Image Classification

Luca Iocchi

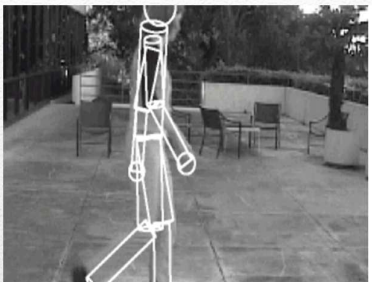
DIAG, Sapienza University of Rome, Italy

Machine Learning 2017/2018

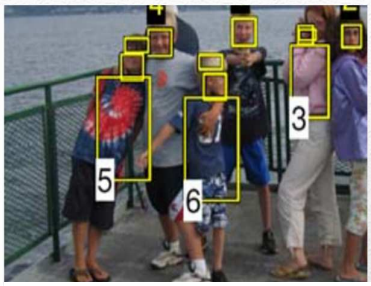
ML in Image Analysis

- Person/object detection, recognition, identification, and tracking
- Face detection, recognition
- Activity recognition
- Gestures recognition
- Expressions classification
- Crowd analysis
- ....

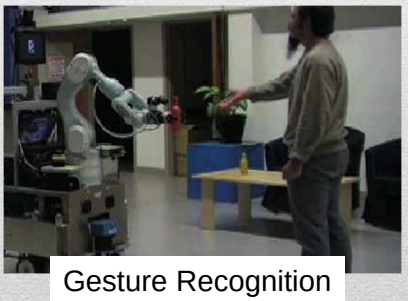
ML & Computer Vision



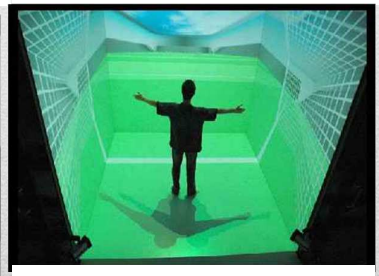
Person detection/tracking



Face detection/recognition



Gesture Recognition



Activity Recognition

Some images from Computer Vision: Algorithms and Applications
© 2010 Richard Szeliski, Microsoft Research

Flow of processing

input data

OBJECT DETECTION

OBJECT RECOGNITION

time → T_i T_{i+1}

id

TRACKING

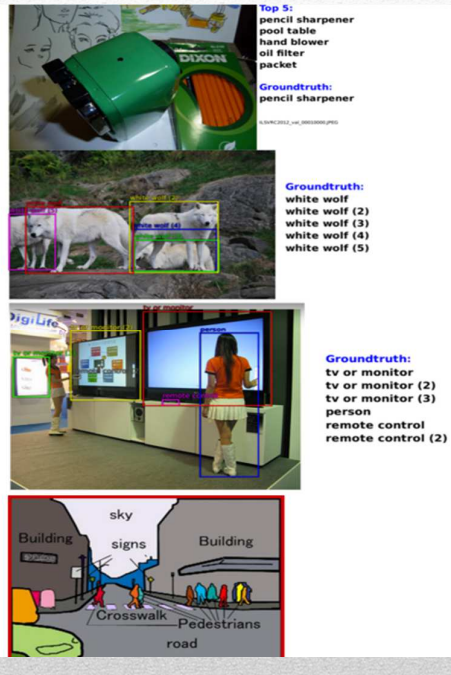
BEHAVIOR AND ACTIVITIES ANALYSIS

John is leaving the scene

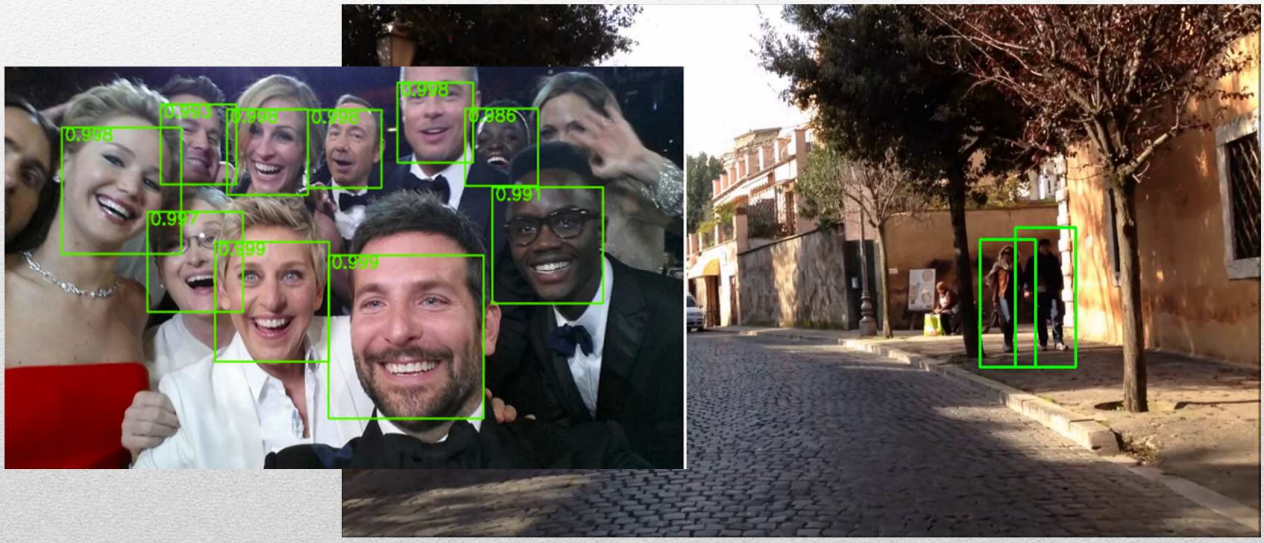
Intelligent distributed surveillance systems: a review
M. Valera and S.A. Velastin

What is object detection?

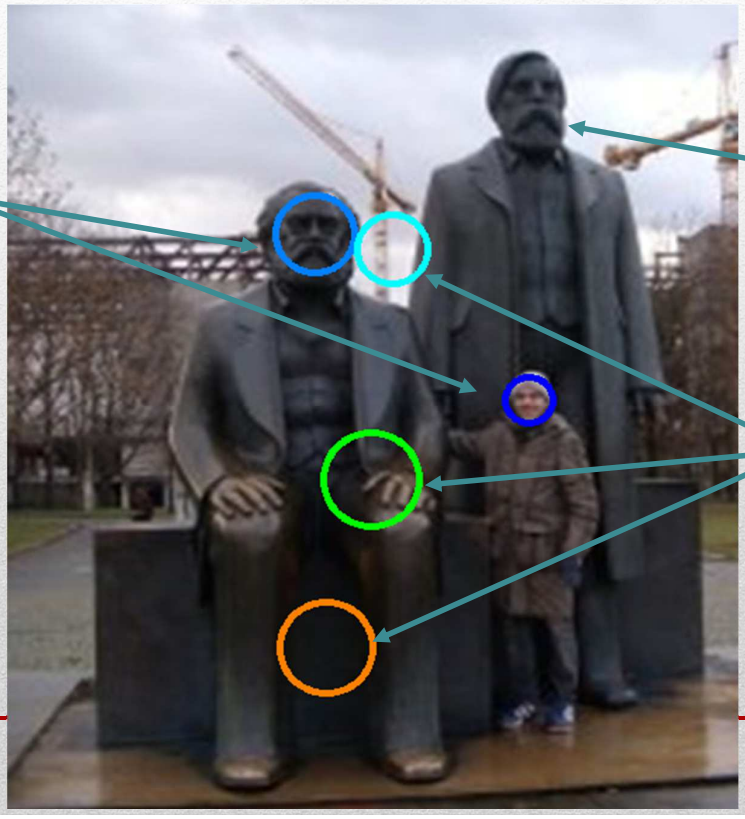
- classification
- localization
- detection
- segmentation



