International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

RECOMMENDER SYSTEMS IN THE HEALTHY FOOD DOMAIN

THAMATAM CHINNI [1], D.VAJIDAPARVEEN [2]
[1]PG Scholar, Dept of ECE, SIR C.V. RAMAN Institute of Technology & Science, AP, India
[2] Assistant Professor, Dept of ECE, SIR C.V. RAMAN Institute of Technology & Science, AP, India

ABSTRACT: We propose a new dataset for the evaluation of food recognition algorithms designed for dietary monitoring. Each image depicts a real canteen tray with dishes and foods arranged in different ways. Each tray contains multiple instances of food classes. We collected a set of 1,027 canteen trays for a total of 3,616 food instances belonging to 73 food classes. The food on the tray images have been manually segmented using carefully drawn polygonal boundaries. We benchmark the dataset designing an automatic tray analysis pipeline that takes a tray image as input, finds the regions of interest, and predicts for each region the corresponding food class. We experimented three different classification strategies using also several visual descriptors. In the experiments, we have achieved about 79% of food and tray recognition accuracy using Convolutional-NeuralNetworks-based features. The dataset, as well as the benchmark framework, are made available to the research community.

INTRODUCTION

HEALTH care on food and good practices in dietary behavior are drawing people’s attention recently. Nowadays technology can support the users in keep tracks of their food consumption, and to increase the awareness in their daily diet by monitoring their food habits. In the recent years many research works have demonstrated that machine learning and computer vision techniques can help to build systems to automatically recognize diverse foods and to estimate the food quantity [1], [2], [3], [4], [5]. To be useful for dietary monitoring, food recognition systems should also be able to operate on “wild” environments such as restaurants, canteens, and such. Obviously, the fair benchmarking of these systems, requires the availability of suitable datasets that actually pose the challenges of the food recognition task in unconstrained environments.

Recent years there is a Speedy increase in searching engines along with Bing photo search: Microsoft's CBIR engine (Public Company), Google's CBIR machine, note: does not work on all photographs (Public Company), CBIR search engine, via Gazopa (Private Company), Imense Image Search Portal (Private Company) and Like.Com (Private Company), photograph retrieval has turn out to be a challenging mission. The hobby in CBIR has grown because of the retrieval problems, barriers and time consumption in metadata-based structures. We can search the textual statistics very effortlessly with the aid of the present era, but this searching methods calls for humans to describe each pix manually within the database, which isn't possible nearly for very massive databases or for the pics with the intention to be generated mechanically, e.g. Pictures generated from surveillance cameras. It has extra drawbacks that there may be a chance to miss photographs that use unique equal word inside the description of pics. The systems based on categorizing snap shots in semantic classes like “tiger” as a subclass of “animal” can debar the miscatergorization trouble, but it's going to requires extra attempt with the aid of a use to pick out the pix that is probably “tigers” , but they all are categorized handiest as an “animal”. Content-based totally photo retrieval (CBIR) is a software of methods of acquisition, pre-processing, analyzing, representation and also know-how pictures to the picture retrieval trouble, that is the trouble of exploring for digital photographs from massive databases. The CBIR device is opposed to traditional approaches, which is understood on concept-based totally strategies i.e., concept-based totally photo indexing (CBII) [1].

Representation of capabilities and similarity measurements are crucial for the retrieval overall performance of a CBIR system. Various strategies were counseled, however even then, it stays as a challenging undertaking because of the semantic gap gift among the photo pixels and high-stage semantics perceived with the aid of human beings. One favorable approach is ML that objectives to remedy this problem within the lengthy term. Deep gaining knowledge of represents a class of ML methods where numerous layers of information processing steps in hierarchical layouts are utilized for type undertaking and study of features [2]. Deep studying frameworks have attained incredible achievements in image classification. However, the ranking of similar pics is inconsistent with the type of photographs. For type of pics, "black boots," "white boots" and "dark-grey boots" are all boots, but for rating of comparable photographs, if a query photograph is a "black boot," we conventionally want to rank the "darkish grey boot" better than the "white boot." CNNs [2] are a specific type of ANN for managing statistics that capabilities a grid-like topology like, image facts, that is a 2D grid of pixels. CNNs are merely ANNs that involves using convolution as opposed to traditional matrix multiplication operation in at least one in every of their layers. Convolution supports three crucial concepts which can facilitate in enhancing a ML system: parameter sharing, equivariant representations, and sparse interactions. CNNs are eminent for their ability to study shapes, textures, and shades, making this problem appropriate for the application of neural networks.

