Parsing Natural Scene Images Using Recursive Neural Networks

CS365:Artificial Intelligence

Instructor: Dr. Amitabha Mukerjee

Students: Shubham Gupta
Vedant Mishra

Summary

- Motivation
- Work done Before
- Overview of the process
- Image Segmentation and Feature Extraction
- Algorithm For Image Parsing
- Results
- References

Motivation

- This approach is general in nature and not limited to scene images only.
- It not only classifies the scene into discrete units but also helps in understanding the way these interact to form a whole scene.
- It has outperformed the state-of-the-art methods for image classification and segmentation on the Stanford dataset.

Work Done Before

- 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

- "Decomposing a Scene into Geometric and Semantically Consistent Regions " by Gould, S., Fulton, R., and Koller, D. also uses merging operations for scene classification.

- Hinton, G. E. and Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science*,313, 2006 used Deep learning to find lower dimensional representations for fixed size input images.

Overview Of the Process

- We would be testing the algorithm on some images of nearby locality.

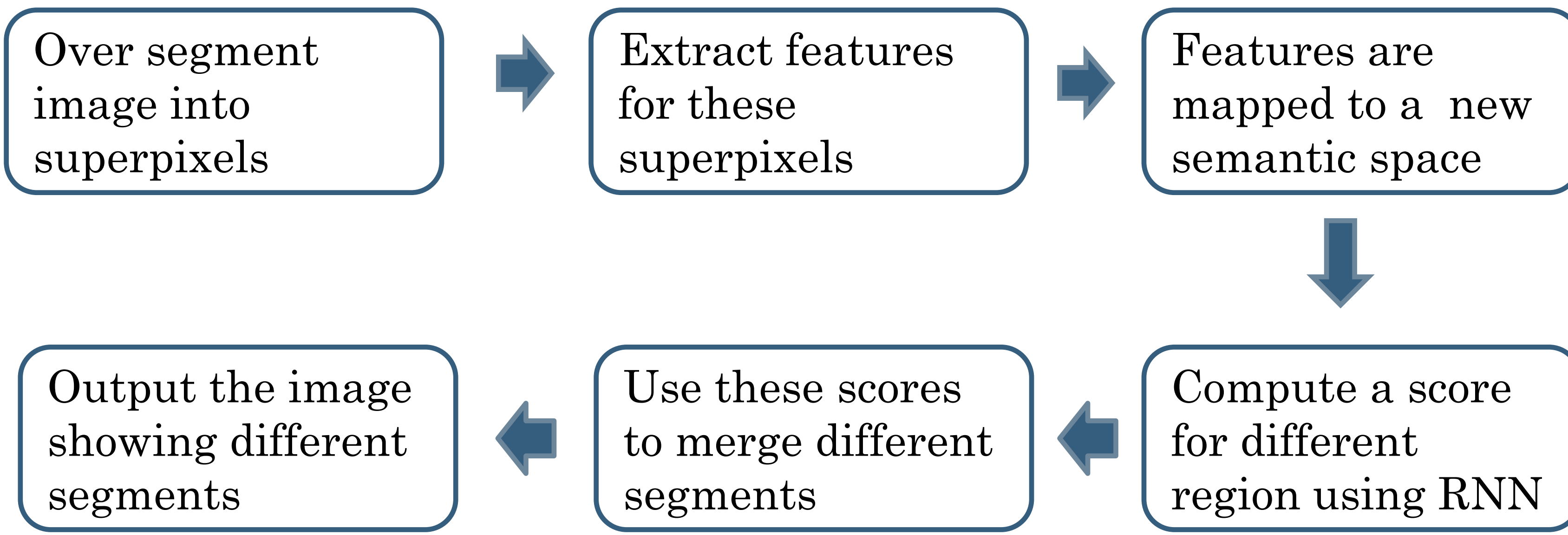


Image Segmentation and Feature Extraction

1. Over segment the image using Mean Shift Algorithm (using [4])

- 2. Extract Features for each segment (Using [3])

