SALPROP: SALIENT OBJECT PROPOSALS VIA AGGREGATED EDGE CUES

Prerana Mukherjee*[1], Brejesh Lall[1], Sarvaswa Tandon[2]

*e-mail: eez138300@ee.iitd.ac.in

[1]Department of Electrical Engineering, Indian Institute of Technology, Delhi, India; [2]National Institute of Technology Goa, India

at IEEE International Conference on Image Processing (ICIP 2017), Beijing China

PROBLEM STATEMENT

GOAL: We propose a novel object proposal generation scheme by formulating a graph-based salient edge classification framework that utilizes the edge context.

Key Features:

- **Fewer number of bounding boxes for good coverage of the prominent objects contained in the image.**
- **Maintain order of saliency in the object proposals.**
- **Tight bound on the objects.**

INTRODUCTION

- **Object localization with high degree of precision is a challenging task.**
- **It is usually solved by**
  - **Using feature statistics**
  - **Generic object region proposals**
  - **Deep learning**
  - ***Exploit Edges***
- **Edges capture most of the shape information thus preserving important structural properties contained in the image.**

PROPOSED APPROACH

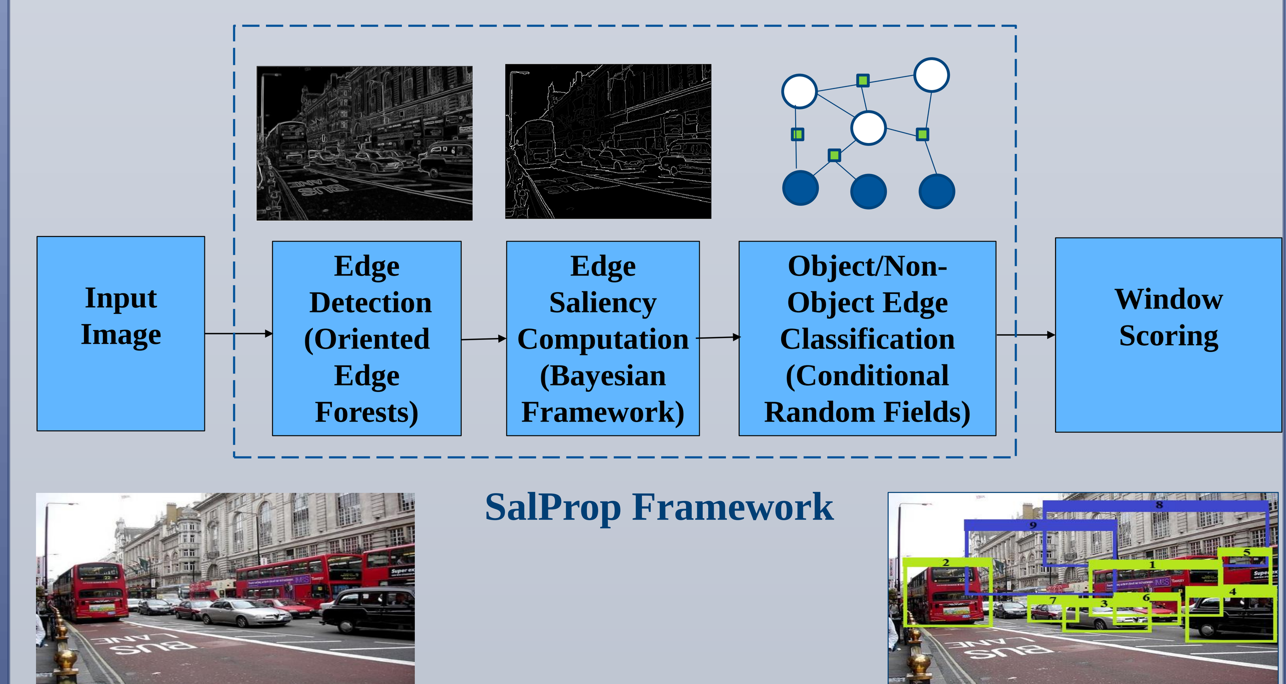

Fig. 1. The SalProp Framework. Given any RGB image, we generate proposals ranked in the order of saliency. Green boxes contain the most salient objects having higher rank and blue boxes contain less salient objects and are ranked lower in the proposal set. The number assigned to each box indicates its saliency ranking in the proposal pool.

