Meta-learning and Knowledge Extraction from Multimedia

Petr Pulc[1] Martin Holeňa[2]

[1]Faculty of Information Technology,
Czech Technical University
Prague, Czech republic

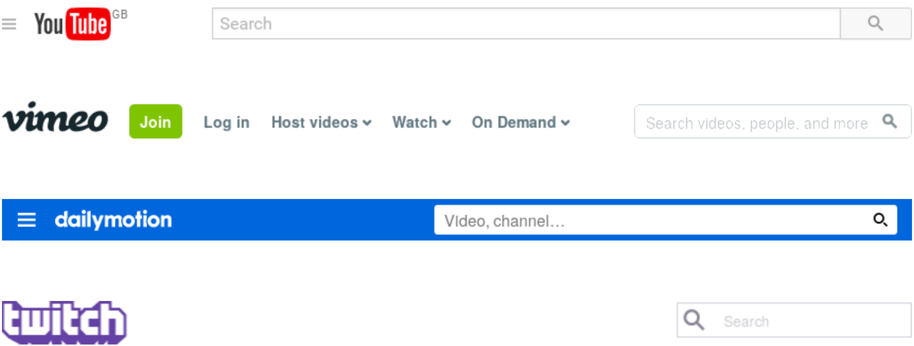
[2]Institute of Computer Science,
Academy of Sciences
Prague, Czech republic

January 5, 2017

Outline

How do you search for multimedia content?

How do you search for multimedia content?



How do you search for multimedia content?

IMG 7206 MOV

PetaPixel

Subscribe 3,698

4,814 views

Add to   Share   ••• More   0   0

Published on 30 Jul 2015

Category	Science & Technology
Licence	Standard YouTube Licence

Annotate the multimedia item . . .

How would someone search?

Alice – General public

- Protest in Turkey
- Gezi Parki
- Crowd walking down the street

How would someone search?

Bob – Director

- 3rd June 2013
- Istiklal Caddesi boulevard
- Protesters walking towards Gezi Parki

How would someone search?

Charlie – Editor

- Static, wide shot
- Around 15:00 local time
- Longer than 5 seconds

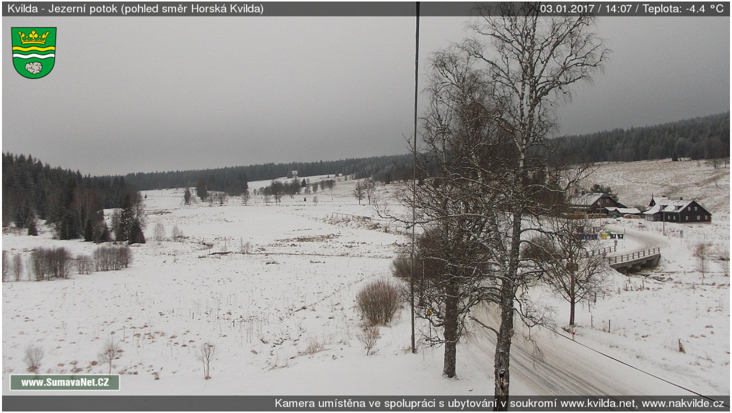
Multimedia content is very diverse . . .

– courtesy of AVC-ČVUT

Multimedia content is very diverse . . .

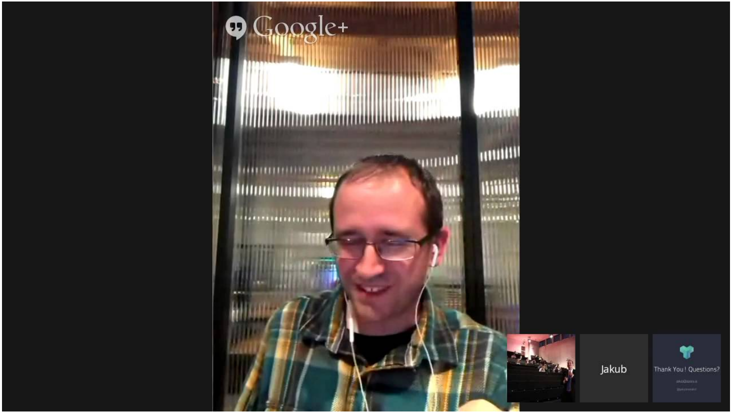
– courtesy of AVC-ČVUT

Multimedia content is very diverse . . .



– courtesy of ŠumavaNet.cz

Multimedia content is very diverse . . .



– courtesy of AVC-ČVUT

Multimedia content is very diverse . . .

