Machine Learning and Computer Vision

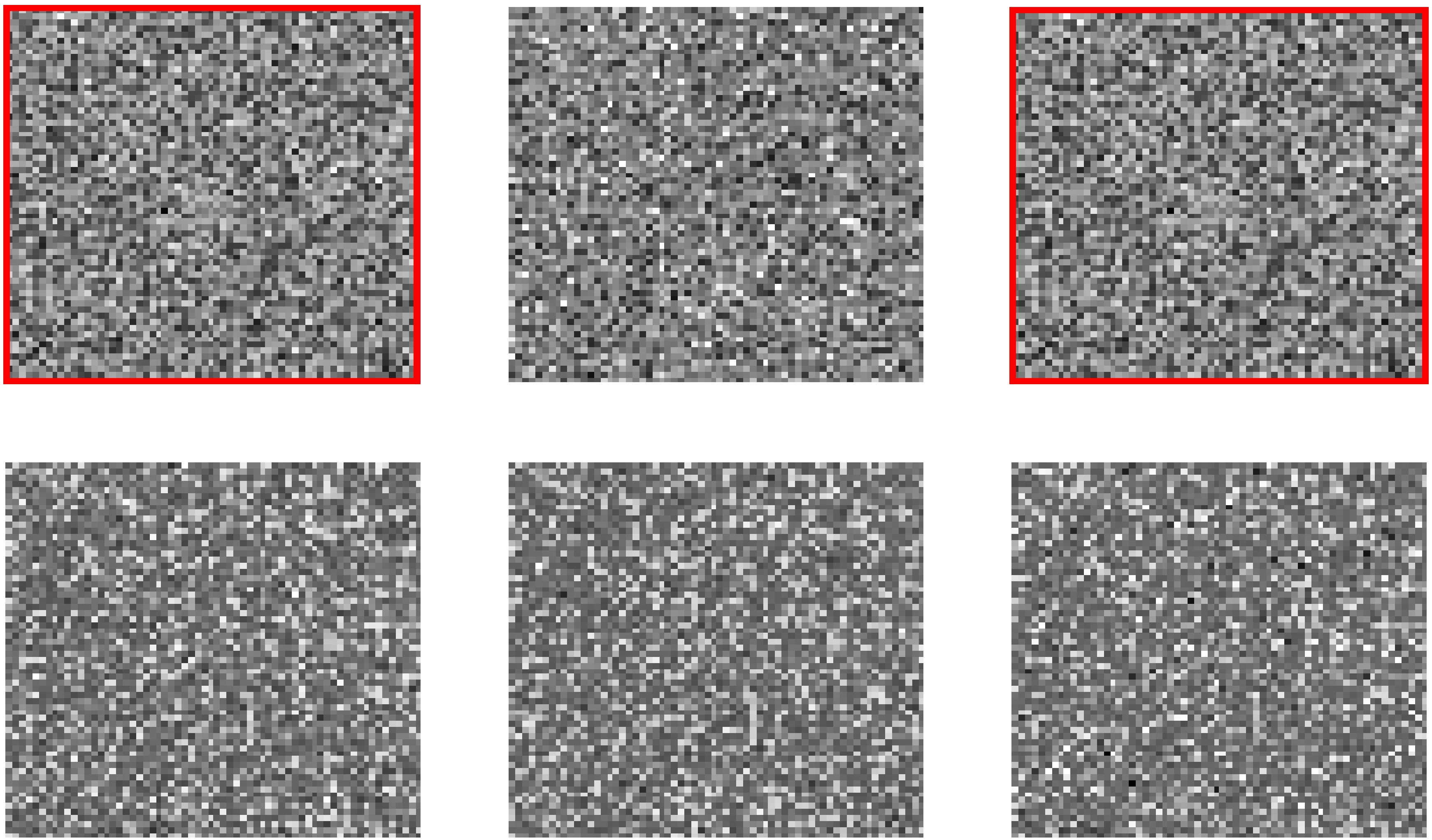
Ashish Agrawal
Department of Computer Science and Engineering

SIGML
20th March 2010

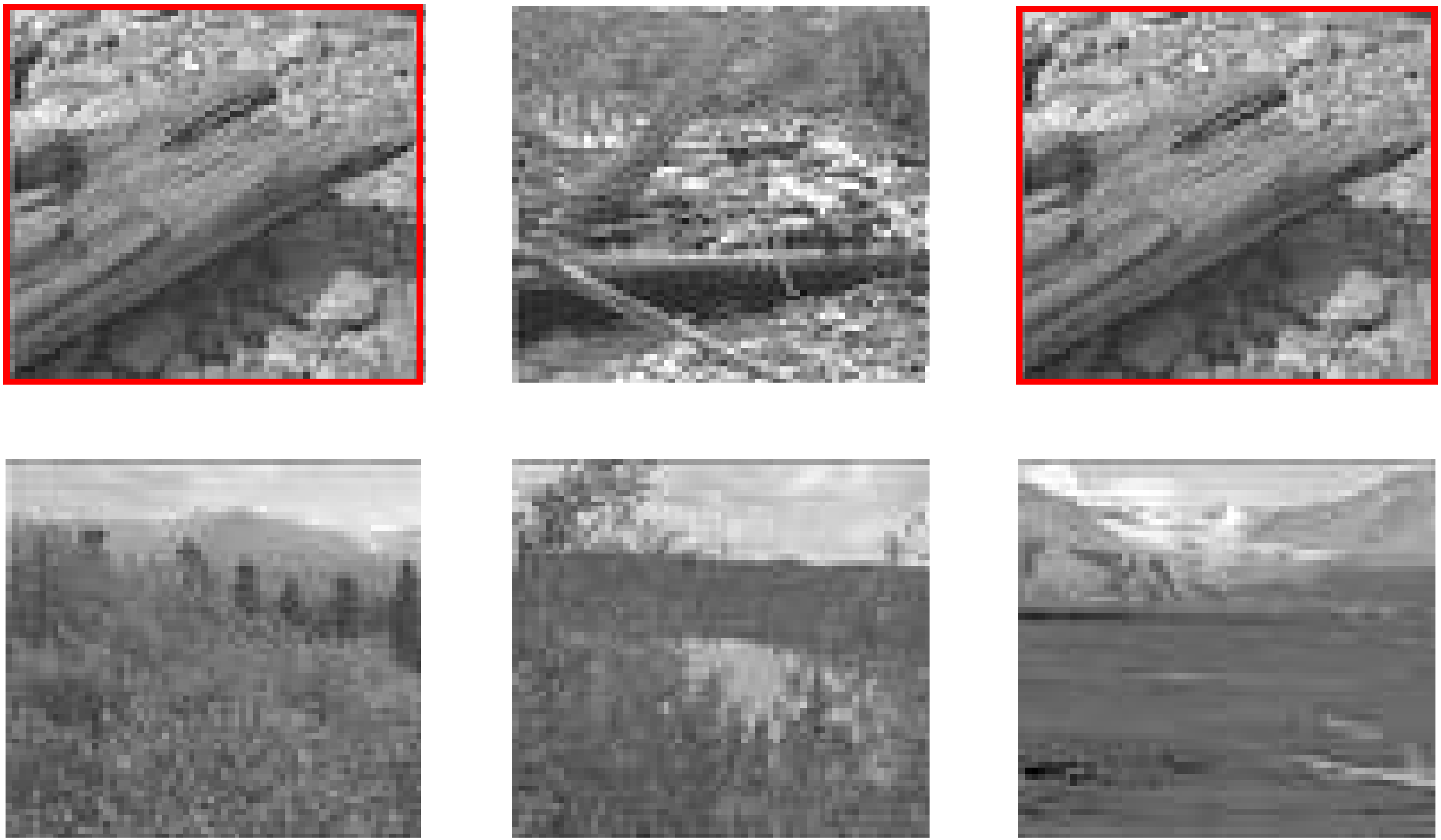
A glance at Computer Vision Problems

- Object Segmentation
- Object Recognition
- Image clustering
- Object tracking
- Denoising images/ Videos

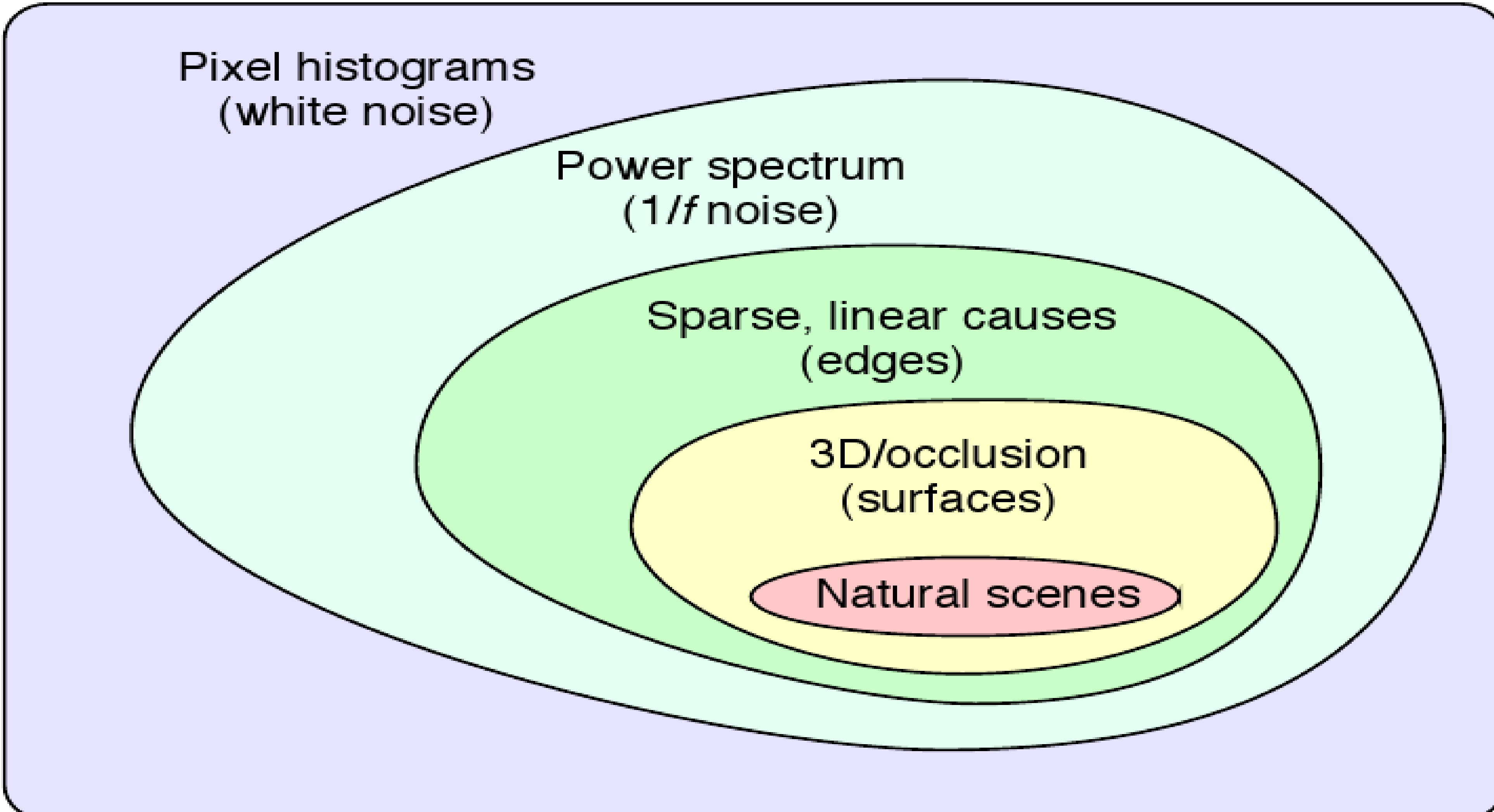
Images ??



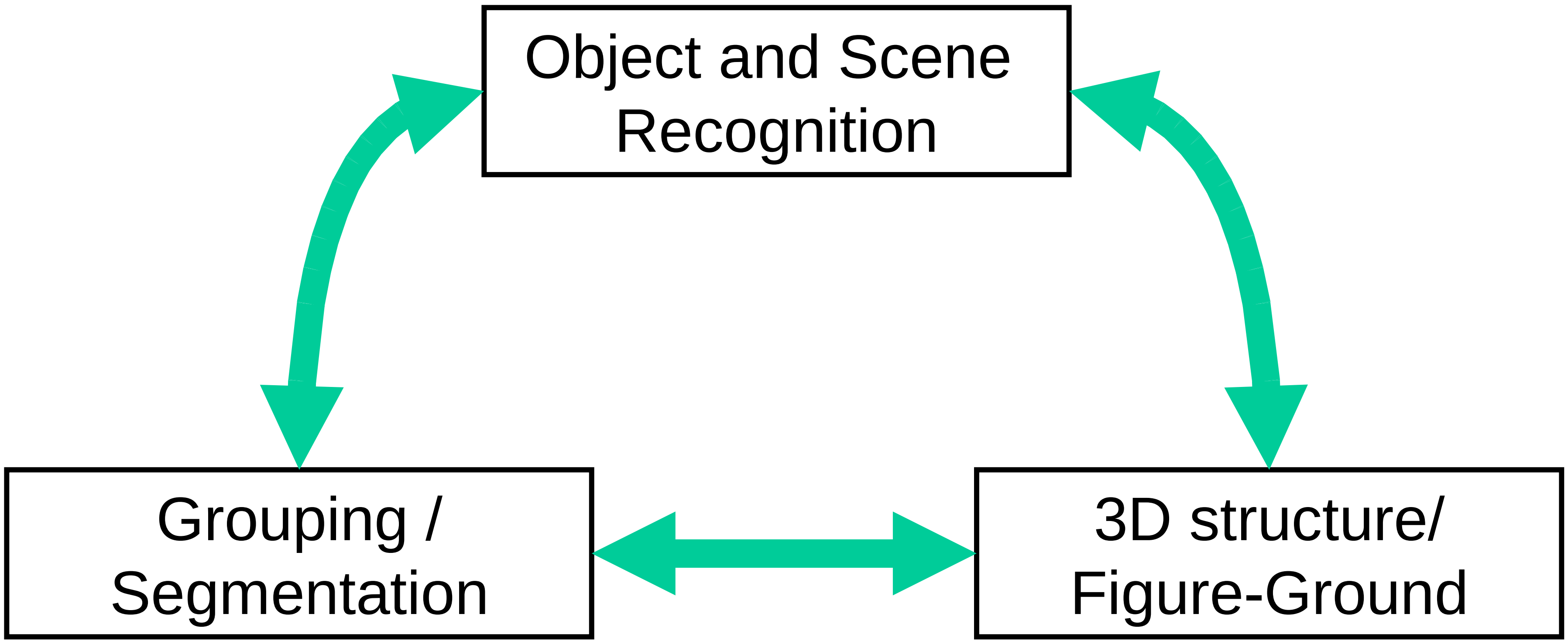
Natural Images



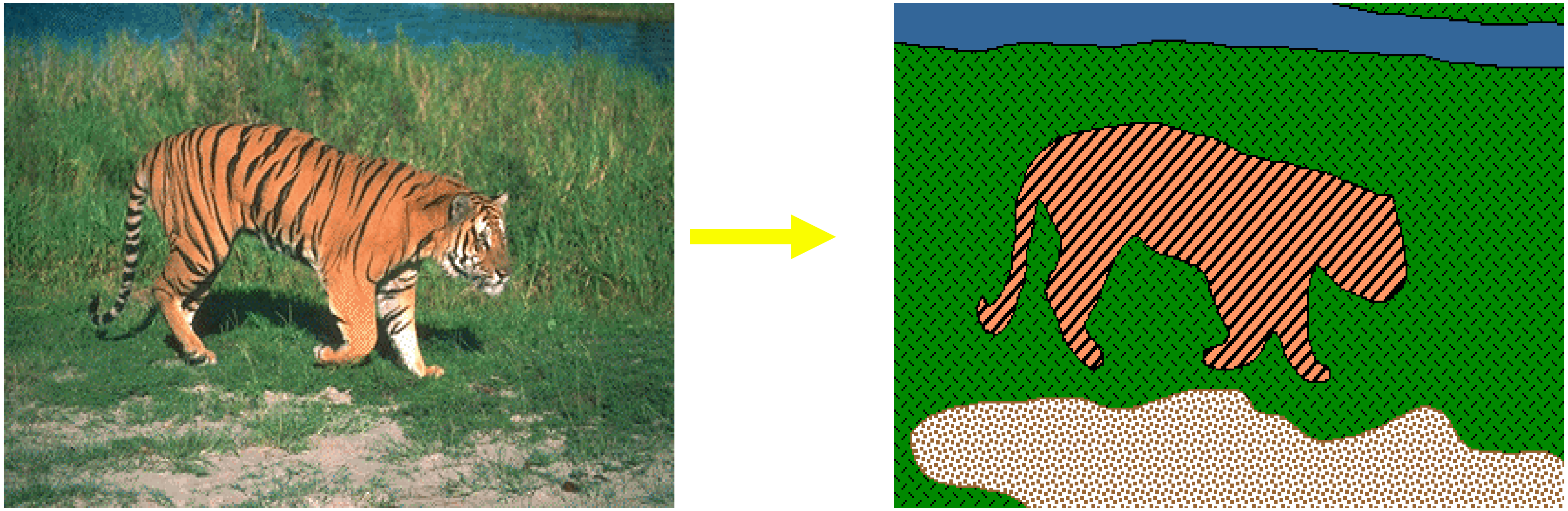
Where do Natural Images lie?