EXPERIMENTAL RESULTS

Table 1. Comparison of top 1000 proposals with state-of-the-art techniques on AUC% (higher the better), number of proposals (N) at 75% recall (lower the better) and recall% (higher the better). '-' indicates that the particular recall rate is not reached.

Method	IoU=0.5			IoU=0.6			IoU=0.7			Time(in s)
	AUC	N@75%	Recall	AUC	N@75%	Recall	AUC	N@75%	Recall	
EdgeBoxes70	65.82	86	93.45	60.52	141	90.73	53.03	294	84.15	0.25
Perceptual Edge	1.8	-	10.4	0.08	-	4.7	0.02	-	1.2	7.2
MCG	71	37	94.6	62.8	95	90.2	62.5	366	83	34
Objectness	62	145	89	52	504	78	30	-	41	3
Rahtu	57	278	84	50	551	79	43.5	-	73.5	3
Randomized Prim's	59.3	129	89	50	315	83	40.7	1000	75	1
Rantalankila	25.14	511	86.38	21.63	718	79.77	17.76	-	70.75	10
Selective Search	62.3	105	93	54	207	88	45.3	544	80	10
Rigor	40.39	-	67.43	32.05	-	54.5	23.44	-	40.73	6.84
GOP	47.8	155	93	41	272	87	33.4	705	76	0.9
SalProp	67.5	74	91	58.1	244	84	44	-	71.3	7

- **SalProp is the best technique at lower number of proposals achieving over 25% and 19% recall with only 1 window at IoU=0.5 and 0.6 respectively.**
- **At IoU=0:7, SalProp outperforms Rahtu [3] by 3.46%, Selective Search [5] by 5.16%, Objectness [2] by 7.32%, Randomized Prim's [4] by 8.71%, GOP [1] by 22.36%, Rigor [8] by 23.46%, Rantalankila [10] by 30.05% and Perceptual Edge [6] by 30.35% at top-10 proposals.**
- **Outperforms objectness [2] by 2%, 6% and 30% at IoU thresholds 0.5, 0.6 and 0.7 respectively.**
- **Comparable performance to EdgeBoxes [9] while having a computational speedup of 5x over MCG [7].**