– courtesy of FOX Entertainment

It would take an individual more than 5 million years to watch the amount of video that will cross global IP networks each month in 2020.
– Cisco[1]

[1] White paper: Cisco VNI Forecast and Methodology, 2015-2020

Our current scope

NARRA

Open platform for creation of open narratives.

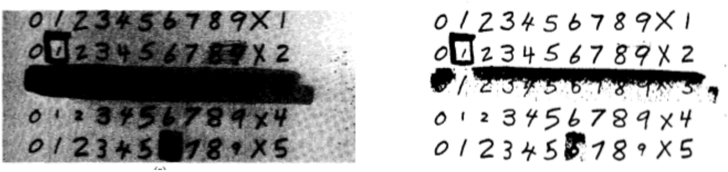
Contains mainly raw documentary material.

It started with text

Providing that the friction of the device and the load on the shaft S are low enough the pressure on the piston will move it to the position N in the direction of the arrow, thus accomplishing mechanical work. The motion of the piston revolves the crank to which it is connected by the rod X from D to E.

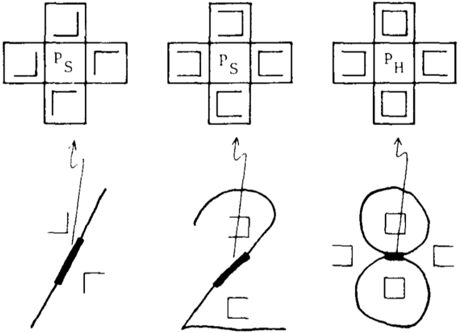
CAULFIELD, H. J.; MALONEY, W. T. Improved discrimination in optical character recognition. Applied Optics, 1969, 8.11: 2354-2356.

It started with text



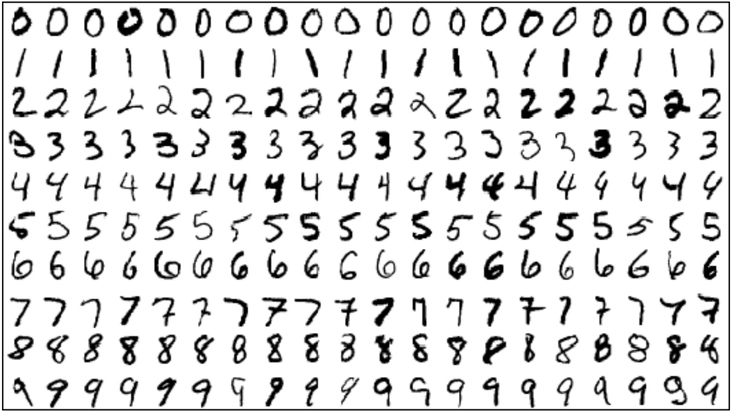
WHITE, James M.; ROHRER, Gene D. Image thresholding for optical character recognition and other applications requiring character image extraction. IBM Journal of research and development, 1983, 27.4: 400-411.

It started with text



MORI, Shunji; NISHIDA, Hirobumi; YAMADA, Hiromitsu. Optical character recognition. John Wiley & Sons, Inc., 1999.

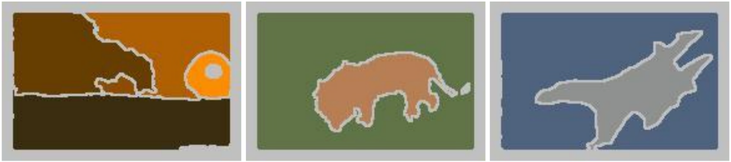
MNIST



LECUN, Yann; CORTES, Corinna; BURGES, Christopher JC. The MNIST database. URL `http://yann.lecun.com/exdb/mnist`, 1998.

LOWE, David G. Object recognition from local scale-invariant features. In: Computer vision, 1999. The proceedings of the seventh IEEE international conference on. IEEE, 1999. p. 1150-1157.

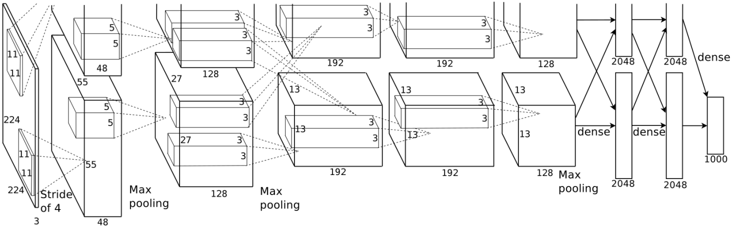
Continued with objects



sea sky sun waves   cat forest grass tiger   jet plane sky

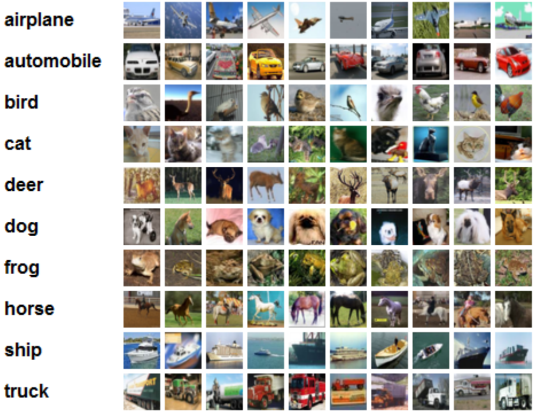
DUYGULU, Pinar, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: European conference on computer vision. Springer Berlin Heidelberg, 2002. p. 97-112.