In this, we investigated an structure of deep studying for CBIR structures through applying a sophisticated deep getting to know gadget, this is, CNNs for reading feature representations from picture information. Overall, our approach is to retrain the pre-trained CNN model, that is, on our dataset. Then, the trained network is used to perform responsibilities: classify items into its appropriate classes and perform a nearest-neighbors analysis to go back the most similar and most relevant photographs to the enter photo [3-4].

BACKGROUND

The use of images in human communication is hardly new. The use of maps and buildings plans to convey information almost certainly dates back to pre-roman times. But, the twentieth century has witnessed unparalleled growth in the number, availability and importance of images in all walks of life. Images now play a crucial role in the fields as diverse as medicine, journalism, advertising, design, education and entertainment. Technology, in the form of inventions such as photography and television, has played a major role in facilitating the capture and communication of image data. But, the real engine of imaging revolution has been the computer, bringing with it a range of techniques for digital image capture, processing, storage and transmission. Once computerized, imaging became affordable and it soon penetrated into areas that were traditionally depending heavily on images for communication, such as engineering, architecture and medicine. Photograph libraries, art galleries and museums, too, began to see the advantages of making their collections available in electronic form.

The creation of the World-Wide Web in the early 1990's, enabling users to access data in a very variety of media from anywhere on the planet, has provided a further massive stimulus to the exploitation of digital images. The number of images on the Web was recently estimated to be between 10 and 30 million. The process of digitization does not in itself make image collections easier to image. Some form of cataloguing and indexing is still necessary to manipulate relevant images. The size of image databases had increased dramatically in recent years. Causes include the development of image capturing devices such as digital cameras and the internet. New techniques and tools need to be proposed with efficient results for sorting, browsing, searching and retrieving images. The text-based approach can be tracked back to 1970's for retrieving images by using annotations. In 1980's content based image retrieval - CBIR was introduced to overcome some disadvantages of the text-based approach.

Content-based image retrieval (CBIR) has become an important practicable technique to support effective searching and browsing of larger and larger collections of unstructured images and videos. Content-based image retrieval - CBIR uses visual content (low-level features) of images such as color, texture, shape, etc. to represent and to index images. These features are described by multi-dimensional vectors called feature vectors that are used in the process of retrieve similar images. Extensive experiments on CBIR show that low-level features not represent exactly the high-level semantic concepts and can fail when used to retrieve similar images. In order to overpass this problem, different approaches aim to propose new methods that use different techniques combined with low level descriptors.

Different CBIR systems have been developed such as simplicity , clue and others. More specifically, the discrepancy between the limited descriptive power of low-level image feature and the richness of user semantics, is referred to as the semantic gap. In order to bridge this gap, different approaches aim to propose new methods by combining low level features and other techniques as textual annotations for creating new descriptors that improve the results in image retrieval. However, the retrieve process become more complex and any method does not warranty the absolute accuracy of results.

LITERATURE SURVEY

Image Retrieval system is an effective and efficient tool for managing large image databases. Content based image retrieval system allows the user to present a query image in order to retrieve images stored in the database according to their similarity to the query image. Content Based Image Retrieval (CBIR) is a technique which uses visual features of image such as color, shape, texture, etc. to search user required image from large image database according to user's requests in the form of a query. In this paper content-based image retrieval method is used retrieve query image from large image database using three features such as color, shape, texture etc. The main objective of this paper is classification of image using K-nearest neighbors Algorithm (KNN).

