International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 16th Aug 2017. Link :http://www.ijiemr.org/downloads.php?vol=Volume-6&issue=ISSUE-7

Title: AN EFFECTIVE CONTENT BASED IMAGE RETRIEVAL BY USING CNN CLASSIFIER

Volume 06, Issue 07, Pages: 219 – 228.

Paper Authors

P.SIVA JYOTHI, SK.MD.SHAREEF

Eswar College Of Engg, Narasaraopet, Guntur Dt.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

AN EFFECTIVE CONTENT BASED IMAGE RETRIEVAL BY USING CNN CLASSIFIER

[1]P.SIVA JYOTHI, [2]SK.MD.SHAREEF

[1]PG Scholar, Dept Of E.C.E, Eswar College Of Engg, Narasaraopet, Guntur Dt.

[2]Asst Professor, Dept Of E.C.E, Eswar College Of Engg, Narasaraopet, Guntur Dt.

ABSTRACT: An effective content-based image retrieval (CBIR) system depends on the discriminative feature which represents an image. In this work, we explore deep convolutional features for a CBIR system. We first show the effectiveness of deep convolutional channel features for a CBIR system. Then we introduce a Multi- Level Pooling method (MLP) to obtain object-aware features from convolutional layers and finally the features extracted from different layers are incorporated to a short representation vector. Through multiple experiments, we show that our approach can achieve state-of-art results on several benchmark retrieval datasets

INTRODUCTION: An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.The first microcomputer-based image database retrieval system was developed at MIT, in the 1980s, by Banireddy Prasaad, Amar Gupta, Hoo-min Toong, and Stuart Madnick.

Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc.

- Image meta search - search of images based on associated metadata such as keywords, text, etc.
- Content-based image retrieval (CBIR) – the application of computer vision to the image retrieval. CBIR aims at avoiding the use of textual descriptions and instead retrieves images based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query image or user-specified image features.
  - List of CBIR Engines - list of engines which search for images based image visual content such as color, texture, shape/object, etc.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Data Scope

It is crucial to understand the scope and nature of image data in order to determine the complexity of image search system design. The design is also largely influenced by factors such as the diversity of user-base and expected user traffic for a search system. Along this dimension, search data can be classified into the following categories:

- Archives - usually contain large volumes of structured or semi-structured homogeneous data pertaining to specific topics.
- Domain-Specific Collection - this is a homogeneous collection providing access to controlled users with very specific objectives. Examples of such a collection are biomedical and satellite image databases.
- Enterprise Collection - a heterogeneous collection of images that is accessible to users within an organization's intranet. Pictures may be stored in many different locations.
- Personal Collection - usually consists of a largely homogeneous collection and is generally small in size, accessible primarily to its owner, and usually stored on a local storage media.

Web - World Wide Web images are accessible to everyone with an Internet connection. These image collections are semi-structured, non-homogeneous and

CBIR attracts much attention both academically and commercially for decades. The retrieval performance of a content-based image retrieval system crucially depends on the feature representation and similarity measurement. In early CBIR systems, images are indexed by their visual content, which is represented by low-level information, including color features, texture features and shape features. Although a variety of techniques have been proposed, the well-known "semantic gap" issue laying between low-level image pixels captured by machines and high-level semantic concepts perceived by human is still one of the most challenging problems in current CBIR research. In last few years, there were many important advances in machine learning. One important breakthrough is known as the deep learning technique, which includes several different types of deep architectures composed of multiple linear and nonlinear transformations. Deep learning has led to state-of-art performance on various problems such as image classification object detection face recognition , etc. Among different types of deep neural networks, convolutional neural networks have been most extensively studied. Convolutional neural network (CNN) is first introduced by in and improved in . However, due to the lack of training data and computing power in early days, it is extremely hard to train a large high-capacity convolutional neural network. Recently, with the rapid growth of data size and the increasing computing power of graphics processor unit, many researchers used convolutional neural network to achieve

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

start-of-art results on various tasks performance of CNN features on retrieval task. For example, has extensively evaluated the performance of the deep features extracted from fully connected layers with and without fine-tuning on related dataset, and overall reported that the features from fc layers outperform traditional SIFT-like features. proposed a method called Convolutional Channel Features (CCF), which transfers low-level features from pre-trained CNN models to feed the boosting forest model. The work shows that convolutional channel features serve as a good way of tailing pre-trained CNN models to diverse tasks without fine-tuning the whole network to each task by achieving state-of-art performances in pedestrian detection, face detection, edge detection and object proposal generation.

