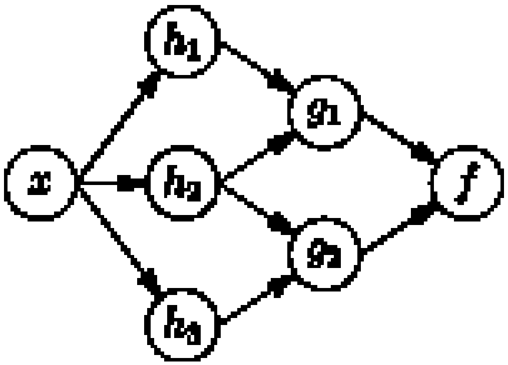
Neural networks in modern image processing

Petra Budíková

DISA seminar, 30. 9. 2014

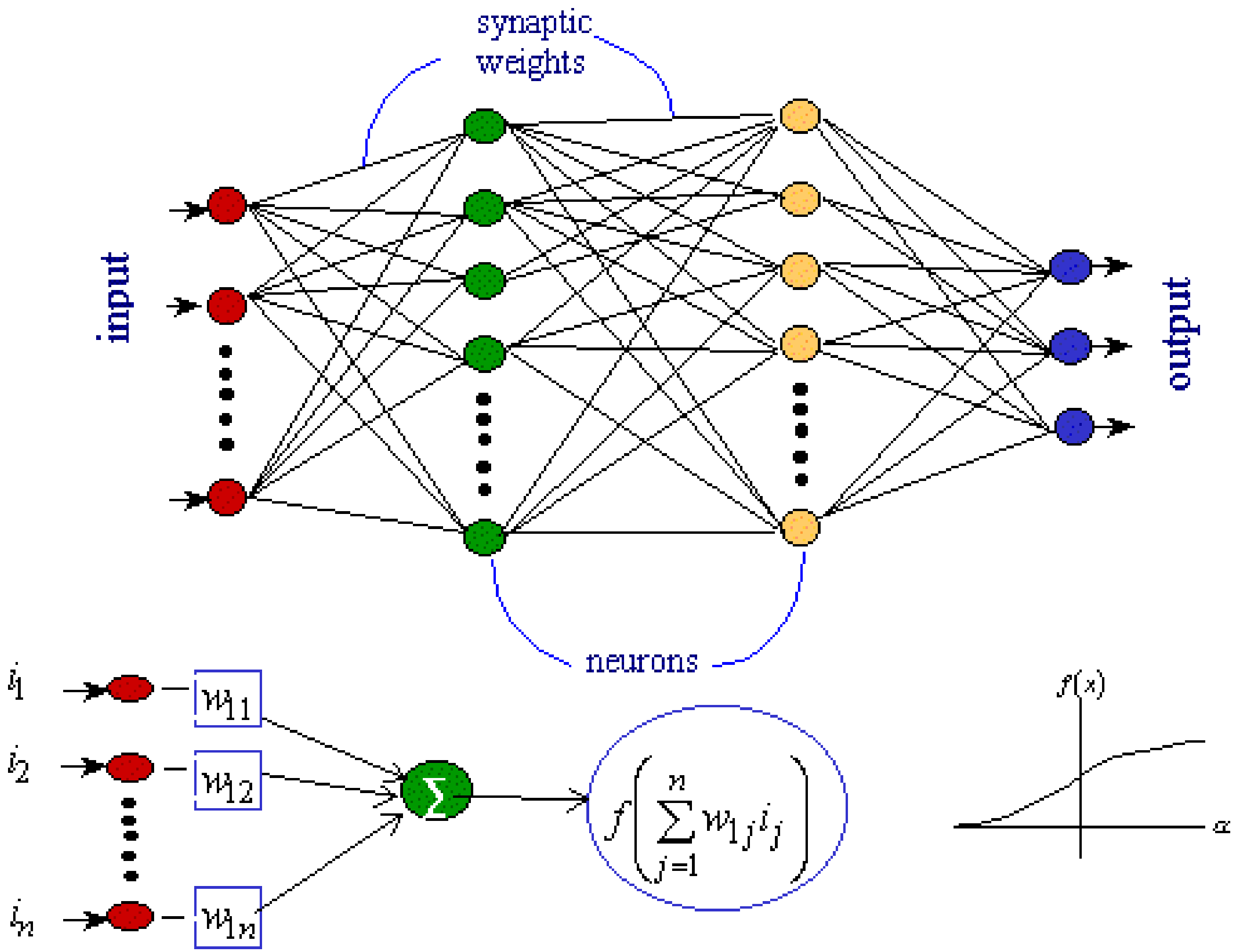
Artificial neural networks

- Computational models inspired by an animal's central nervous systems
- Systems of interconnected "neurons" which can compute values from inputs
- Are capable of approximating non-linear functions of their inputs