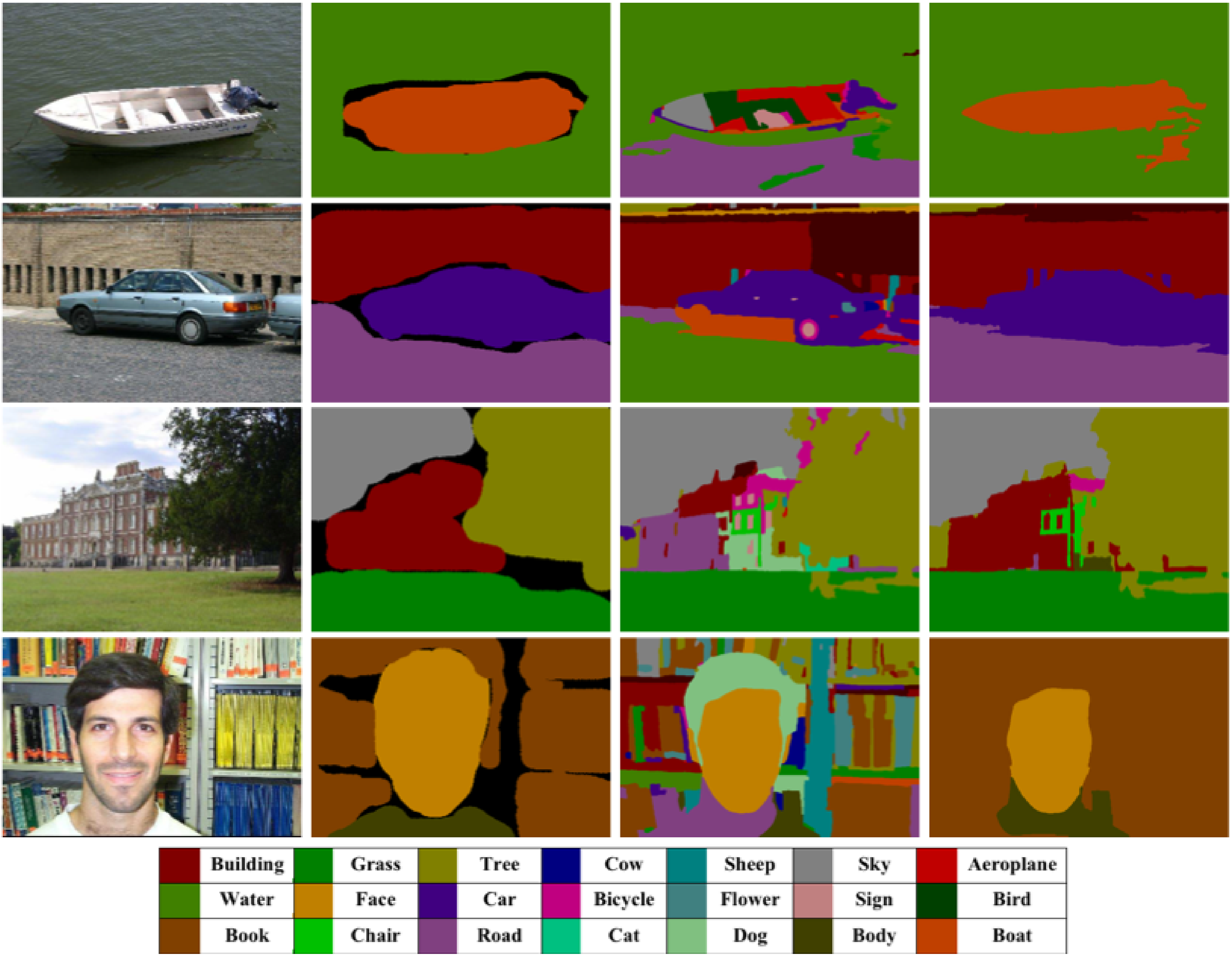
The central problems of images



Segmentation:
From images to objects



MSRC dataset

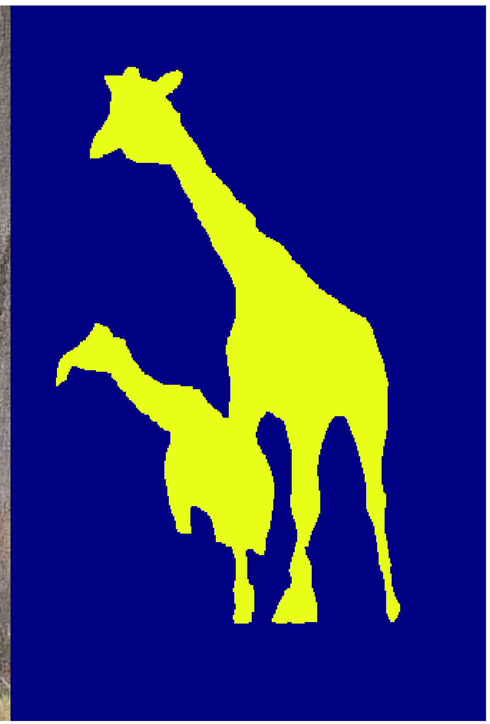


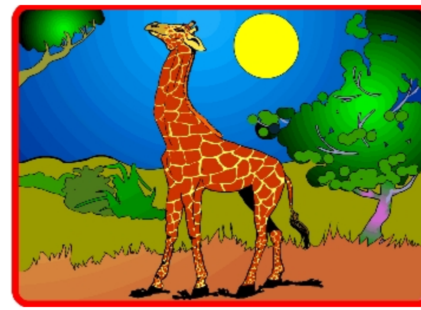
Detection and Segmentation: Bottles

Orig. Image Segmentation Orig. Image Segmentation

Detection and Segmentation: Giraffes

Orig. Image | Segmentation | Orig. Image | Segmentation

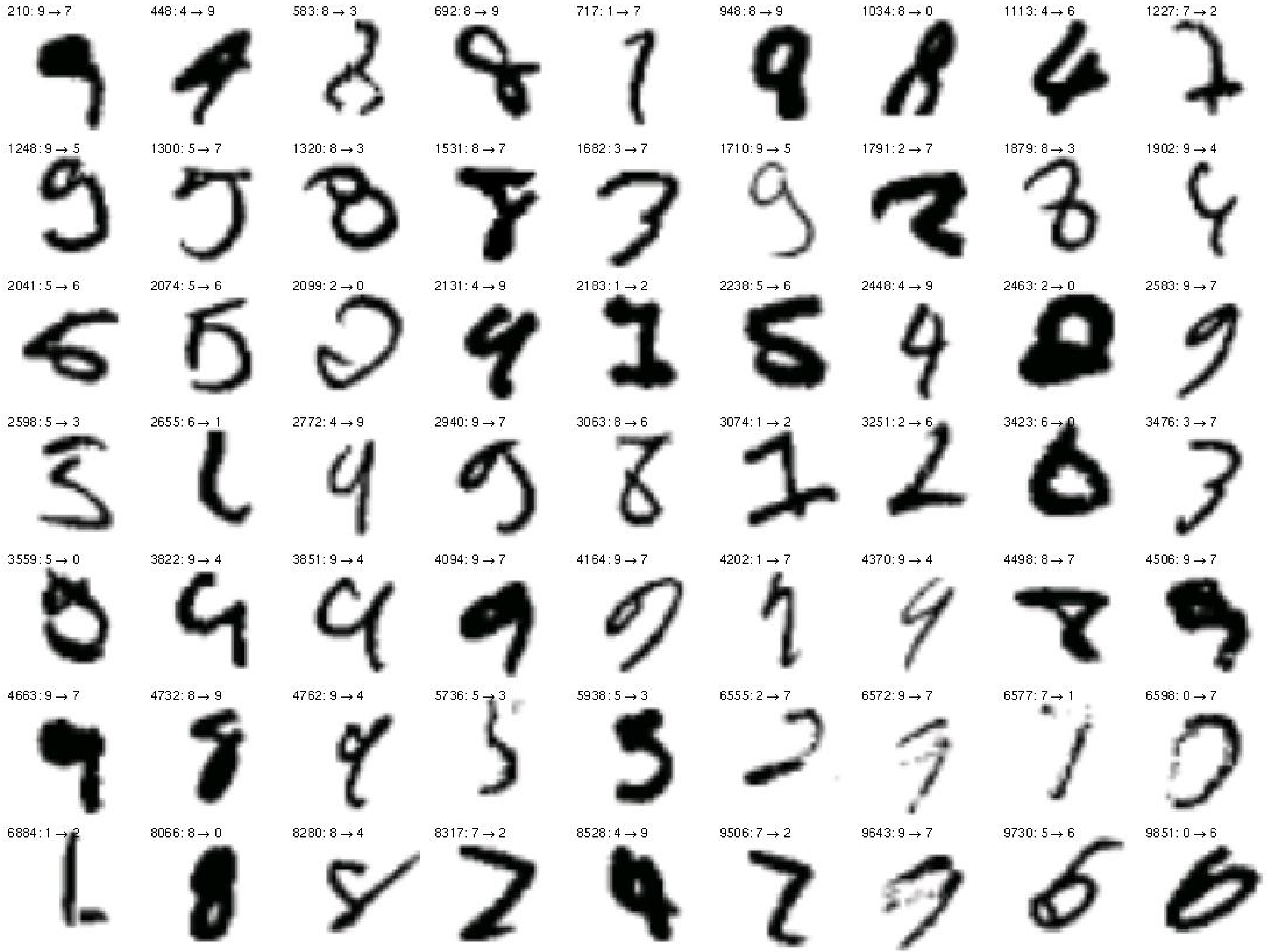




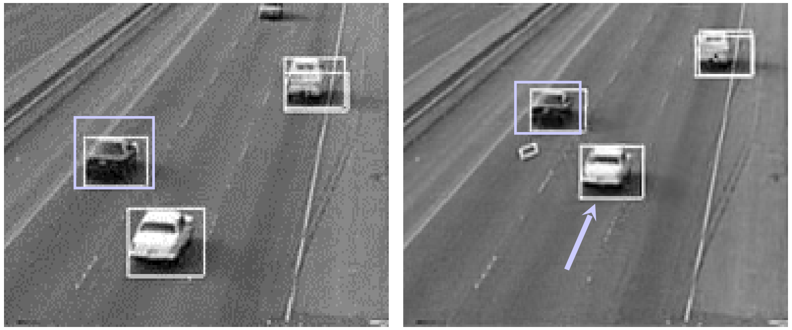
Detection and Segmentation: Mugs

Orig. Image Segmentation

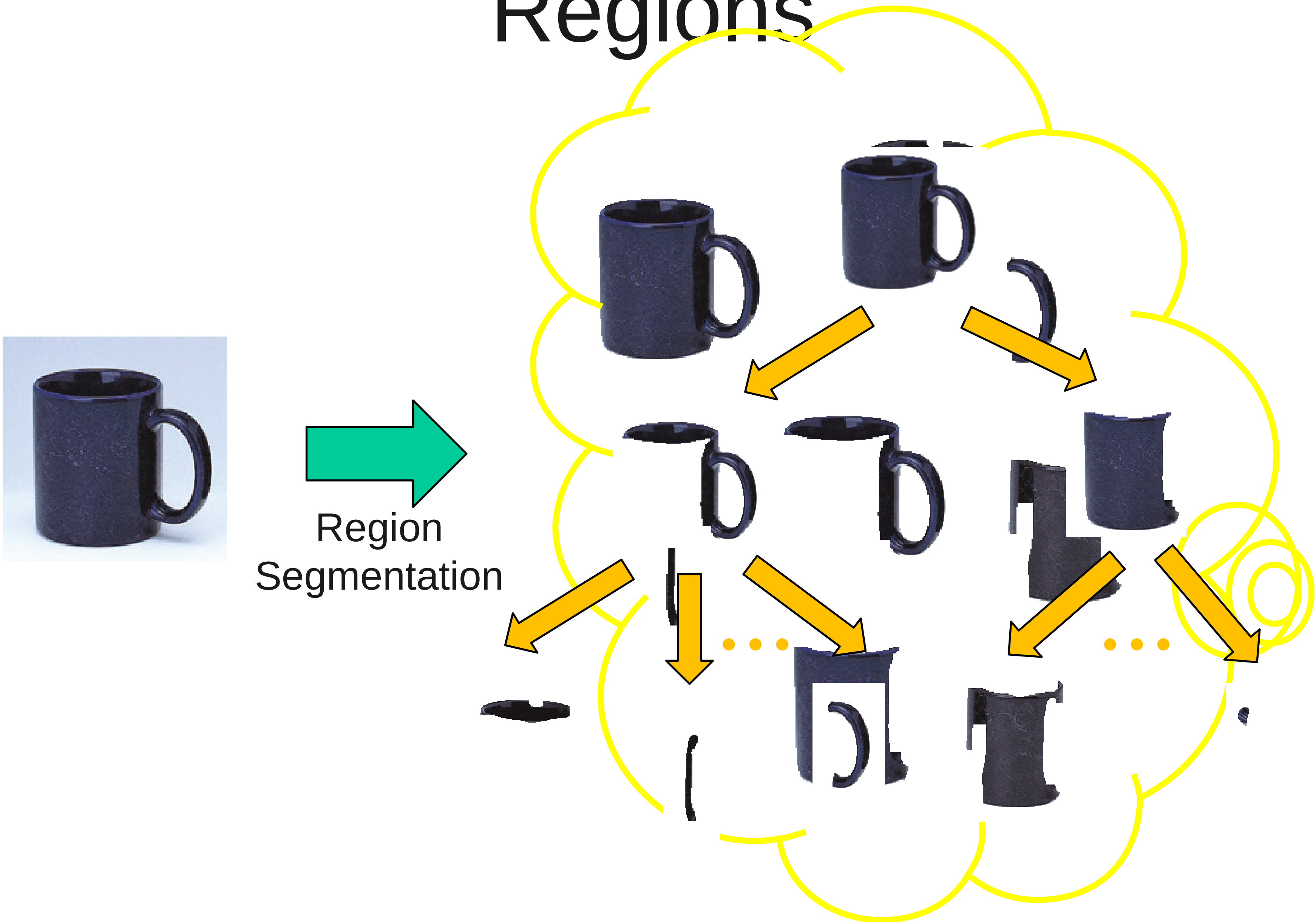
Orig. Image Segmentation


210: 9 → 7
448: 4 → 9
583: 8 → 3
692: 8 → 9
717: 1 → 7
948: 8 → 9
1034: 8 → 0
1113: 4 → 6
1227: 7 → 2
1248: 9 → 5
1300: 5 → 7
1320: 8 → 3
1531: 8 → 7
1682: 3 → 7
1710: 9 → 5
1791: 2 → 7
1879: 8 → 3
1902: 9 → 4
2041: 5 → 6
2074: 5 → 6
2099: 2 → 0
2131: 4 → 9
2183: 1 → 2
2238: 5 → 6
2448: 4 → 9
2463: 2 → 0
2583: 9 → 7
2598: 5 → 3
2655: 6 → 1
2772: 4 → 9
2940: 9 → 7
3063: 8 → 6
3074: 1 → 2
3251: 2 → 6
3423: 6 → 0
3476: 3 → 7
3559: 5 → 0
3822: 9 → 4
3851: 9 → 4
4094: 9 → 7
4164: 9 → 7
4202: 1 → 7
4370: 9 → 4
4498: 8 → 7
4506: 9 → 7
4663: 9 → 7
4732: 8 → 9
4762: 9 → 4
5736: 5 → 3
5938: 5 → 3
6555: 2 → 7
6572: 9 → 7
6577: 7 → 1
6598: 0 → 7
6884: 1 → 2
8066: 8 → 0
8280: 8 → 4
8317: 7 → 2
8528: 4 → 9
9506: 7 → 2
9643: 9 → 7
9730: 5 → 6
9851: 0 → 6

