Image Features and Categorization

Computer Vision

CS 543 / ECE 549

University of Illinois

Derek Hoiem

Jia-Bin Huang

Where are we now?

- Object instance recognition

- Face recognition

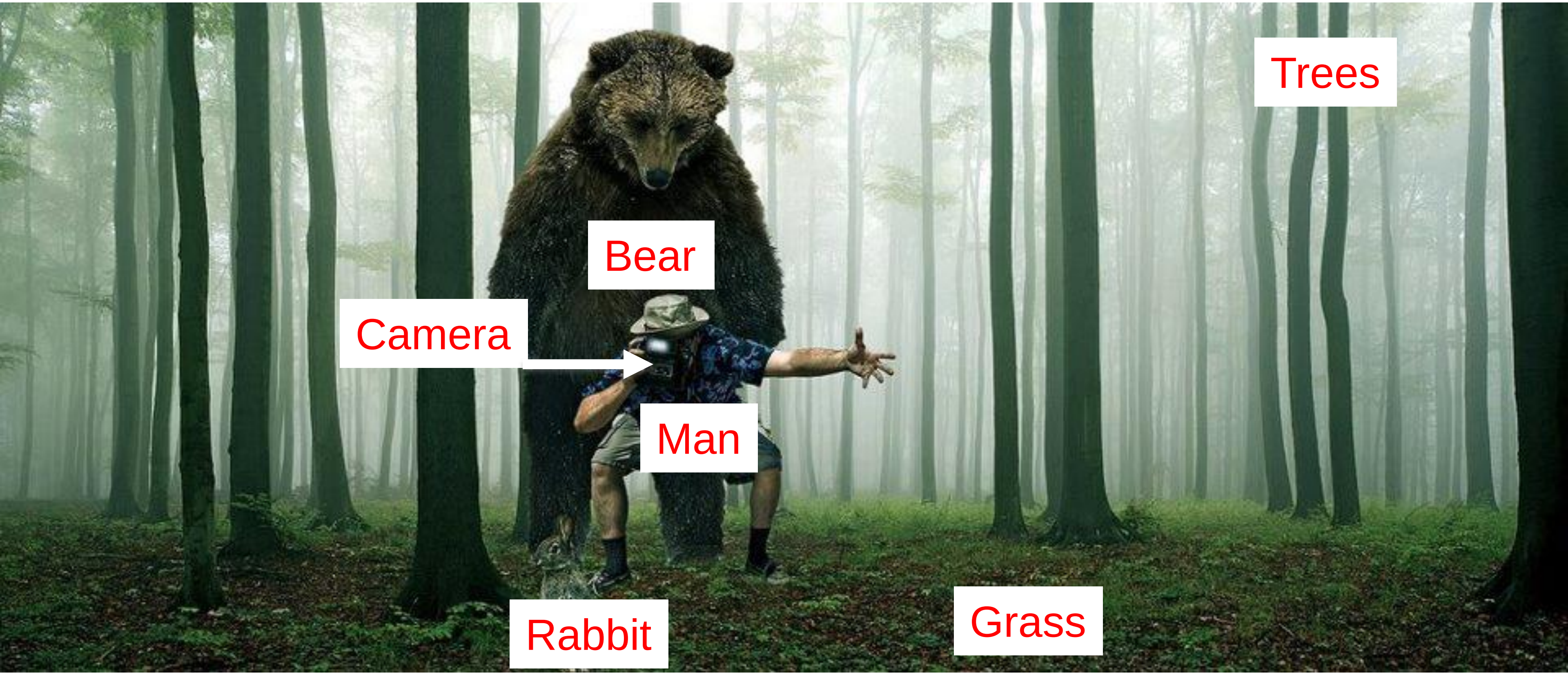
- **Today: Image features and categorization**

- Object detection
- Object tracking

Today: Image features and categorization

- General concepts of categorization
  - Why? What? How?
- Image features
  - Color, texture, gradient, shape, interest points
  - Histograms, feature encoding, and pooling
  - CNN as feature
- Image and region categorization

What do you see in this image?



Forest

Describe, predict, or interact with the object based on visual cues

Is it **dangerous**?

Is it **alive**?

How **fast** does it run?

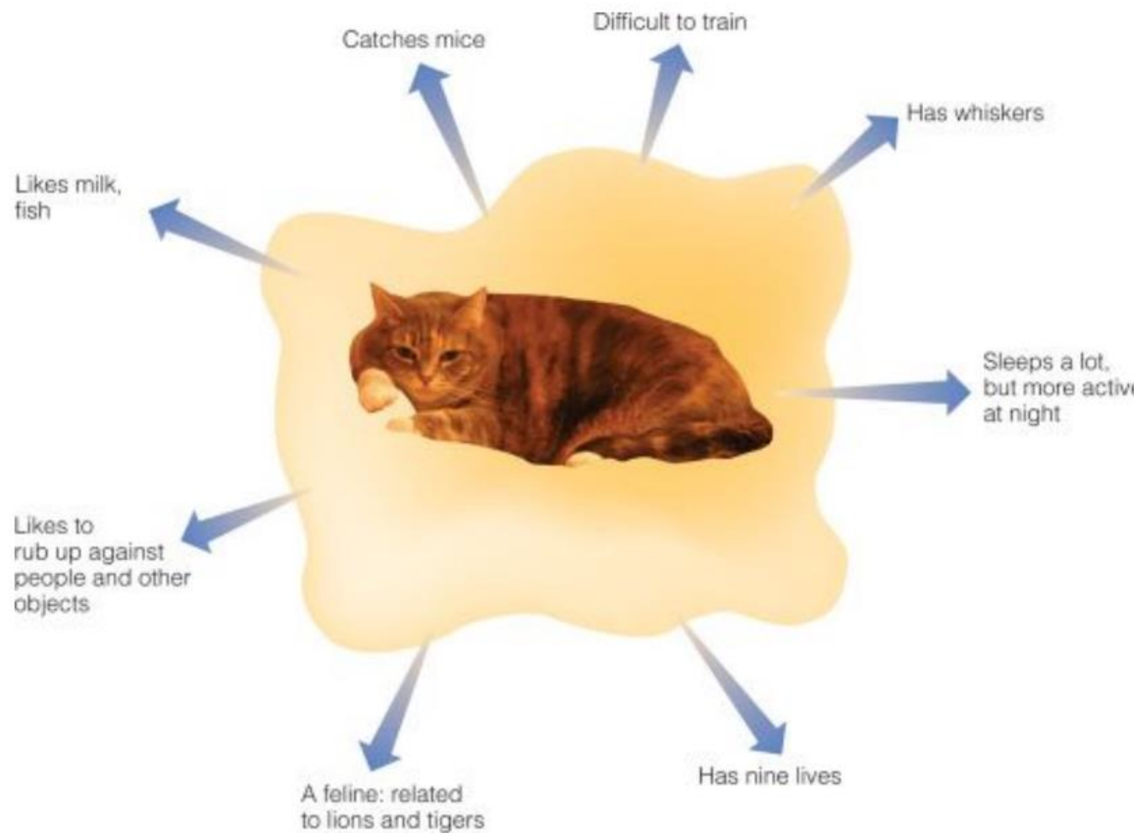
Is it **soft**?

Does it have a **tail**?

Can I **poke with it**?

Why do we care about categories?

- From an object's category, we can make predictions about its behavior in the future, beyond of what is immediately perceived.
- Pointers to knowledge
  - Help to understand individual cases not previously encountered
- Communication



Theory of categorization

How do we determine if something is a member of a particular category?

- Definitional approach
- Prototype approach
- Exemplar approach

Definitional approach: classical view of categories

- Plato & Aristotle
  - Categories are defined by a list of properties shared by all elements in a category
  - Category membership is binary
  - Every member in the category is equal

Aristotle by Francesco Hayez

The Categories (Aristotle)

Slide Credit: A. A. Efros

Prototype or sum of exemplars ?

Prototype Model

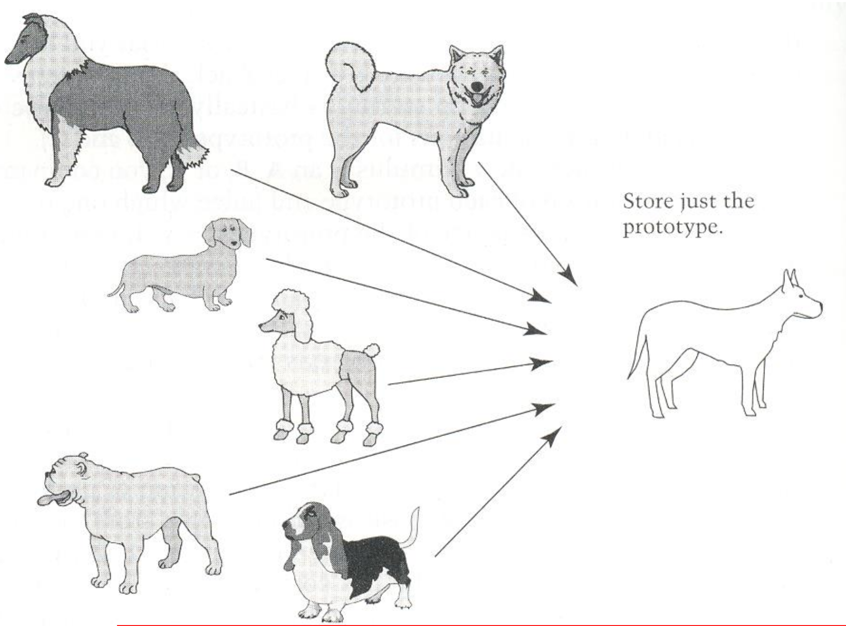


Figure 7.3. *Schematic of the prototype model. Although many exemplars are seen, only the prototype is stored. The prototype is updated continually to incorporate more experience with new exemplars.*

Category judgments are made by comparing a new exemplar to the prototype.

Exemplars Model

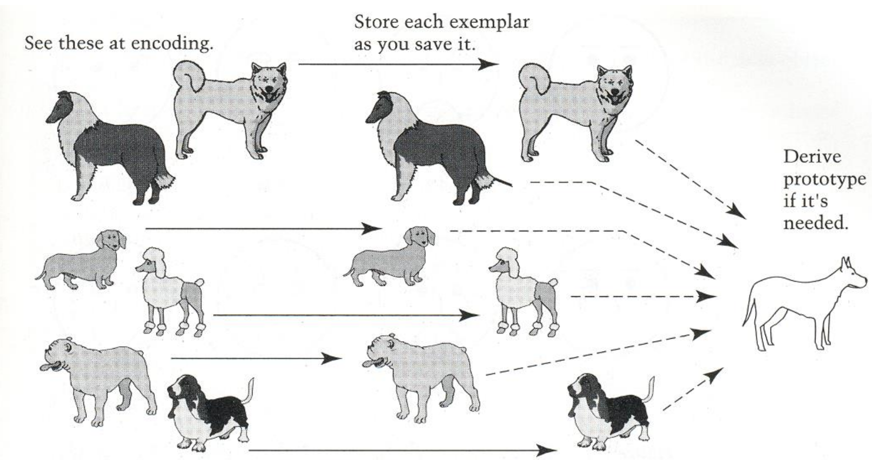


Figure 7.4. *Schematic of the exemplar model. As each exemplar is seen, it is encoded into memory. A prototype is abstracted only when it is needed, for example, when a new exemplar must be categorized.*

Category judgments are made by comparing a new exemplar to all the old exemplars of a category or to the exemplar that is the most appropriate

Slide Credit: Torralba

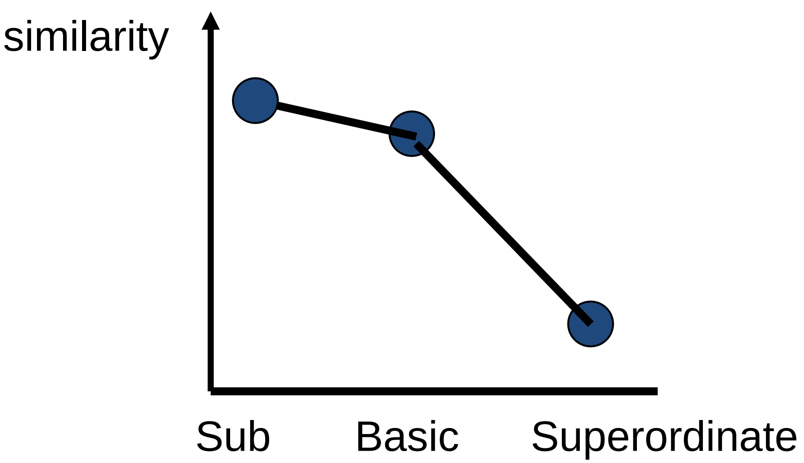
Levels of categorization [Rosch 70s]

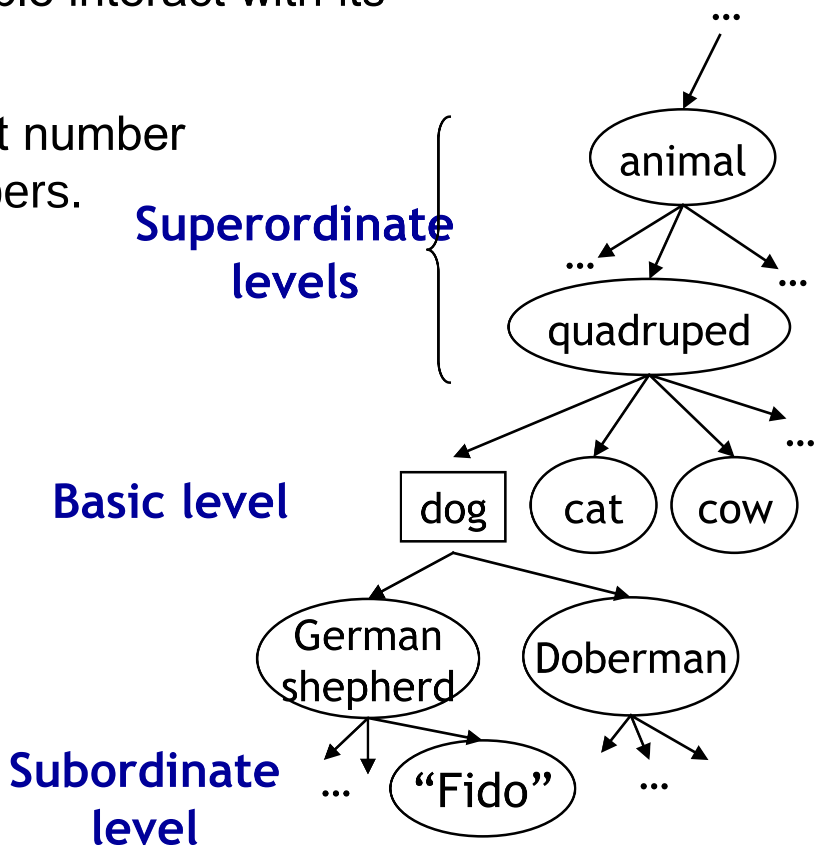
Definition of Basic Level:

• **Similar shape**: Basic level categories are the highest-level category for which their members have similar shapes.

• **Similar motor interactions**: … for which people interact with its members using similar motor sequences.

• **Common attributes**: … there are a significant number of attributes in common between pairs of members.





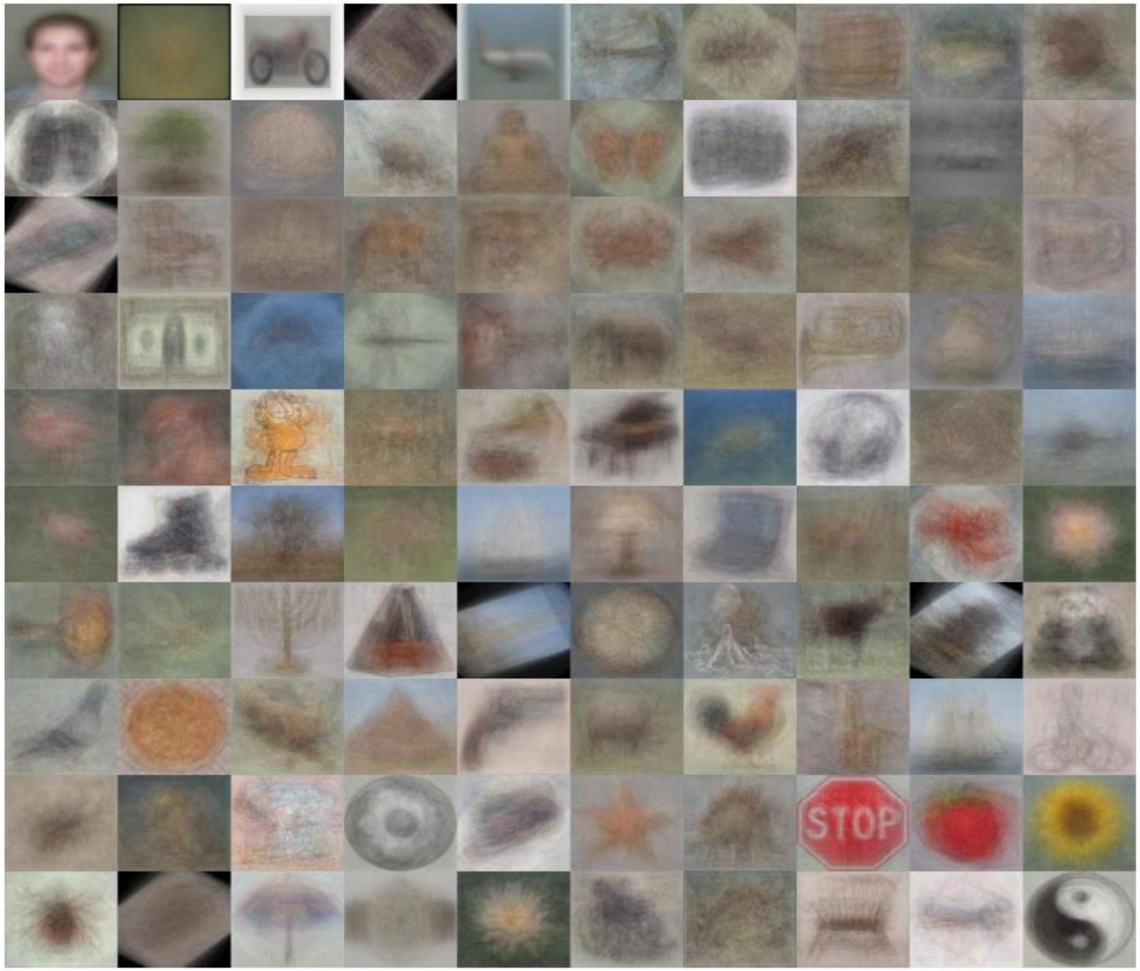
Rosch et a. Principle of categorization, 1978