Person/Face Detection

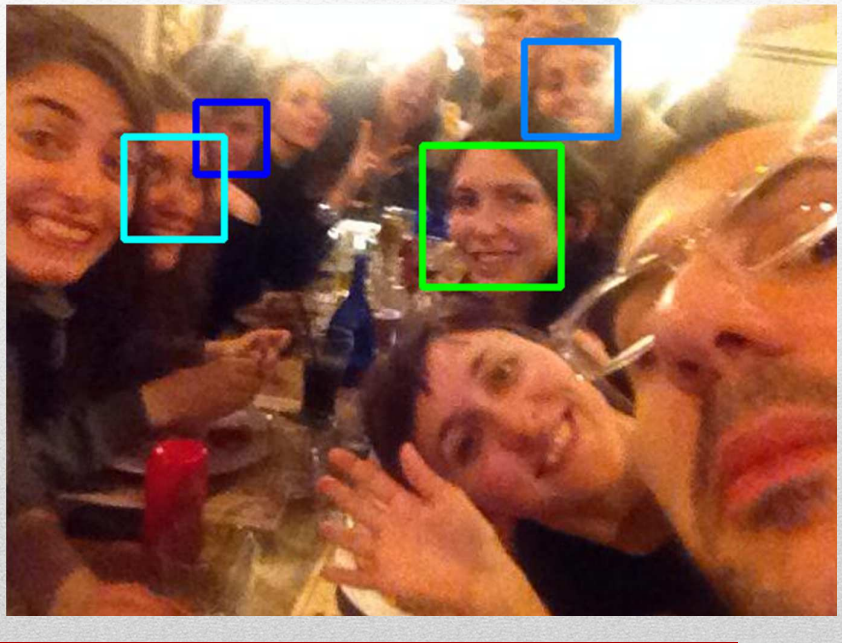


Detection Issues

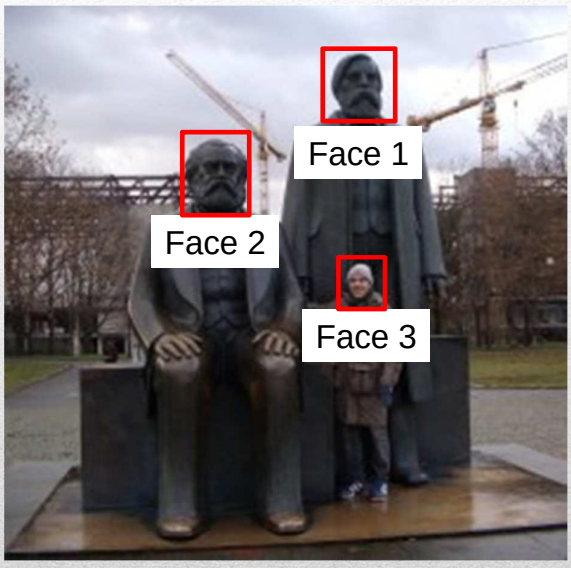


Additional Issues

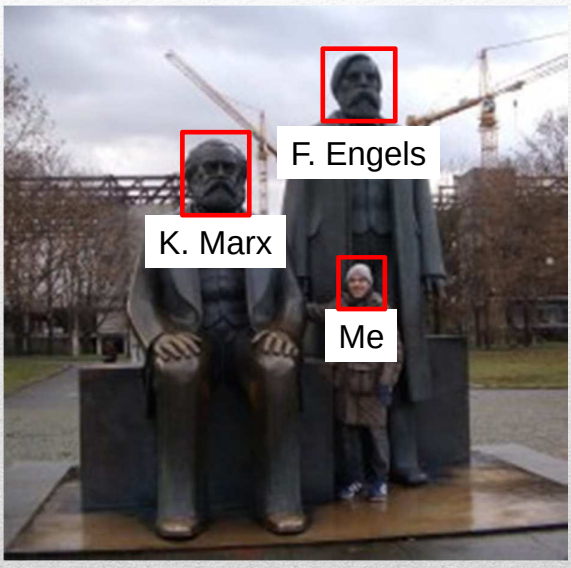
- Rotation
- Blurring
- Illumination
- Occlusion
- Glasses
- …



Detection vs. Recognition

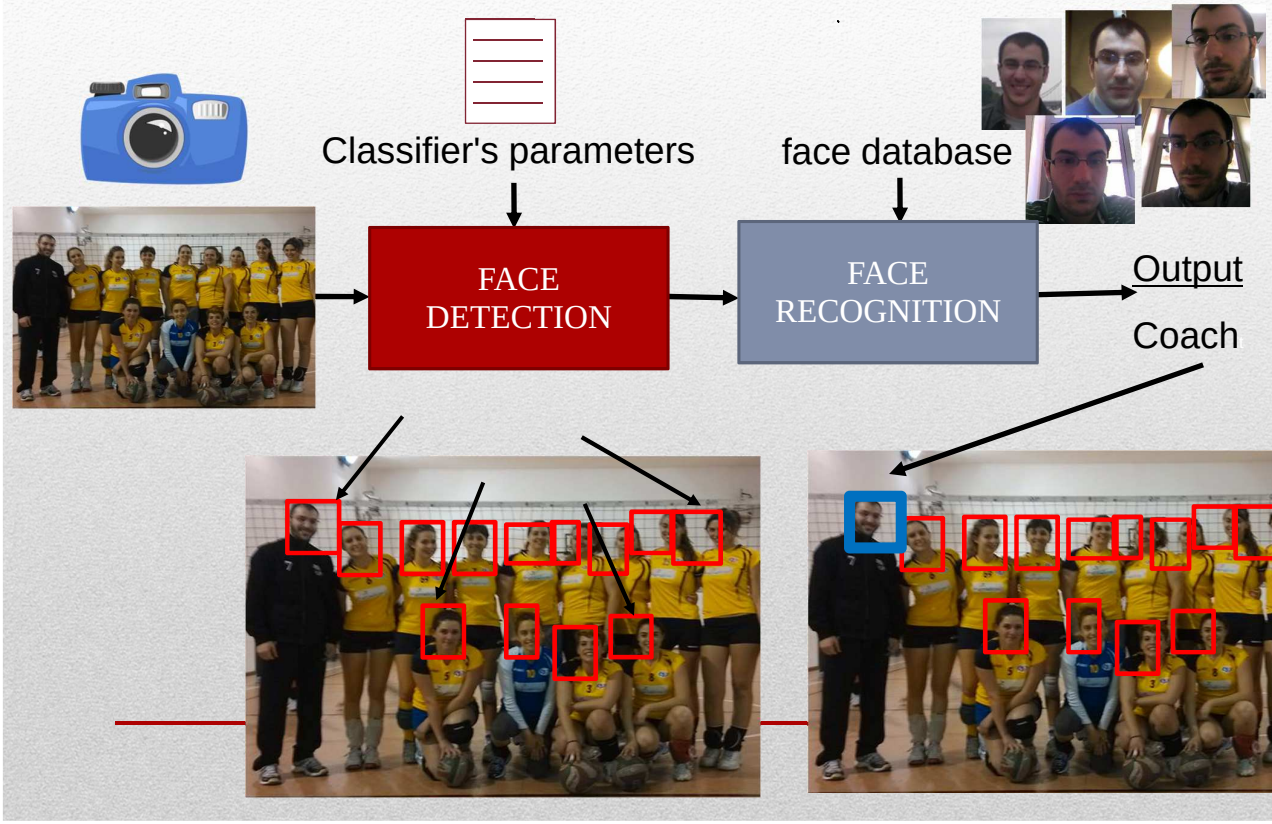


detection



recognition

Detection and Recognition



Face Detection in

Bayesian Learning

$$p(zebra \mid image)$$

vs.