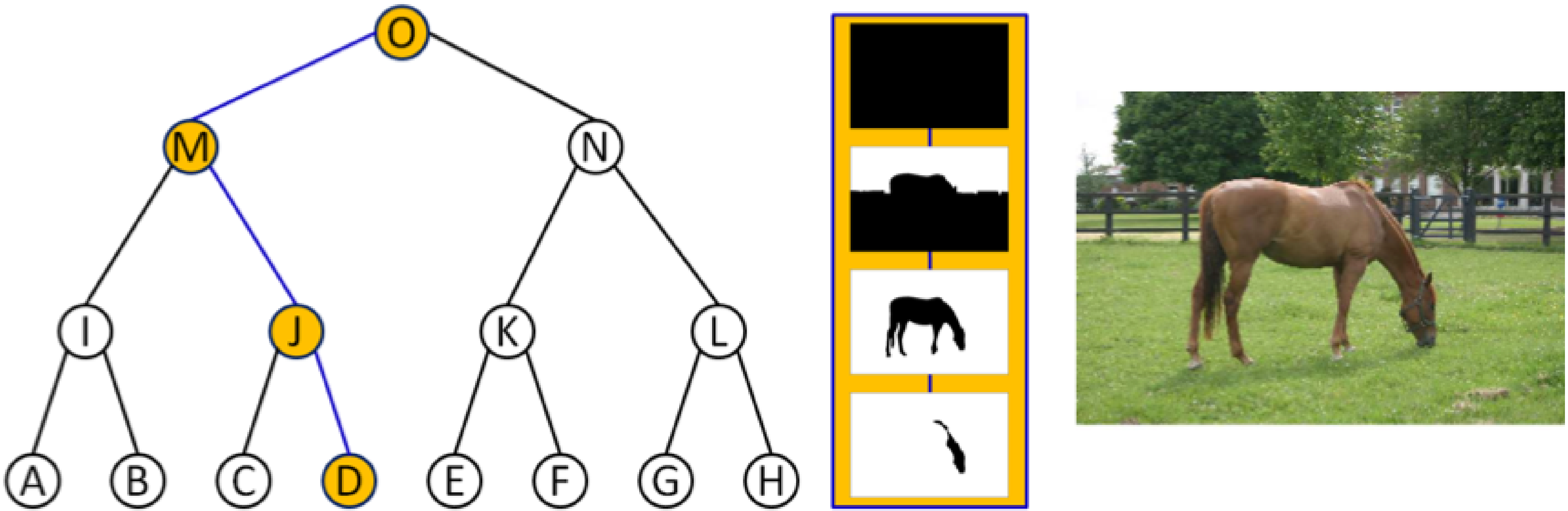
Temporal Segmentation: Tracking



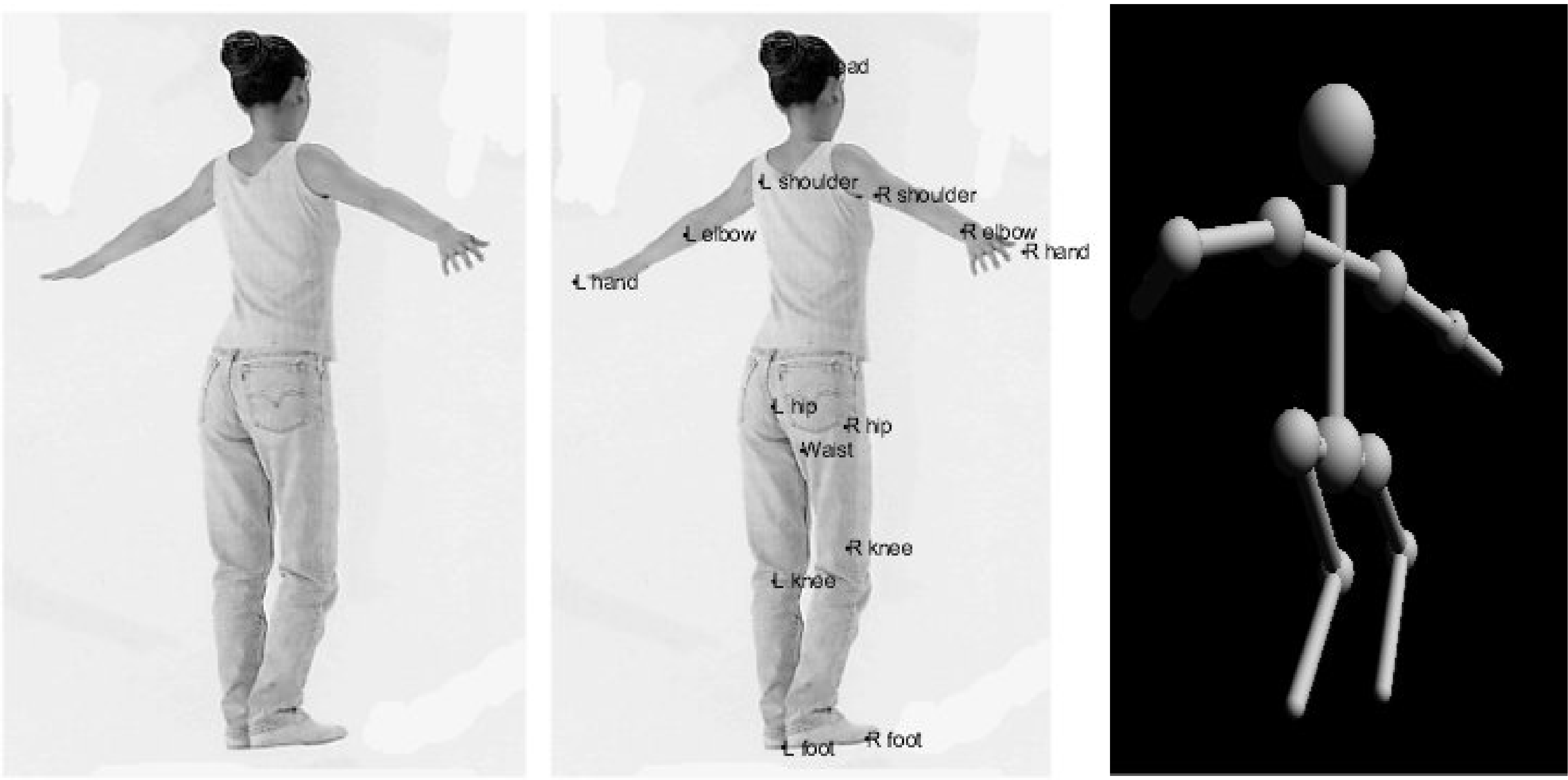
Object Representation using Regions



Context from region tree



Human body configurations



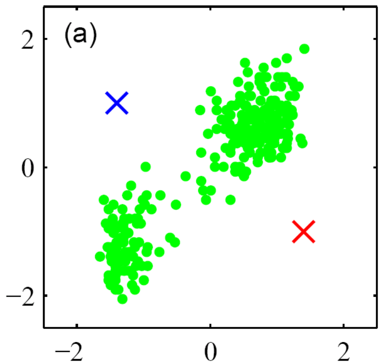
3D detection - Tsukaba

Some Machine Learning Techniques

- K-means clustering
- SVM
  - Kernel Trick
- Belief Propagation
- Markov Random Fields

K-means clustering

Data and initial (random) means



Assign each point to its nearest mean

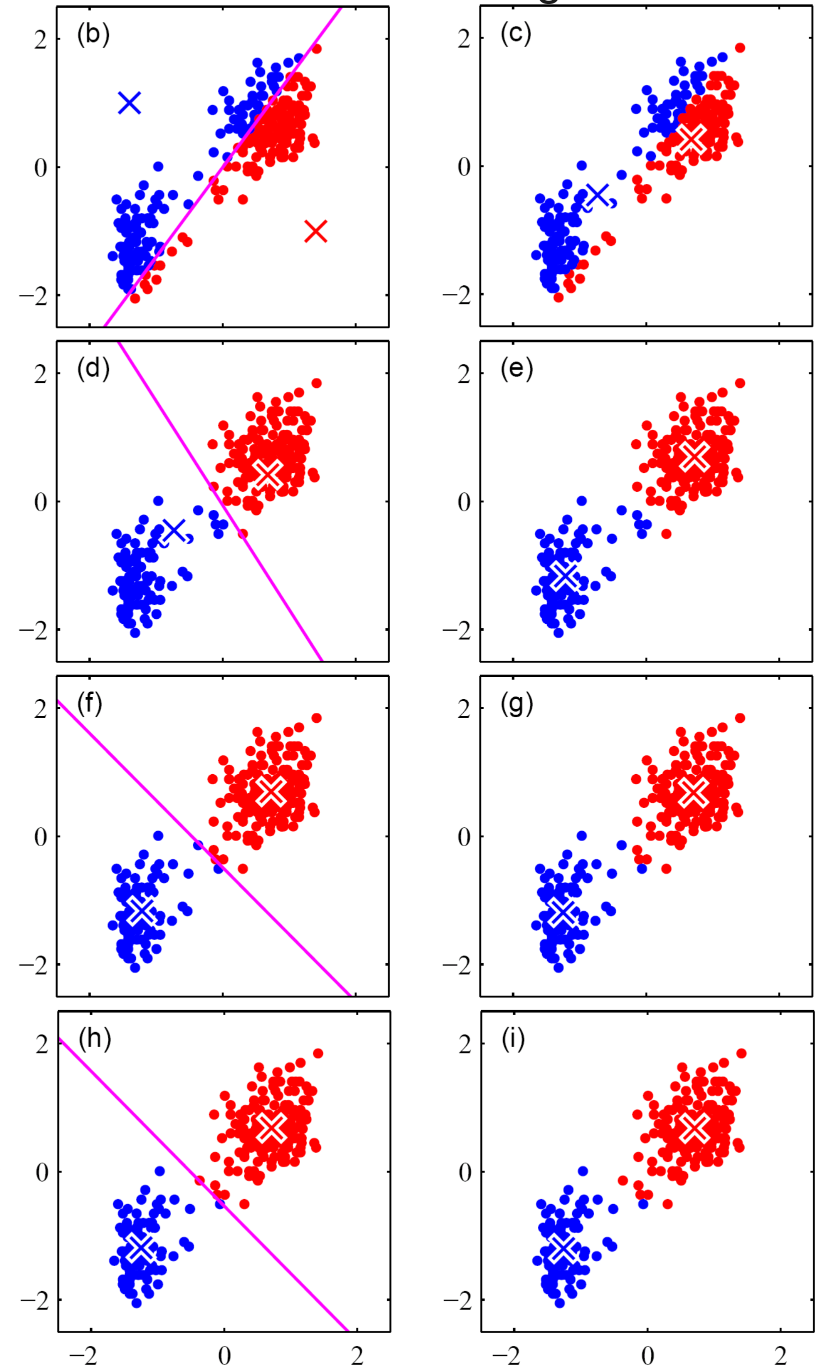
Set each mean to the average of its data

