Machine learning in computer vision

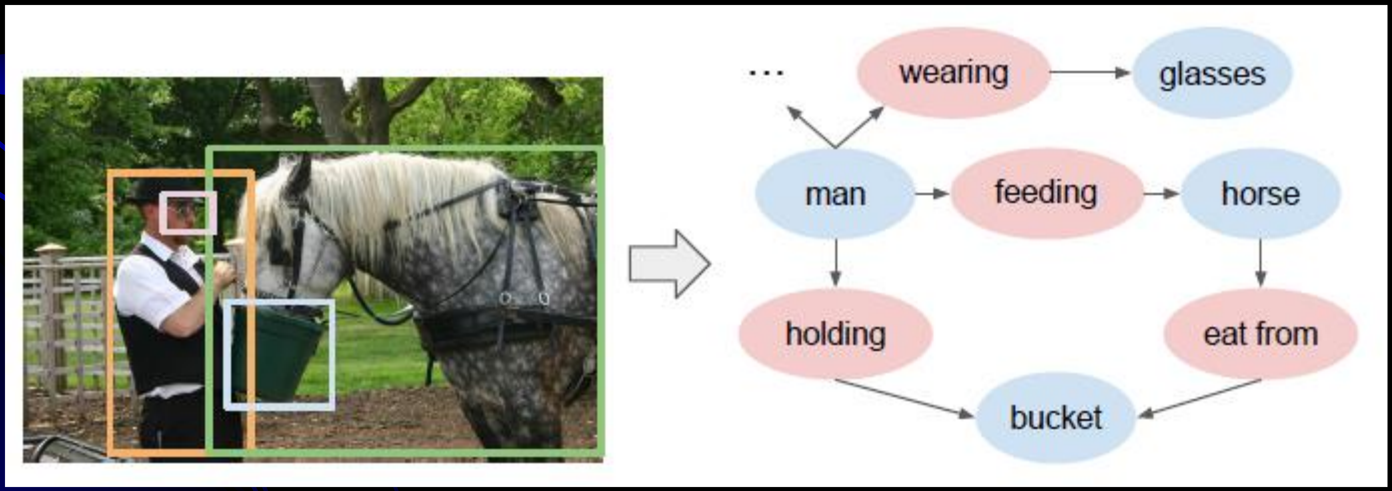
Elena Šikudová

Computer vision tasks

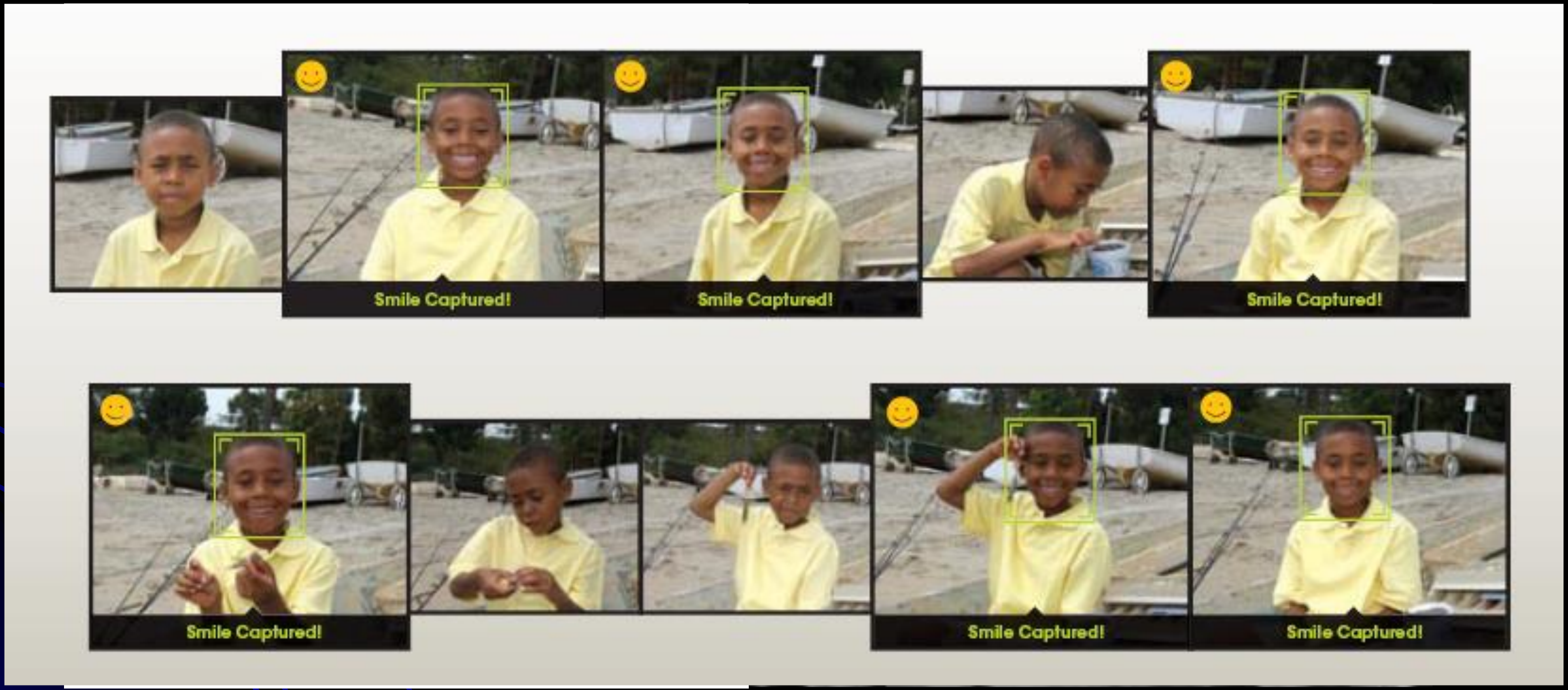
Object detection

Object localization → Scene analysis

Object recognition

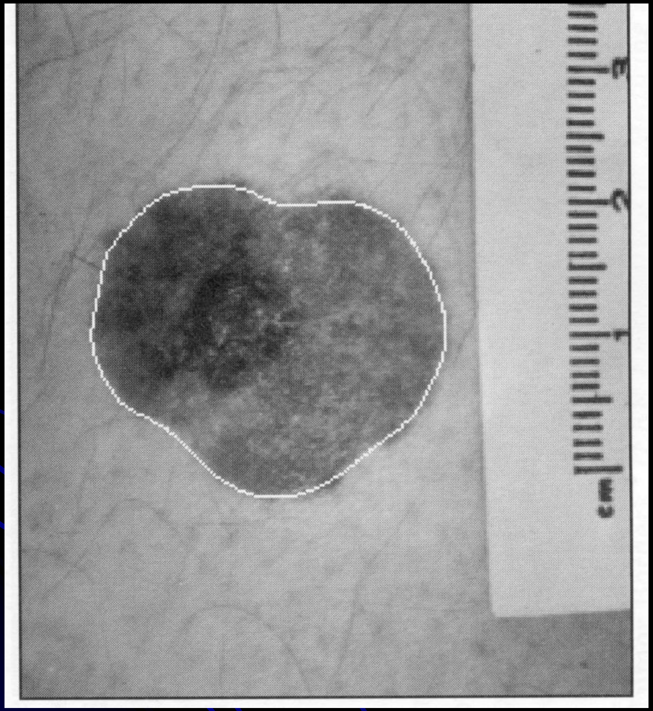


Face detection

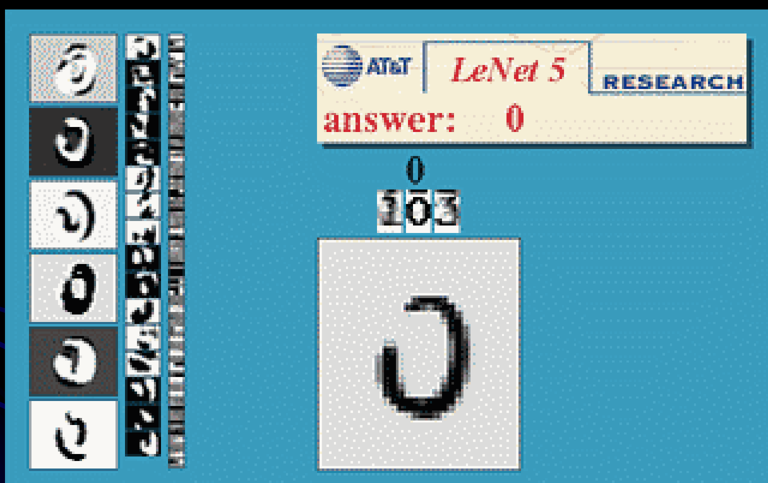


Tumor detection

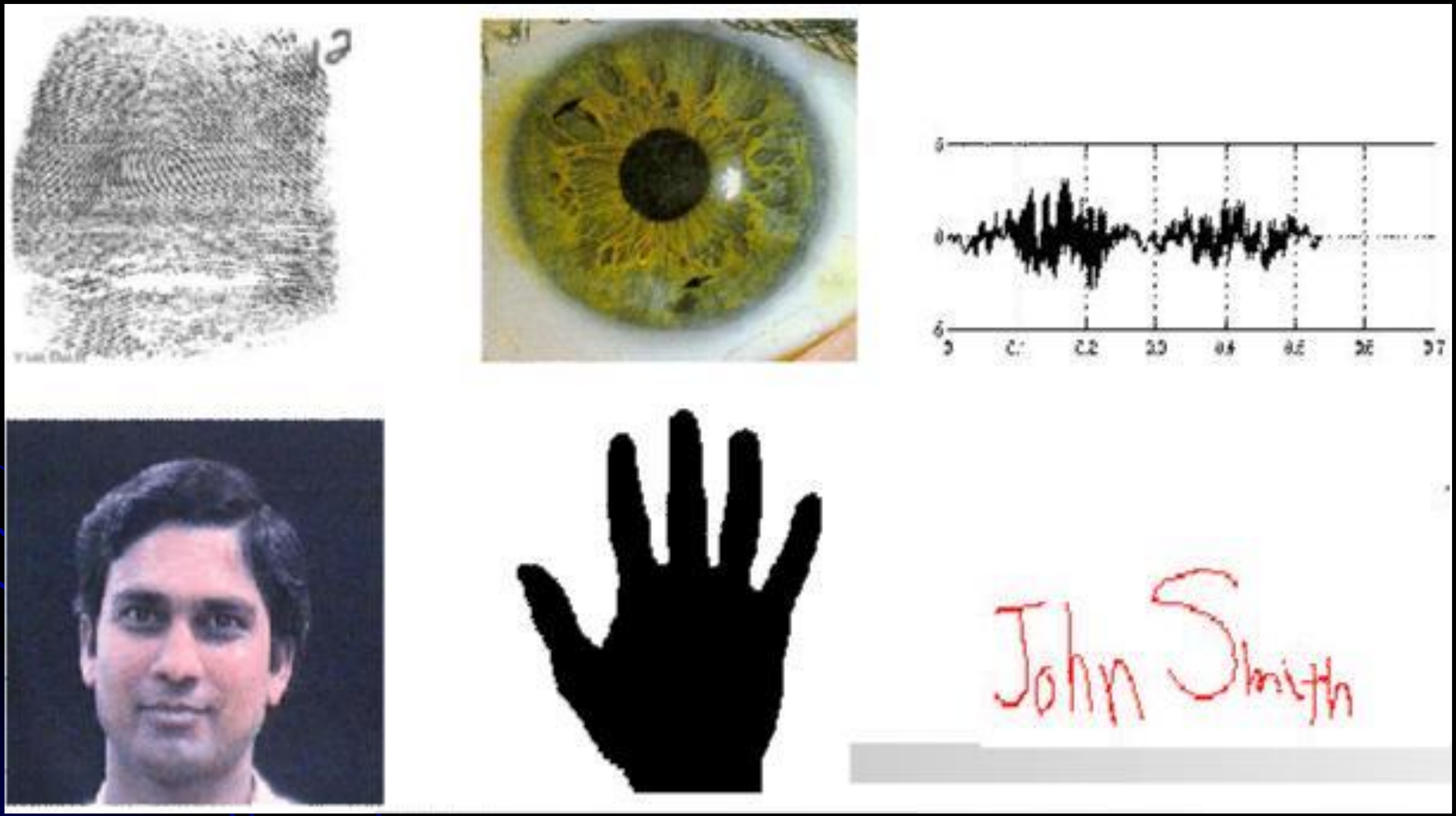
Medical applications



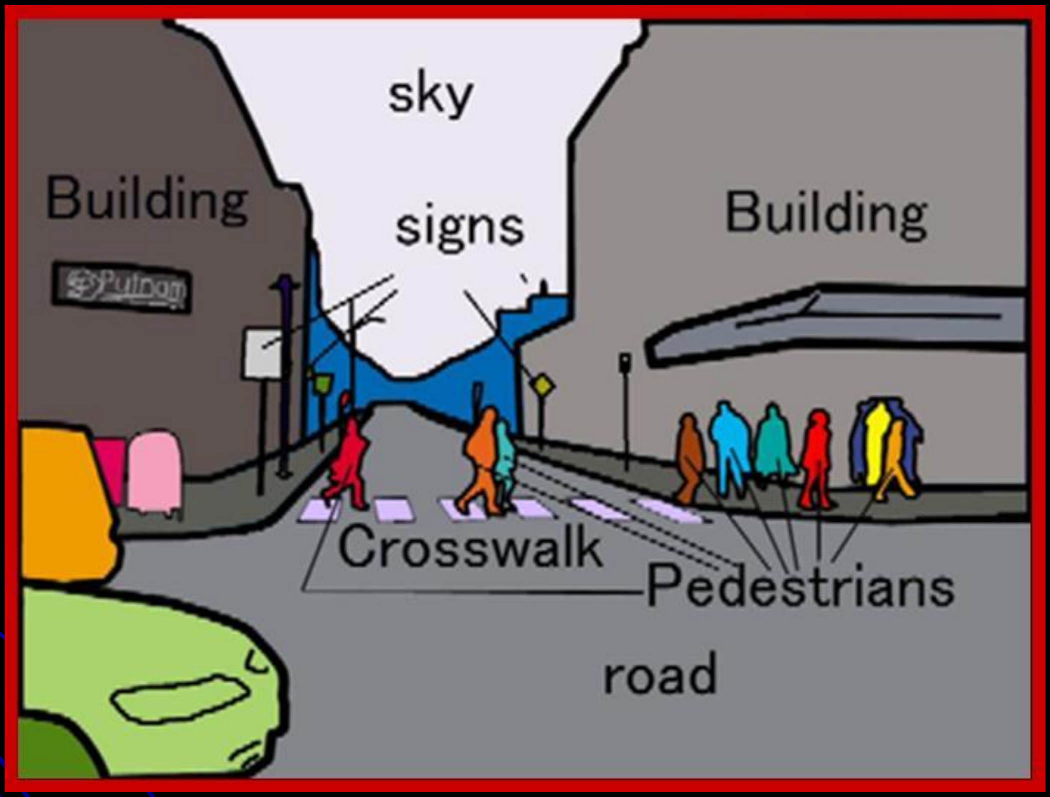
Optical character recognition (OCR)



