CENPARMI
Centre for Pattern Recognition and Machine Intelligence

Université Concordia University

Paper ID: 157

Object cosegmentation using deep Siamese network

Prerana Mukherjee, Brejesh Lall and Snehith Lattupally
Indian Institute of Technology, Delhi

Introduction

Cosegmentation refers to such class of problems which deals with the segmentation of the common objects from a given set of images without any priori knowledge about the foreground.

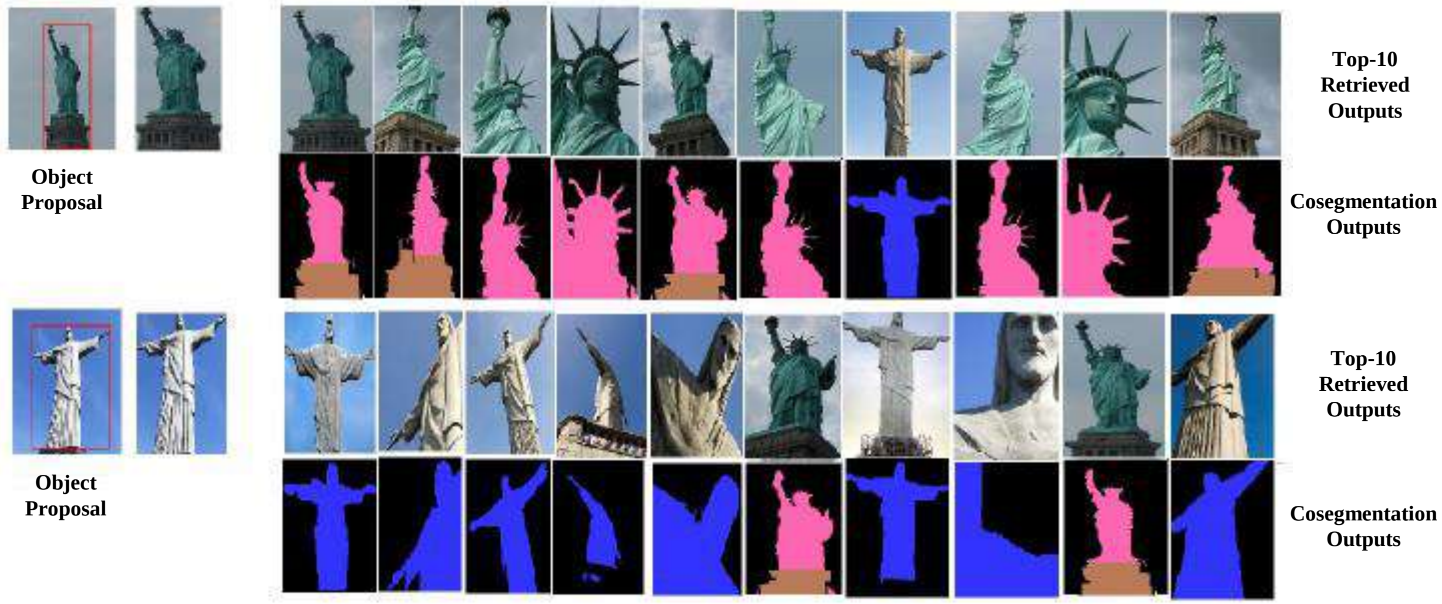


Fig. 1. Cosegmentation on Top-10 Retrieved outputs for query images.

Problems ?

- Huge variations in objects in terms of scale, rotation, illumination and affine changes.
- Lack of sufficient information about the foreground objects makes it highly complex to deal with it