Iteration 1

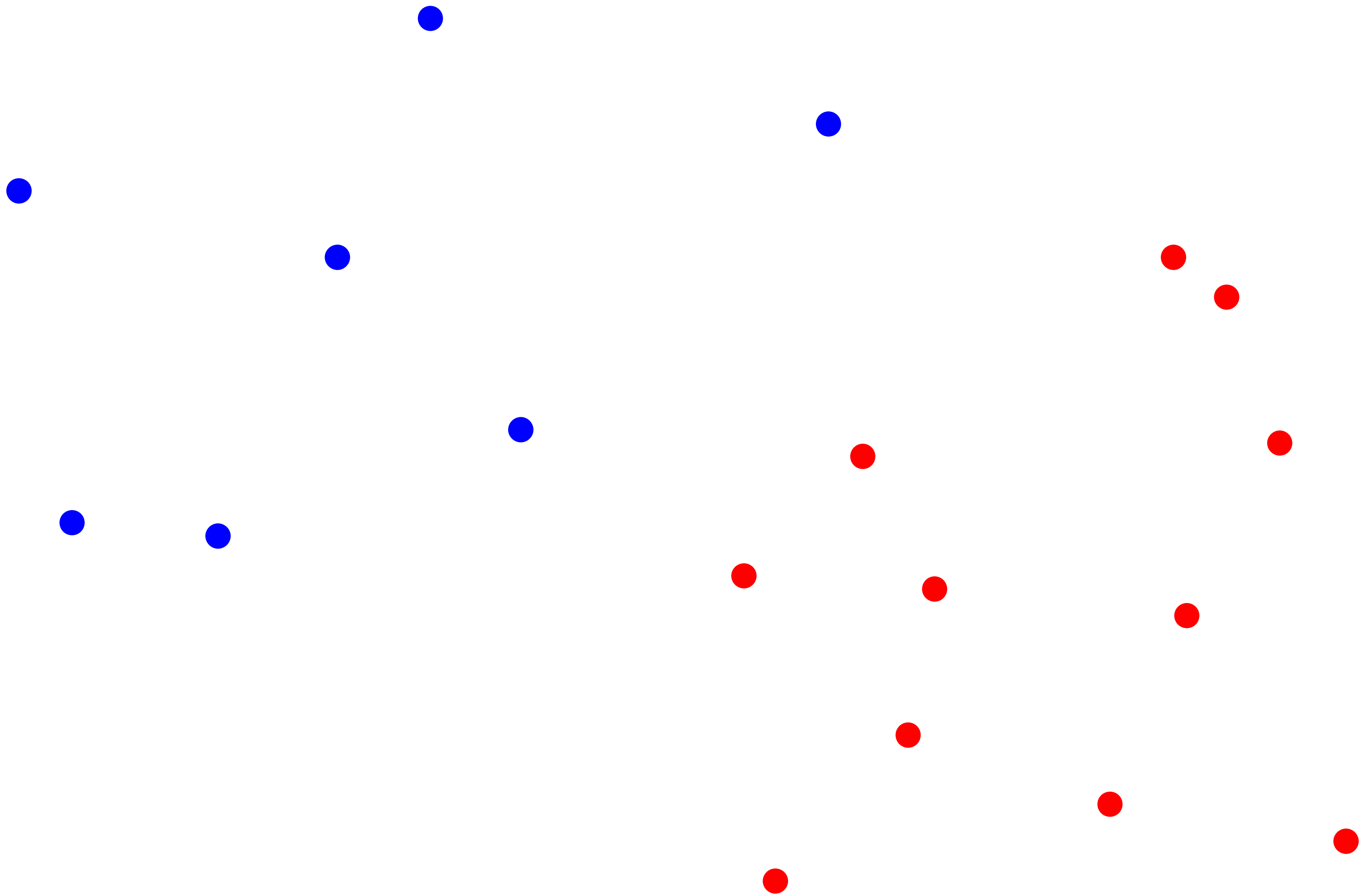
Iteration 2

Iteration 3

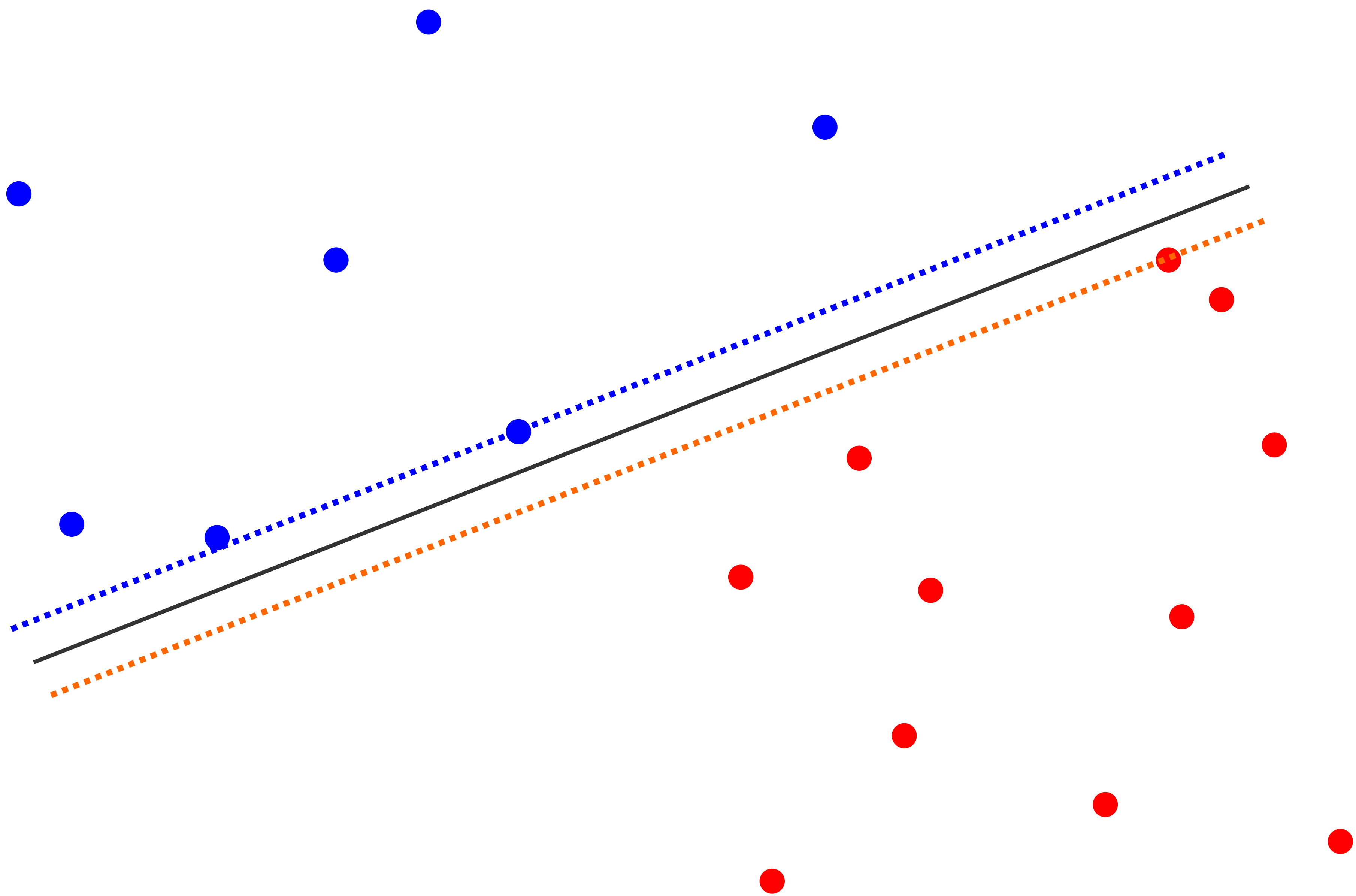
Iteration 4 and convergence



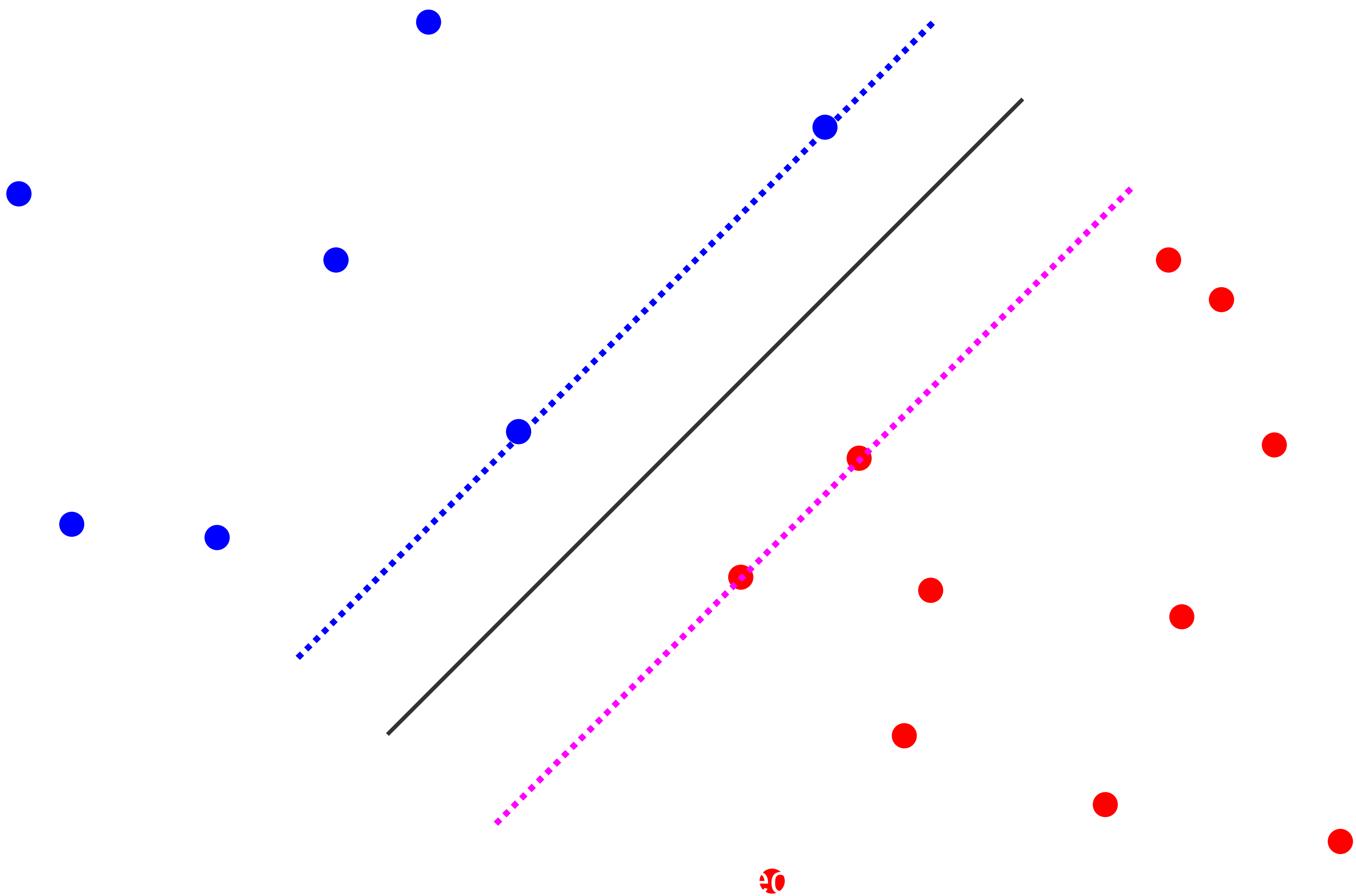
Binary Classification



A Separating Hyperplane



Maximum Margin Hyperplane



Some Popular Kernels

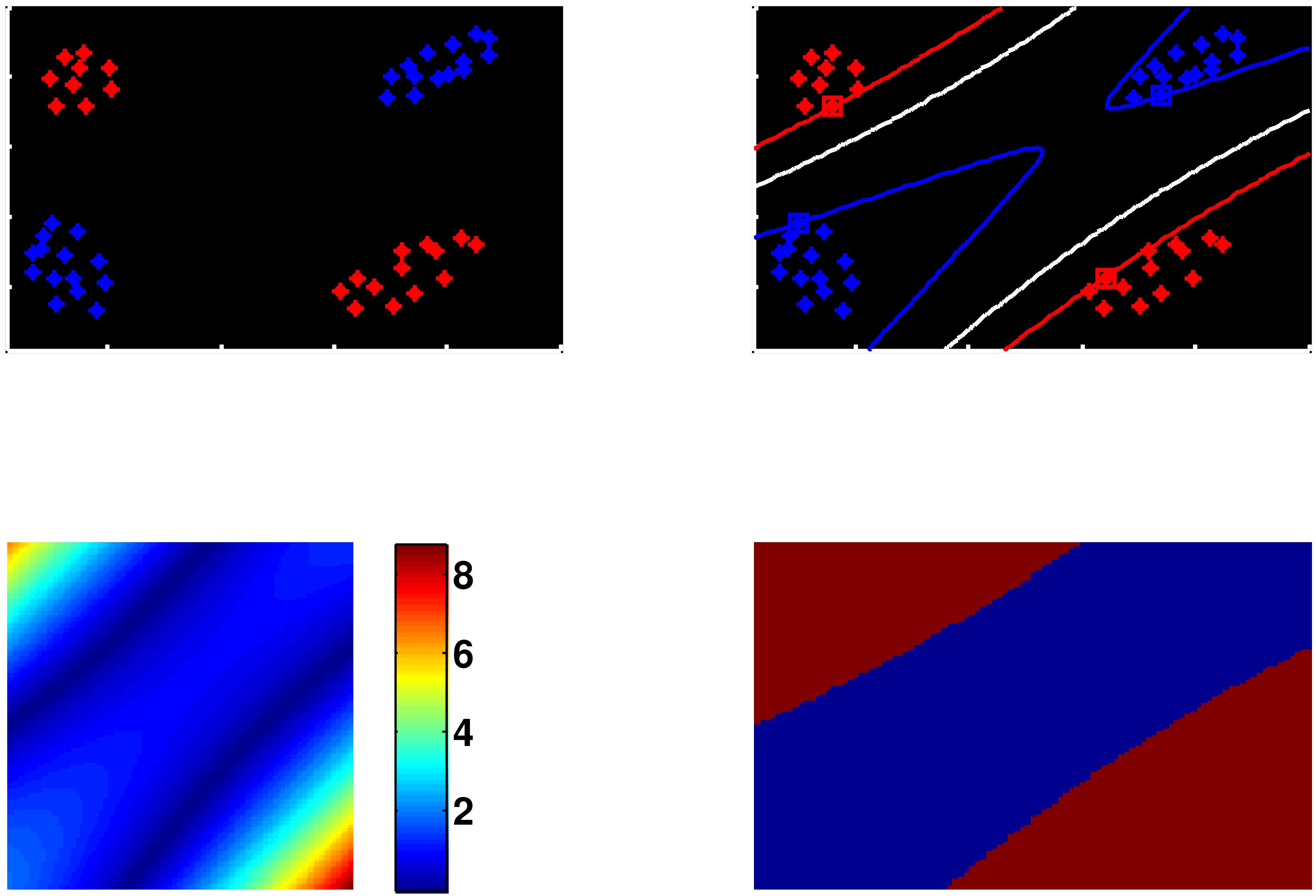
Linear : $K(\mathbf{x}_i,\mathbf{x}_j) = \mathbf{x}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x}_j$

Polynomial : $K(\mathbf{x}_i,\mathbf{x}_j) = (\mathbf{x}_i^t \boldsymbol{\Sigma}^{-1} \mathbf{x}_j + c)^d$

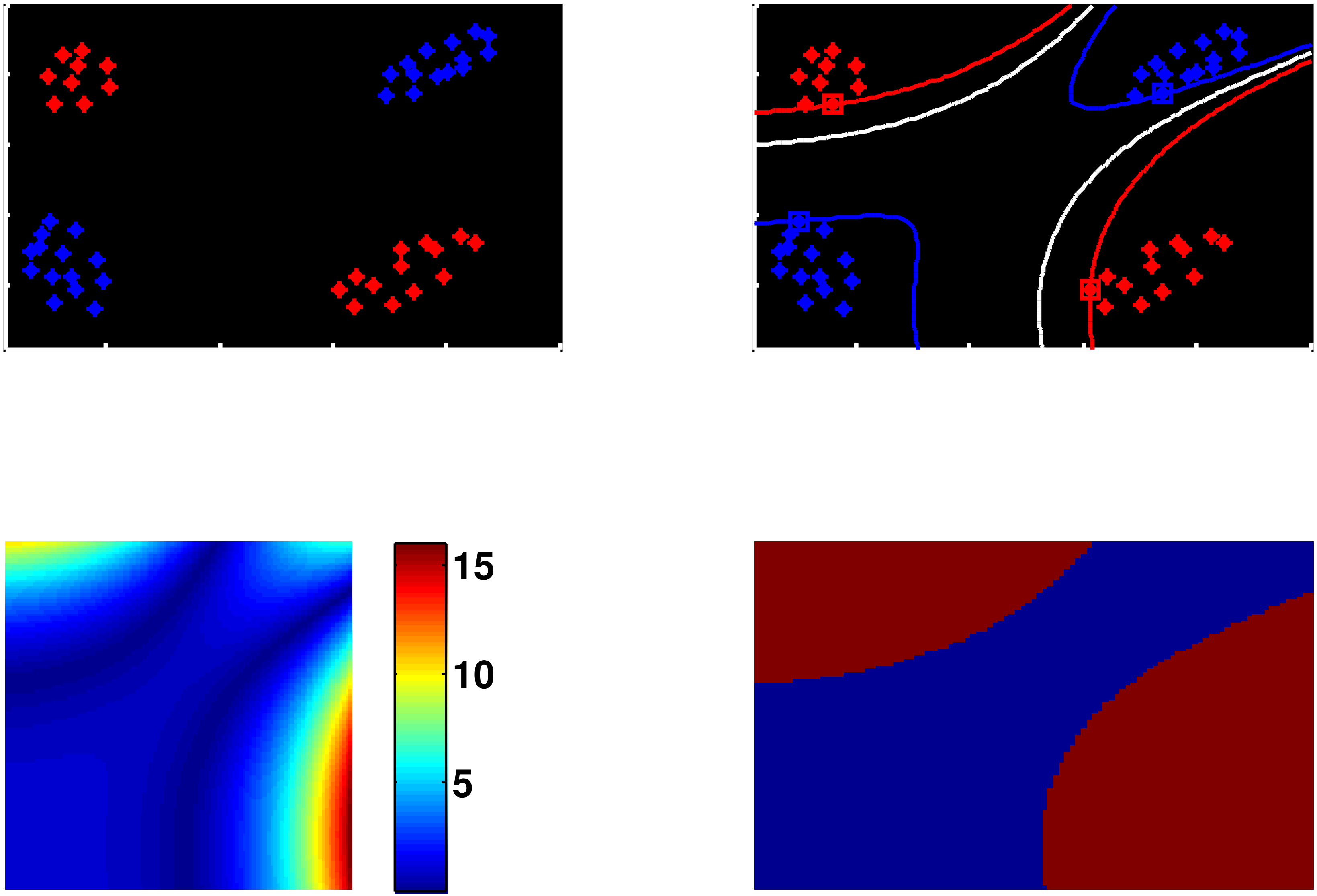
Gaussian (RBF) : $K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\Sigma_k \gamma_k (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2)$

Sigmoid : $K(\mathbf{x}_i,\mathbf{x}_j) = \tanh(\mathbf{x}_i^t \mathbf{x}_j - c)$