Biometrics

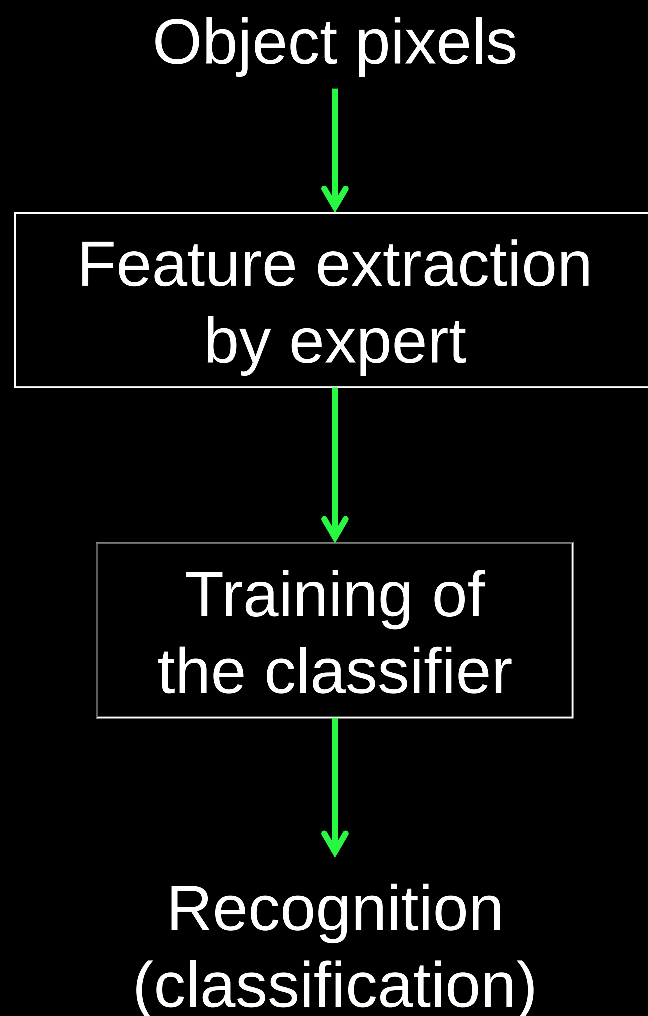


Scene segmentation



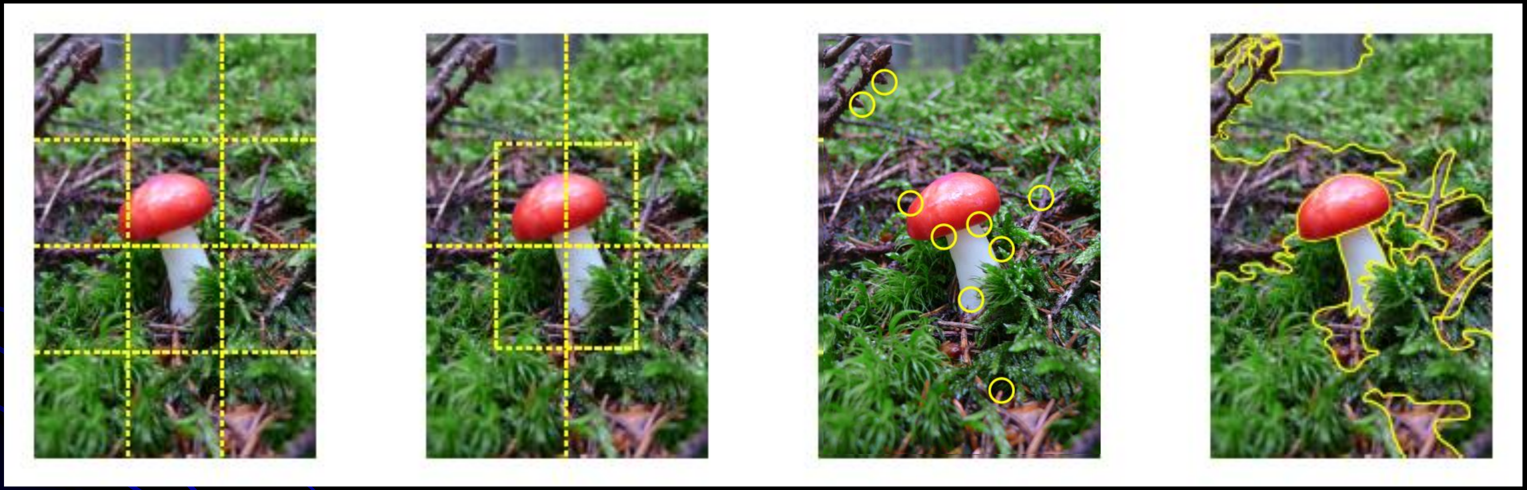
Object recognition

Traditional approach



Object

A segmented area



12 5 9 17

# of segmented areas

Features

Measurements quantifying some object properties

Grouped to feature vectors

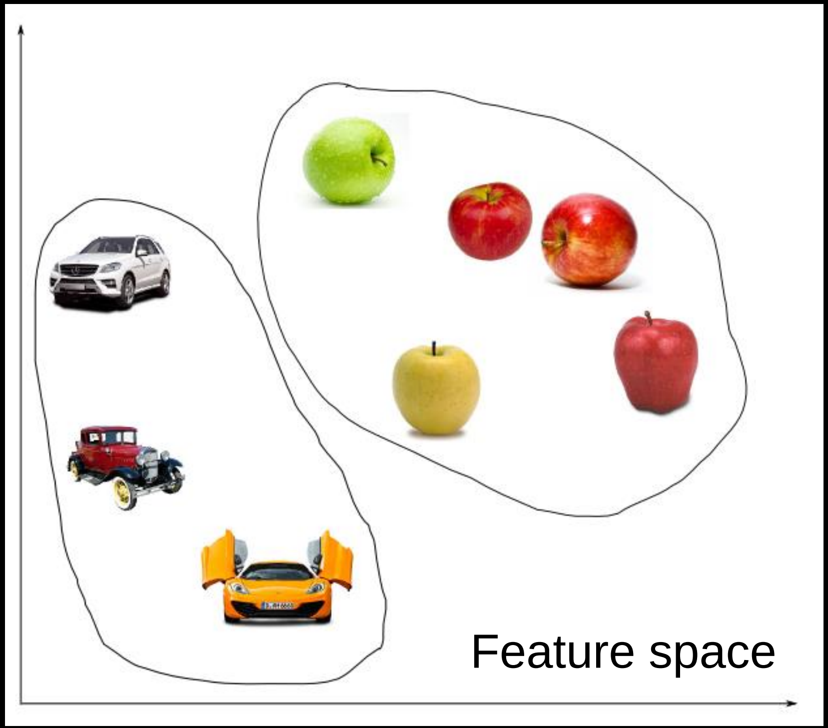
Domain expert knowledge

Feature vector = object descriptor

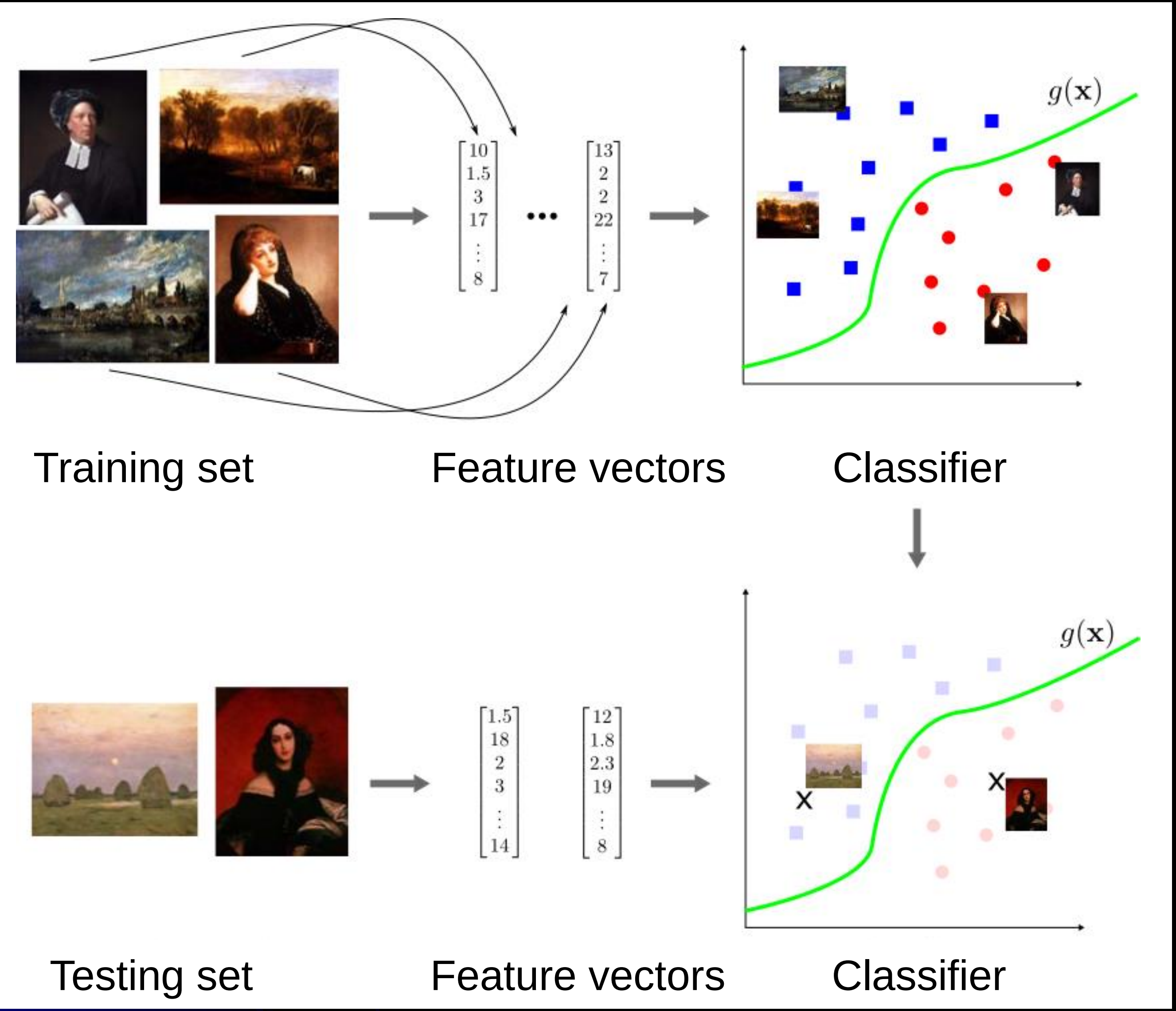
Invariant

Discriminative

Compact



Object (class) recognition



Feature-based classification

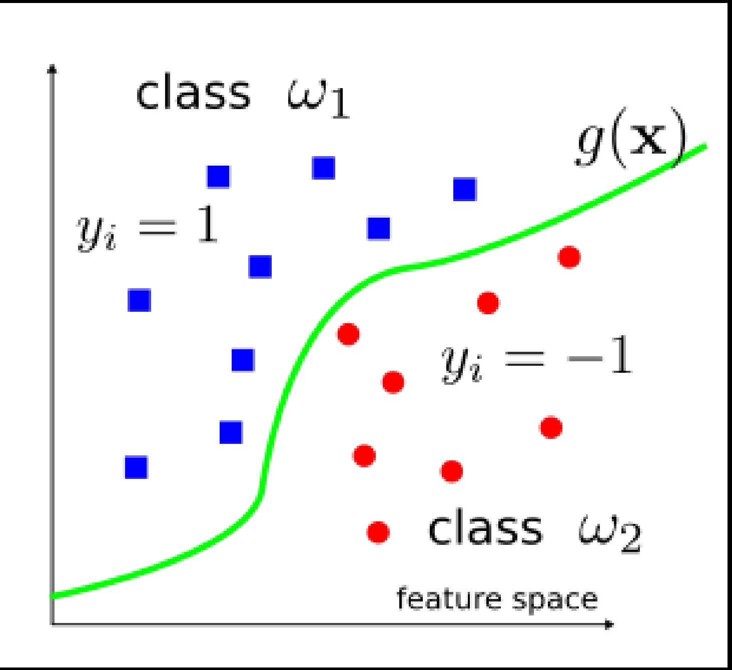
Statistics	Bayesian decision theory
Rules	Decision trees
Metrics (distance)	Nearest neighbour techniques Discriminant analysis Support vector machines
Biologically inspired	Neural networks

Supervised classification

Training set

N observations
$(\mathbf{x}_1, \cdots \mathbf{x}_N), \mathbf{x}_i \in \mathrm{R}^d$

Correct classification
$(\mathrm{y}_1, \cdots \mathrm{y}_N), \mathrm{y}_i \in \{-1,1\}$



Classification problem:

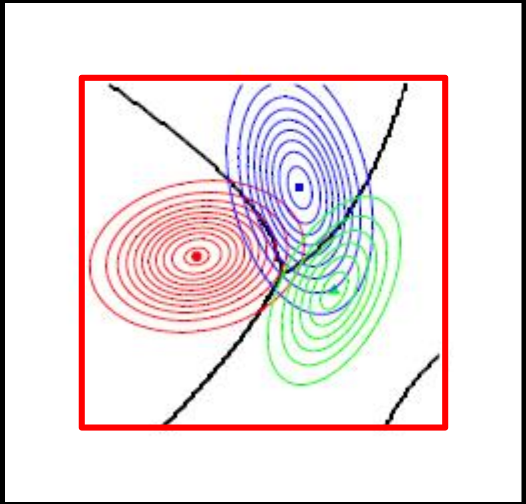
find $f(\mathbf{x})$ s.t. $f(\mathbf{x}_i)=y_i$

determine $g(\mathbf{x})$ from $f(\mathbf{x})$

Naïve Bayes classifier

Bayes rule $$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})}$$

$f(\mathbf{x}) = \omega_i,$ where $i = \arg\max_j P(\mathbf{x}|\omega_j)P(\omega_j)$



Fake smile recognition

Features:

std, max, min, mean

Accuracy 91.3%

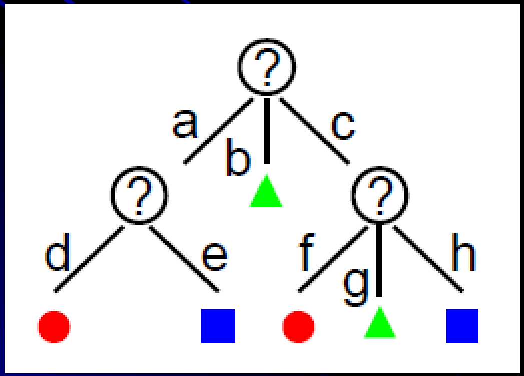
Dibeklioglu, H. et al. 2010. Eyes do not lie: spontaneous versus posed smiles. MM '10.

