Relationship Aware Context Adaptive Feature Selection Framework for Image Parsing

Basim Azam
PhD Candidate

This paper was presented at IEEE International Joint Conference on Neural Networks (IJCNN 2021)

Outline

Introduction

Proposed Approach

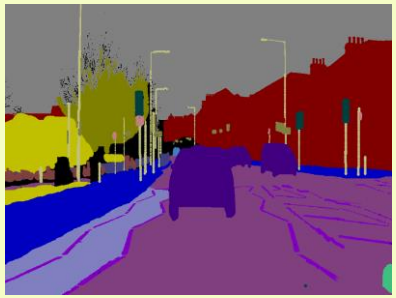
Experimental Evaluation

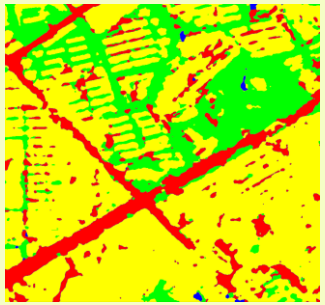
Conclusions

Introduction

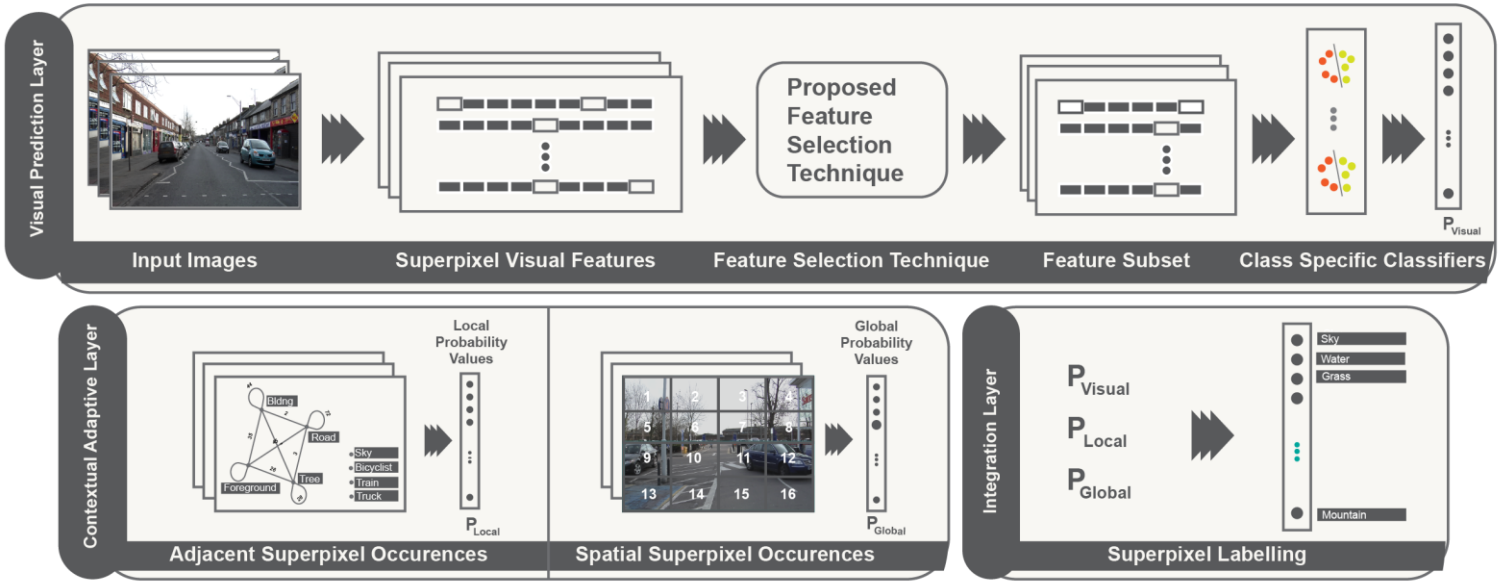
Image parsing refers to segmentation of an image into regions with object category labels such as tree, building, car and road.