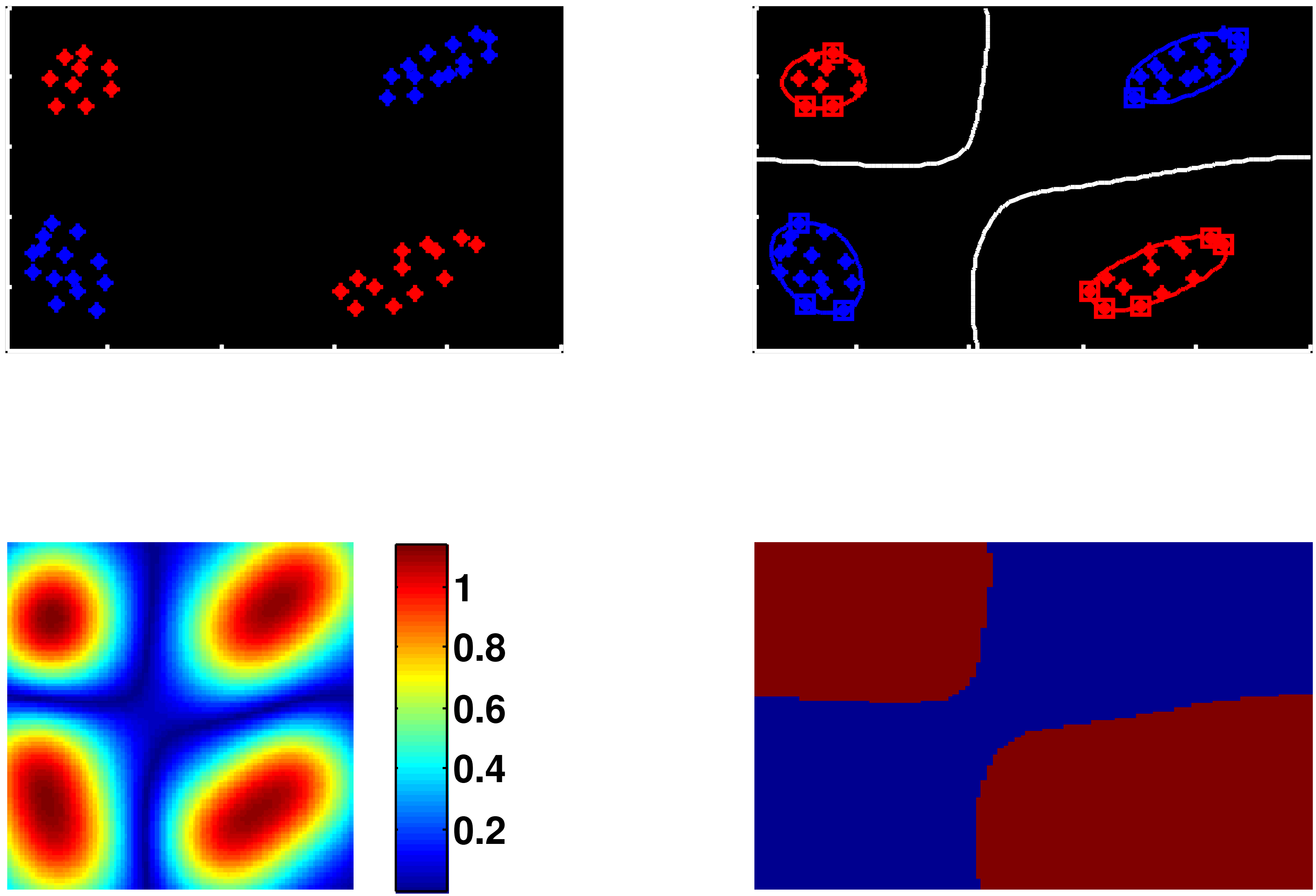
Polynomial Kernel of Degree 2



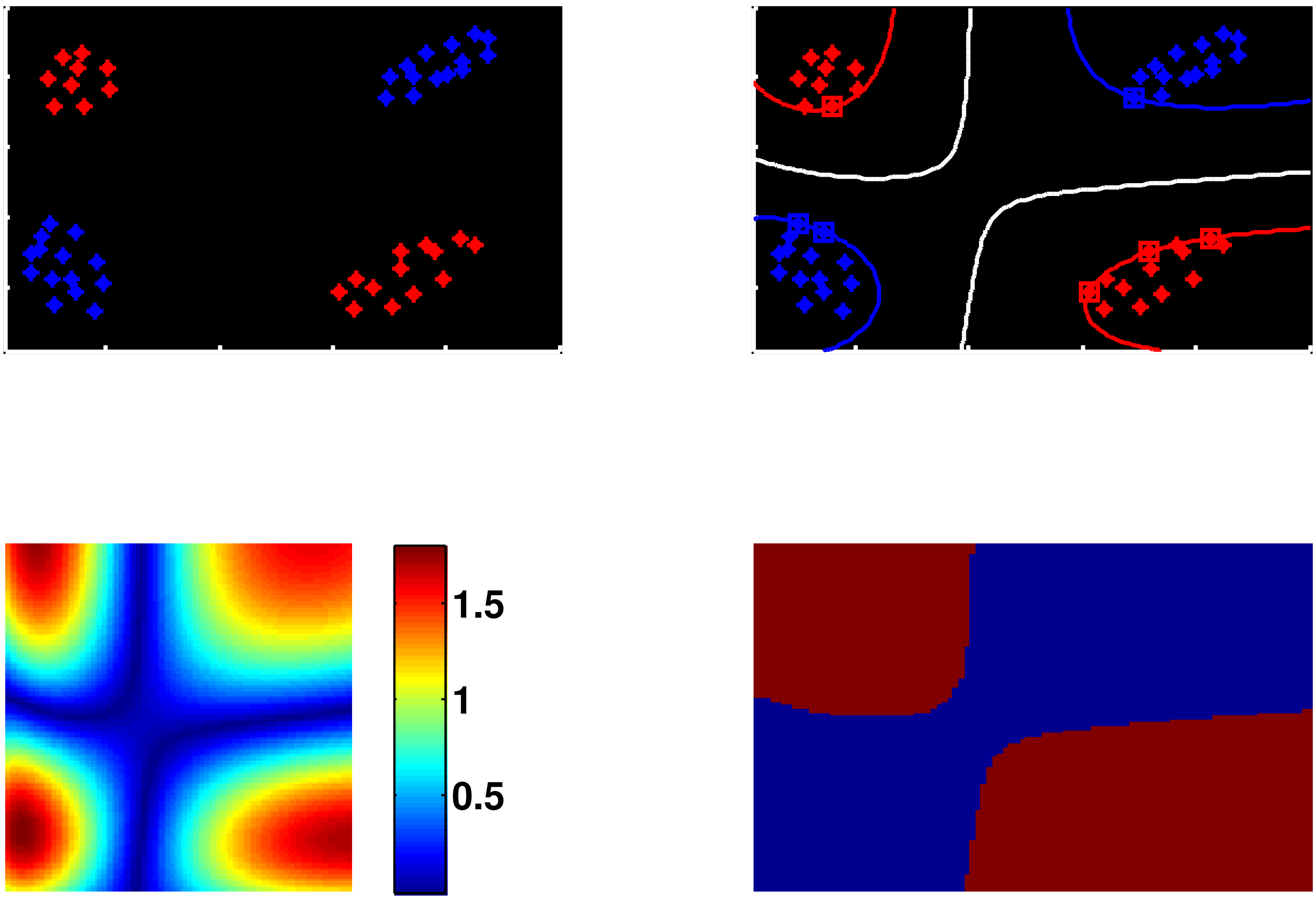
Polynomial Kernel of Degree 5



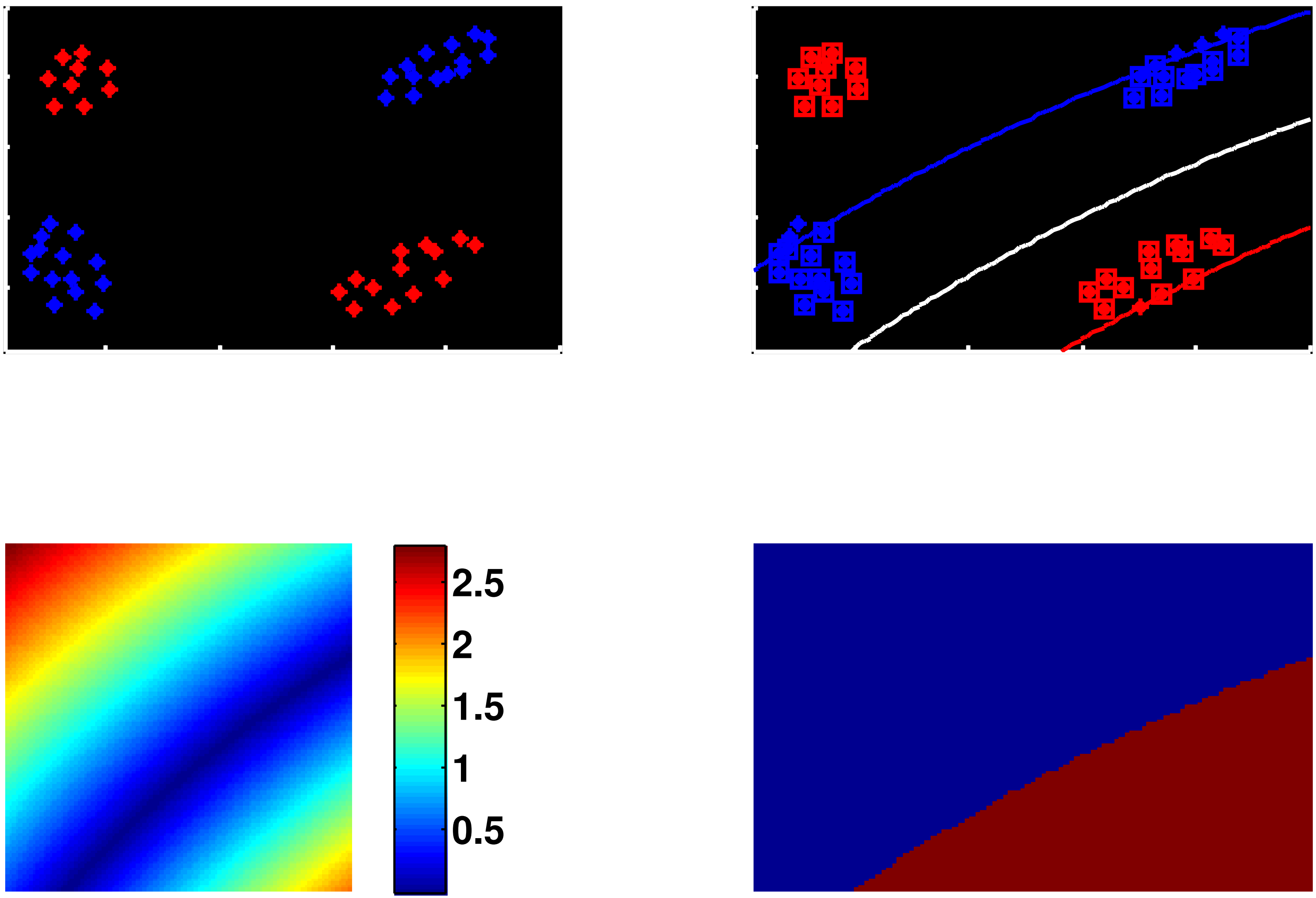
RBF Kernel



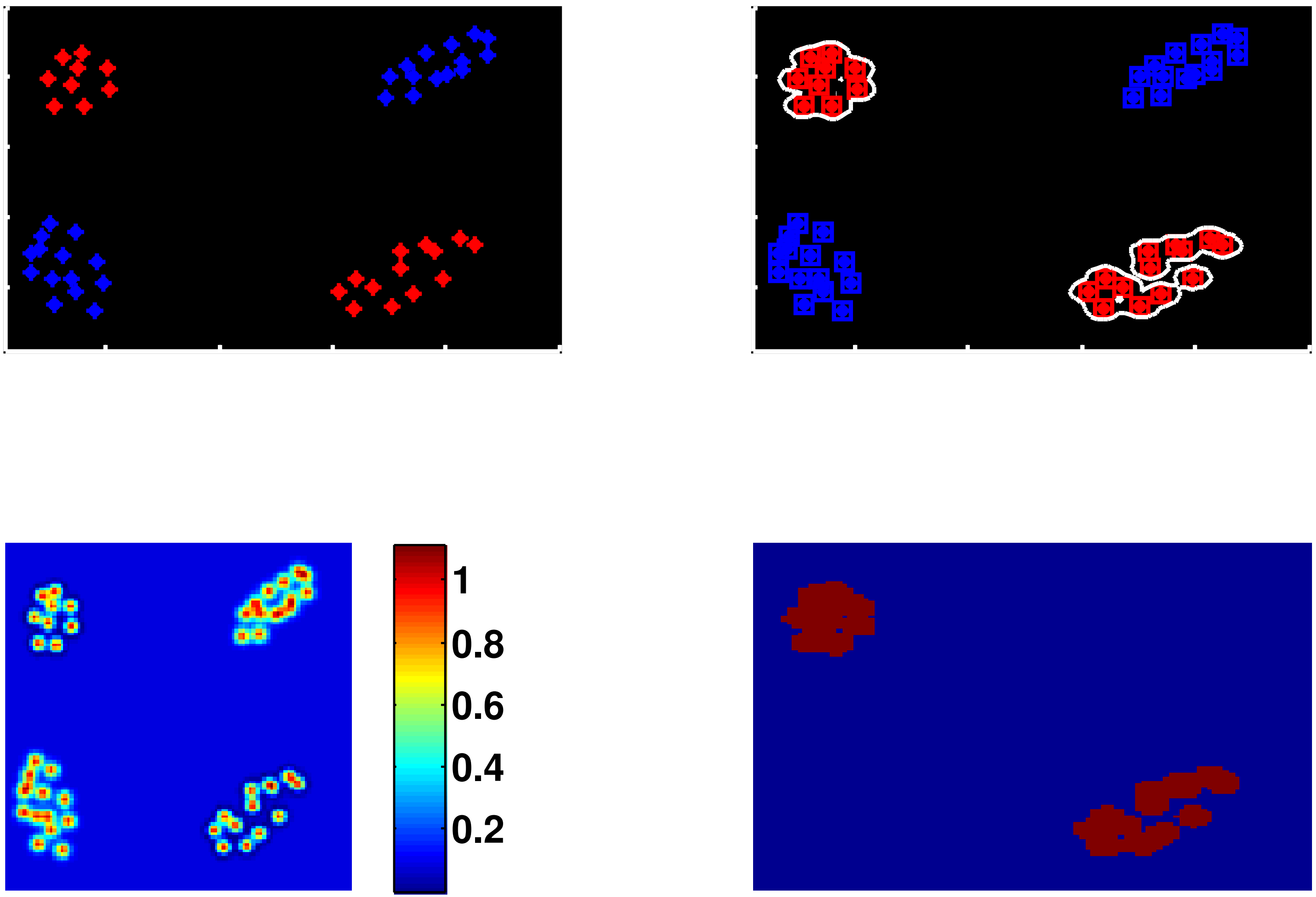
Exponential χ^2 Kernel



Kernel Parameter Setting - Underfitting



Kernel Parameter Setting – Overfitting



Intersection Kernel

Histogram Intersection kernel between histograms *a*, *b*

$$K(a,b) = \sum_{i=1}^{n} min(a_i, b_i)$$

$$a_i \geq 0$$
$$b_i \geq 0$$

K small -> *a*, *b* are different
K large -> *a*, *b* are similar

New Kernel

Decision function is $\operatorname{sign}(h(x))$ where:

Linear:

$$h(x) = w'x + b = \sum_{i=1}^{\#\mathrm{dim}} w_i x_i + b$$

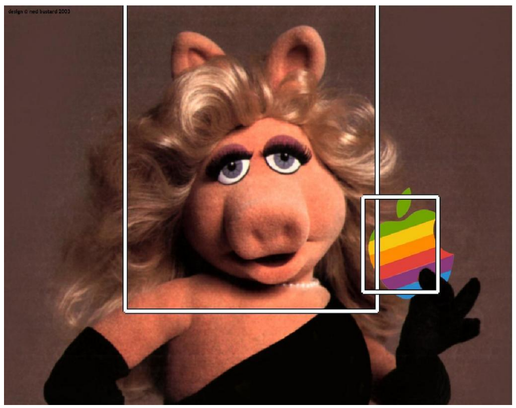
Non-linear Using Kernel

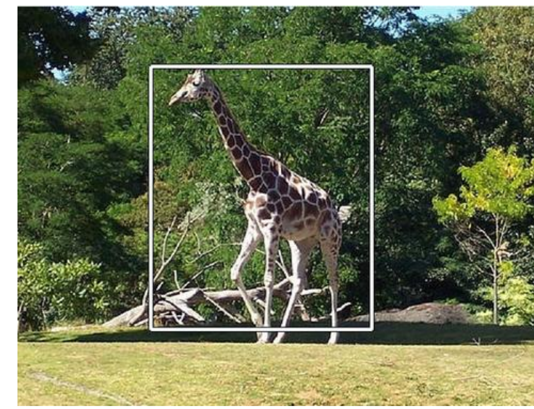
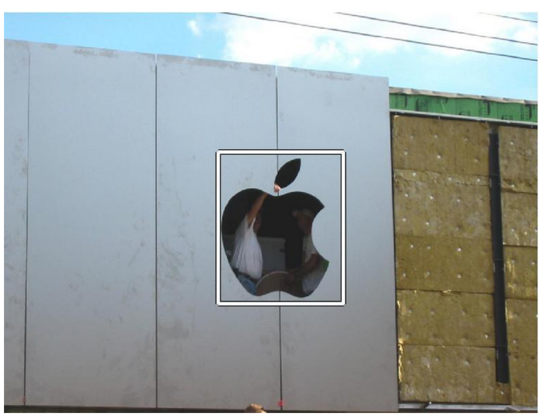
$$h(x) = \sum_{j=1}^{\#\mathrm{sv}} \alpha^j K(x, x^j) + b$$

Histogram Intersection Kernel

$$= \sum_{j=1}^{\#\mathrm{sv}} \left(\alpha^j \sum_{i=1}^{\#\mathrm{dim}} \min(x_i, x_i^j) \right) + b$$

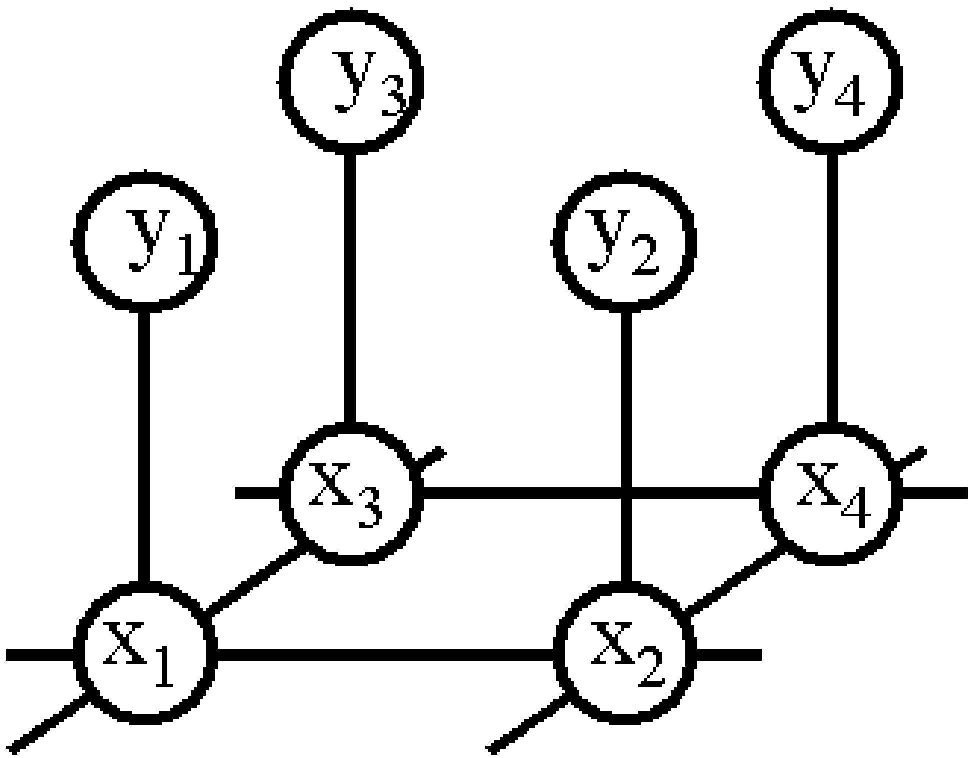
Results – ETHZ Dataset




Markov Random Fields

- Allows rich probabilistic models for images.
- But built in a local, modular way. Learn local relationships, get global effects out.



MRF nodes as pixels

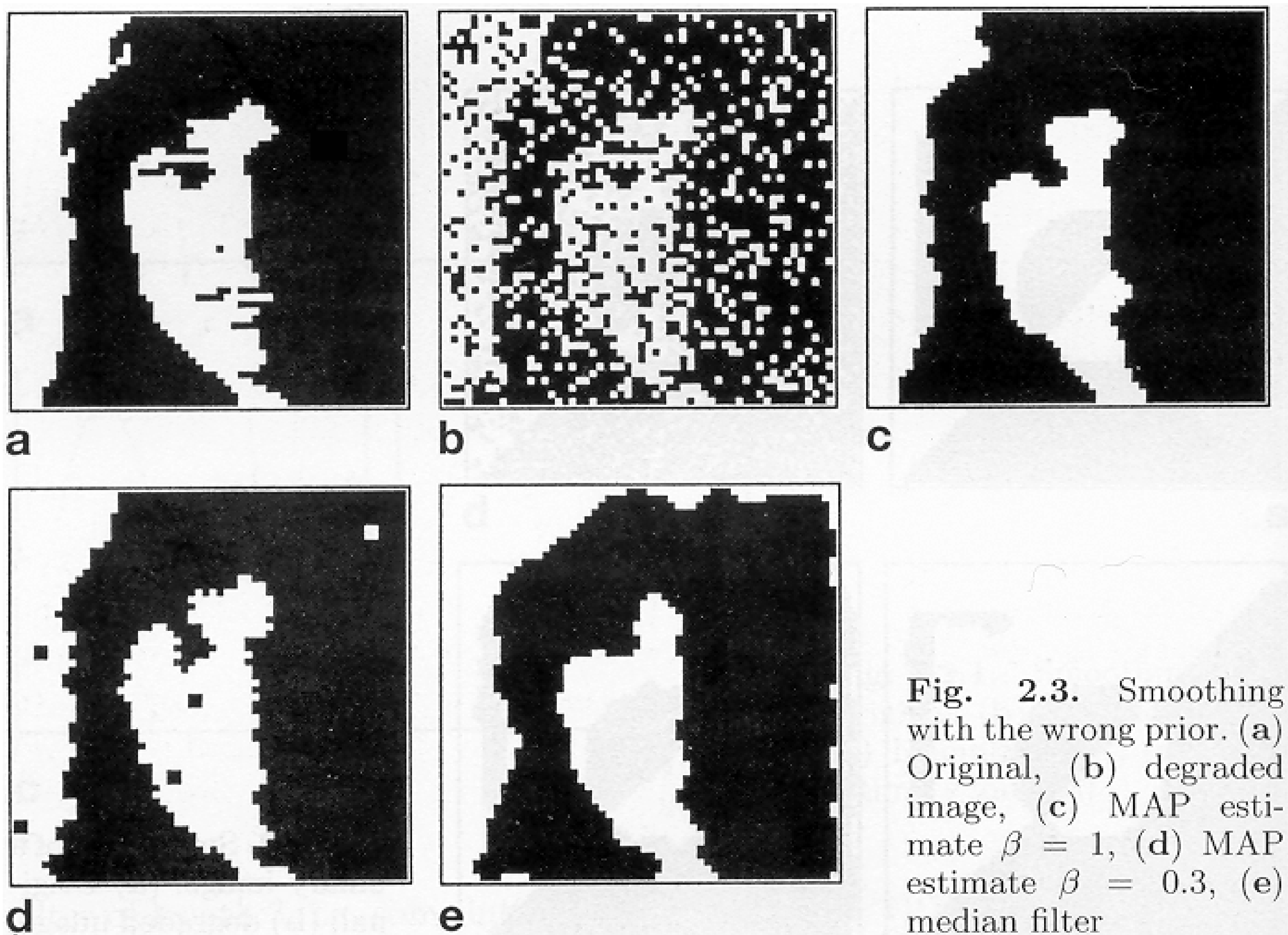
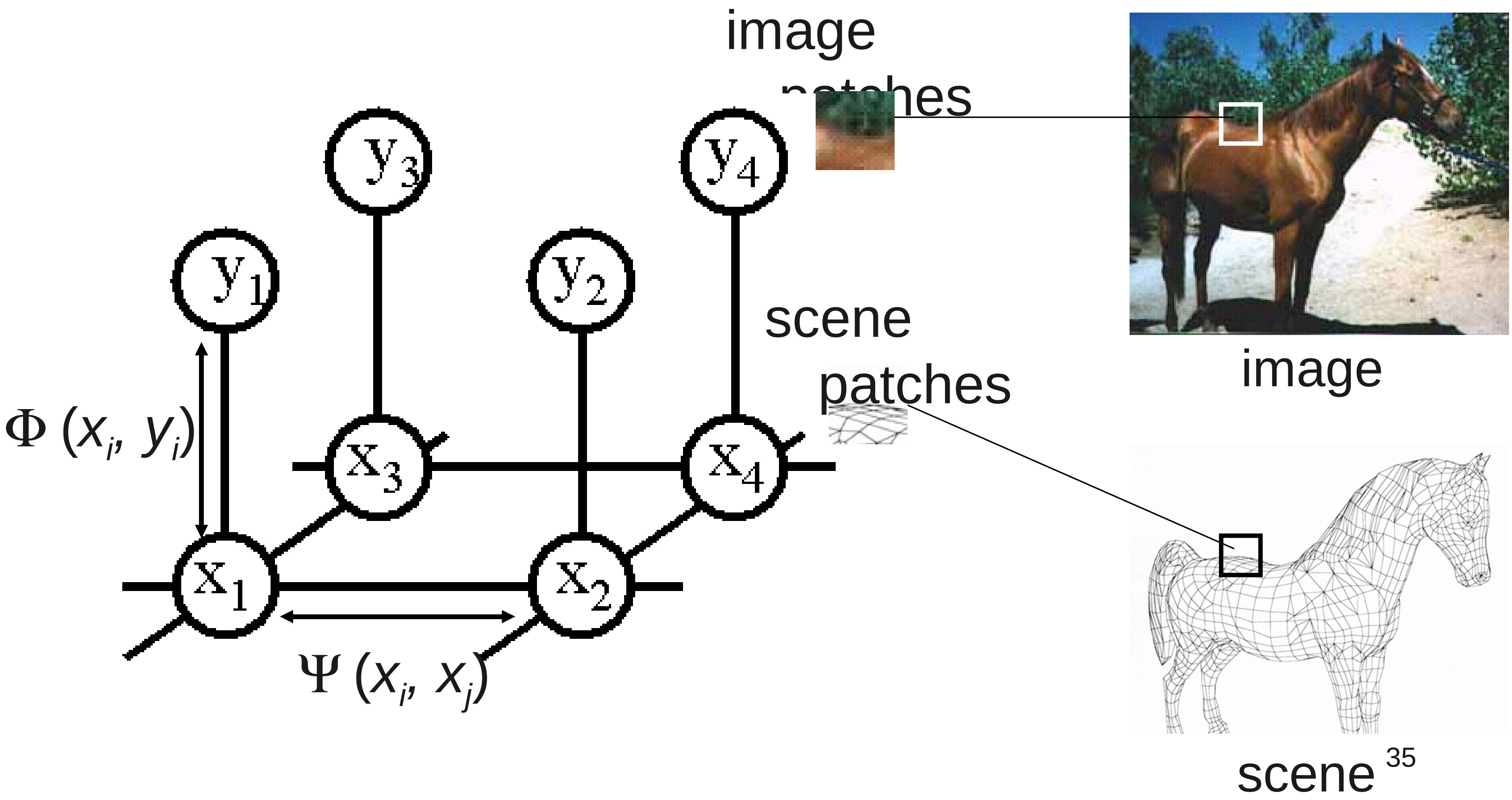


Fig. 2.3. Smoothing with the wrong prior. (**a**) Original, (**b**) degraded image, (**c**) MAP estimate $\beta = 1$, (**d**) MAP estimate $\beta = 0.3$, (**e**) median filter

Winkler, 1995, p. 32

MRF nodes as patches



Super-resolution

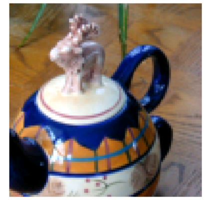
- Image: low resolution image
- Scene: high resolution image

ultimate goal...