a) Color Histogram

b) Shape Features

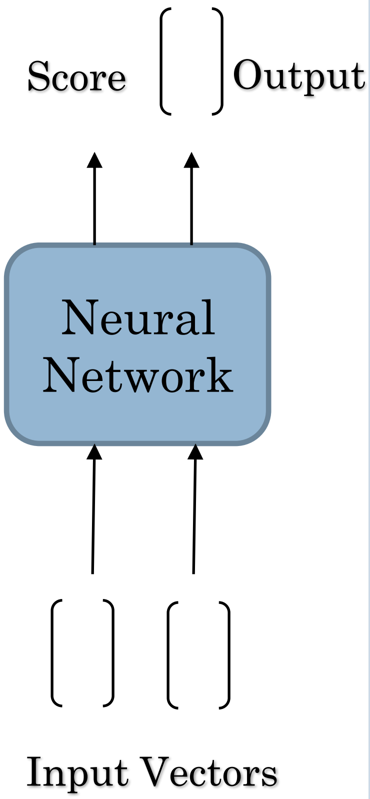
c) Area

d) Textures

Algorithm for Parsing Images

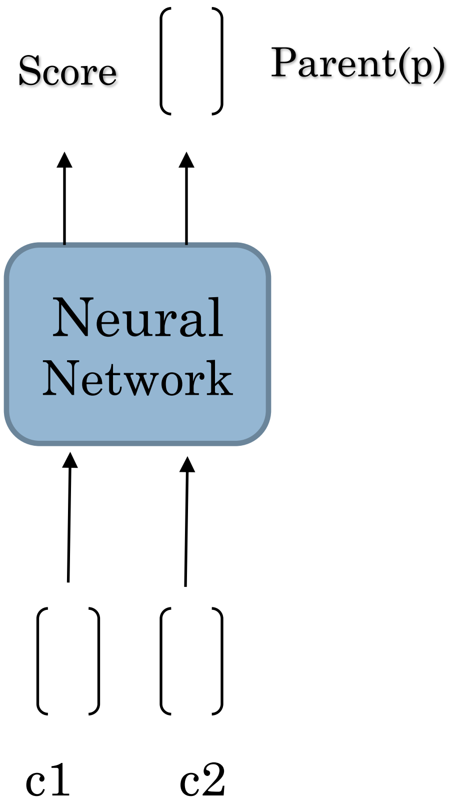
- Input
  - Set of vector which represent image segment
  - Adjacency matrix A(i,j) which is 1 if segment i is neighbor of j.

- Output
  1. The semantic representation if the 2 vectors are merged.
  2. Score of how plausible the new node is.
  3. Correct Parse Tree



Recursive Neural Network is used for Parsing.

Recursive Neural Network



1. $p = \text{sigmoid}\left(W \begin{bmatrix} c1 \\ c2 \end{bmatrix} + b \right)$, Using[1]

where sigmoid outputs between 0 and 1 and c1 and c2 are input vectors.

$\text{Score} = (w^T)_{score}\, p$ Using [1]