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

BAYES:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

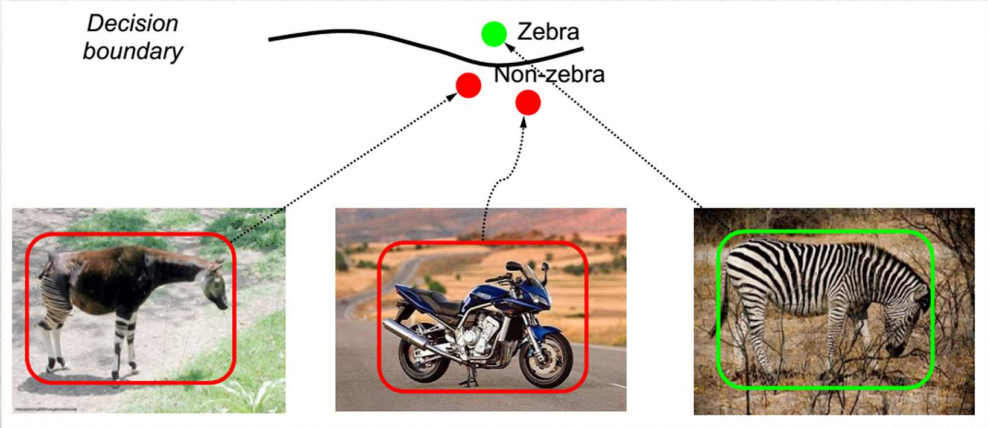
we look for (posterior)

we measure (likelihood)

we know (priors)

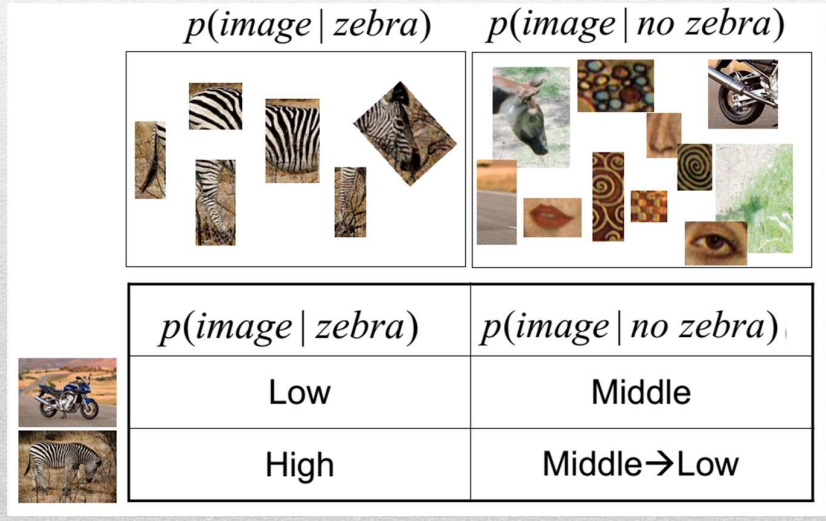
Bayesian Learning

Discriminative models: estimate P(zebra|image)



Bayesian Learning

Generative models: estimate P(image|zebra)



Object Detection Steps

1. **Feature Computation**

   What features?
   How can they be computed as quickly as possible?

2. **Feature Selection**

   What are the most discriminating features?

3. **Detection (in real time)**

   Must focus on potentially positive areas

Image Features

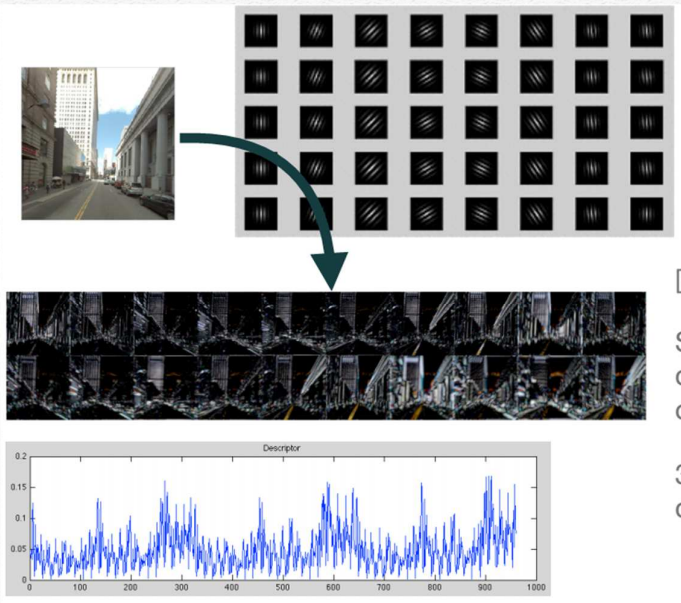
Global features

- GIST

Local features

- Corners (Harris, Shi-Tomasi, FAST, …)
- Scale invariant (SIFT, SURF, …)
- Binary (BRIEF, ORB, …)

GIST

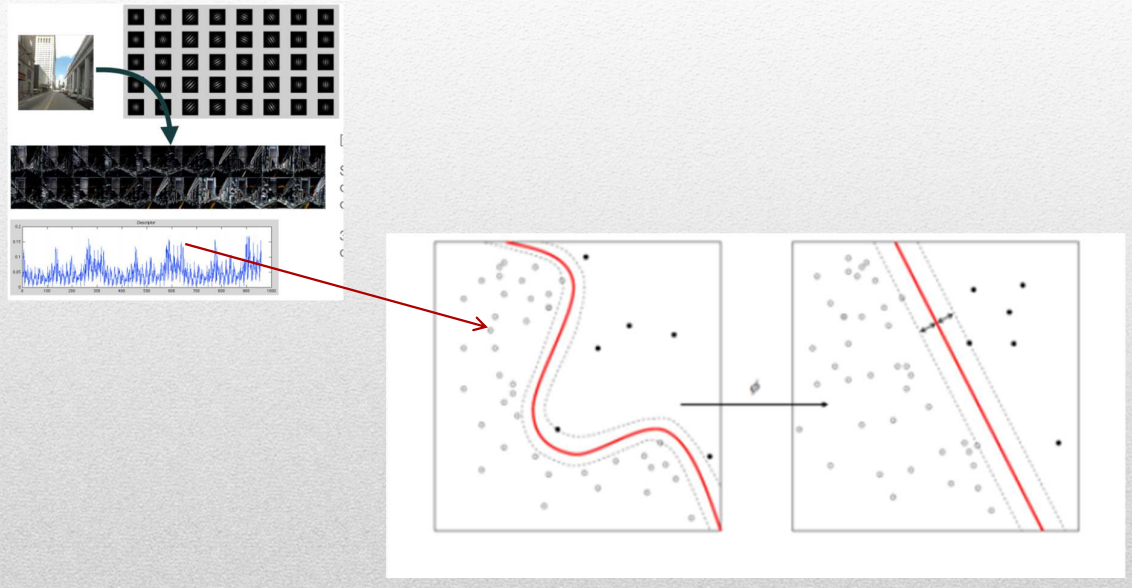


[Oliva&Torralba'01,06]

Several Gabor filters different scales & orientations

320 x 3 (RGB) descriptor

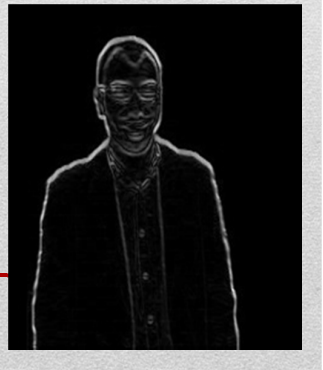
GIST + SVM



Histogram Of Oriented Gradients (HOG)

1-D mask centered operator
- X-filtering and Y-filtering

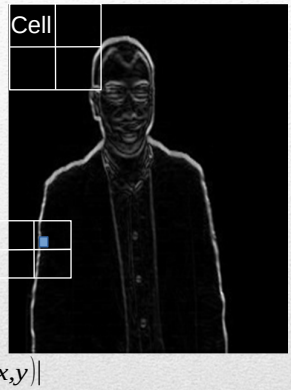
Gradient Vector
- magnitude and direction



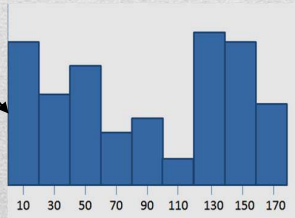
Orientation Binning

Cell creation

- **Cell** : group of pixels used to collect gradient information locally (e.g., 8x8 pixels for people detection)



Θ ; $|\nabla f(x,y)|$



Histograms channels

- Unsigned gradients
  - Separation over 0 to 180 degrees
  - 9 bins and 20 degrees for each bin
- Signed gradients
  - Separation over 0 to 360 degrees

Descriptor Block

Normalized histogram for a group of cells located in the same block.

