DeepBox: Learning Objectness with Convolutional Networks

Paper by Weicheng Kuo, Bharath Hariharan, Jitendra Malik

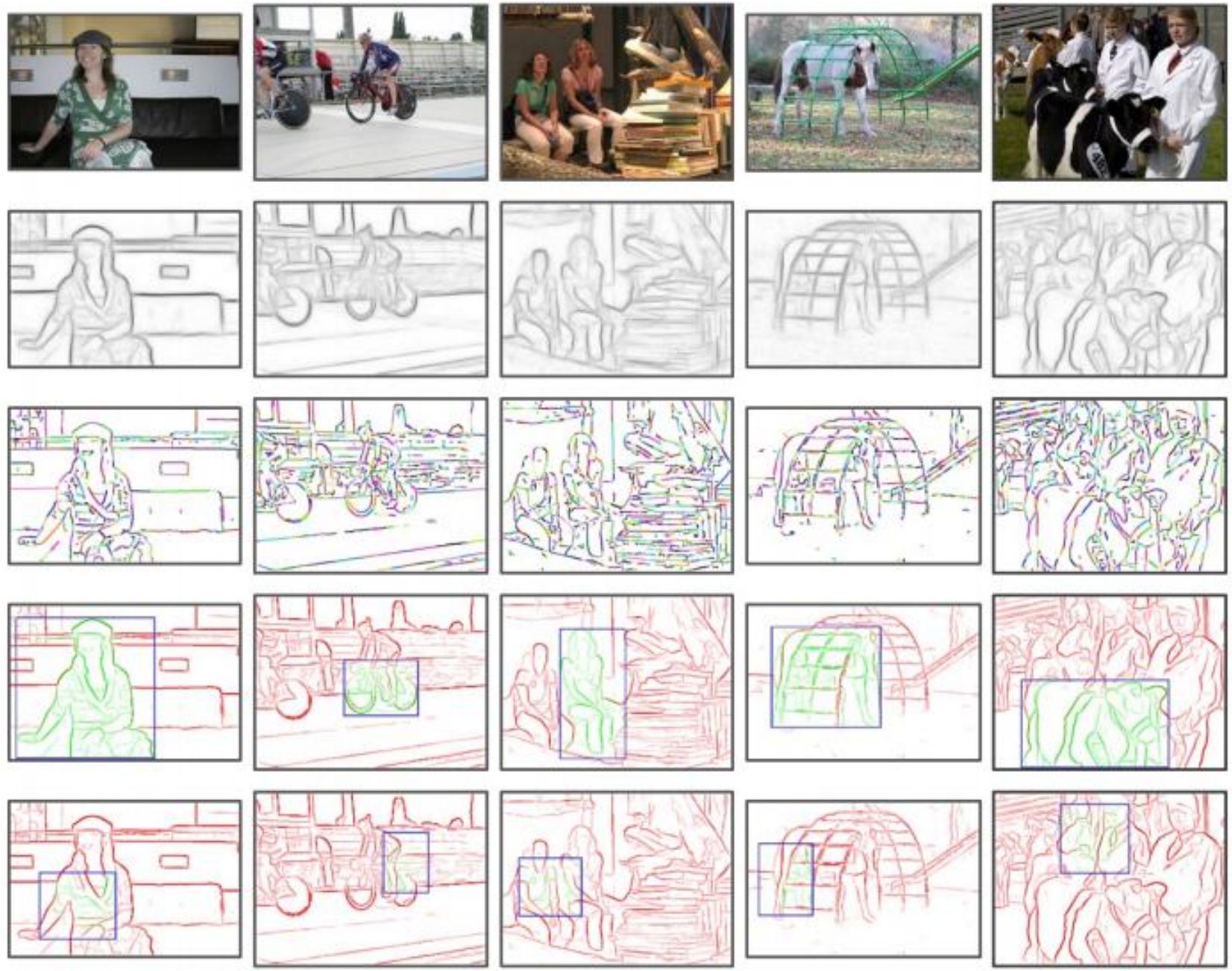
Presented by Divya Patel

10/12/2016

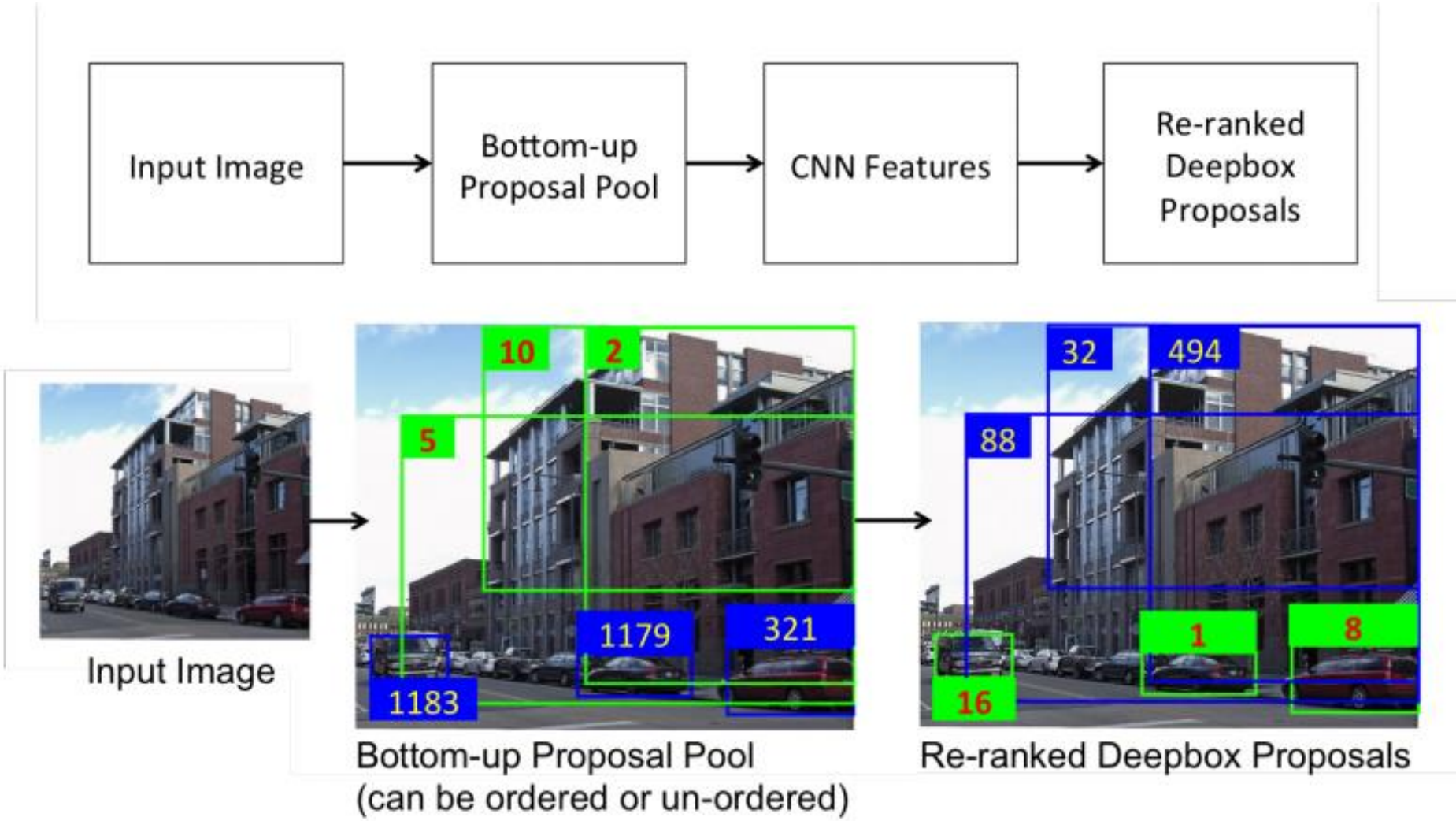
Task

- Find a way to exploit the semantic notion of objectness in order to speed up the task of ranking object proposals, while maintaining similar object detection performance as state-of-the-art methods