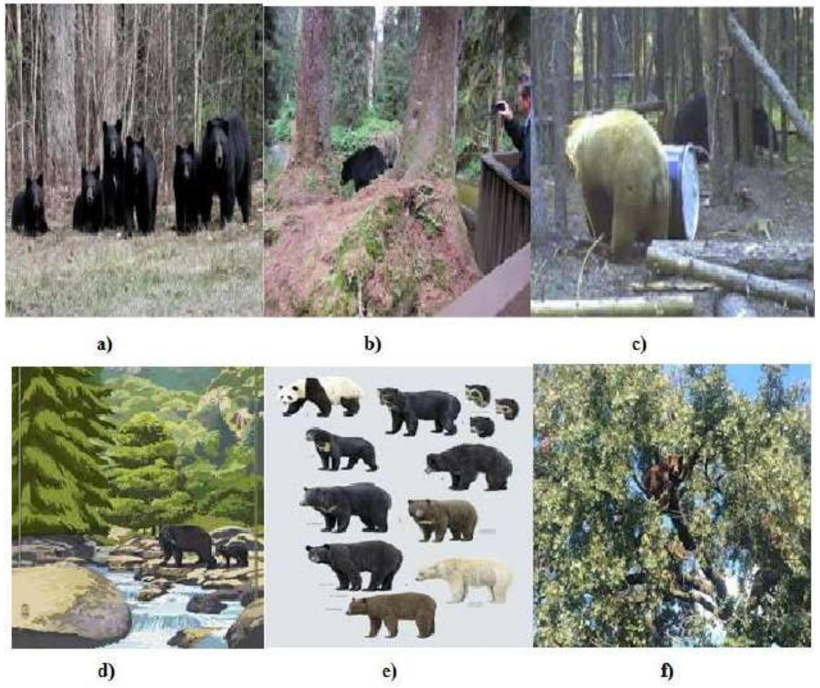


Fig. 2. Variational changes in appearance of the common object (bear): (a) Multiple instances of bear (b) Scale change (c) High occlusion (d) Synthetic changes (e) Intra-class variation (f) Cluttered background.

- Hence exploit commonness prior.

KEY CONTRIBUTIONS

- Cosegmentation is posed as a clustering problem to align the similar objects using Siamese network and segmenting them. We also train the Siamese network on non-target classes with no to little fine-tuning and test the generalization capability to target classes.
- Generation of visual summary of similar images based on relative relevance.

Dataset

We performed various experiments on available cosegmentation datasets. We used MSRC, ICoseg, Pascal, Coseg-Rep and animals datasets in our experiments.

MSRC dataset consists of 14 categories. Each category consists of 30 images, 213x320. It consists of categories like cow, car, chair, plane etc.

ICoseg dataset consists of 38 categories. Each categories consists of about 20 to 30 images, 300x500. It consists of categories like landmarks, sports, animals etc.

Methodology

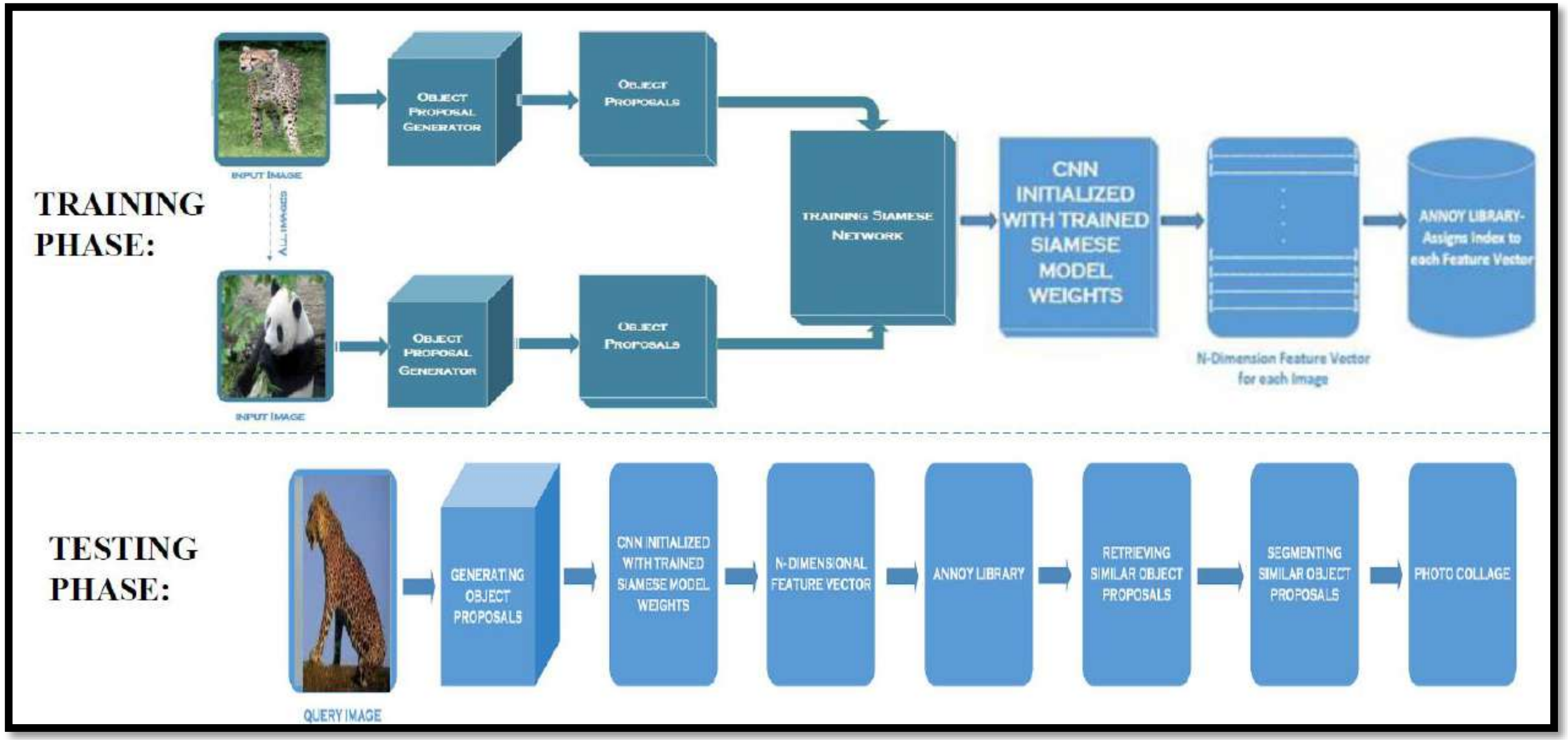
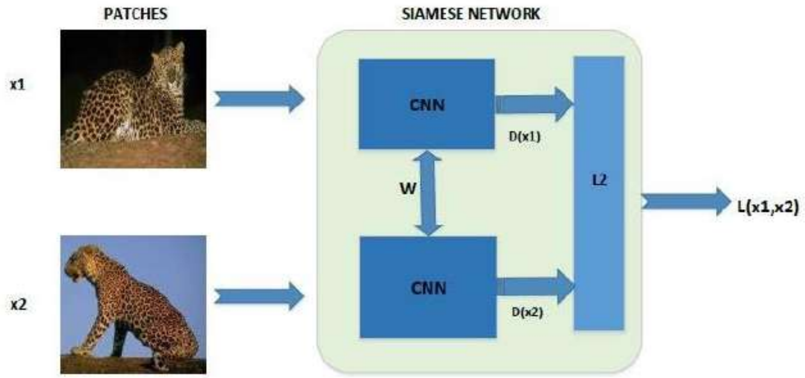

