A REVIEW OF THE METHODS FOR QUALITATIVE UNDERSTANDING OF A SCENE IMAGE

Janhavi Shirke¹ and N. M. Shahane²

¹Student, Department of Computer Engineering

²Associate Professor, Department of Computer Engineering
K. K. Wagh College of Engineering- Nashik, Maharashtra, India.
{Email: ¹borse.janhavi@gmail.com }

Abstract

Ambiguity in nature gets preserved in its captured images also. Preserving it while classifying an image, is a challenging task. A scene image also can belong to multiple categories at a time which makes a task of classification much more difficult and often leads to classification errors. Binary classification fails to capture this ambiguity while classifying the scene image into one of mutually exclusive classes. Fuzzy classifier handles this problem by considering fuzzy membership with non-mutually exclusive class categories. This paper provides a review of the existing methods for qualitative understanding of a natural scene image.

Keywords: Fuzzy membership, binary classification, fuzzy classifier.

1. Introduction

A scene image consists of many categories exhibiting different semantic meanings and hence they can be discriminated. It means that all the patterns belonging to a particular category, although they are different, follow similar characteristic behaviour. Depending on the above discussion some may conclude that the natural scene images are conveying multiple semantic meanings and hence their classification task is a multi label one. An image can be assigned a single label or multiple labels depending upon its characteristics. Many of the approaches deal with scene understanding problem as multi label classification problem. All those approaches try to solve the problem based on the assumption that scene categories are mutually exclusive. It is equivalent to the task of assigning a single image to multiple class labels which seems to be quite illogical and unnatural. It is because the real scenario is exactly opposite.

In reality scene categories are overlapped with each other which depict ambiguity/uncertainty in the nature. Hence it is more natural and logical to rank the objects in the images according to their degree of membership rather than assigning an image to multiple class labels.

For example, in an image containing two objects, mountain & ocean, it is hardly possible to locate the boundaries of each class representing the two objects. This is because the boundaries are no more sharpened as in case of synthetic images or images containing manmade objects. Rather they are uncertain due to presence of non-mutually exclusive natural objects in a scene image.

In the literature it is found that some uncertainties can

be represented using probability theory. The ambiguity can be best modeled by using fuzzy membership functions [1]. It specifies the degree of membership of an object to a particular class label.

Again, one can note that, as all scene images are captured from large distances to fit the entire scene in the images, most of the details are lost as it is moved away. That is the color component, shape and size of natural objects cannot be quantified. So what remains is the texture component and hence in case of natural scene images we assume that if an image is represented and described by its texture, it would result in better discriminative analysis of its objects.

2. Supervised Learning

In *supervised learning* [2], input space is characterized by the most discriminating image features and an input image is represented by its corresponding feature vector. An output space is the set of labels out of which any subset may be associated with an unseen image depending on its input characteristics. The task is to model a classifier function by analyzing the training samples from input data for which associated set of labels are known. The trained classifier or model is then used to classify the unseen data samples. The job is to predict the set of labels for an unseen image based on the function built using training images.

3. Fuzzy Membership

The only purpose for which fuzzy set theory came into existence is to capture uncertainty of certain type [1]. It was gifted by the great scientist Zadeh.

A fuzzy set A is a set with imprecise boundary as

compared to the crisp set which has definite boundary. Membership of a variable x to the set A is not being a binary decision depending on whether x is relevant to A or not, but it is a matter of degree. It states by how much degree a variable x belongs to set A . Hence the values a variable x can take necessarily lies between the interval $[0, 1]$.

A fuzzy set [1] is uniquely and completely defined by its characteristic function μ_A or A .

$$\mu_A: X \rightarrow [0, 1] \text{ or} \\ A: X \rightarrow [0, 1]$$

It is called a membership function which maps each element of a given universal set X into real numbers of the interval $[0, 1]$. The universal set is always a crisp set.

The following section contains some recent approaches proposed in the literature focusing on the problem of scene understanding.

