VRG Prague in "Large-Scale Landmark Recognition Challenge"

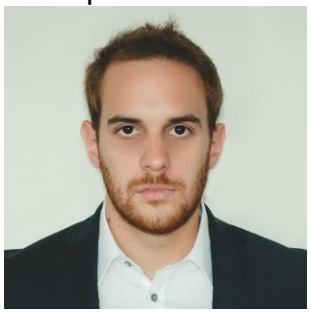
CVPR 2018 workshop: Large-Scale Landmark Recognition: A Challenge

Presenter: Giorgos Tolias

Visual Recognition Group, CTU in Prague

The team

Filip Radenović



Ahmet Iscen

Giorgos Tolias

Ondřej Chum

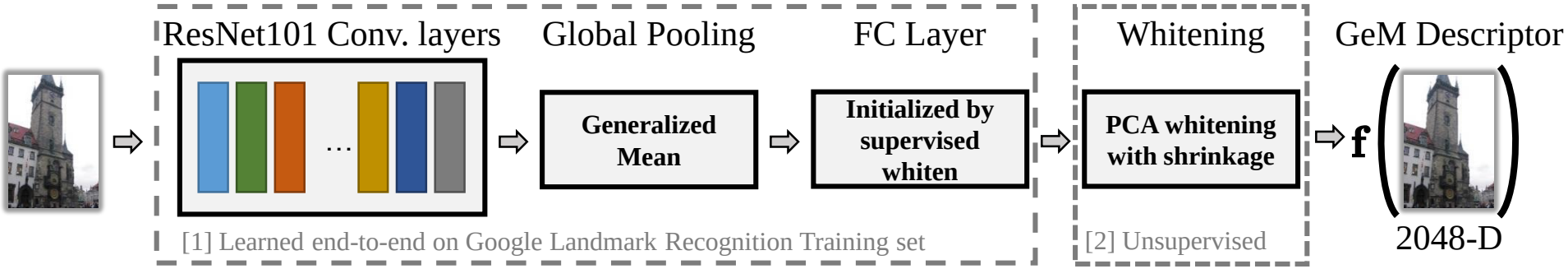
Visual Recognition Group, Czech Technical University in Prague

Our approach: an overview

- Landmark recognition
  - Huge number of classes
  - Low intra-class variability
- Don't train a landmark classifier
  - Metric learning for training descriptors or use pre-trained
  - k-NN classifier
- Combine CNNs with classical approaches
  - Global CNN-based descriptors
  - Local features and spatial verification

CNN-based global descriptors

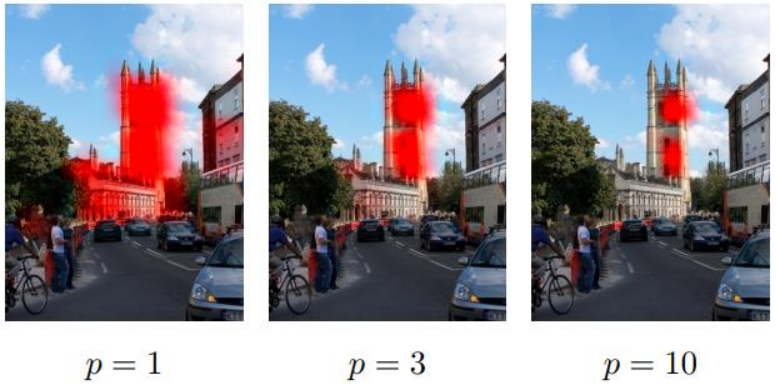
GeM descriptor



Generalized-mean pooling (GeM):

$$f_k = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^p \right)^{\frac{1}{p}}$$