where w^T is a parameter vector

Parsing an Image

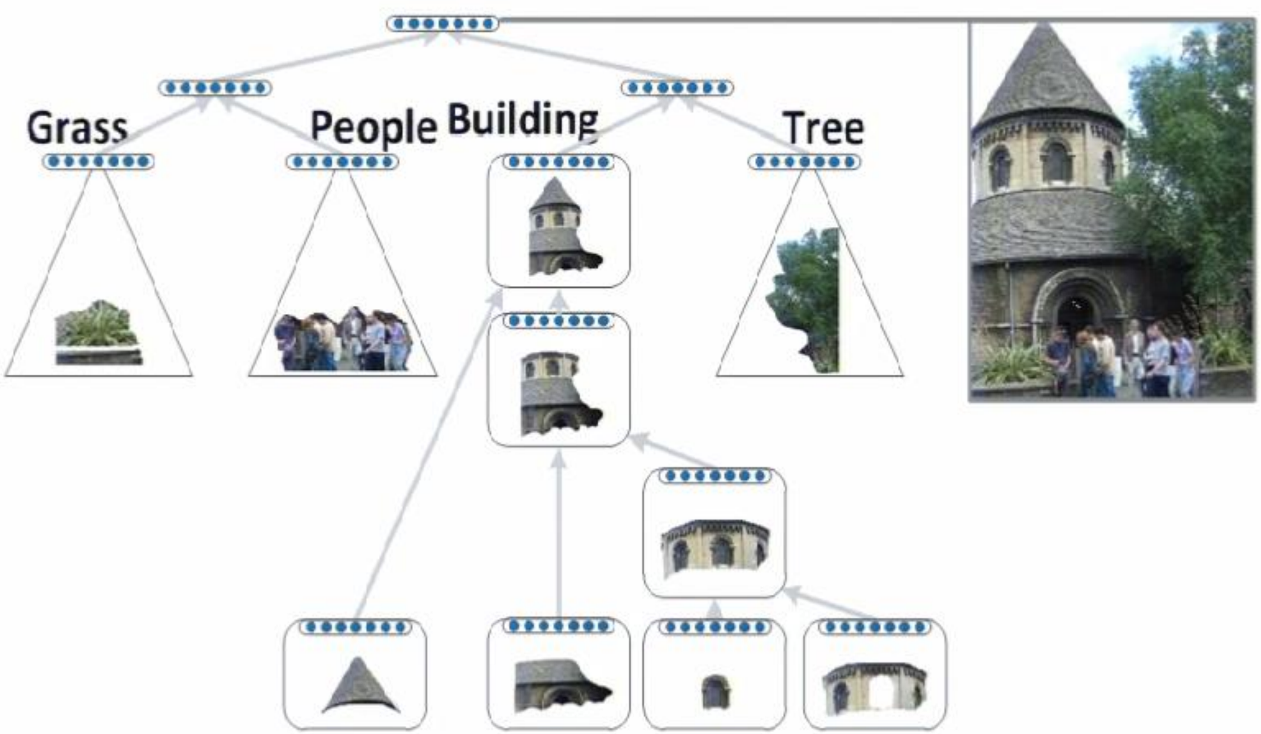


Image from www.socher.org

RESULTS

279	279	279	279	279	9	9	9	9	7	7	7	7	7	7	7	7		
279	279	279	279	279	9	9	9	9	7	7	7	7	7	7	7	7		
279	279	279	279	9	9	9	9	9	7	7	7	7	7	7	7	7	7	
279	279	279	9	9	9	9	9	9	7	7	7	7	7	7	7	7	7	7
279	279	279	9	9	9	9	9	9	7	7	7	7	7	7	7	7	7	5
279	279	279	9	9	9	9	9	9	7	7	7	7	7	5	5	5	5	5
279	279	279	9	9	9	9	9	9	9	5	5	5	5	5	5	5	5	5
278	278	278	9	9	9	9	9	9	9	5	5	5	5	5	5	5	5	5
278	278	278	9	9	9	9	9	9	9	1	5	5	5	5	5	5	5	5
278	278	278	278	9	9	9	9	9	9	1	1	5	5	5	5	5	5	
278	278	278	278	9	9	9	9	9	10	10	1	1	5	5	5	5		
278	278	278	278	278	9	10	10	10	10	10	1	1	1	5				
278	278	278	278	278	10	10	10	10	10	10	1	1	1					
278	278	278	278	276	276	10	10	10	10	10	1	1	1					
278	278	278	276	276	276	276	10	10	10	10	1	1						
278	276	276	276	276	276	276	10	10	10	10	1	1						
276	276	276	276	276	276	276	276	10	10	10	1	1						
276	276	276	276	276	276	276	276	276	10	10	1							
281	276	276	276	276	276	276	276	276	10	10	10							
281	276	276	276	276	276	276	276	276	276	10	1							
281	276	276	276	276	276	276	276	276	276	271								

Test Image
(Faculty Building)

Image after
Segmentation

Matrix denoting different
segments of the image

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

[4] Comaniciu, D. and Meer, P. Mean shift: a robust approach toward feature space analysis. *IEEE PAMI*, 24(5):603–619, May 2002.