Automatic linguistic indexing of pictures is an important but highly challenging problem for researchers in computer vision and content-based image retrieval. In this paper, we introduce a statistical modeling approach to this problem. Categorized images are used to train a dictionary of hundreds of statistical models each representing a concept. Images of any given concept are regarded as instances of a stochastic process that characterizes the concept. To measure the extent of association between an image and the textual description of a concept, the likelihood of the occurrence of the image based on the characterizing stochastic process is computed. A high likelihood indicates a strong association. In our experimental implementation, we focus on a particular group of stochastic processes, that is, the two-dimensional multiresolution hidden Markov models (2D MHMMs). We implemented and tested our ALIP (Automatic Linguistic Indexing of Pictures) system on a photographic image database of 600 different concepts, each with about 40 training images. The system is evaluated quantitatively using more than 4,600 images outside the training database and compared with a random annotation scheme. Experiments have demonstrated the good accuracy of the system and its high potential in linguistic indexing of photographic images.

Interests to accurately retrieve required images from databases of digital images are growing day by day. Images are represented by certain features to facilitate accurate retrieval of the required images. These features include Texture, Color, Shape and Region. It is a hot research area and researchers have developed many techniques to use these features for accurate retrieval of required images from the databases. In this paper we present a literature survey of the Content Based Image Retrieval (CBIR) techniques based on Texture, Color, Shape and Region. We also review some of the state-of-the-art tools developed for CBIR.

EXISTING SCHEMES

HSV color space

Basically, there are three properties or three dimensions of color that being hue, saturation and value HSV means Hue, Saturation and Value. It is important to look at because it describes the color based on three properties. It can create the full spectrum of colors by editing the HSV values. The first dimension is the Hue. Hue is the other name for the color or the complicated variation in the color. The quality of color as determined by its dominant wavelength. This Hue is broadly classified into three categories. They are primary Hue, Secondary Hue and Tertiary Hue. The first and the foremost is the primary Hue it consists of three colors they are red, yellow and blue. The secondary Hue is formed by the combination of the equal amount of colors of the primary Hue and the colors of the secondary Hue which was formed by the primary Hue are Orange, Green and violet. The remaining one is the tertiary Hue is formed by the combination of the primary Hue and the secondary Hue. The limitless numbers of colors are produced by mixing the colors of the primary Hue in different amounts. Saturation is the degree or the purity of color. Then the second dimension is the saturation. Saturation just gives the intensity to the colors. The saturation and intensity drop just by mixing the colors or by adding black to the color. By adding the white to the color in spite of more intense the color becomes lighter. Then finally the third dimension is the Value. The value is the brightness of the color. When the value is zero the color space is totally black with the increase in the color there is also increase in the brightness and shows the various colors. The value describes the contrast of the color. That means it describes the lightness and darkness of the color. As similar to the saturation this value consists of the tints and shades. Tints are the colors with the added white and shades are the colors with the added black.

Properties of the HSV color space

Sensing of light from an image in the layers of human retina is a complex process with rod cells contributing to scotopic or dim-light vision and cone cells to photopic or bright-light vision (Gonzalez and Woods, 2002). At low-levels of illumination, only the rod cells are excited so that only gray shades are perceived. As the illumination level increases, more and more cone cells are excited, resulting in increased color perception. Various color spaces have been introduced to represent and specify colors in a way suitable for storage, processing or transmission of color information in images. Out of these, HSV is one of the models that separate out the luminance component (Intensity) of a pixel color from its chrominance components (Hue and Saturation). Hue represents pure color, which is perceived when incident light is of sufficient illumination and contains a single wavelength. Saturation gives a measure of the degree by which a pure color is diluted by white light. For light with low illumination, corresponding intensity value in the HSV color space is also low.

The HSV color space can be represented as a Hexacone, with the central vertical axis denoting the luminance component, I (often denoted by V for Intensity Value). Hue, is a chrominance component defined as an angle in the range[0,2p] relative to the red axis with red at angle 0, green at 2p/3, blue at 4p/3 and red again at 2p. Saturation, S, is the other chrominance component, measured as a radial distance from the central axis of the hexacone with value between 0 at the center to 1 at the outer surface. For zero saturation, as the intensity is increased, we move from black to white through various shades of gray. On the other hand, for a given intensity and hue, if the saturation is changed from 0 to 1, the perceived color changes from a shade of gray to the most pure form of the color represented by its hue. When saturation is near 0, all the pixels in an image look alike even though their hue values are different.

