Parsing Natural Scene Images Using Recursive Neural Networks

Shubham Gupta 10699   Vedant Mishra 10792
Advisor: Dr. Amitabha Mukerjee
CS365:Artifiicial Intelligence

Motivation

- Active research field
- Outperformed state-of-the-art methods
- Discovers recursive structure of natural scene images.
- Method not limited to images only

Related Work

- Parsing Natural Scenes and Natural Language with RNN
  - Richard Socher, Andrew Y. Ng, Christopher D. Manning, Cliff Chiung-Yu Lin[2011]

- Decomposing a Scene into Geometric and Semantically Consistent Regions
  - Gould, S., Fulton, R., and Koller, D. [2009]

- Reducing the dimensionality of data with neural networks
  - Hinton, G. E. and Salakhutdinov, R. R.[2006]
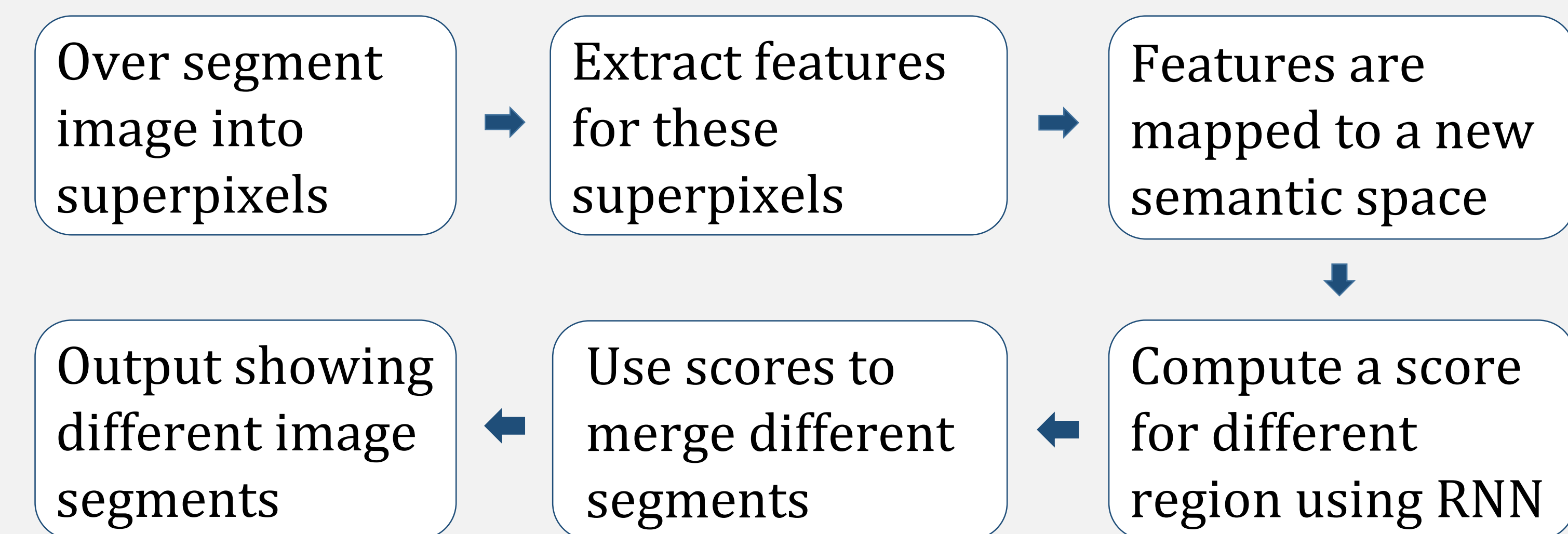
Our Approach