  - Mathematically, a neuron's network function $f(x)$ is defined as a composition of other functions $g_i(x)$, which can further be defined as a composition of other functions.



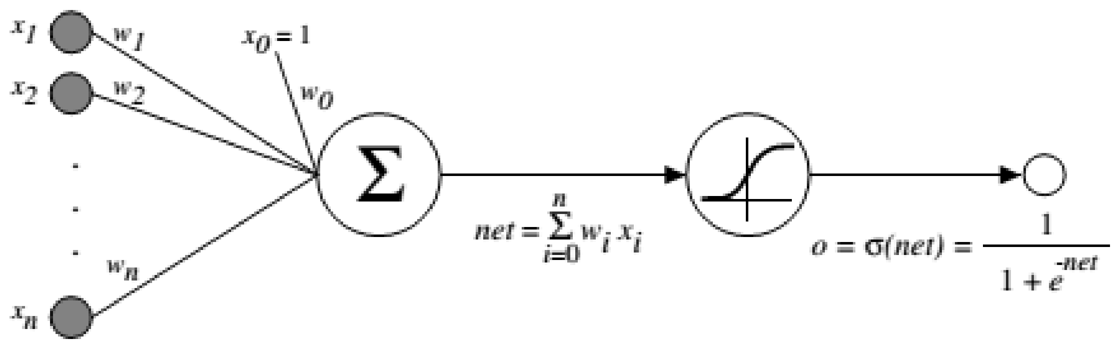
- Known since 1950s
- Typical applications: pattern recognition in speech or images

Artificial neural networks (cont.)



Artificial neural networks (cont.)

