CS688: Web-Scale Image Retrieval

Intro to Object Recognition

Sung-Eui Yoon
(윤성의)

Course URL:
http://sglab.kaist.ac.kr/~sungeui/IR

KAIST

Class Objectives

- **Introduction to object detection**
  - **Representation (features)**
  - **Learning**
  - **Recognition**

- **Recently performed within deep neural net with an end-to-end optimization**

KAIST

What are the different visual recognition tasks?

Classification:

Does this image contain a building? [yes/no]

Classification:

Is this an beach?

Image Search

flickr

webshots

Ask Images

Cydral

picsearch

Google Image Search

altavista

Organizing photo collections

Detection:

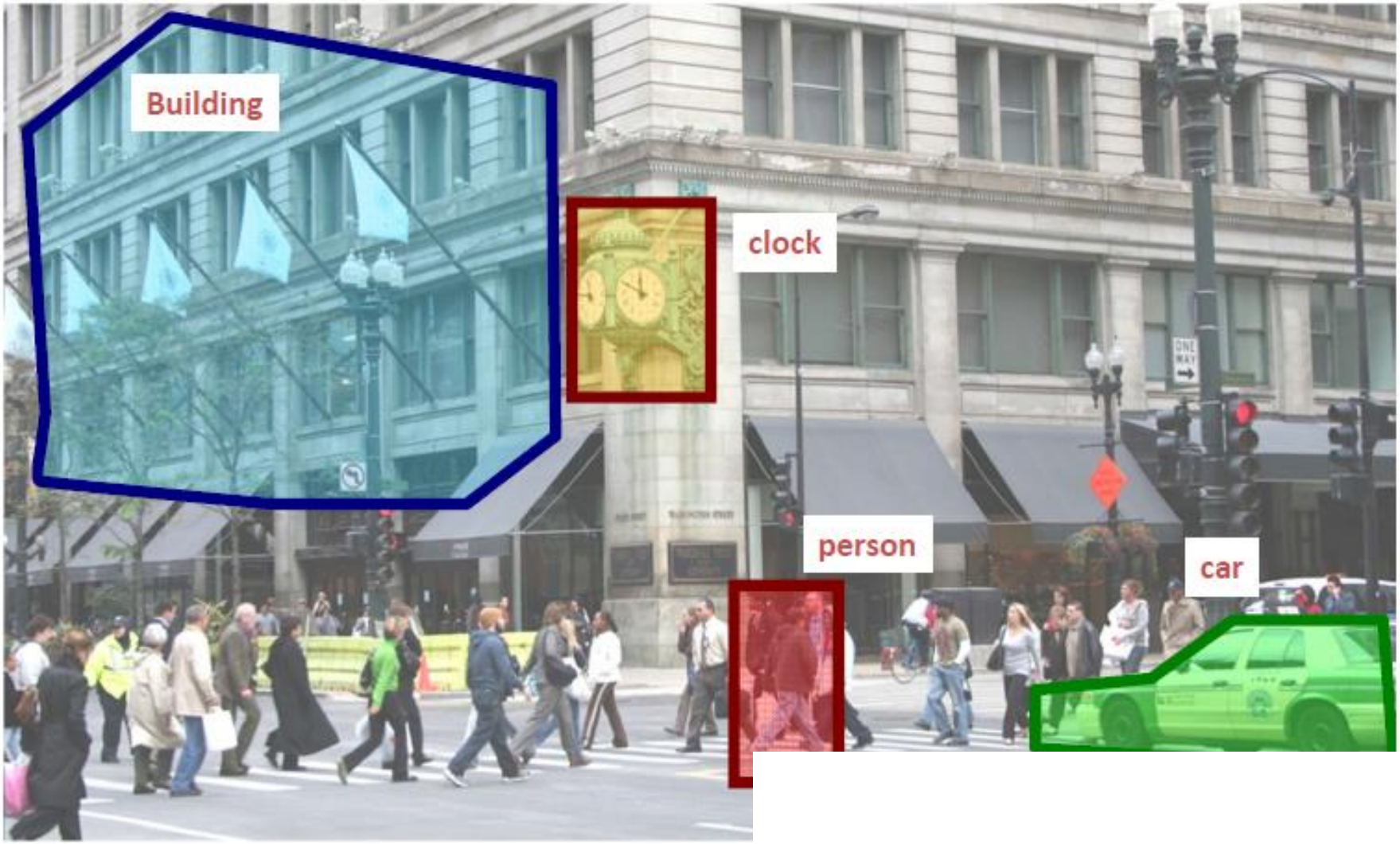
Does this image contain a car? [where?]

Detection:

Does this image contain a car? [where?]

