Insight DCU @ TRECVid 2014

Kevin McGuinness, Eva Mohedano, Zhenxing Zhang, Feiyan Hu, Rami Albatal, Cathal Gurrin, Noel E. O'Connor, Alan F. Smeaton, Amaia Salvador, Xavier Giro-i-Nieto, Carles Ventura

Centre for Data Analytics

Insight

Overview

Insight-DCU participated in the instance search (INS) and semantic indexing (SIN) tasks in 2014. Two very different approaches were submitted for INS, one based on features extracted using pre-trained deep convolutional neural networks (CNNs), and another based on local SIFT features, large vocabulary visual bag-of-words aggregation, inverted index-based lookup, and geometric verification on the top-N retrieved results. Two interactive runs and two automatic runs were submitted, the best interactive runs achieved a mAP of 0.135 and the best automatic 0.12. Our semantic indexing runs were based also on using CNN features, and on SVM classifiers with linear and RBF kernels. One run was submitted to the main task, two to the no annotation task, and one to the progress task. Data for the no-annotation task was gathered from Google Images and ImageNet. The main task run has achieved a mAP of 0.086, the best no-annotation runs had a close performance to the main run by achieving a mAP of 0.080, while the progress run had 0.043.

Instance Search using CNN features

A target image dataset was built by uniformly extracting keyframes for every shot with a sample rate of 1/4 fps. The resulting dataset contained 647,628 keyframes.

Features

The system used *Caffe* and the pretrained ImageNet models extract 4096D CNN features for each keyframe from layer 7 of the network. CNN features extracted globally for whole query image and also locally on region around target object mask. SCG object candidates also extracted for some of the queries, and local CNN features computed on these candidates.

(a) Cropping (b) Square crop (c) Zeroing (d) Blurring

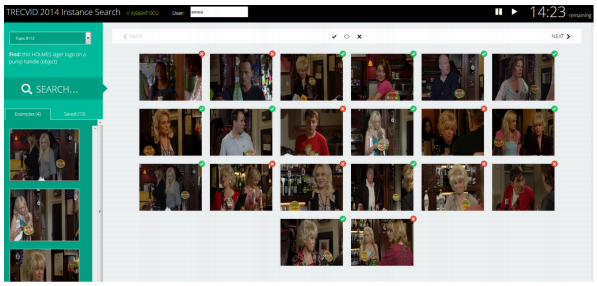
Indexing

Approximate nearest neighbours based on 50 forests of random prodjections (Spotify annoy implementation) used for matching with cosine similarity

User interface

Web-based UI for interactive instance search developed allowing users to annotate results as "positive," "negative," or "neutral."

Results can be refined whenever the user desires by requerying using images marked as positive and merging the results.



Runs

Final list expanded using two strategies: requerying and linear SVM. Interactive runs achieved mAP of 0.135 and 0.126. Automatic run achieved accuracy of 0.062.

Instance Search using Bag-of-words

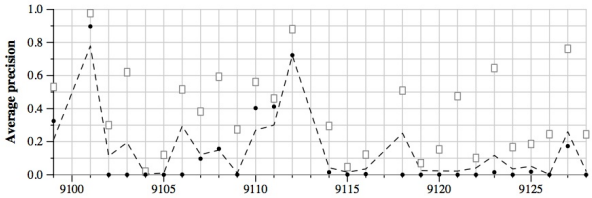
Approach uses the widely-used bag-of-word representation and vector-space model of information retrieval. We used high-dimensional vocabulary size to increase the visual words' discriminative power, and then spatial verification and query expansion technologies to further improve the performance. We focused on increasing the initial ranked accuracy by using a query-adaptive weighting function for visual objects similarity measurement which is a crucial point for spatial verification and query expansion. Product quantization for ANN Search is also used to improve the accuracy and speed performance in the feature quantization step.

Details

- Harris-Laplace detector, SIFT descriptor
- Vocabulary Size: 1M
- Inverted index, Lp-Norm IDF weighting
- Multiple query and jointed average scoring
- Geometric verification on top 100 results
- Query expansion using top 20 results

Runs

One automatic run, mAP 0.120.



Semantic Indexing

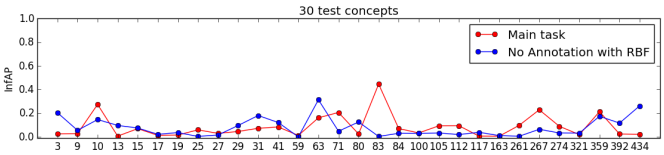
4096D CNN features from pre-trained ImageNet 2011 model, extracted from layer 7 using Caffe. Standard mirroring and cropping expansion, followed by average pooling applied.

No annotation task

For this no annotation task single-term queries were posted to two data sources: Google Images, and ImageNet. To ensure a high-quality training set, we first search for the concept in ImageNet; if the concept does not exist in as an ImageNet visual category, we use images retrieved using a search for the term on Google Images. We gathered 36 concepts from ImageNet and 24 concepts from Google Image search.

Experiments

When we compare the results of the main task (mAP = 0.086) and no annotation task (best mAP = 0.080), about 20 concepts has similar or better infAP score in no annotation task than the result in main task. While the following concepts (Beach, Chair, Instrumental musician, News studio, Singing, stadium, flags, forest) have better score in main task.



This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289. Programme material copyrighted by BBC.

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD DUBLIN sfi Science Foundation Ireland