4. Related Works

Section 2.1 enlists the methods which assume that classes contained in scene images are not overlapping (mutually exclusive) and convert the scene understanding problem into multi label classification problem. Section 2.2 enlists the methods which assume that the classes in scene images are overlapped. Section 2.3 enlists the methods which focus on the tasks such as feature extraction, representation and description.

Scene understanding as Multi label Problem

Gjorgji Madjarov & Dragi Kocev [3] studied 12 different multi label learning methods from three different types such as problem transformation, algorithm adaptation and ensembles based on the underlying algorithm used such as Support Vector Machines [2] (SVM), decision trees and nearest neighbor. They critically evaluated their performances using metrics like hamming loss, one error, precision, recall, F-measure, macro & micro recall/precision etc. depending on example-based, label based, ranking based or efficiency based measures. It is found that SVM based methods give best performance in accuracy for the domains with large number of features as compared to decision trees & nearest neighbors. But decision trees are better than SVMs for large datasets. Rather SVMs are good for smaller datasets.

It is also found that Hierarchy of Multi label classifiers (HOMER) was best when evaluated on recall (completeness) while Random Forest of Predictive Clustering Trees (RF-PCT) was best when evaluated on precision (exact predictions). Further Classifier Chains (CC) & Binary Relevance (BR) are ranked next according to their performances evaluated on selected measures. We are only interested in comparing the multi label methods for scene dataset. It can be seen that for scene dataset Ensemble of Classifier Chain (ECC) performs best for

precision, recall, F-measure & accuracy measures. Also ranking loss & one error is lesser when BR is used. It means that for scene dataset RF-PCT & HOMER approaches do not work well.

Min-Ling & Zhi-Hua [4] studied the multi label problem with the assumption that the multi label learning algorithms are categorized depending on the label correlations they use either first order, second order or higher order. First order means they ignore the existence of other class labels thus making the classification task simple binary one, e.g. Multi Label Decision Trees (ML-DT), Multi Label K-nearest neighbours (ML-KNN) & BR. Second order algorithms exploits label correlations pair wise, e.g. Calibrated Label Ranking (CLR), Rank SVM, Collective Multi label Classifier (CML) and higher order algorithms check influence of one label on rest all of the labels, e.g. CC, Random k-Label sets (RAKEL). Most of the algorithms use binary classification as their intermediate method where all of them solved the multi label problem by combining solutions of many binary problems using various state-of-art methods like AdaBoost, stacked aggregation, Error correcting output codes (ECOC).

It is found that multi label learning,

1. Models the complex semantics in output space.
2. Assumes relevance ordering of each class label such that a binary decision in classification is converted into an ordered/graded membership.

Random Forest of Predictive Decision Trees (RF-PDT) is found to be the best performing algorithm with both classification and ranking metrics.

In the analysis [4], label correlations are used to categorize the multi label methods but still there are no specific benchmarks for exploiting label correlations especially with domains that have large output space.

Jian & Victor [5] used BR-KNN as their base classifier. They trained classifier using labeled set and randomly choose some images from unlabeled set as test data, performed labeling on them and then put those newly tested images into trained set. With this newly formed training set they re-trained the classifier. They iterated the process for 10 times and evaluated accuracy which was found more in all datasets but they did not compare their results with other state of art classifiers.

Liping & Michael [6] converted the image classification problem into the optimization problem with objective function using sparseness in label indicator so that relevant & irrelevant images with respect to labels can be classified. They used graph based method depending on random walk for the same where nodes represent each labeled and unlabeled image sample and the edges represent the relation between the two nodes. Initially they formed such a graph and then tried to convey label

information among the nodes. For multi label images they tried to constrain each image such that it must have a very small set of relevant labels associated.