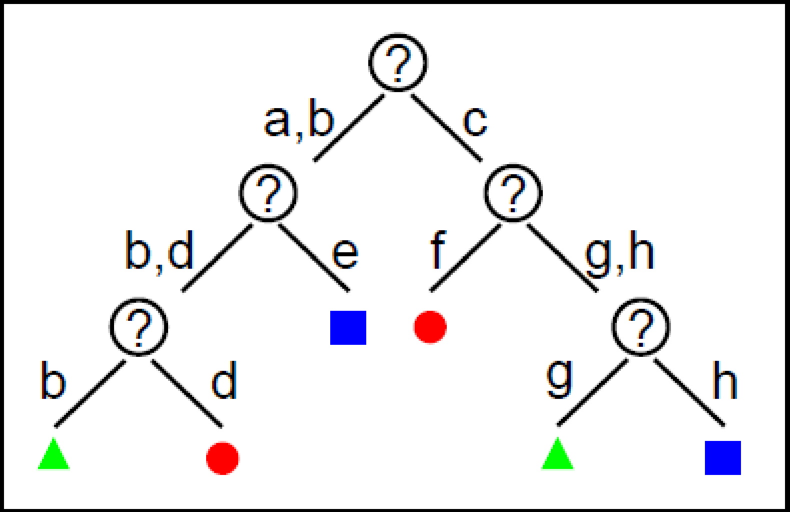
Decision trees

Nominal data – no interpretation of distance

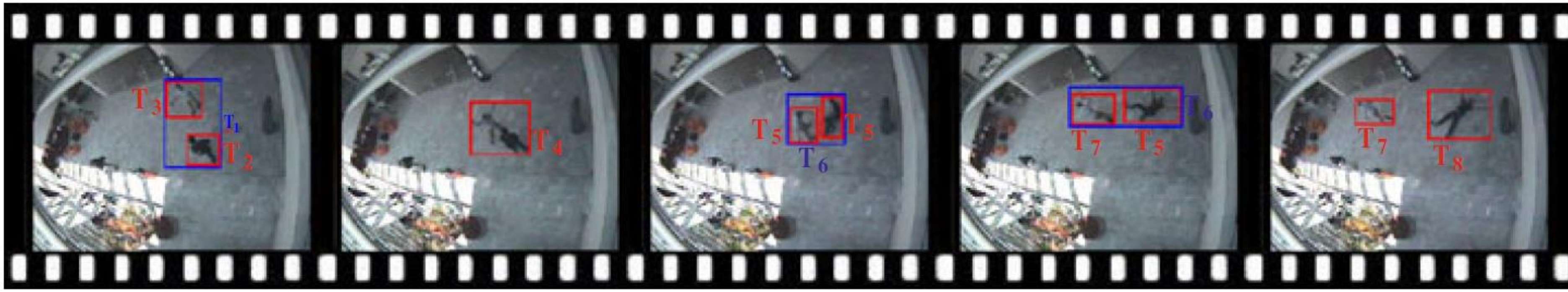
Rules

Tree:
node = test, branches = possible outcomes
leaf = object class





Action recognition in video



Labels definition:

T_1: Two person are approaching
T_2: Person quickly going up
T_3: Person slowly moving in the middle
T_4: Blob quite big with great agitation
T_5: Agitated Persons on the center right of the scene
T_6: Two persons very close and quite agitated
T_7: Person quickly moving to the right
T_8: Person not moving

Questions asked by a decision tree for this sequence:

- Does T_6 exist in the sequence? ✓
- Is there T_1 before T_6? ✓
- Is there T_5 at the same time as T_6? ✓
- Is there T_8 after T_5? ✓
- Is there T_7 at the same time as T_8? ✓

Conclusion:

There might be a fight in this sequence

Various scene
Accuracy 70% – 100%

Simon, C. et al.. 2010. Visual event recognition using decision trees. *MTA,* 50, 1, 95-121.

K nearest neighbours

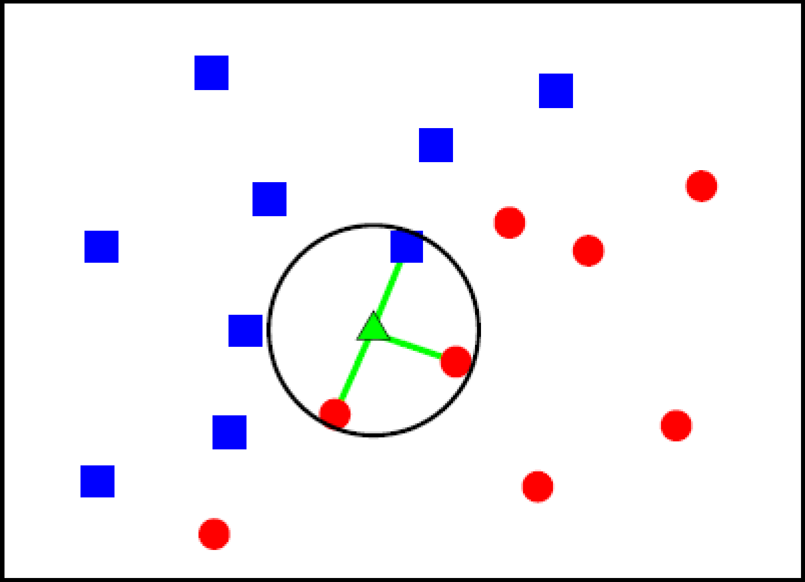
K=1

$$f(\mathbf{x}) = y_i, \qquad \text{where } i = \arg\min_j \|\mathbf{x}_j - \mathbf{x}\|$$

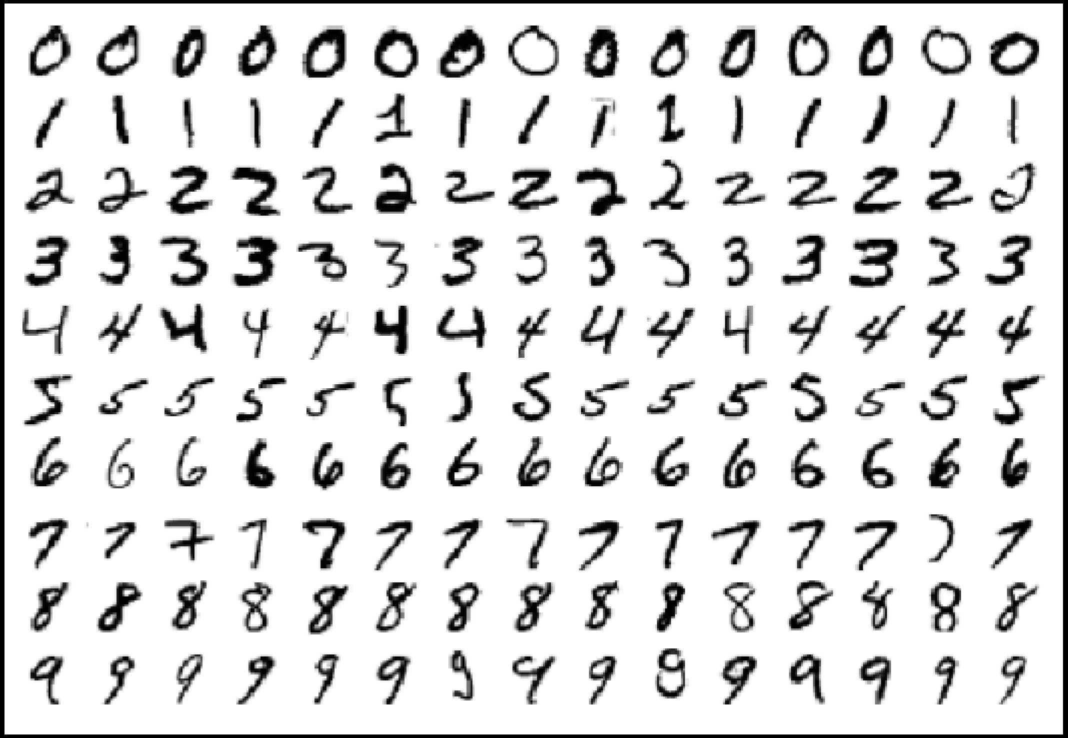
K>1

majority classification

K = 3



Written sign recognition



Accuracy

Numbers 99%

Capital letters 94%

Small letters 89%

Pérez-Cortes J.C., et al, 2000 *Fast and Accurate Handwritten Character Recognition Using Approximate Nearest Neighbours Search on Large Databases*. SSPR /SPR 2000

Linear classifier

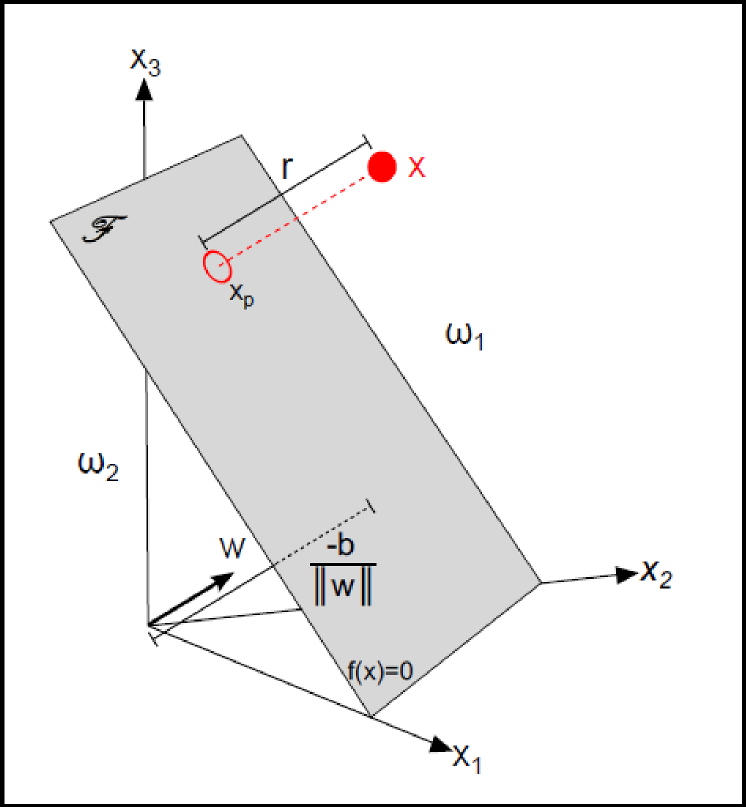
Hyperplane $\mathbf{w}^T\mathbf{x} + b = 0$

$\mathbf{w}^T\mathbf{x} + b \geq 0$ for $\mathbf{x} \in \omega_1$
$\mathbf{w}^T\mathbf{x} + b < 0$ for $\mathbf{x} \in \omega_2$