- Overview

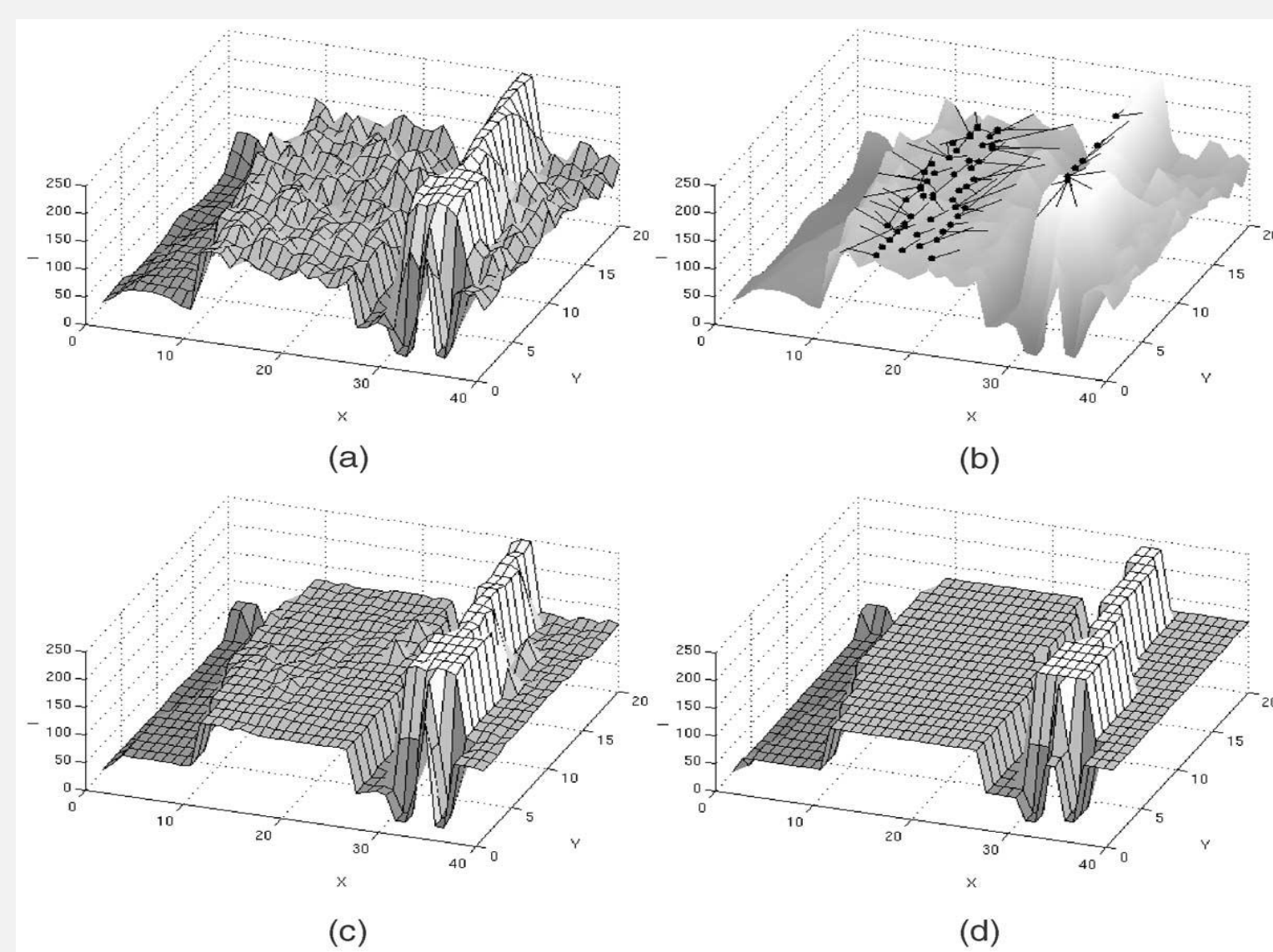| Over segment image into superpixels | → | Extract features for these superpixels | → | Features are mapped to a new semantic space |
| Output showing different image segments | ← | Use scores to merge different segments | ← | Compute a score for different region using RNN |

- Image Segmentation
  - Using Mean Shift Algorithm

Image from Source[3]

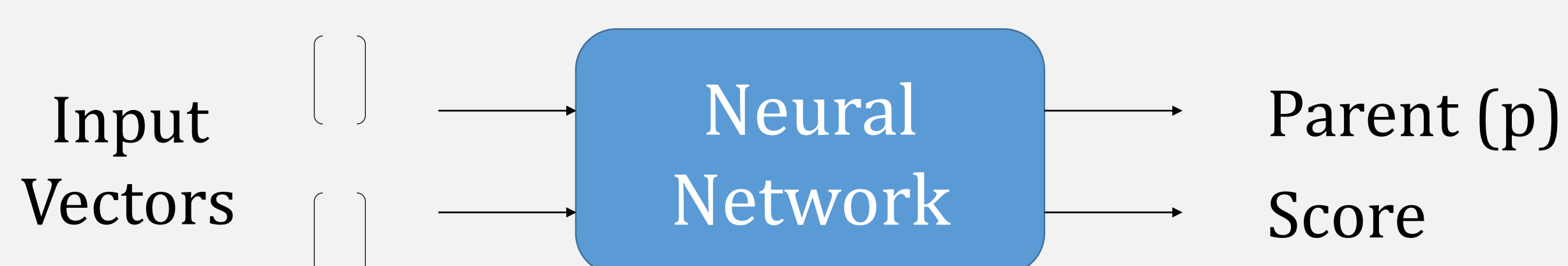- Extract Features for each segment   "From [3]"
  - Color Histogram
  - Shape Features
  - Area
  - Texton map