Detection:

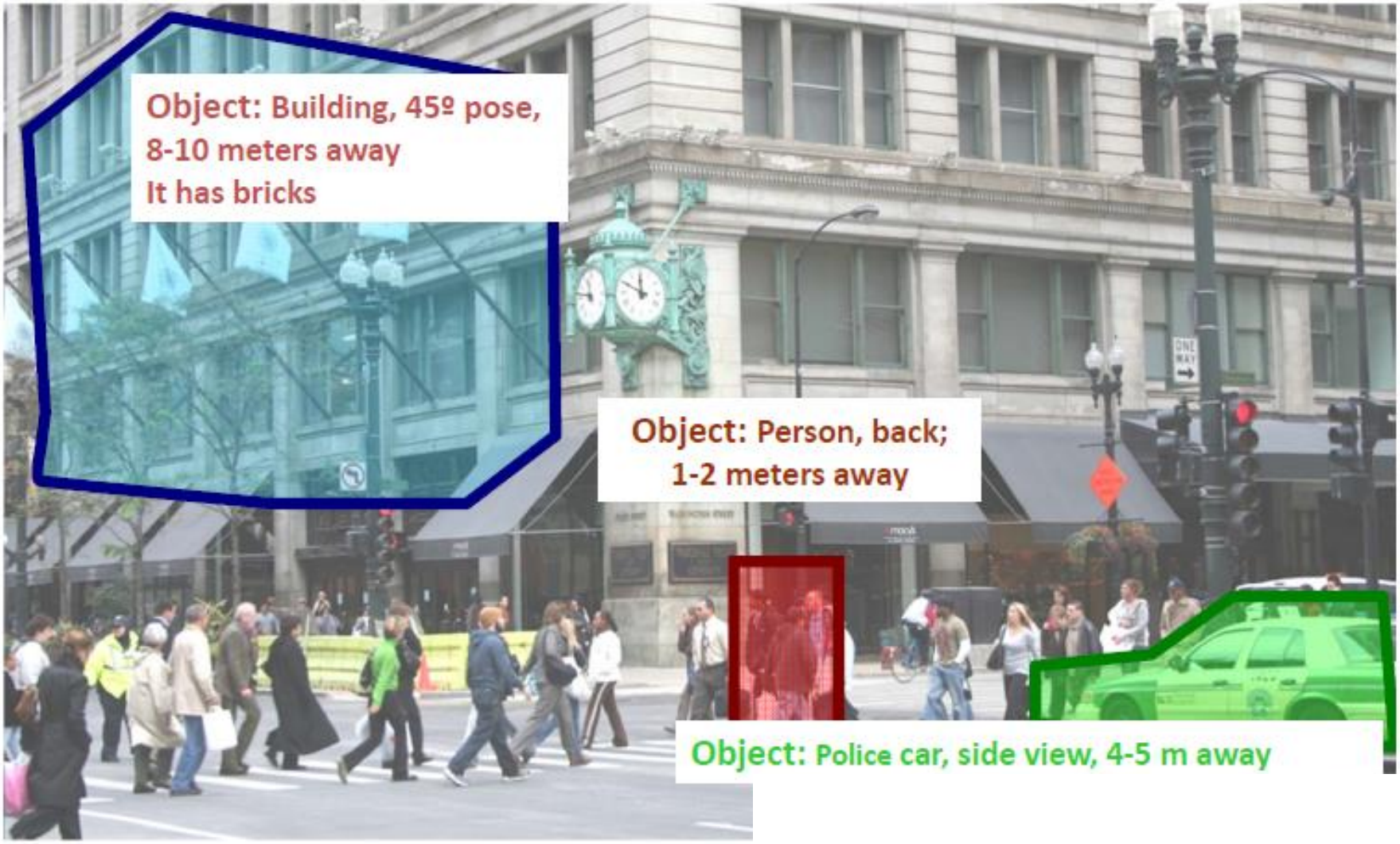
Which object does this image contain? [where?]



Detection:

Accurate localization (segmentation)

