SEMANTIC LABELLING OF IMAGES USING SUPPORT VECTOR MACHINES

Chandana Satya Prakash
University at Buffalo, State University of New York
csatyapr@buffalo.edu

Abstract

In this project, we perform semantic labeling of images from a benchmark dataset and evaluate the labeling procedure. There are eight classes representing the semantic labels to which super pixel must be assigned. These classes are: sky, tree, road, grass, water, building, mountains, and foreground objects. We perform 5-fold cross-validation with the dataset. We initially train our classifier using a set of training data which contains a set of images which have already been labelled and then we test our classifier on some test images which uses the trained classifier to label the test image.

1. Introduction

A scene can include various details about objects, regions, geometry, location, activities, and even nonvisual attributes and can be described in many ways. For example, a typical urban scene can be described by specifying the location of the foreground animal object and background grass, sky, and road regions. Alternatively, the image could be summarized as a street scene. We would like a computer to be able to detect about all these aspects of the scene and provide both coarse image-level tags and detailed pixel level annotations describing the semantics and geometry of the scene.

Object classification is assigning semantic labels to an object. It can be used as a building block for many other tasks such as localization, detection, and scene parsing. Current approaches typically adopt one of the two classification models: multiclass classification, which predicts one label out of a set of mutually exclusive labels or binary classifications, which make binary decisions for each label independently. The classification model used in carrying out our work is the multiclass classification model.

2. Related work

Gould, Fulton and Koller addressed the problem of decomposing a scene into geometrically and semantically coherent regions. They reasoned over both pixels and regions through a unified energy function. They also proposed an effective inference technique for optimizing this energy function and showed how it could be learned from data. We have done a major part of our work inspired by them.

P.Kontschieder, S. Rota Bul`o, H. Bischof and M.Pelillo_presented a simple and effective way to integrate ideas from structured learning into the popular random forest framework for the task of semantic image labelling. In particular, they incorporated the topology of the local label neighborhood in the training process and therefore intuitively learned valid labelling transitions among adjacent object categories. Also, they provided several experiments on the challenging CamVid and MSRCv2 databases and found superior results when compared to standard random forest or conditional random field (using pairwise potentials) classification results

D. Baby and S. Raju addressed the problem of task-specific image partitioning by supervised training and a region based image retrieval by creating a database of partitioned images and inputting a query image. The correlation clustering model aimed to merge super pixels into regions of homogeneity with respect to the solution of any particular image labeling problem. The used LP relaxation to approximately solve the correlation clustering over a super pixel graph where a rich pair wise feature vector was defined based on several visual cues. They used S-SVM for supervised training of parameters in correlation clustering, and the cutting plane algorithm with LP-relaxed inference was applied to solve the optimization problem of S-SVM.

D. Depalov and B. Gandhi presented a new approach for semantic classification that utilizes perceptual models for image segmentation and classification. The main

innovations of the proposed approach are the use of an algorithm that produces perceptually uniform segments and the selection of perceptually-motivated region-wide color and texture features. The features of these regions are used as medium level descriptors and are the key to bridging the gap between low-level image primitives and high-level image semantics. Their results indicate that the proposed approach offers significant performance improvements over the existing literature.

X. Yu and Hong Liu put forward a method to classify image regions based on their semantics. It can reduce the gap between human's perception and description of image content. Because the pre-defined semantic class hierarchy reflects in the semantics by human's subject, so it is flexible and intuitive query by novice. The use of the binary SVM classifiers that classify image regions using different features at different levels in the hierarchy were the main reasons behind the high classification precision that we achieved in our experiments. Currently, we are looking adding more feature extraction methods to get high precision and put more classifiers to include more classes into the system.