Edge Boxes



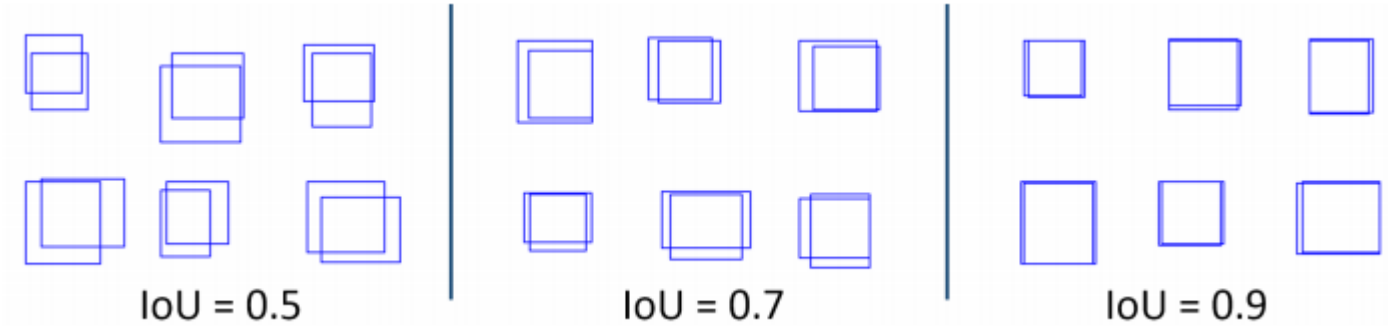
Process


Input Image
Bottom-up Proposal Pool
CNN Features
Re-ranked Deepbox Proposals
10
2
5
1179
321
1183
32
494
88
1
8
16
Input Image
Bottom-up Proposal Pool
(can be ordered or un-ordered)
Re-ranked Deepbox Proposals

Network Architecture

Conv(11,96,4) -> ReLI -> pool(3,2) -> ReLU -> conv(5,256,1) -> ReLU fc(1024) -> ReLU -> fc(2)

Final AUC of 0.74



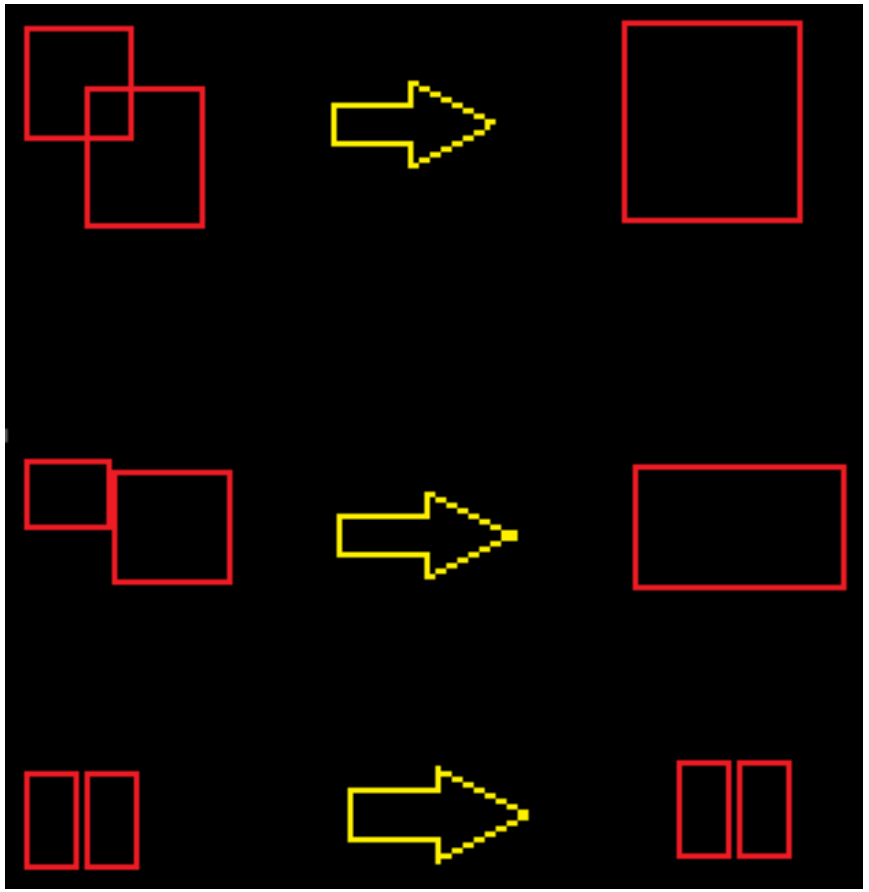
	AUC on PASCAL VOC (IoU = 0.5 / 0.7)
Original	0.76 / 0.62
Outputs of fc6 changed to 1024	0.76 / 0.62
Image crop from 227x227 to 120x120	0.745 / 0.591
Remove fc6	0.737 / 0.591
Remove conv5 and conv4	0.654 / 0.514
Remove conv5 and conv4 and conv3	0.693 / 0.554