TOPIC: Textural Features for Image Classification

Texture is one of the important characteristics used in identifying objects or regions of interest in an image, whether the image be a photomicrograph, an aerial photograph, or a satellite image. This paper describes some easily computable textural features based on gray tone spatial dependencies, and illustrates their application in category identification tasks of three different kinds of image data: photomicrographs of five kinds of sandstones, 1:20 000 panchromatic aerial photographs of eight land-use categories, and Earth Resources Technology Satellite (ERTS) multi special imagery containing seven land-use categories. We use two kinds of decision rules: one for which the decision regions are convex polyhedral (a piecewise linear decision rule), and one for which the decision regions are rectangular parallelepipeds (a min-max decision rule). In each experiment the data set was divided into two parts, a training set and a test set. Test set identification accuracy is 89 percent for the photomicrographs, 82 percent for the aerial photographic imagery, and 83 percent for the satellite imagery. These results indicate that the easily computable textural features probably have a general applicability for a wide variety of image-classification applications.We have described a class of quickly computable textural features which seem to have general applicability to many kinds of image data. The textural features are based on statistics which summarize the relative frequency distribution (which describes how often one gray tone will appear in a specified spatial relationship to another gray tone on the image). We have used these features in category-identification tasks of three different kinds of image data. Identification accuracy on independent test sets are 89 percent for the photomicrograph image set (five categories of sandstones), 82 percent for the aerial photographs (eight landuse categories), and 83 percent for the satellite imagery (seven land-use categories). These initial experimental results are promising. Much work needs to be done, however, on gray-tone normalization on the imagery and the use of features which are invariant under monotonic gray-tone transformations. The reason is that in one important sense, texture is independent of tone. Two people examining photographs of the same texture

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

may actually be seeing two different, though related, kinds of tones in the texture. One photograph may have been developed such that its tones are light and thin, and the other photograph may have been developed such that its tones are dark and heavy. Most people could easily make the observation that the texture on the two images is the same. For a machine to find that the textures are the same, either the images must be probability quantized or the features computed from the probability quantized images (which are invariant under monotonic gray-tone transformations), or the features themselves must be invariant under monotonic gray-tone transformations. Of the textural features described in Appendix I, the angular second-moment, the entropy, the sum entropy, the difference entropy, the information measure of correlation, and the maximal-correlation features have the invariance property. We intend to repeat the experiments reported here using these kinds of features. We expect that these features will provide more generalized results

TOPIC: Image Indexing Using Color Correlograms, Conference on Computer Vision & Pattern Recognition

We define a new image feature called the color correlogram and use it for image indexing and comparison. This feature distills the spatial correlation of colors, and is both effective and inexpensive for content-based image retrieval. The correlogram robustly tolerates large changes in appearance and shape caused by changes in viewing positions, camera zooms, etc. Experimental evidence suggests that this new feature outperforms not only the traditional color histogram method but also the recently proposed histogram refinement methods for image indexing/retrievalWe have described a new image feature that can be used to index and compare images. Since this feature captures the spatial correlation of colors in an image, it is effective in discriminating images. It thus rectifies the major drawbacks of the classical histogram method. The correlogram can also be computed efficiently. Our experiments on a large image database evaluated using fair performance measures show that the correlogram performs very well. The correlogram is powerful and needs to be explored in further detail. One practical question is, can the correlogram be compressed with only a minor loss in quality? It will also be interesting to study the use of correlograms for target search and open-ended browsing of image databases . Some of our results show that when there is a large lighting change between a query and its correct answer, auto correlograms rank the answer within the top 15 (in these cases, the histogram and CCV/C fail). It will be interesting to try correlograms on color spaces which are stable under lighting change and are also perceptually uniform. Our experiments so far have been based on color quantization in the RGB color space.

TOPIC: Visual pattern recognition by moment invariant

In this paper a theory of two-dimensional moment invariants for planar geometric figures is presented. A fundamental theorem is established to relate such moment

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

invariants to the well-known algebraic invariants. Complete systems of moment invariants under translation, similitude and orthogonal transformations are derived. Some moment invariants under general two-dimensional linear transformations are also included. Both theoretical formulation and practical models of visual pattern recognition based upon these moment invariants are discussed. A simple simulation program together with its performance are also presented. It is shown that recognition of geometrical patterns and alphabetical characters independently of position, size and orientation can be accomplished. It is also indicated that generalization is possible to include invariance with parallel projection.asked on a comprehensive study with both handwritten numerals and aircraft data, the authors show that the new method of deriving Zernike moment invariants along with the new normalization scheme yield the best overall performance even when the data are degraded by additive noise

