

A Neural Network Approach to Improve the Efficiency of Image Annotation

Pankaj Savita

Prof. Deepshikha Patel

Prof. Amit Sinhal

Abstract: Image Annotation is one of the most challenging problems in Computer Vision today. The manual annotation of images is not only expensive, but also time consuming and sometimes inaccurate as well one of the biggest applications of Automatic Image annotation is in the field of image search and retrieval. There is a huge urge in the Computer Vision community today to find ways to automatically annotate images and it is with this motivation that we propose Neural Network based a novel approach to the problem of image annotation in this paper. These approach are applied to the Image data set. Our main work is focused on the image annotation by using multilayer perceptron, which exhibits a clear-cut idea on application of multilayer perceptron with special features. By using this algorithm we can save more memory space, and in case of web applications transferring of images and download should be fast.

Keywords: Image Annotation, Neural Network, MLP

1. INTRODUCTION

Image annotation can be defined as the process of modeling the work of a human annotator when assigning words to images based on their visual properties. up to now most of the image annotation systems are based on the combination of Image analysis and statistical machine learning techniques. To improve retrieval accuracy, the research focus has been shifted from designing sophisticated low level feature extraction algorithm to reducing the semantic gap between the visual features and the richness of human semantics. Nowadays the number of digital images is growing with an incredible speed. Describing images by their semantic contents can facilitate users to index, retrieve, organize and interact with huge data using existing text searching techniques. As the majority of the images are barely documented, current research on semantic image retrieval is closely related automatic image annotation (auto-annotation) that works toward finding a solution to the problem of automatically linking keywords to an unlabelled image. Image annotation or automatically annotate images with keywords is a solution to this problem. It is based on some machine learning techniques, which learn the correspondence between visual features and semantics of images. That is, image annotation systems can recognize or classify visual features into some pre-defined classes. Figure 1 shows a general architecture of image annotation systems. The segmentation component partitions images into local contents via either some block or region based method. Then, the feature extraction component extracts low-level features from the segmented images. That is, each segmented block

or region is represented by feature vectors. Next, the annotation component assigns the (low-level) feature vectors to some pre-defined categories. This performs like the pattern classification task. Finally, the post-processing component (dependent on the application) uses the output of the annotation component to decide on some recommended action for the final decision.

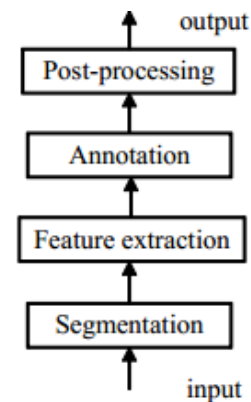


Fig. (1). Block diagram of an image annotation system.

The basic premise of auto-annotation approaches is that a model describing how low-level image features like color, texture and shape are related to keywords, can be learnt from a training set of images. Obtained model is then applied to un-annotated images in order to automatically generate keywords that describe their content. Usually, the keywords with the highest probability are chosen to annotate the image. For solving the problem of automatic image annotation, many different approaches have been used. Here we will

mention some referent methods to point out different approaches used for automatic image annotation. Methods based on translation model and several extensions have assumed auto-annotation to be analogous to translation problem between languages. Models which use Latent Semantic Analysis transform the features to a vocabulary of visual terms, which represent a purely visual language. Renewal methods based on classifications are used for classifying images into a large number of categories. Under the assumption that the basic goal of annotation is to facilitate and improve image retrieval

2. IMAGE SEGMENTATION

In general, visual content of an image can be represented by either global or local features. Global features take all the pixels of an image into account. Color histograms, for examples, can be extracted to represent or describe the global color content of images. In this case, an image can be described that it contains 40% of blue, 37% of yellow and so on. However, as global features consider the visual features of the whole image, they cannot completely describe different parts of an image. On the other hand, image segmentation into local contents (i.e. different regions or areas) is able to provide more detailed information of images. In general, there are two strategies for extracting local features. The first one is to partition a set of fixed sized blocks or tiles (see Fig. 2 for some examples) and the second for a number of variable shaped regions of interest. After performing block and/or region based segmentation, low-level features can be extracted from the tiles or regions for local feature representation.