3. Approach

3.1 Super pixels

Our model decomposes a scene into and semantically consistent regions using a classifier over pixels and regions. In the work carried out, we have labelled the images using a fixed number of labels. One approach that could be used to assign these labels to an image is pixel wise but there are some disadvantages of doing so such as Classifying large regions rather than individual pixels we can compute more robust features and reduce inference complexity. Also, the use of individual pixels makes it difficult to utilize more global cues, including both robust statistics about the appearance of larger regions, which can help average out the random variations of individual pixels, and relationships between regions, which are hard to "transmit" by using local interactions at the pixel level. Instead of performing pixel wise labelling we have adopted a more improved approach which is super pixels based labelling. Super pixels correspond to small, nearly-uniform regions in the image they are constructed in advance using simple procedures. Super pixel preserve boundary. Also, they can significantly reduce computational cost and allow feature extraction to be conducted from a larger homogeneous region.

3.2. Features

Features can also be used for segment classification.

One of the features we have used is color and we have used two different color spaces namely RGB color space and HSV color space. Color is not a strong clue for image classes, and few object categories are associable to a single color. On the other side, color can provide strong information on object boundaries: a sharp color change between two parts gives a strong clue that those patches can belong to different objects. Additionally, up to a certain extent, objects instances tend to be color homogeneous. The color composition feature exploits the fact that the HSV cannot simultaneously perceive a large number of colors. In addition, it accounts for the varying image characteristics and the adaptive nature of the HSV. The RGB color space is not as informative as the HSV color space.

Another feature used is the texture component. In texture we use LMFilters to give us useful information and we take the mean and max of these Reponses returned. The segmentation algorithm combines the color composition and texture features to obtain segments of uniform texture. We use texture to determine the major structural composition of the image and combines it with color, first to estimate the major segments, and then to obtain accurate And precise localization of the border between regions.

We also use the position descriptor to label our images. Normalized y coordinate is used to detect elements such as sky which normally tend to be at a higher height than road. The cues given by this position descriptor are quite helpful while labelling images.

3.3 Support Vector Machine

In our study, the label of an object is not known initially. Only the label of a set of objects is known, which is called the training data. We map an object to its label according to the information learned from the training data.

The way of constructing a hyper plane to get binary classifiers done that can separate members of one class from others, but most real data hardly separate because the hyper plane that can successfully separate the members of the two classes in most case does not exist. One measure to solve this problem is to map the data into a higher dimensional space, where the members of the two classes can separate by a hyper plane. However, the traditional classifier is not good at in high dimensional vector. It is extremely expensive in terms of memory and time.

The way of constructing a hyper plane to get binary classifiers done that can separate members of one class from others, but most real data hardly separate because the hyper plane that can successfully separate the members of the two classes in most case does not exist. One measure to solve this problem is to map the data into a higher

dimensional space, where the members of the two classes can separate by a hyper plane. However, the traditional classifier is not good at in high dimensional vector. It is

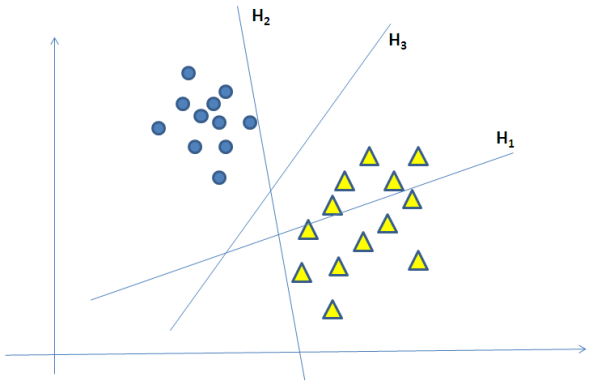

Figure 1. Separation hyper planes. H1 does not separate the two classes; H2 separates but with a very tinny margin between the classes.

extremely expensive in terms of memory and time. Support Vector Machines can solve this problem. SVM avoid over fitting the data by choosing a hyper plane from the many that can separate the data. That maximizes the minimum distance from the hyper plane to the closest training point. Such a hyper plane call the maximum margin hyper plane. Another advantage of the SVM is the compact representation of the decision boundary, so the number of support vectors is small as compared to the number of points in the training set.

Learning the semantics for each class through using SVM based on different features of training sample regions of each class. After training the SVM, binary classifiers that can classify image regions based on their semantics create. Then, we use these classifiers to classify our database of image regions leading to the classification to determine the class of an input image region and this image region map into its class in the semantic class hierarchy.