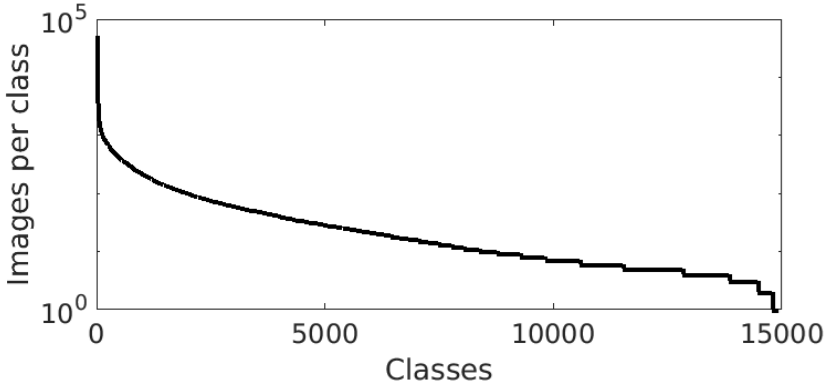
$p \to \infty$ max pool MAC
$p = 1$ avg pool SPoC



[1] Radenovic, Tolias, Chum: Fine-tuning CNN Image Retrieval with No Human Annotation, PAMI 2018 (github.com/filipradenovic/cnnimageretrieval-pytorch)
[2] Mukundan, Tolias, Bursuc, Jegou, Chum: Understanding and Improving Kernel Local Descriptors, submitted to IJCV

GeM-based recognition

- Simple k-NN classifier
- Accumulate similarity with top-N training images
- IDF-like normalization per class



GeM recognition












GeM common mistakes












GeM common mistakes









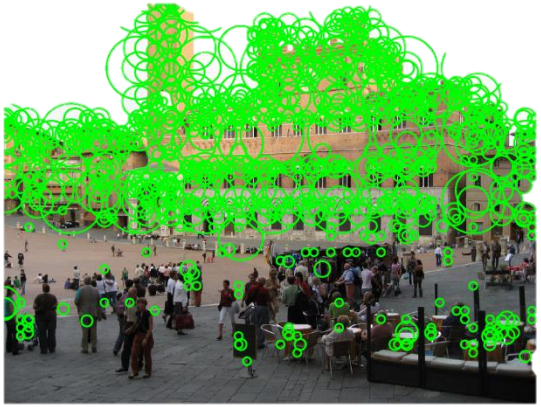
Local features – Spatial matching

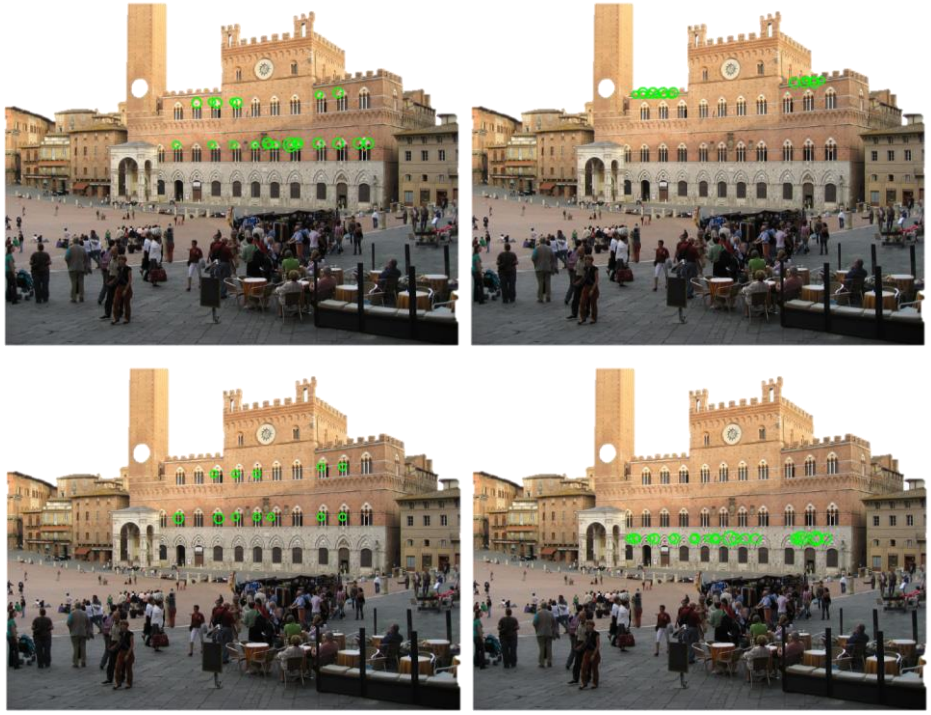
Local features extraction & indexing

Aggregated Selective Match Kernel (ASMK) [3]