Image categorization

- Cat vs Dog

Image categorization

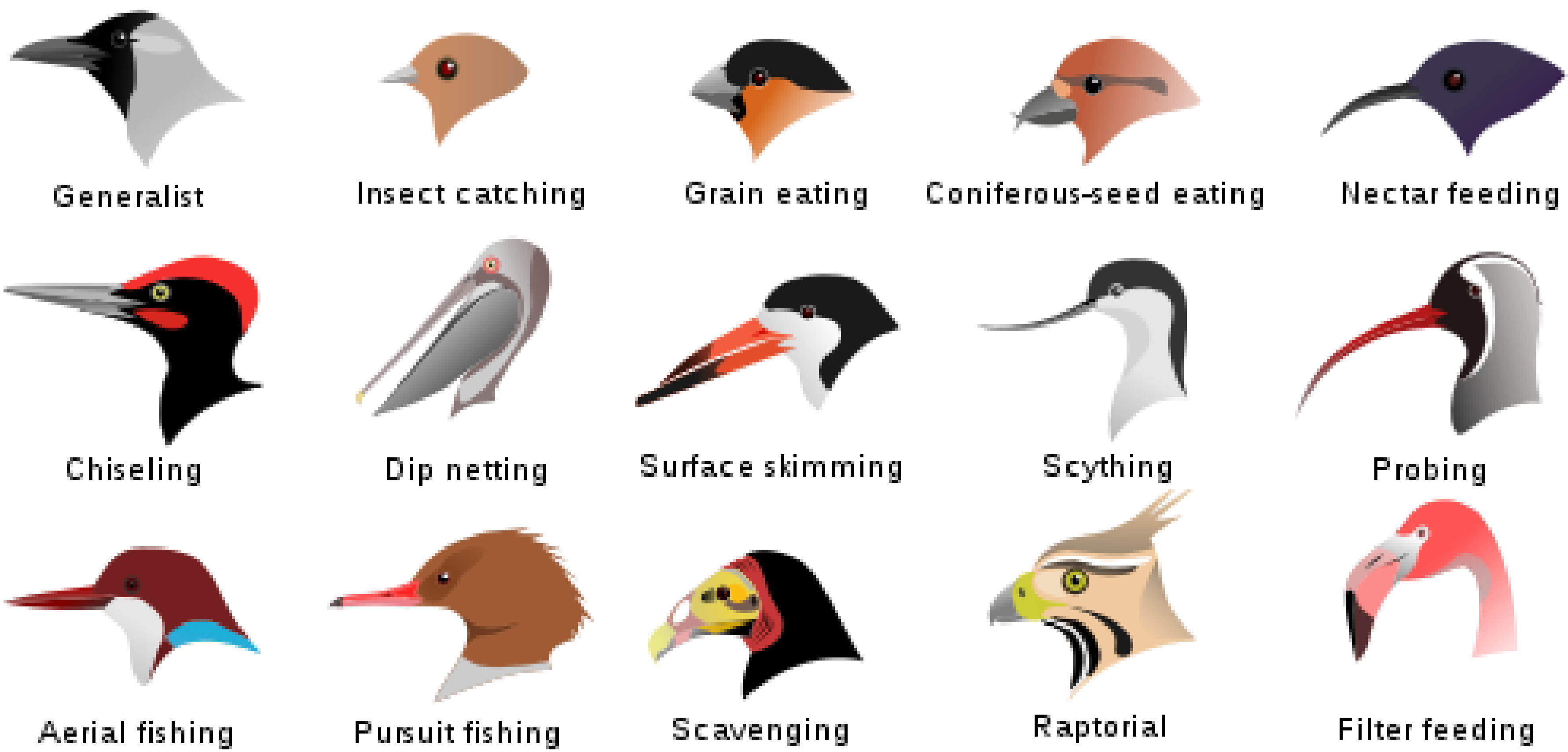
- Object recognition



Caltech 101 Average Object Images

Image categorization

- Fine-grained recognition



Visipedia Project

Image categorization

- Place recognition

Places Database [Zhou et al. NIPS 2014]

Image categorization

- Visual font recognition

[Chen et al. CVPR 2014]

Image categorization

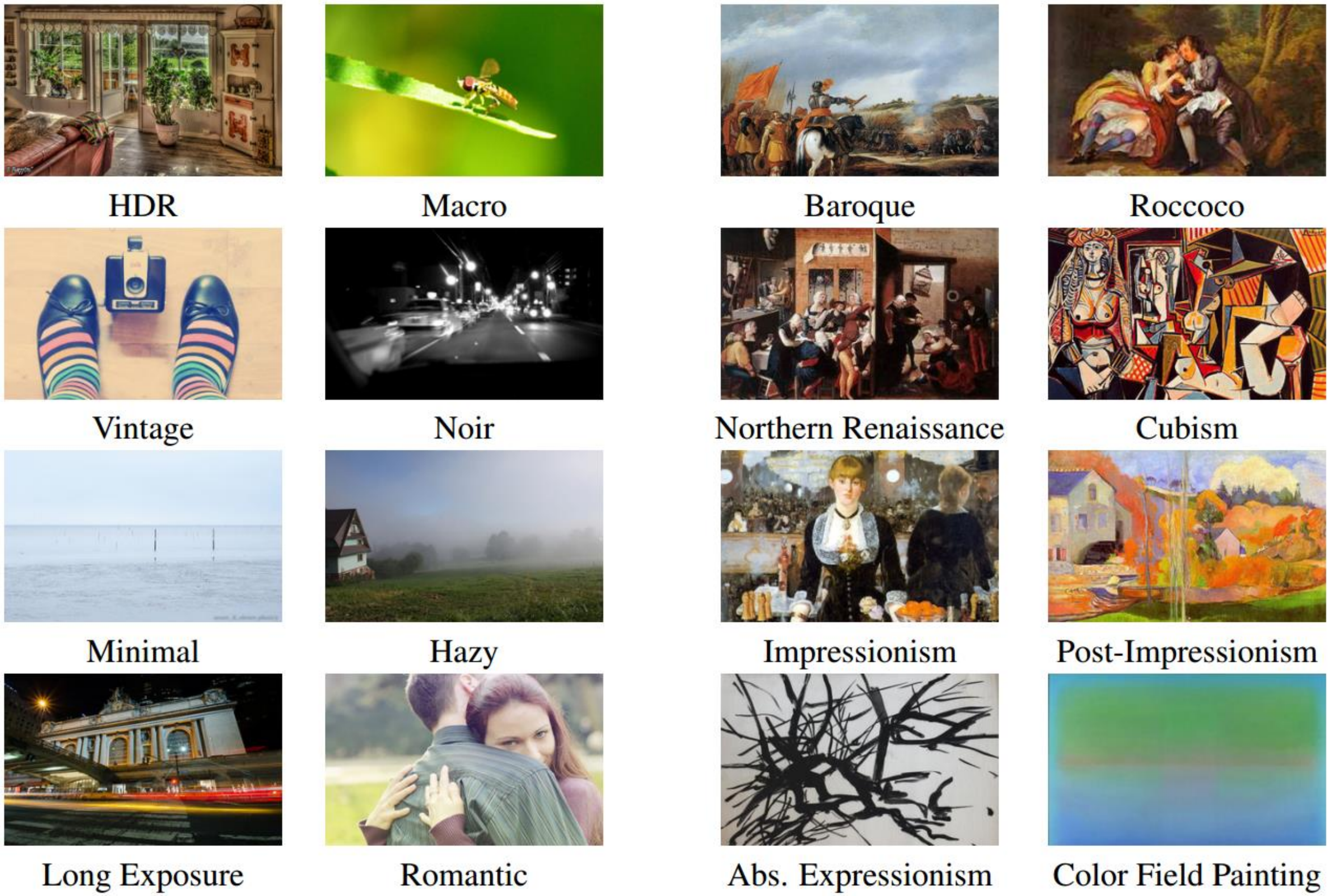
- Dating historical photos

1940 1953 1966 1977

[Palermo et al. ECCV 2012]

Image categorization

- Image style recognition

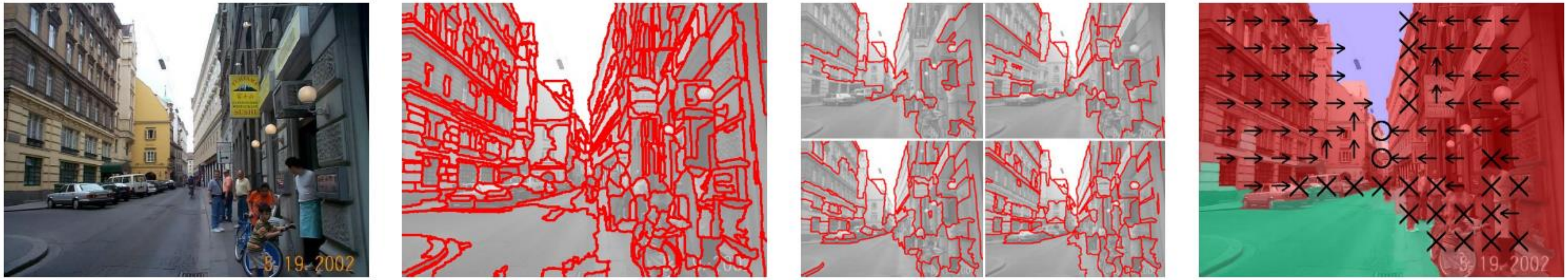


Flickr Style: 80K images covering 20 styles. Wikipaintings: 85K images for 25 art genres.

[Karayev et al. BMVC 2014]

Region categorization

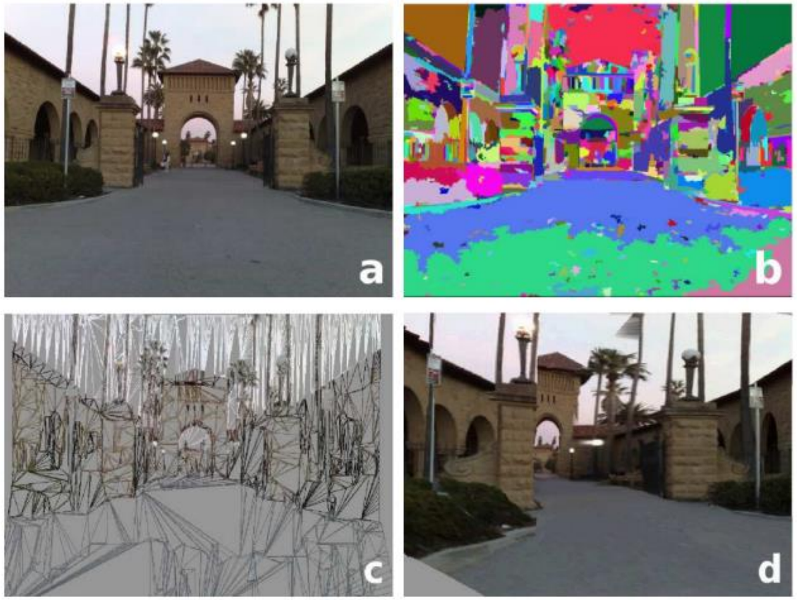
- Layout prediction



Input Superpixels Multiple Segmentations Surface Layout

Assign regions to orientation

Geometric context [Hoiem et al. IJCV 2007]

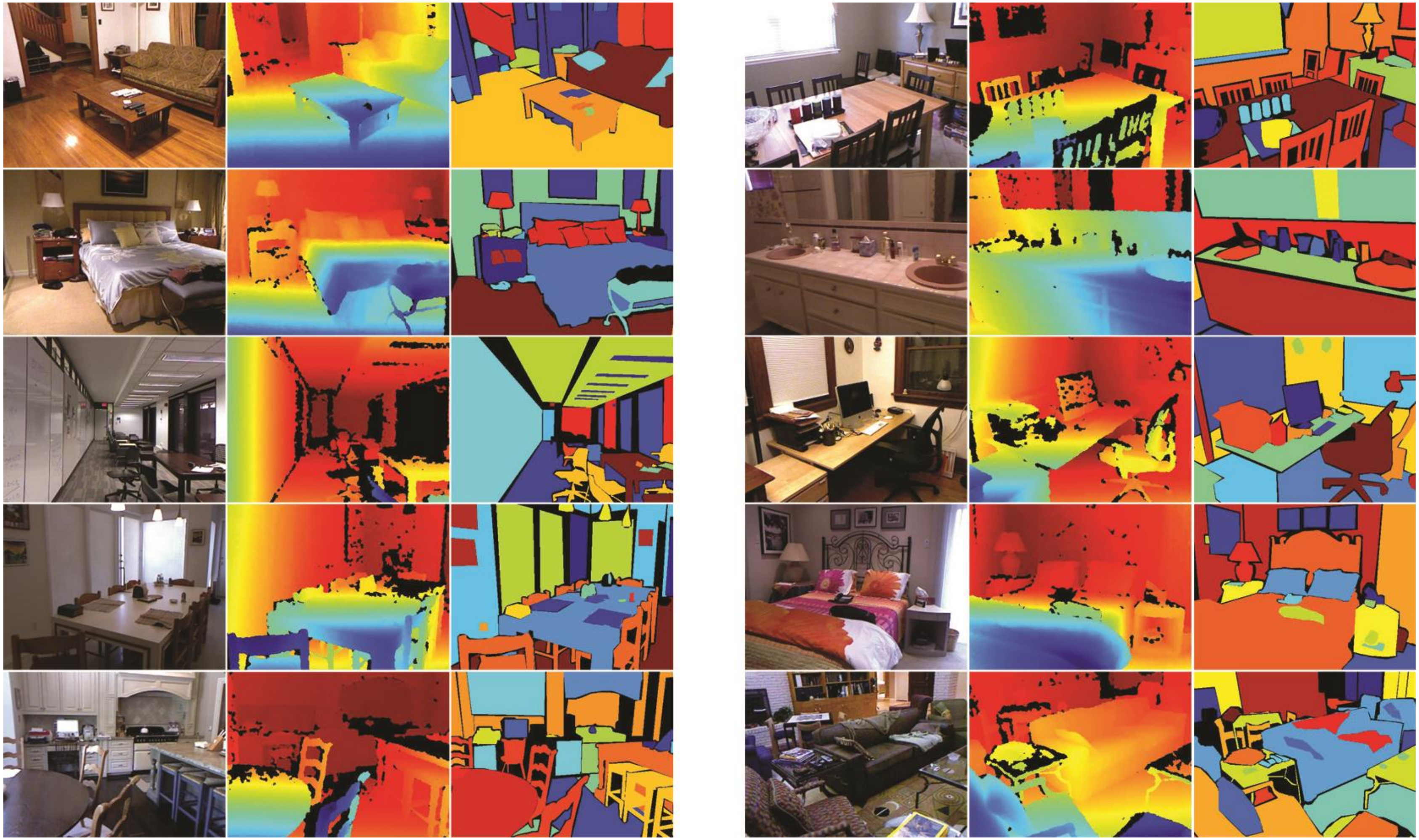


Assign regions to depth

Make3D [Saxena et al. PAMI 2008]

Region categorization

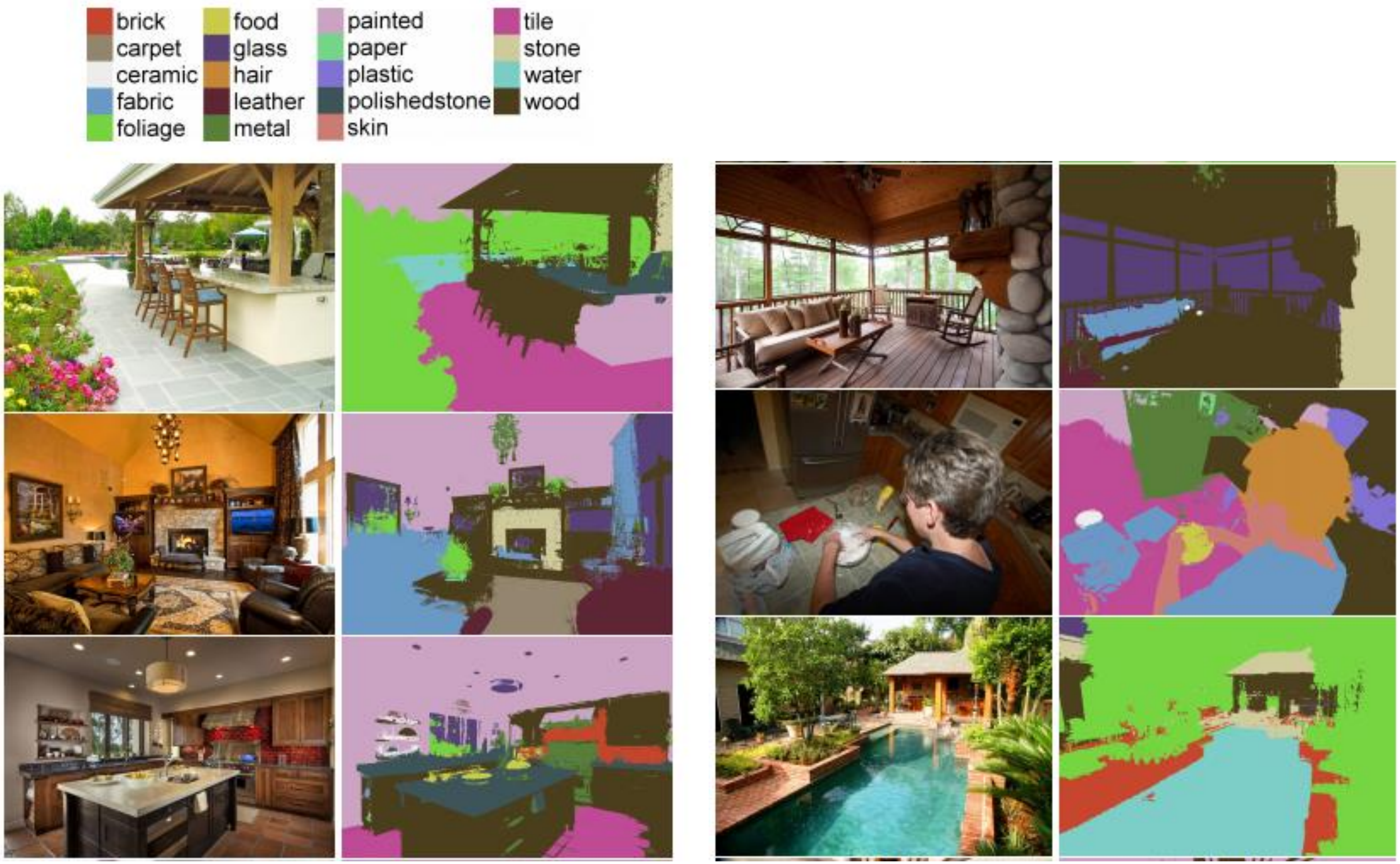
- Semantic segmentation from RGBD images



[Silberman et al. ECCV 2012]