TRAINING PHASE:
INPUT IMAGE
ALL IMAGES
INPUT IMAGE
OBJECT PROPOSAL GENERATOR
OBJECT PROPOSALS
OBJECT PROPOSAL GENERATOR
OBJECT PROPOSALS
TRAINING SIAMESE NETWORK
CNN INITIALIZED WITH TRAINED SIAMESE MODEL WEIGHTS
N-Dimension Feature Vector for each Image
ANNOY LIBRARY- Assigns Index to each Feature Vector
TESTING PHASE:
QUERY IMAGE
GENERATING OBJECT PROPOSALS
CNN INITIALIZED WITH TRAINED SIAMESE MODEL WEIGHTS
N-DIMENSIONAL FEATURE VECTOR
ANNOY LIBRARY
RETRIEVING SIMILAR OBJECT PROPOSALS
SEGMENTING SIMILAR OBJECT PROPOSALS
PHOTO COLLAGE

Fig. 3 Overall Architecture

- Generating n-dimensional feature vector for each object proposal, from trained model.
- The feature vectors will be given to annoy(approximate nearest neighbors) library, which assigns an index to
each vector.
- Generating object proposals for test images and retrieving
the similar object proposals from trained Siamese network.
- Segmenting out the objects from retrieved similar object
proposals.

Siamese Network

- The input to the Siamese network are two input patches (object proposals) along with a similarity label. Similarity label '1' indicates that patches are similar while '0' indicates dissimilar patches.
- Two CNNs generate a N-Dimensional feature vector in forward pass. The N-Dimensional vectors are fed to the contrastive loss layer which helps in adjusting the weights such that positive samples are closer and negative samples are far from each other. Contrastive loss function penalizes the positive samples that are far away and negative samples that are closer.
- After training the Siamese Network, we deployed the trained model on test images. First we extracted the object proposals for the test images. A N-Dimensional feature vector is generated for each of the proposals. (N=256 in our experiments)
- The features extracted from the test image proposals are given to ANNOY library. ANNOY assigns indices to each of the features. To retrieve nearest neighbor for any of the feature, it measures the Euclidean distance to all other features and indices of neighbors are assigned in the increasing order of their Euclidean distance.



Segmentation

- Segmentation is performed on the retrieved similar object proposals. We used Fully convolutional Networks for semantic segmentation.
- It utilizes a skip architecture which combines the semantic information from deep (coarse information) and shallow (fine appearance information) layers.

Visual Summary based on relative importance

- A visual summary is created from the segmented proposals. While retrieving the similar object proposals using ANNOY library, we preserved the Euclidean distances corresponding to each of the proposals.
- A basic collage is formed with 10 slots constituting the most similar proposal (least Euclidean distance) getting a larger block.
- The remaining segmented objects are placed in the other slots and a background is added to the image.

Experimental Results

- The first baseline involves training the Siamese network with pretrained ILSVRC Imagenet models. The weights are fine-tuned for target classes as in the datasets and then segmentation is performed on the clustered test set data.
- In the second baseline, we train the network on non-target classes and test the generalization ability on target classes.