Dataset and source code:

[1]nlp.stanford.edu/~socherr/cppFeatures.tar.bz2

[2]http://www.socher.org/index.php/Main/ParsingNaturalScenesAndNaturalLanguageWithRecursiveNeuralNetworks

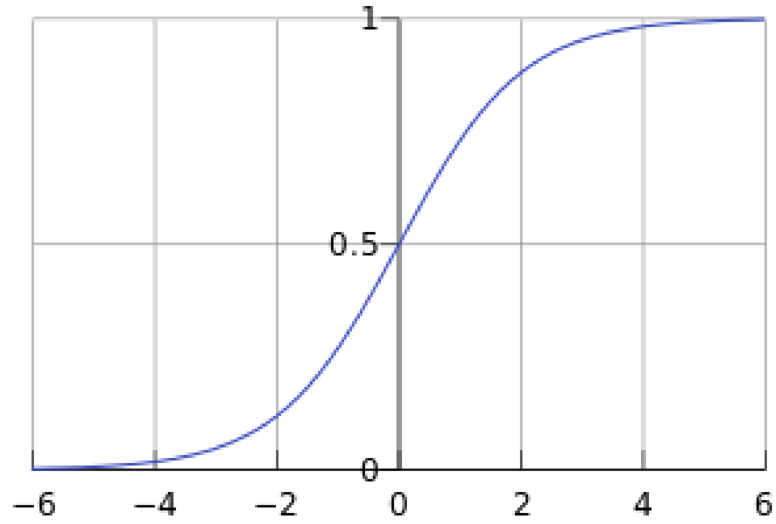
THANK YOU!!

Questions??

Sigmoid Function

- Map elements to 0 and 1.

- $S(t) = \dfrac{1}{1 + e^{-t}}$



Max Margin Framework

- This framework is for training the data.
- The score of the tree is computed by sum of the parsing decision scores at each node.
- $s(x_i , y_j)$ is the score for the correct parse y_i corresponding to the input image $x_{i.}$
- Max margin framework is defined as :

$$J = \sum s(x_i , y_j) - \max (s(x_{i\,,} y) + \Delta(y , y_i))$$

- The loss function Δ penalizes all incorrect decision.

LABELLING IN RECURSIVE NEURAL NETWORK

- We can use softmax function for calculating labels in neural network.
  - $Label_p = softmax\,(W^{label}\,p\,)$ ref [1]

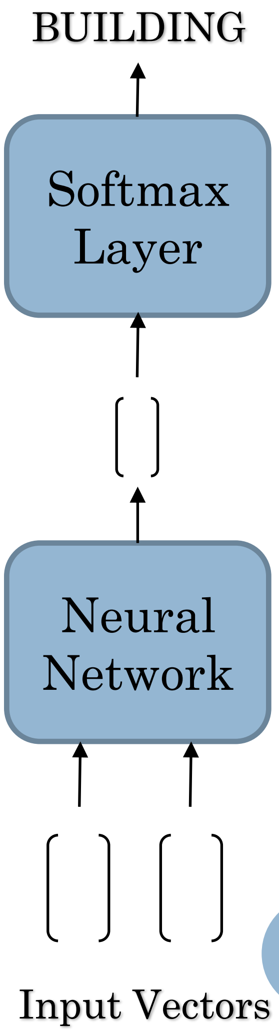
$$p_i = \frac{\exp(q_i)}{\sum_{j=1}^{n} \exp(q_j)}$$