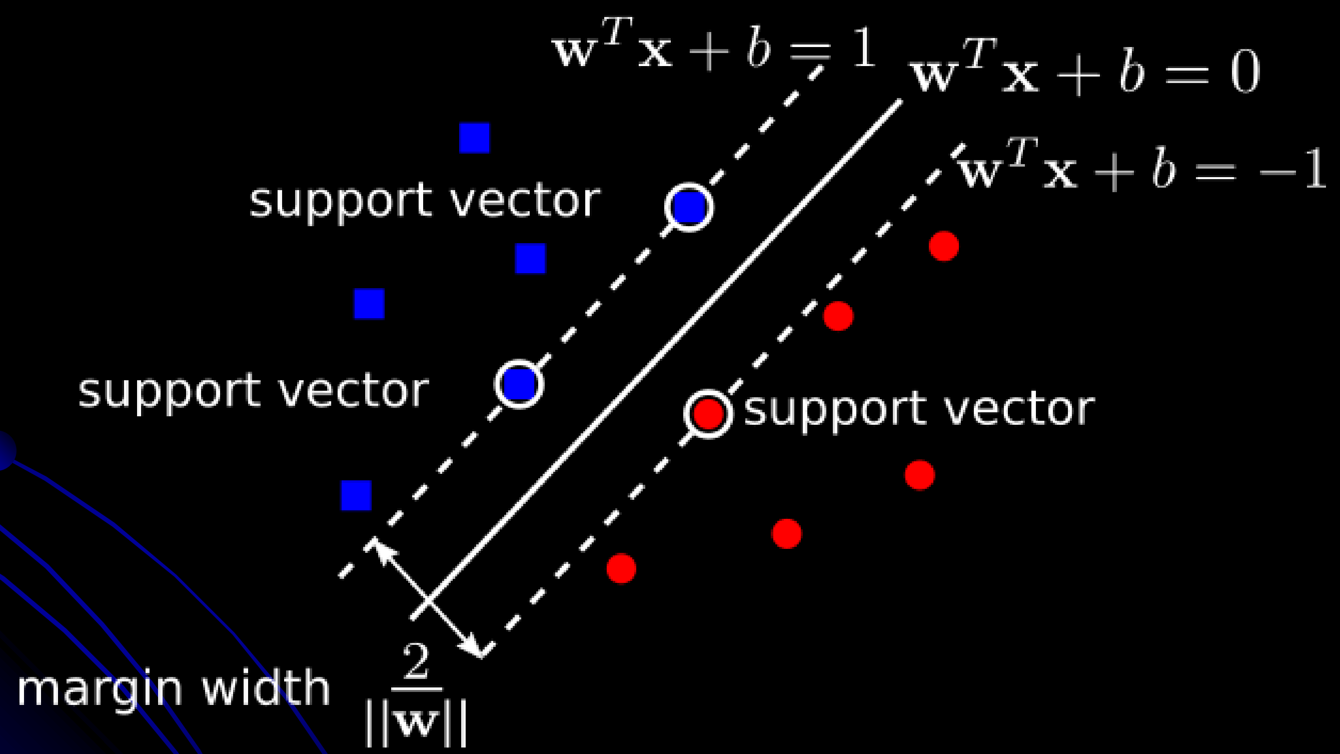
$$\mathbf{u} = [b, \mathbf{w}^T]$$
$$\mathbf{z} = [1, \mathbf{x}^T]^T$$

$$\mathbf{u}^T\mathbf{z} = 0$$

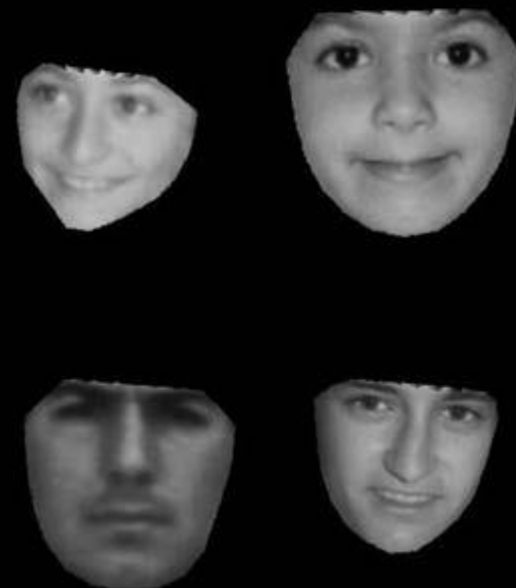
$$\mathbf{u}_{i+1} = \mathbf{u}_i - \alpha(i)\nabla O|_{\mathbf{u}_i}$$



Support vector machines



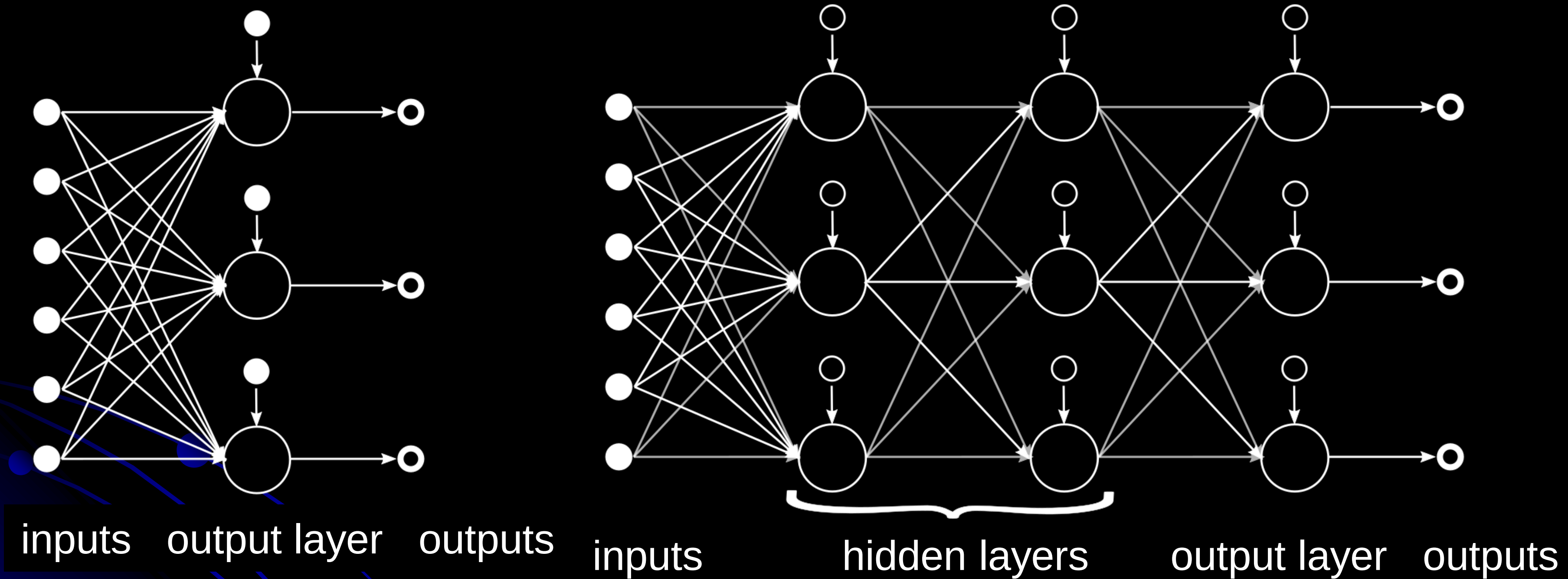
Person and age identification



Method	# of face pairs	Person identification	Age identification
Human (Color/Hair)	100	77.8 %	75.7 %
Human (Gray/Cut-out)	100	66.9 %	63.8 %
SVM	100	80.0 %	75.5 %

Lanitis, A., 2008 *Evaluating the performance of face-aging algorithms*, *ICAFGR2008*, 1-6

Artificial neural networks



Feed-forward neural networks

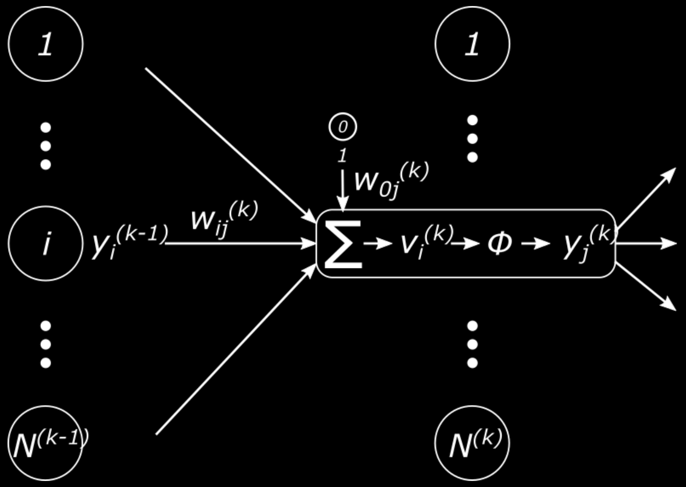
Outputs of neurons in the previous layer are weighted and summed

$$v_j^{(k)} = \sum_{i=1}^{N^{(k-1)}} w_{ij}^{(k)} y_i^{(k-1)} + w_{0j}^{(k)}$$

Non-linear activation function is applied

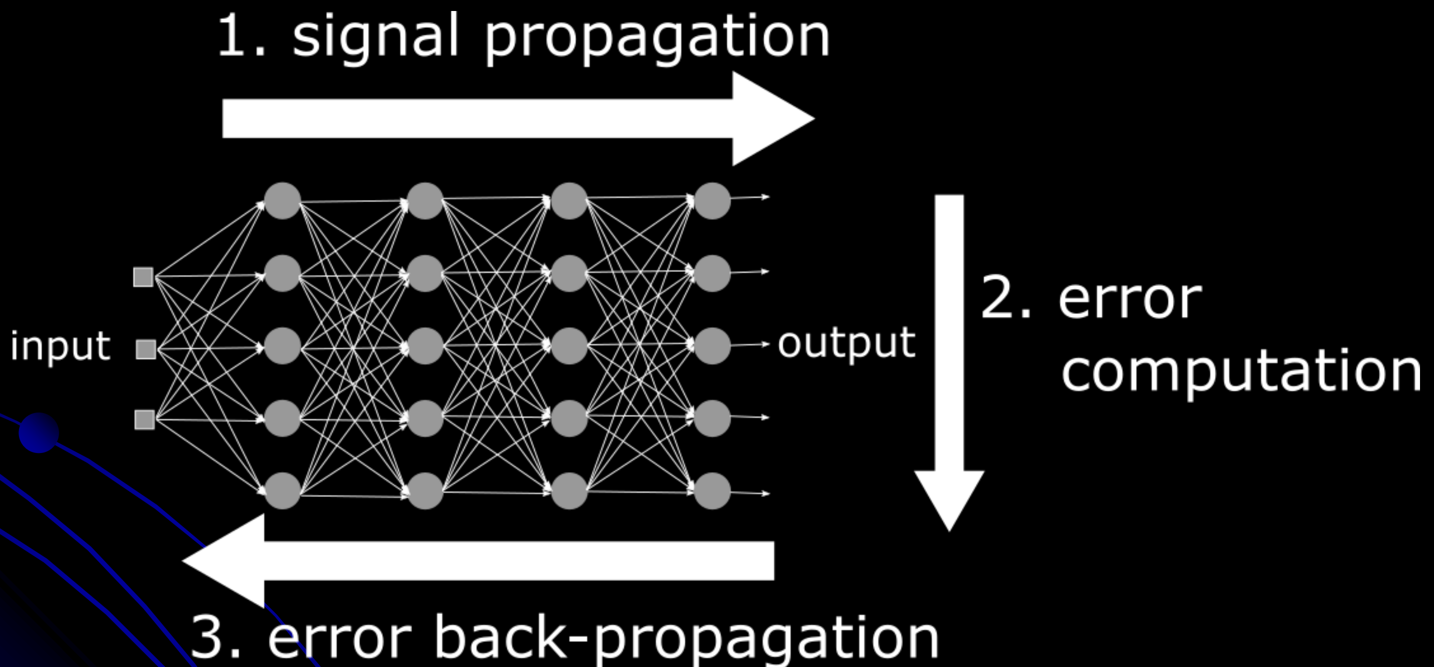
$$y_j^{(k)} = \phi^{(k)} \left(v_j^{(k)} \right)$$