Experimental Results

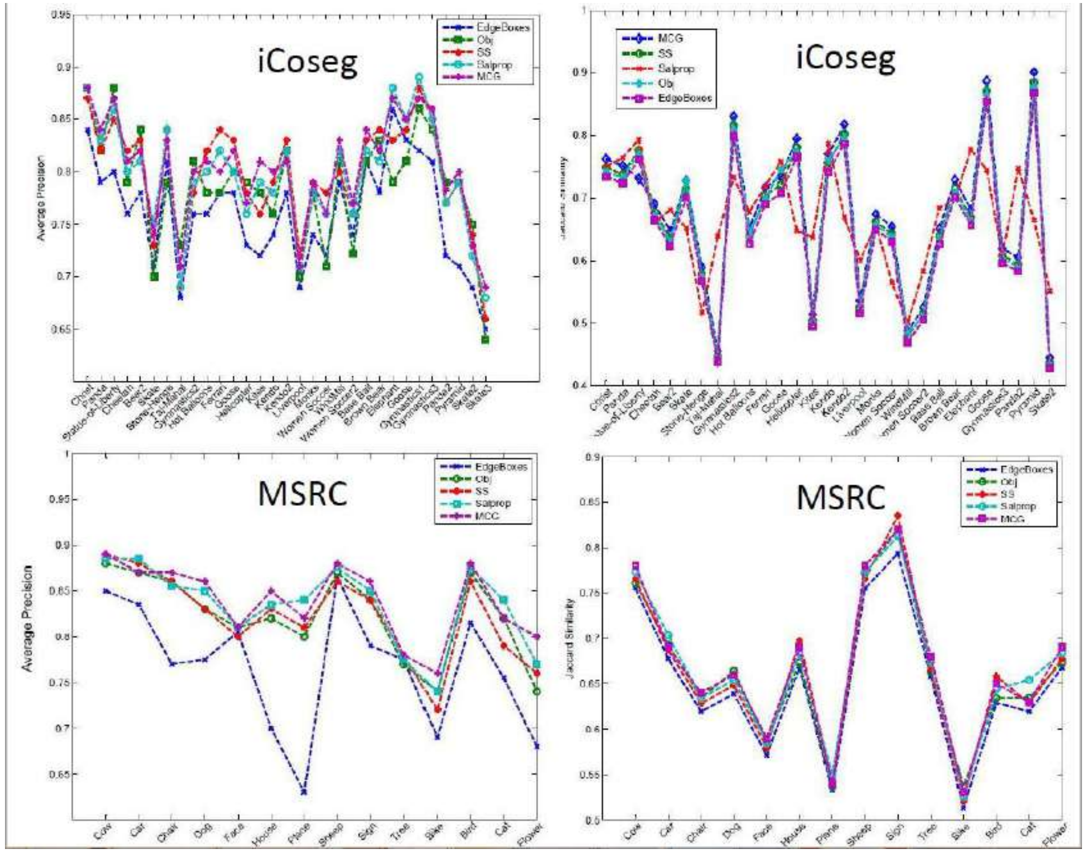


Fig. 4 Performance analysis of various object proposal generation methods with proposed architecture.