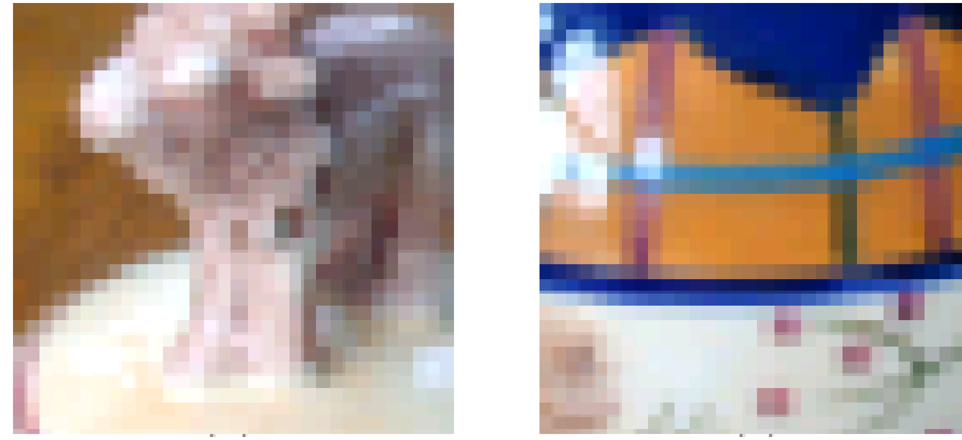
image

scene

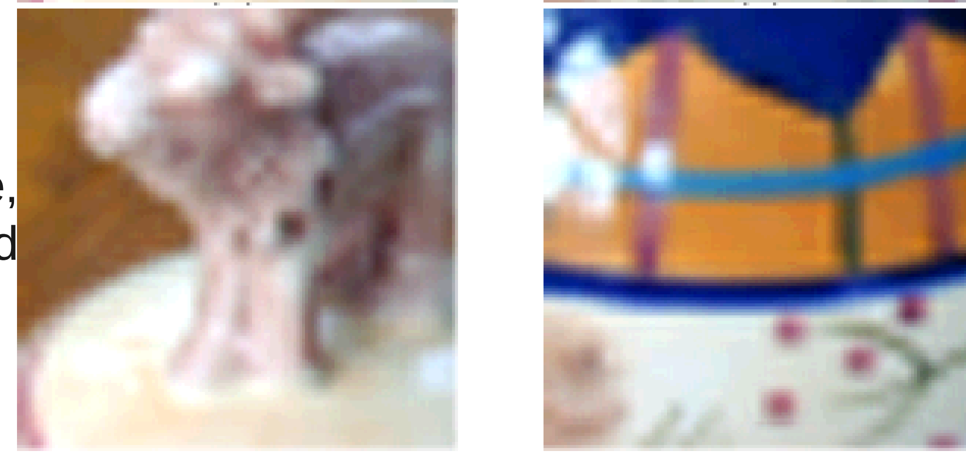
Pixel-based images are not resolution independent



Pixel replication



Cubic spline, sharpened



Training-based super-resolution

Polygon-based graphics images are resolution independent

Training images, **~100,000 image/scene patch pairs**

Images from two Corel database categories:
"giraffes" and "urban skyline".

Do a first interpolation

Zoomed low-resolution

Low-resolution

Zoomed low-resolution

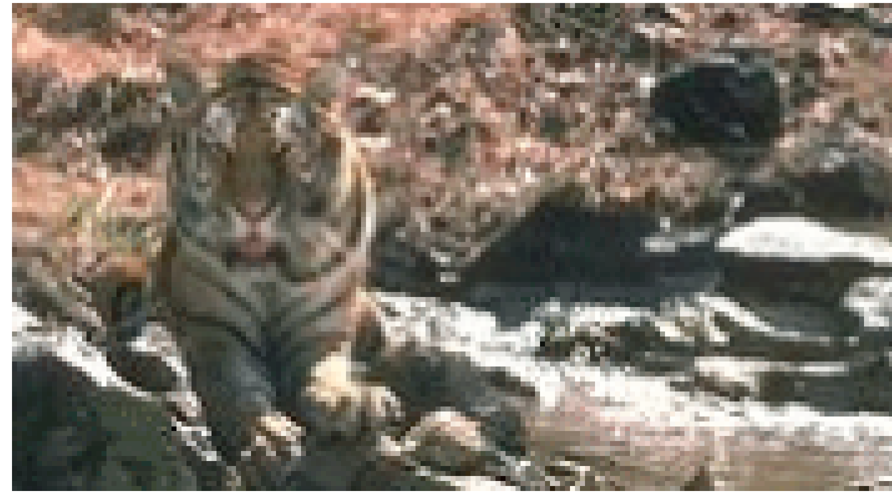
Full frequency original

Low-resolution

Representation

Zoomed low-freq.

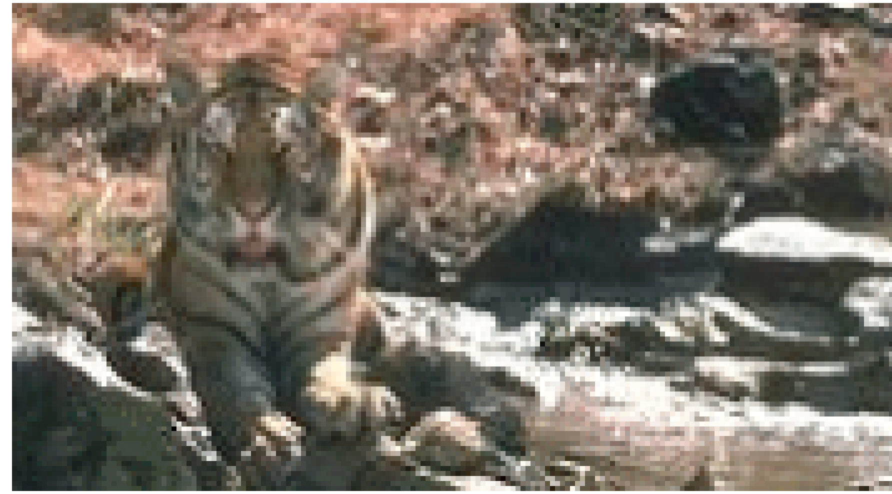
Full freq. original

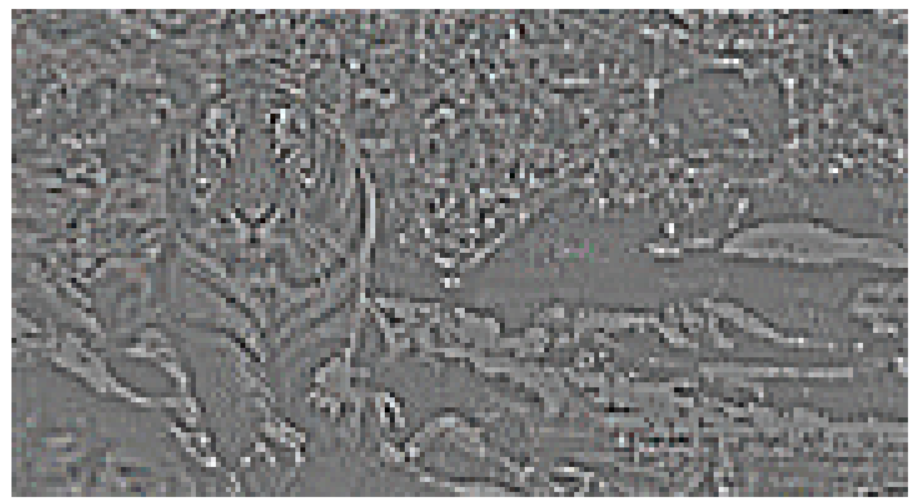


Representation

Zoomed low-freq.

Full freq. original



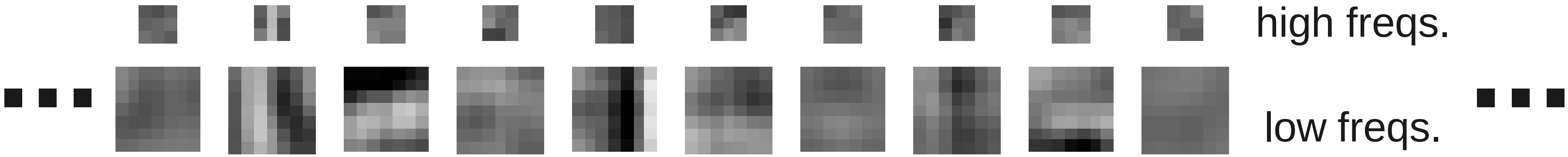


True high freqs

Low-band input
(contrast normalized,
PCA fitted)

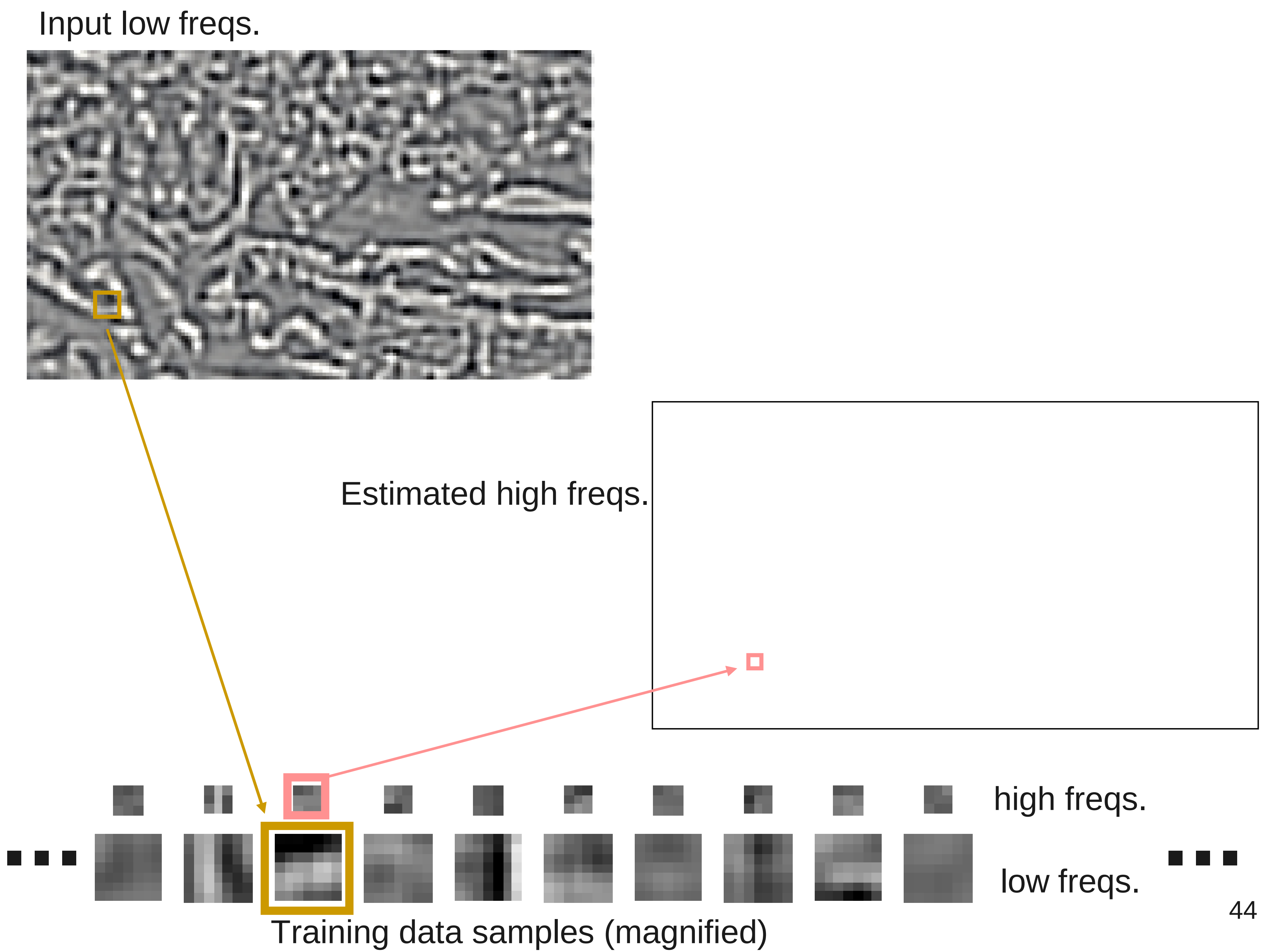
(to minimize the complexity of the relationships we have to learn,
we remove the lowest frequencies from the input image,
and normalize the local contrast level).

Gather ~100,000 patches

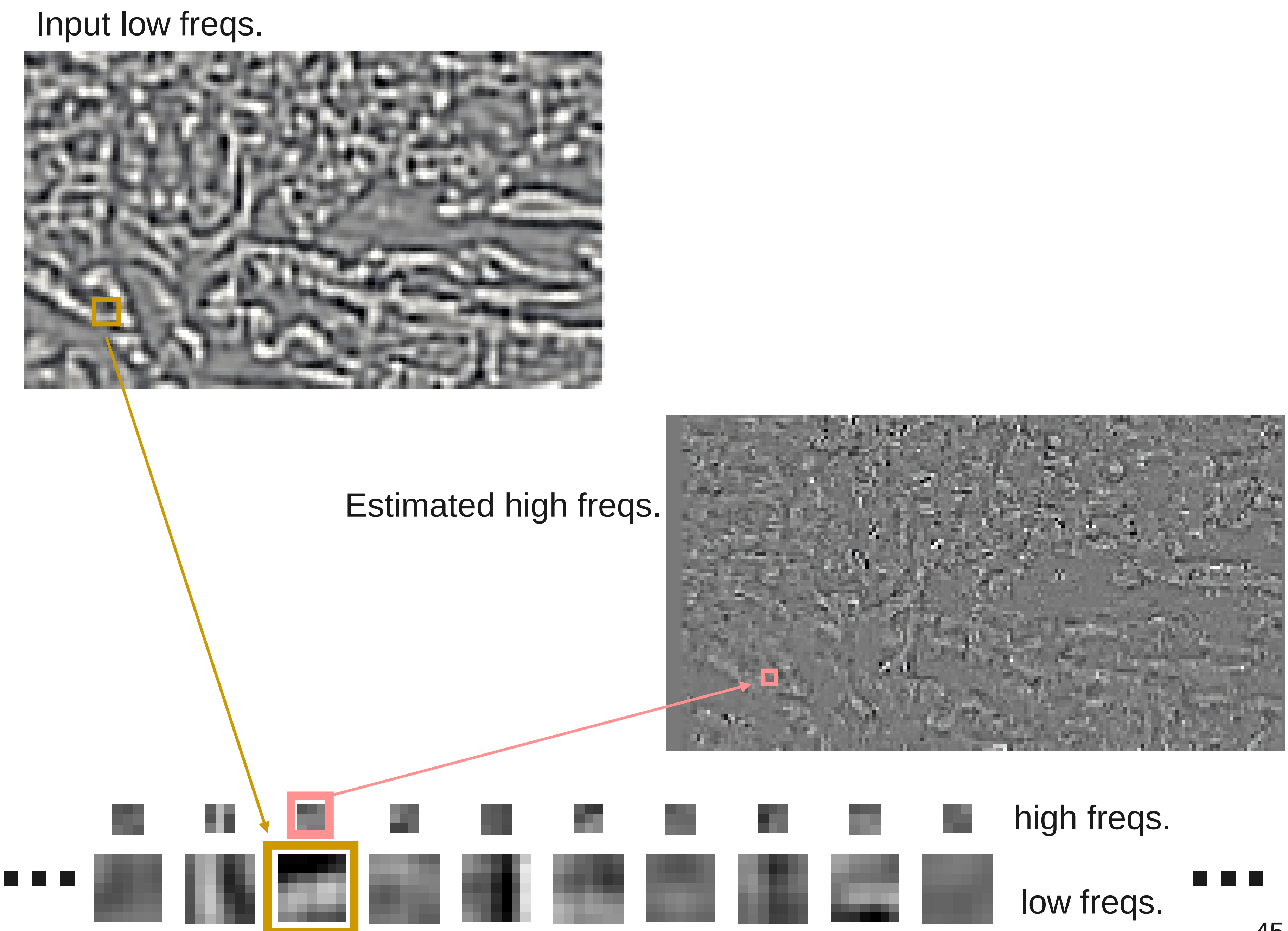


Training data samples (magnified)