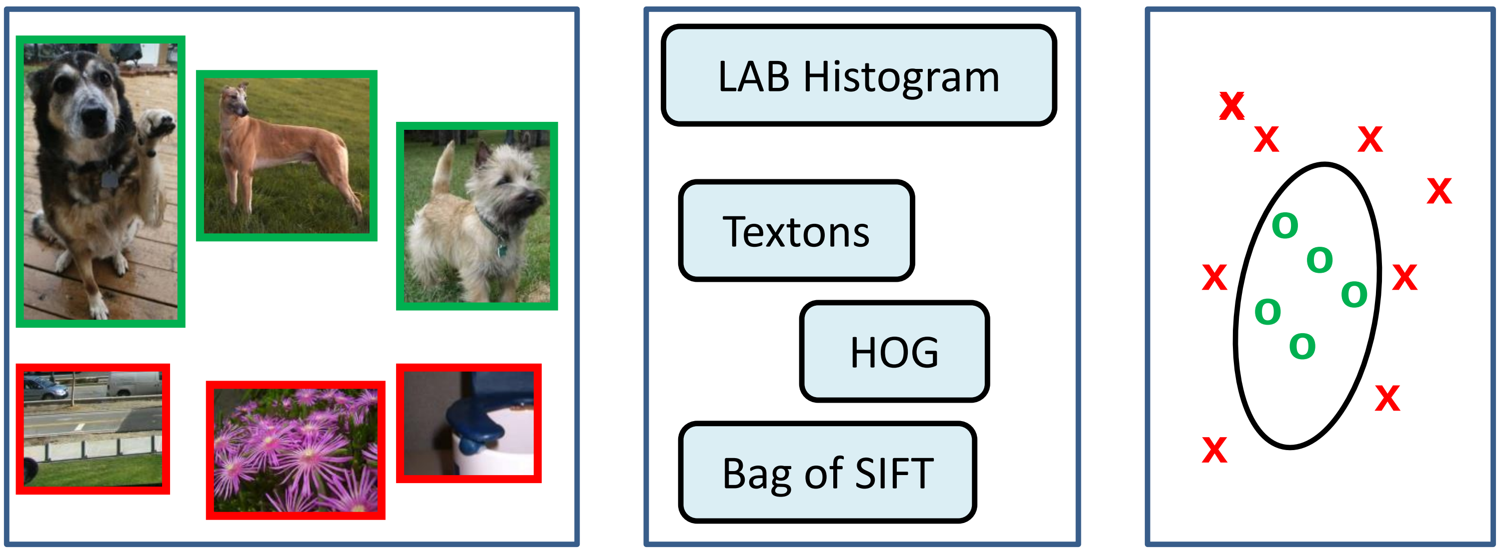
Region categorization

- Material recognition



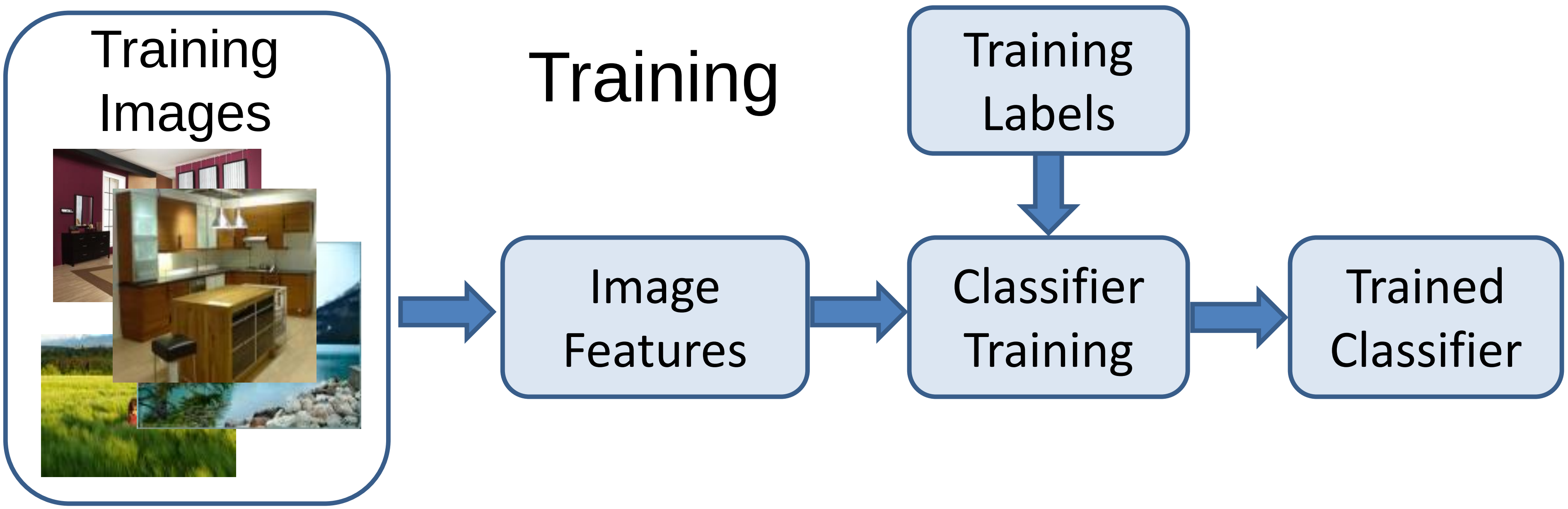
[Bell et al. CVPR 2015]

Supervised learning

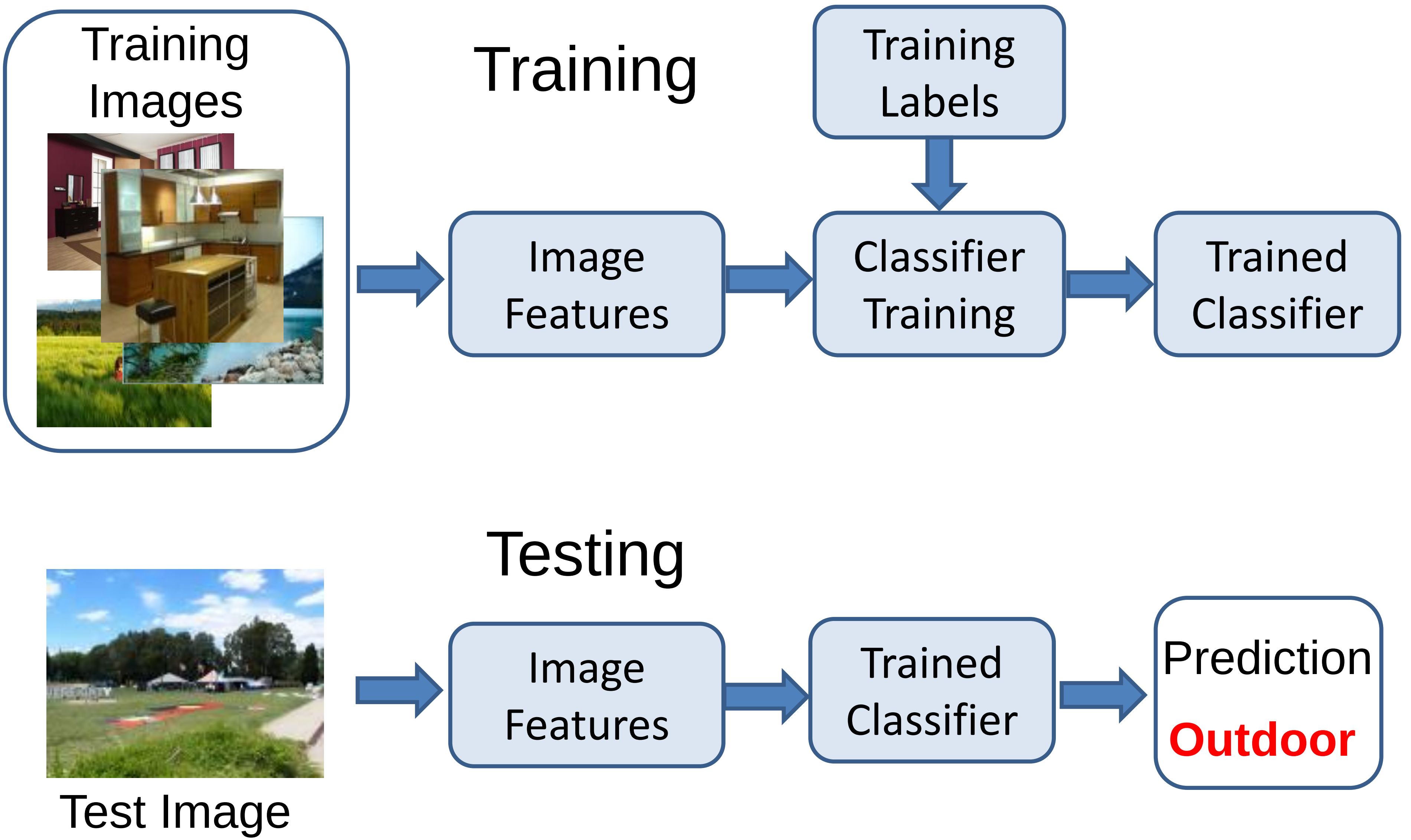


Examples + Image Features + Classifier = Category label

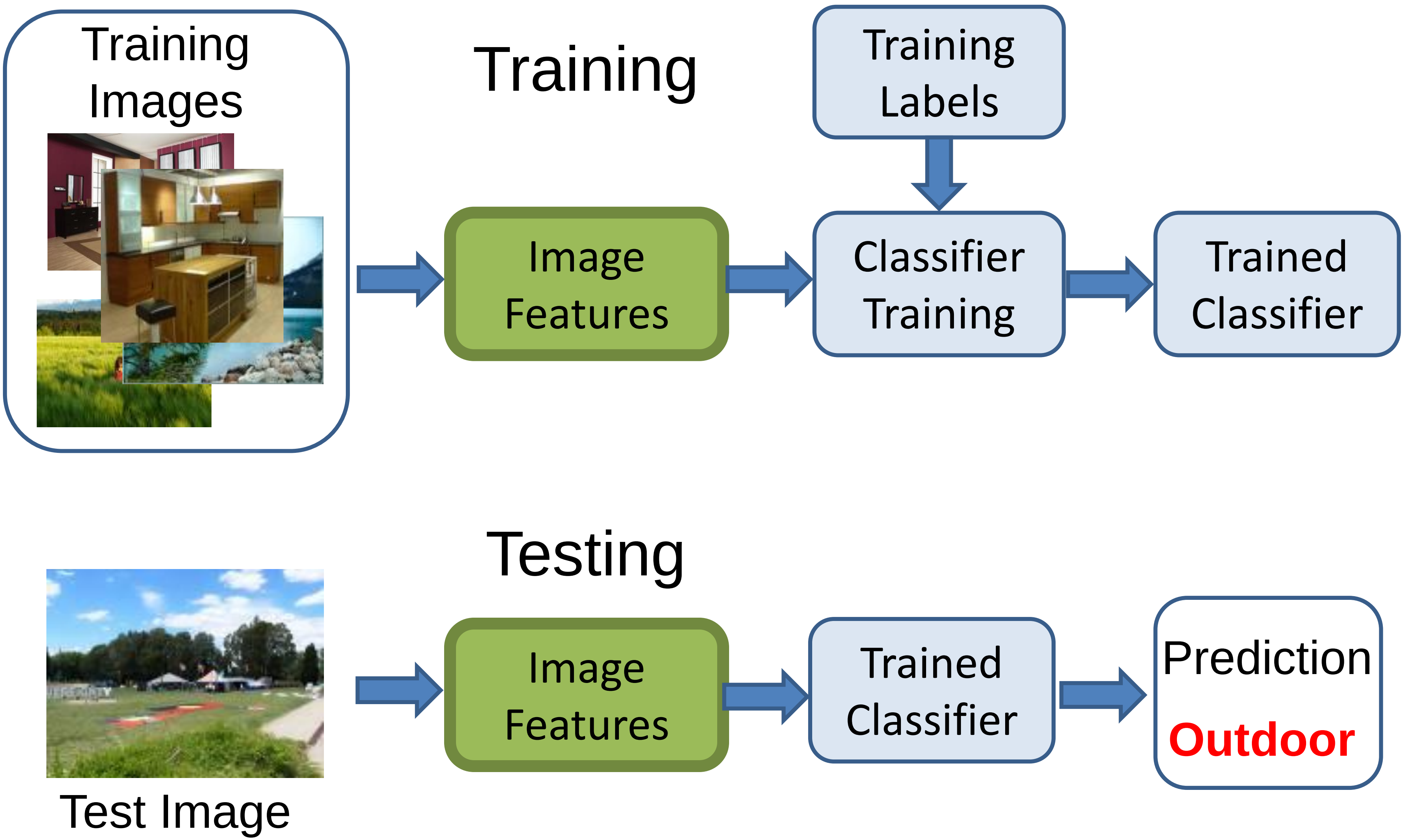
Training phase



Testing phase



Testing phase



Q: What are good features for…

- recognizing a beach?






Q: What are good features for…

- recognizing cloth fabric?






Q: What are good features for…

- recognizing a mug?






What are the right features?

Depend on what you want to know!

- Object: shape
  - Local shape info, shading, shadows, texture
- Scene : geometric layout
  - linear perspective, gradients, line segments
- Material properties: albedo, feel, hardness
  - Color, texture
- Action: motion
  - Optical flow, tracked points

General principles of representation

- Coverage
  - Ensure that all relevant info is captured
- Concision
  - Minimize number of features without sacrificing coverage
- Directness
  - Ideal features are independently useful for prediction

Image representations

- Templates
  - Intensity, gradients, etc.
- Histograms
  - Color, texture, SIFT descriptors, etc.
- Average of features

Image Intensity

Gradient template

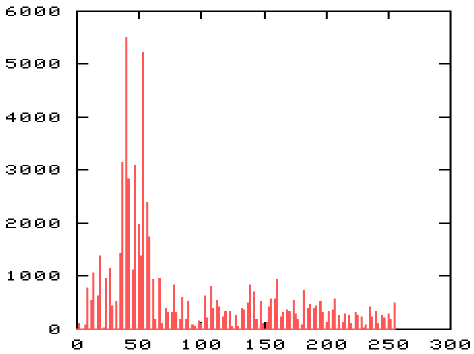
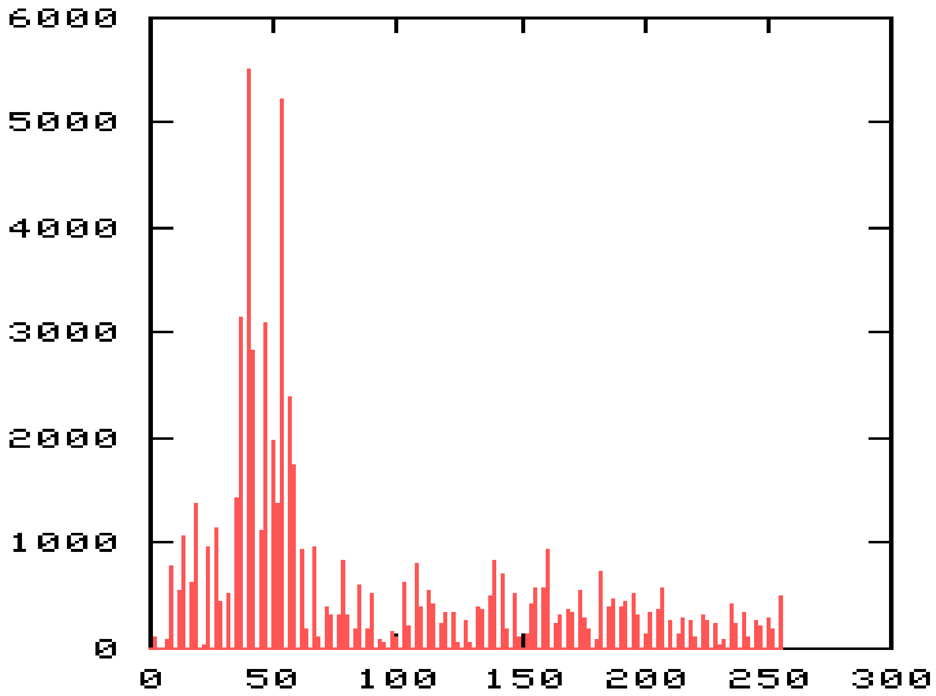
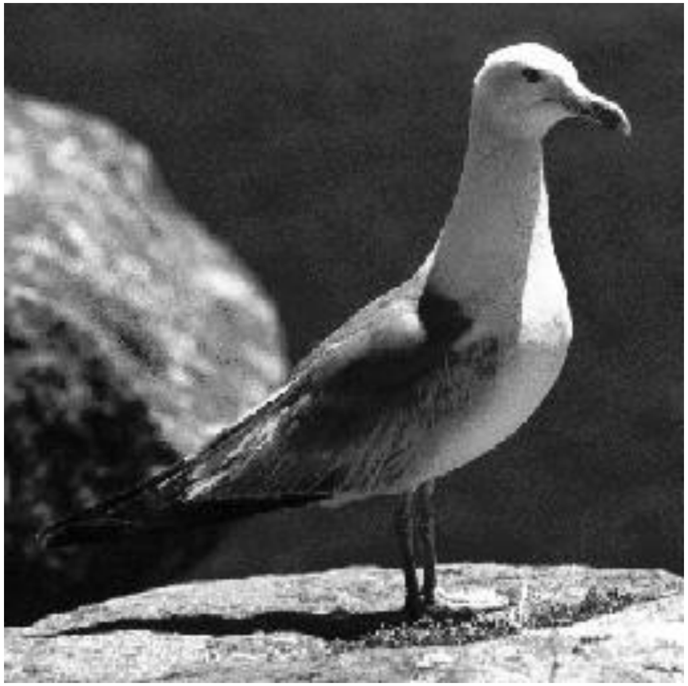