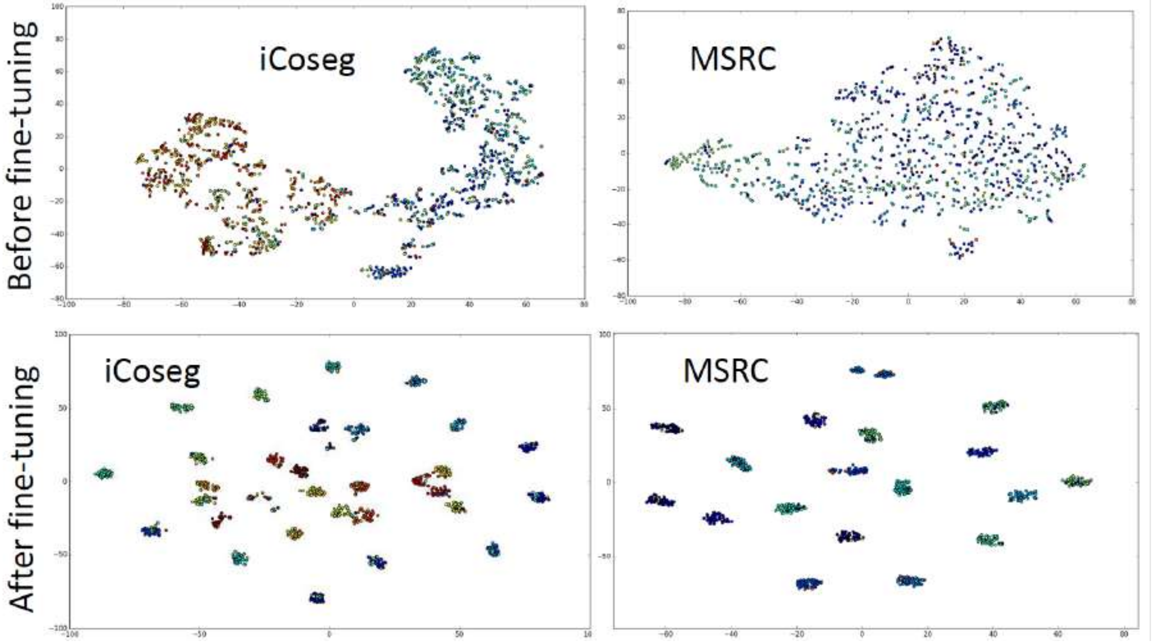
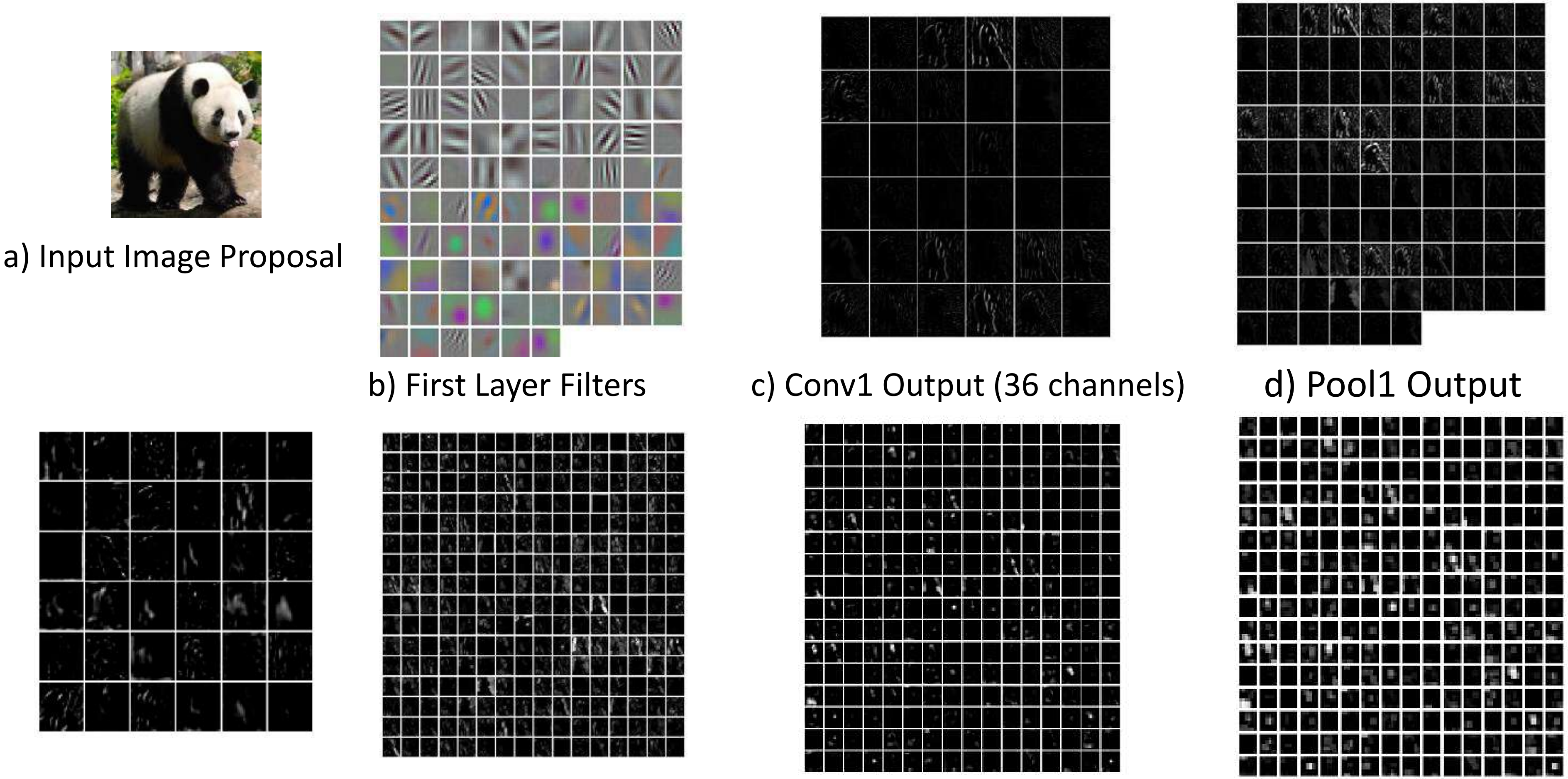


Fig. 5 Visualization of iCoseg and MSRC Training set using t-SNE

Experimental Results

Visualization of output of different layers for given input proposal.