- Network node in detail

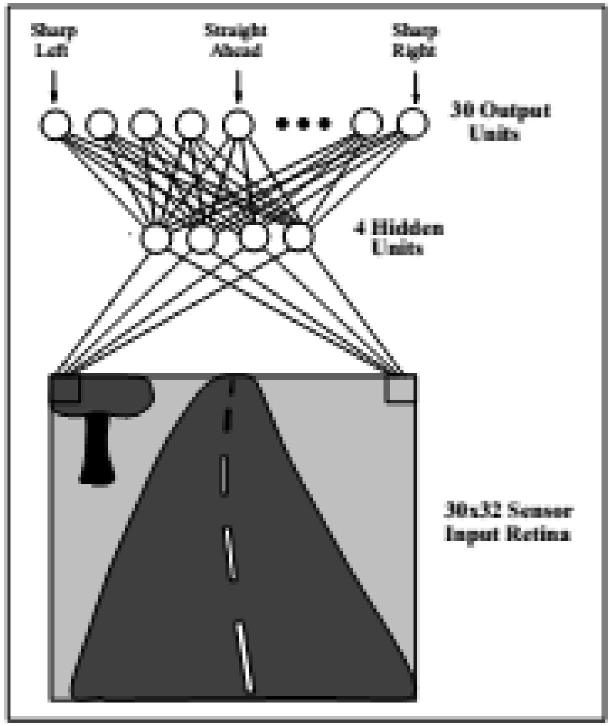


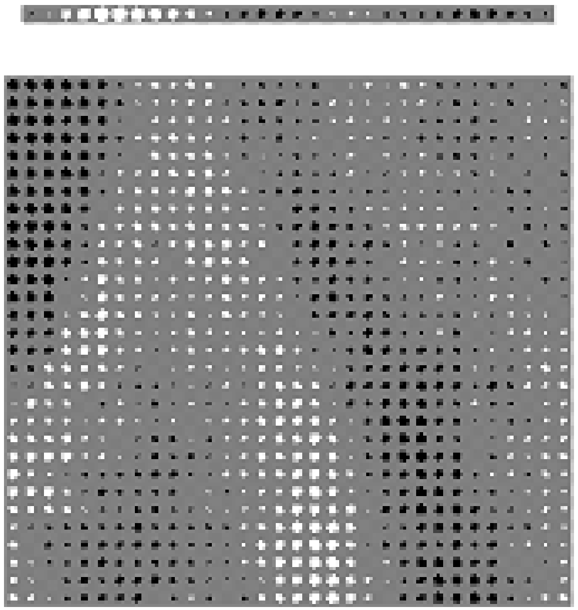
$\sigma(x)$ is the sigmoid function

- Network learning process = tuning the synaptic weights
  - Initialize randomly
  - Repeatedly compute the ANN result for a given task, compare with ground truth, update ANN weights by *backpropagation* algorithm to improve ANN performance

Artificial neural networks – example

- ALVINN system for automatic car driving (ANN illustration form [Mitchell97])



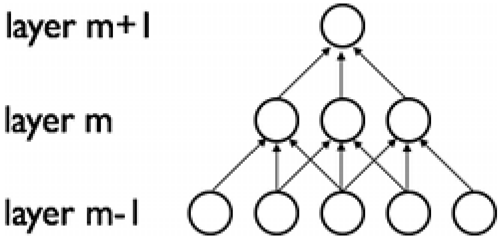


Neural networks before 2009(+-) and after

- Before 2009: ANNs typically with 2-3 layers
  - Reason 1: computation times
  - Reason 2: problems of the backpropagation algorithm
    - Local optimization only (needs a good initialization, or re-initialization)
    - Prone to over-fitting (too many parameters to estimate, too few labeled examples)
  - => Skepticism: A deep network often performed worse than a shallow one
- After 2009: Deep neural networks
  - Fast GPU-based implementations
  - Weights can be initialized better (Use of unlabeled data, Restricted Boltzmann Machines)
  - Large collections of labeled data available
  - Reducing the number of parameters by weight sharing
  - Improved backpropagation algorithm
  - Success in different areas, e.g. traffic sign recognition, handwritten digits problem

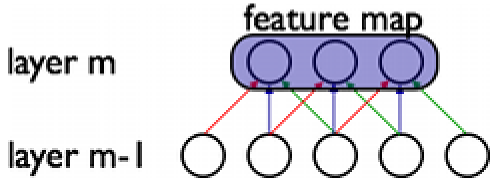
Convolutional neural networks

- A type of feed-forward ANN where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field
  - Inspired by biological processes
  - Widely used for image recognition
- Multiple layers of small neuron collections which look at small portions of the input image
  - The input hidden units in the m-th layer are connected to a local subset of units in the (m-1)-th layer, which have spatially contiguous receptive fields



Convolutional neural networks

- Shared weights: each sparse filter h_i is replicated across the entire visual field. The replicated units form a feature map, which share the same parametrization, i.e. the same weight vector and the same bias.
  - Weights of the same color are shared, i.e. are constrained to be identical
  - Replicating units allows for features to be detected regardless of their position in the visual field.