Output goes to the next layer



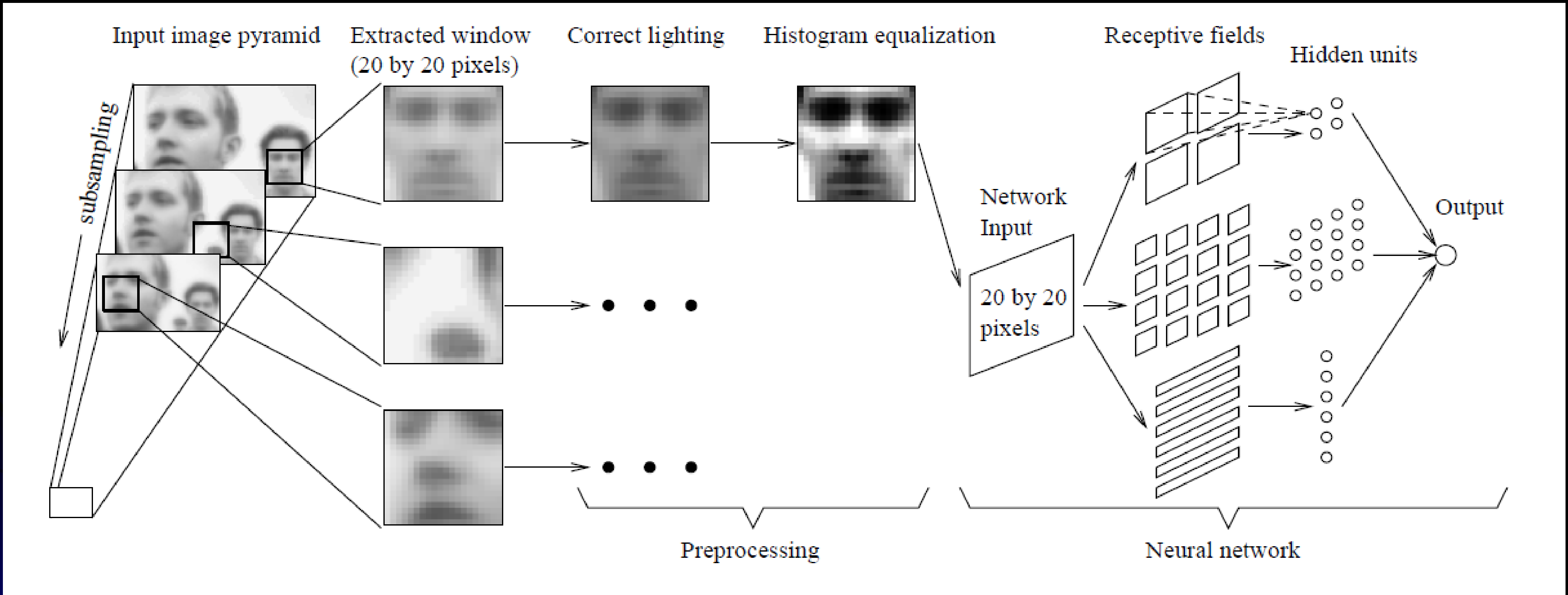
Network training

Weight matrix (weight vectors in all layers)



$$\mathbf{W} \leftarrow \mathbf{W} + \Delta\mathbf{W}, \qquad \text{where } \Delta\mathbf{W} = -\alpha \nabla E|_{\mathbf{W}}$$

Face detection



90% accuracy

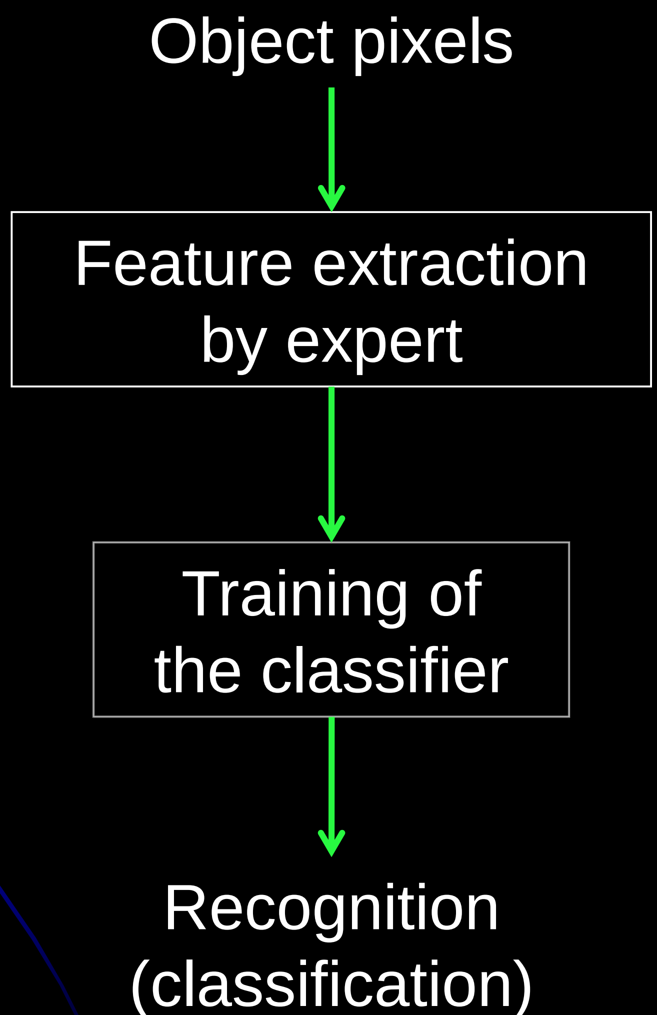
Rowley, H. A., et al. 1998, *Neural network-based face detection*, PAMI, vol. 20, no. 1, pp. 23-38

Identification of portraits

2 hidden layers
8 geometrical features
80% accuracy

Šikudová E. et al., 2006, *Extracting semantic information from art images*. CVG Computational Imaging and Vision, vol 32. Springer

Traditional approach



Deep learning

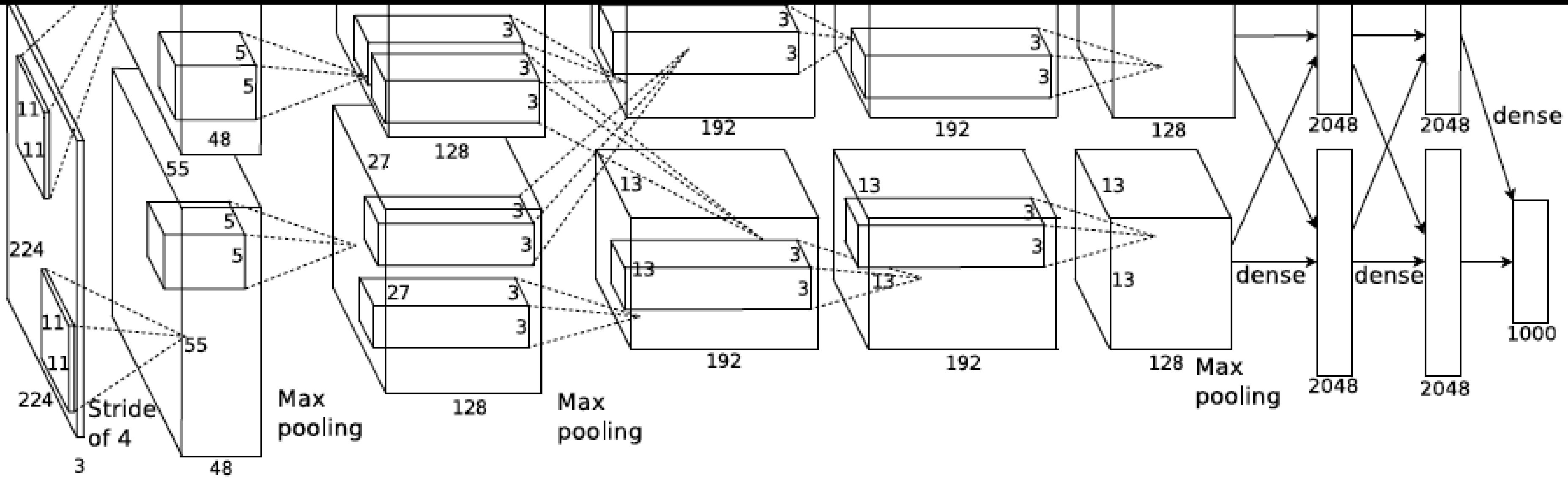
Object pixels

Features learned during training

Recognition (classification)