TOPIC: Histograms of Oriented Gradients for Human Detection

We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection. We study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results. The new approach gives near-perfect separation on the original MIT pedestrian database, so we introduce a more challenging dataset containing over 1800 annotated human images with a large range of pose variations and backgroundsWe have shown that using locally normalized histogram of gradient orientations features similar to SIFT descriptors in a dense overlapping grid gives very good results for person detection, reducing false positive rates by more than an order of magnitude relative to the best Haar wavelet based detector from . We studied the influence of various descriptor parameters and concluded that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good performance. We also introduced a new and more challenging pedestrian database, which is publicly available. Convolutional neural network (CNN) is a type of feed-forward artificial neural network where the individual neurons are tiled in such a way that they respond to overlapping re-gions in the visual field" [11]. They are biologically-inspired invariant of Multilayer Per-ceptrons (MLP) which are designed for the purpose of minimal preprocessing. These mod-els are widely used in image and video recognition. When CNNs are used for image recog-nition, they look at small portions of the input image called receptive fields with the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

help of multiple layers of small neuron collections which the model contains [11]. The results we get from this collection are tiled in order for them to overlap such that a better represen-tation of the original image is obtained; every such layer repeats this process. This is the reason they are able if the input image is translated in any way. The outputs of neuron clus-ters are combined by local or global pooling layers which may be included in convolutional networks. Inspired by biological process, convolutional networks also contain various com-binations of fully connected layers and convolutional layers, with point-wise nonlinearity applied at the end of or after each layer [11]. The convolution operation is used on small regions so as to avoid the situation when if all the layers are fully connected billions of pa-rameters will exist. Convolutional networks use shared weights in the convolutional layers i.e. for each pixel in the layer same filter (weights bank) is used which is advantageous because it reduces the required memory size and improves performance. CNNs use rela-tively less amount of pre-processing as compared to other image classification algorithms, meaning that the network learns the filters on its own which are traditionally manually-engineered in other algorithms. CNNs have a major advantage over others due to the lack of a dependence on prior-knowledge and the difficult to design hand-engineered features.

4.1.1 Sparse Connectivity

CNNs enforce a local connectivity pattern between neurons of adjacent layers to exploit spatially-local correlation [6]. We have illustrated in fig.4.1 that in layer m the inputs of hidden units are from a subset of units in layer m-1, units containing spatially adjoining receptive fields.

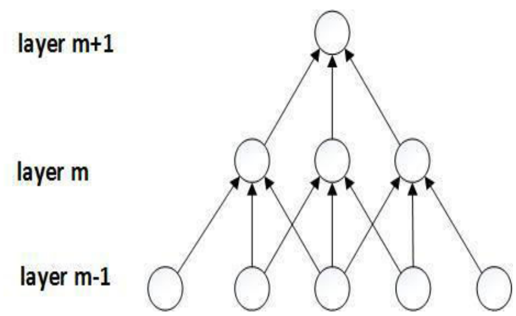


Figure 4.1: Sparse Connectivity [6]

Let us consider layer m-1 as an input retina. It can be seen in the figure that the layer m have receptive fields of width 3 in the input retina and are thus connected only to 3 adjacent neurons in the retina layer [6]. There is similar connectivity between the units in layer m+1 and the layer below. It can be said that their with respect to the input receptive field is larger where as with respect to the layer below their receptive field is 3. There is no response in the each unit to variations which are outside their receptive fields with respect to the retina thus ensuring that the strongest response to a spatially local input pattern is produced by the learnt filter

4.1.2Shared Weights

Every filter h_i in CNNs is duplicated across the complete visual field. The duplicated filters consists of the same parameters i.e. weights and bias that form a feature map.

We can see in fig.4.2 that same feature map contains 3 hidden units. The weights of same color are shared that are constrained to be identical [6]. We can still use gradient descent to learn such shared parameters by altering the original algorithm by a very small margin. When the gradients of the shared parameters are summed, then it gives the gradient of a shared weight. We can detect the features regardless of their location in the visual field by duplicating the units. The huge reduction of the number of free parameters being learnt can lead to weight sharing increasing the learning efficiency. CNNs achieve better generalization on vision problems due to the constraints on these models.