As we increase saturation towards 1, the colors get separated out and are visually perceived as the true colors represented by their hues. Low saturation implies presence of a large number of spectral components in the incident light, causing loss of color information even though the illumination level is sufficiently high. Thus, for low values of saturation or intensity, we can approximate a pixel color by a gray level while for higher saturation and intensity, the pixel color can be approximated by its hue. For low intensities, even for a high saturation, a pixel color is close to its gray value. Similarly, for low saturation even for a high value of intensity, a pixel is perceived as gray. We use these properties to estimate the degree by which a pixel contributesto color perception and gray level perception.

One possible way of capturing color perception of a pixel is to choose suitable thresholds on the intensity and saturation. If the saturation and the intensity are above their respective thresholds, we may consider the pixel to have color dominance; else, it has gray level dominance. However, such a hard thresholding does not properly capture color perception near the threshold values. This is due to the fact that there is no fixed level of illumination above which the cone cells get excited. Instead, there is a gradual transition from scotopic to photopic vision. Similarly, there is no fixed threshold for the saturation of cone cells that leads to loss of chromatic information at higher levels of illumination caused by color dilution. We, therefore, use suitable weights that very smoothly with saturation and intensity to represent both color and gray scale perception for each pixel.

PROPOSED IMPLEMENTATION

Researches in the literature have often focused on different aspects of the food recognition problem. Many works address the challenges in the recognition of food by developing recognition strategies that differ in terms of features and classification methodologies. With respect to the features, the work of He et al. [6] describes the food image by combining both global and local features, while the work of Farinella All the authors are with the Department of Informatics, System and Communication, University of Milano-Bicocca, Italy et al [7] uses a vocabulary built on textons. SIFT and local binary patterns are used in [8], while in [9], the context of where the pictures were taken is also exploited along with the visual features. With respect to the classification strategies, the most widely used are k-NN classifiers [6], [10], and Support Vector Machines [7], [8]. An evaluation of different classification methodologies is reported in [5] where SVM, Artificial Neural Networks and Random Forest classification methods are analyzed. Recently, Convolutional Neural Network (CNN) are also being used in the context of food recognition [11], [12], [13]. Other works in the literature focus on the design of a complete system for diet monitoring in real contexts. Often these systems exploit mobile application for food recognition, assessment, and logging. Examples of such systems are FoodLog [14], DietCam [15], Menu-Match [16], FoodCam [17], and those described in [18], [19], and [10]. Food quantity estimation is very important in the context of a dietary monitoring since on it depends the assessment of the food intakes. Works that tackle this problem are for example [20], [21], [22], [23], [24], [25], [26], [27]. All these works require a reference information to be able to estimate the quantity of food on the plate. This information may came from markers or token for camera calibration, the size of a reference objects (e.g. thumb, or eating tools), or from the specific location where the food is consumed (e.g. canteen). Other works, instead of estimating the amount of food from 2D images, use 3D techniques coupled with template matching or shape reconstruction algorithms [28], [29], [20]. Very few works specifically consider the problem of leftover estimation. Often the problem is theoretically treated as a special case of the problem of food recognition and quantity estimation [23], [18]. Only one work to date explicitly tackles the problem with assessment experiments on a dedicated dataset

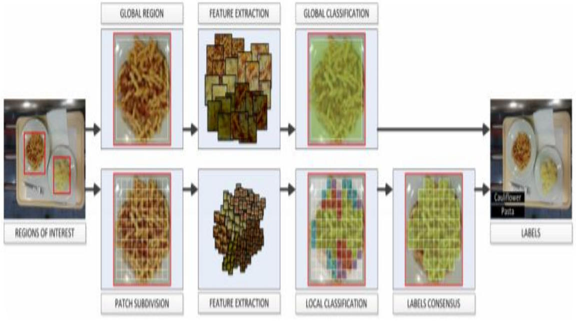


Fig: Processing pipeline for the food classification.