(a) Input image   (b) Texton map   (c) Feature pair = (r,t)   (d) Superimposed rectangles
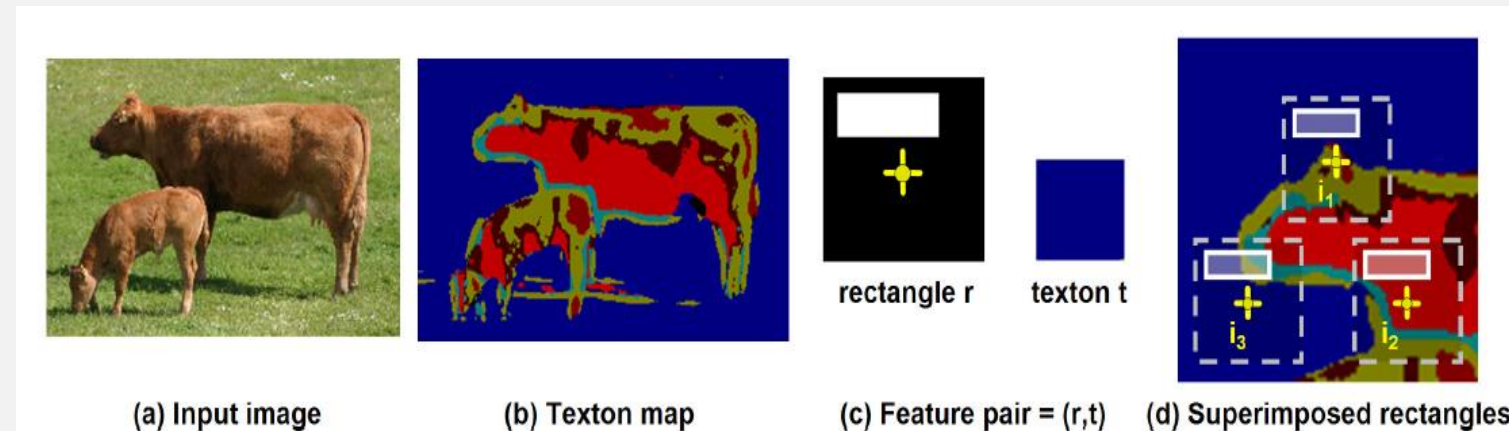
- Algorithm for Parsing Images

Input Vectors → Neural Network → Parent (p) / Score

Recursive Neural Network

- Architecture

$$p = sigmoid(W \begin{bmatrix} c1 \\ c2 \end{bmatrix} + b)$$ "From [1]"

here p is the parent vector and c1 and c2 are the input vectors

$$Score = (W^t)_{score}\ p$$ "From [1]"

where w^t is the parameter matrix
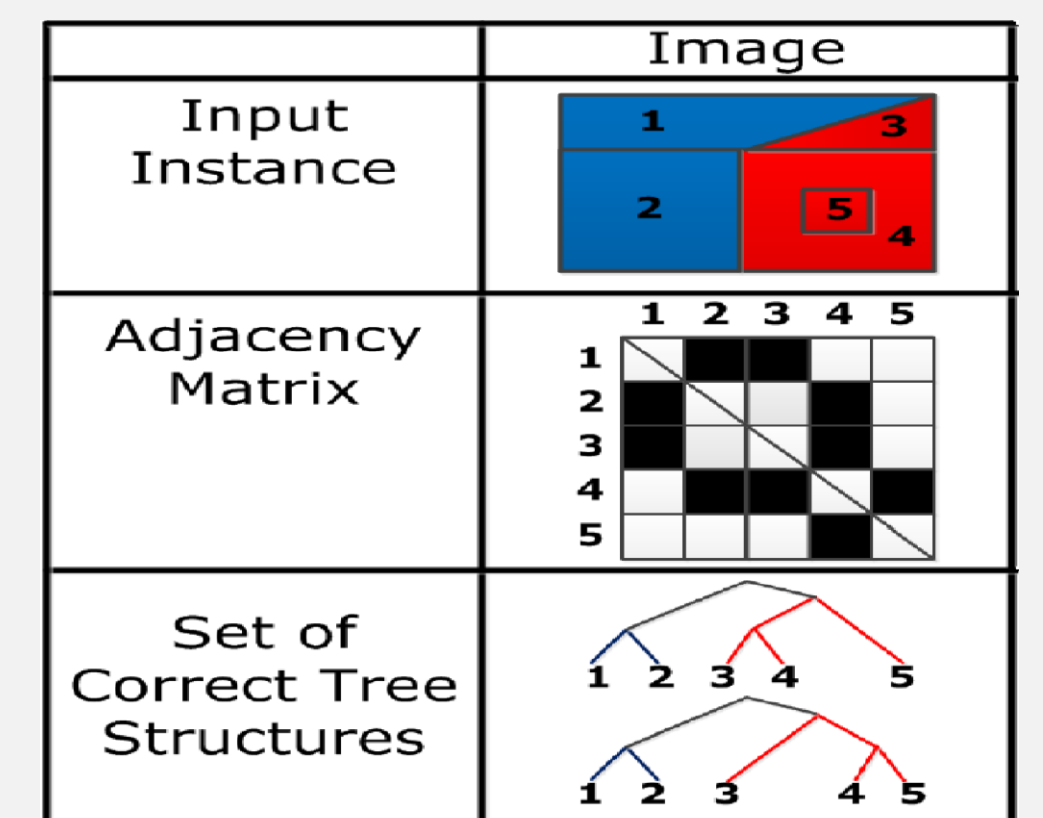
- Training