Same Cell contributes many times through the overlapping proceeds

- Block stride = 8x8 pixels.

Detection window : size of 64x128 pixels.

Block descriptor : size of 2 cells = 16 x 16 pixels.

Cell : size of 8x8 pixels.

The Viola and Jones Method

- **Very popular method**
- **Recognition is very fast**
  (e.g., real-time for digital cameras)
- **Key contributions**
  1. **Integral image for fast feature extraction**
  2. **Boosting (Ada-Boost) for face detection**
  3. **Attentional cascade for fast rejection of non-face sub-windows**

Training may take a long time

[1] P. A. Viola, M. J. Jones. Rapid object detection using a boosted cascade of simple features, CVPR, pp.511-518, 2001
[2] P. A. Viola, M. J. Jones. Robust Real-Time Face Detection. IJCV 57(2), pp. 137-154, 2004

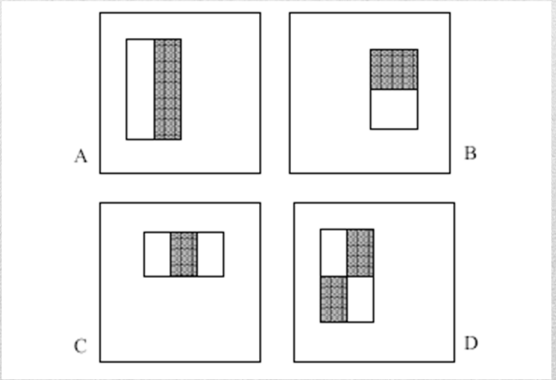
Viola and Jones Steps

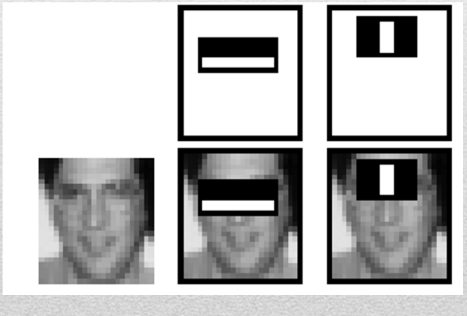
1. **Feature Computation**
   - **Quick Feature Computation**
     - Rectangle features
     - Integral image representation
2. **Feature Selection**
   - **Efficient classification**
     - Ada-Boost training algorithm
3. **Detection (in real time)**
   - **Real-timeliness**
     - A cascade of classifiers

Features

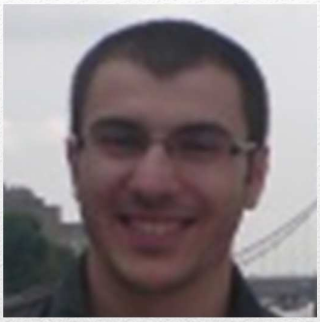
Four basic types

- Easy to calculate
- White areas are subtracted from the black ones
- Integral image representation makes feature extraction faster

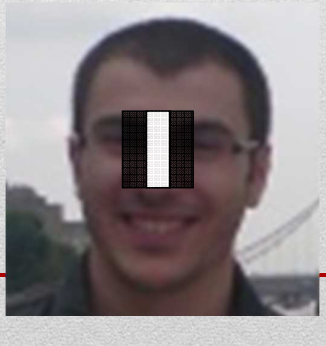


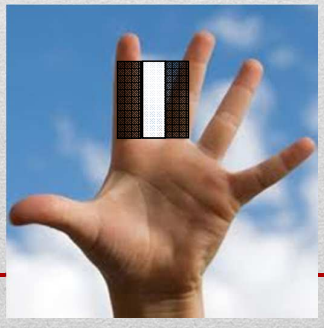


Rectangle Features



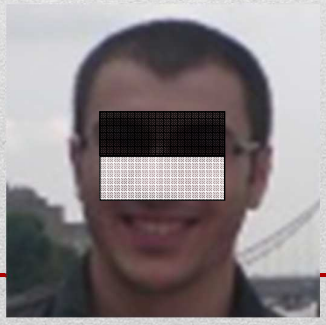
Value = ∑ (pixels in white area) – ∑ (pixels in black area)

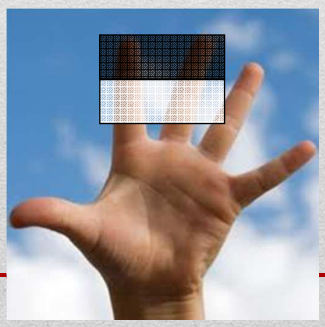




Rectangle Features

Value = ∑ (pixels in white area) – ∑ (pixels in black area)



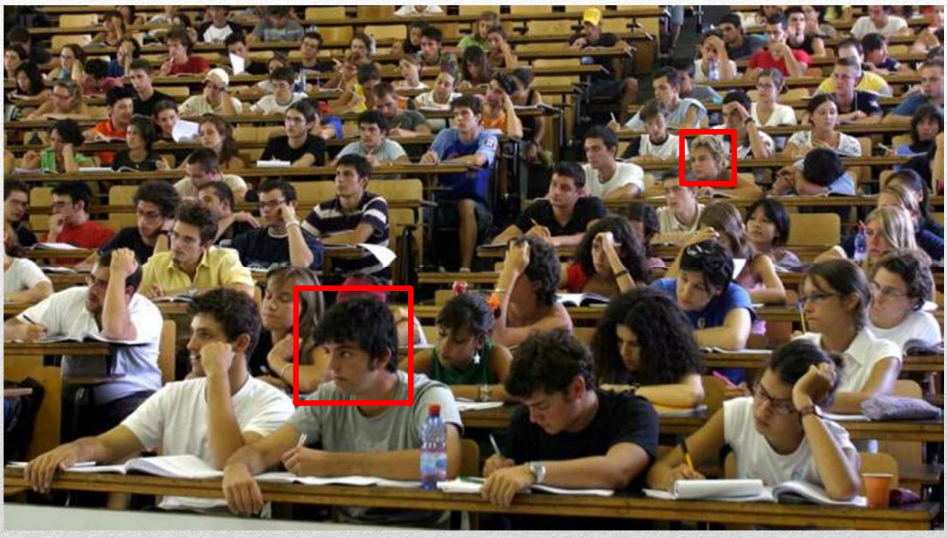


Sliding window search

Slide a window (e.g., 30x30) across the image and evaluate the current portion of the image w.r.t. the object model at every location

We assume that the number of locations where the object is present is (very) small

Multiple scale search



Multiple scale search

Image pyramid

Readings

P. Viola and M.J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001.

P. Viola and M.J. Jones. Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade. Advances in Neural Information Processing System, pp. 1311-1318, 2001