a) Input Image Proposal

b) First Layer Filters

c) Conv1 Output (36 channels)

d) Pool1 Output

e) Conv2 Output

f) Pool2 Output

g) Conv5 Output

h) Pool5 Output

Experimental Results

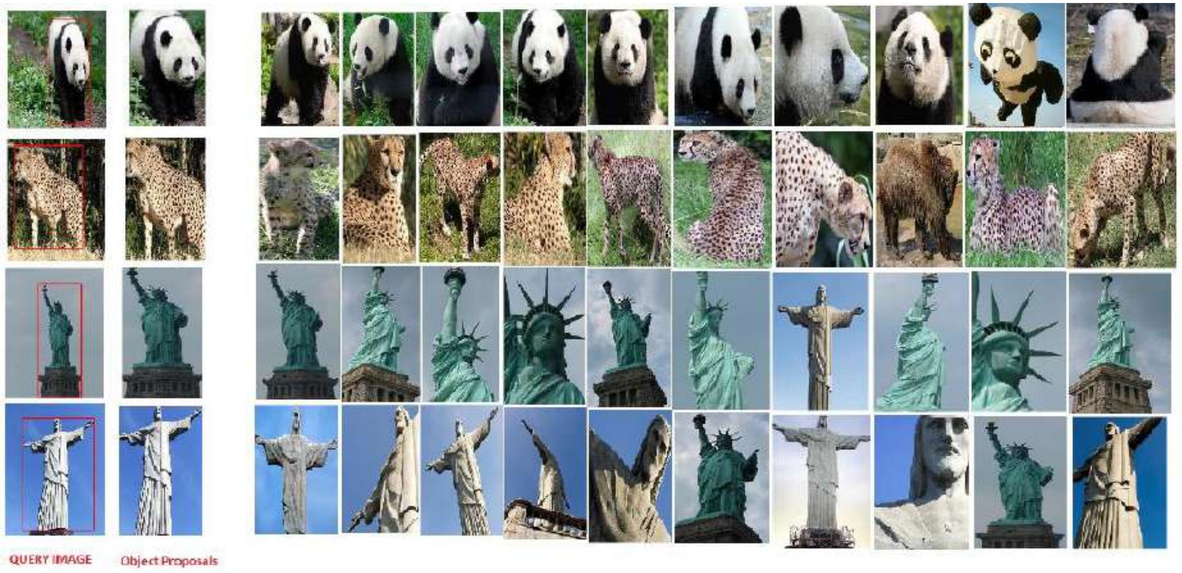


Fig. 6 Top 10 Retrieved Results

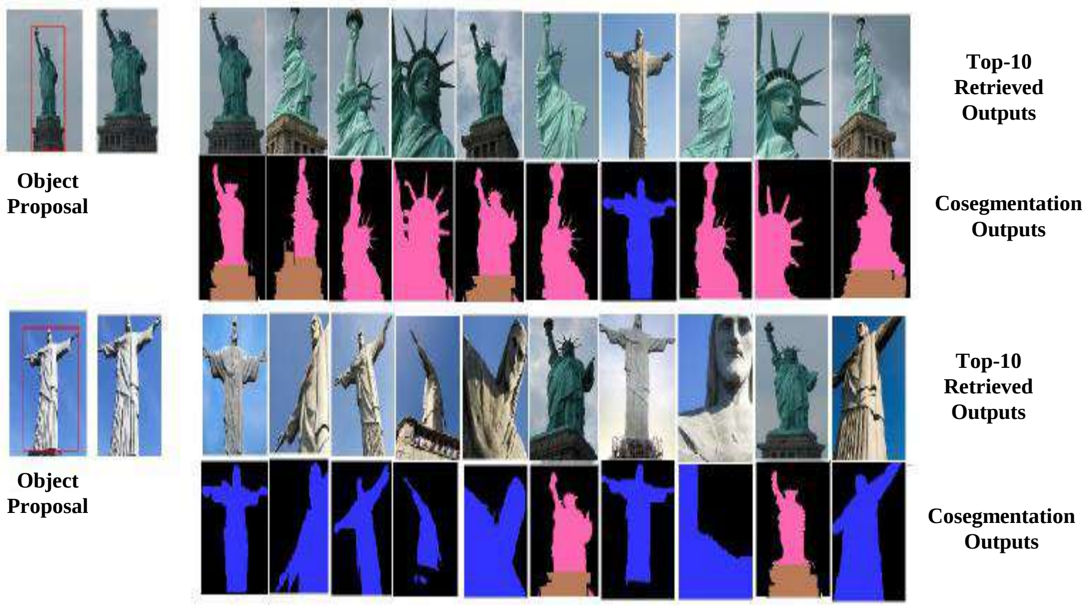


Fig. 7 Cosegmentation on Top-10 Retrieved outputs for query images.