Image representations: histograms





Global histogram

- Represent distribution of features
  - Color, texture, depth, …

Images from Dave Kauchak

Image representations: histograms

- Data samples in 2D

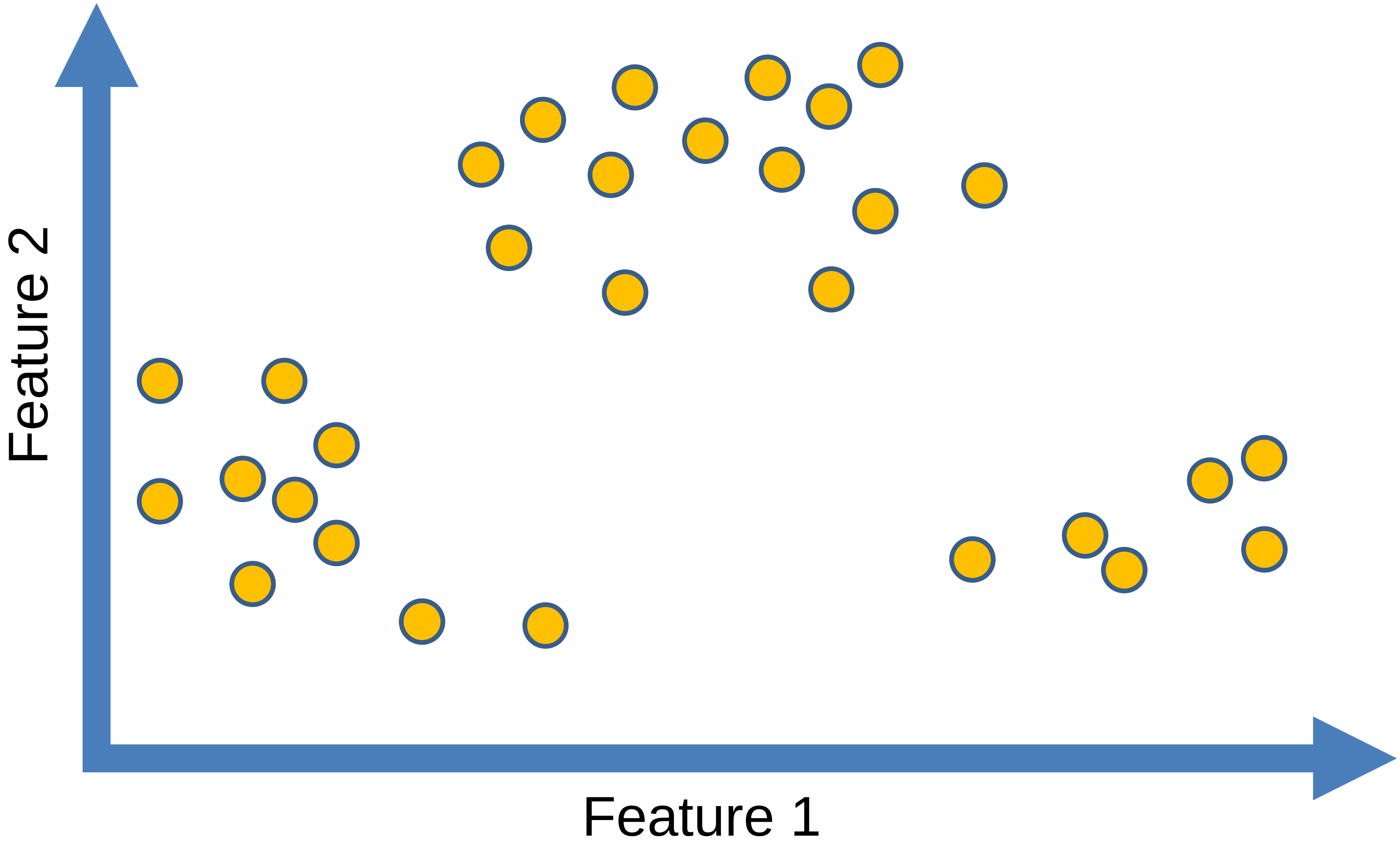


Image representations: histograms

- Probability or count of data in each bin
- Marginal histogram on feature 1

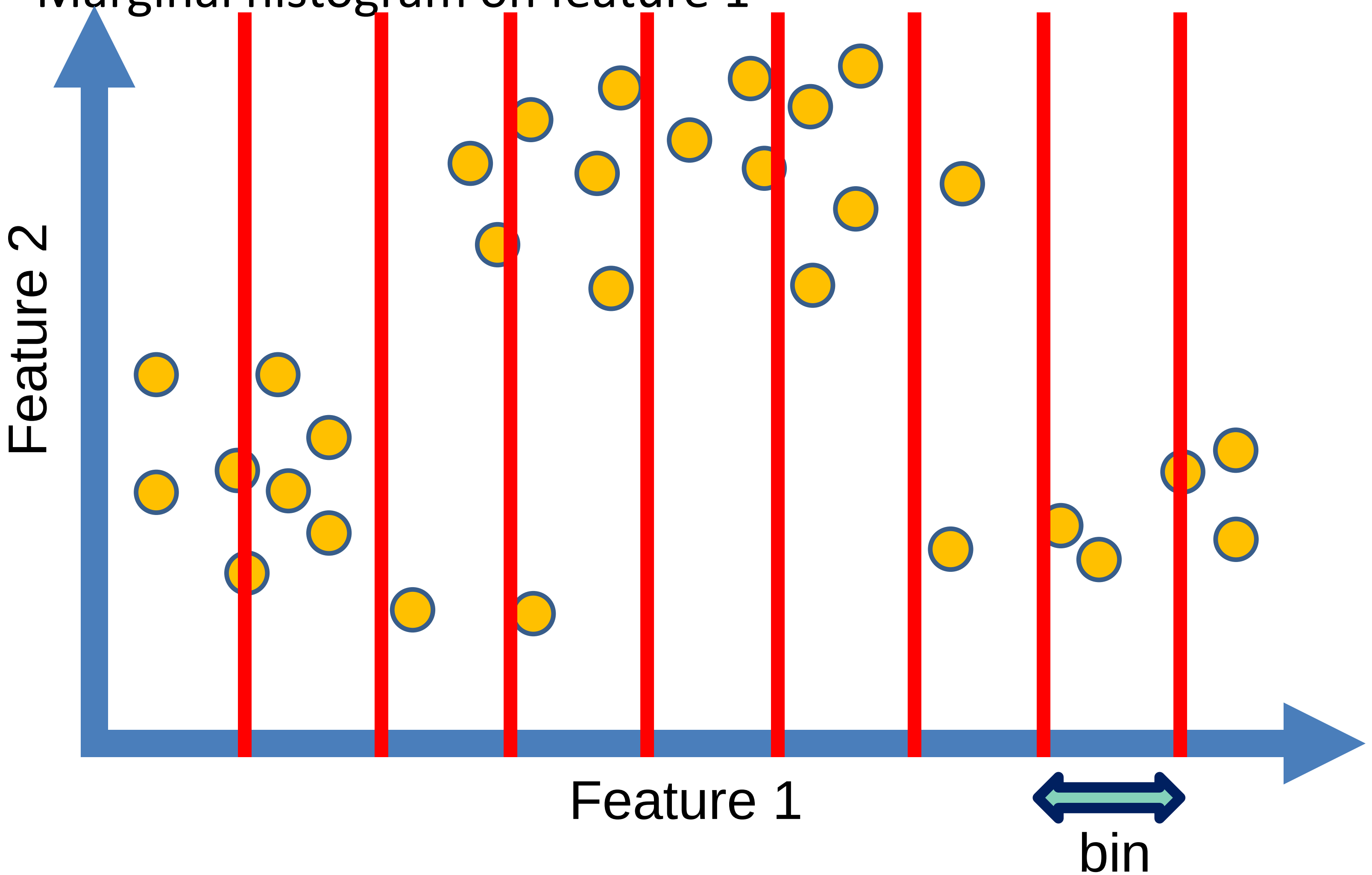


Image representations: histograms

- Marginal histogram on feature 2

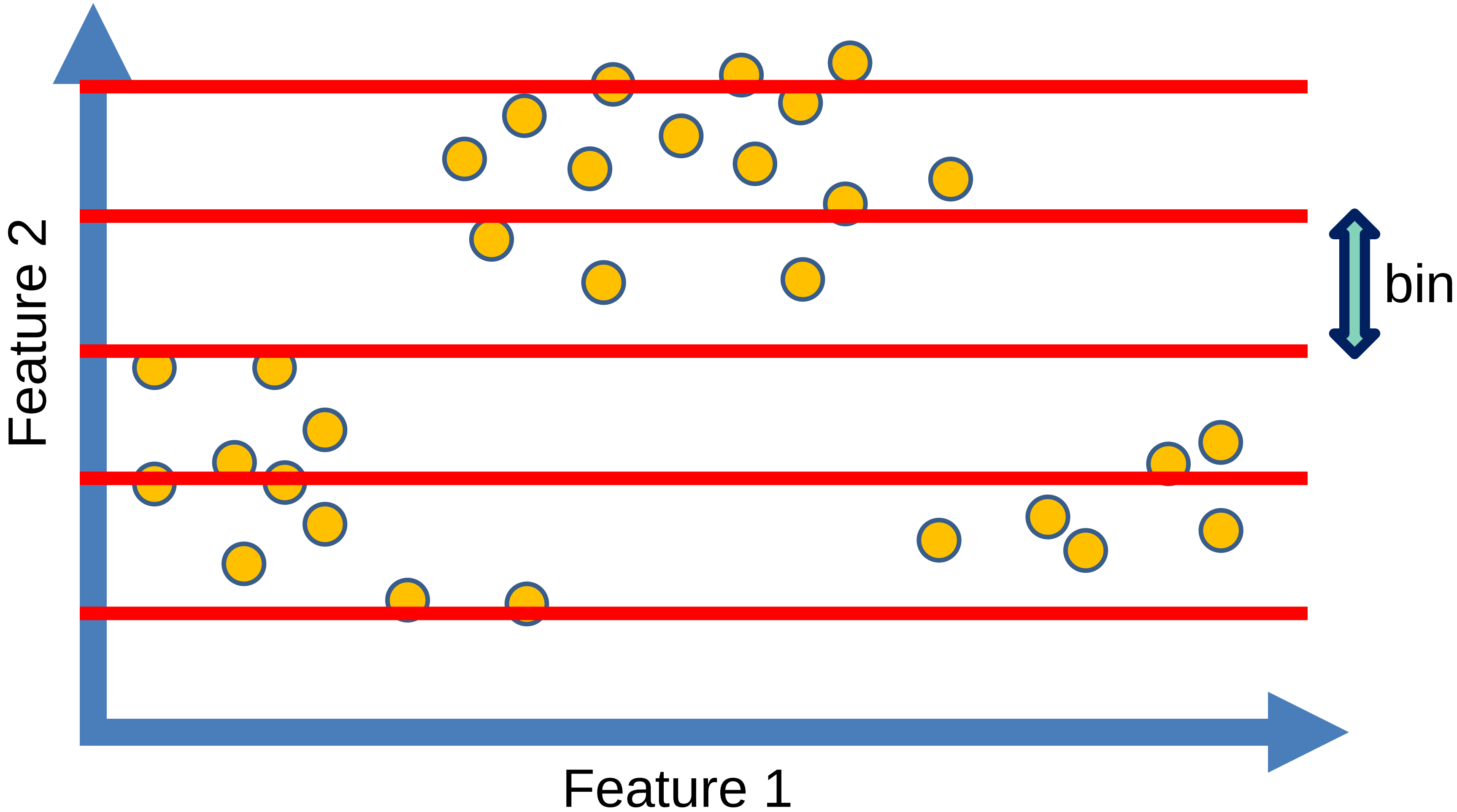
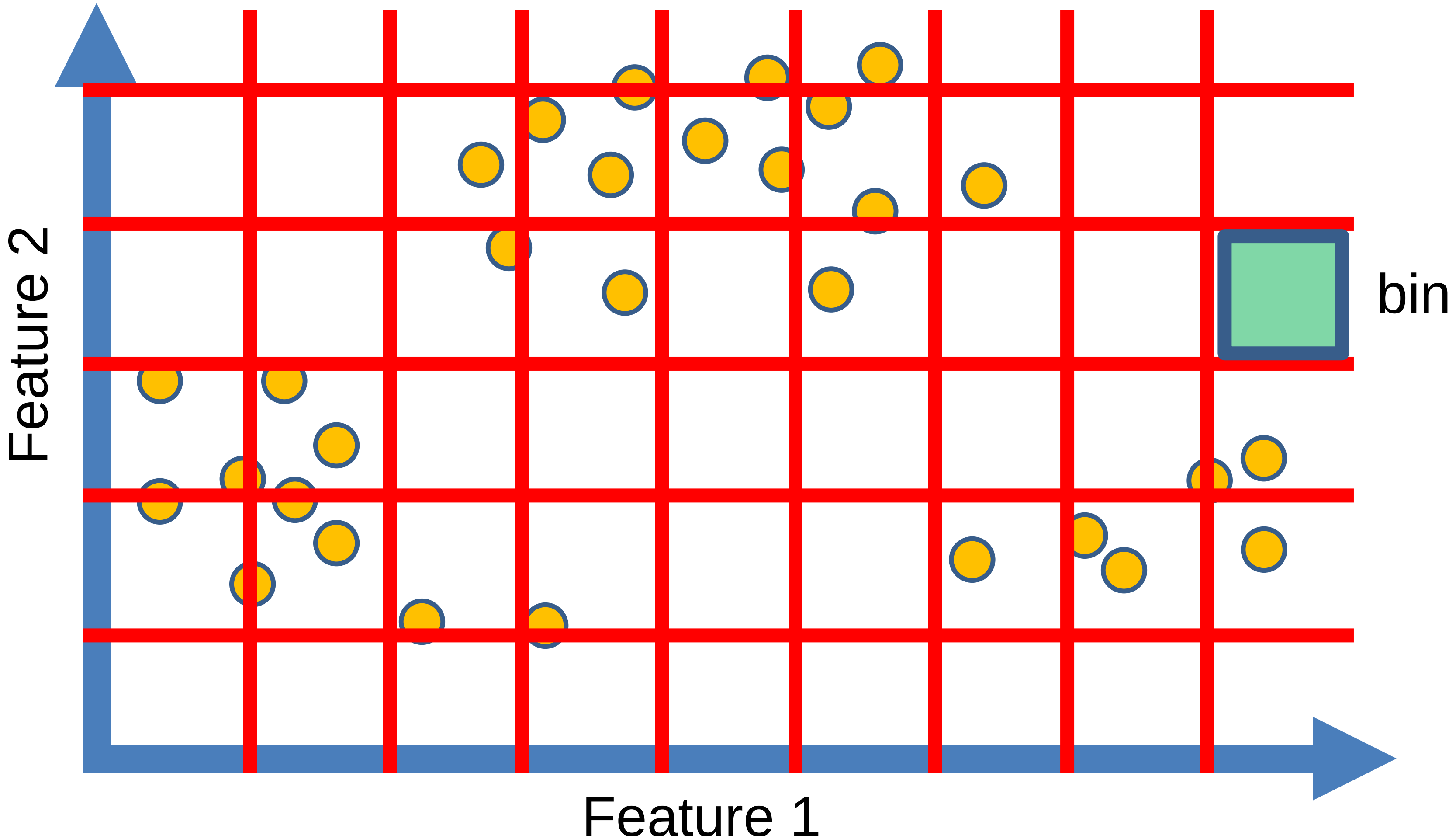
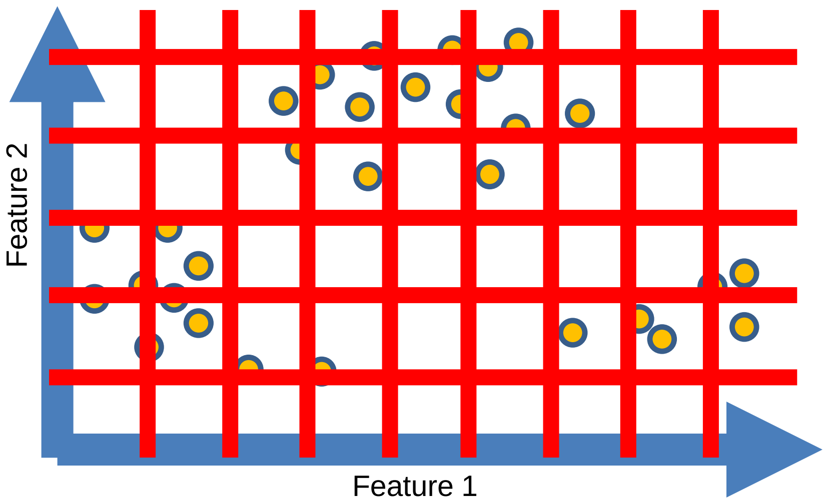


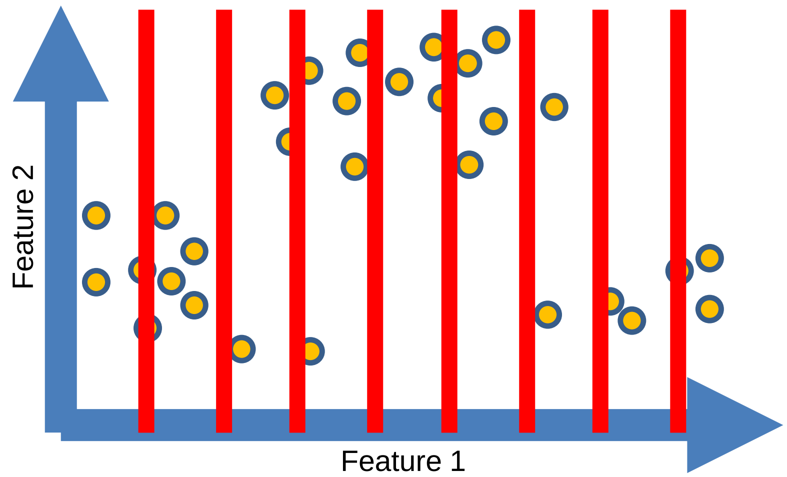
Image representations: histograms

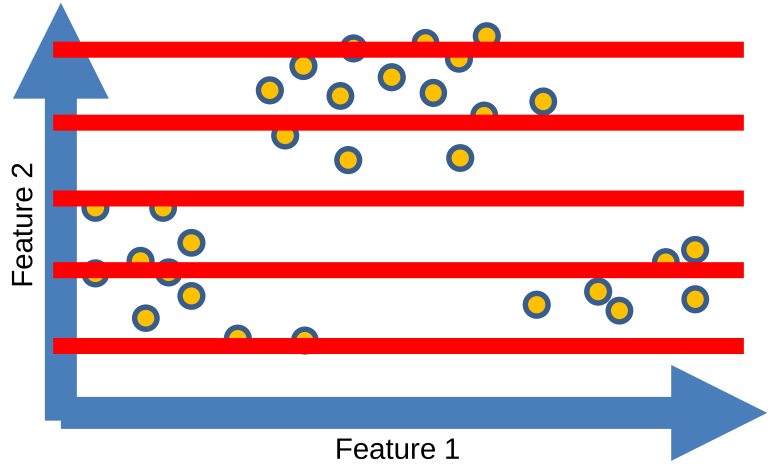
- Joint histogram



Modeling multi-dimensional data







Joint histogram

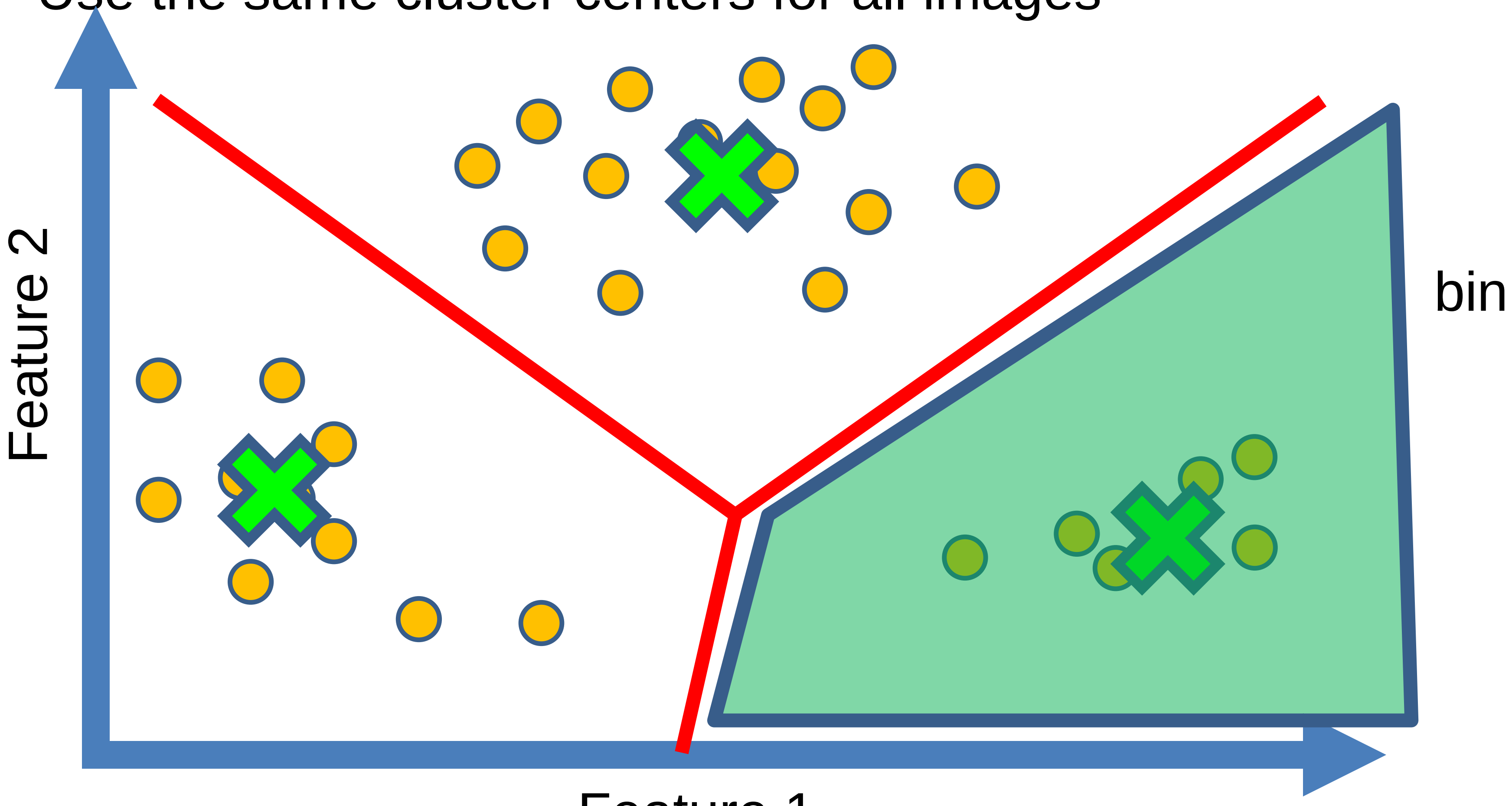
- Requires lots of data
- Loss of resolution to avoid empty bins