3. Experimental analysis

Below are the results of the conducted work. We first performed the 5-fold cross validation for 10000 super pixels. We can see that we have obtained an average accuracy of 13. In some folds class 3 seems to have the most accuracy and in some class 6 seems to have the most accuracy.

We have performed these experiments on MATLAB 2013a using libsvm. One of the main reasons for a low accuracy is the use of the above mentioned version of MATLAB. Using MATLAB 2014b which uses fitcsvm we can obtain an accuracy of above 40.

Also, tried implementing SIFT features but it did not show any improvement in the accuracy.

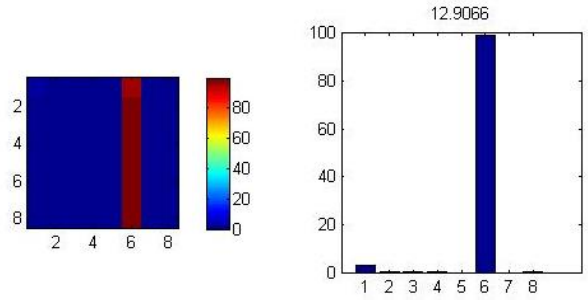

Fig 3. Fold 1 of the 5-Fold cross validation for 10000 super pixels

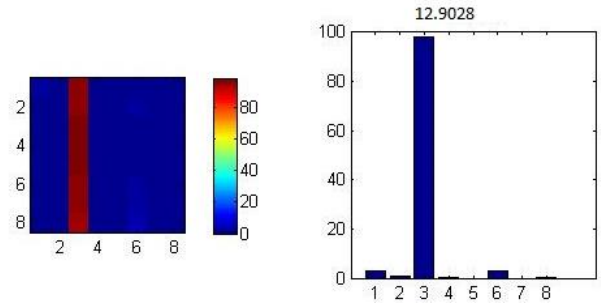

Fig 4. Fold 2 of the 5-Fold cross validation for 10000 super pixels

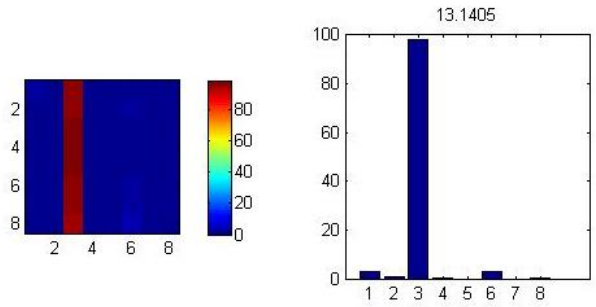

Fig 5. Fold 3 of the 5-Fold cross validation for 10000 super pixels

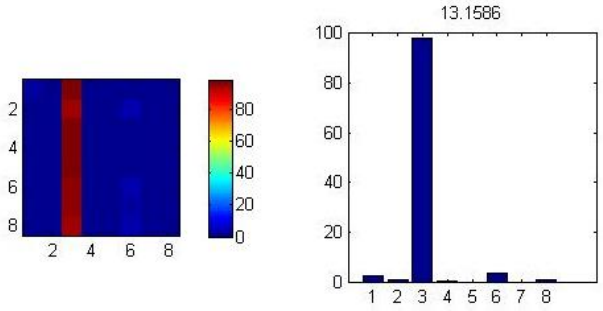

Fig 6. Fold 4 of the 5-Fold cross validation for 10000 super pixels

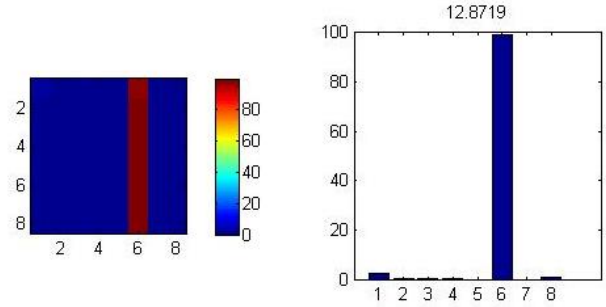

Fig 7. Fold 5 of the 5-Fold cross validation for 10000 super pixels

Below are the results by varying the number of super pixels.