Richard Cabral & Torre [7] tried to solve the multi-label classification problem for input set with some of the labels missing for some images. They used the property of image histogram as being equal to the sum of its object histograms including background and thus converted the multi-label classification problem into the rank minimization problem. They were also able to find the class histograms representations thus providing image

5. Approaches Which Measure Ambiguity in Nature

Miguel & Carlos [8] tried to tackle the problem of representing uncertainties in nature. They tried to help in selecting the most appropriate fuzzy membership function for scene images which could be useful for representing uncertainties in the nature. For each intensity level they assigned some k number of fuzzy sets in an image and then constructed a single interval type-2 fuzzy set (IT2FS). They calculated entropy of each such IT2FS. Minimum value of entropy is selected as a threshold value for better segmentation. Here one restriction is placed on the membership function is that, it has two local maxima, one for background and other for foreground (objects). They tested their method on every possible combination of fuzzy sets with three entropies and found the improved results with more no. of basic membership functions integrated to form interval type-2 fuzzy membership function.

There is only one approach by *Lim & Chan* [9], to the best of our knowledge, in which scene images are considered as containing non-mutually exclusive data. They have integrated fuzzy reasoning with qualitative reasoning and then mapped semantics of an image onto the output space of predefined classes. They ranked the classes according to their membership degree (confidence value) with respect to an image. They proposed Fuzzy Qualitative Rank Classifier (FQRC) and created a 4-tuple membership function to find dominant region, which solely belongs to a particular class and to find the overlapped region between two classes. For this they have used histogram representation of each attribute for each class.

Their results are found to be more towards reality with classification accuracy more than 70%. But the overall time required is more as compared to other traditional approaches as it needs to calculate membership value of each feature with respect to each and every class. Although it has measured the ambiguity in natural scene images very precisely, their algorithmic time complexity needs to be reworked on. This is because it is directly

proportional to the no. of classes and no. of features used. Also they have used leave one out strategy for categorizing the whole input dataset into training data and testing data which means that they have tested only a single part of whole dataset but still their time complexity is more. So if more and more data has to be classified it would require more time which would be hazardous to any real time applications. Fig. 1 shows the comparison of scene understanding approaches in the literature.

6. Methods for Feature Representation and Description

It is found that statistical methods like co-occurrence matrix has major difficulties involving high time complexity if an image is represented with high number of intensity levels. Model based methods like fractals [10] can efficiently describe roughness in natural scene images. But as natural surface is not deterministic but would always have some statistical variation, it makes the computation of fractal dimension much more difficult.

B.S. Manjunathi and W.Y. Ma [11] used Gabor wavelet filters for feature extraction, analysis and compared their results with other multi-resolution texture features for image retrieval problem.

It is found that the results with Gabor are more robust than the others mentioned earlier.

S. E. Grigorescu, N. Petkov [12] evaluated three texture feature extraction operators, mean, variance and standard deviation using number of filters. The filters used are derived from discrete transform, Gabor filters & Laws' masks.

They evaluated all local statistics in a square neighborhood of size 17×17 using non-linear point operations. All the similar textures were clustered in the same cluster. Pair-wise separability of cluster feature vectors is calculated using Mahalanobis distance. As it is evident that greater the separability, better is the classification possible.

It is found that all the sinusoidal transforms & Laws' [13] mask provide comparable results with misclassification probability of $10^{-7}\%$. Gabor filters results in misclassification probability of $10^{-12}\%$ which is much lesser than using previously stated filters. Again sinusoidal filters have only two possible orientations 0° and 90° , but Gabor filters have eight possible orientations. It seems that if textures are used for describing images, Gabor filter would be a better choice. But if implementation is concerned Laws' masks are better as its misclassification probability is negligible although more than that of Gabor filters.

Mihran Tuceryan & Jain in an article [14] gives detailed analysis of various texture based segmentation methods such as statistical methods, geometrical methods, signal

processing methods like spatial domain filtering, Fourier transform filters, Fourier domain filters, Gabor & wavelet filters etc. Spatial filtering methods such as LOG can work on many textures and can discriminate both natural & synthetic textures by controlling some parameters of estimation. Fourier domain also gives similar results. Gabor & wavelet models integrate frequency analysis into spatial domain thereby localizing the global frequency analysis. Also it is found that integrating a region based method with boundary based method obtain more robust and clean segmentation.