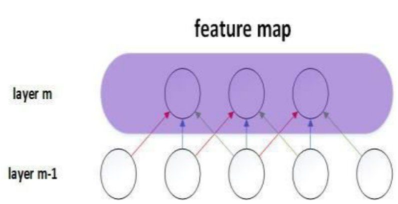


Figure 4.2: Shared Weights [6]

4.4System Design

The fig.4.6 illustrates how the system of retrieval works for this study. The query image is pre-processed and is evaluated with the trained neural network and the regions are clas-sified. It is then matched against the annotation index with images on which the neural network was trained. All the images in the dataset which are similar to the query image are returned to the user based on the number of images required by him. In other words, top N images similar to the query image are retrieved. This section brieflyexplains the major components of the system design

Query Image

The query image is a user input image which he wants to use as a sample to retrieve images from the dataset. The query image can be from any source and need not be from our dataset. An example of query image can be seen in The image has to be pre-processed before it is evaluated by the trained neural network because the dataset on which the network was trained had images pre-processed and works with specific constraints. The image is converted to grayscale and is resized to 28x28 pixels as a part of pre-processing step. Once the query image is a 28x28 pixel grayscale image it can be evaluated with the train model.

Trained Neural Network

I have discussed in section 4.3 how I trained the neural network. The result that is returned after training is a train model which is a Theano function. After the query image is con-verted to grayscale and is resized it is evaluated with the train model. Based on the training results the regions of the query image are classified according to the class labels. This information is stored and is used for matching against the annotation index.

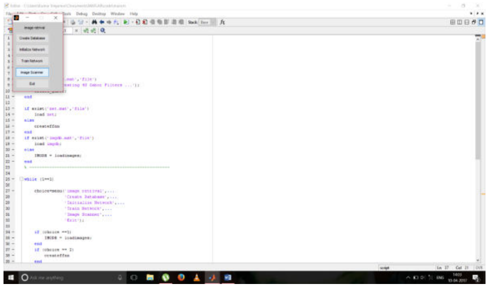
Annotation Index

I built an annotation index with the help of the annotations provided with every image. The index contains the regions that have been annotated and classified by the users

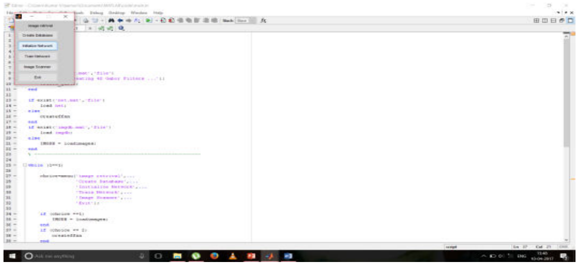
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

for each image. I check for labels from valid regions to get new labels. These labels are added to the index and for each label only once an image is added to the index. This results in an index which contains the information about all the labels present in each image. The annotation index is also used for generating a mapping for labels based on the existing active labels. I maintain a synonym list which contains the synonyms for all the 8 classes that I am using. These synonyms are names used by users to annotate the images. For example in some cases sky is annotated as cloudy sky, clear sky, sky etc. To handle this situation mapping is done for the labels to be mapped to their synonyms. Once the annotation index is ready the dataset is compressed using the annotation index, images and the mapping.

Create data base



Intalize network



Train data

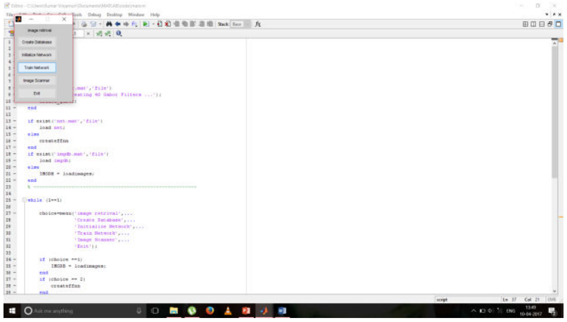
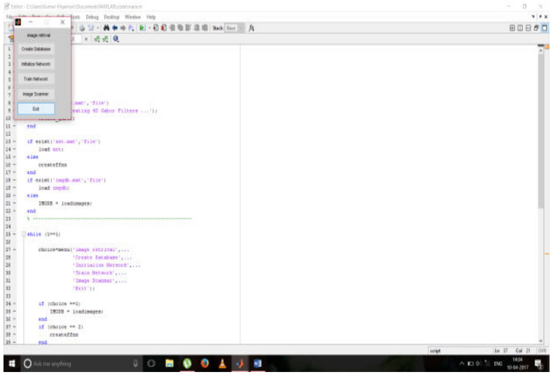
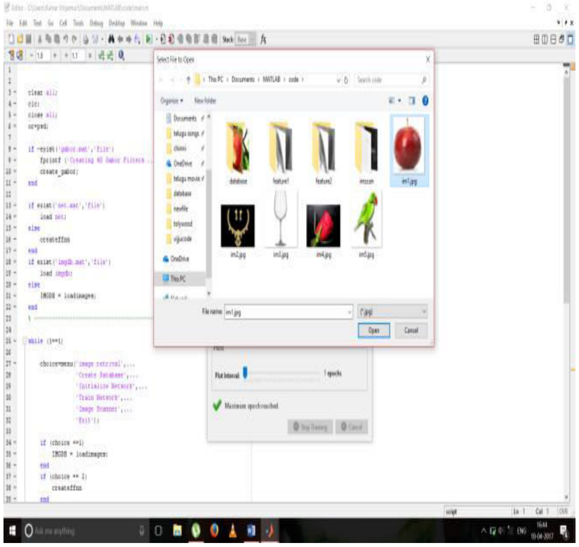