Continued with objects



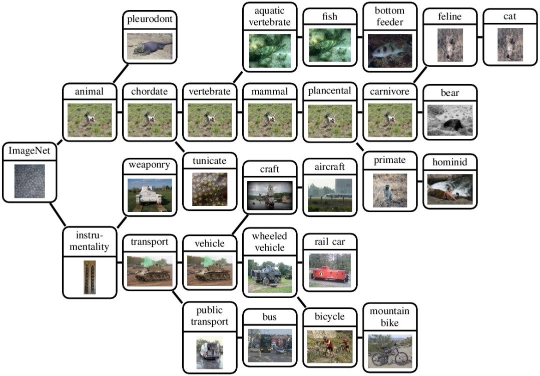
KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012. p. 1097-1105.

CIFAR-10



KRIZHEVSKY, Alex; HINTON, Geoffrey. Learning multiple layers of features from tiny images. 2009.

ImageNet



DENG, Jia, et al. Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. p. 248-255.

~10 % of WordNet synsets

How deep goes the rabbit hole?

HE, Kaiming, et al. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

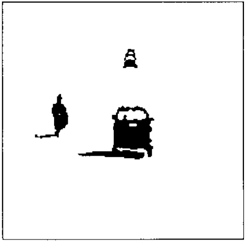
Spoiler: 152 layers

How deep goes the rabbit hole?

HE, Kaiming, et al. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

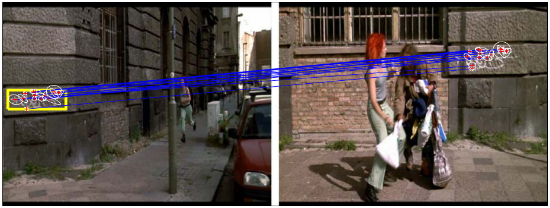
Spoiler: 152 layers
Usually “only” 50 layer version used.
http://ethereon.github.io/netscope/#/gist/db945b393d40bfa26006
https://transcranial.github.io/keras-js
https://ml4a.github.io/demos/keras.js

And video?



FORESTI, Gian Luca. Object recognition and tracking for remote video surveillance. IEEE Transactions on circuits and systems for video technology, 1999, 9.7: 1045-1062.

And video?



SIVIC, Josef; ZISSERMAN, Andrew. Video Google: A text retrieval approach to object matching in videos. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003. p. 1470-1477.

And video?

And many more specialised applications
very limited in description power:
CCTV, law enforcement, forewarning, astronomy, etc.

To sum up . . .

Input:

Audiovisual sequence
(and technical metadata, possibly faulty)

To sum up . . .

Input:

Audiovisual sequence
(and technical metadata, possibly faulty)

Goal:

Human-comprehensible description

To sum up . . .

Input:

Audiovisual sequence
(and technical metadata, possibly faulty)

Goal:

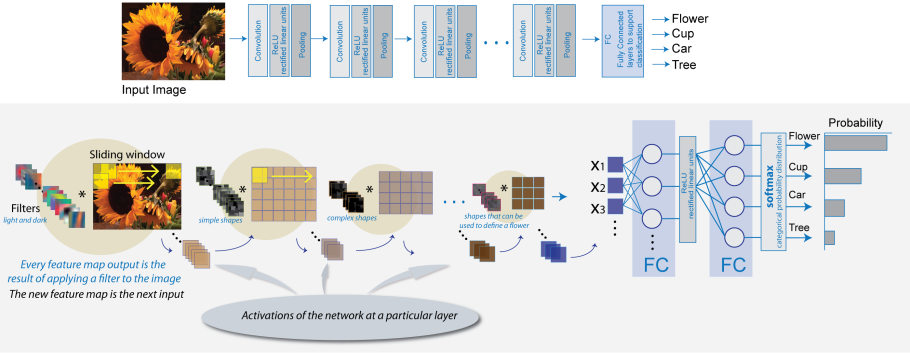
Human-comprehensible description

Reason:

Multimedia archives
(currently curated by hand)

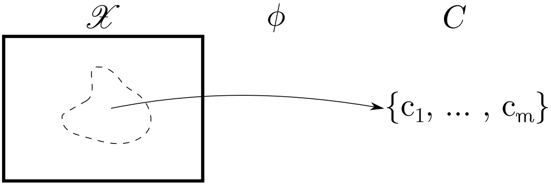
Neural Networks, Neural Networks Everywhere

Neural Networks, Neural Networks Everywhere



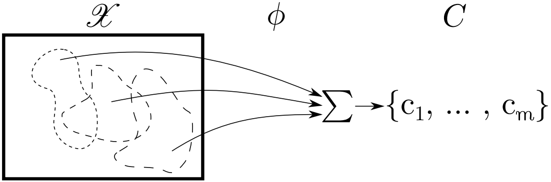
– courtesy of MathWorks

Single classifier:



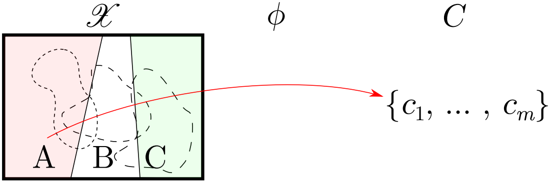
Meta-learning introduction

Team of classifiers:



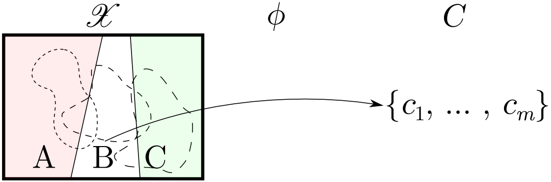
Meta-learning introduction

Meta-trained classifier selection:



Meta-learning introduction

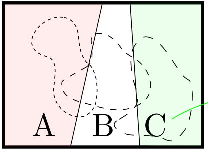
Meta-trained classifier selection:



Meta-learning introduction

Meta-trained classifier selection:

$\mathscr{X}$ ϕ C



$\{c_1, \ldots, c_m\}$

Meta-learning example

Possible approach for flower subspecies recognition (illustration):

SVM of bulb size and flower hue

k-NN of hue

k-NN of pistil length

SVM of petal and sepal dimensions

Naïve Bayes of thorn count

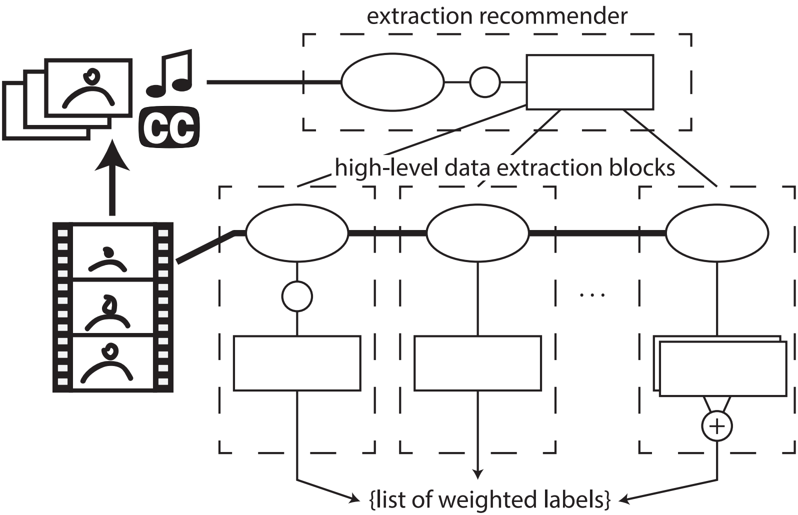
Decision tree of saturation

Gaussian mixture of bloom shape

Some features of meta-learning

- transferable meta-knowledge
- possible distributed parallel processing
- inductive transfer of knowledge
- faster than running all methods
- cheaper than creating single complex model
- higher human-comprehensibility

Proposed approach for multimedia



1. Object segmentation

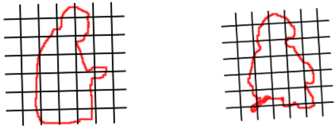
Based on motion and visual contours

1. Object segmentation

Based on motion and visual contours

1. Object segmentation

Based on motion and visual contours



3. Selection of more detailed descriptors

Upper body:

- clothes
- hair style
- carried objects
- emotions

Background:

- daytime
- season
- environment

Current status

- Finding and implementing the processing blocks.
- Improvements of the segmentation algorithm and meta-feature extraction.
- Proposing the method of precision assessment.

Thank you for your attention.

petr.pulc@fit.cvut.cz