Marginal histogram

- Requires independent features
- More data/bin than joint histogram

Modeling multi-dimensional data

- Clustering
- Use the same cluster centers for all images



Computing histogram distance

- Histogram intersection

$$\text{histint}(h_i, h_j) = 1 - \sum_{m=1}^{K} \min\left(h_i(m), h_j(m)\right)$$

- Chi-squared Histogram matching distance

$$\chi^2(h_i, h_j) = \frac{1}{2} \sum_{m=1}^{K} \frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)}$$

- Earth mover's distance
(Cross-bin similarity measure)
  - minimal cost paid to transform one distribution into the other

[Rubner et al. The Earth Mover's Distance as a Metric for Image Retrieval, IJCV 2000]

Histograms: implementation issues

- Quantization
  - Grids: fast but applicable only with few dimensions
  - Clustering: slower but can quantize data in higher dimensions

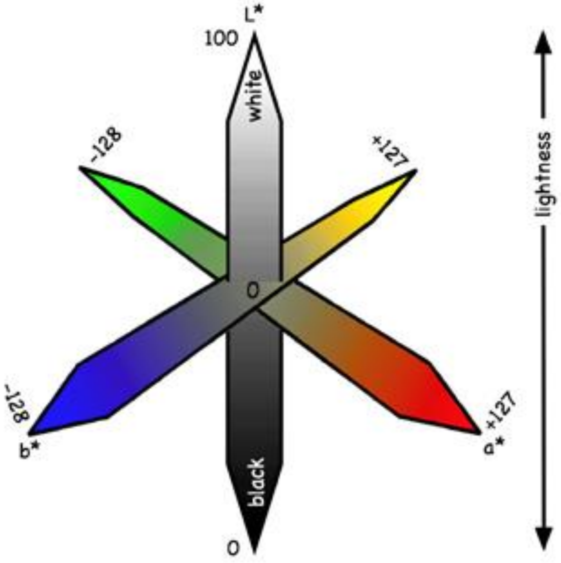
Few Bins
Need less data
Coarser representation

Many Bins
Need more data
Finer representation

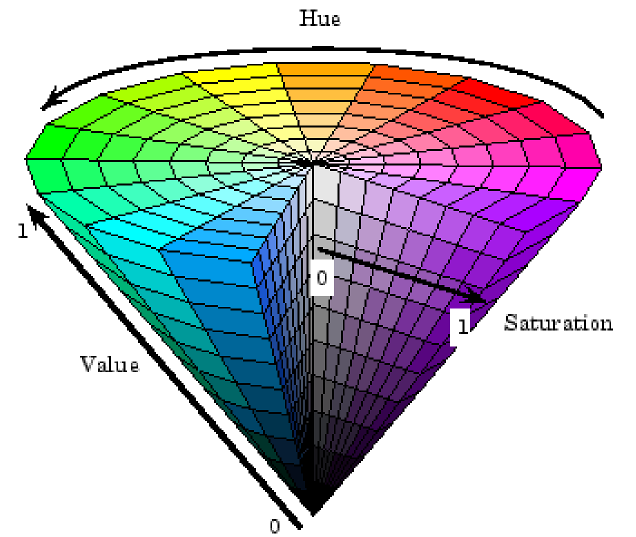
- Matching
  - Histogram intersection or Euclidean may be faster
  - Chi-squared often works better
  - Earth mover’s distance is good for when nearby bins represent similar values

What kind of things do we compute histograms of?

- Color

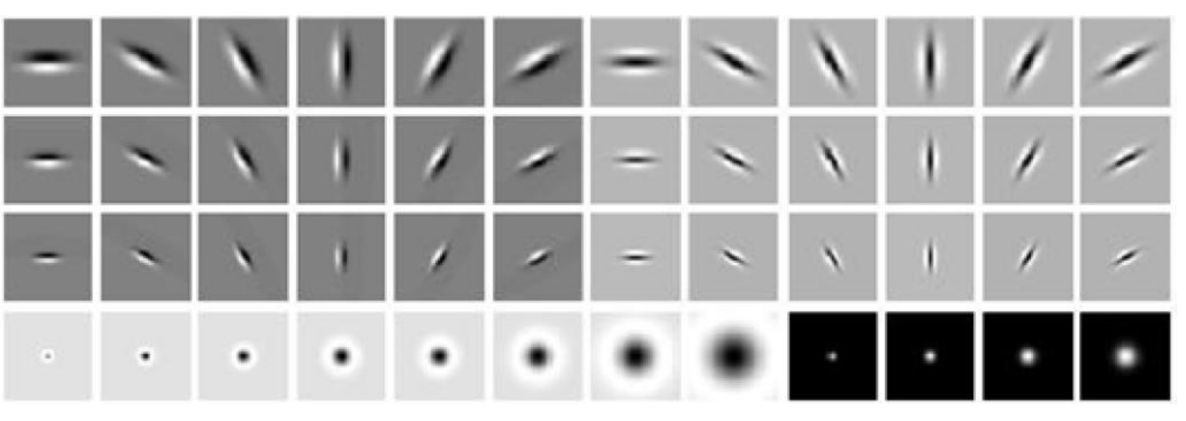


L*a*b* color space



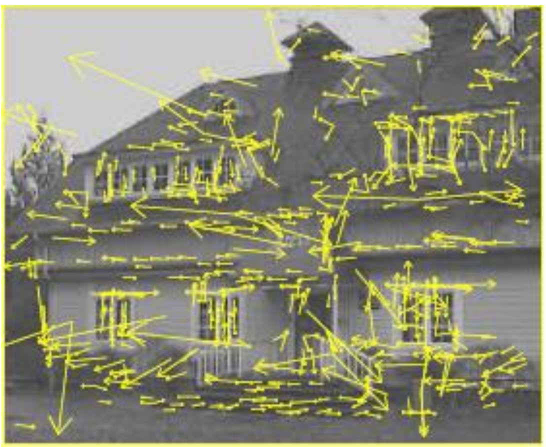
HSV color space

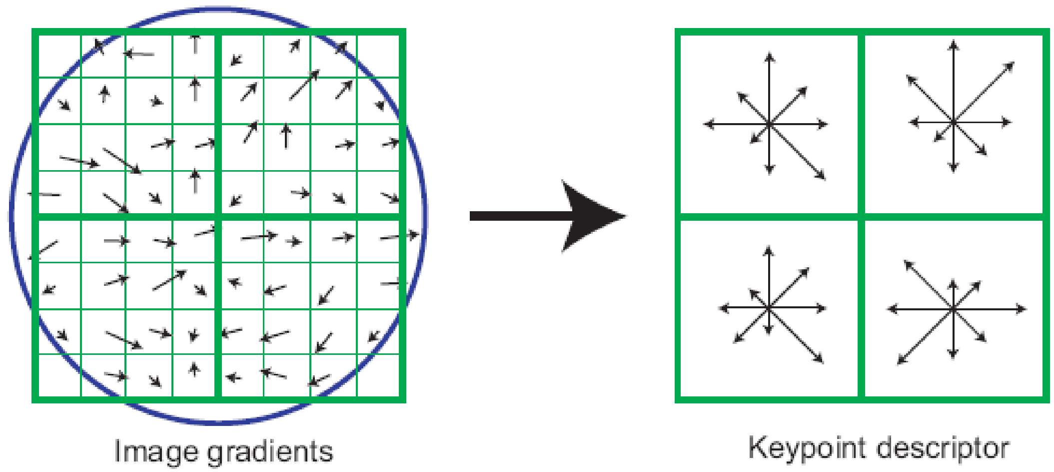
- Texture (filter banks or HOG over regions)



What kind of things do we compute histograms of?

- Histograms of descriptors

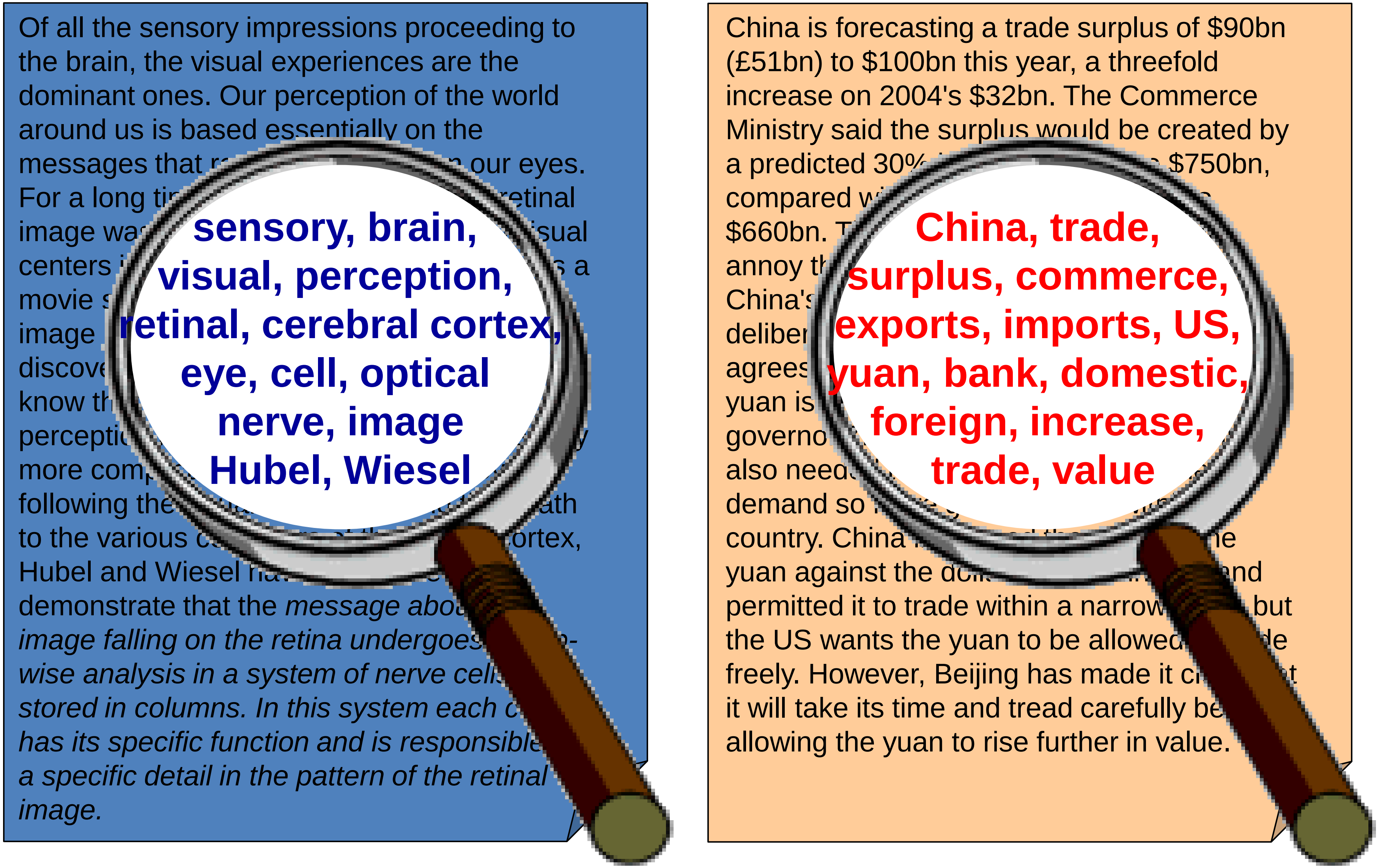




SIFT – [Lowe IJCV 2004]

- “Bag of visual words”

Analogy to documents



Bag of *visual words*

- Image patches
- BoW histogram
- Codewords

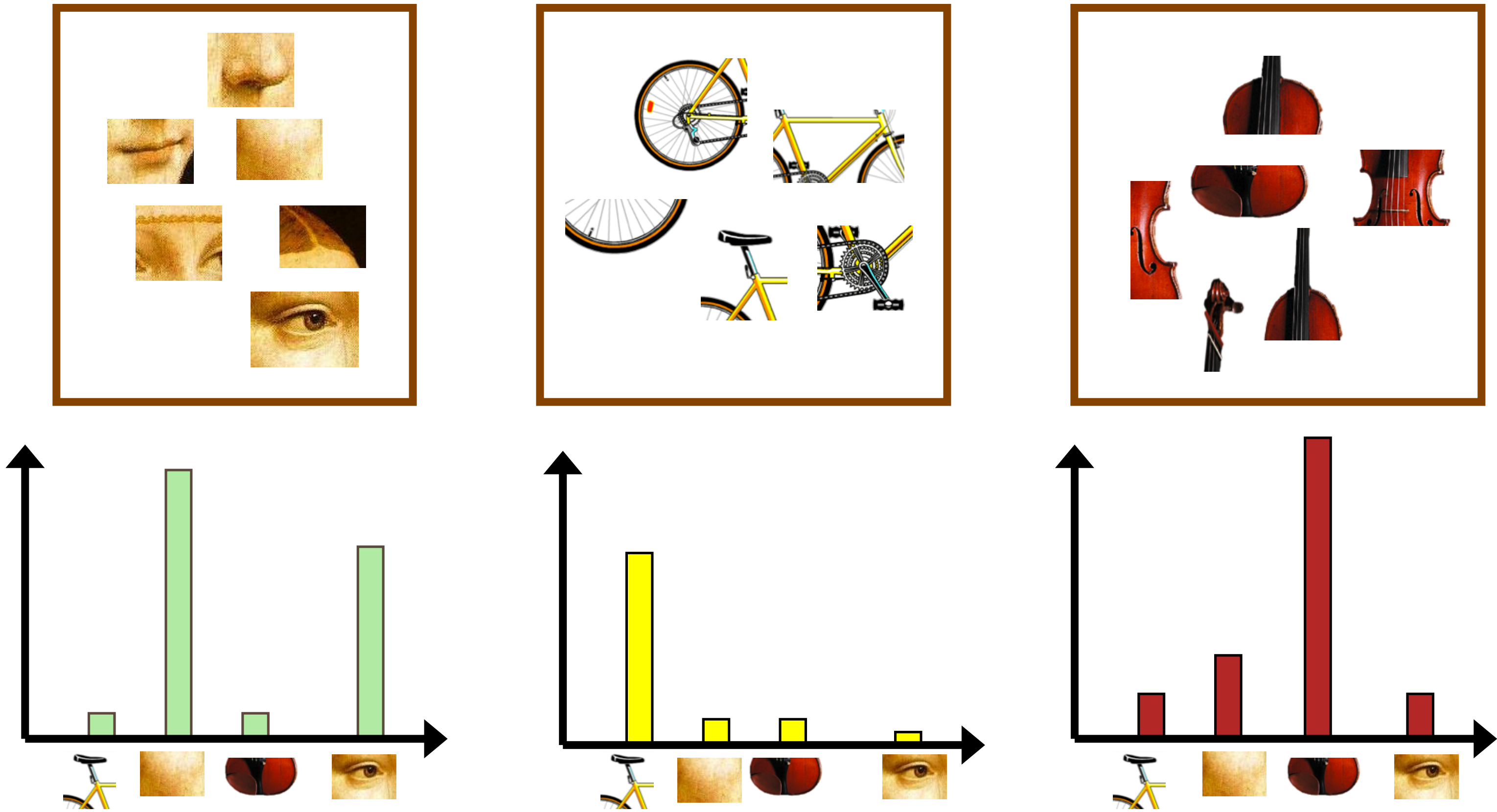


Image categorization with bag of words

Training

1. Extract keypoints and descriptors for all training images
2. Cluster descriptors
3. Quantize descriptors using cluster centers to get "visual words"
4. Represent each image by normalized counts of "visual words"
5. Train classifier on labeled examples using histogram values as features

Testing

1. Extract keypoints/descriptors and quantize into visual words
2. Compute visual word histogram
3. Compute label or confidence using classifier

HW5 - Prob2 scene categorization

- Training and testing images for 8 categories

- Implement repres
- BoW model: preprocessed descriptors
  - Learning dictionary using K-Means
  - Learning classifier (NN, SVM)
- Report results