Image scannerexit



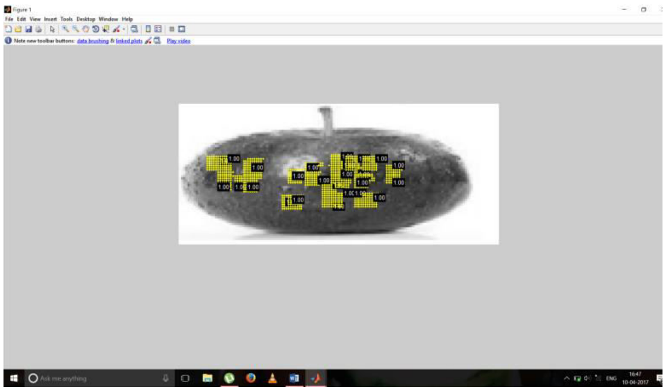
cnn layer

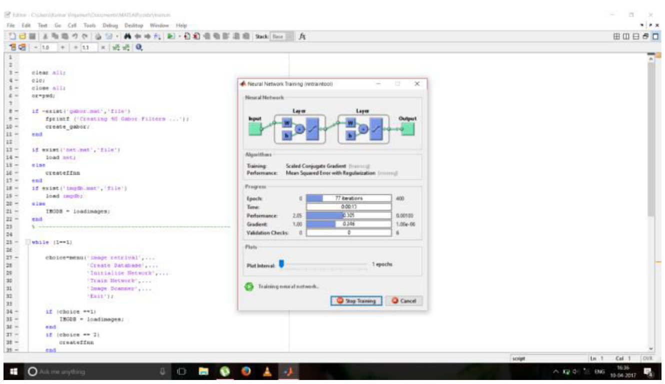
Selection of Input image



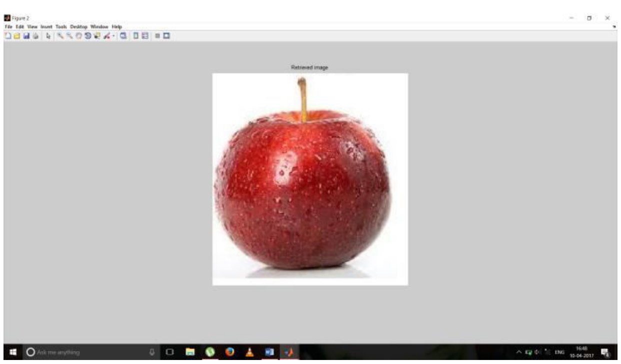
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Weights of input image





Output or retrival image



CONCLUSION

In this work, we evaluate the performances of different layers in deep CNNs for content-based image retrieval system, and propose a multi-level pooling method to obtain object-aware representation. We aggregate features from the low-level and high-level layers to form our final representation, which contains both vision and semantic information. We achieve state-of-art results on several benchmark retrieval datasets and a large-scale practical commodity dataset with the statical parameters compress operation on the final representation.

BIBLIOGRAPHY

[1] Fred´eric´ Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfel-low, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.

[2] James Bergstra, Olivier Breuleux, Fred´eric´ Bastien, Pascal Lamblin, Razvan Pas-canu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010. Oral Presentation.

[3] Yining Deng and B. S. Manjunath. Unsupervised segmentation of color-texture re-gions in images and video. IEEE Trans. Pattern Anal. Mach. Intell., 23(8):800–810, August 2001.

[4] Yahong Han, Fei Wu, Qi Tian, and Yueting Zhuang. Image annotation by input 2013;output structural grouping sparsity. Image Processing, IEEE Transactions on, 21(6):3066–3079, June 2012.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[6] University of Montreal LISA Lab. Deep learning tutorial, 2008.

[7] Ying Liu, Dengsheng Zhang, and Guojun Lu. Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recogn., 41(8):2554–2570, August 2008.

[8] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. Pattern Recogn., 40(1):262–282, January 2007.