- Bag-of-Words based
- Single binary descriptor per word
- Inverted file structure

- DELF local features [4]



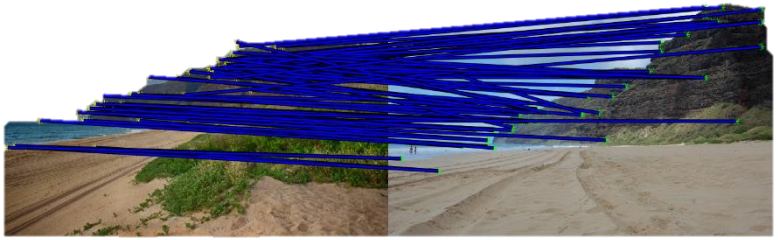


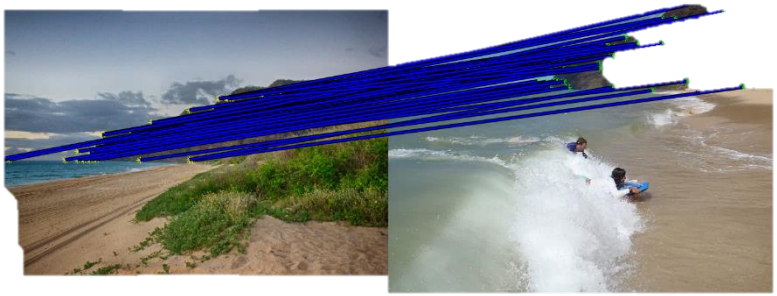
features in the same word

[3] Tolias, Avrithis, Jegou: To aggregate or not to aggregate: Selective match kernels for image search, ICCV 2013
[4] Noh, Araujo, Sim, Weyand, Han: Large-Scale Image Retrieval with Attentive Deep Local Features, ICCV 2017

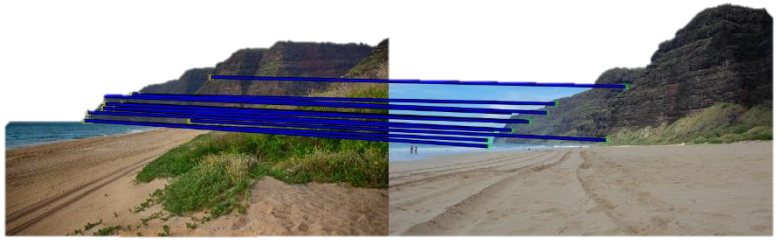
Matching & spatial verification

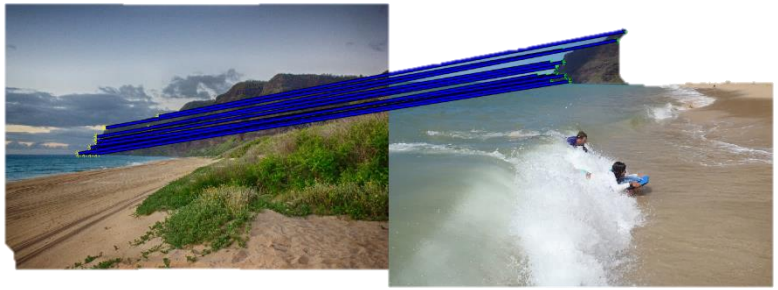
Tentative matches (ASMK)





Spatial verification (SP) [5]