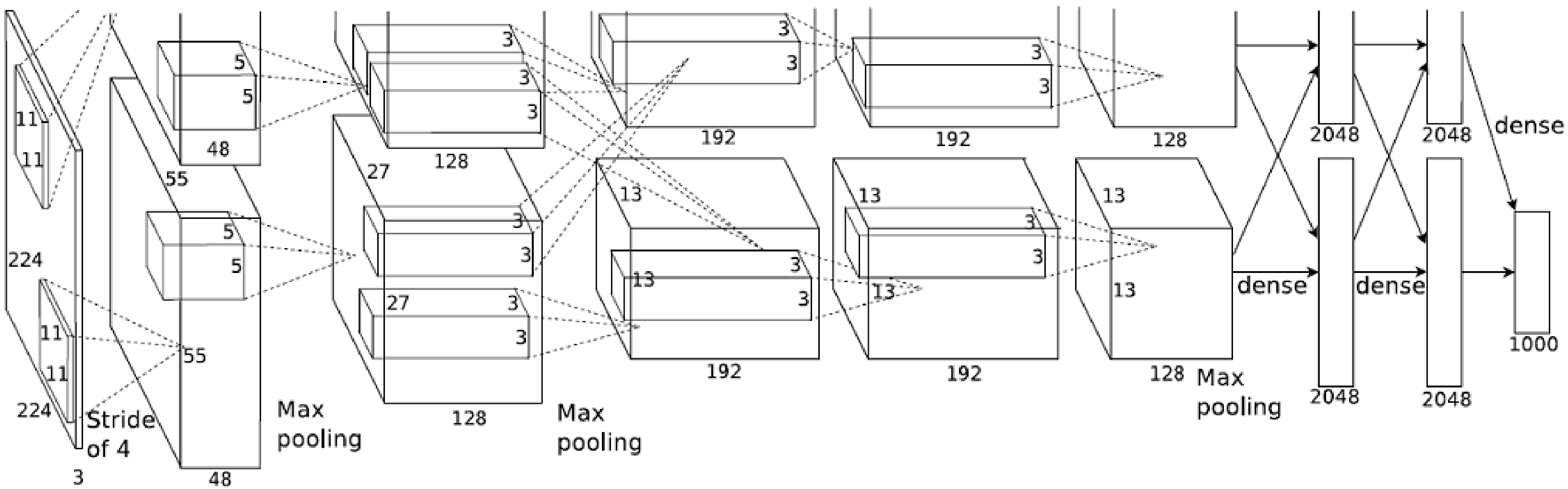


  - Weight sharing greatly reduces the number of free parameters to learn.
- MaxPooling: another important concept of CNNs
  - non-linear down-sampling – the input image is partitioned into a set of non-overlapping rectangles and maximum value is taken for each such sub-region
  - Advantages:
    - It reduces the computational complexity for upper layers
    - It provides a form of translation invariance

Krizhevsky 2012: ImageNet neural network

- The ImageNet challenge: recognize 1000 image categories
  - Training data: 1.2M manually cleaned training images (obtained by crowdsourcing)
- Krizhevsky solution: deep convolutional neural network
  - 5 convolutional layers, 3 fully connected layers
  - 60 million parameters and 650,000 neurons
  - New function for nodes (Rectified Linear Units)
  - Efficient GPU implementation of NN learning, highly-optimized implementation of 2D convolution
  - Data augmentation
    - generating image translations and horizontal reflections
    - five 224 × 224 patches (the four corner patches and the center patch) as well as their horizontal reflection from each 256×256 image
    - => transformation invariance, reduces overfitting
  - Additional refinements such as the “dropout” regularization method

Krizhevsky 2012 (cont.)



- Great success!!!

INRIA/Xerox	33%,
Uni Amsterdam	30%,
Uni Oxford	27%,
Uni Tokyo	26%,
Uni Toronto	**16% (deep neural network)** [Krizhevsky-NIPS-2012]

Krizhevsky 2012 – more than just classification?