This section describes proposed methodology which employs DConvNet for CBIR system. Working of CNN can be explained as follows: A 2-D convolutional layer applies sliding filters to the input. The layer convolves the input by moving the filters along the input vertically and horizontally and computing the dot product of the weights and the input, and then adding a bias term. A ReLU layer performs a threshold operation to each element of the input, where any value less than zero is set to zero.A max pooling layer performs down-sampling by dividing the input into rectangular pooling regions and computing the maximum of each region.A fully connected layer multiplies the input by a weight matrix and then adds a bias vector.

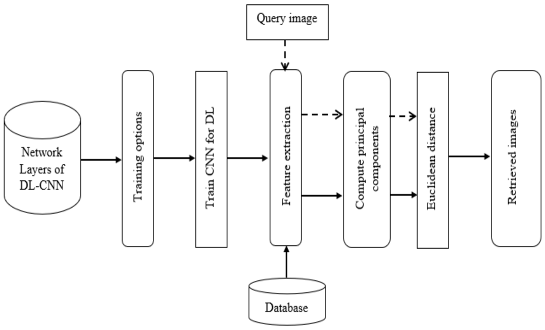


Fig. 5.1. Proposed DConvNet for CBIR system

DL-CNN

According to the facts, training and testing of DL-CNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from [0,1]. Figure 1 discloses the architecture of DL-CNN that is utilized in proposed methodology for CBIR system for enhanced feature representation of word image over conventional retrieval systems.

Convolution layer as depicted in Figure 5.1 is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It’s a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F\left(k_x, k_y, d\right)$.

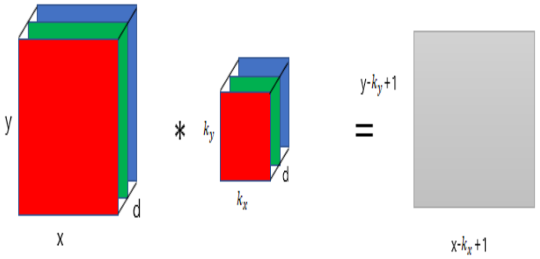
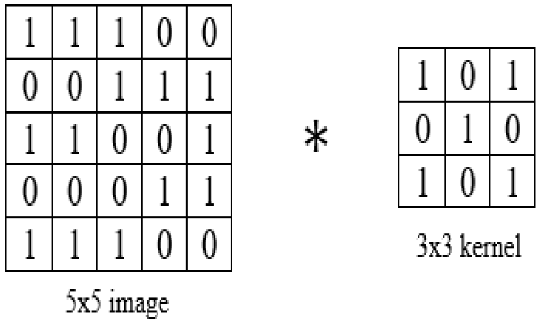


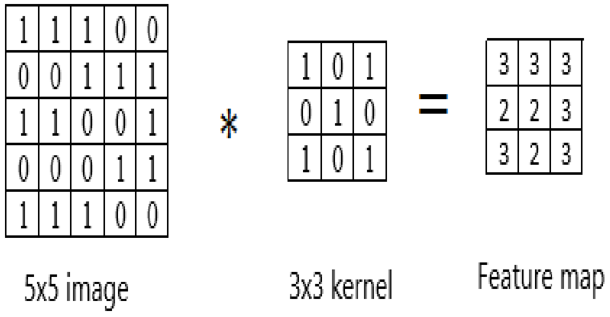
Fig. 5.1. Representation of convolution layer process

The output obtained from convolution process of input image and filter has a size of $C\left((x-k_x+1),\left(y-k_y+1\right),1\right)$, which is referred as feature map. An example ofconvolution procedure is demonstrated in Figure 5.2. Let us assume an input image with a sizeof 5 × 5 and the filter having the size of 3 × 3. The feature map of input image is obtained by multiplying the input image values with the filter values as given in Figure 3.



$$\begin{array}{|c|c|c|c|c|}\hline 1&1&1&0&0\\\hline 0&0&1&1&1\\\hline 1&1&0&0&1\\\hline 0&0&0&1&1\\\hline 1&1&1&0&0\\\hline\end{array} \ast \begin{array}{|c|c|c|}\hline 1&0&1\\\hline 0&1&0\\\hline 1&0&1\\\hline\end{array}$$