[5] Philbin, Chum, Isard, Sivic, Zisserman: Object retrieval with large vocabularies and fast spatial matching, CVPR 2007

- Recognition (similar to GeM)
  - Simple k-NN classifier
  - Rank with ASMK
  - Accumulate #inliers for top-100 training images with $\geq$8 inliers
  - IDF-like normalization per class

Recognition with spatial verification










Spatial verification – common mistakes








Spatial verification – common mistakes








Combined classifier

Combined class similarity

GeM class similarity

SP class similarity

$$C = f(A) + \lambda \cdot g(B)$$

Combined classifier

Combined class similarity

GeM class similarity

SP class similarity

$$C = f(A) + \lambda \cdot g(B)$$

identity

0.5

$g(x) = x^{0.1}$

Combined classifier

Combined class similarity

GeM class similarity

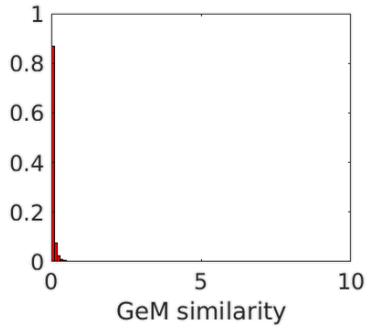
SP class similarity

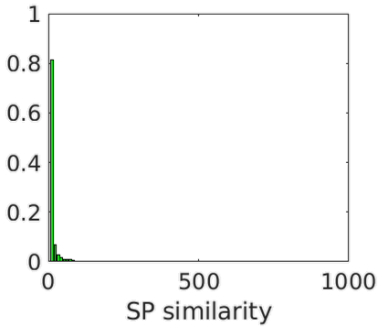
$$C = f(A) + \lambda \cdot g(B)$$

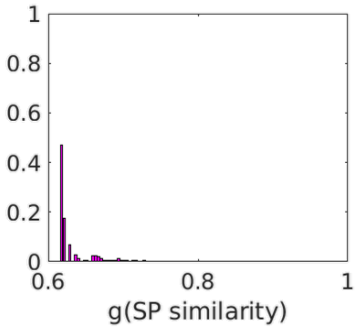
identity

0.5

$g(x) = x^{0.1}$







Combined classifier

Combined class similarity

GeM class similarity

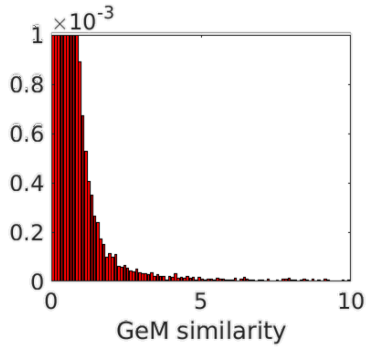
SP class similarity

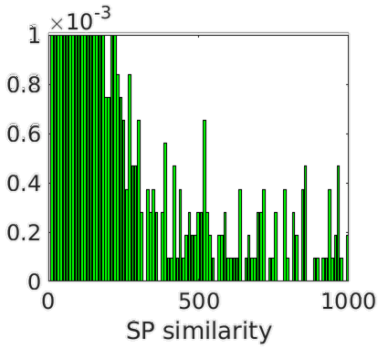
$$C = f(A) + \lambda \cdot g(B)$$

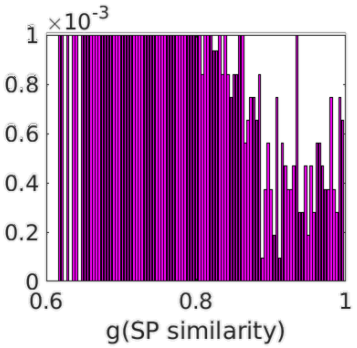
identity

0.5

$g(x) = x^{0.1}$







Combined classifier

Combined class similarity

GeM class similarity

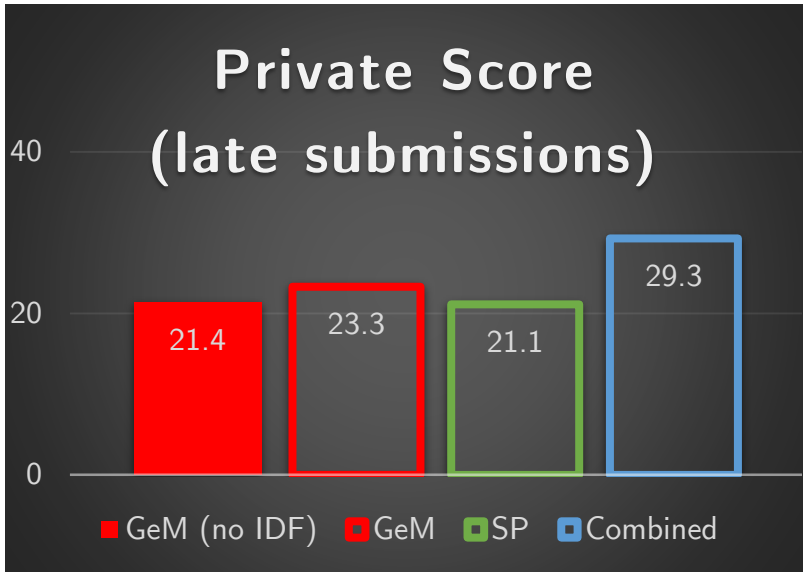
SP class similarity

$$C = f(A) + \lambda \cdot g(B)$$

identity

0.5

$g(x) = x^{0.1}$



Link to code and references

- CNN training for retrieval (MatConvNet & Pytorch)
  http://cmp.felk.cvut.cz/cnnimageretrieval/

- ASMK Matlab/mex implementation
  http://cmp.felk.cvut.cz/~toliageo/soft.html