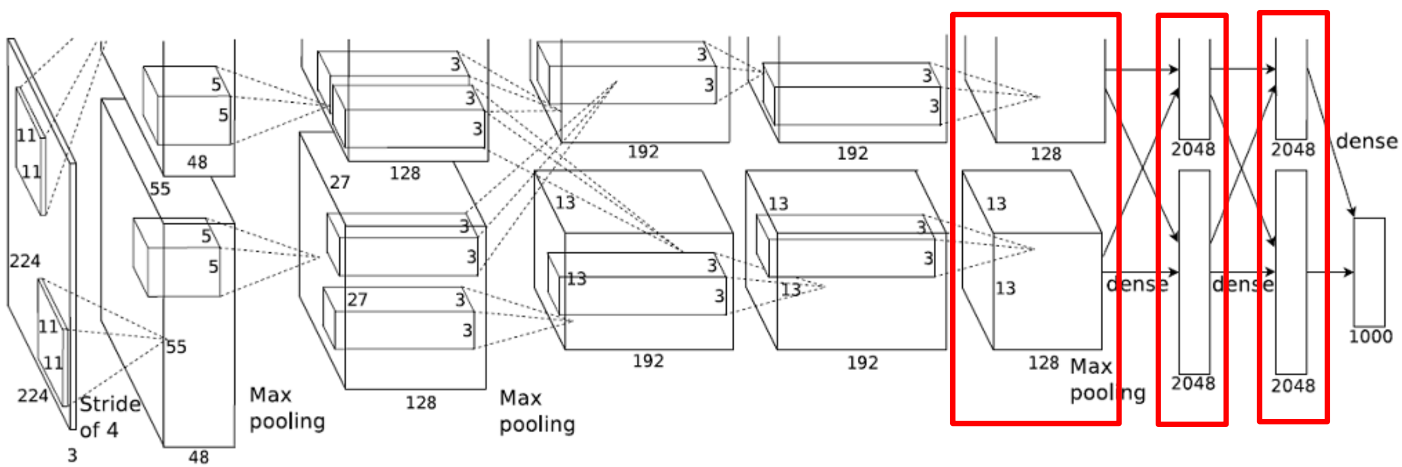
- Indications that the last hidden layers carry semantics!
- Suggestion in [Krizhevsky12]:
  - Responses of the last hidden layer can be used as a compact global image descriptor
  - Semantically similar images should have small Euclidean distance

Convolutional neural network implementations

- cuda-convent
  - Original implementation by Alex Krizhevsky
- decaf
  - Python framework for training neural networks
- Caffe
  - Convolutional Architecture for Fast Feature Embedding
  - Berkeley Vision and Learning Center
  - C++/CUDA framework for deep learning and vision
  - An active research and development community
  - Main advantage in comparison with other implementations: it is FAST
  - Wrappers for Python and MATLAB

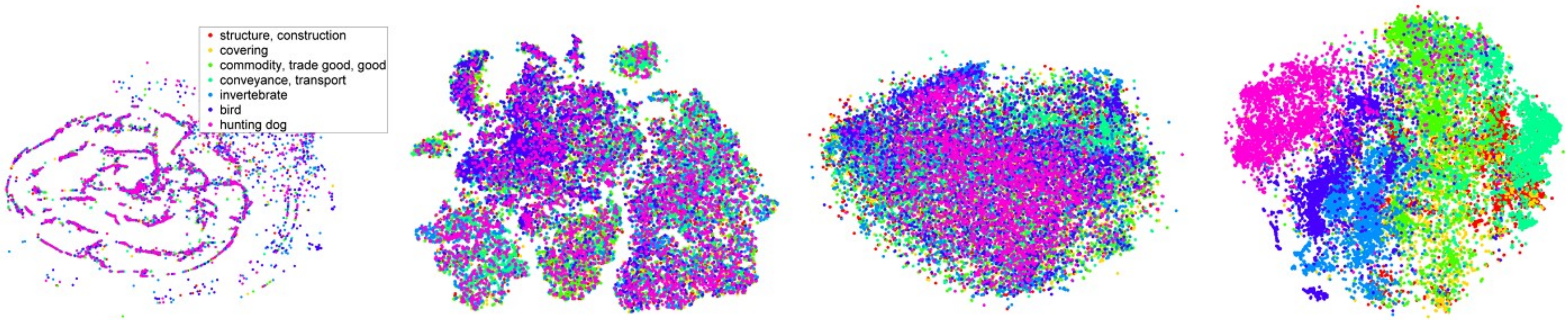
DeCAF

- decaf
  - Python framework for training neural networks
  - Deprecated, replaced by Caffe
- DeCAF
  - Image features derived from neural network trained for the ImageNet competition
  - 3 types: $DeCAF_5$, $DeCAF_6$, $DeCAF_7$
    - Derived from last 3 hidden layers of the ImageNet neural network
    - Descriptor sizes: ??? dimensions for $DeCAF_5$, 4096 dimensions for $DeCAF_6$, $DeCAF_7$



DeCAF (cont.)