(a)



$$\begin{array}{|c|c|c|c|c|}\hline 1&1&1&0&0\\\hline 0&0&1&1&1\\\hline 1&1&0&0&1\\\hline 0&0&0&1&1\\\hline 1&1&1&0&0\\\hline\end{array} \ast \begin{array}{|c|c|c|}\hline 1&0&1\\\hline 0&1&0\\\hline 1&0&1\\\hline\end{array} = \begin{array}{|c|c|c|}\hline 3&3&3\\\hline 2&2&3\\\hline 3&2&3\\\hline\end{array}$$

(b)

Fig. 5.2. Example of convolution layer process (a) an image with size **5 × 5** is convolving with **3 × 3** kernel (b) Convolved feature map

RESULT

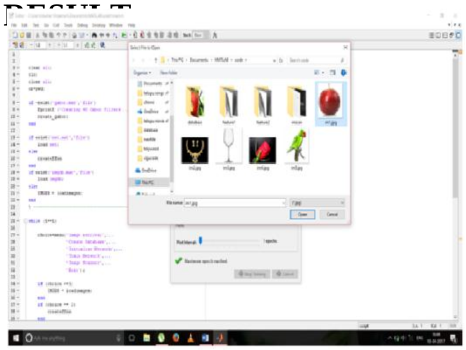


Figure 7.6: Selection of input image

7

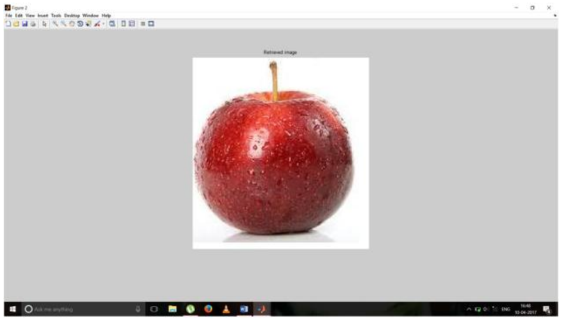


Figure 7.8: Output or retrieval image

CONCLUSION

In the recent years, it has been demonstrated that visual recognition and machine learning methods can be used to develop systems that keep tracks of human food consumption. The actual usefulness of these system heavily depends on the capability of recognizing foods in unconstrained environments. In this paper, we propose a new dataset for the evaluation of food recognition algorithms designed for dietary monitoring. The images have been acquired in a real canteen and depict a real canteen tray with foods arranged in different ways. Each tray contains multiple instances of food classes. We collected a set of 1,027 canteen trays for a total of 3,616 food instances belonging to 73 food classes. The tray images have been manually segmented using carefully drawn polygonal boundaries. We designed a suitable automatic tray analysis pipeline that takes a tray image as input, finds the regions of interest, and predicts for each region the corresponding food class. We evaluated three different classification strategies using several visual descriptors. The best performance has been obtained by using Convolutional-Neural-Networks-based features. The dataset, as well as the benchmark framework, are made available to the research community. Thanks to IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS 10 the way it has been annotated, this database along with the UNIMIB2015 can be used for food segmentation, recognition and quantity estimation.

REFERENCE

[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.

[9] Chih Fong Tsai. Bag-of-words representation in image annotation: A review. ISRN Artificial Intelligence, 2012(1), 2012.

[10] Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. Deep learning for content-based image retrieval: A comprehen-sive study. In Proceedings of the ACM International Conference on Multimedia, pages 157–166. ACM, 2014.

[11] Wikipedia. Convolutional neural network — wikipedia, the free encyclopedia, 2015. [Online; accessed 9-May-2015].

[12] Jianxiong Xiao, Jianxiong Xiao, James Hays, J. Hays, Krista A. Ehinger, K. A. Ehinger, Aude Oliva, A. Oliva, A. Torralba, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. pages 3485–3492. IEEE, 2010.

[13] Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo. Evaluat-ing bag-of-visual-words representations in scene classification. In Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval, MIR ’07, pages 197–206, New York, NY, USA, 2007. ACM.

[14] Dengsheng Zhang, Md Monirul Islam, Guojun Lu, and Jin Hou. Semantic image retrieval using region based inverted file. In Proceedings of the 2009 Digital Image Computing: Techniques and Applications, DICTA ’09, pages 242–249, Washington, DC, USA, 2009. IEEE Computer Society