- Image parsing has a variety of applications and is being fundamentally applied in
  - Autonomous Driving
  - Medical Image Segmentation
  - Hyperspectral Image Analysis
  - Remote Sensing Applications



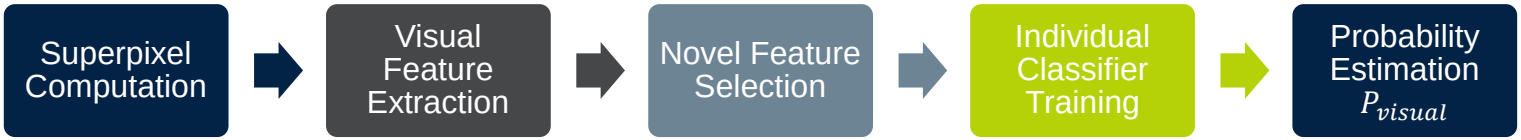
Proposed Approach

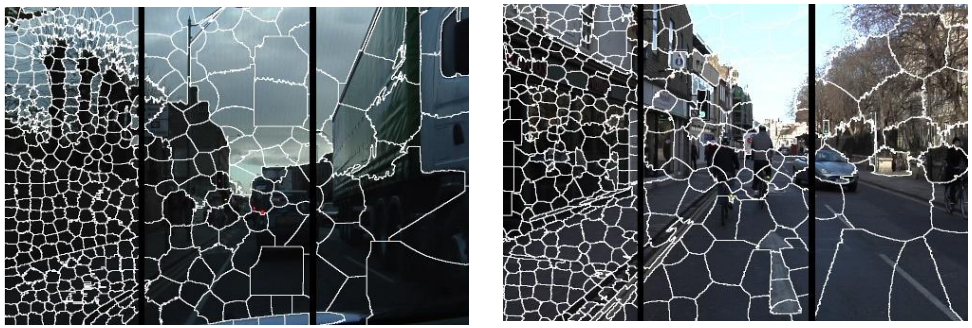


Proposed Image Parsing Framework. **Visual Prediction Layer, Contextual Adaptive Layer** and **Integration Layer.**

The architectures contemplates the contextual properties of the object categories and offers enhanced performances.

Visual Prediction (VP) Layer



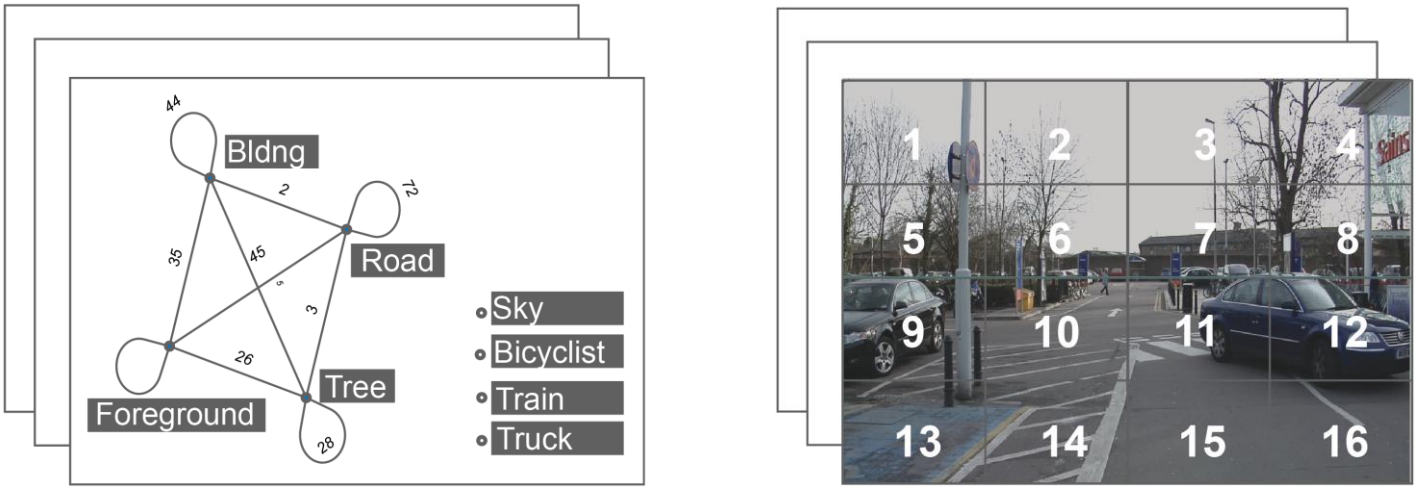


Images segmented using SLIC Superpixel algorithm of size 1024, 512 and 64 pixels.

