A Neural Network Model for Automatic Image Annotation and Annotation Refinement: A survey

[1]Alpesh Dabhi, [2]Bhavesh Prajapati
[1]M.E. (Computer science & Engineering), [2]Associate Professor
[1]Computer Science and Engineering, L.D. College of Engineering, Ahmedabad , Gujarat
[2] Computer Engineering, L.D. College of Engineering, Ahmedabad, Gujarat
[1]and.alpesh@gmail.com, [2]b.b.prajapati@gmail.com

***Abstract* - To retrieve an image from large image database is somehow very difficult task of image retrieval system. There are many methods proposed in the past to retrieve an image but still research has been going on to build an efficient method. Image can be retrieve by using visual low-level content such as shape, color, and texture or by using tags or keywords which are described by the semantic meaning of given image. To retrieve images using low-level visual features user needs to give an input as a query image and image retrieval gives set of images which are visually similar to given query image. But it is very difficult for many users to get query image each time which suffice their requirement. Content based image retrieval (CBIR) is a method which retrieves image based on low-level visual features. So to overcome problem of CBIR another method is to classify semantically all the images of the database as keywords. The entire database images are classified as a set of keywords and images can be retrieved based on these keywords. The main advantage of such method is that user can retrieve image in the same manner as they retrieve text document. One method is to manually classify all images; but it is very difficult and time consuming to classify large quantity of images manually, so some sort of automated method is required to perform this task. Automatic Image Annotation (AIA) is an automated method which maps low-level visual features for the high-level semantic features of the given image. This research paper is based on our survey of various AIA methods, where All the AIA methods consist of Artificial Neural Networks as a classification network.**

I.INTRODUCTION

Nowadays; with the frequent and easy access to the digital gadgets such as Digital Cameras and Mobiles etc., information in the form of digital images is increasing. Image based database is increasing leaps and bounds. To retrieve the unique image from the large image database, effective and efficient method is required.

One method is content based image retrieval (CBIR) in which image is retrieved based on low level features like shape, color and texture. In this method user needs to apply query image in CBIR and similar images based on sample query image is retrieve by the system. But there is a semantic gap between low level visual feature and high level semantic concept that are used by the user. In manual image annotation, images are annotated manually by the user, so that images can be retrieved as easy as retrieving text document. This method is accurate but it is also inefficient because of the manual assignment of keywords to image, which is cumbersome and time consuming process. To overcome such problems of manual image annotation and to bridge the semantic gap, research in this area shifted to Automatic Image Annotation [1].

Automatic Image Annotation (AIA) is a technique in which keywords are automatically assigned to image by using some machine learning techniques. The main concept of AIA is to automatically learn semantic concept model from large number of image database and use the concept model to label new images. Basically AIA techniques can be divided into two parts: (1) Visual feature extraction and (2) classification and Annotation. A surveys on different AIA techniques is given in zhang et al[1] and hafidha et al[2]. From various known techniques; to build a classification network for AIA, One method is to learn semantic concept is using Artificial Neural Networks (ANN).

ANNs are information processing system inspired by the ability of the human brain to learn from observation. ANN is a learning network which consists of multiple layers of interconnected nodes, which are also known as neurons or perceptions. An ANN can learn from example and make decision for a new sample. Different from other common classifiers which usually learn one class at a time, ANN can learn multiple classes at a time.

The rest part of paper is arranged in four sections. Section II describes an AIA system. Section III shows various low level visual features. Section IV described how neural network is used in AIA and Section V focus on annotation refinement methods, finely conclusion is given in section VI.

II.AUTOMATIC IMAGE ANNOTATION SYSTEM

Automatic Image Annotation system is mainly divides into two parts: the first part is training a classifier network with use of a training image database, and the second past is testing a network for sample query images. These two parts are described in figure 1.