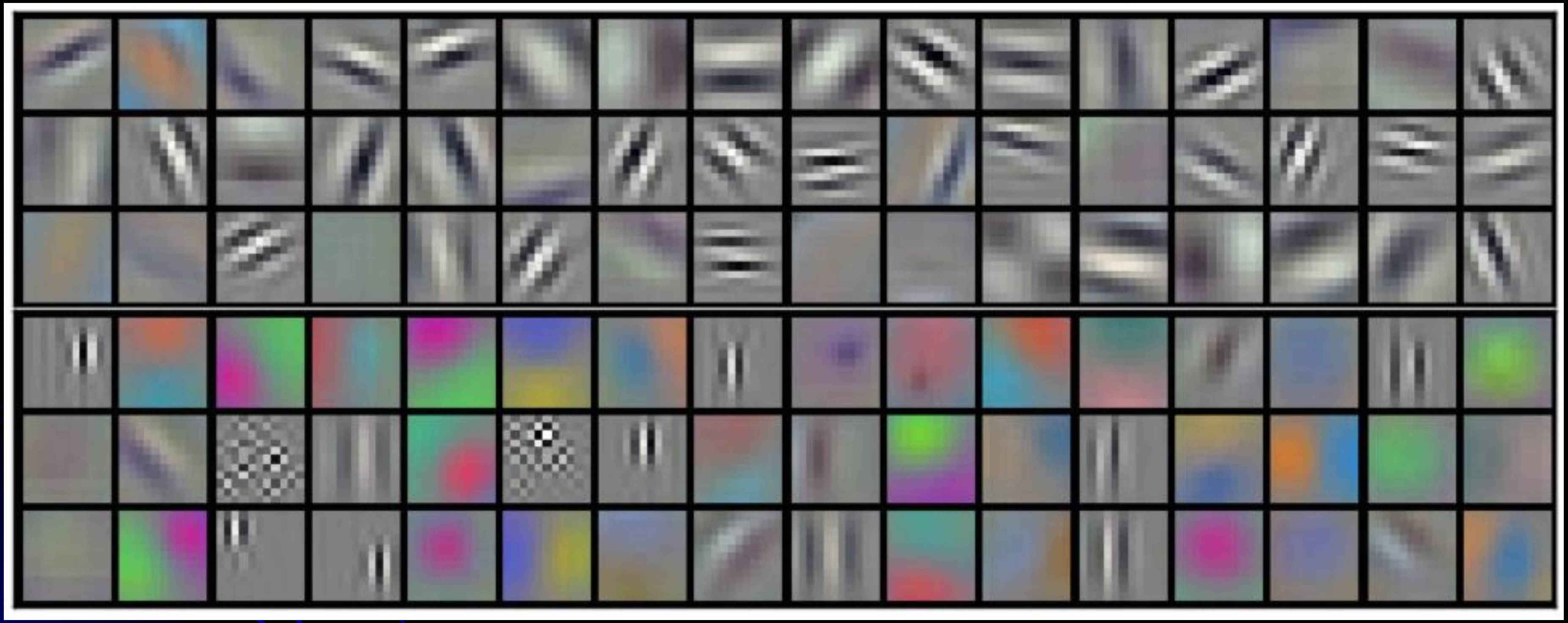
AlexNet

Convolutional network, 7 hidden layers,
650 000 neurons 60 000 000 parameters
Trained on 2 GPU for a week



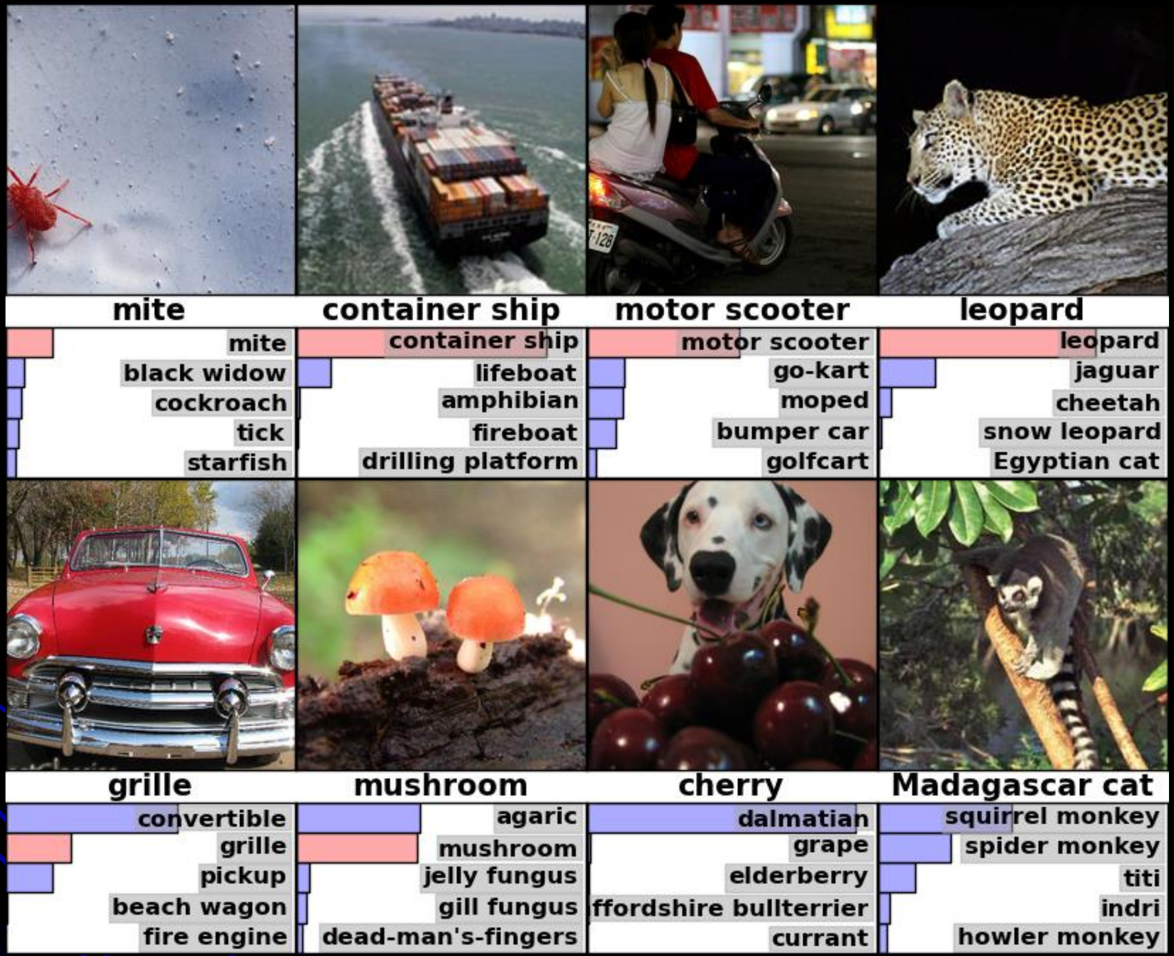
Krizhevsky, A., et al. 2012, ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012