3. IMAGE LOW-LEVEL FEATURES

In general, low-level features such as color, texture, shape, and spatial relationship are extracted to represent image features.

3.1. Color

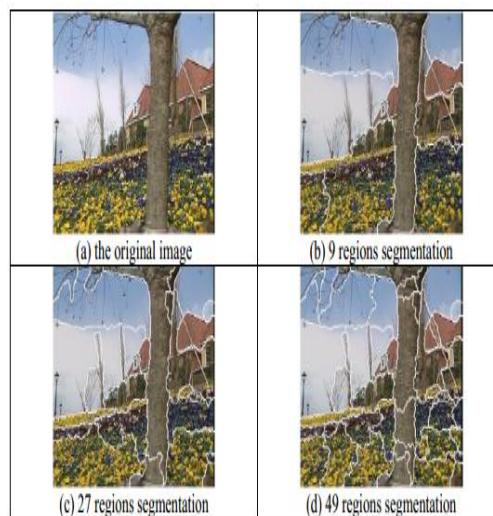
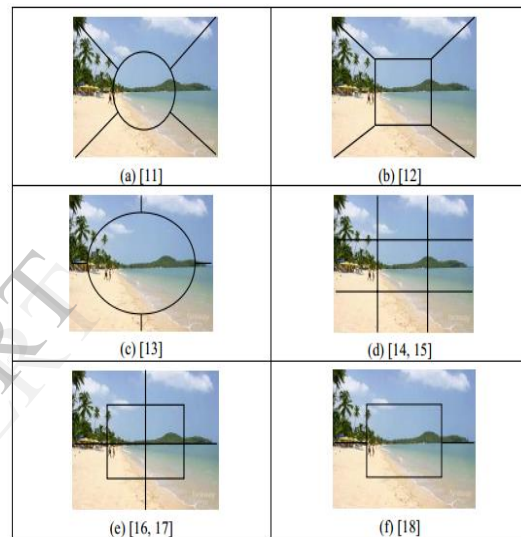
Color is the most used visual feature for image retrieval due to the computational efficiency of its extraction. All colors can be represented variable combinations of the three so-called primary colors: red (R), green (G), and blue (B). There are some other color spaces for representing the color feature, such as HSV, $L^*u^*v^*$, 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.

3.2. 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.

3.3 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



Fig(2): Shape after segmentation 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.

3.4. Spatial Relationship

Objects and the spatial relationships (such as left of, inside, and above) among objects in an image are used to represent the image content. That is, an image can be divided into a number of sub-blocks and color, texture, and/or shape features are extracted from each of the sub-blocks. Then, we can project them along the x and y axes, such as 'left/right', 'below/above' relationships between them. Koetal. consider spatial color histograms which show better performances than the traditional one, i.e. global color histogram

4. LITERATURE SURVEY

Vailaya et al.[9] proposed a hierarchical classification scheme to first classify images into indoor or outdoor categories, then, outdoor images are further classified as city or landscape; finally, landscape images are classified into sunset, forest, and mountain classes. In other words, three Bayes classifiers are used for the three-stage classification.

Wang and Li[15] propose Semantics sensitive integrated matching on pattern analysis and machine intelligence.

J.Li [5] assumed that image annotation could be viewed as analogous to the cross-lingual retrieval problem and proposed the cross-media relevance model for this problem

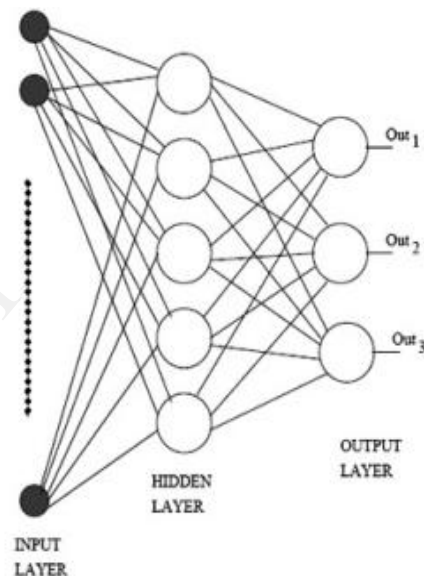
Blei and Jordan[12] propose location (LDA) model which finds conditional relationships between latent variable representations of image regions and words. This model is compared with a Gaussian-multinomial mixture model and a Gaussian-multinomial LDA model.