Detection: Estimating object semantic & geometric attributes



Applications of Object Recognitions and Image Retrieval

Computational photography

Assistive technologies

Surveillance

Security

Assistive driving

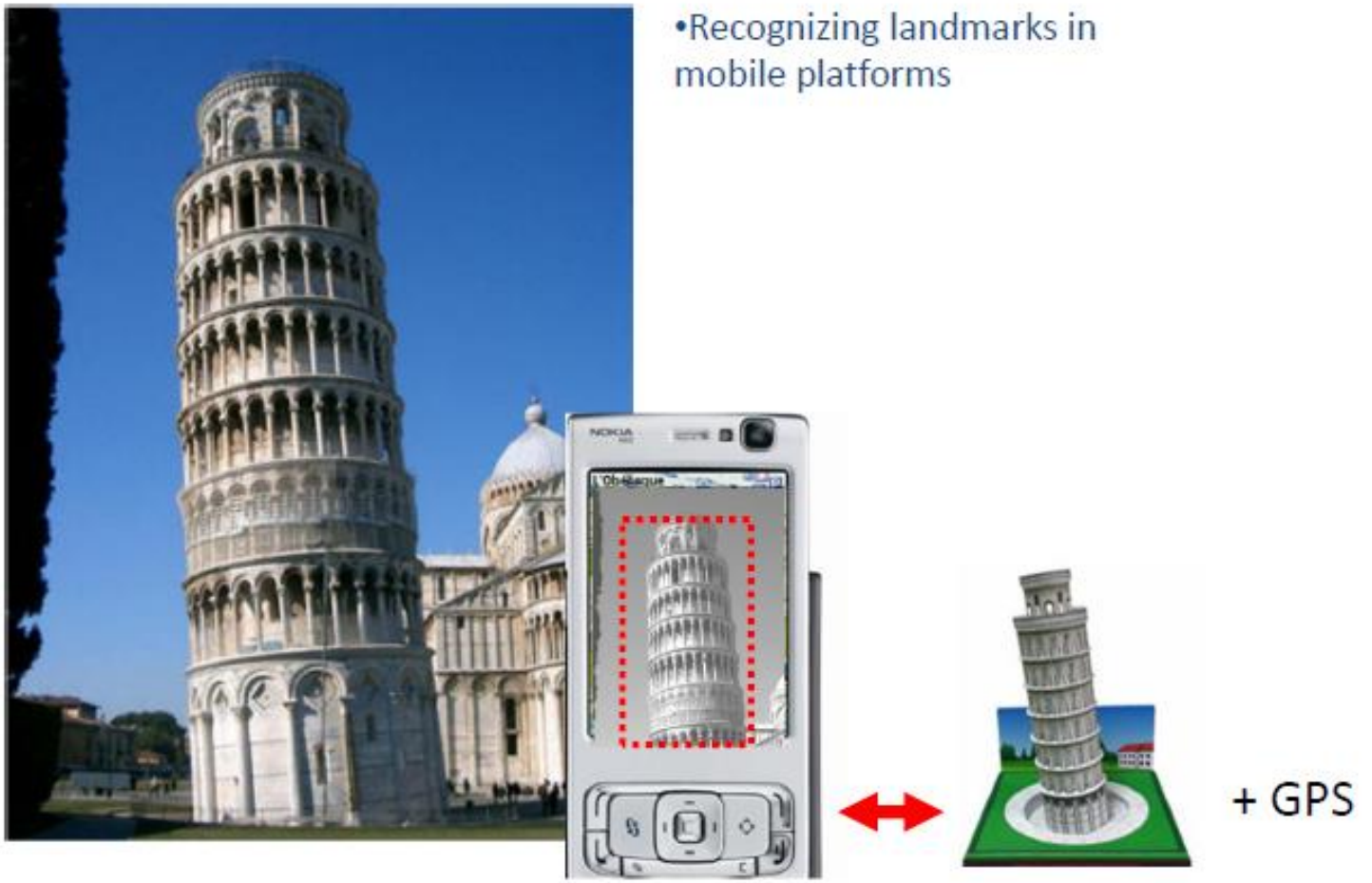
Categorization vs Single instance recognition

Does this image contain the Chicago Macy building's?

Categorization vs Single instance recognition

Where is the crunchy nut?

Applications of Object Recognitions and Image Retrieval



Fei-Fei Li

Activity or Event recognition

What are these people doing?

Visual Recognition

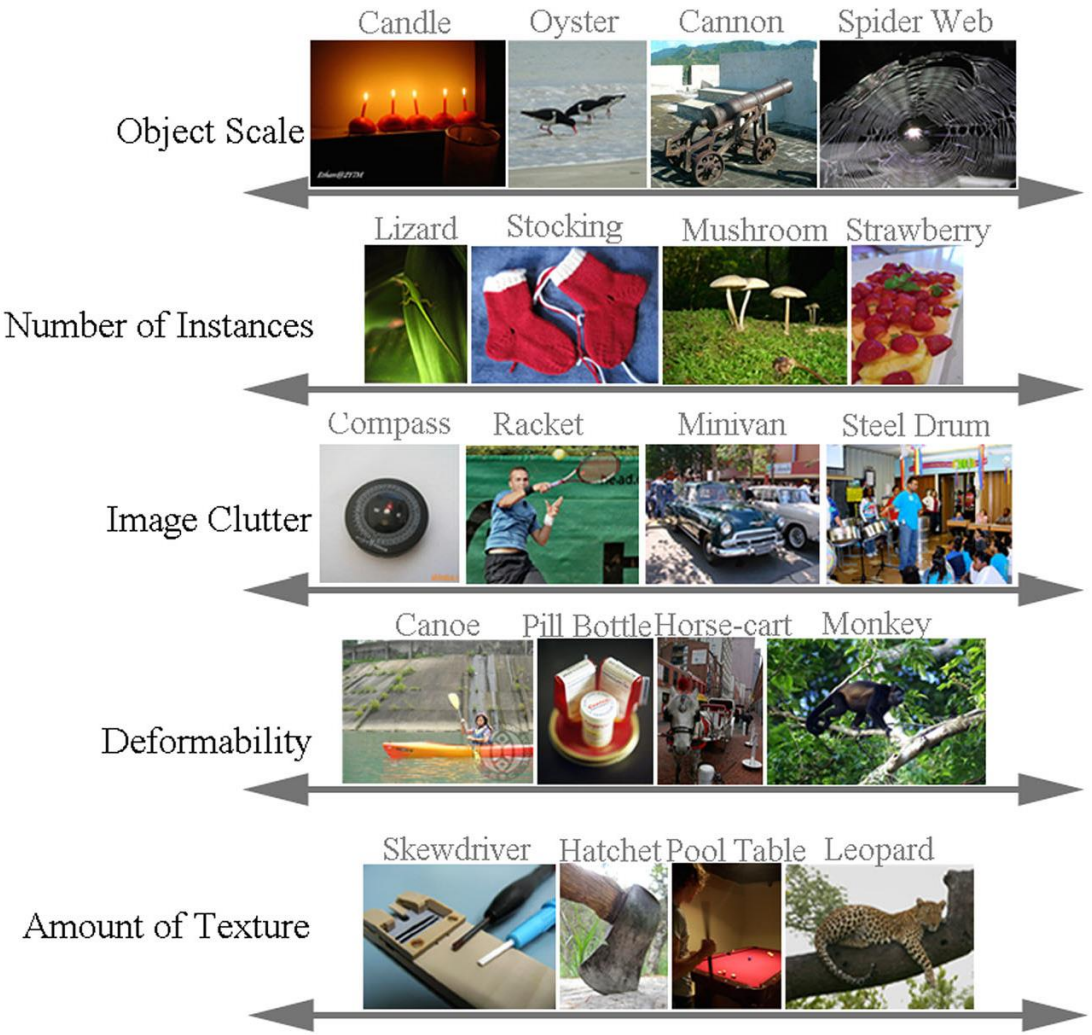
- Design algorithms that are capable to
  - Classify images or videos
  - Detect and localize objects
  - Estimate semantic and geometrical attributes
  - Classify human activities and events

