Large-scale Image Retrieval using Neural Net Descriptors

SIGIR 2015, August 9-13, 2015, Santiago, Chile

David Novak
david.novak@fi.muni.cz

Michal Batko
batko@fi.muni.cz

Pavel Zezula
zezula@fi.muni.cz

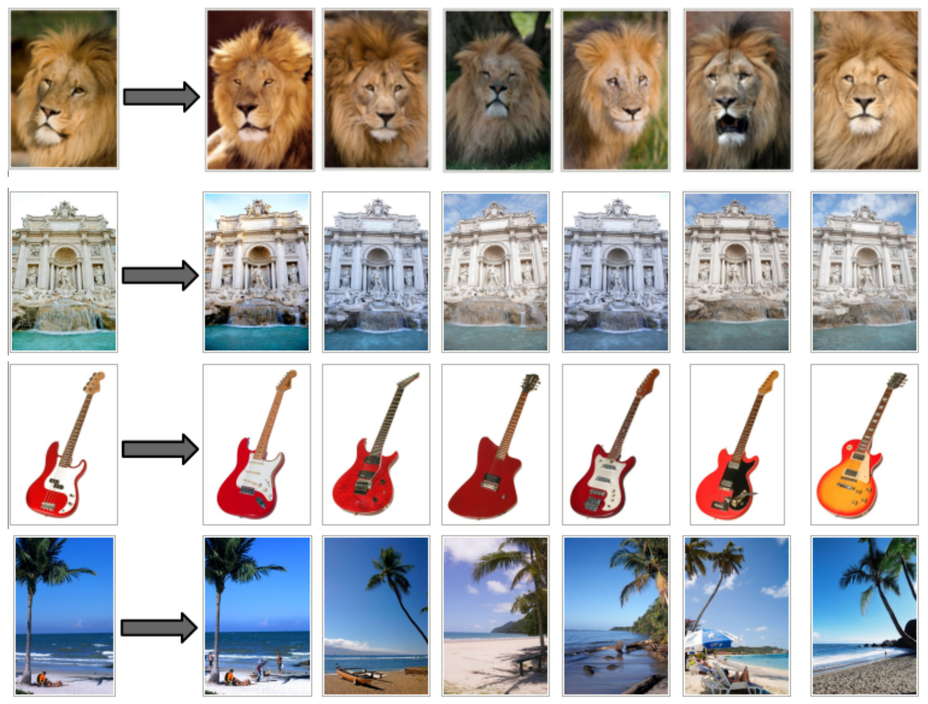
Laboratory of Data Intensive Systems and Applications (DISA), Masaryk University, Brno, Czech Republic

Demo Overview

- image collection organized for query-by-example search
  - neural network visual descriptor extracted from each image
  - descriptors organized by a distance-based similarity index
  - given an (external) query image, the system finds the *k* most similar images to the query, i.e. *k* closest descriptors