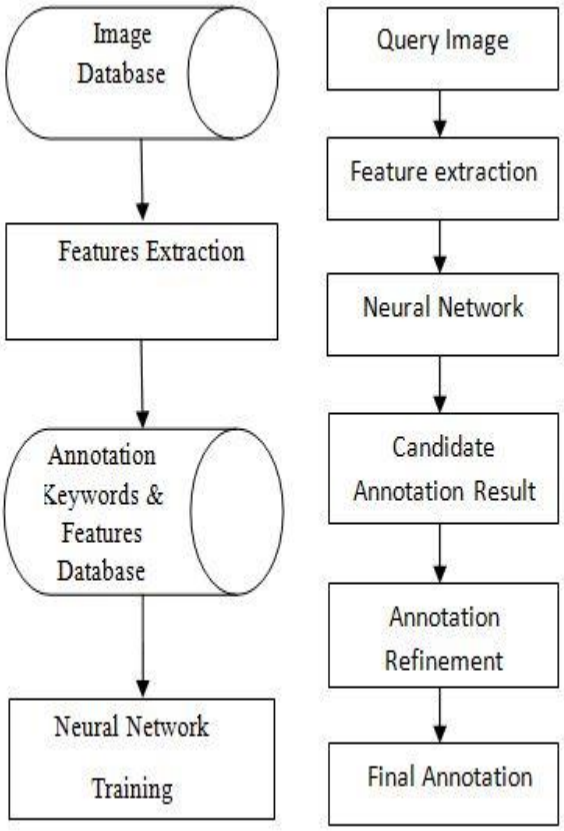

Figure 1: Automatic Image Annotation System

In the first part training images are already annotated by experts using manual annotation. We can train classifier network using this database. In this AIA system we use Neural Network as a classifier network, which can classify more than one class at a time. When a neural network is trained we can use this trained network to classify sample query images, which described in second part of fig. 1. Result produced by neural network is some time not accurate i.e. resulting candidate annotation might have some unnecessary keywords. So with the help of Annotation Refinement network we can refine the candidate annotation result and produce final annotation result.

III.VISUAL FEATURE EXTRACTION

The first step in automatic classification of images is to extract visual low-level features of image. Images in AIA system is represented by low-level features like color, shape and texture. These features are described below:

Color:

Color is the most used visual feature for image retrieval due to the computational efficiency of its extraction. All colors can be represented variable combinations of the three so-called primary colors: red (R), green (G), and blue (B). There are some other color spaces for representing the color feature, such as HSV, LUV, YIQ, etc. In particular, color histogram is one common method used to represent color contents for indexing and retrieval. It shows the proportion of pixels of each color within the image, which is represented by the distribution of the number of pixels for each quantized bin.

Shape:

Shape is one of the most important features for describing the content or object(s) of an image. Compared with color and texture features, shape features are usually described after images have been segmented into regions or objects. The shape representations can be divided into two categories, boundary-based (or edge detection) and region-based. The former uses only the outer boundary of the shape, such as the chain code method, while the later uses the entire shape region. However, to effectively extract shape features depends on segmentation methods.

Texture:

Texture is an important element of images for surface, object identification, and region distinctions. In addition to colors, it has been extracted to classify and recognize objects and scenes. Texture can be regular or random. Most natural textures are random. Regular textures are composed of textures that have a regular or almost regular arrangement of identical or at least similar components. Irregular textures are composed of irregular and random arrangements of components related some statistical properties.

This feature also can be further divided into local features and global features. Local features are the computed features over the segmented regions of the image whereas Global features are the computed feature over the entire image. Bouyerbou et al[2] survey on the hybrid methods, which combine both global and local features for image representation. A feature vector, is a measurement of different features of image, would be the input of classification network. To create a feature vector using local features, first image shall segment into different region before extracting features. Various image segmentation techniques proposed in the past but still not found proper method. Computational cost of local method is higher compared to global method because in global method no need to segment an image.

IV.IMAGE ANNOTATION USING ARTIFICIAL NEURAL NETWORK

Artificial neural network is a learning network that can learn from examples. When network is trained by learning methods with training sample data, we can make decision for new sample. Generally ANN is a multi-layer network build by collection of interconnected nodes called neuron. Figure 1 shows a three layer neural network. First layer is input layer which has neurons equal to dimension of input sample. Second layer is hidden layer, in hidden layer the choice of number of hidden layer and the number of neuron at each layer is open issues in ANN approaches [3] and the output layer contain neuron equal to number of classes. Each connecting edge between neuron of different layer is associated with weight. An activation function which generates output based on weight of connecting edge and output of previous layers neurons. Learning method is needs to train neural network such as back-propagation method.