- DELF
  https://github.com/tensorflow/models/tree/master/research/delf

[1] Radenovic, Tolias, Chum: Fine-tuning CNN Image Retrieval with No Human Annotation, PAMI 2018
[2] Mukundan, Tolias, Bursuc, Jegou, Chum: Understanding and Improving Kernel Local Descriptors, (submitted) IJCV
[3] Tolias, Avrithis, Jegou: To aggregate or not to aggregate: Selective match kernels for image search, ICCV 2013
[4] Noh, Araujo, Sim, Weyand, Han: Large-Scale Image Retrieval with Attentive Deep Local Features, ICCV 2017
[5] Philbin, Chum, Isard, Sivic, Zisserman: Object retrieval with large vocabularies and fast spatial matching, CVPR 2007

Our work at CVPR 2018

- Related to the Retrieval Challenge
  - Paper 2730 Wed
    about: **Large-scale retrieval benchmark**
  - Paper 2778 Thu
    about: **Efficient manifold search**
  - Paper 2779 Thu
    about: **Manifold search for unsupervised learning**

Implementation details

- GeM
  - Based on ResNet101 pre-trained on ImageNet
  - 1 positive, 5 negatives per anchor
  - 300 epochs (1 week of training, 1 x Tesla P100)
  - Unsupervised whitening on ≤3 images/landmark + 60k external
  - 2048D descriptor
  - Multi-scale extraction (724, 1024, 1448): sum-aggregate
  - Use top-8 images for recognition
- DELF
  - ≤1000 features/image
  - 128D per descriptor
- ASMK
  - 65k words,
  - Assignment to 5 words
  - Accumulated similarity: $(descriptor\ similarity)^3$
  - Threshold at hamming distance: 48
- Spatial verification
  - Re-projection error $\leq 2 \cdot 25^2$
  - Verified if ≥8 inliers

GeM

Memory for 1.2M images
9.3 GB (2.3 GB 1byte/dim)

Recognition time
210 ms CPU machine

SP

Memory for 1.2M images
17 GB (11+6)

Recognition time
340 ms CPU machine