Experimental Results

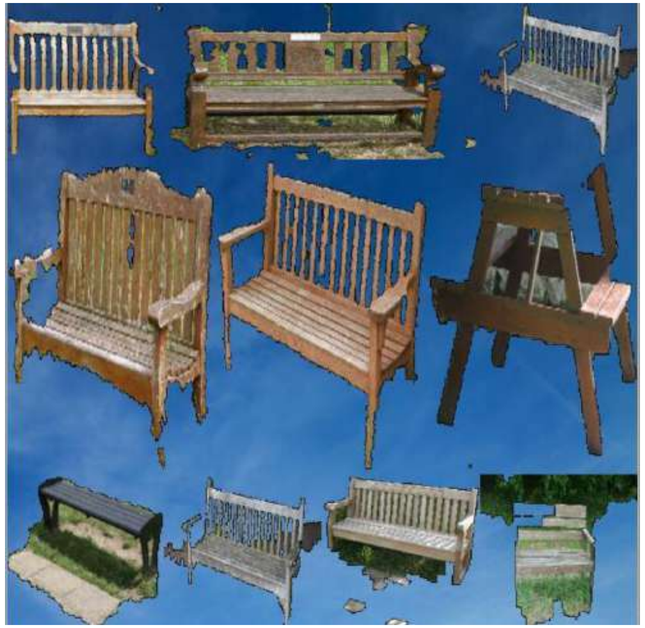


Fig. 8 Example of Collage results for Chair class (MSRC).

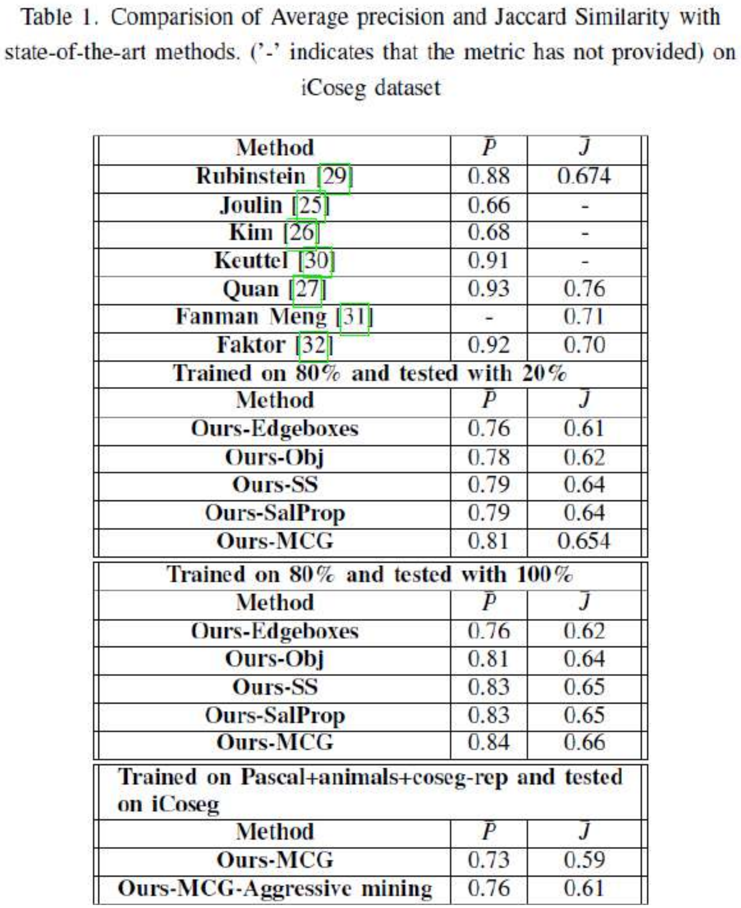


Table 1. Comparision of Average precision and Jaccard Similarity with state-of-the-art methods. ('-' indicates that the metric has not provided) on iCoseg dataset

Method	P	J
Rubinstein [29]	0.88	0.674
Joulin [25]	0.66	-
Kim [26]	0.68	-
Keuttel [30]	0.91	-
Quan [27]	0.93	0.76
Fanman Meng [31]	-	0.71
Faktor [32]	0.92	0.70
Trained on 80% and tested with 20%		
Method	P	J
Ours-Edgeboxes	0.76	0.61
Ours-Obj	0.78	0.62
Ours-SS	0.79	0.64
Ours-SalProp	0.79	0.64
Ours-MCG	0.81	0.654
Trained on 80% and tested with 100%		
Method	P	J
Ours-Edgeboxes	0.76	0.62
Ours-Obj	0.81	0.64
Ours-SS	0.83	0.65
Ours-SalProp	0.83	0.65
Ours-MCG	0.84	0.66
Trained on Pascal+animals+coseg-rep and tested on iCoseg		
Method	P	J
Ours-MCG	0.73	0.59
Ours-MCG-Aggressive mining	0.76	0.61

