Swap Retrieval: Retrieving images of cats when the query shows a dog

Amir Ghodrati, Xu Jia, Marco Pedersoli, Tinne Tuytelaars

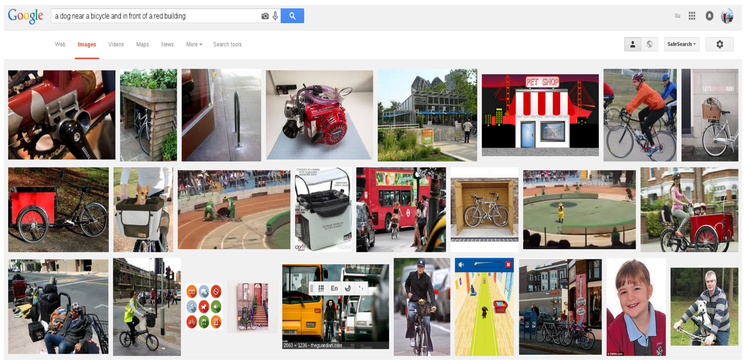
ESAT-PSI / iMinds,
KU Leuven

Introduction

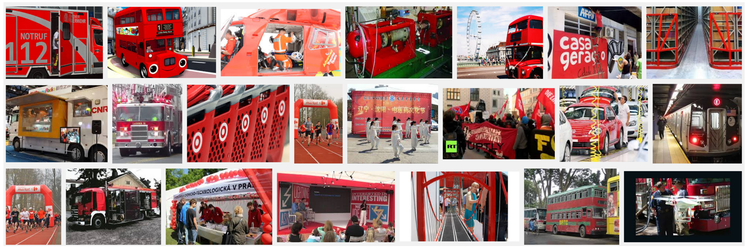
Motivation:

What is in user's mind

Cross-modal text-based image retrieval:



Query-by-example image retrieval:



Task:

Frst attempt towards the hybrid-query-by-example image retrieval - category-swap image retrieval **(swap retrieval)**

Given a query image containing an object from one category, retrieve the images with similar context but containing an object from another category.

Solutions

Baselines:

1. Similarity based on visual features
The similarity score between the query image from a given category and the candidate images from the swapped category is computed directly based on DeCAF feature or Classemes feature representation.

2. Domain Adaptation
Images with similar context but from two different object categories are considered as source and target domains.

3. Metric Learning
Learn a metric using textual similarity and dissimilarity as constraint

Attribute-based method

attribute collection

attribute selection
textually discriminative and visually compact

luggage	television	naps	remote
suitcase	tv	asleep	control

paired attributes
represent more complex concepts, similar to bi-concept and visual phrase

motorcycle, man	camera, close	keyboard, sleep
motorcycle, riding	close, stare	computing, sleep

attributes for swap retrieval

class-sensitive paired attributes	hat, cat	hat, dog
common paired attributes	motorcycle, man	People, kite

attribute detector
train an SVM classifer for each attribute

Ranking
based on the similarity between the query and candidates computed in the attribute space

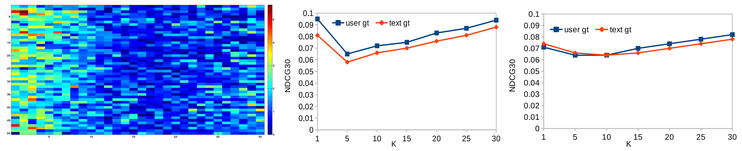
Experiments

Evaluation metric

NDCG: measures the ranking quality of a retrieved image based on its relevance and position in the ranking list.

$$NDCG_k = \frac{1}{Z} \sum_{i=1}^{k} \frac{2^{rel_i}-1}{\log_2(i+1)}$$

relevance
textual similarity between the query and candidates;
learning a ranking function based on human judgement



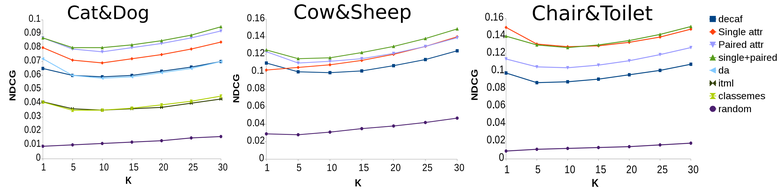
matrix of agreement between human ranking and text based

performance of attribute based method and DeCAF for user ranking and text ranking as groundtruth

Dataset
Three pairs of categories from COCO dataset:
(cat, dog), (cow, sheep) and (toilet, chair)

Results

Quantitative results



Qualitative results