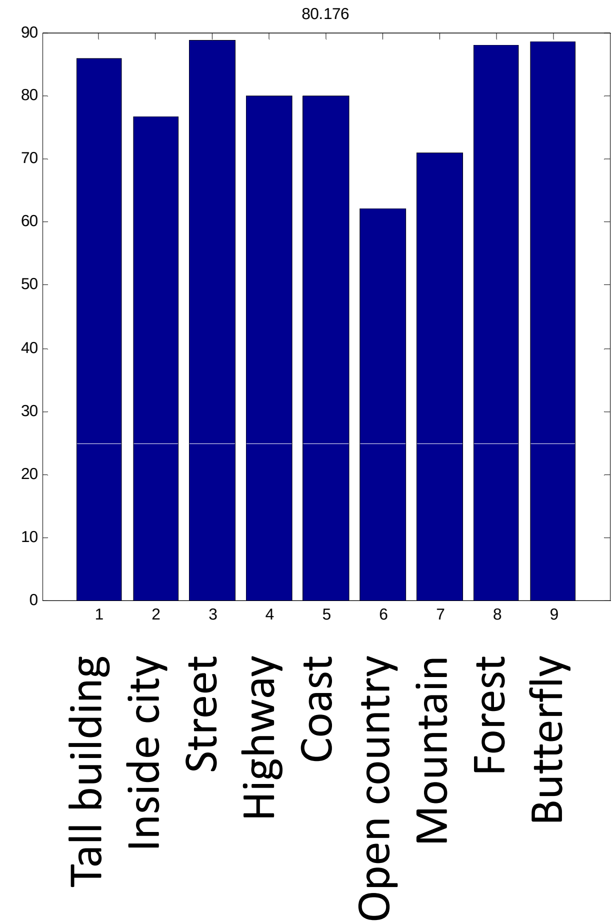
Break damn it. Break!

BUTTERFLIES

Butterflies!

Performance

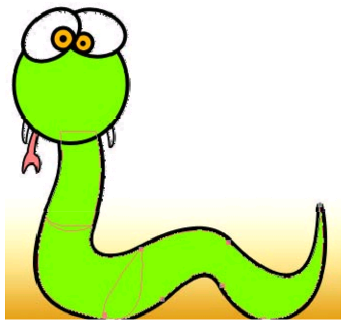


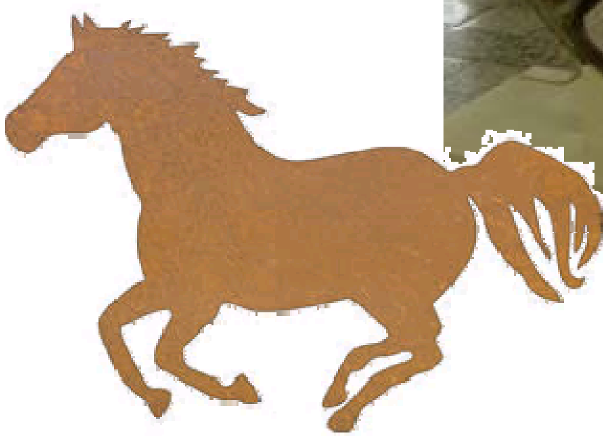


Off on a tangent!

A PHILOSOPHICAL NOTE

Nature knows best?





Still awake?

DISCUSSION

Links

- http://people.csail.mit.edu/torralba/code/spatialenvelope/

- Support vector machine toolbox:
  http://ida.first.fraunhofer.de/~anton/software.html

- Additional images:
  http://www.vision.caltech.edu/Image_Datasets/Caltech101/
  http://www.vision.caltech.edu/Image_Datasets/Caltech256/

Thank you!