Why is this challenging?


How many object categories are there?
~10,000 to 30,000

ImageNet Large Scale Visual Recognition Challenge [IJCV 15]

- **Contains 1k classes and about 1M images**
- **Annotations**
  - **Image-level: its class**
  - **Object-level: bounding box w/ label**



WordNet and ImageNet [CVPR 09]

- **ImageNet is based on WordNets**
- **ImageNet**
  - **Contains 14 M images as 2014**
  - **21k synonym set, synset**
  - **Each synset is populated about 650 images**

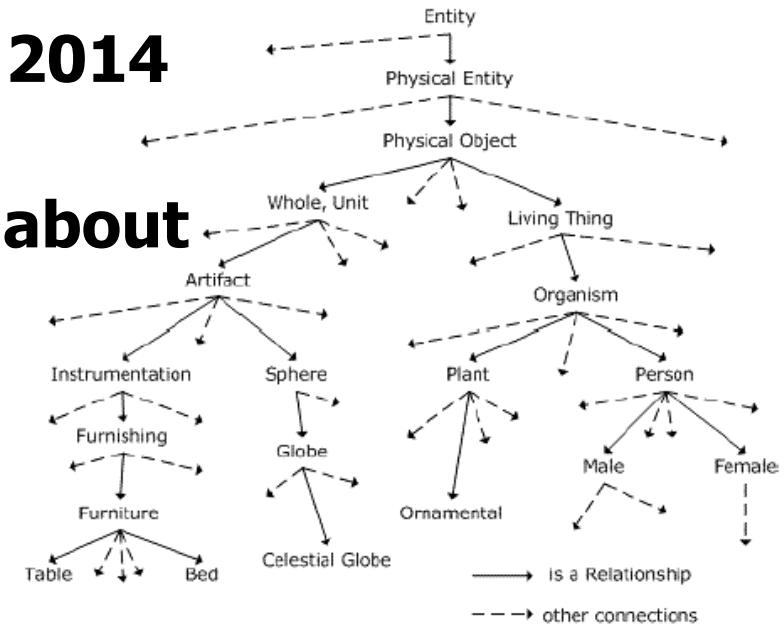


Fig. 2. An example of WordNet nouns taxonomy

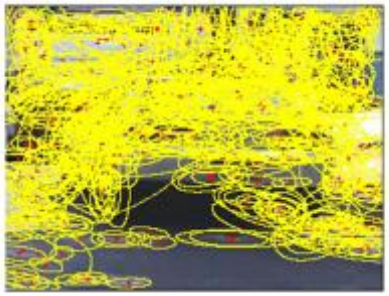
KAIST

Basic issues

- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

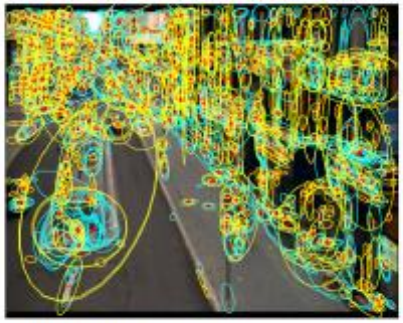