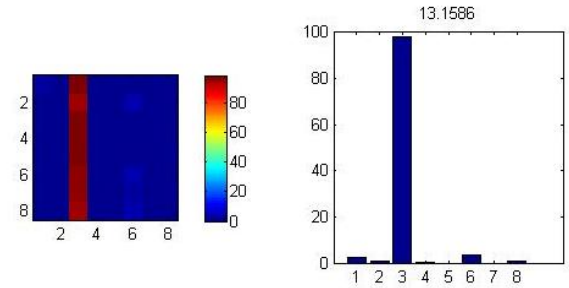

Fig 8. Accuracy for 1000 super pixels

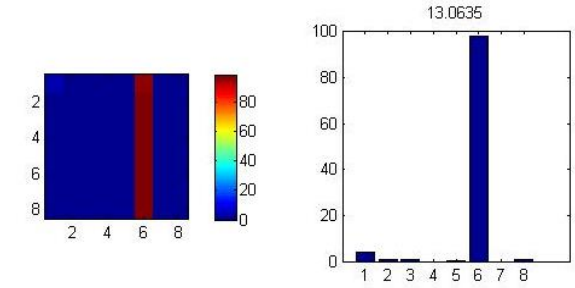

Fig 9. Accuracy for 10000 super pixels

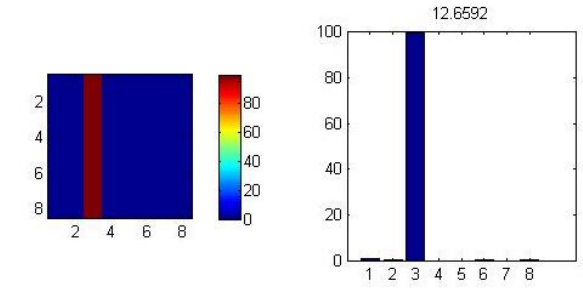

Fig 10. Accuracy for 25000 super pixels

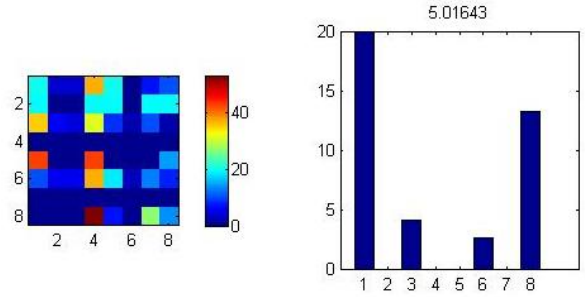

Fig 11Accuracy for 10000 super pixels for linear.

4. Conclusion

In this paper, we have performed 5-fold cross validation on training set and then tested it on a set of images. We have also tried to increase the accuracy by using SIFT features and also compared it against linear and varied number of super pixels.

We can increase the accuracy by using many other tactics such as using LAB color space, hog and SIFT features.

References

[1] D. Comaniciu, P.Meer, and S.Member. Mean shift: A robust approach toward feature space analysis. PAMI, 2002.

[2] A. Criminisi. Microsoft research cambridge object recognition image databasehttp://research.microsoft.com/vision/ cambridge/recognition, 2004.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

N.E.Ayat, M.Cheiet, C.Y.Suen, Automatic model selection For the optimization of SVM kernels-pattern recognition 1733-1745 (2005)

[4] Zaher Al Aghbari, Region-based semantic image classification –International Journal of Image and Graphics
Vol.6. No.3 (3006) 357-375

[5] A. Criminisi. Microsoft research cambridge object recognition image database. http://research.microsoft.com/vision/cambridge/recognition, 2004.
[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, 2007.
[8] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. IJCV, 2008.
[9] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2004.
[10] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. IJCV, 2007.
[11] D. Hoiem, A. A. Efros, and M. Hebert. Closing the loop on scene interpretation. CVPR, 2008.
[12] D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert. Recovering occlusion boundaries. ICCV, 2007.
[13] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In CVPR, 08.
[14] B. C. Russell, A. A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In CVPR, 06.
[15] B. C. Russell, A. B. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. IJCV, 2008.
[16] A. Saxena, M. Sun, and A. Y