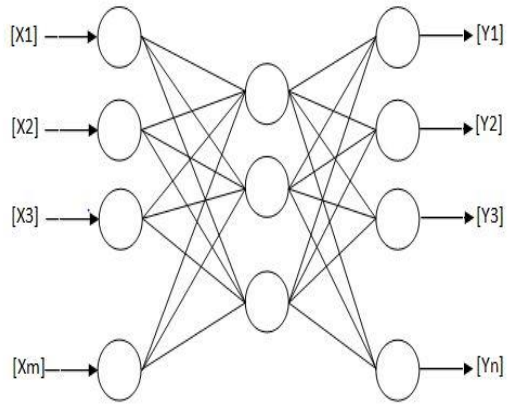

Figure 1: Three layer feed-forward neural network

Neural Network models have certain common characteristics. They are given a set of inputs X=(X1, X2,…,Xm) and their corresponding set of outputs Y=(Y1, Y2,…, Yn) for a certain process. Here the input X is a visual feature vector of the image and output Y is a keyword vector which is use to label image. The output produce by a neuron for given input X is calculated by following formula:

$$Y = f(WX + B) \qquad (1)$$

Here, W is the weight matrix, which is the weight matrix of connection link weight (synaptic weight) between neurons. B is a bias vector of the layer of network. The final output depends on the transfer function f (.), which process the incoming information from other neurons. With the correct choice of parameter W and B we can get desired target output. This choice is done in training phase of network. Training a neural network means adapting its connections so that the model gives the desired computational behavior for all inputs.

Many models have proposed to train neural network for image annotation in the past. Chen et. al. [3, 4] proposed a neural network model with adaptive structure for image annotation. In this model adaptive structure enables both global and local visual features. This model is divided into two networks, Recognition Network and Correlation Network. Recognition network has set of

sub network and the number of sub network is determined by the number of segmented region of an image. Recognition network produces a keyword vector for the image and Correlation network enhances annotation performance by using the keyword correlation information. Both genetic algorithm and back propagation algorithm are used to train recognition network whereas only back propagation algorithm is used to train correlation network and it is experimented on a synthetic image dataset. In the process of image annotation, image is first segmented into several regions which create a visual feature vector for each region. The input layer of recognition network receives the feature vector of a region and output layer would generate a keyword vector. This keyword vector indicates which keyword should be selected to label the input region. Finally correlation network receives keyword vector from recognition network and refine annotation using keyword correlation information.

Pankaj savita et. al.[5] propose a novel technique to image annotation using neural network. In this work they use a neural network with 2 hidden layer each hidden layer has 25 neuron. The number of neuron used in input and output layer is 50 and 2 respectively. This model is experimented on 50 images and result show that neural network classifier outperform conventional (Discrete Sine Transform) DST technique.

Mustapha Oujaoura et. al.[6] presented an image annotation system using region growing as image segmentation algorithm, moments and Multilayer Neural Network. For input feature vector for neural network they use hu moments, legendre moment and Zernike moment as feature extraction methods. This method is experimented on ETH-80 dataset.

Yufeng zhao et. al.[7] proposed a novel image annotation scheme based on neural network. In this work; for discovering the latent contextual correlation among the keywords, Latent semantic analysis (LSA) is employed. The proposed annotation method is composed of three parts. First, latent semantic analysis is introduced to reveal the keyword correlation. Second, neural network is obtained for mapping visual content of image to textual keyword with the use of training data and third, given test image, the learnt neural network is able to provide the keywords to be annotated. Annotation scheme was experimented on corel dataset and result show that the high annotation accuracy can be achieved at image level.

Fabio del et. al.[8] use a neural network for automatic classification of high resolution satellite images. In this work they first assessed and optimized the neural network approach for pixel based classification of a single very high resolution image and then they investigated on the capability of supervised neural network in providing automatic classification on a collection of image.

V.ANNOTATION REFINEMENT

Result obtained by classification network may not be satisfactory because it may contain noise i.e. keywords which are not related to the image. It is necessary to refine the current annotation result because many of annotation keywords are inappropriate for image content. Many researchers have strived to invent novel algorithms of automatic annotation to improve the quality of annotation [9].

In order to fully utilize the confidence scores of the candidate annotations C. Wing et. al. [10] reformulates the image annotation refinement process as a graph ranking problem and solves it with random walk with restart. In this method graph is constructed for candidate annotation, in which each candidate annotation w_i is considered as a vertex of graph G. All vertexes are connected with proper edge weight based on co-occurrence similarity as follow:

$$sim(w_i, w_j) = \begin{cases} \frac{num(w_i, w_j)}{\min(num(w_i), num(w_j))} & num(w_i, w_j) > 0 \\ 0 & num(w_i, w_j) = 0 \end{cases} \quad (2)$$

Where, $num(w_i)$ is defined as the number of images annotated by annotation w_i , and $num(w_i, w_j)$ is defined as the number of images annotated by both w_i and w_j. After that Random Walk with Restart (RWR) algorithm is applied to graph for re-rank the candidate annotation. In RWR algorithm, the initial probability of all the keywords is considered as the restart vector $v(i)$. Starting from $v(i)$, at each time-tick there are two choices. One is to randomly choose an available edge to follow; the other choice is to jump to w_j with probability $c \times v(i)$, where c is the probability of restarting the random walk.