The feature selection technique , making use of mutual information (MI) and fisher score (FI), is proposed to rank the features in descending order.

- MI determines the amount by which information provided by feature decreases
- FI facilitates to verify the separability of an individual feature from other features.

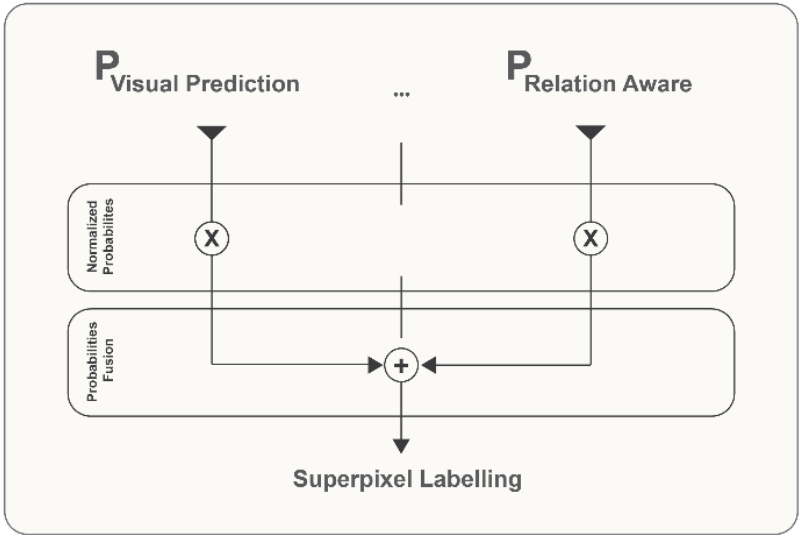
Context Adaptive (CA) Layer



(a) The adjacent superpixel vote computations and (b) global computation of votes for object occurrences.

The term contextual adaptation represents the information of the objects occurring within an image in varying ranges.

Integration Layer

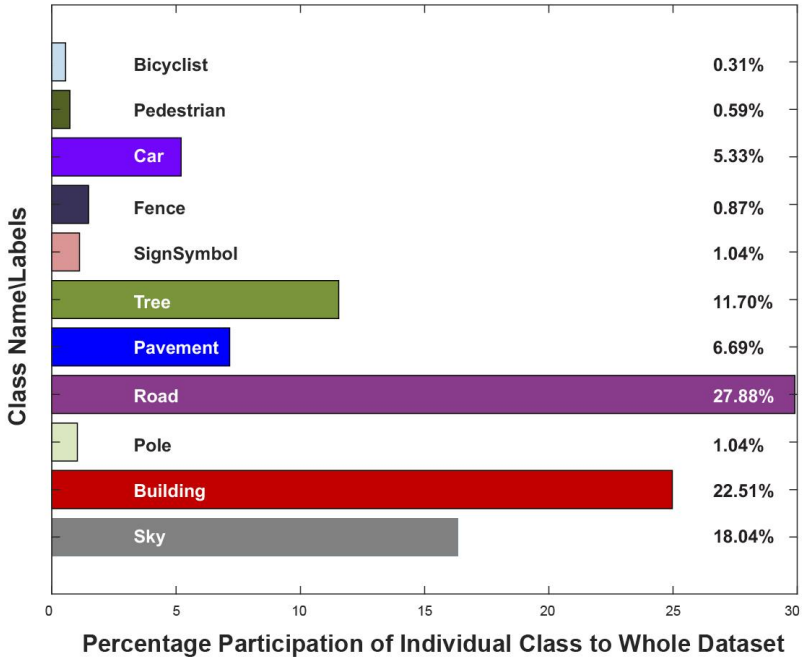


The weights of integration layer are computed using probability values estimated in VP layer and CA layer.