Representation

- **Building blocks: sampling strategy**



Interest operators

Dense, uniformly



Multiple interest operators

Randomly

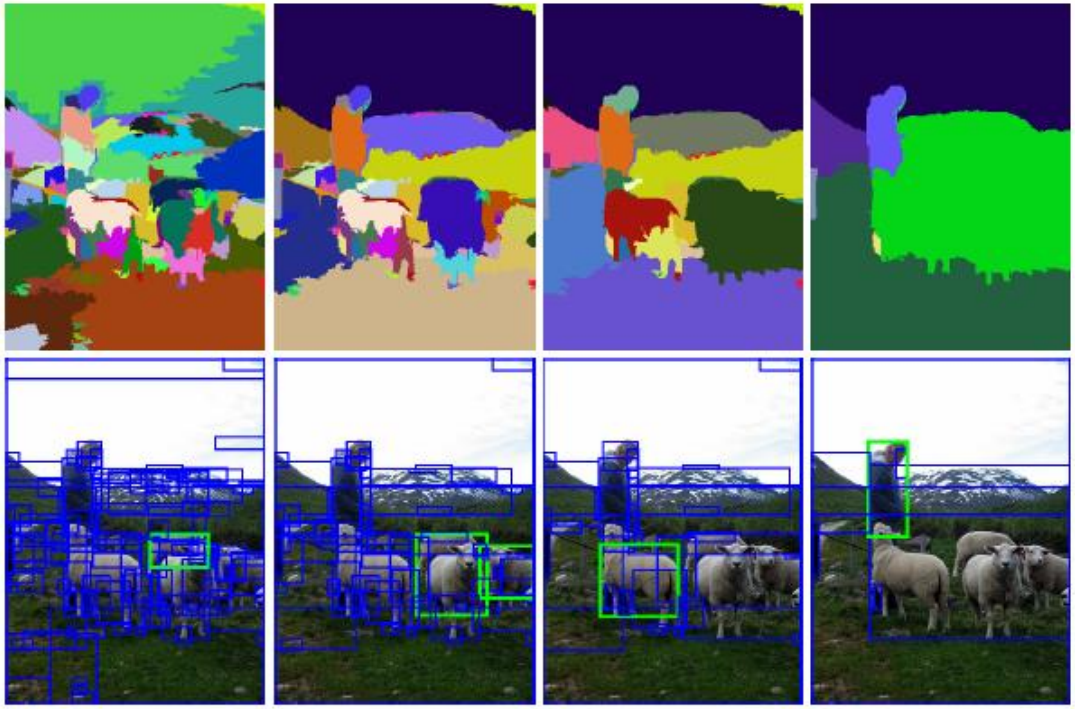
Image credits: L. Fei-Fei, E. Nowak, J. Sivic

- **Recently, features from convolution neural nets**

KAIST

Region Proposals

- **Adopted commonly by many recognition approaches**

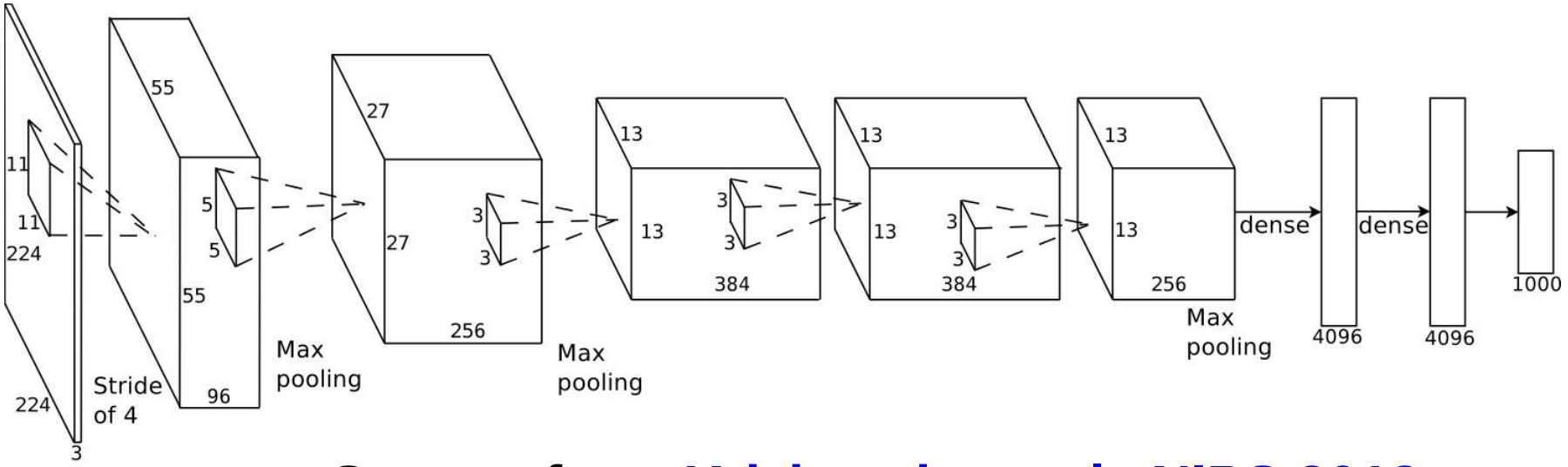


Identify different regions as candidates of objects
Selective Search, Uijlings et al.

KAIST

Convolutional Neural Network (CNN)

- **Features from some layers of CNNs**

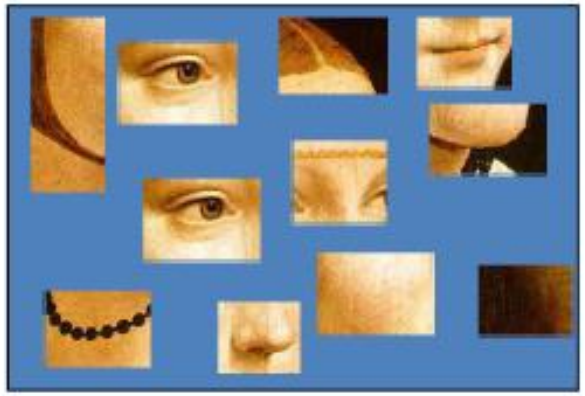


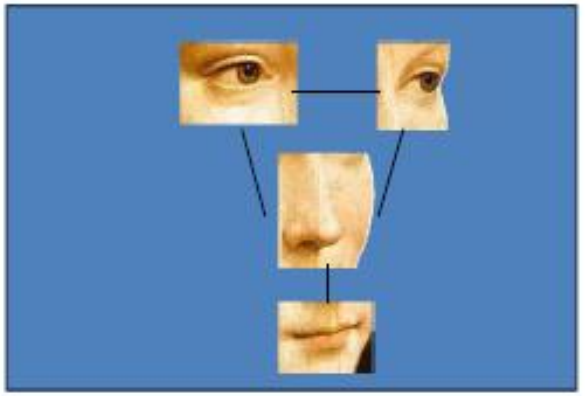
System from Krizhevsky et al., NIPS 2012