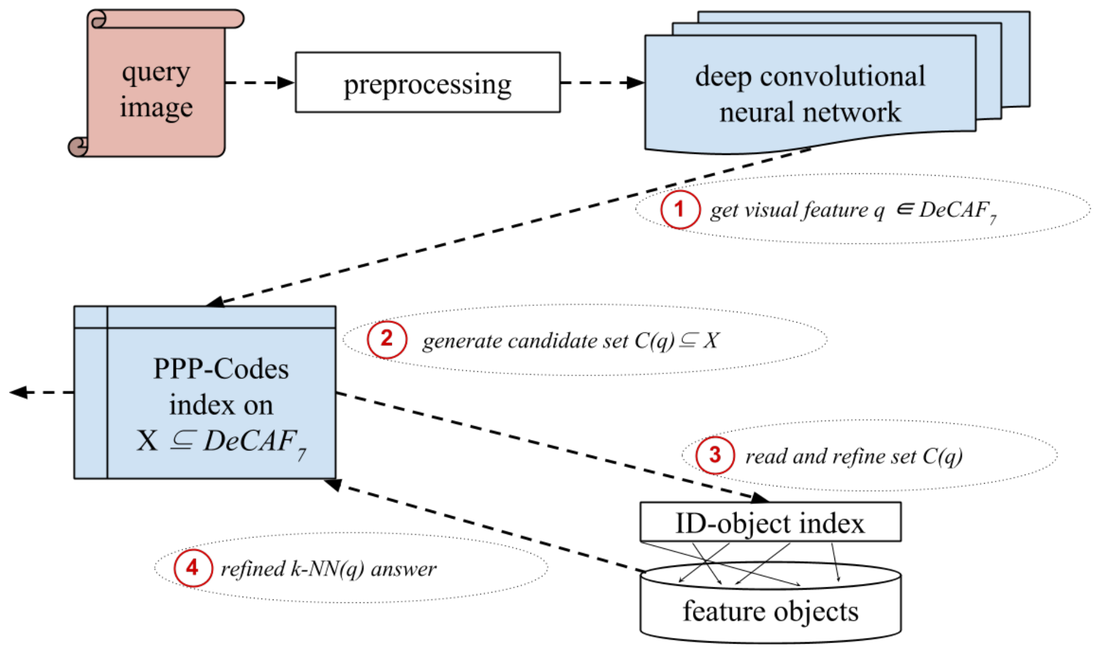
Data

- 20 million photos and graphics of various topics provided by a photo stock company for research purposes [1]
  - image thumbnails
  - text annotations
  - neural network descriptors
- publicly available for research purposes
  `http://disa.fi.muni.cz/profiset/`



`http://disa.fi.muni.cz/demos/profiset-decaf/`

System Architecture



Similarity Indexing & Searching

- indexing based on distance space: (D, δ)
  - D is domain of objects, δ is distance function δ: $D \times D \rightarrow \mathbb{R}$
    - in our case, *D is domain of DeCAF$_7$ descriptors and δ is Euclidean distance*
  - the index stores collection $X \subseteq D$
  - given a query k-NN(q), $q \in D$, the index returns k objects with minimum distance $\delta(q,x)$, $x \in X$
- distance-based index PPP-Codes [4] has two components:
  - main memory index that keeps compressed info about objects
  - disk store to keep all objects (descriptors) stored by their IDs
- given k-NN(q) query
  i. the memory index identifies candidate set $C(q) \subseteq X$
  ii. the $C(q)$ set is read from the disk and refined by $\delta(q,x)$, $x \in C(q)$

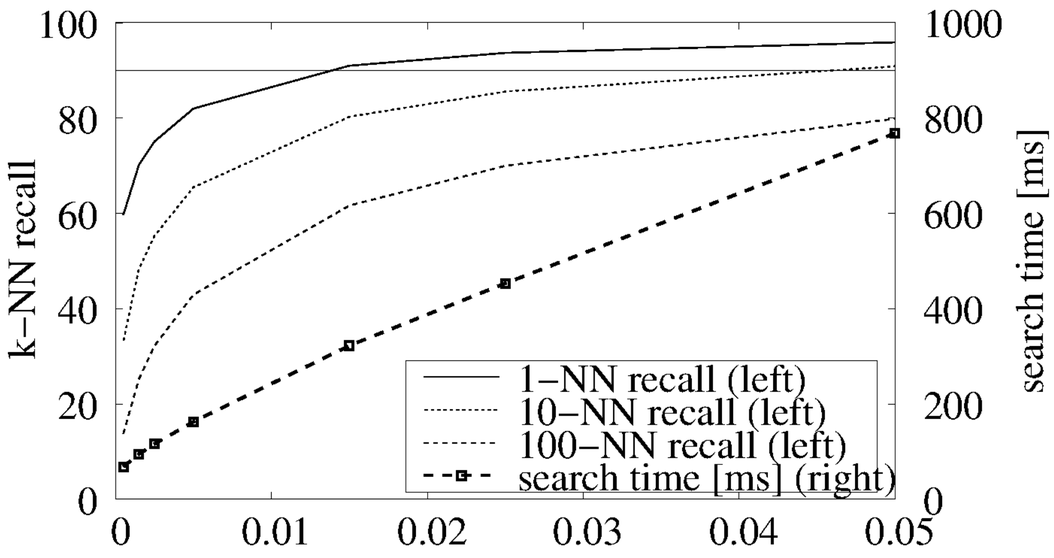
Try it yourself

`http://disa.fi.muni.cz/demos/profiset-decaf/`

Deep Convolutional Neural Networks

- we use a trained model provided by Caffe framework
  - BVLC Reference CaffeNet: trained like Krizhevsky et al. [2]
  - no retraining nor fine tuning
  - `http://caffe.berkeleyvision.org/`
- we use output from the last hidden layer as descriptor [3]
  - the DeCAF$_7$ descriptor is 4096-dimensional float vector
  - Euclidean distance on DeCAF$_7$ vectors expresses visual similarity
- 20 million images: 320 GB of uncompressed DeCAF$_7$ descriptors

Performance Evaluation



- online response to k-NN queries with recall about 80% [1]

References

[1] Novak, Cech, Zezula (2015). Efficient Image Search with Neural Net Features. To appear in *Proceedings of SISAP* 2015.

[2] Krizhevsky, Sutskever, Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012, pp. 1106–1114.

[3] Donahue et al. (2014). DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In Proceedings of ICML 2014, pp. 647–655.

[4] Novak, Zezula (2014). Rank Aggregation of Candidate Sets for Efficient Similarity Search. DEXA 2014, pp. 42-58. (Best Paper of DEXA 2014)