

Prof. Guojun Lu
Gippsland Campus
FedUni

Research proposal

This document describes a research proposal for the project "Image Search and Retrieval" in charge of Prof. Guojun Lu at FedUni, Australia.

Overview

Approaches to image information retrieval based solely on natural language descriptions of the images not only suffer from the problem that such descriptions may not always be available but also the process of matching images according to user descriptions may lead to undesired results. Users may produce different statements even for the same image according to their mood, the context of the image or other factors that are hard to capture without the intervention of a human.

An effort to overcome this high variability in results obtained by crowd sourcing text description of images is explored by Socher[1] et al., where the authors propose a method to segment, annotate and classify images based on deep architectures.

Besides outperforming state of the art methods in similar image tasks, deep architectures provide also a means of learning hierarchical levels of features with each level providing a further sophistication in the abstraction used to describe the data. In this way, detected parts of the image in low layers can be recursively composed in higher layers to convey a higher level of representation. These abstractions can be used to better match one image to another.

Although this approach removes the problems of relying only on low level features to understand an image, there is still room for improvements that can be applied to the solution.

Deep architectures are good learning high level features over low level ones in an unsupervised manner. However, they still require labelling of the data to be trained successfully. To overcome this problem one may think of latent variable methods such as the one proposed in[2] for natural language documents. In these models, the different topics in which documents are classified are learnt as part of the training process, not depending on hand labelled data. Then the similarity between two documents can be assessed by assigning a vector to each document representing how much each document belongs to a topic class and computing distances between these vectors.

Objectives

So, a desiderata for a method in image similarity and retrieval would include the hierarchical abstraction power of deep architectures, the unsupervised learning capabilities of latent variable models, and also a method to retrieve images fast. A really promising approach has been proposed by Hinton and Krizhevsky[3] where the authors not only propose a means to address image similarity in an hierarchical unsupervised fashion, but also a way of retrieving similar images using a technique called semantic hashing[5].

In this approach, many levels of features are learnt, and these features are used to initialize autoencoders. The autoencoders are then used to map each image to a binary word. Similar images map to similar binary words. To retrieve images that are similar to a query image, the binary code is computed for a query image, and small modifications are performed in the code to obtain the addresses of similar images. This leads to a constant time method for image retrieval.

Another approach to image similarity would be to train deep neural networks to classify objects and then use the activity vectors from the last hidden layer to compare between images. Although this leads to excellent results as can be seen in [4], it is a supervised method, and does not have the advantages from semantic hashing.

Another related problem is how to include the captions in the retrieval process when they are available, how to decide when the captions are relevant, and what to do when captions are missing.

Consequently, a method that meets all the above mentioned requirements still needs to be found, and needs to be tested in an experimental environment to assess for generalization in new datasets and suitability in adapting to practical computing infrastructures.

Research Plan

Work will begin with bibliographic research of image retrieval and similarity methods and their suitability to be extended or combined to meet the above mentioned goals. Developing new methods from scratch may be a possibility, although it is not likely to be necessary.

Implementations of state of the art methods will be created or modified, in case they are already available, with the intention of evaluating new ideas by comparing their performance with a baseline technique.

A key part in the evaluation of any Machine Learning technique is to test it for generalization in new data, so it will be really important to test the techniques on new datasets.

References

- [1] Parsing Natural Scenes and Natural Language with Recursive Neural Networks, Richard Socher, Cliff Lin, Andrew Y. Ng, and Christopher D. Manning.

- [2] Probabilistic topic models. Steyvers, M. & Griffiths, T. (2007).
- [3] Using Very Deep Autoencoders for Content-Based Image Retrieval. Krizhevsky, A. and Hinton, G.E.(2011).
- [4] ImageNet Classification with Deep Convolutional Neural Networks. Krizhevsky, A. and Hinton, G.E.(2011).
- [5] Semantic Hashing. Salakhutdinov R. R, and Hinton, G. E.(2007).