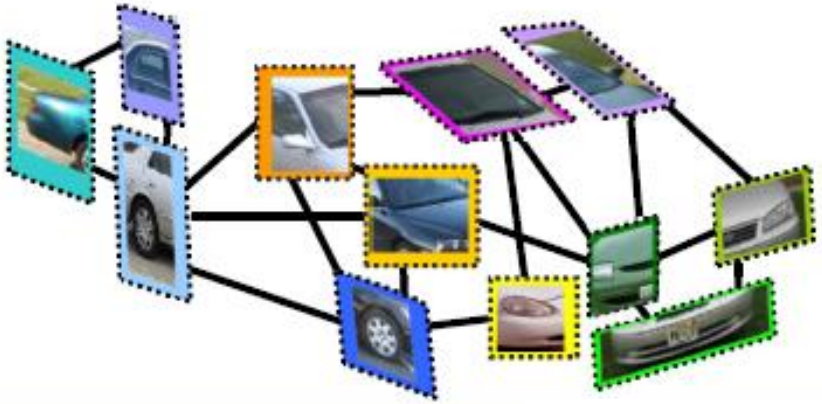
KAIST

Representation

– Appearance only or location and appearance







Object categorization: the statistical viewpoint

$$p(zebra \mid image)$$

vs.

$$p(no\ zebra \mid image)$$

- Bayes rule: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

$$\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$$

Object categorization: the statistical viewpoint

$$p(zebra \mid image)$$

vs.

$$p(no\ zebra \mid image)$$

- Bayes rule: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

Object categorization: the statistical viewpoint

- Discriminative methods model posterior
- Generative methods model likelihood and prior

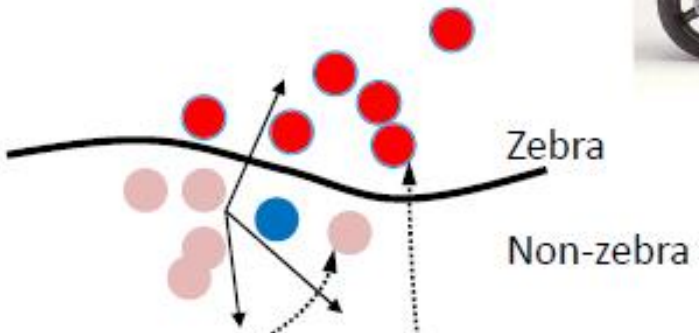
- Bayes rule:

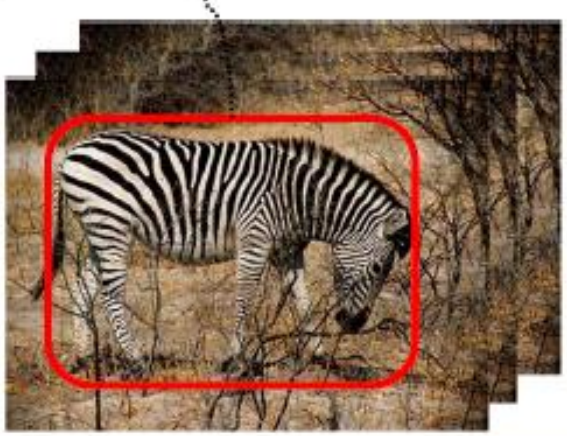
$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

Discriminative models

- Modeling the posterior ratio:

$$\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$$

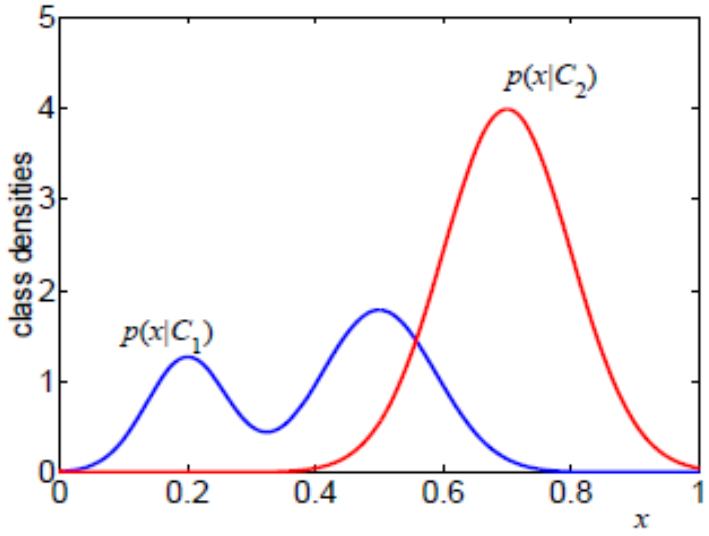




Generative models

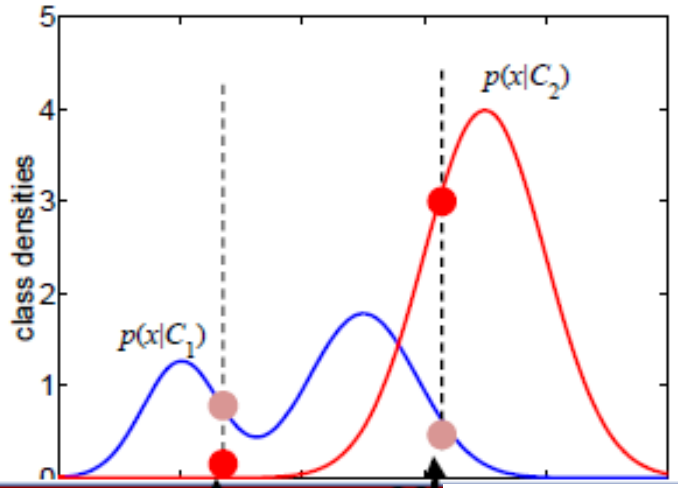
- Modeling the likelihood ratio:

$$\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}$$



Generative models

$p(image \mid zebra)$	$p(image \mid no\ zebra)$
High	Low
Low	High



Basic issues

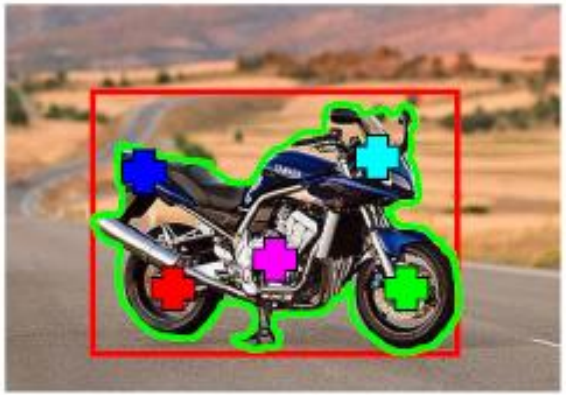
- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Learning

- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)

Learning

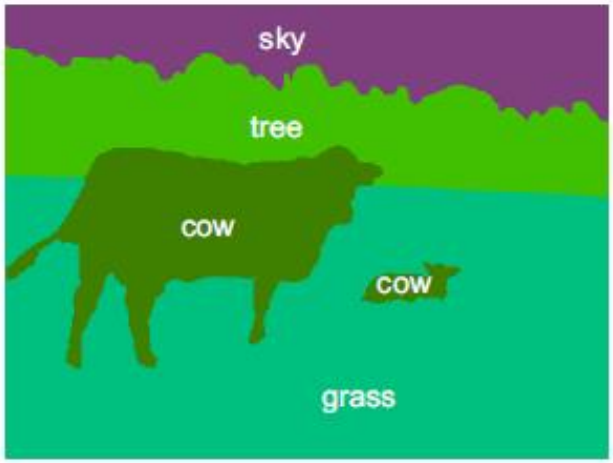
- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental
- Priors



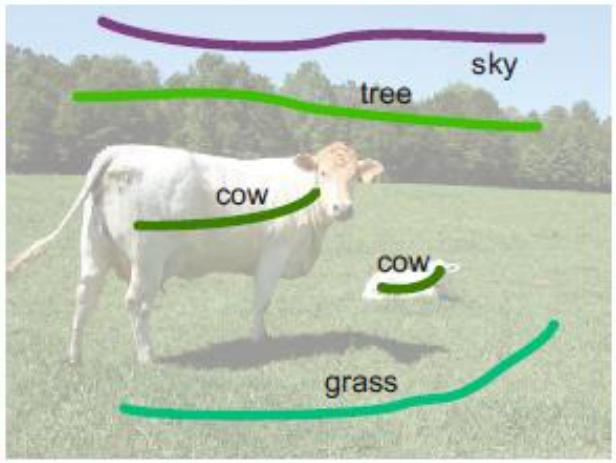
Scribble-Supervised Convolutional Networks for Semantic Segmentation [CVPR 16]

- **Deep learning requires lots of data, but how can we prepare such data?**

(a) image



(b) mask annotation



(c) scribble annotation

- **Allow users just a few strokes, and learn segmentation from them**
- **How about image/video search?**

KAIST

Basic issues

- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Recognition

– Recognition task: classification, detection, etc..

Recognition

– Recognition task

– Search strategy: Sliding Windows

Viola, Jones 2001,

- Simple
- Computational complexity (x,y, S, θ, N of classes)