Huang et al.[6] construct a classification tree for hierarchical 11-category classification. They report that using the color histogram performs better than the general histogram, and the classification tree outperforms the traditional nearest neighbour classifier.

5. ARTIFICIAL NEURAL NETWORKS

Neural networks (or artificial neural networks) learn by experience, generalize from previous experiences to new ones, and can make decisions. A neural network can be thought of as a black box non-parametric classifier. That is, different from naïve Bayes. We do not need to make assumptions about the distribution densities. Neural networks are therefore more flexible. A

multilayer perceptron (MLP) network consists of an input layer including a set of sensory nodes as input nodes, one or more hidden layers of computation nodes, and an output layer of computation nodes. The input nodes/neurons are the feature values of an instance, and the output nodes/neurons (usually lying in the range) represent a discriminator between its class and all of the other classes. That is, each output value is a measure of the network's confidence that the class corresponding to the highest output value is returned as the prediction for an instance. Each interconnection has associated with it a scalar weight which is adjusted during the training phase. Figure 2 shows an example of a three-layer feed-forward network having input, output, and one hidden layers



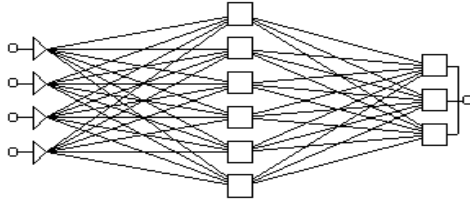
Fig(3):the three-layer neural network

6. PROPOSED TECHNIQUES FOR IMAGE ANNOTATION

6.1 Multilayer perceptron algorithm

This section presents the architecture of a feed word neural network that is used to compress image in the research works. Multilayer-perceptron algorithm is a widely used learning algorithm in Artificial Neural Networks. The Feed-Forward Neural Network architecture is capable of approximating most problems with high accuracy and generalization ability. This algorithm is based on the error-correction learning rule. Error propagation consists of two passes through the different layers of the network, a forward pass, and a backward pass. In the forward pass the input vector is applied to the sensory nodes of the network and its effect propagates through the network layer by layer. Finally a set of outputs is produced as the actual response of

the network. During the forward pass the synaptic weight of the networks are all fixed. During the back pass the synaptic weights are all adjusted in accordance with an error-correction rule. The actual response of the network is subtracted from the desired response to produce an error signal. This error signal is then propagated backward through the network against the direction of



Input layer Hidden Layer output Layer

synaptic conditions. The synaptic weights are adjusted to make the actual response of the network move closer to the desired response.

6.2 Algorithm:

The algorithm for Perceptron Learning is based on the back-propagation rule discussed previously. This algorithm can be coded in any programming language, and in the case of this tutorial, Java for the applets. In this case we are assuming the use of the sigmoid function $f(net)$ described earlier in the tutorial. This is because it has a simple derivative.

Algorithm:

1. Initialize weights and threshold.

Set all weights and thresholds to small random values.

2. Present input and desired output

Present input $\mathbf{X}_p = x_0, x_1, x_2, \dots, x_{n-1}$ and target output $\mathbf{T}_p = t_0, t_1, \dots, t_{m-1}$ where n is the number of input nodes and m is the number of output nodes. Set w_0 to be $-\phi$, the bias, and x_0 to be always 1. For pattern association, \mathbf{X}_p and \mathbf{T}_p represent the patterns to be associated. For classification, \mathbf{T}_p is set to zero except for one element set to 1 that corresponds to the class that \mathbf{X}_p is in.

3. Calculate the actual output

Each layer calculates the following:

$$y_{pj} = f[w_0x_0 + w_1x_1 + \dots + w_nx_n]$$

This is then passes to the next layer as an input. The final layer outputs values o_{pj} .