- Performance of DeCAF features analyzed in [Donahue14] in context of several image classification tasks
  - DeCAF_5 not so good
  - DeCAF_6 and DeCAF_7 very good, in many cases outperform state-of-the-art descriptors
  - DeCAF_6 typically more successful, but only by small margin



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

Utilization of DeCAF descriptors

- Recognition of new (unseen in ImageNet) categories by training (a linear) classifier on top of the DeCAF descriptors
  - [Donahue14]
  - [Girshick14]
  - Two solutions of ImageCLEF 2014 Scalable Concept Annotation Challenge
  - …
- Very good results reported

Similarity search: MPEG7 vs. $DeCAF_7$

- Similarity search in 20M images; 1st image is the query

MPEG7 descriptors

$DeCAF_7$ descriptors

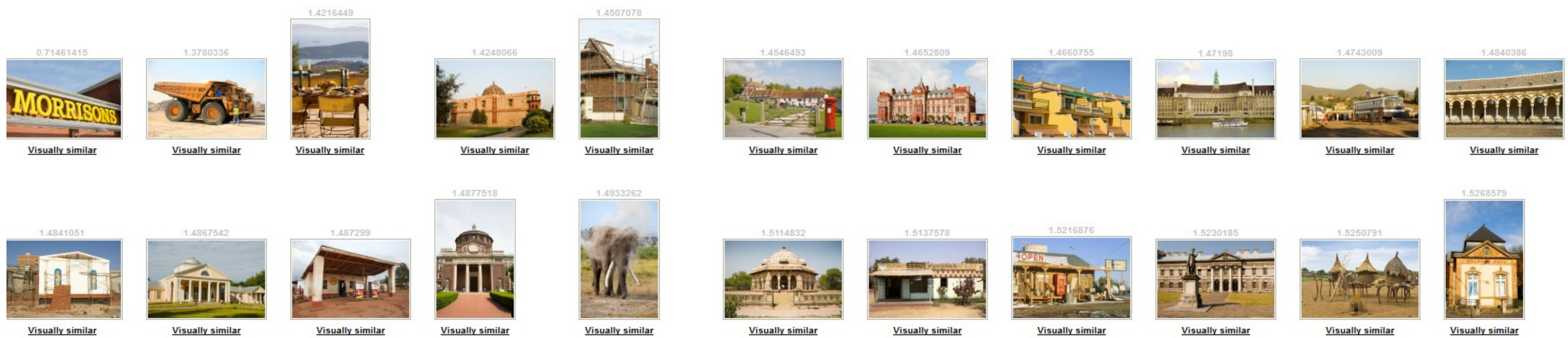
Similarity search: MPEG7 vs. DeCAF$_7$

MPEG7 descriptors

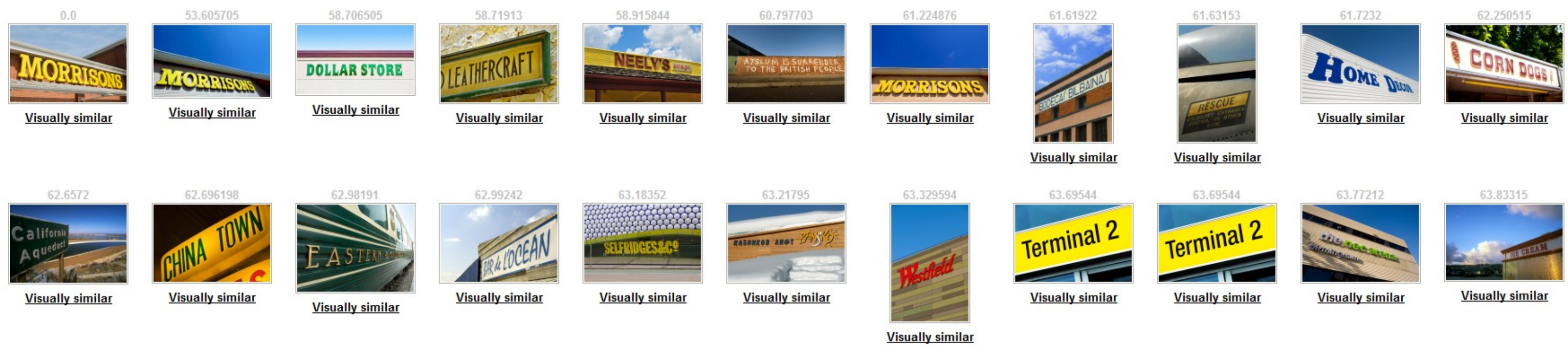
DeCAF$_7$ descriptors

Similarity search: MPEG7 vs. DeCAF$_7$

MPEG7 descriptors



DeCAF$_7$ descriptors



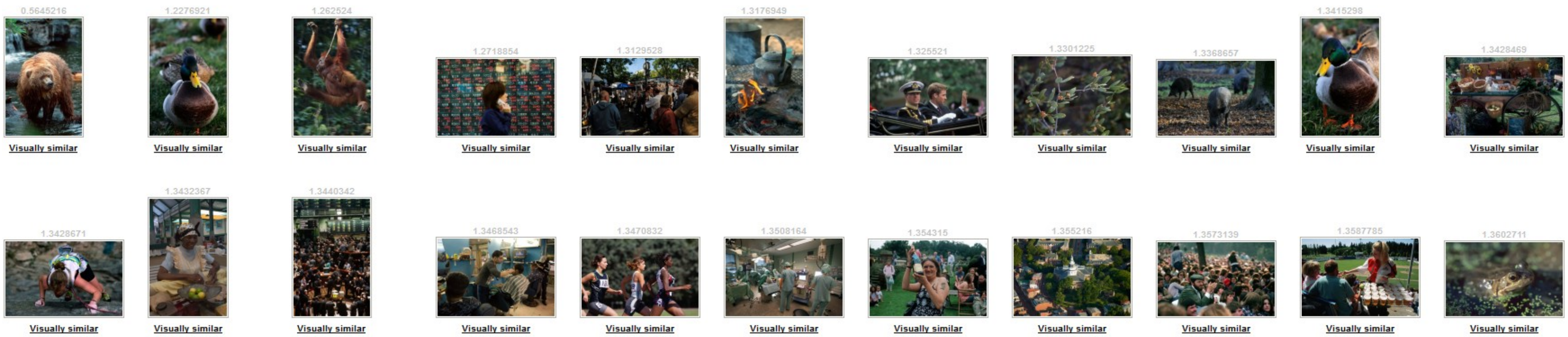
Similarity search: MPEG7 vs. DeCAF$_7$

MPEG7 descriptors

DeCAF$_7$ descriptors

Similarity search: MPEG7 vs. DeCAF$_7$

MPEG7 descriptors



DeCAF$_7$ descriptors