AlexNet Features



1. layer: 96 convolutional kernels

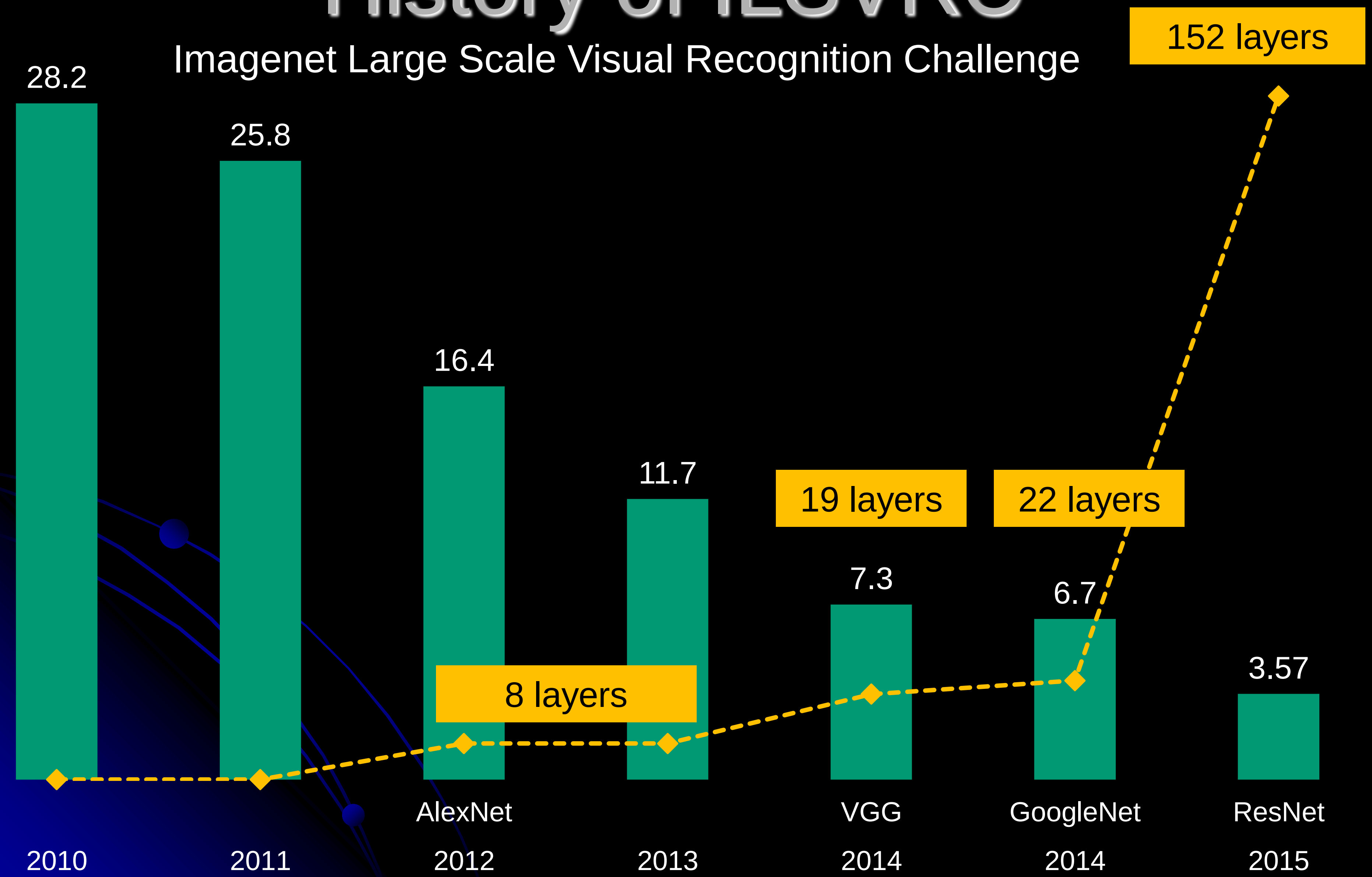
AlexNet results



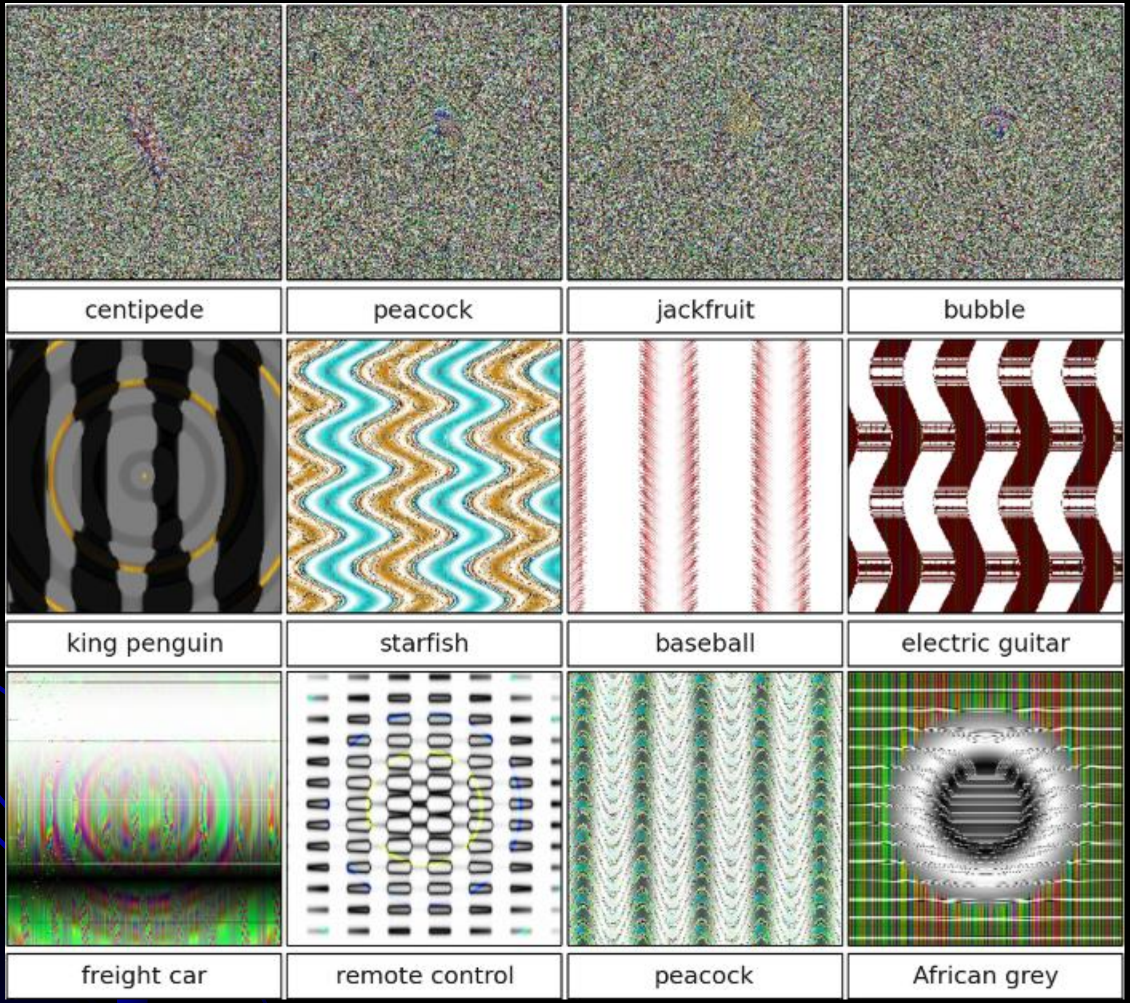
top 5 error 16,4%

History of ILSVRC

Imagenet Large Scale Visual Recognition Challenge

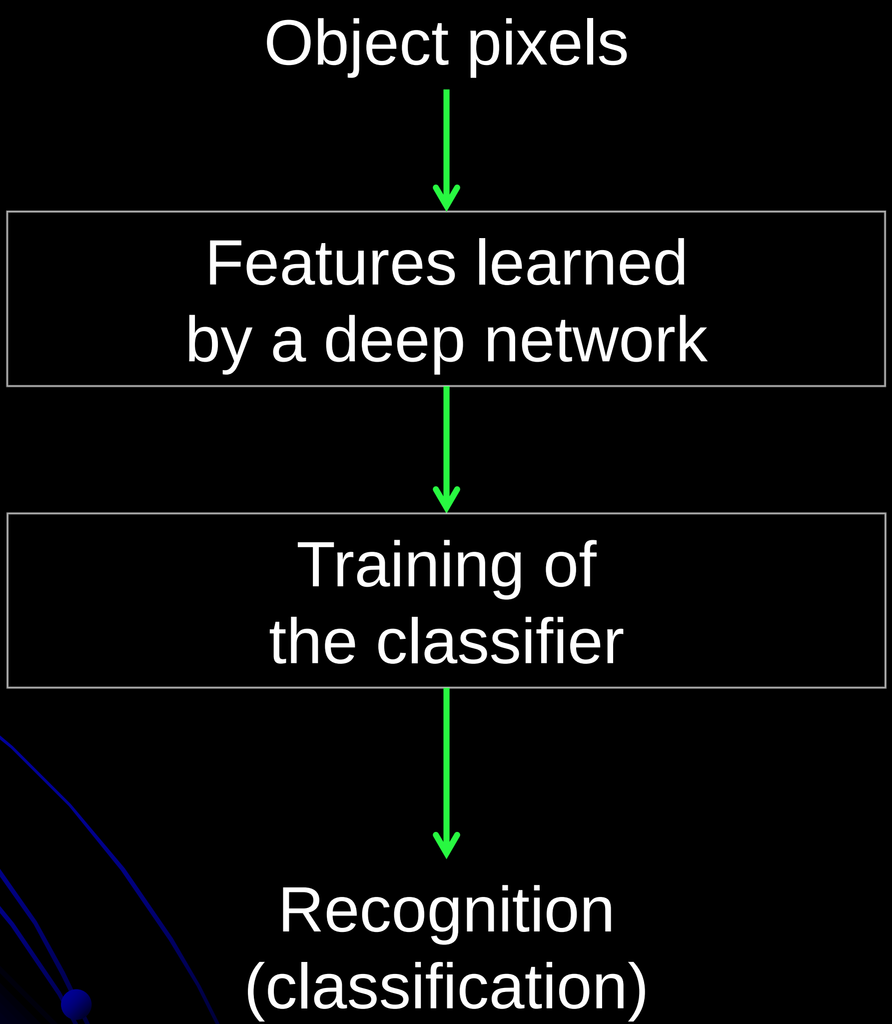


Wrong classification



Nguyen, A. Et al. 2015, *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images,* CVPR '15

Combined approach



Combined approach

OverFeat network (NYU), no additional training + SVM

Task	**Database**	**mAP**
Image classification	Pascal VOC 2007	73.9%
Scene classification	MIT-67	58.4%
Fine grained recognition	Caltech-UCSD Birds 200-2011	53.3%
	Oxford 102 Flowers	74.70%

Razavian et al, 2014, *CNN Features Off-the-Shelf: An Astounding Baseline for Recognition*, CVPRW '14, 512-519

Combined approach

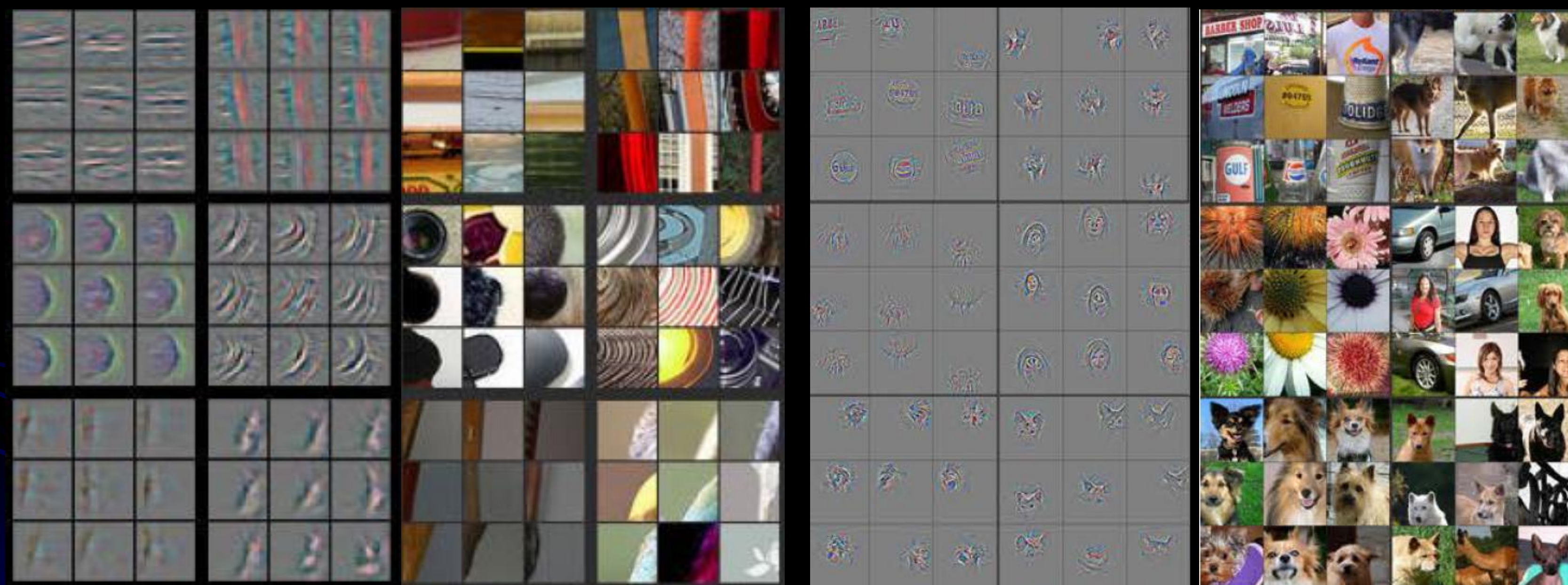
AlexNet + SVM

Layer	Accuracy	
	Caltech 101	Caltech 256
SVM (1)	44.8%	24.6%
SVM (2)	66.2%	39.6%
SVM (3)	72.3%	46.0%
SVM (4)	76.6%	51.3%
SVM (5)	86.2%	65.6%
SVM (7)	85.5%	71.7%

Zeiler, M. D. et al. 2014, *Visualizing and Understanding Convolutional Networks*, ECCV2014, pp. 818–833

Combined approach

AlexNet + SVM



response image
3. layer

response image
5. layer

Zeiler, M. D. et al. 2014, *Visualizing and Understanding Convolutional Networks*, ECCV2014, pp. 818–833

Other covered topics

- Feature selection and preprocessing
- Unsupervised learning
- Hidden Markov models
- Classification quality evaluation

Conditions

written exam
project

- at least 50% from each

Bibliography

Christopher M. Bishop

Pattern Recognition and Machine Learning

Richard O. Duda, Peter E. Hart, David G. Stork

Pattern classification

Trevor Hastie, Robert J. Tibshirani, J. Jerome H. Friedman

The elements of statistical learning

Šikudová a kol.

Počítačové videnie: detekcia a rozpoznávanie objektov

Takehome message

No algorithm solves all the problems

Machine learning is a methodology, not a tool

Domain knowledge

https://xkcd.com/1838/