4. Adapts weights

Starting from the output we now work backwards.

$w_{ij}(t+1) = w_{ij}(t) + \tilde{n}b_{pj}o_{pj}$, where \tilde{n} is a gain term and b_{pj} is an error term for pattern p on node j .

For output units

$$b_{pj} = ko_{pj}(1 - o_{pj})(t - o_{pj})$$

For hidden units

$$b_{pj} = ko_{pj}(1 - o_{pj})[(b_{p0}w_{j0} + b_{p1}w_{j1} + \dots + b_{pk}w_{jk})]$$

where the sum(in the [brackets]) is over the k nodes in the layer above node j .

7. CONCLUSION

Automatic image annotation has emerged as an alternative which can enhance image management and retrieval. The aim is to annotate image with concepts of a higher semantic level, which will correspond to keywords which users intuitively use during image retrieval. It is hard to infer high-level semantics from the image features, because it is necessary to explore all image objects and their relations, and include knowledge necessary for semantic interpretation of overall image. In this paper, the KRFPN formalism based on Fuzzy Petri Net theory was used for knowledge representation. This representation uses simple graphical notation with just a few types of elements and has a well-defined semantics so the model is easily understood. The well-defined inference algorithms can be used for image annotations at various semantic levels of abstraction.

REFERENCES: -

- [1] K. Zagoris, S. Chatzichristofis, N. Papamarkos, and Y. Boutalis, "Img (anaktisi): a web content based image retrieval system," in SISAP, 2009, pp. 154–155.
- [2] S. A. Chatzichristofis, Y. S. Boutalis and Mathias Lux, "Img (rummager): An interactive content based image retrieval system," in SISAP, 2009, pp. 151–153.
- [3] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, 2008 vol. 40(2), pp. 1–60.
- [4] X. Qi and Y. Han, "Incorporating multiple svms for automatic image annotation," Pattern Recognition, 2007, vol. 40, no. 2, pp. 728–741.

- [5] J. Li and J. Wang, "Real-time computerized annotation of pictures," in Proceedings of the 14th annual ACM international conference on Multimedia. ACM, 2006, p. 920.
- [6] J. Huang, S. Kumar, and R. Zabih, "An automatic hierarchical image classification scheme," in Proceedings of the sixth ACM international conference on Multimedia. ACM New York, NY, USA, 1998, pp. 219–228.
- [7] M. Szummer and R. Picard, "Indoor-outdoor image classification," in Proceedings of the 1998 International Workshop on Content-Based Access of Image and Video Databases (CAIVD'98), 1998, p. 42.
- [8] O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram based image classification," IEEE transactions on Neural Networks, 1999, vol. 10, no. 5, p. 1055.
- [9] Dirichl, A. Vailaya, M. Figueiredo, and H. Zhang, "Image classification for content-based indexing," IEEE Transactions on Image Processing, 2001, vol. 10, no. 1, pp. 117–130.
- [10] Ghoshal, E. Chang, K. Goh, G. Sychay, and G. Wu, "Cbsa: content based soft annotation for multi-model image retrieval using Bayes point machines," IEEE Transactions on Circuits and Systems for Video Technology, 2003, vol. 13, no. 1, pp. 26–38.
- [11] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Transactions on pattern analysis and machine intelligence, 2000, vol. 22, no. 12, pp. 1349–1380.
- [12] Blei and Jordan, "Visual seek: a fully automated content-based image query system," in Proceedings of the fourth ACM international conference on Multimedia. ACM, 1997, p. 98.
- [13] A. Yavlinsky, Behold: a content based image search engine for the World Wide Web. Citeseer, 2006.
- [14] S. A. Chatzichristofis, Y. S. Boutalis, and M. Lux, "Selection of the proper compact composite descriptor for improving content based image retrieval," in SPPRA, 2009, pp. 134–140.
- [15] J. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics sensitive integrated matching for picture libraries," IEEE Transactions on pattern analysis and machine intelligence,