Yae cao et. al.[11] proposed a neural network based method to combine the measures of word to word semantic similarity for image annotation. In this model candidate annotation is generated by max-min posterior pseudo probability (MMP) and this candidate annotation is refined using feed-forward neural network based method. Random walk with restart method is used to refine the annotation. Particle swarm optimization (PSO) is used to train a neural network for achieving the optimal annotation accuracy. The image annotation experimented on the corel5k dataset and results show that the proposed method is effective and promising. J. Liu et. al. [12] developed a novel method to estimate the word correlation based on the improved Nearest Spanning Chain (NSC). Using this enhanced word correlation they utilized the graph ranking method to perform the annotation refinement.

VI.CONCLUSION

Automatic Image Annotation (AIA) gives better performance in image retrieval system. In image retrieval system, image can be easily retrieve using keywords if images are annotated by appropriate keywords. In AIA method image is automatically classify by keywords using classification network. In this paper, we focused on a neural network model as a classification network of AIA system. Neural Network gives better performance when image is classifies by more than one class i.e. keywords. We can easily train neural network than other classification network and once a neural network is trained we can easily annotate sample images. It is not necessary that AIA system gives satisfactory result each time, so various annotation refinement algorithms proposed to refine the result of classification network. Overall AIA system is a very challenging task and research has been going to develop efficient AIA system.

REFERENCES

[1] D. Zhang, Md. M. Islam, G. Lu, "A review on automatic image annotation techniques" ELSEVIER Pattern Recognition 45 (2012) 346-362

[2] H. Bouyerbou, S. Oukid, N. Benblidia, K. Bechkoum, "Hybrid Image Representation Methods for Automatic Image Annotation: A Survey" ICSES 2012- International Conference on Signals and Electronic Systems WROCLAW, POLAND, sep-2012

[3] Z.chen, H. Fu, Z. Chi, D. Feng, "A Neural Network Model with Adaptive Structure For Image Annotation", in 11th Int. Conf. Control, Automation, Robotics and Vision, Singapore, Dec. 2010.

[4] Z.chen, H. Fu, Z. Chi, D. Feng, "An Aduptive Recognition Model for Image Annotation", IEEE Transaction on Systems, Man, and Cybernetics—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 6, Nov. 2012

[5] Pankaj Savita, Deepshikha Patel, Amit Sinhal, "A Novel Technique to Image Annottion using Neural Network" International Journal of computer Applications & Information Technology Vol.2, Issue II feb-March 2013

[6] M. Oujaoura, B. Minaoui, M. Fakir, "Image Annotation using Moments and Multilayer Neural Networks", Special Issue of Internationa Journal of Computer Applications (0975-8887) on Software Engineering, Database and Expert systems-SEDEXS, sep-2012

[7] Y. Zhao, Y. Zhao, Z Zhu, J. Pan, "A Novel Image Annotation Scheme based on Neural Network" Eighth International Conference on Intelligent System Design and Applications

[8] F. Frate, F. Pacifici, G. Schiavon, C. solimini, "Use of Neural Networks for Automatic Classification from High-resolution Images", IEEE Transactions on Geoscience and Remote sensing, VOL. 45, NO. 4, april-2007

[9] Haiyu Song, Xiongfei Li, Pengjie Wang, "Image Annotation Refinement Using Dynamic Weighted Voting Based on Mutual Information" JOURNAL OF SOFTWARE, VOL. 6, NO. 11, NOVEMBER 2011

[10] C. Wang, F. Jing, L. Zhang, H. Zhang, "Image Annotation Refinement using Random Walk with Restarts" *MM'06*, October 23–27, 2006, Santa Barbara, California, USA. ACM 1-59593-447-2/06/0010

[11] Y. Cao, X. Liu, J. Bing, L. Song, "Using Neural Network to Combine Measure of Word semantic Similarity for Image Annotation", Proceeding of the IEEE International Conference on Information and Automation, Shenzhen- China June-2011

[12] J. Liu, M. Li, Q. Liu, H. Lul, S. Mal, "Image Annotation Refinement using NSC-Based Word Correlation" ICME 2007