- Different label for different segments according to p . For eg

  red - Building

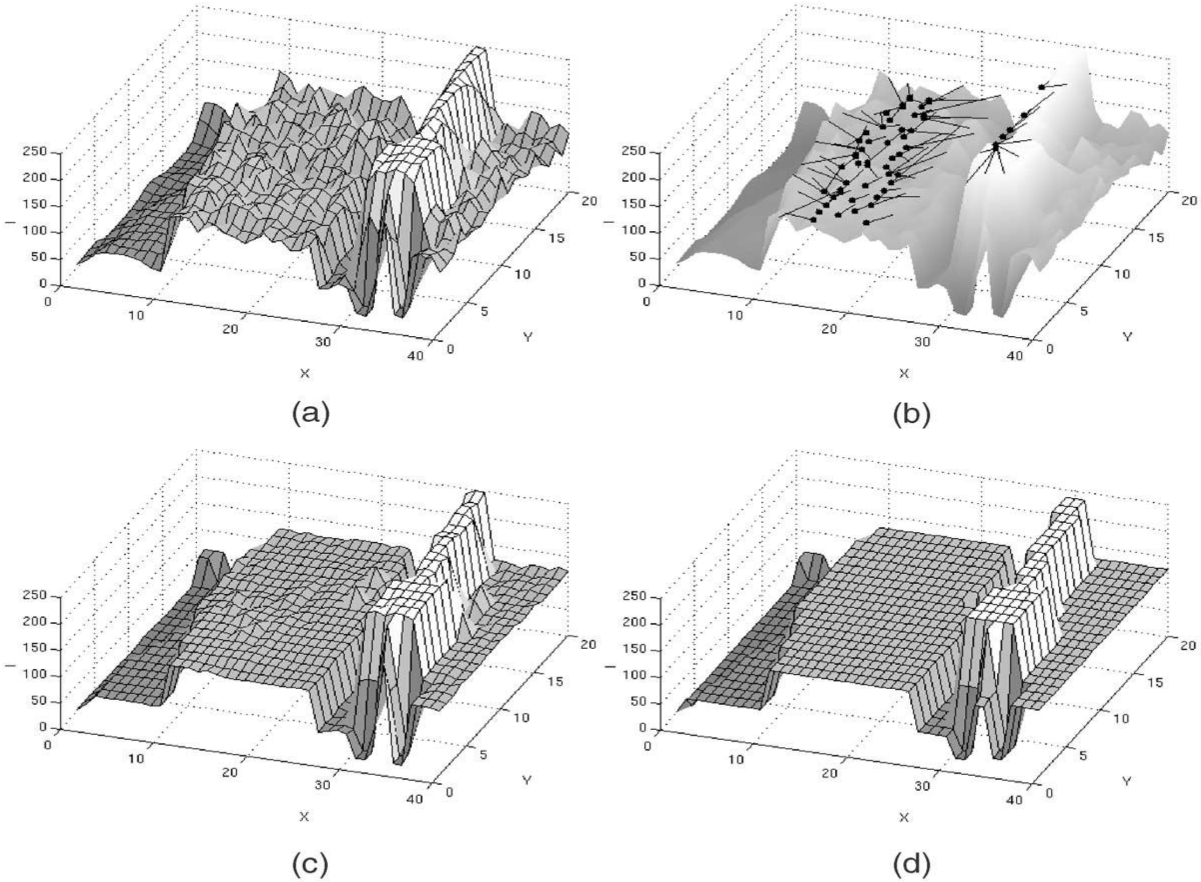
  blue - Sky

  green - Grass



OVERSEGMENTATION

- Kernel is a function of feature vector.



Mapping in the semantic space

- $a_j = f(W^{sem} F_i + b^{sem})$

  where a_i is the map in semantic space, F_i is the feature vector and W is the parameter matrix