Parsing Natural Scenes with Recursive Neural Networks

Shubham Gupta, Vedant Mishra
Advisor: Dr. Amitabha Mukerjee
CS365: Artificial Intelligence
II Semester, Year 2012-13

1 Introduction

Object detection and classification in an image has been one of the most fascinating applicative problems in AI. Many methods have been proposed in this respect. We will be using RNN (recursive neural networks) so as to identify different classes in an image (for e.g. sky, building, water etc.). Humans are highly efficient in classifying different objects in an image. Although much of the work has been done in this field but human like efficiency and accuracy is yet not achieved.

2 Motivation

Natural scene images not only consists of certain distinct units but also has some sort of relationship with its neighboring units. There is "nested hierarchical structuring in scene images that capture both part-of and proximity relationships". For example, bicycles are mostly on street regions or buildings or tree would be on side of the street. Parsing the scenes by understanding this hierarchical structure not only allows to classify the scene into different units but also helps to understand the way these interact to form a whole scene.

The recursive neural network approach used in [1] is general in nature and is not limited to scene classification. It can be extended to discover the recursive structures of other form of inputs like natural language sentences

3 Related Work

Main work on which our project would be based is "**Parsing Natural Scenes and Natural Language with Recursive Neural Networks"** by Richard Socher, Andrew Y. Ng, Christopher D. Manning, Cliff Chiung-Yu Lin of Stanford University.

Work done by Gould, S., Fulton, R., and Koller, D. in "**Decomposing a Scene into Geometric and Semantically Consistent Regions** "also uses merging operations for scene classification. Other work in the recent years on scene classification include "Li, L-J., Socher, R., and Fei-Fei, L. **Towards total scene understanding: classification, annotation and segmentation in an automatic framework**".

Lee, H., Grosse, R., Ranganath, R., and Ng, A. in the paper "**Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations**" used deep learning for the classification of
Images with more realistic sizes.

4 Methodology

Our main work is to implement the **natural scenes image parsing** using the source code available at www.socher.org on a dataset of images of IIT Kanpur. On the basis of the results obtained, we would modify the code so as to improve the performance of the algorithm.

ALGORITHM ([1])

The outline of the algorithm is on the similar lines as described in [1].
Images are divided into small segments with features (like color, texture etc.) associated with them. These features are mapped into a semantic space using a NN. With these representations and adjacency matrix as input, RNN computes:
(i) A score that supports neighboring segments being merged to form larger segments.
(ii) Modified semantic space and adjacency matrix for the merged region.
(iii) Label of each node (defines the object categories such as building or street based on the training results).
The model is trained, using Max margin estimators and greedy approach, so that the score is high when neighboring regions have the same class label. After regions with the same object label are merged, neighboring objects are merged to form the whole image. These merging decisions can be visualized as a tree structure in which each node has its label and score, and higher nodes represent larger elements of the image. The tree with highest cumulative score is taken as the representation of the input image.

5 POSSIBLE EXTENSIONS

a.) No. of classes in which a scene image is parsed could be increased. Right now the code parses into 8 classes namely grass, building, water, road, mountain, tree, sky and all other objects.
b.) Also the code can be modified to incorporate local context of the neighborhood pixels for improved performance.

6 SOURCE CODE AND DATASET

Source Code for RNN Model: "www.socher.org"

Dataset: We would be creating our own dataset for testing purposes.
But for initial training and understanding purposes the dataset used is:
http://www.socher.org/index.php/Main/ParsingNaturalScenesAndNaturalLanguageWithRecursiveNeuralNetworks

7 REFERENCES

[1] **Main Paper**: socher-linCC-NgA-11_parsing-natural-scenes-w-RNNs
http://nlp.stanford.edu/pubs/SocherLinNgManning_ICML2011.pdf
[2] Video related to Parsing Images: http://techtalks.tv/talks/54422/
[3] Decomposition of scene into geometric regions and semantically consistent regions:
http://robotics.stanford.edu/~koller/Papers/Gould+al:ICCV09.pdf