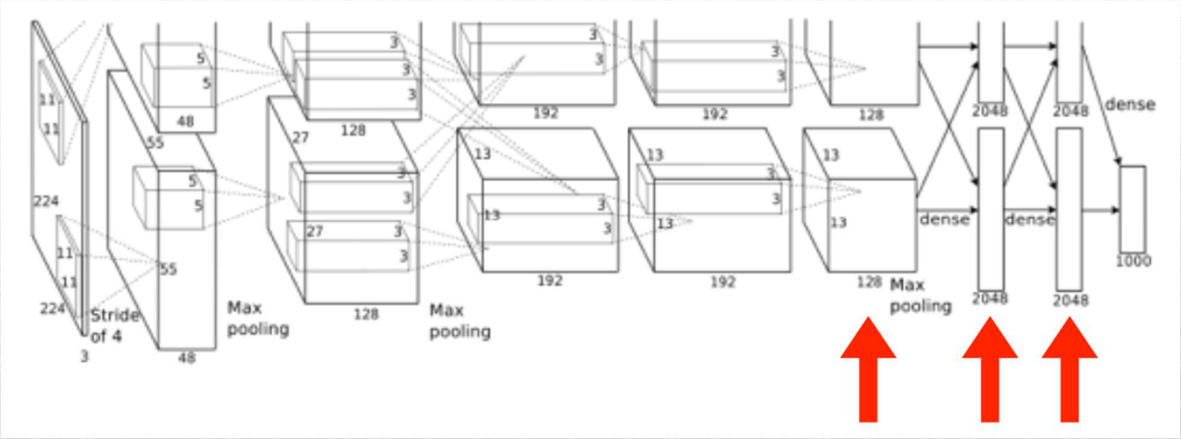
Image Representation Features

Global features

L

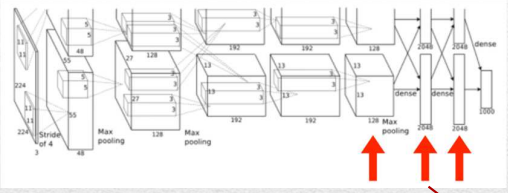
- Scale invariant (SIFT, SURF, …)
- Binary (BRIEF, ORB, …)