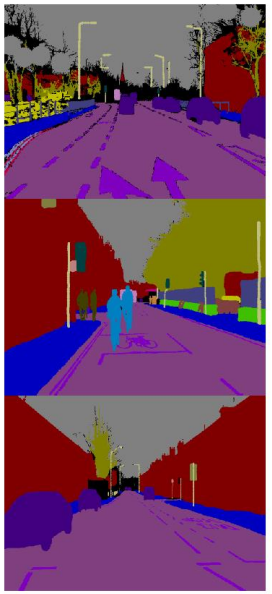
The Multilayer-Perceptron (MLP) is applied to learn integration weights.

Dataset

CamVid is a challenging benchmark, having crowded scenarios , varying illumination conditions and multiple types of roads.



Images



Labels

Ablation Study

Table 1. Layer-wise Performance Evaluation in terms of Accuracy (%) for 512 superpixels

	VP Layer	CA Layer	Proposed Approach
512 Superpixels	68.25	87.61	89.79
Sky	95.79	92.66	94.52
Building	67.16	95.15	93.44
Pole	0.08	0	0
Road	95.52	98.22	96.29
Pavement	22.75	70.59	83.69
Tree	39.43	83.59	86.42
SignSymbol	0.06	0	31.85
Fence	26.76	61.43	71.47
Car	30.72	77.07	86.73
Pedestrian	11.83	0.02	26.88
Bicyclist	33.21	28.1	62.16

Table 2. Layer-wise Performance Evaluation in terms of Accuracy (%) for 256 superpixels

	VP Layer	CA Layer	Proposed Approach
256 Superpixels	66.57	81	85.63
Sky	96.6	88.62	94.45
Building	62.15	92.35	89.93
Pole	0.05	0	0
Road	97.05	98.22	95.69
Pavement	0	38.03	68.3
Tree	46.56	76.13	80.64
SignSymbol	0.04	0	11.73
Fence	25.69	32.92	57.43
Car	27.26	61.14	79.04
Pedestrian	27.24	0	10.9
Bicyclist	0.09	0.02	45.06

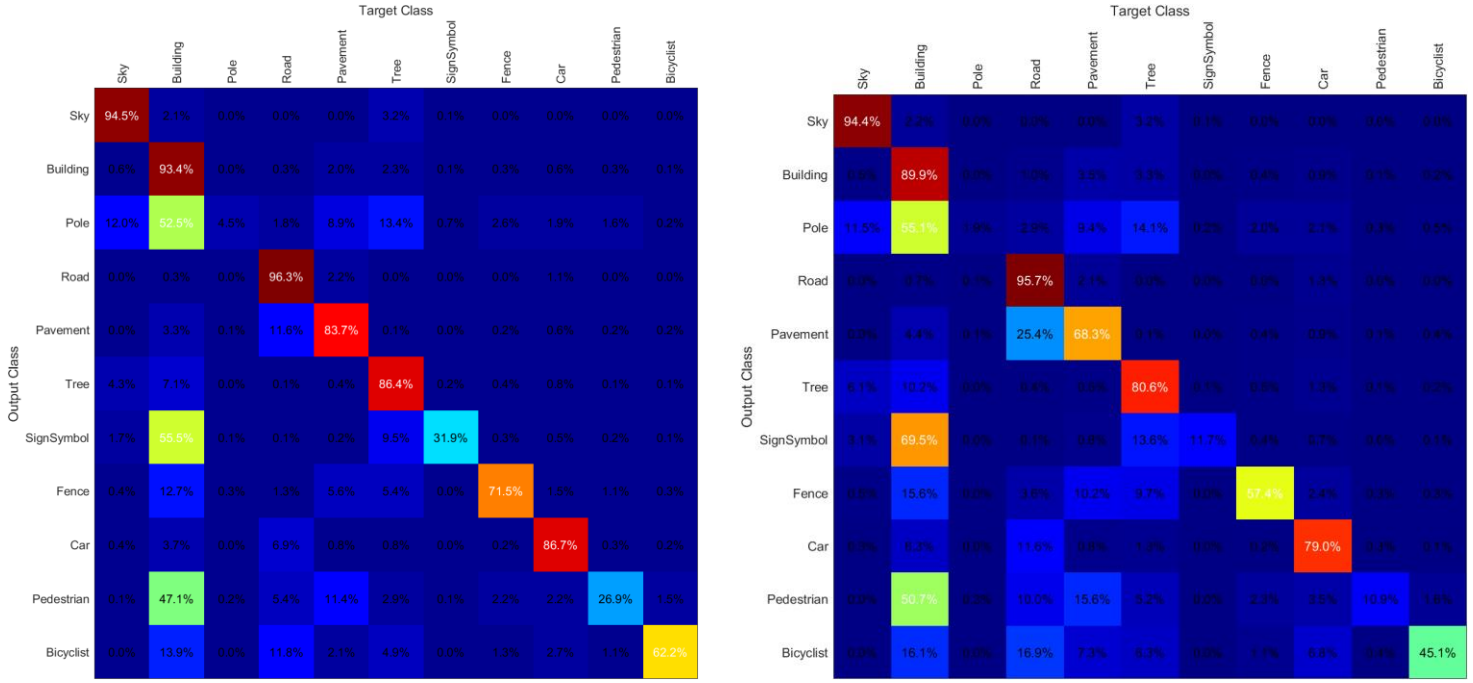
The architecture achieves better and higher scores by integrating both the VP layer and CA layer. Also , higher number of superpixels increase the overall accuracy score.