- BSW by Lampert et al 08
- Also, Alexe, et al 10

Recognition

– Recognition task

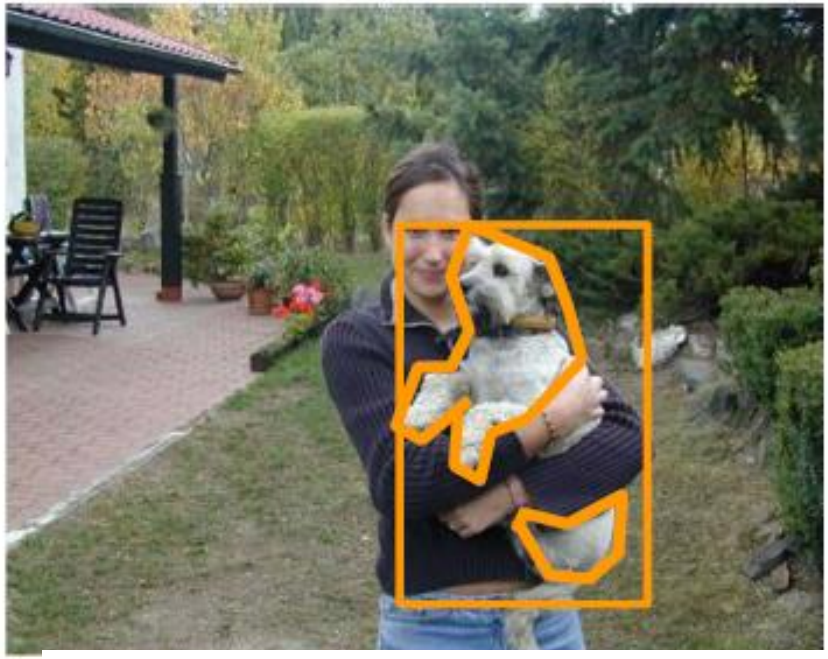
– Search strategy: Sliding Windows

Viola, Jones 2001,

- Simple
- Computational complexity (x,y, S, θ, N of classes)

  - BSW by Lampert et al 08
  - Also, Alexe, et al 10

- Localization
  - Objects are not boxes



Recognition

– Recognition task

– Search strategy: Sliding Windows

Viola, Jones 2001,

- Simple
- Computational complexity (x,y, S, θ, N of classes)

  - BSW by Lampert et al 08
  - Also, Alexe, et al 10

- Localization
  - Objects are not boxes
  - Prone to false positive

Non max suppression:
Canny '86
....
Desai et al , 2009

Recognition

- Recognition task
- Search strategy
- Attributes

•Savarese, 2007
•Sun et al 2009
• Liebelt et al., '08, 10
•Farhadi et al 09

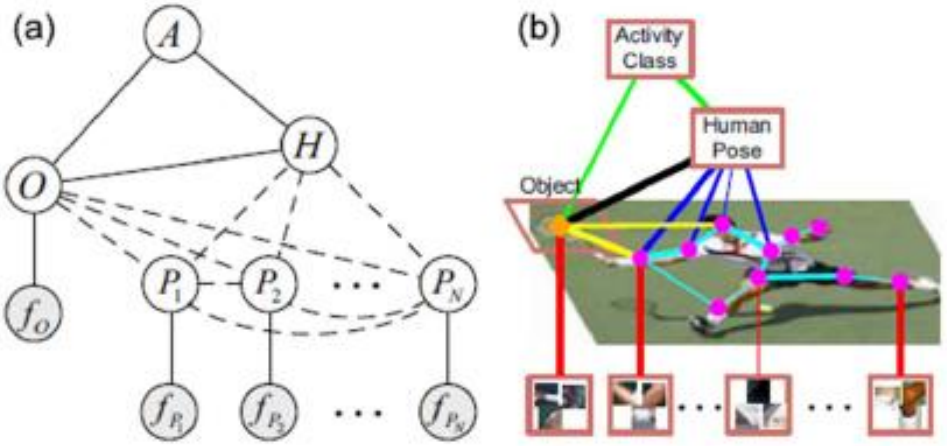
Category: car
Azimuth = 225º
Zenith = 30º

- It has metal
- it is glossy
- has wheels

•Farhadi et al 09
• Lampert et al 09
• Wang & Forsyth 09

Recognition

- Recognition task
- Search strategy
- Attributes
- Context

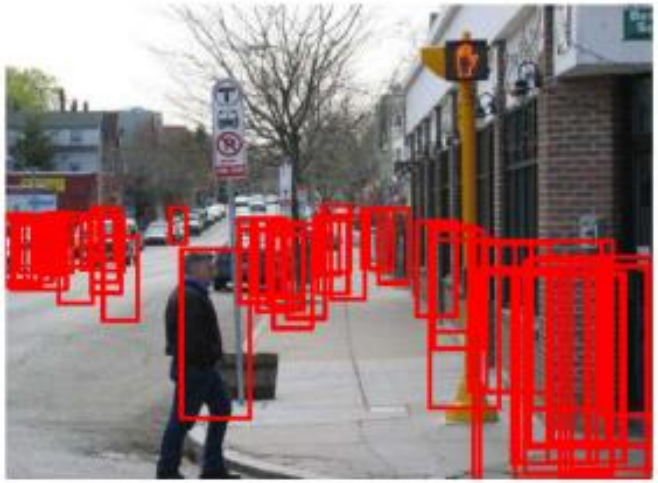


Semantic:

- Torralba et al 03
- Rabinovich et al 07
- Gupta & Davis 08
- Heitz & Koller 08
- L-J Li et al 08
- Yao & Fei-Fei 10

Geometric

- Hoiem, et al 06
- Gould et al 09
- Bao, Sun, Savarese 10



Class Objectives were:

- **Introduction to object detection**
  - **Representation (features)**
  - **Learning**
  - **Recognition**

- **Recently performed within deep neural net with an end-to-end optimization**

KAIST

Next Time and Homework

- **Bag of visual words approach**

- **Go over the next lecture slides**
- **Come up with one question on what we have discussed today**
  - **1 for typical questions (that were answered in the class)**
  - **2 for questions with thoughts or that surprised me**

- **Write questions at least 4 times**

KAIST