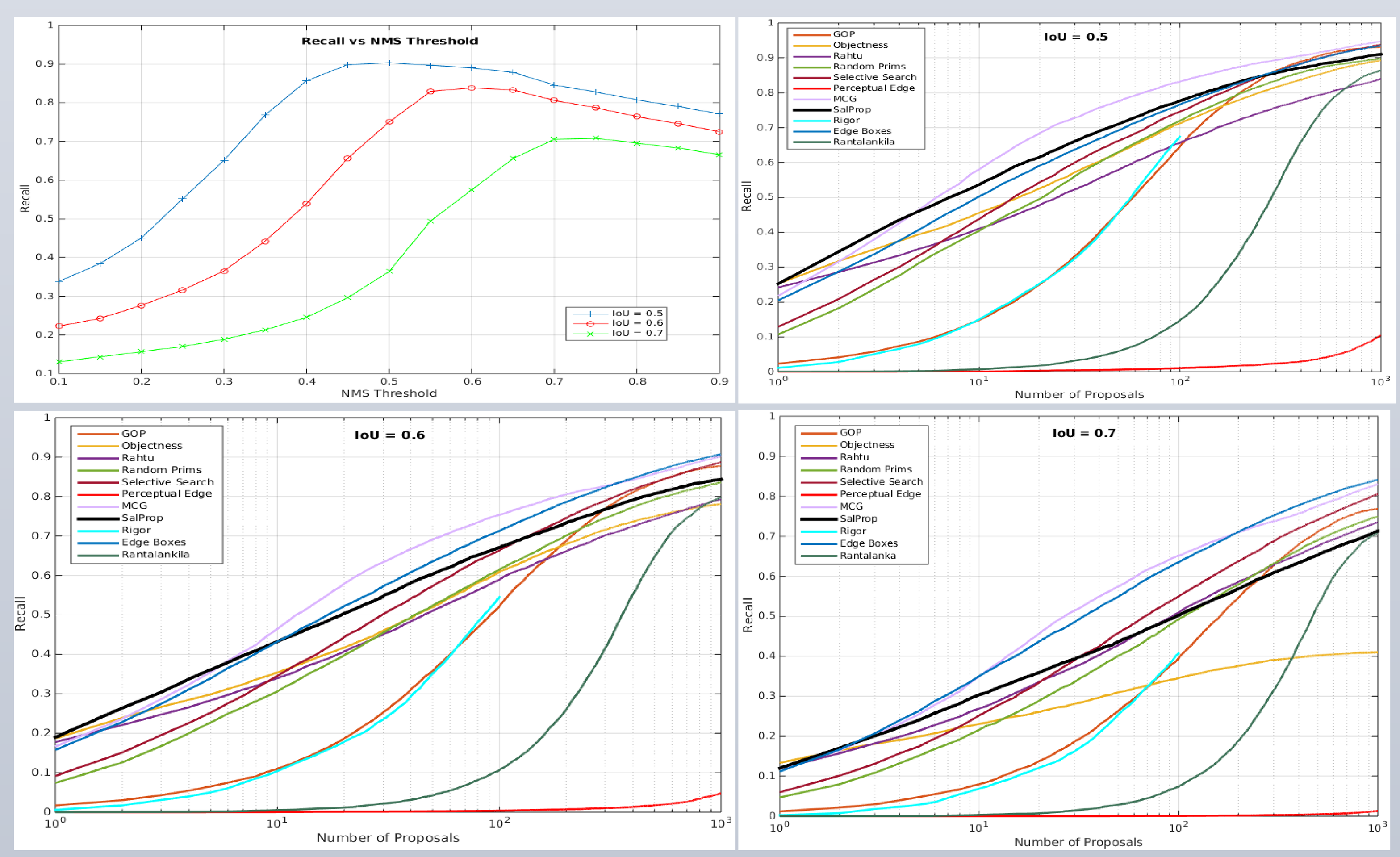

Fig. 3. (a) NMS cut-off threshold for highest recall value at varying IoU on validation set images. (b)-(d) The detection rate vs. the number of bounding box proposals for varying IoU = 0.5, 0.6 and 0.7 on test set images.