Take a break…

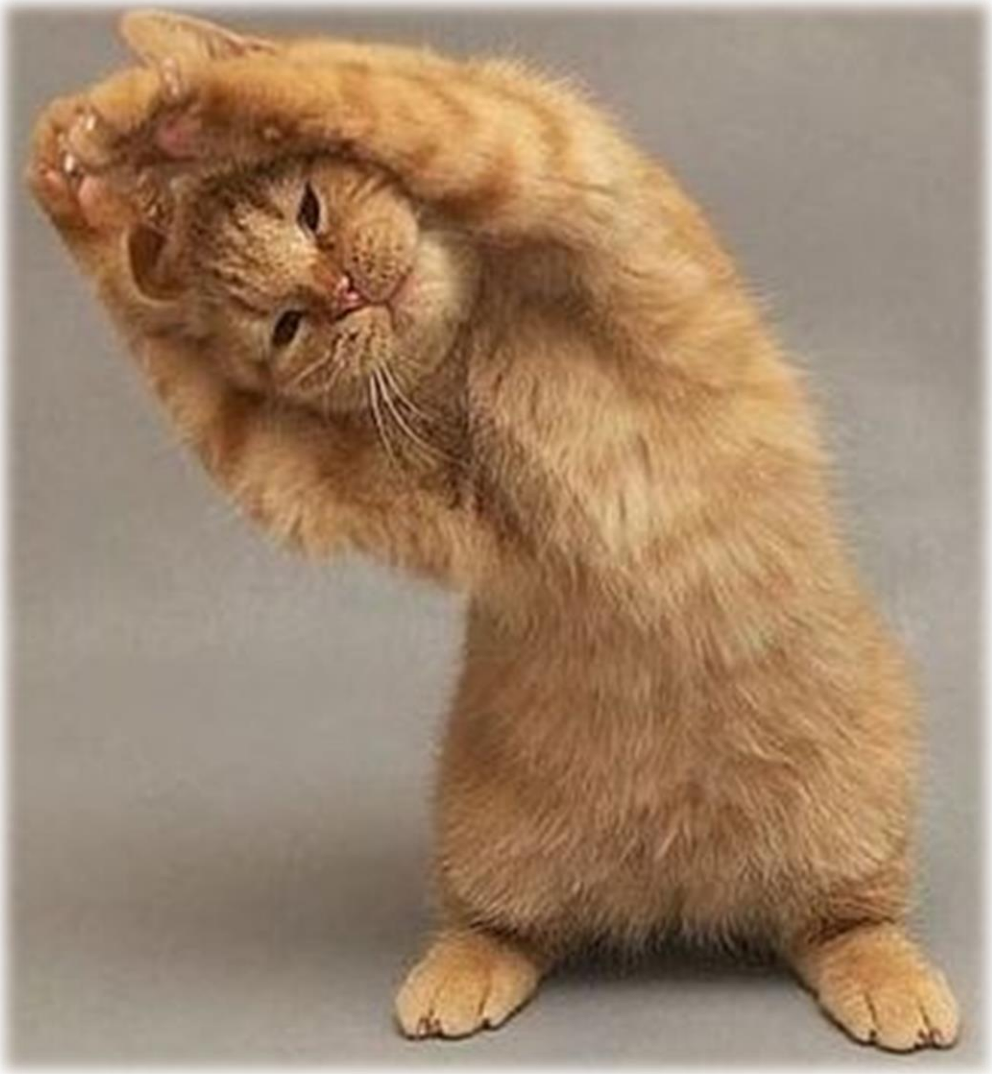
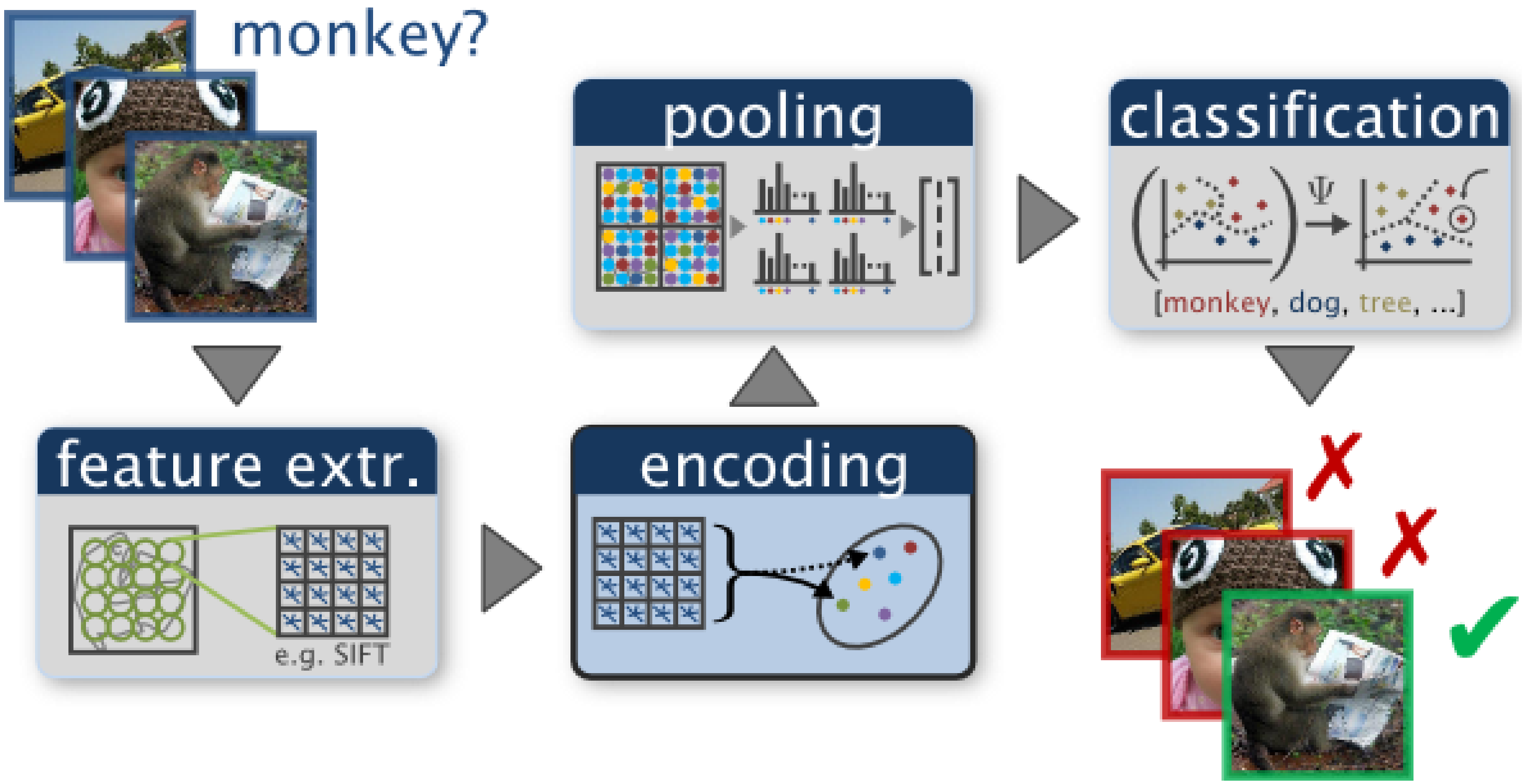


Image source: http://mehimandthecats.com/feline-care-guide/

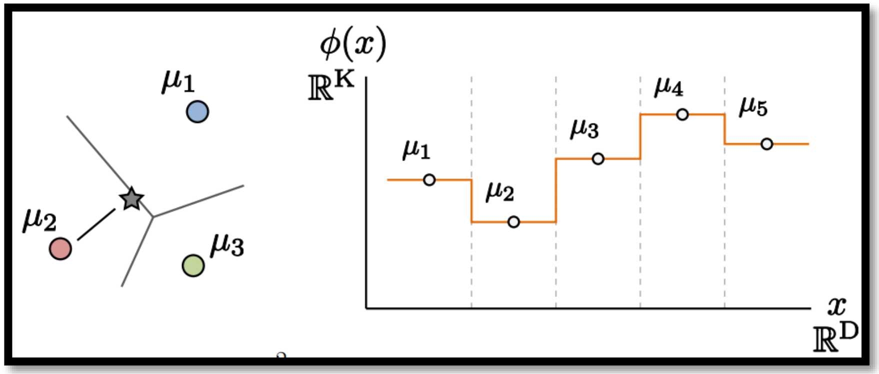
Bag of visual words image classification



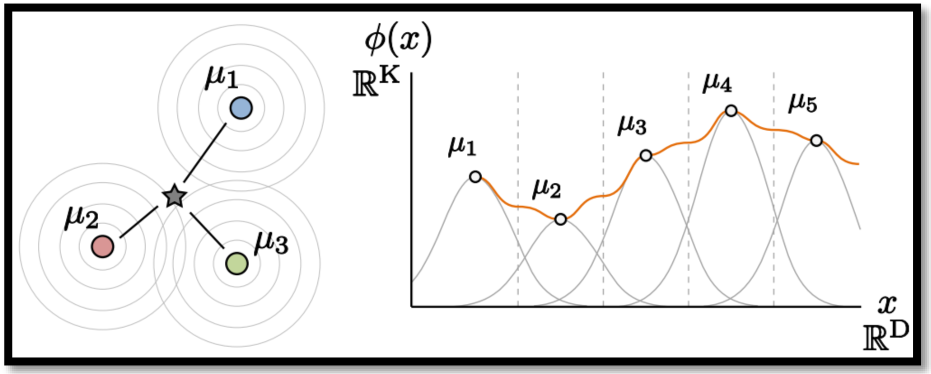
[Chatfieldet al. BMVC 2011]

Feature encoding

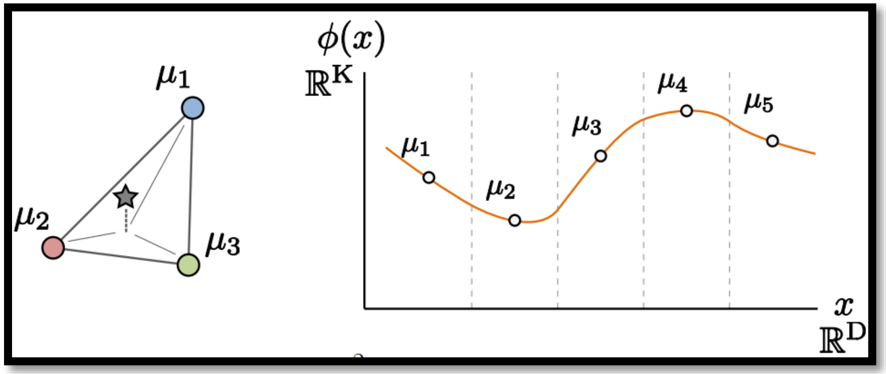
- Hard/soft assignment to clusters



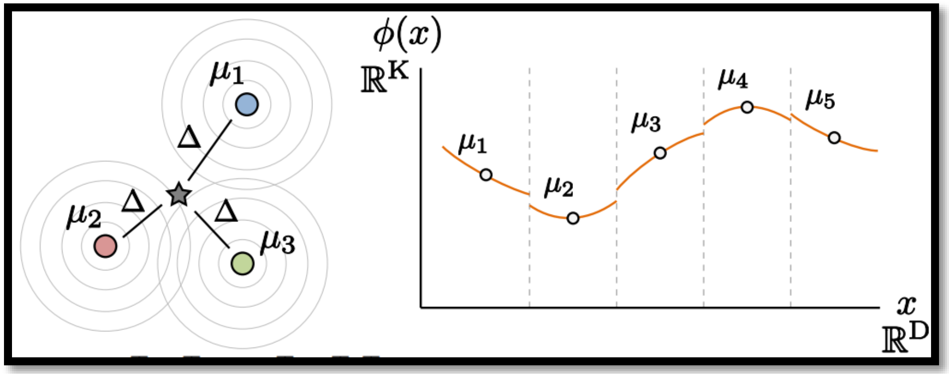
Histogram encoding



Kernel codebook encoding



Locality constrained encoding



Fisher encoding

[Chatfieldet al. BMVC 2011]

Fisher vector encoding

- Fit Gaussian Mixture Models

$$\Theta = (\mu_k, \Sigma_k, \pi_k : k = 1, \dots, K)$$

- Posterior probability

$$q_{ik} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_k)\right]}{\sum_{t=1}^{K} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_t)^T \Sigma_k^{-1} (\mathbf{x}_i - \mu_t)\right]}$$

- First and second order differences to cluster k

$$u_{jk} = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}},$$

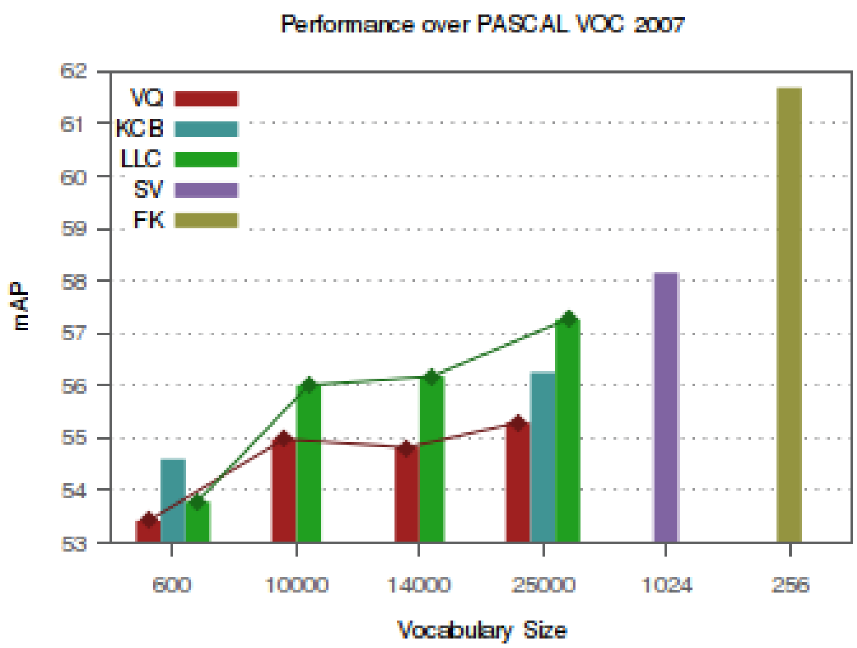
$$v_{jk} = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik} \left[\left(\frac{x_{ji} - \mu_{jk}}{\sigma_{jk}}\right)^2 - 1\right]$$

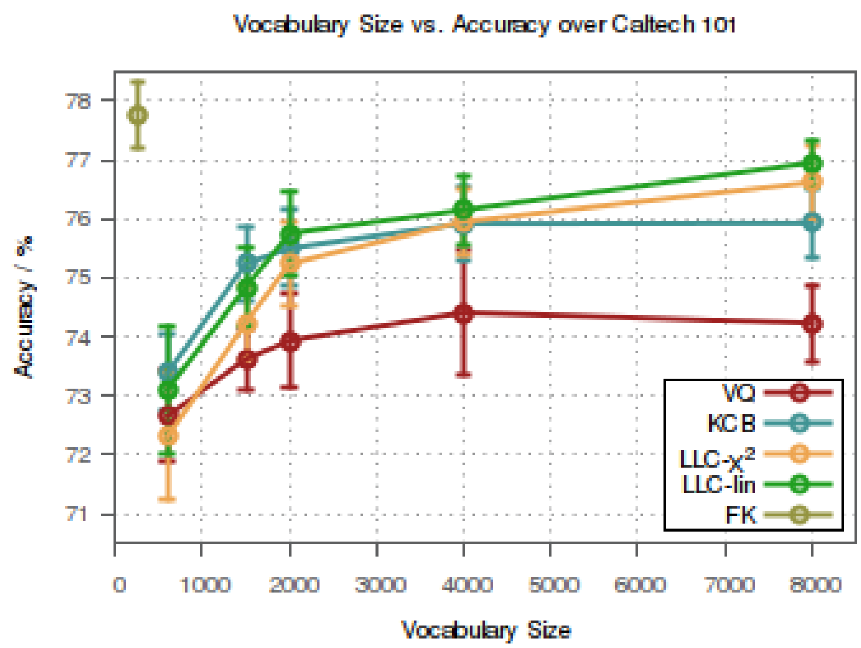
$$\Phi(I) = \begin{bmatrix} \vdots \\ \mathbf{u}_k \\ \vdots \\ \mathbf{v}_k \\ \vdots \end{bmatrix}$$

[Perronnin et al. ECCV 2010]

Performance comparisons

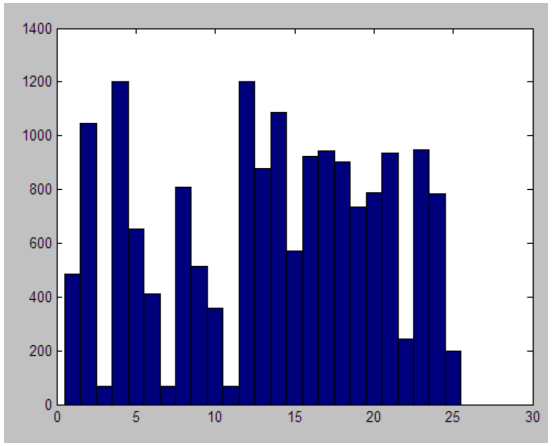
- Fisher vector encoding outperforms others
- Higher-order statistics helps

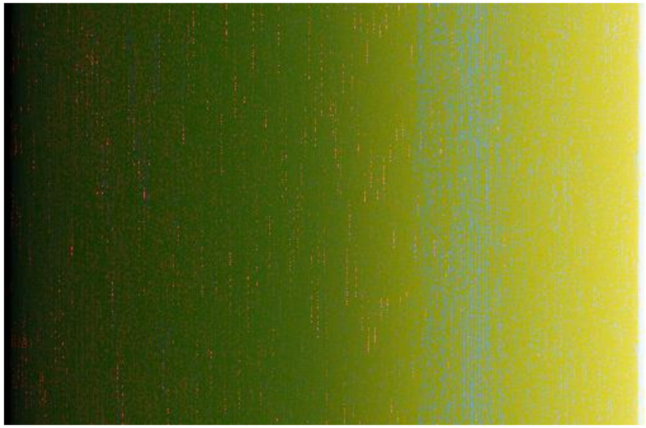




[Chatfieldet al. BMVC 2011]

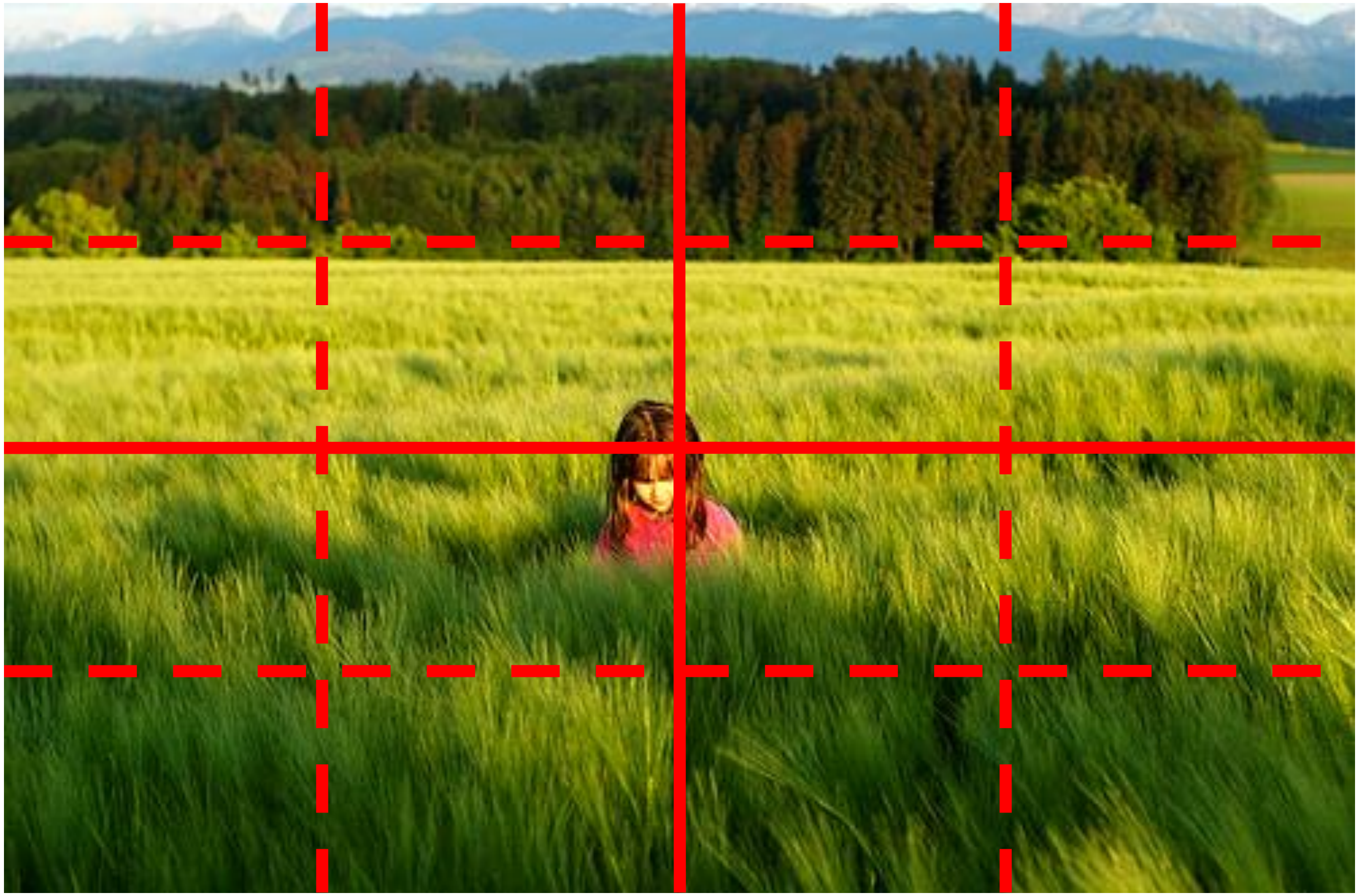
But what about spatial layout?