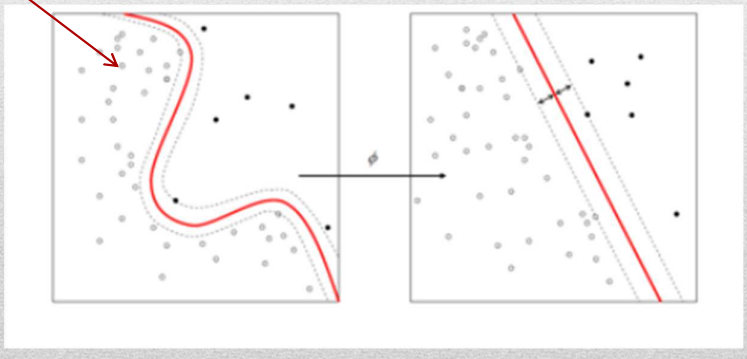
CNN



Deep Features

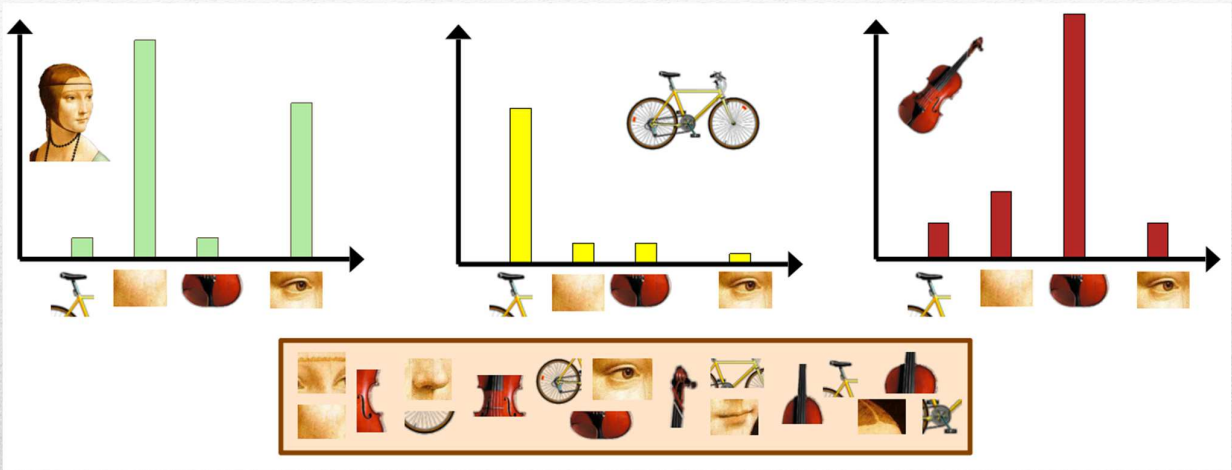
CNN + SVM



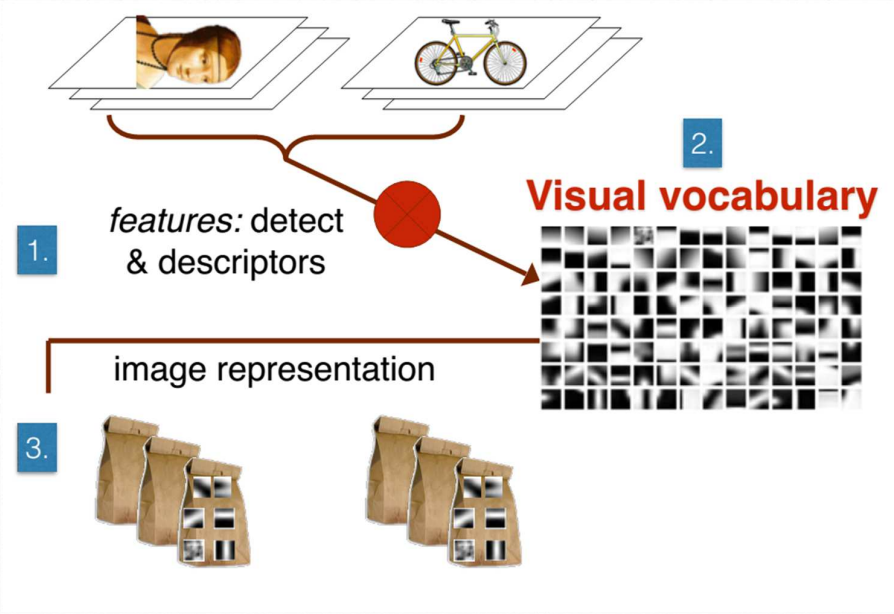


Bag of words

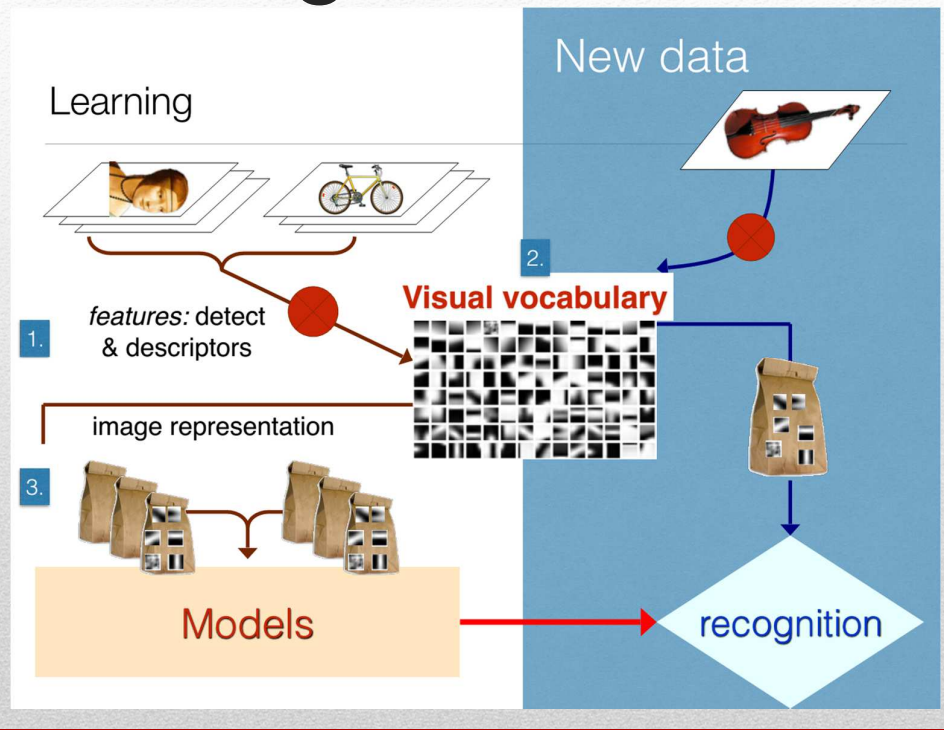
Bag of words



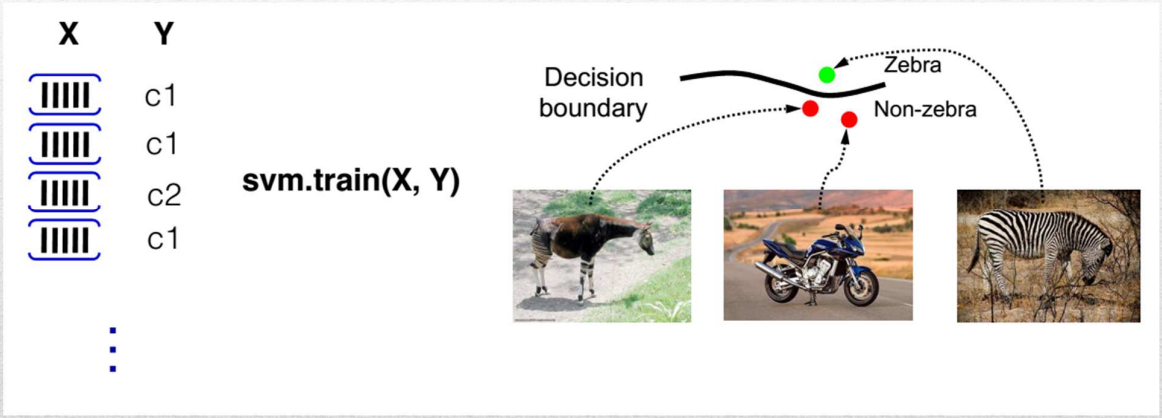
Bag of words



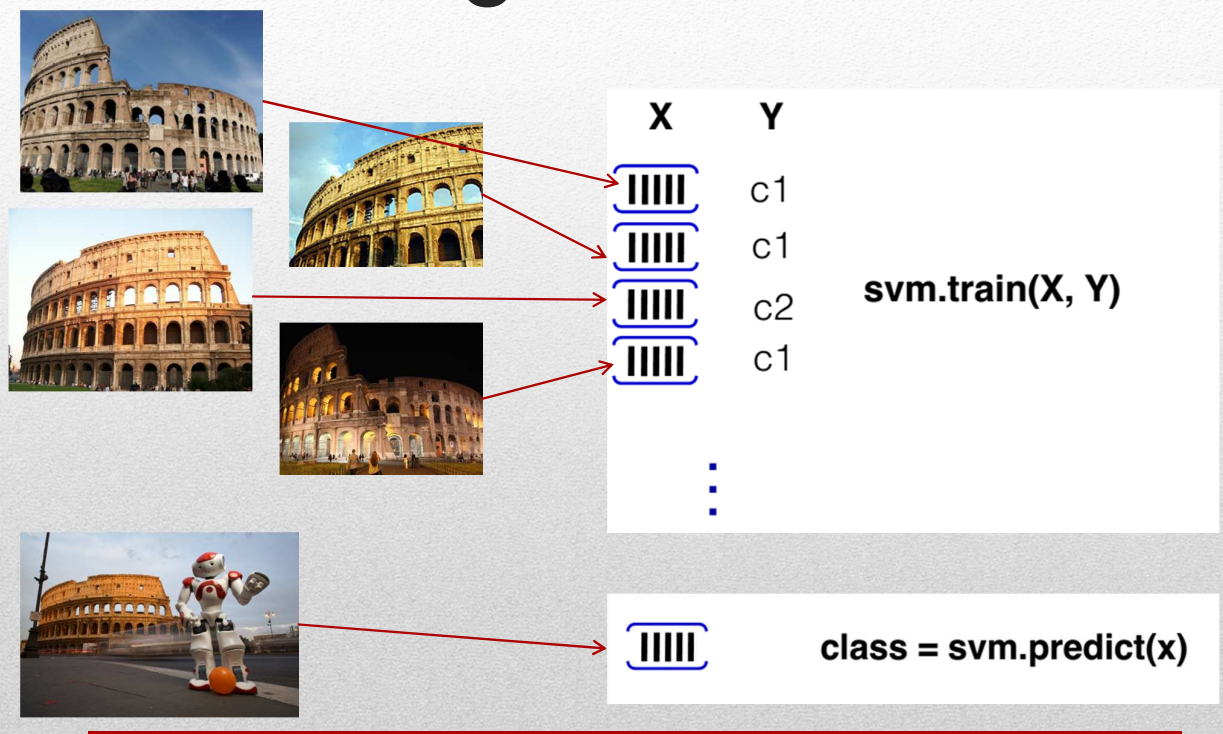
Bag of words



Bag of words



Bag of words



Readings

Aude Oliva, Antonio Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. IJCV, 2001.

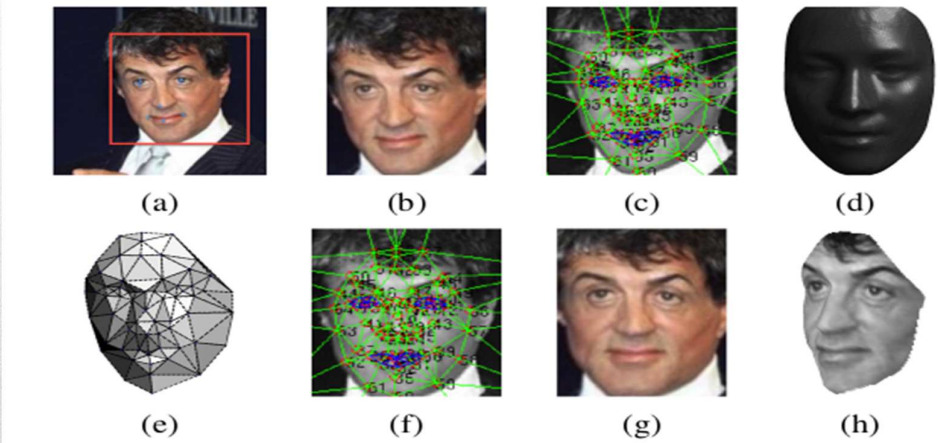
Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton . ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012

Evan Shelhamer, Jeff Donahue, Jon Long, Yangqing Jia, and Ross Girshick. Deep Learning for Vision: a Hands On Tutorial with Caffe

Fei-Fei, Karpathy, Johnson. Convolutional Neural Networks for Visual Recognition (http://cs231n.stanford.edu)