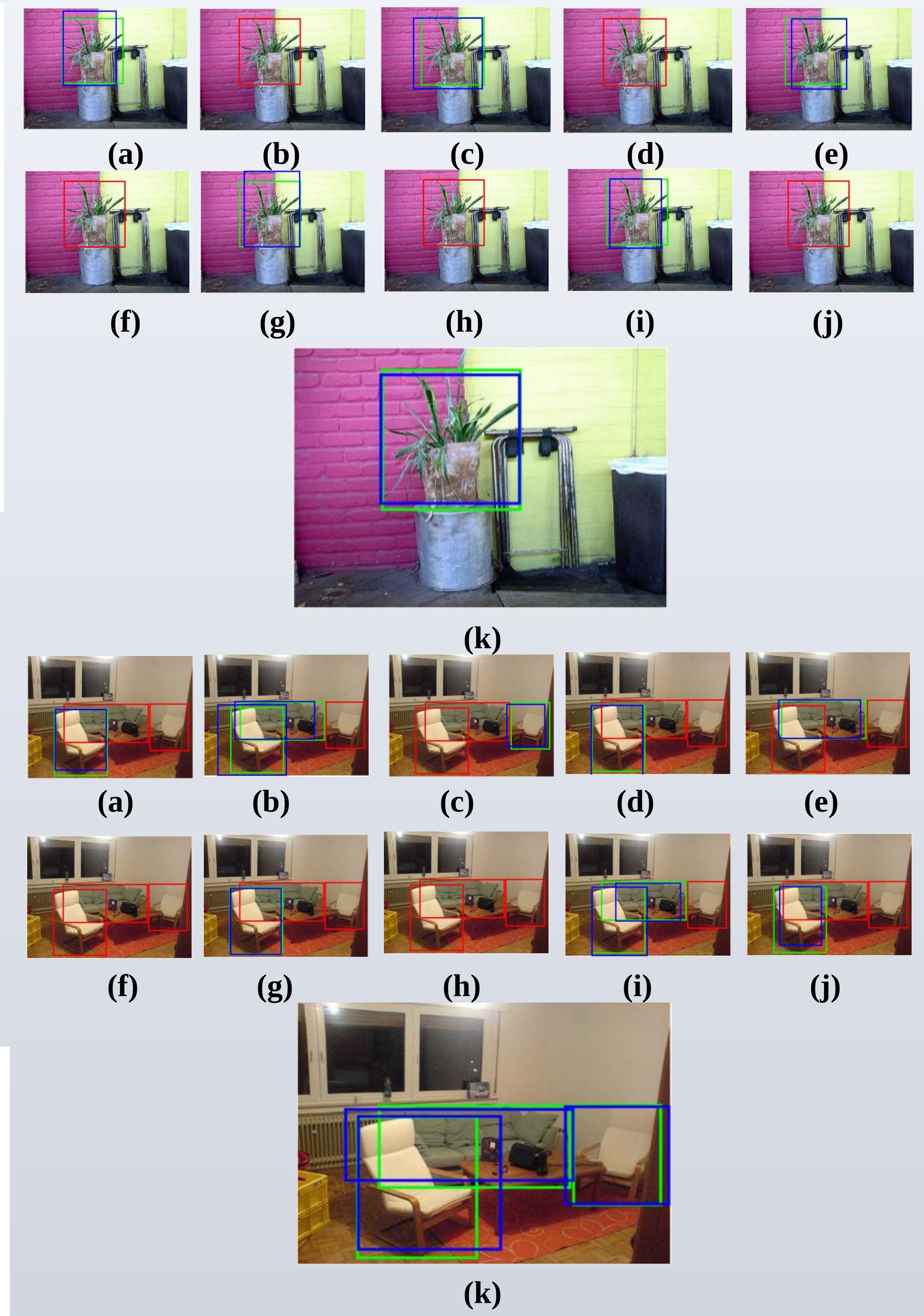

Fig. 4. Qualitative examples of our object proposals with other state-of-art techniques. (a) Geodesic Object Proposals [1] (b) Objectness [2] (c) Rahtu [3] (d) Randomized Prim [4] (e) Selective Search [5] (f) Perceptual Edge [6] (g) Multiscale Combinatorial Grouping [7] (h) Rigor [8] (i) EdgeBoxes70 [9] (j) Rantalankila [10] (k) SalProp. Blue bounding boxes are the closest produced object proposals to each ground truth bounding box shown in green. Missed objects are shown with bounding boxes indicated in red meaning that the object was not found. IoU threshold=0.7 was used to determine correctness for all examples.

REFERENCES

[1] Geodesic object proposals, ECCV'14.
[2] Measuring the objectness of image windows, TPAMI'12.
[3] Learning a category independent object detection cascade, ICCV'11.
[4] Prime object proposals with randomized prim's algorithm, ICCV'13.
[5] Selective search for object recognition, IJCV'13.
[6] Making better use of edges via perceptual grouping, CVPR'15.
[7] Multiscale combinatorial grouping for image segmentation and object proposal generation, TPAMI'17.
[8] RIGOR: Reusing inference in graph cuts for generating object regions, CVPR'14.
[9] Edge boxes: Locating object proposals from edges, ECCV'14.
[10] Generating object segmentation proposals using global and local search, CVPR'14.