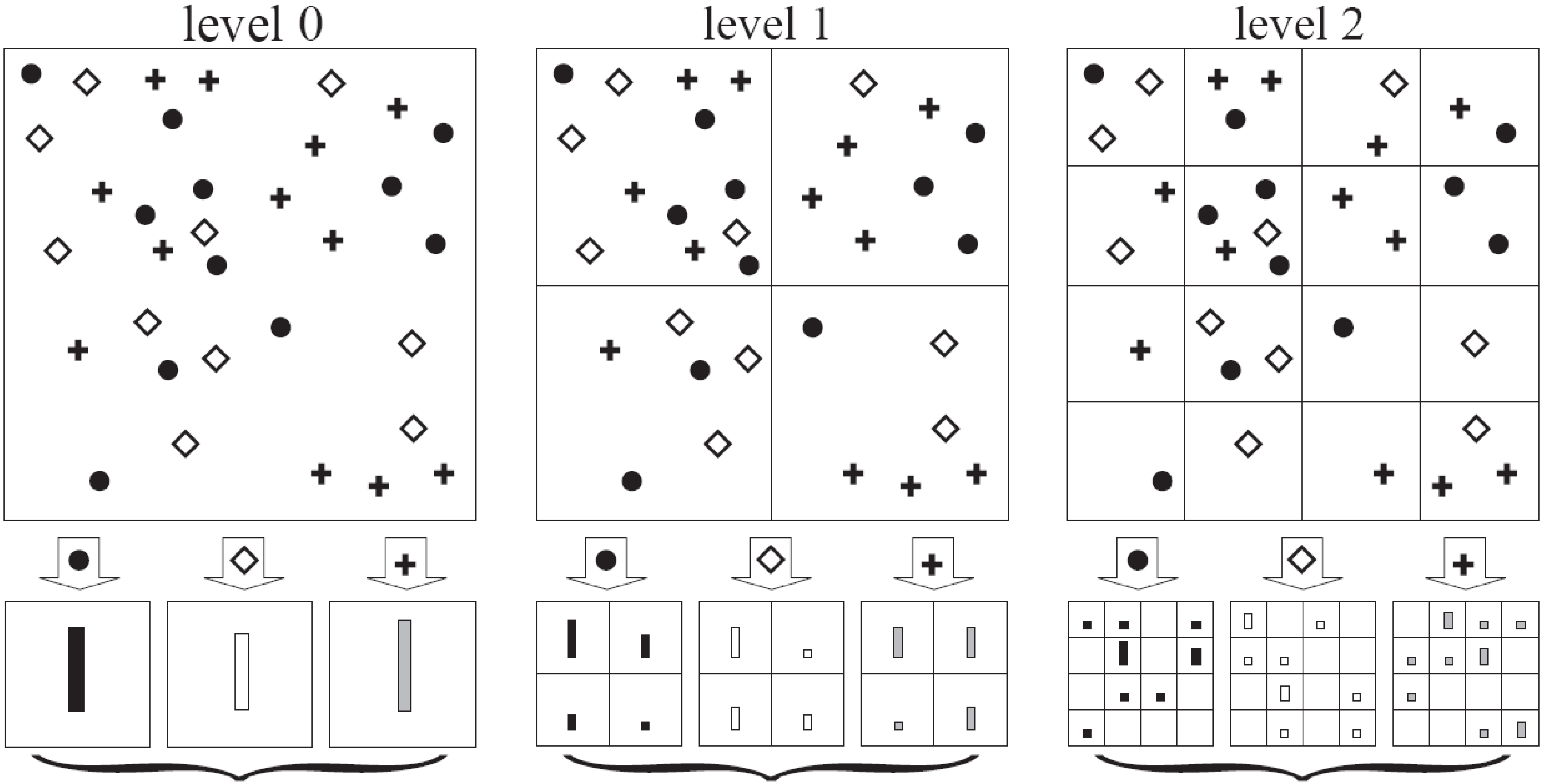
All of these images have the same color histogram

Spatial pyramid



Compute histogram in each spatial bin

Spatial pyramid

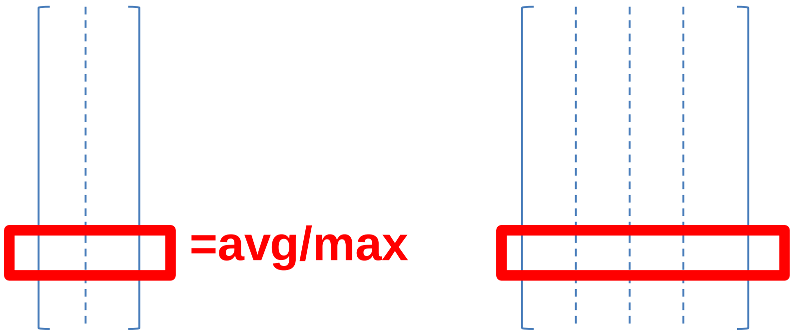


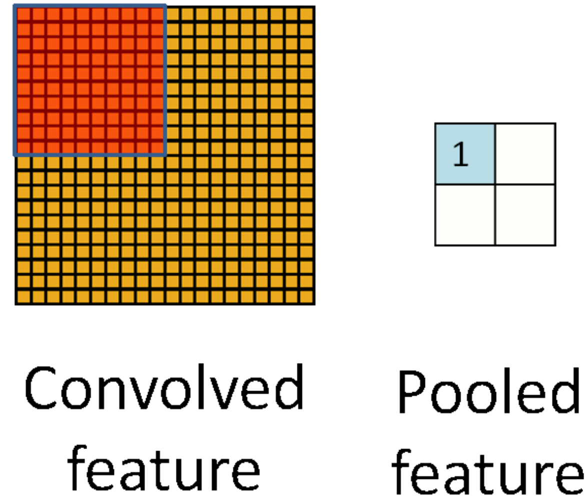
High number of features – PCA to reduce dimensionality

[Lazebnik et al. CVPR 2006]

Pooling

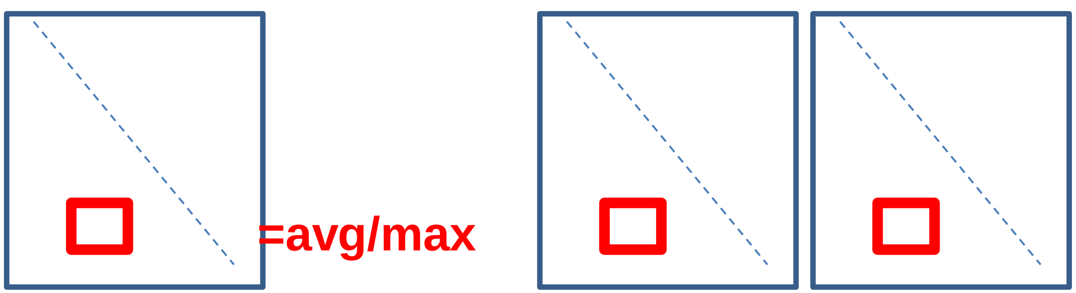
- Average/max pooling





Source: Unsupervised Feature Learning and Deep Learning

- Second-order pooling [Joao et al. PAMI 2014]

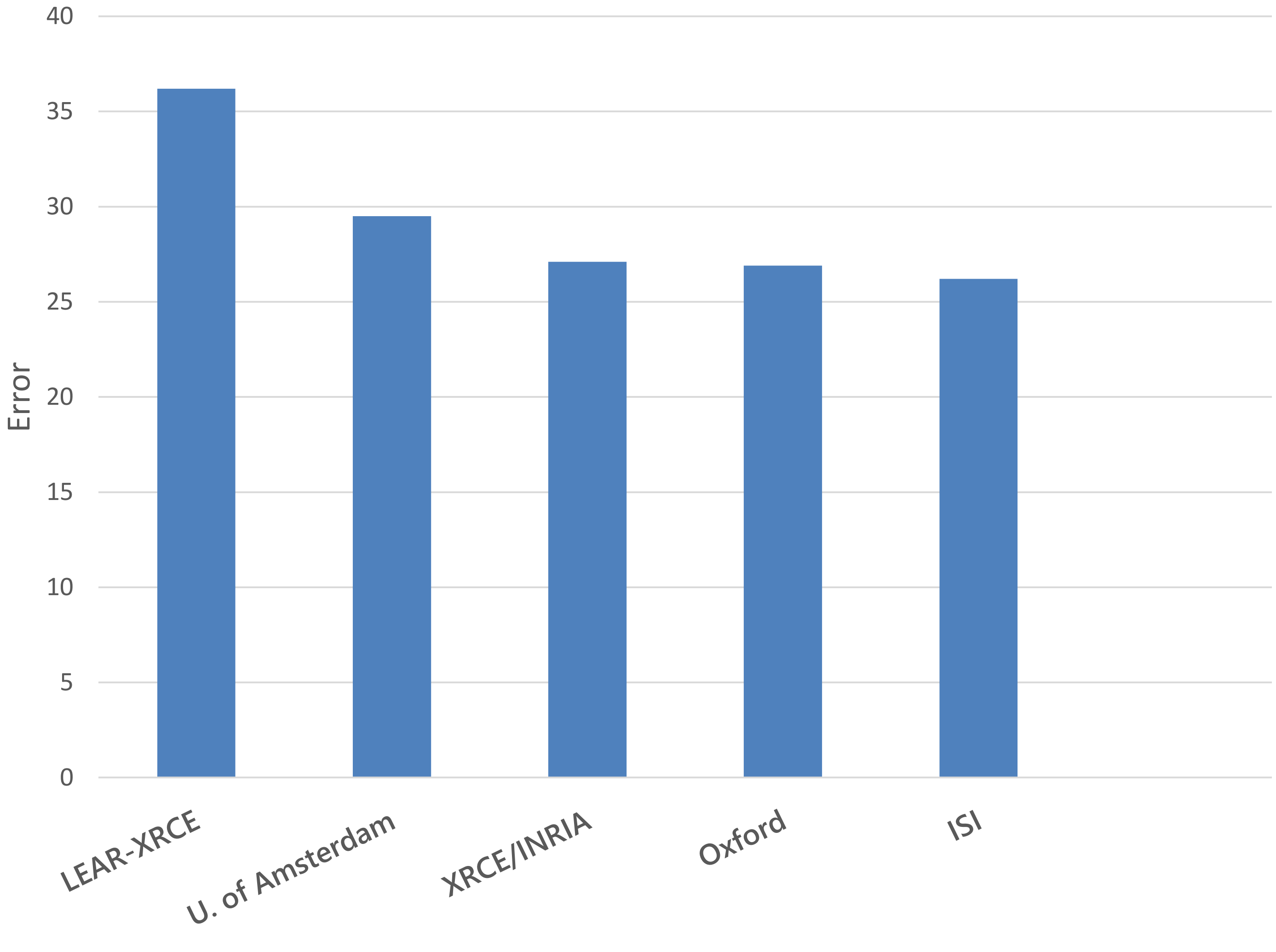


$$\mathbf{G}_{avg}(R_j) = \frac{1}{|F_{R_j}|} \sum_{i:(\mathbf{f}_i \in R_j)} \mathbf{x}_i \cdot \mathbf{x}_i^\top$$

$$\mathbf{G}_{max}(R_j) = \max_{i:(\mathbf{f}_i \in R_j)} \mathbf{x}_i \cdot \mathbf{x}_i^\top$$

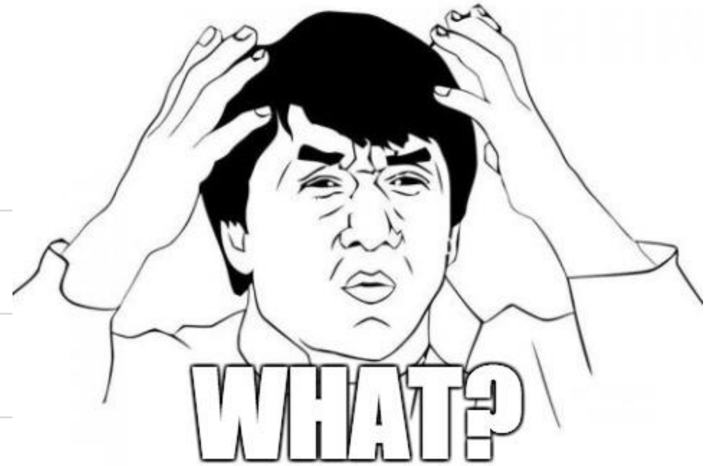
2012 ImageNet 1K

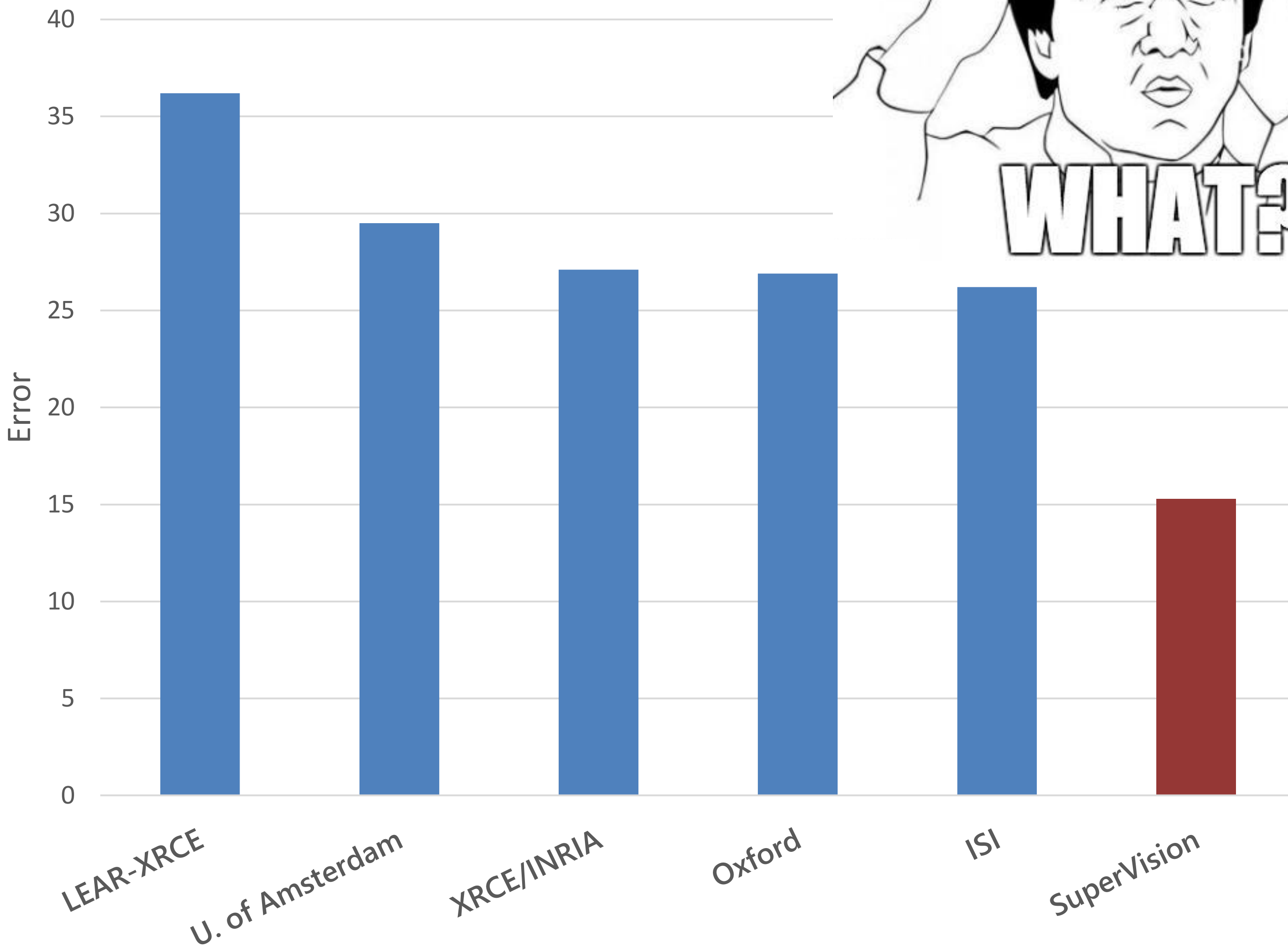
(Fall 2012)



2012 ImageNet 1K

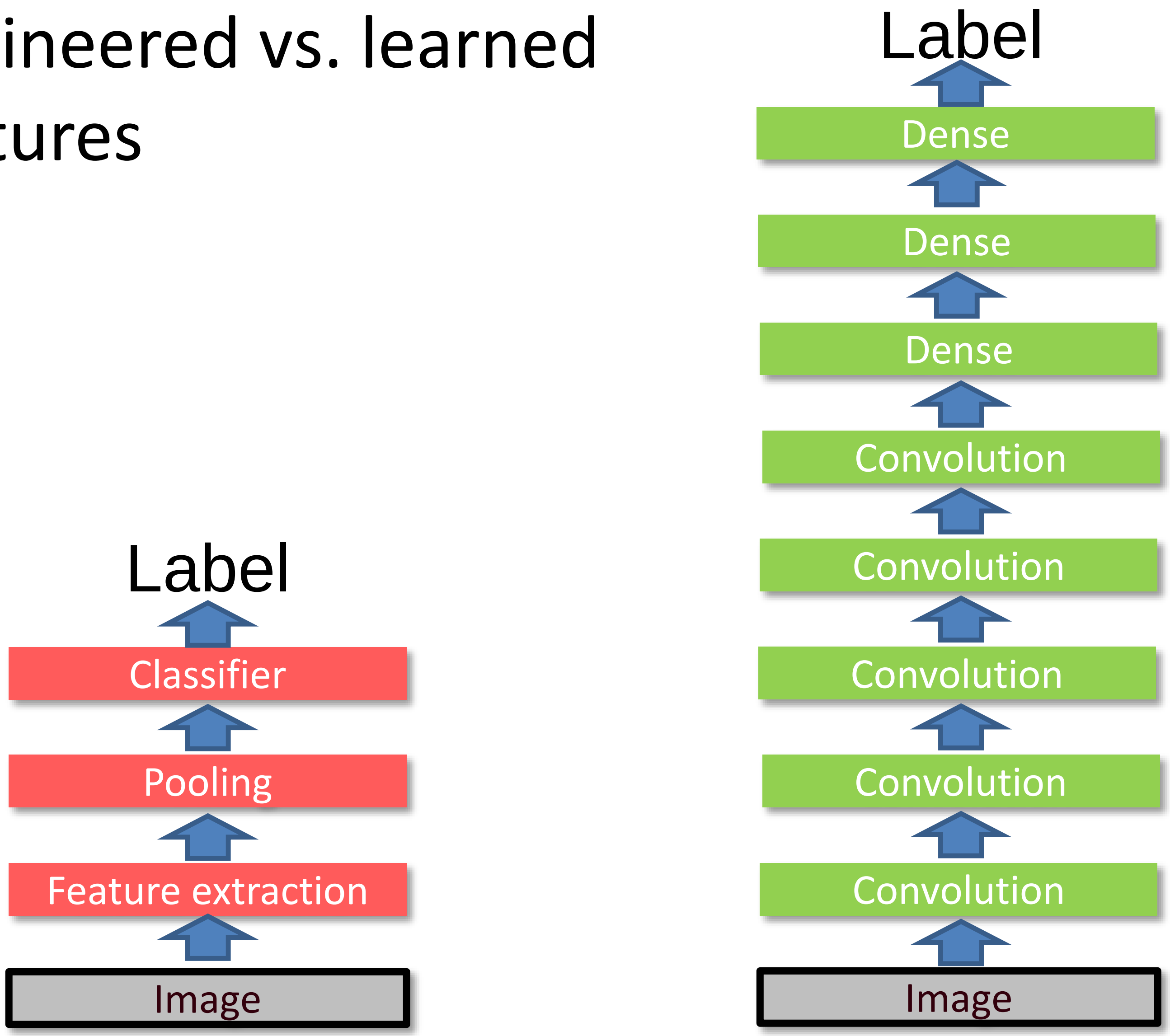
(Fall 2012)