Nearest neighbor estimate



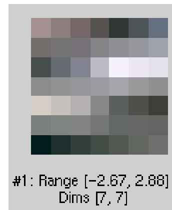
Nearest neighbor estimate



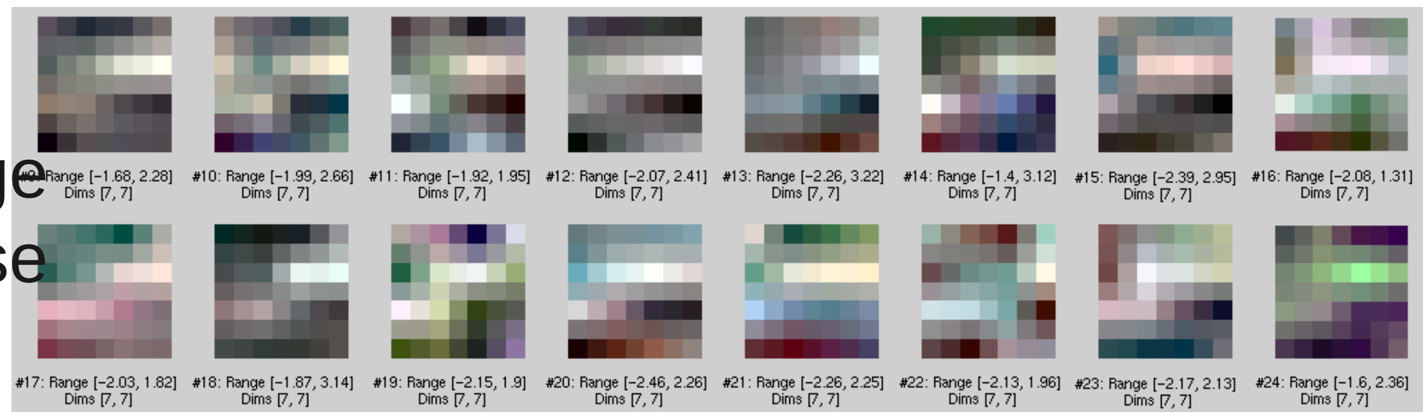
Training data samples (magnified)

Example: input image patch, and closest matches from database

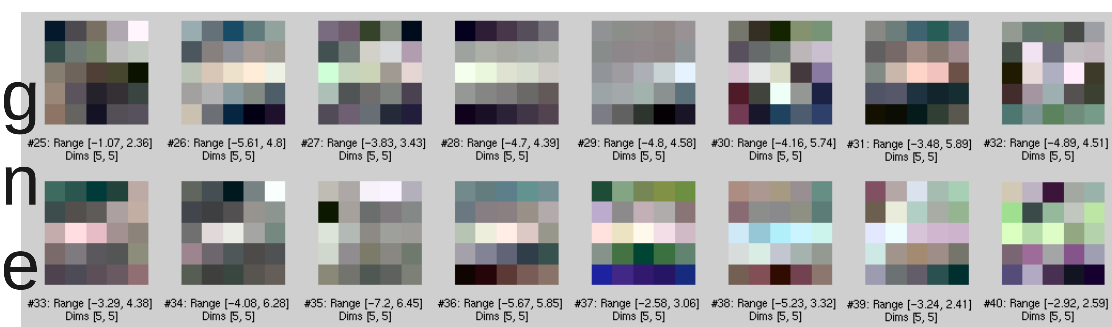
Input patch



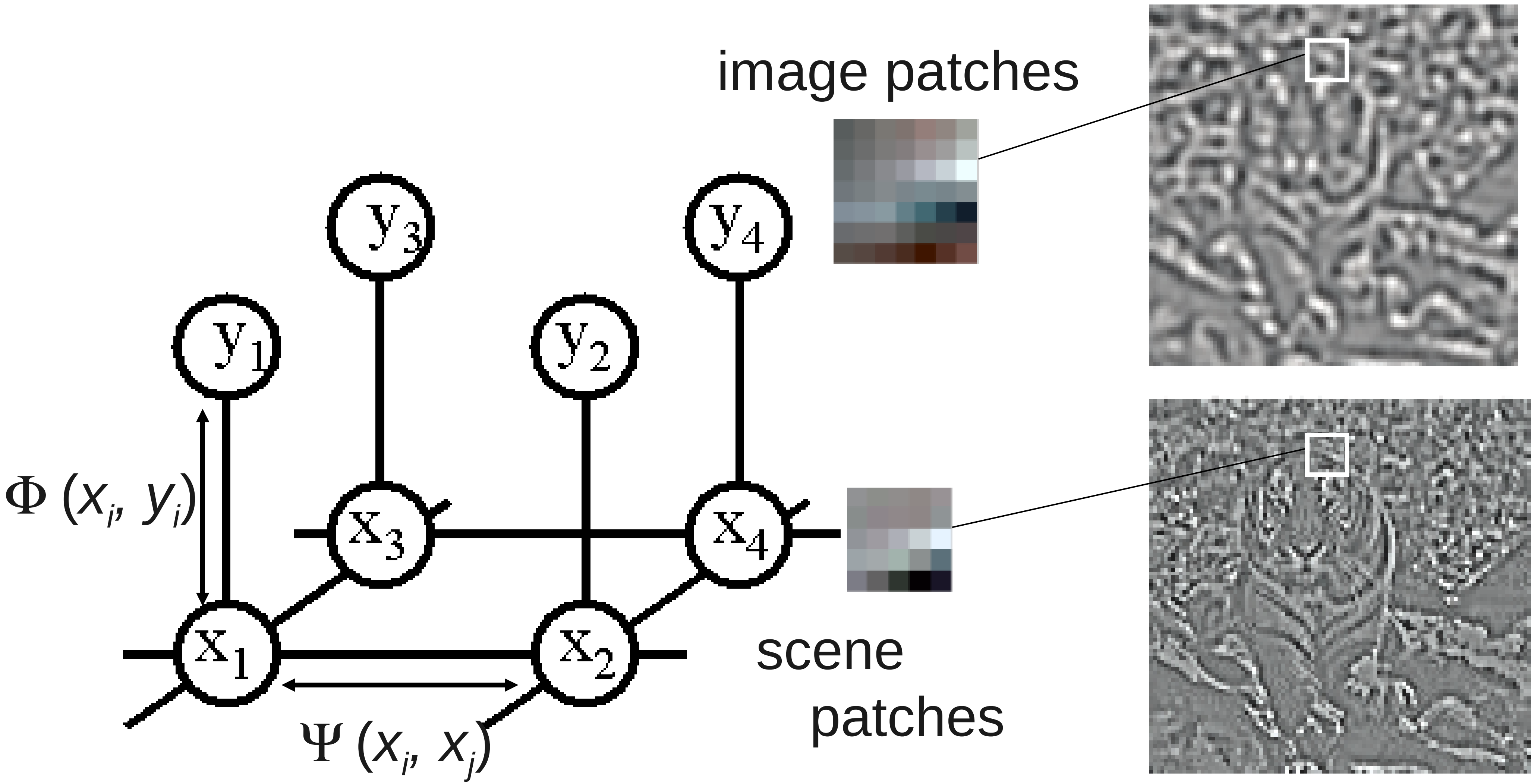
Closest image patches from database



Corresponding high-resolution patches from database



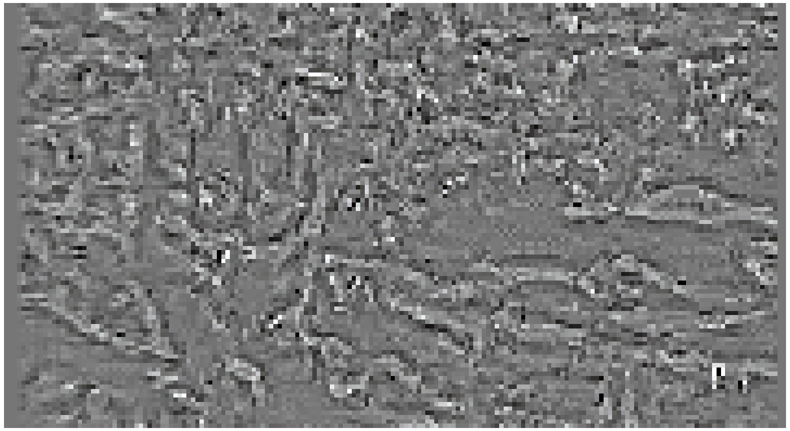
Markov network



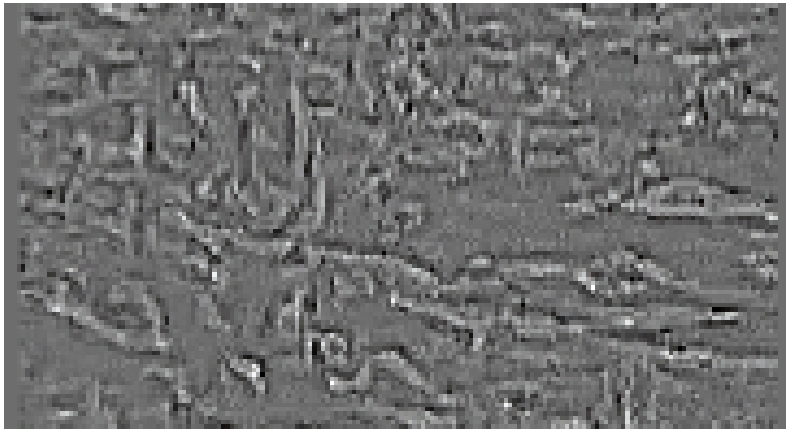
Belief Propagation

After a few iterations of belief propagation, the algorithm selects spatially consistent high resolution interpretations for each low-resolution patch of the input image.

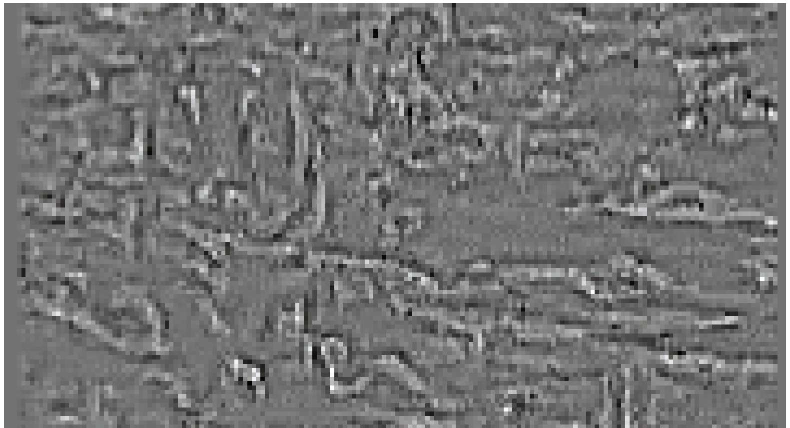
Input



Iter. 0



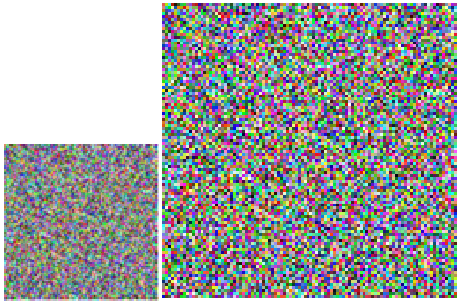
Iter. 1



Iter. 3

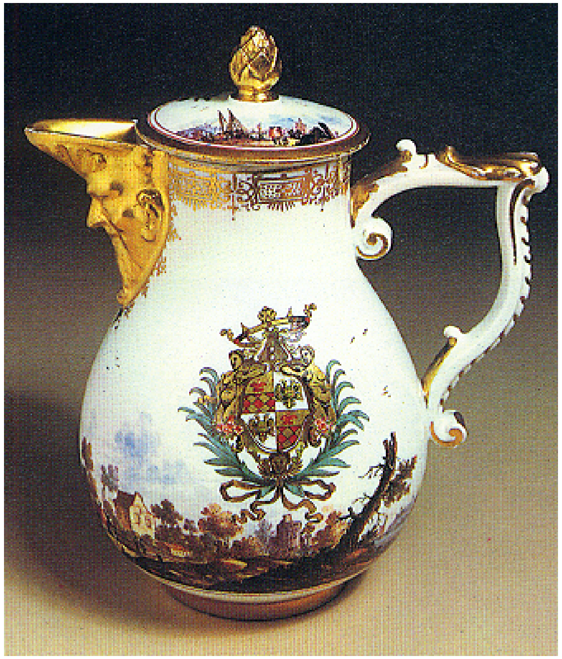
The algorithm learns that, in such a world, we add random noise when zoom to a higher resolution.

Original
50x58



Training images

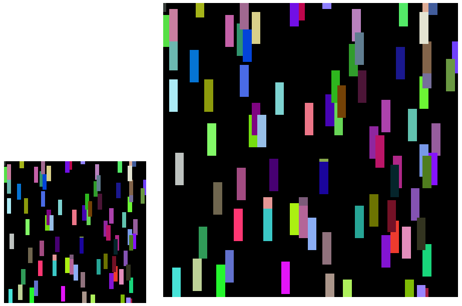
Markov
network



True

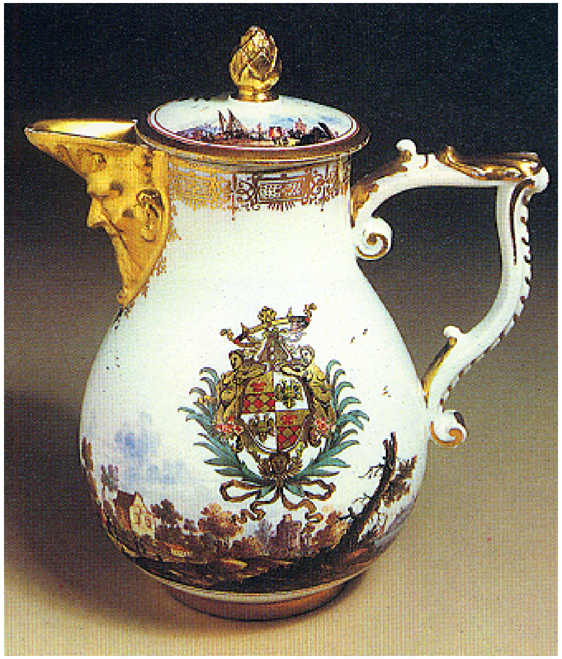
The Markov network algorithm hallucinates those vertical rectangles that it was trained on.

Original
50x58



Training images

Markov
network



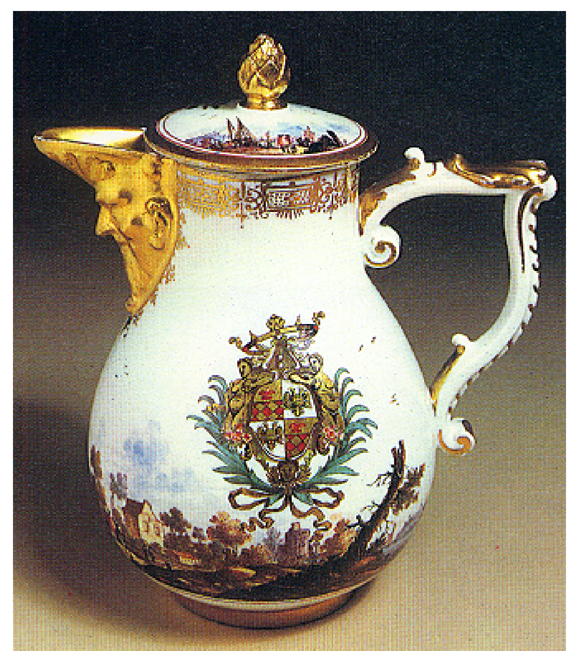
True

The algorithm makes a reasonable guess at the high resolution image, based on its training images.

Original
50x58

Training images

Markov
network



True

Image Blurring

Blur kernel

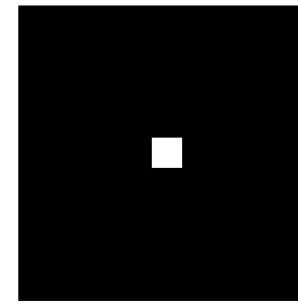
Input to algorithm

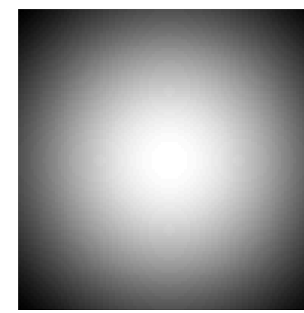
Desired output

Multiple possible solutions

Sharp image **Blur kernel**

Blurry image

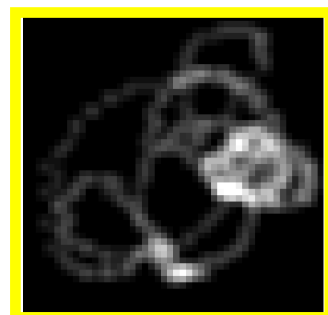




Three sources of information

1. Reconstruction constraint:

Estimated sharp image

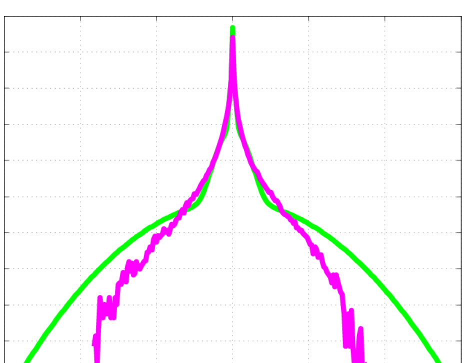


Estimated blur kernel

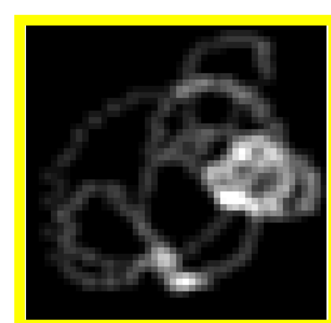
=

Input blurry image

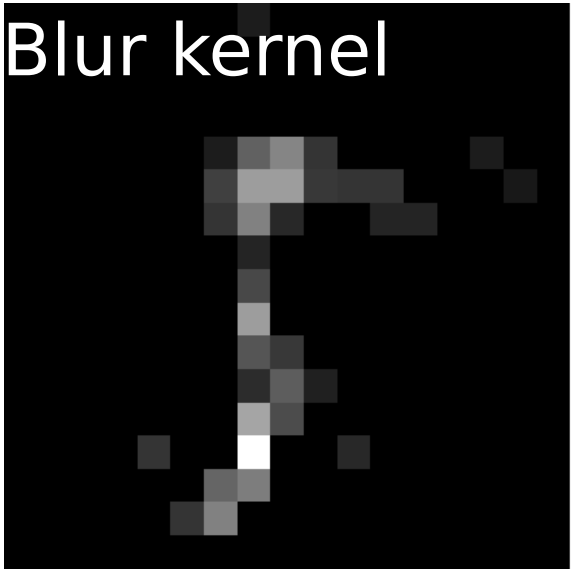
2. Image prior:



3. Blur prior:



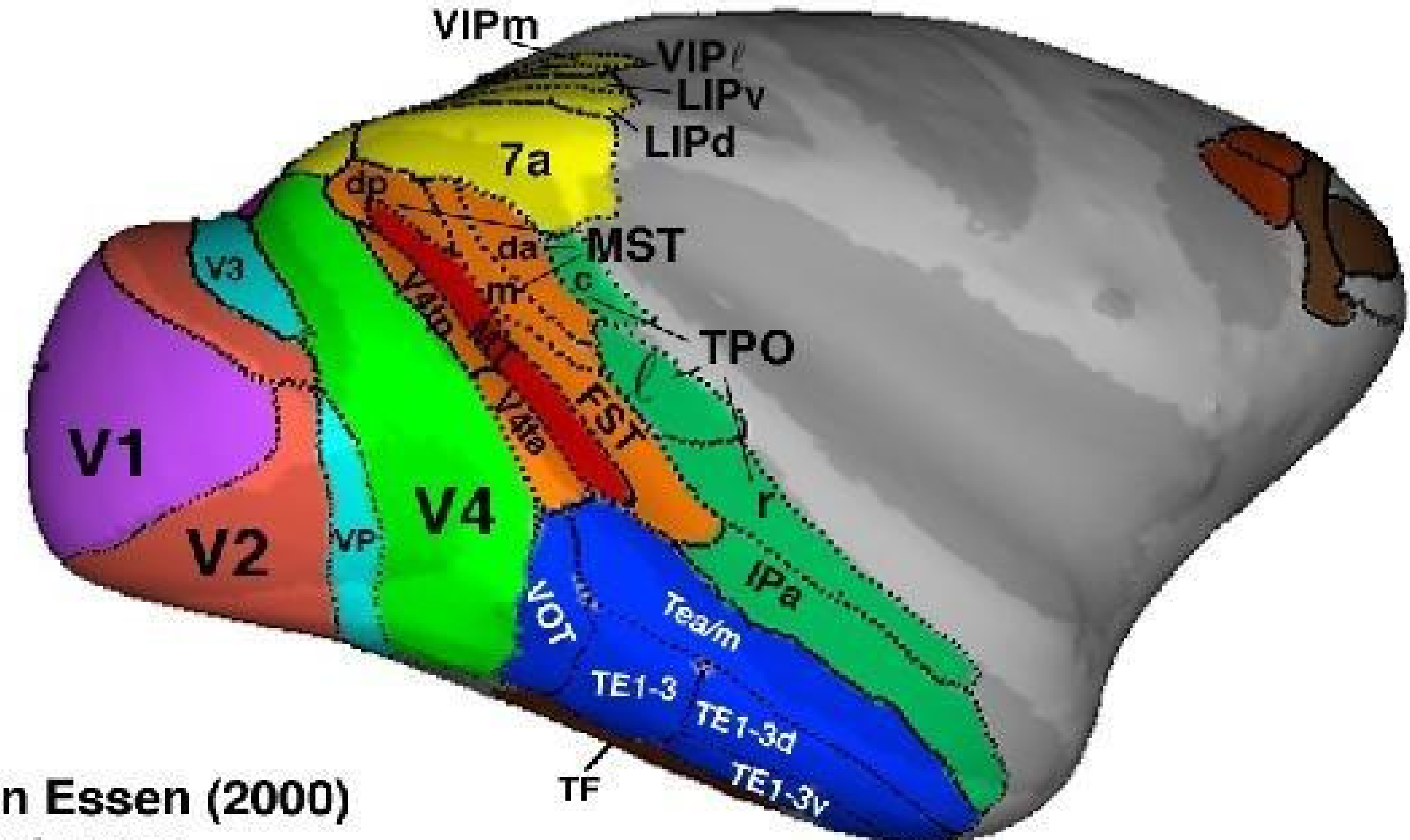
Original photograph


Blur kernel


Output

Human Vision

Primary Visual Cortex



Lewis & Van Essen (2000)
Visual areas

V1 neurons

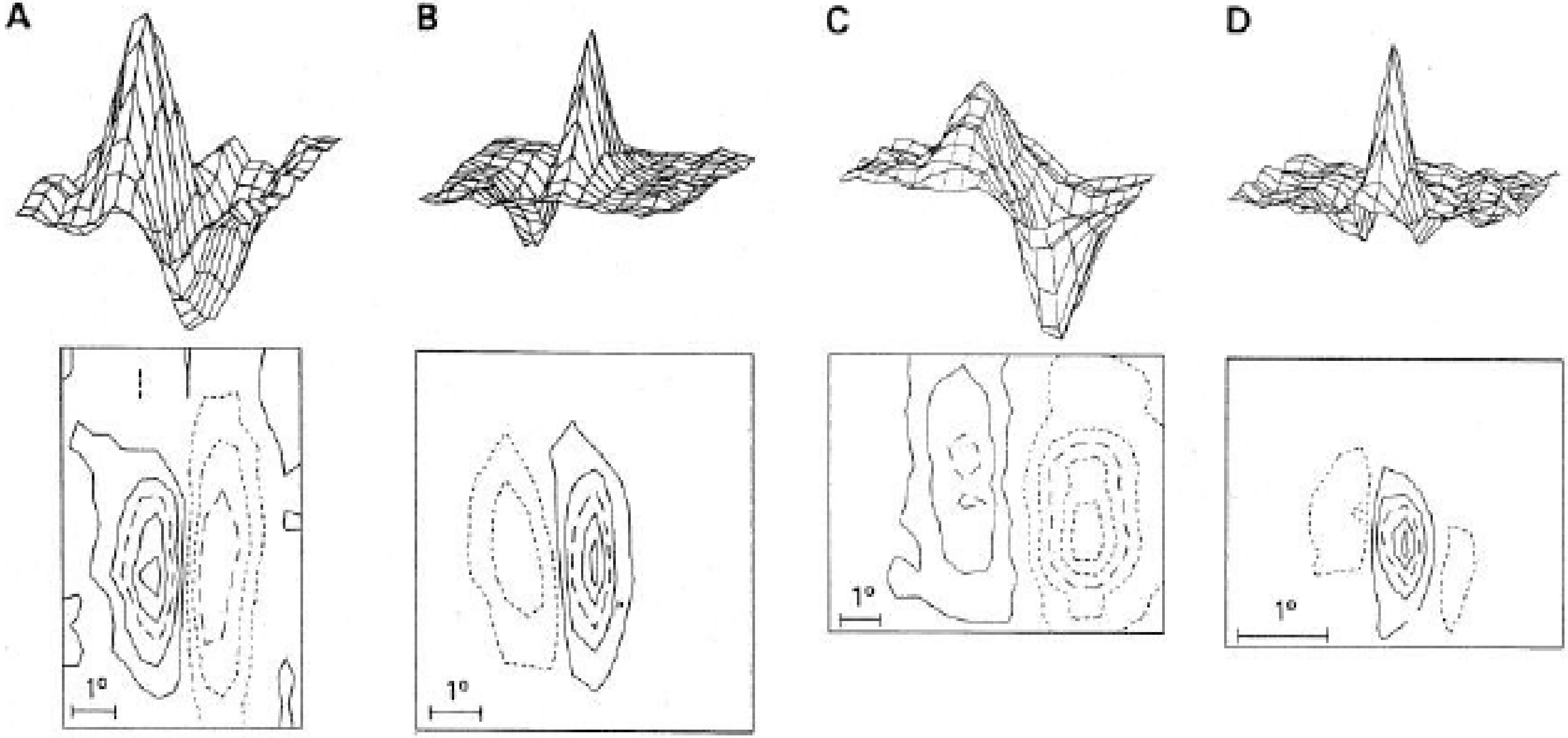
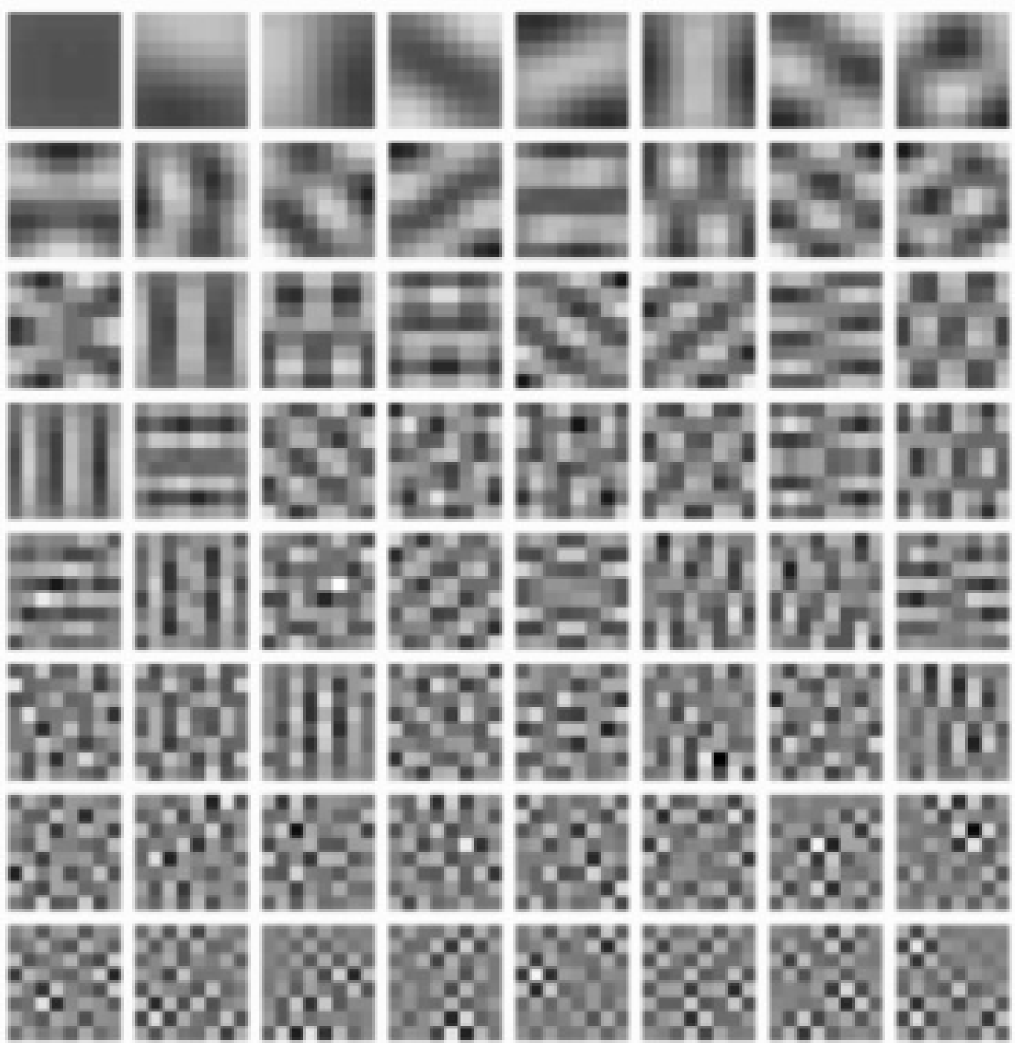
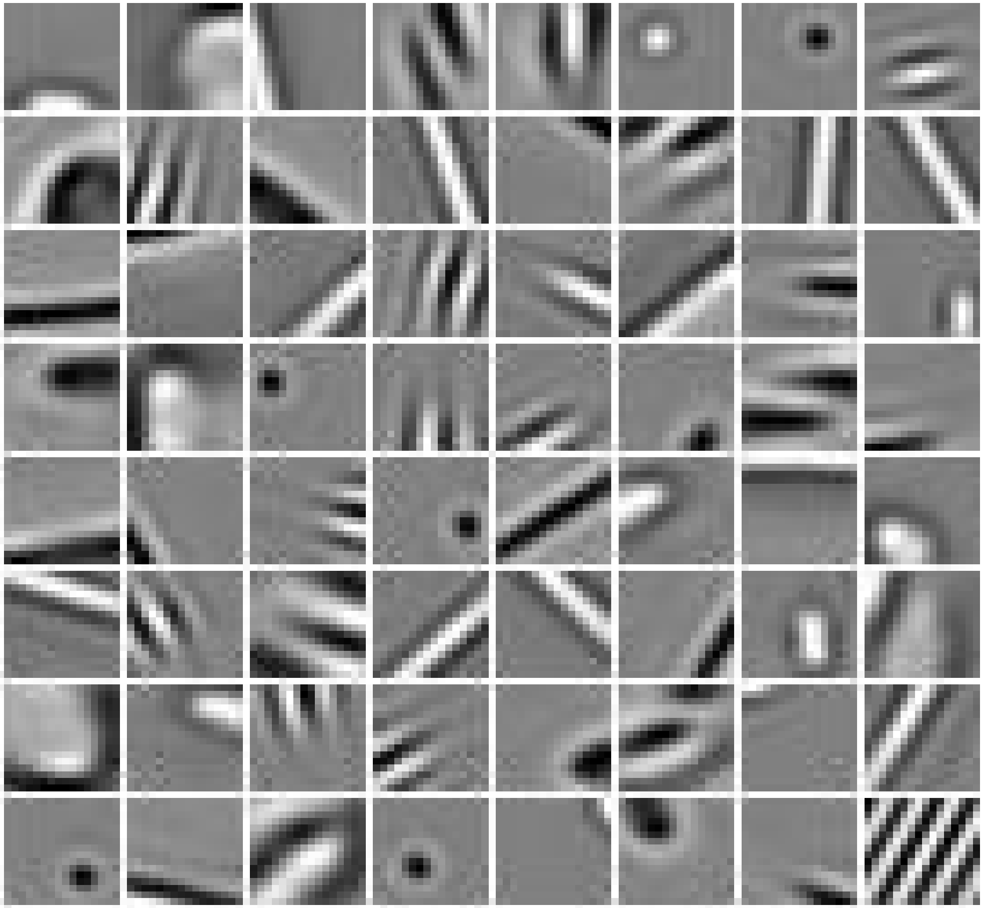


Image Patches



Better patches



The Vision debate

- How do we combine Low-Level and High Level features?
- One side – Understand how vision works and Use the features obtained from the vision research to train classifier systems better.
- Other side – Huge amount of data available, features can be found by the Classifiers themselves.

Gigantic Image Collections

Object Recognition for large-scale image search

Google Image Search

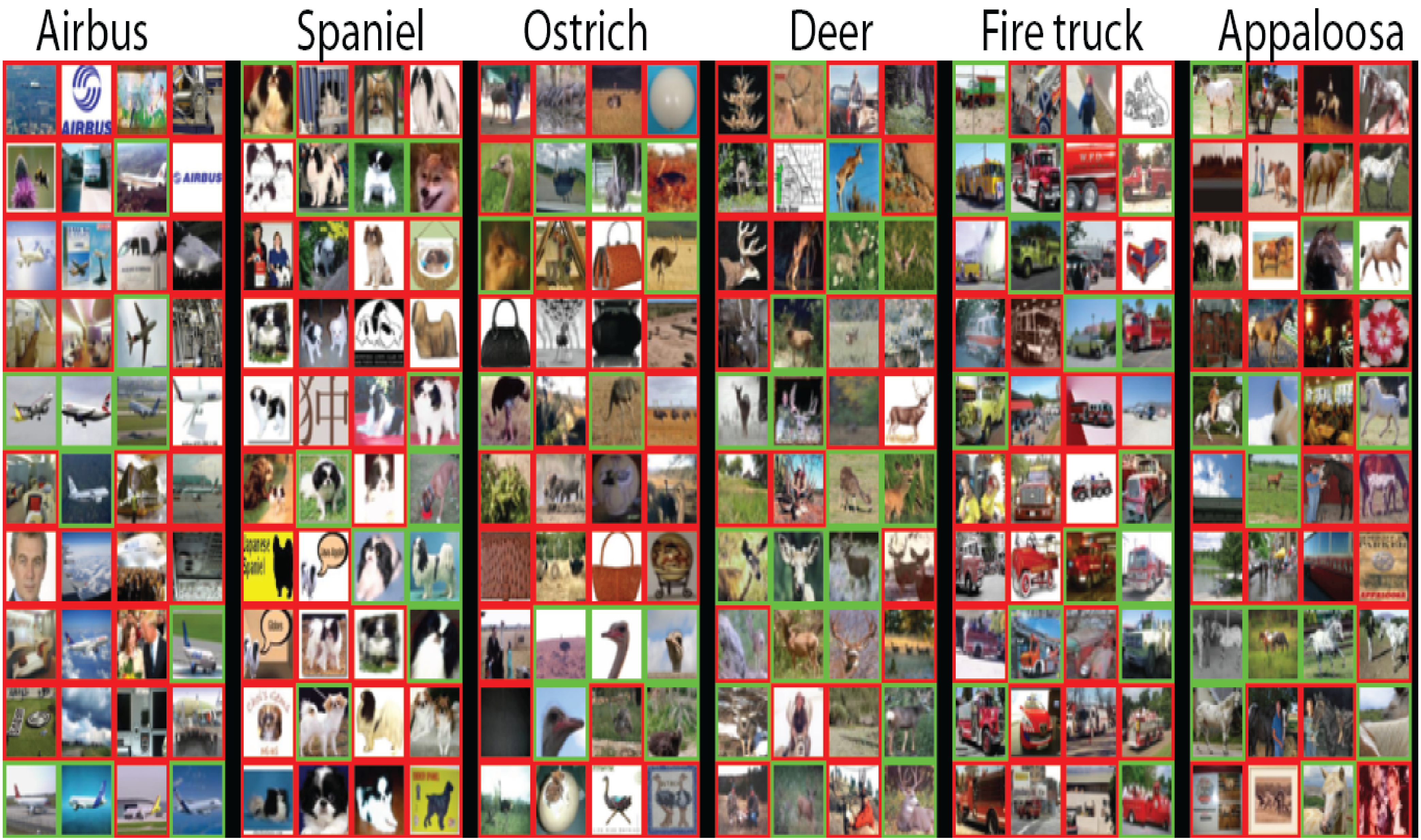
Spectrum of Label Information

Human annotations

Noisy labels

Unlabeled

Clustering



References -

Lectures and Presentations by the speakers of the Microsoft Research Winter School on Machine Learning and Computer Vision.
http://research.microsoft.com/en-us/events/winterschool2010/

Thank You

Questions