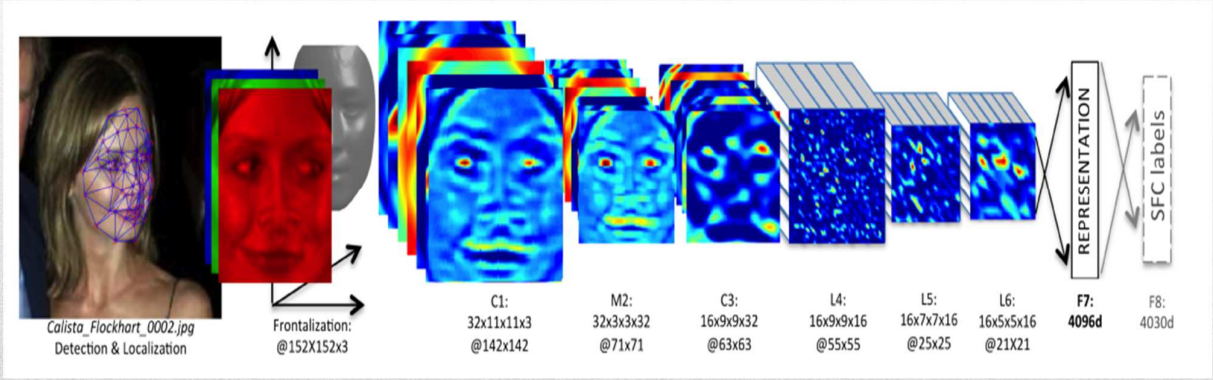
Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. ACM MM Open Source Competition 2014

Face Alignment



Deep Face

Deep Face



DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Open Face

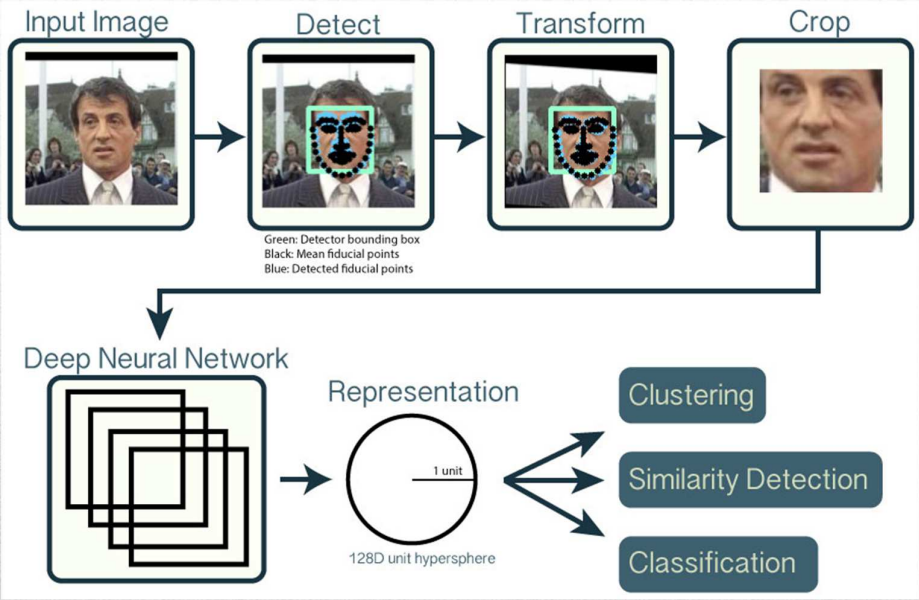
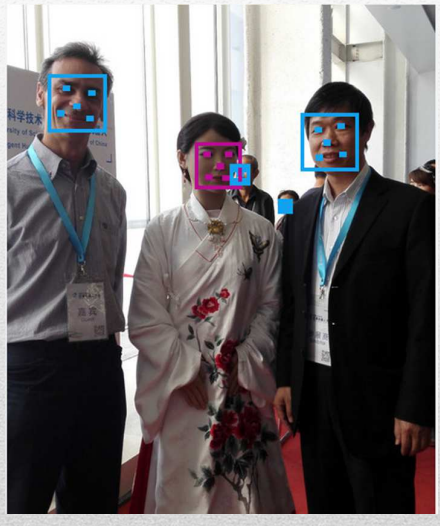


Image from Open Face web ste

Microsoft Cognitive Services

Face API

- Detection
- Verification
- Identification
- Similarities
- Grouping



https://azure.microsoft.com/en-us/services/cognitive-services/face/

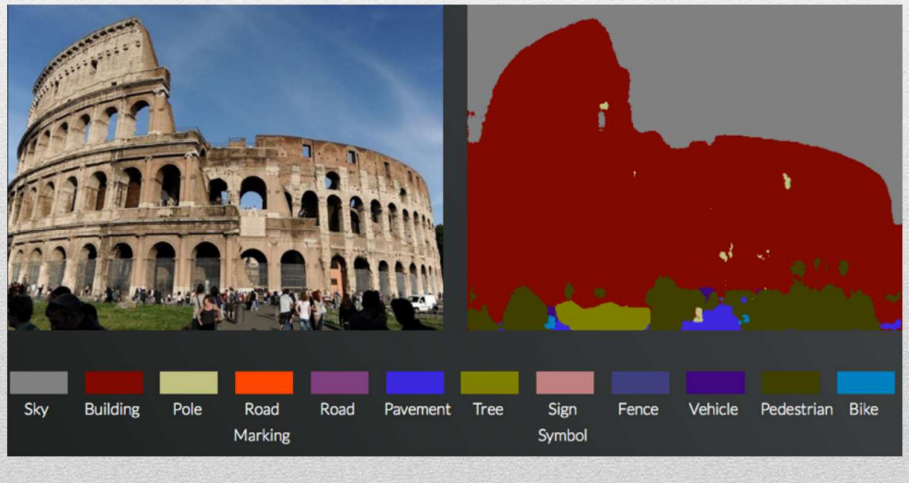
Readings

Y. Taigman, M. Yang, M. Ranzato, L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in IEEE Conference on Computer Vision and Pattern Recognition pp. 1701-1708, 2014

Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proc. of CVPR 2015.