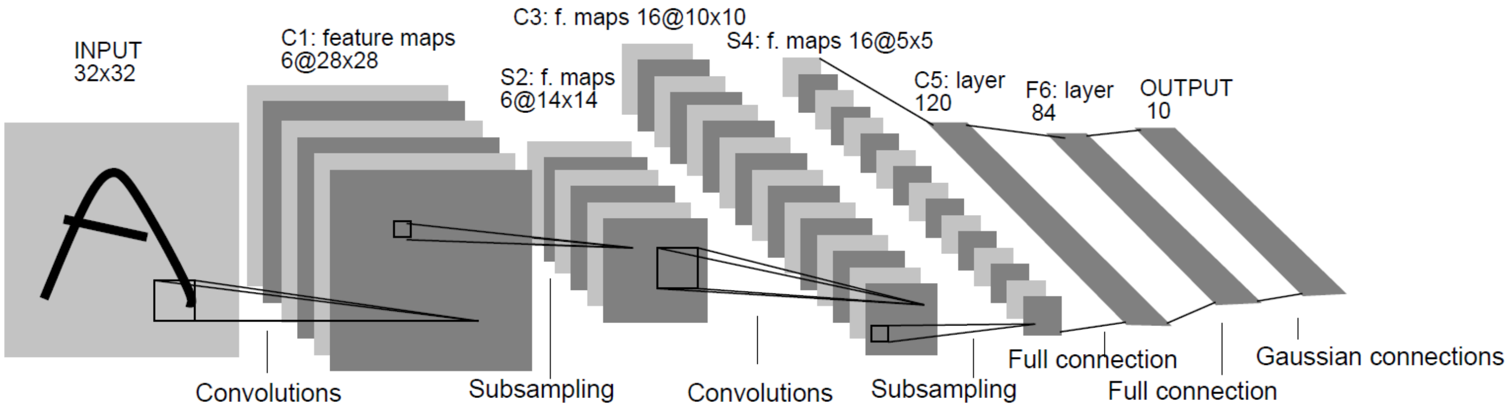




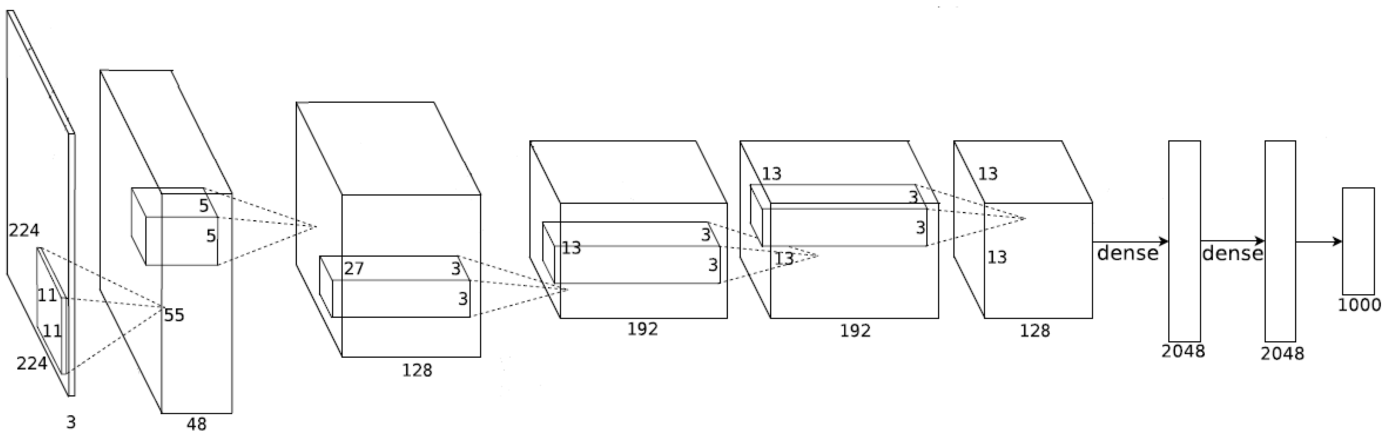
Shallow vs. deep learning

- Engineered vs. learned features



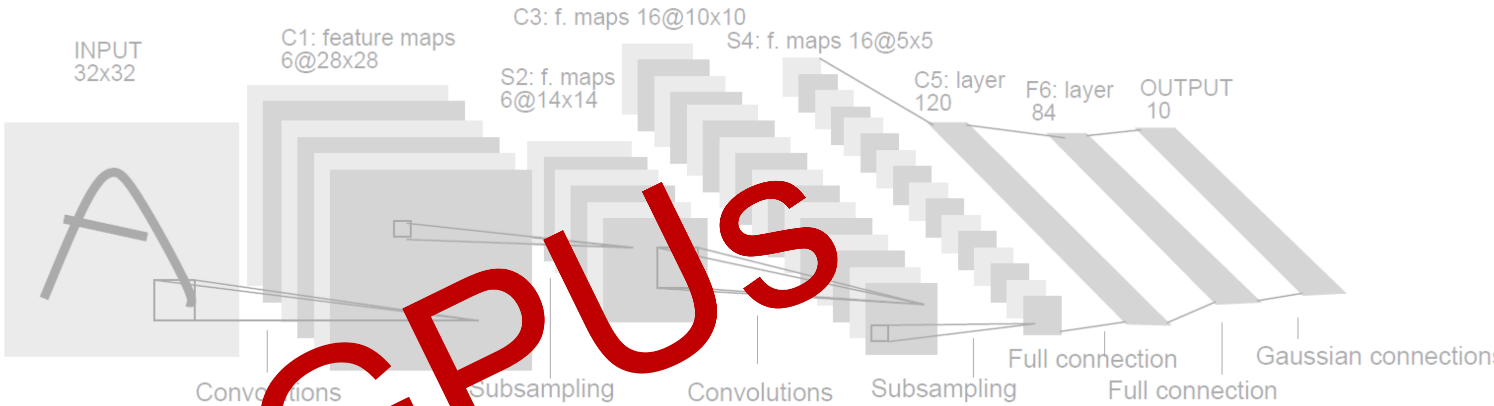


Gradient-Based Learning Applied to Document Recognition, LeCun, Bottou, Bengio and Haffner, Proc. of the IEEE, **1998**



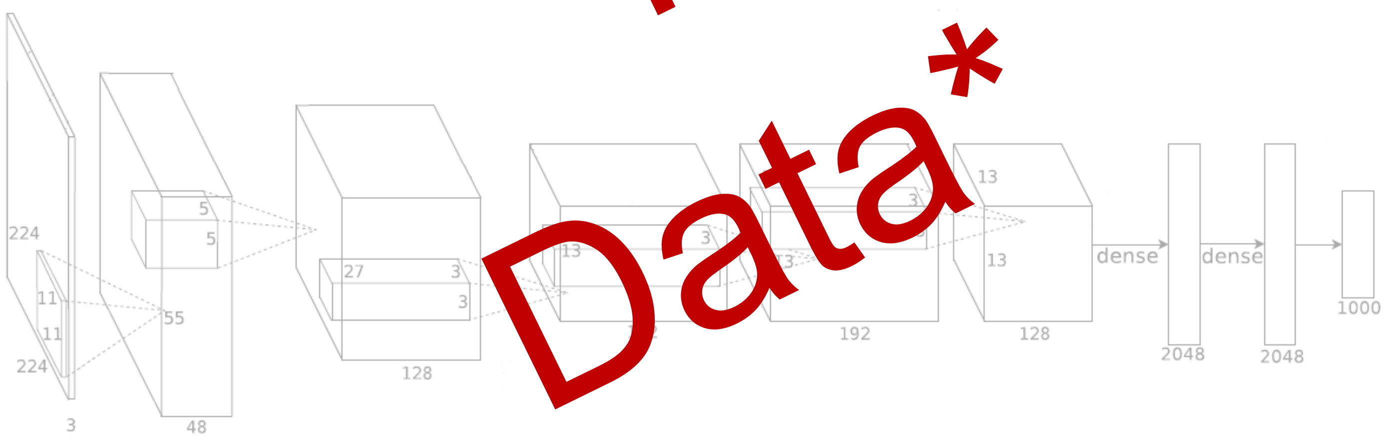
Imagenet Classification with Deep Convolutional Neural Networks, Krizhevsky, Sutskever, and Hinton, NIPS **2012**

Slide Credit: L. Zitnick



Gradient-Based Learning Applied to Document Recognition, LeCun, Bottou, Bengio and Haffner, Proc. of the IEEE, **1998**

GPUs + Data*

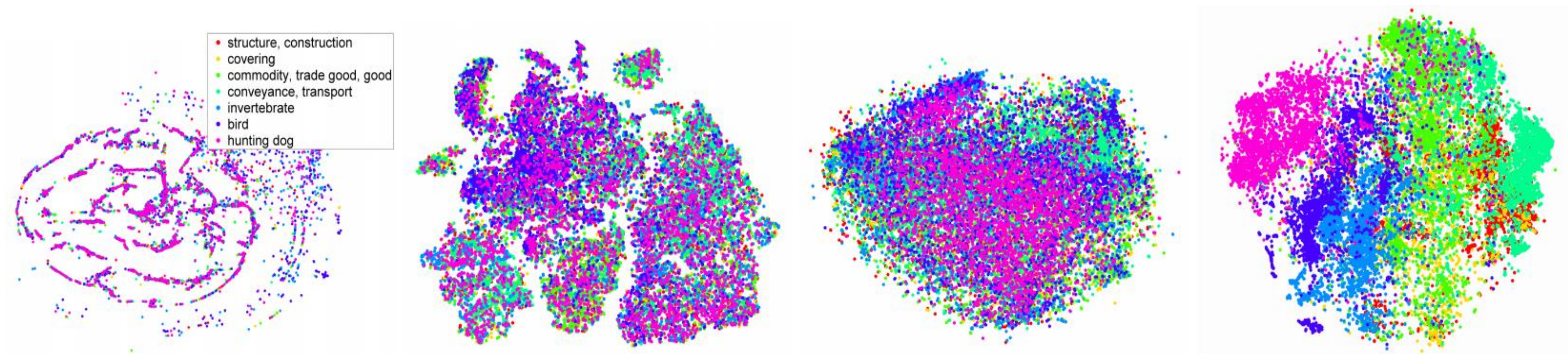


Imagenet Clas
Networks, Kriz

* Rectified activations and dropout

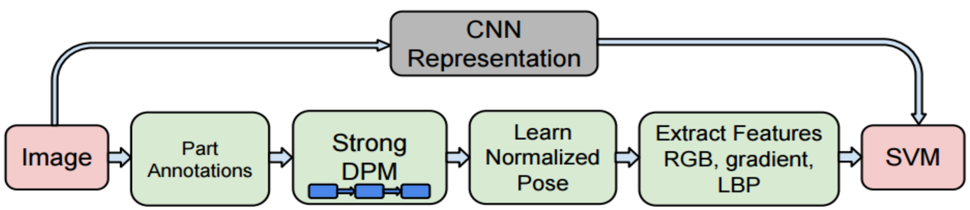
Slide Credit: L. Zitnick

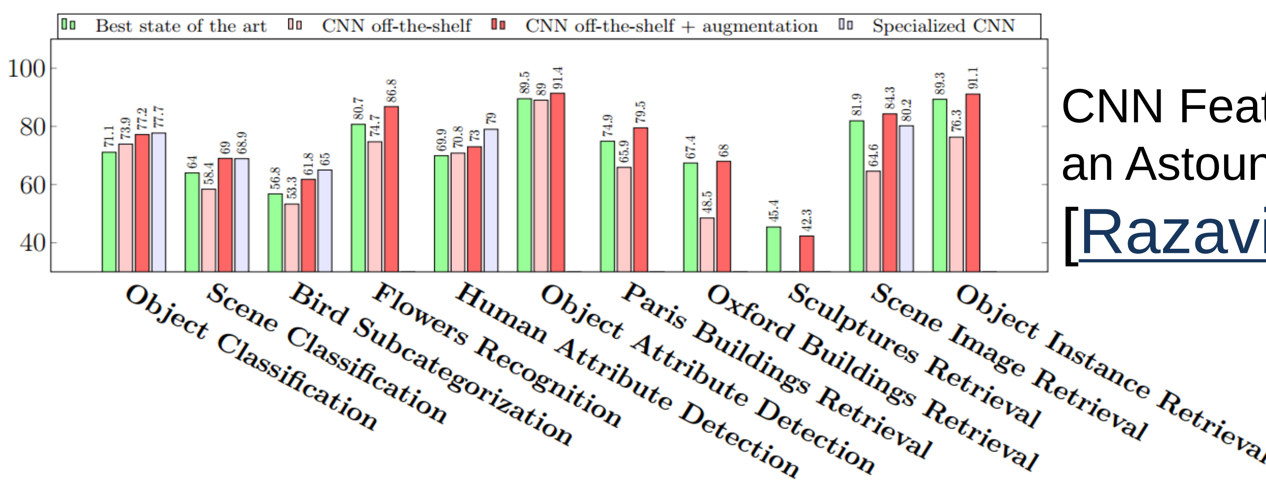
Convolutional activation features



(a) LLC (b) GIST (c) $DeCAF_1$ (d) $DeCAF_6$

[Donahue et al. ICML 2013]

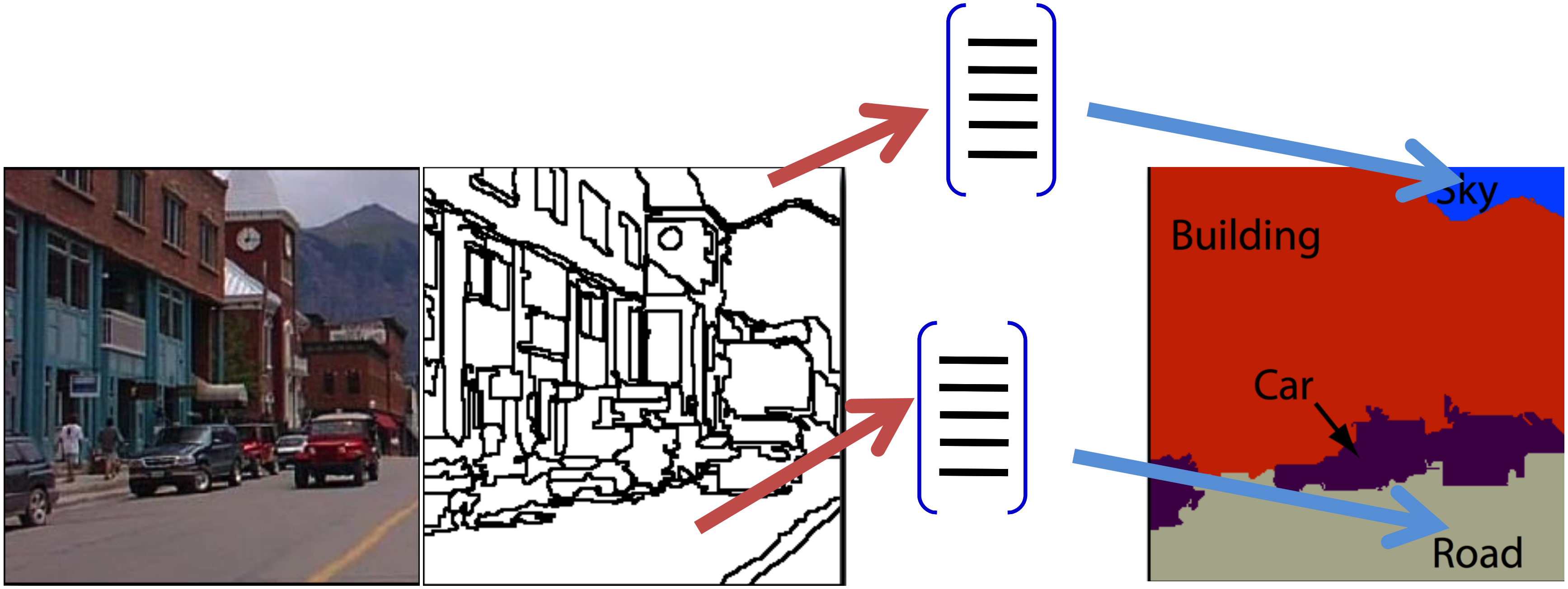




CNN Features off-the-shelf:
an Astounding Baseline for Recognition
[Razavian et al. 2014]

Region representation

- Segment the image into superpixels
- Use features to represent each image segment



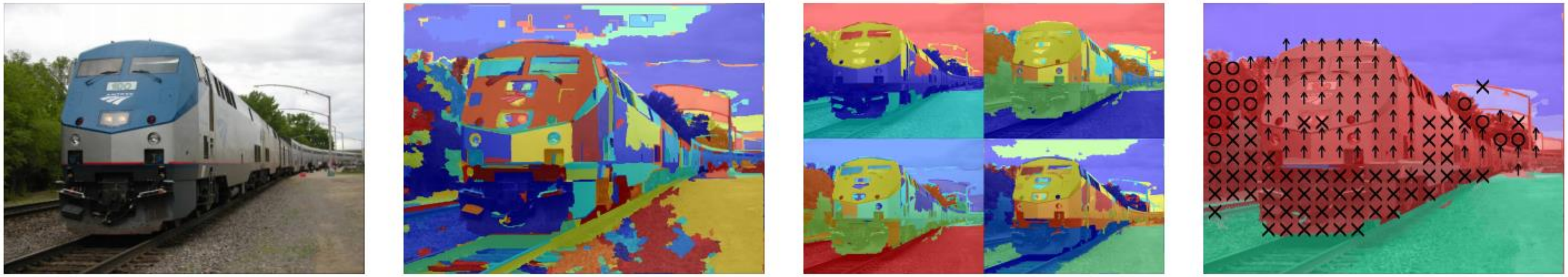
Joseph Tighe and Svetlana Lazebnik

Region representation

- Color, texture, BoW
  - Only computed within the local region
- Shape of regions
- Position in the image

Working with regions

- Spatial support is important – multiple segmentation



(a) Input (b) Superpixels (c) Multiple Hypotheses (d) Geometric Labels

Geometric context [Hoiem et al. ICCV 2005]

- Spatial consistency – MRF smoothing

Beyond categorization

- Exemplar models [Malisiewicz and Efros NIPS09, ICCV11]
  - Ask not "what is this?", ask "what is this like" – Moshe Bar
- A train?

Things to remember

- Visual categorization help transfer knowledge
- Image features
  - Coverage, concision, directness
  - Color, gradients, textures, motion, descriptors
  - Histogram, feature encoding, and pooling
  - CNN as features
- Image/region categorization

Next lecture - Classifiers