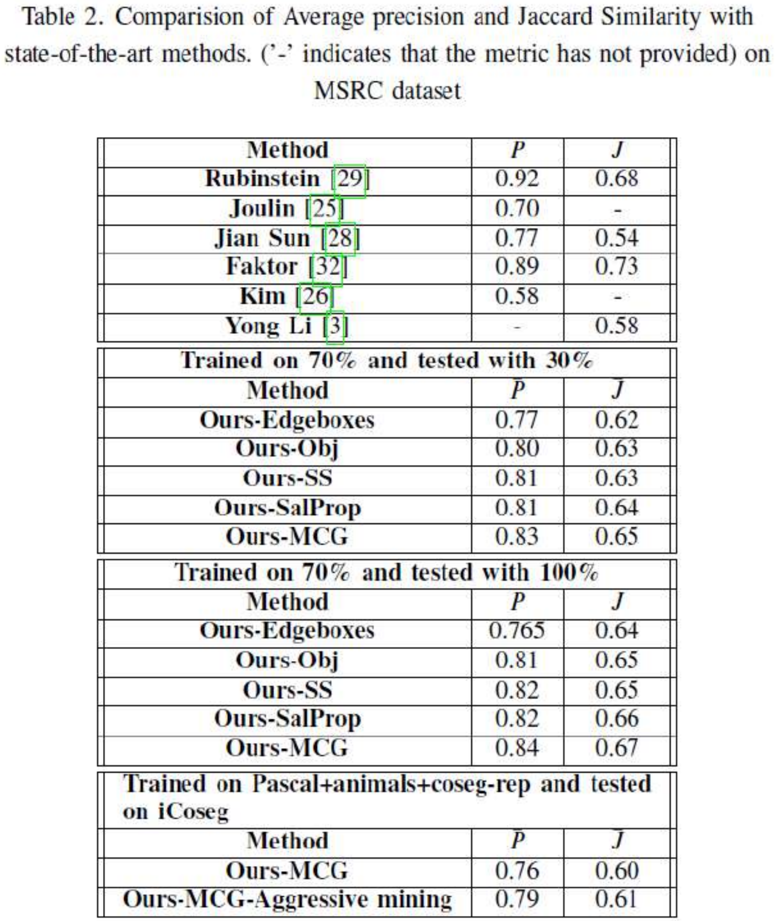


Table 2. Comparision of Average precision and Jaccard Similarity with state-of-the-art methods. ('-' indicates that the metric has not provided) on MSRC dataset

Method	P	J
Rubinstein [29]	0.92	0.68
Joulin [25]	0.70	-
Jian Sun [28]	0.77	0.54
Faktor [32]	0.89	0.73
Kim [26]	0.58	-
Yong Li [3]	-	0.58
Trained on 70% and tested with 30%		
Method	P	J
Ours-Edgeboxes	0.77	0.62
Ours-Obj	0.80	0.63
Ours-SS	0.81	0.63
Ours-SalProp	0.81	0.64
Ours-MCG	0.83	0.65
Trained on 70% and tested with 100%		
Method	P	J
Ours-Edgeboxes	0.765	0.64
Ours-Obj	0.81	0.65
Ours-SS	0.82	0.65
Ours-SalProp	0.82	0.66
Ours-MCG	0.84	0.67
Trained on Pascal+animals+coseg-rep and tested on iCoseg		
Method	P	J
Ours-MCG	0.76	0.60
Ours-MCG-Aggressive mining	0.79	0.61

Conclusion

- We addressed object cosegmentation and posed it as a clustering problem using deep Siamese network to align the similar images which are segmented using semantic segmentation.
- We compared the performance of various object proposal generation schemes on Siamese architecture.
- We performed extensive evaluation on iCoseg and MSRC dataset and demonstrated that the deep features can encode the commonness prior and thus provide a more discriminative representation for the features.

References

- S. Vicente, C. Rother, and V. Kolmogorov, “Object cosegmentation,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 2217–2224.
- Y. Li, J. Liu, Z. Li, H. Lu, and S. Ma, “Object co-segmentation via salient and common regions discovery,” Neurocomputing, vol. 172, pp. 225–234, 2016.
- A. Joulin, F. Bach, and J. Ponce, “Discriminative clustering for image co-segmentation,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 1943–1950.
- G. Kim and E. P. Xing, “On multiple foreground cosegmentation,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 837–844.
- R. Quan, J. Han, D. Zhang, and F. Nie, “Object co-segmentation via graph optimized-flexible manifold ranking,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 687–695.
- J. Sun and J. Ponce, “Learning dictionary of discriminative part detectors for image categorization and cosegmentation,” International Journal of Computer Vision, vol. 120, no. 2, pp. 111–133, 2016.
- M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, “Unsupervised joint object discovery and segmentation in internet images,” IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), June 2013.
- F. Meng, J. Cai, and H. Li, “Cosegmentation of multiple image groups,” Computer Vision and Image Understanding, vol. 146, pp. 67–76, 2016.
- A. Faktor and M. Irani, “Co-segmentation by composition,” in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1297–1304.

THANKYOU