Comparative Analysis

Table 3. Quantitative Performance Comparisons with Previous Approaches on the CAMVID BG Dataset in terms of Accuracy (%), Proposed feature selection base architecture achieves competitive average and Global accuracy, It also performs Extremely well for various challenging classes (Building, Car, Tree, Fence, Bicyclist)

	Building	Tree	Sky	Car	Sign Symbol	Road	Pedestrian	Fence	Pole	Pavement	Bicyclist	Global	Average
Proposed Approach (ADB-MLP)	**93.4**	**86.42**	**94.5**	**86.7**	**31.9**	**96.3**	**26.8**	**71.47**	**4.5**	**68.3**	**62.16**	**89.8**	**67.1**
CNN-CRF [1]	84.3	65.3	95.6	74.6	0.4	93.5	25.6	32.3	13.8	85	54.3	72.9	56.8
SfM + Appearance [2]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	69.1	53.0
Super Parsing [3]	87.0	67.1	96.9	62.7	30.1	95.9	14.7	17.9	1.7	70	19.4	83.3	51.2
Local Label Descriptors [4]	80.7	61.5	88.8	16.4	n/a	98.0	1.09	0.05	4.13	12.4	0.07	73.6	36.3
Boosting + Detector + CRF [5]	81.5	76.6	96.2	78.7	40.2	93.9	43.0	47.6	14.3	81.5	33.9	83.8	59.2
FCN+Comb [6]	79.7	77.2	85.7	86.1	45.3	94.9	45.9	69.0	25.2	86.2	57.9	88.7	63.8
ReSeg [7]	86.8	84.7	93.0	87.3	48.6	98.0	63.3	20.9	35.6	87.3	43.5	88.7	68.1

Confusion Matrices



Confusion matrices over the CamVid test set using Adaboost as class specific classifier and MLP in Integration Layer. Left: 512 Superpixels, Right:256 Superpixels

The misclassification can also be detected as the superpixels of building are confused with pole, sign symbol and pedestrian

Conclusion

A novel image parsing framework with feature selection and context adaption is presented.

Supremacy

- Context adaptive layer produces improvements
- The feature selection technique improves the accuracy and reduces the computational complexity
- Significant improvements to the global accuracy on CamVid Dataset
- The contextual adaptation produces learns category specific features.

Future Works

- Evaluation on more dataset.
- Investigations to improve the by optimizing the learning parameters and fusion of layers

Thank You

Questions?

b.azam@cqu.edu.au

RHD Symposium 2021, School of Engineering Technology
Day 2 – Tuesday 7th December
Session 5 - 1:30 PM AEST