Yousef & Peter [15] tried to focus on the tasks such as

feature extraction and representations so that the application areas like object recognition, image classification and content based image retrieval (CBIR) systems would be benefited with fewer efforts. The aim was to bridge the gap between human perception of scene images and their computer vision counterparts. They integrated visual vocabularies (histograms) of all the classes instead of using traditional universal vocabularies like BOWs (bag of words). In doing so they mixed intensity based BOWs with image color information using key point density based weighting method. They used OSR (outdoor scene recognition) dataset and their classification accuracy was found to be 88.28%.

Fig.1 : Comparison of scene understanding approaches in the literature

6. Summary and Conclusion

Many of the approaches deal with scene understanding problem as multi label classification problem. Most of them solve the problem using various machine learning algorithms. But those approaches try to solve the problem based on the assumption that scene categories are mutually exclusive which is unnatural and hence illogical. In reality scene categories are overlapped with each other which depict ambiguity in the nature. Using Fuzzy membership functions the ambiguity can be measured more precisely. Approaches based on fuzzy

References

- [1] George J. Klir, Bo Yuan, *Fuzzy Sets And Fuzzy Logic: Theory and Applications*, (PHI Learning private limited).
- [2] Richard O. Duda, Peter Hart and David Stork, *Pattern Classification*, (Wiley India Pvt. Ltd.).
- [3] Gjorgji Madjarov, Dragi Kocev, *et. al.*, “An extensive experimental comparison of methods for multi-label learning”, Elsevier Ltd, 2012.
- [4] Min-Ling Zhang, Zhi-Hua Zhou, “A Review on Multi-Label Learning Algorithms”, IEEE transactions on knowledge and data engineering, vol. 26, no. 8, August 2014.
- [5] Jian Wu1, Victor S. Sheng2, *et. al.*, “Multi-label Active Learning for Image Classification”, ICIP 2014.
- [6] Liping Jing, Michael K. Ng, “Sparse Label-Indicator Optimization Methods for Image Classification”, IEEE transactions on image processing, vol. 23, no. 3, March 2014.
- [7] Ricardo Cabral, Fernando De la Torre, *et. al.*, “Matrix Completion for Weakly-Supervised Multi-Label Image Classification”, IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 1, January 2015.
- [8] Miguel Pagola, Carlos Lopez-Molina, *et. al.*, “Interval Type-2 Fuzzy Sets Constructed From Several Membership Functions: Application to the Fuzzy Thresholding Algorithm”, IEEE transactions on fuzzy systems, vol. 21, no. 2, April 2013.
- [9] Chern Hong Lim, Anhar Risnumawan, and Chee Seng Chan, “A Scene Image is Non mutually Exclusive-A Fuzzy Qualitative Scene Understanding”, IEEE transactions on fuzzy systems, vol. 22, no. 6, December 2014.
- [10] Alex P. Pentland: 'Fractal-Based Description Of Natural Scenes', IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. Pami-6, No. 6, November 1984.
- [11] B.S. Manjunathi, W. Y. Ma, “Texture Features for Browsing and Retrieval of Image Data”, IEEE transactions on pattern analysis and machine intelligence, vol. 18, no. 8, august 1996.
- [12] S. E. Grigorescu, N. Petkov, and P. Kruizinga, “A comparative study of filter based texture operators using Mahalanobis Distance”, IEEE, 2000.
- [13] Guillaume Lematre, Miroslav Rodojevi, “Texture segmentation: Co-occurrence matrix and Laws' texture masks methods”.
- [14] Mihran Tuceryam, Anil K. Jain, “Texture Analysis”, Research Gate Article September 2000.
- [15] Yousef Alqasrawi, Daniel Neagu and Peter I. Cowling, “Fusing integrated visual vocabularies-based bag of visual words and weighted colour moments on spatial pyramid layout for natural scene image classification”, Research gate, on signal image and video processing · July 2011.