Literature

Books

[Mitchell97] T. Mitchell. Machine Learning. ISBN 978-0070428072. McGraw Hill, 1997.

Research papers

[Donahue14] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. *Decaf: A deep convolutional activation feature for generic visual recognition*. ICML, 2014.

[Girshick14] R. Girshick, J. Donahue, T. Darrell, and J. Malik. *Rich feature hierarchies for accurate object detection and semantic segmentation*. CVPR, 2014.

[Jia14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell. *Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding*. Submitted to ACM MULTIMEDIA 2014 Open Source Software Competition.

[Krizhevsky12] A. Krizhevsky, I. Sutskever, G. E. Hinton: *ImageNet Classification with Deep Convolutional Neural Networks*. NIPS 2012.

Literature (cont.)

Other

- http://caffe.berkeleyvision.org/
- J. Materna: *Deep Learning: budoucnost strojového učení?* http://fulltext.sblog.cz/2013/01/09/deep-learning-budoucnost-strojoveho-uceni/
- J. Čech: *A Shallow Introduction into the Deep Machine Learning*. https://cw.felk.cvut.cz/wiki/_media/courses/ae4m33mpv/deep_learning_mpv.pdf
- Basic explanation of convolutional neural networks principles http://deeplearning.net/tutorial/lenet.html