Meta-learning and Knowledge Extraction from Multimedia

Petr Pulc\textsuperscript{1}  Martin Holeňa\textsuperscript{2}

\textsuperscript{1}Faculty of Information Technology, Czech Technical University
Prague, Czech republic

\textsuperscript{2}Institute of Computer Science, Academy of Sciences
Prague, Czech republic

January 5, 2017
How do you search for multimedia content?

Petr Pulc, Martin Holeňa (FIT CTU)

Multimedia (Meta-)learning

January 5, 2017 3 / 45
How do you search for multimedia content?

YouTube

vimeo

Dailymotion

twitch
How do you search for multimedia content?
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...
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.

– Cisco¹

¹White paper: Cisco VNI Forecast and Methodology, 2015-2020
Our current scope

**NARRA**

Open platform for creation of open narratives.

Contains mainly raw documentary material.
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.
LECUN, Yann; CORTES, Corinna; BURGES, Christopher JC. The MNIST database. URL \url{http://yann.lecun.com/exdb/mnist}, 1998.
CIFAR-10

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

~10 % of WordNet synsets
How deep goes the rabbit hole?


Spoiler: 152 layers
How deep goes the rabbit hole?


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?

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)
Every feature map output is the result of applying a filter to the image.
The new feature map is the next input.

Activations of the network at a particular layer

--- courtesy of MathWorks
Meta-learning introduction

Single classifier:

\[ \mathcal{X} \quad \phi \quad C \quad \{c_1, \ldots, c_m\} \]
Team of classifiers:

\[ \mathcal{X} \xrightarrow{\phi} C \xrightarrow{\sum} \{c_1, \ldots, c_m\} \]
Meta-trained classifier selection:

\[ \mathcal{X} \xrightarrow{\phi} C \]

\[ \{c_1, \ldots, c_m\} \]
Meta-trained classifier selection:

\[ \mathcal{X} \xrightarrow{\phi} C \rightarrow \{c_1, \ldots, c_m\} \]
Meta-trained classifier selection:

\[ \mathcal{X} \quad \phi \quad C \]

\[ \{c_1, \ldots, c_m\} \]
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 extraction recommender

high-level data extraction blocks

{list of weighted labels}
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