BAYESIAN FRAMEWORK

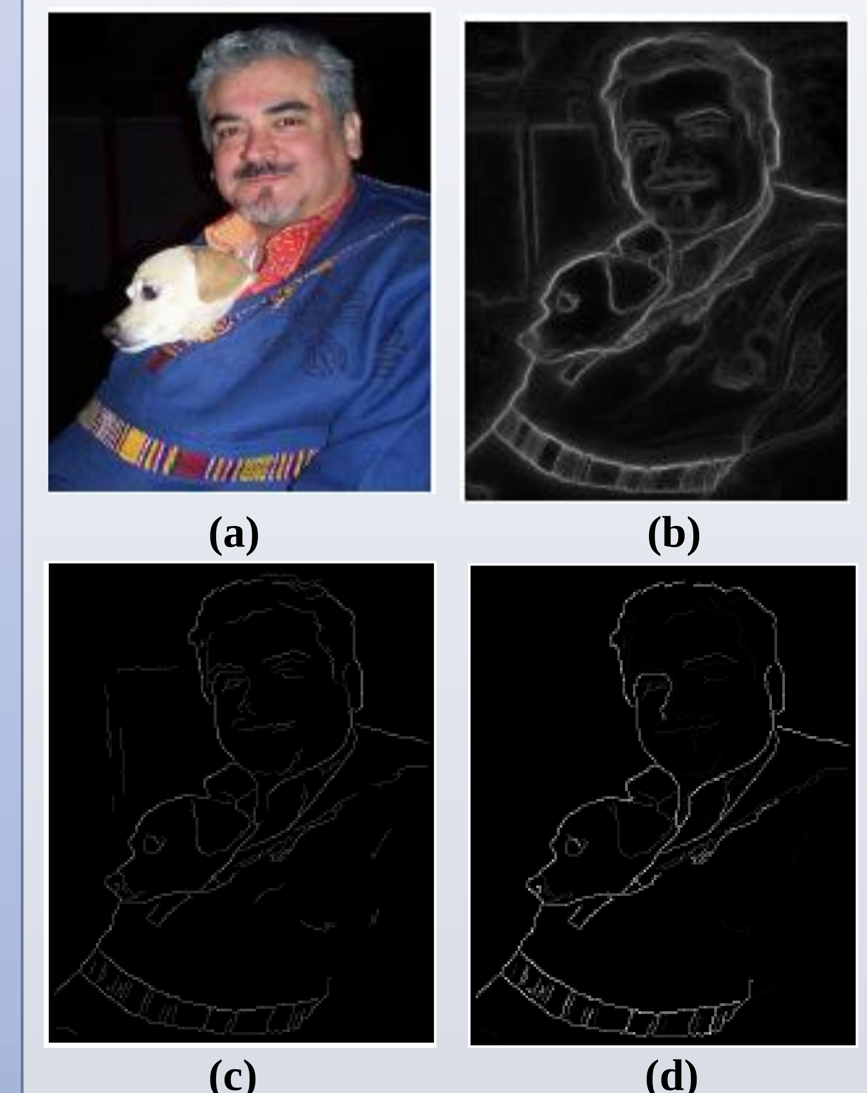

Fig. 2. (a) Original image (b) Edge map using OEF (c) After NMS and thresholding (d) Bayesian Probabilistic edge map (indicating saliency of edge segments).

$$p(sal|\mathfrak{s}) = \frac{p(sal)p(\mathfrak{s}|sal)}{p(sal)p(\mathfrak{s}|sal) + p(bg)p(\mathfrak{s}|bg)}$$

GRAPH CRF FORMATION

Algorithm 1 Algorithm for CRF learning and Prediction

1: **procedure** CRF–STRUCTURED PREDICTION
2: **Input:** $\mathcal{V}$: set of edge segments with 7-D feature vector, $\mathcal{E}$: set of edge links with 4-D feature vector
3: **Output:** Labels L for each node (edge segment)
4: Optimize an objective function (energy) with respect to parameter vector $\mathbf{W} = [\mathbf{W_1} \quad \mathbf{W_2}]$
5: $E(L|X) = \sum_{i\epsilon\mathcal{V}} \phi(l_i, X; \mathbf{W_1}) + \sum_{\{i,j\}\epsilon\mathcal{E}} \psi(l_i, l_j, X; \mathbf{W_2})$
6: where L is the structured label, X is the structured input features, l_i is the label of the node, $\mathbf{W_1}$ are the node parameters, $\mathbf{W_2}$ are the link parameters, $\phi(l_i, X; \mathbf{W_1})$ are unary potentials given as the inner product of the node features with node weights and $\psi(l_i, l_j, X; \mathbf{W_2})$ indicates pairwise potentials given as a linear function of link features and weights (shared over all links). The objective function is optimized using Block-coordinate Frank Wolfe Structured SVM to compute parameter $\mathbf{W}$.
7: Encode the structure of the problem in a joint feature function $\hat{\psi}(\mathbf{x}, \mathbf{y})$ as in prediction using,
8: $\hat{\mathbf{y}} = argmax_{\mathbf{y}\epsilon Y} \mathbf{W}^T \hat{\psi}(\mathbf{x}, \mathbf{y})$
9: where $\mathbf{y}$ is the structured label, $\mathbf{x}$ is the feature vector of a data point (node), Y denotes set of all possible labels {0,1} and $\hat{\mathbf{y}}$ is the prediction of the data point. Solve for $\hat{\mathbf{y}}$. Once the weights are learnt using the training data, predictions are made using AD3 inference algorithm.
10: **end procedure**

CONCLUSION

- **Novel object proposal generation algorithm operating in a *computationally efficient* learning based setting where the salient object edge density inside the bounding box is analyzed to score the proposal set.**
- ***High recall rates with lesser number of proposals* with varying IoU thresholds and subsequently making it more reliable in context of competing methods.**
- ***Ranked the key objects* according to their saliency.**