- Max Margin Principle

$$J = \sum s(x_i , y_j) - max (s(x_i , y) + \Delta(y , y_i)$$ "From [1]"

here $s(x_i , y_j)$ is the score for the correct parse y_i and Δ is the loss function

- Back propagation through Structure

Results

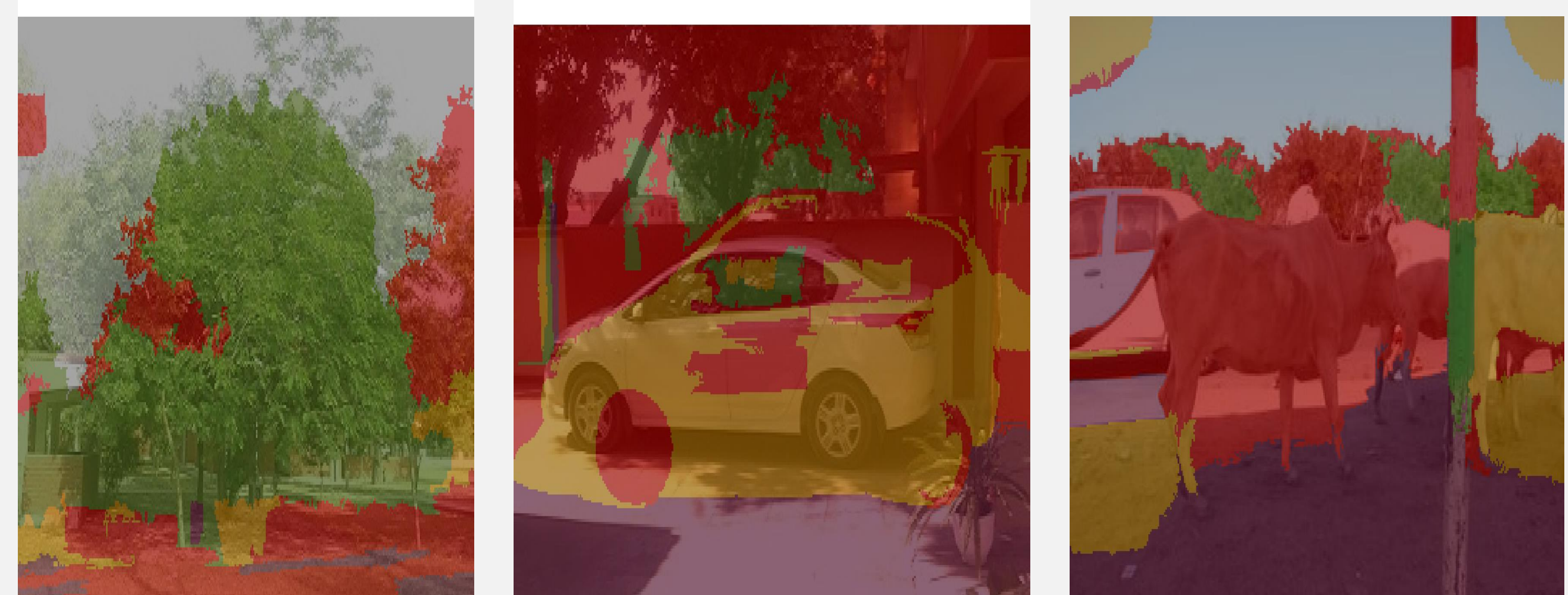- Oversegmentation

Test Image (Faculty Building)   Segmented Image

- Parsed images

Accuracy: i.70% ii.45%   iii.50%

sky tree road grass water bldg

References

[1]. Main Paper : socher-linCC-NgA-11_parsing natural scenes with RNNs
          http://nlp.stanford.edu/pubs/SocherLinNgManning_ICML2011.pdf
[2] Decomposition of scene into geometric regions and semantically
     consistent regions
[3] 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/ParsingNaturalScenesAndNatural
LanguageWithRecursiveNeuralNetworks