Image Segmentation

Pixel-wise classification → assign a semantic label to each pixel in an image



Skin classification

Image Segmentation

Super-pixels: pixels grouped together according to semantic coherence

- Unsupervised
- Efficient processing

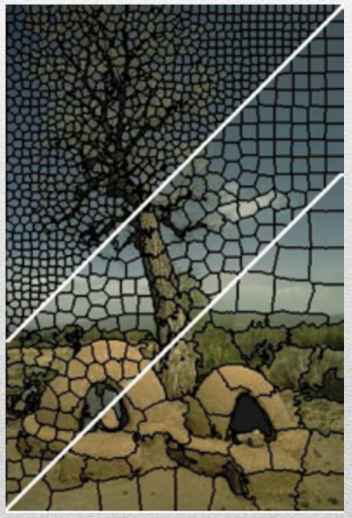


Image Segmentation

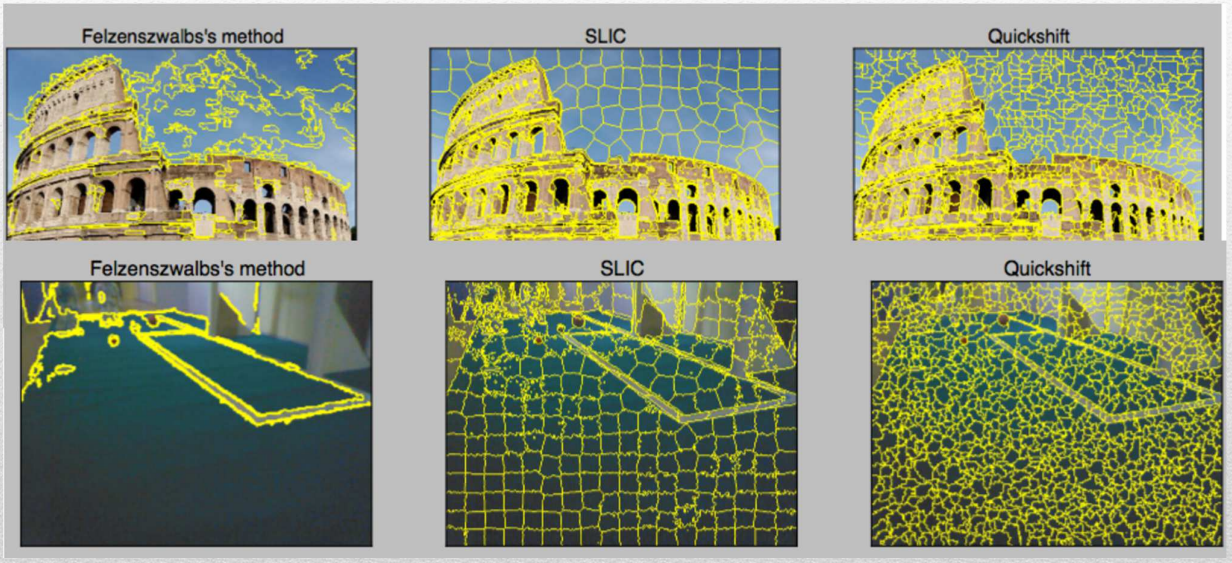


Image Segmentation

Super-pixels + descriptors + semantics

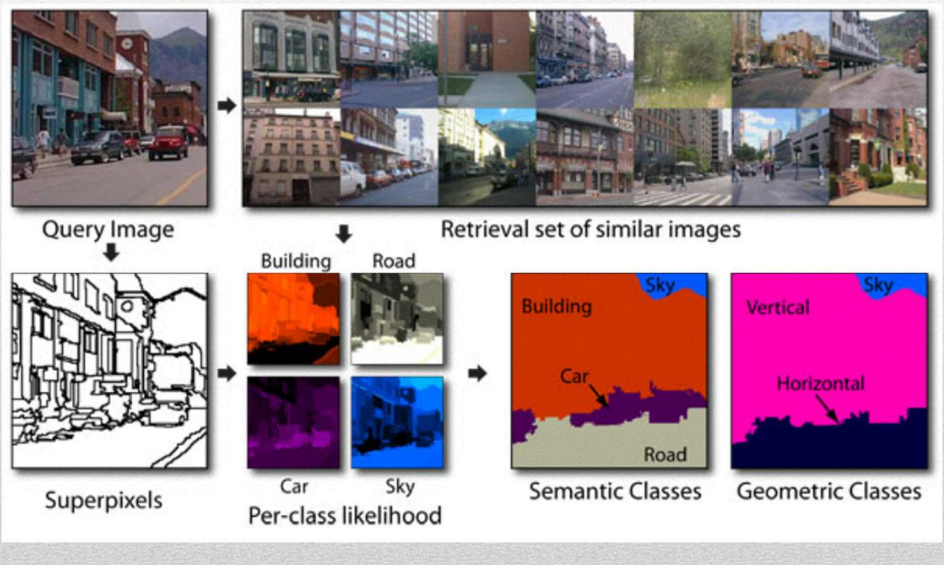
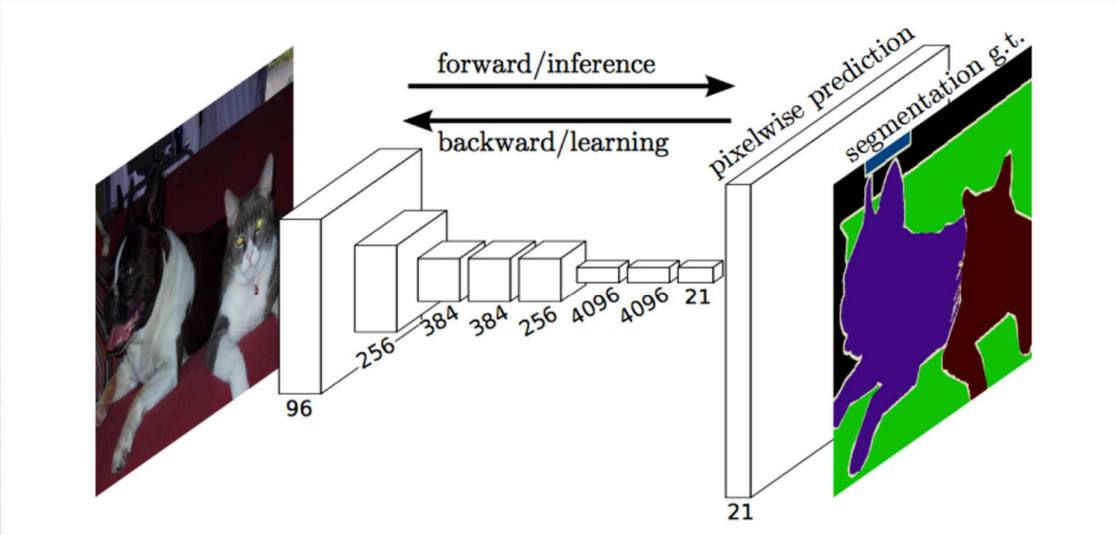


Image Segmentation

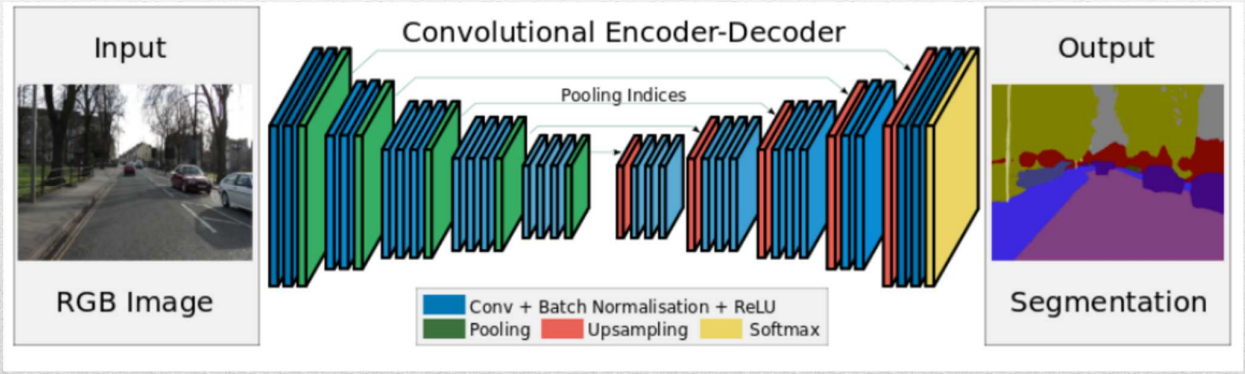
Deep learning



Fully Convolutional Networks for Semantic Segmentation
Jonathan Long, Evan Shelhamer, Trevor Darrell. CVPR 2015

Image Segmentation

Deep learning



Vijay Badrinarayanan, Ankur Handa and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling." CVPR, 2015.

Readings

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR 2014.

Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015

Vijay Badrinarayanan, Ankur Handa and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling." CVPR, 2015.

Credits

Some slides of this presentation adapted from:

- P. Sermanet, “Object Detection with Deep Learning”
- K.H. Wong. “Ch. 6: Face detection”
- P. Viola and T.-W. Yue. “Adaboost for Face Detection”
- D. Miller. “Face Detection & Synthesis using 3D Models & OpenCV”
- S. Lazebnik. “Face detection”
- C. Schmid. “Category-level localization”
- C. Huang and F. Vahid. “Scalable Object Detection Accelerators on FPGAs Using Custom Design Space Exploration”
- P. Smyth. “Face Detection using the Viola-Jones Method”
- K. Palla and A. Kalaitzis. “Robust Real-time Face Detection”
- Ana C. Murillo "Visual Recognition of Human Actions", PhD Course, Sapienza University of Rome, Italy, 14 - 18 March, 2016.