Key

Conv(kernel size, stride, output channels)

Pool(kernel size, stride)

Overlapping Bounding Boxes Optimization



Training - Initialization

- First two convolutional layers were initialized using the ImageNet model
- Fully connected layers were initialized randomly

Training – Sliding Windows

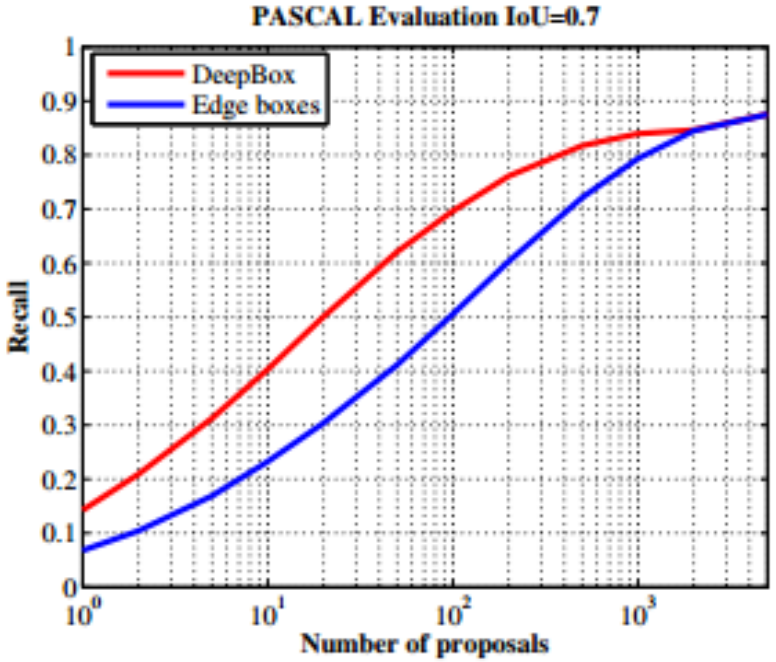
- Given the ground truths objects of an image, we generate negative windows unless a window overlaps over beta = 0.5 with a ground truth object, in which case, it is discarded
- Windows are traversed so that they only overlap one another up to IoU = 0.65
- Teaches the Neural Network to learn to separate the background from objects

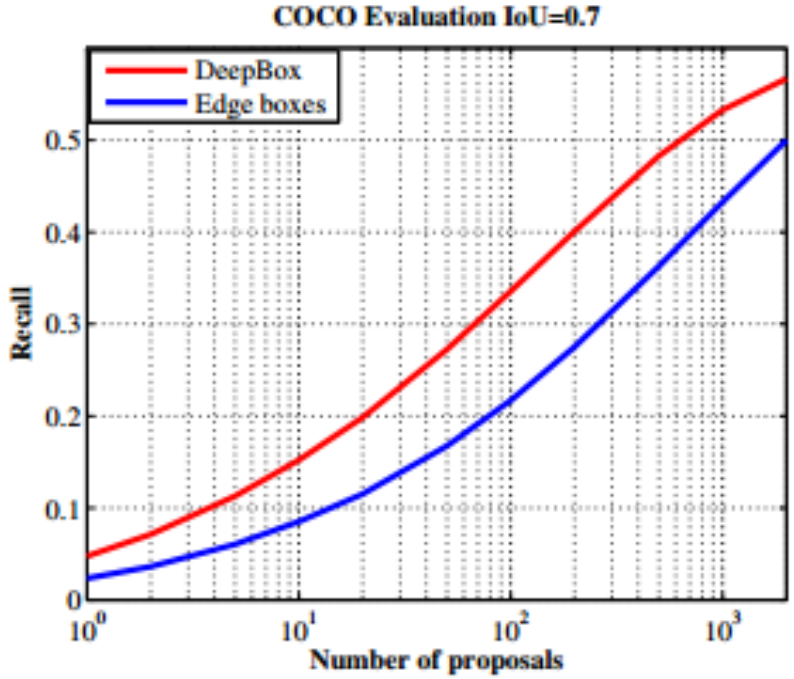
Training – Hard Negatives

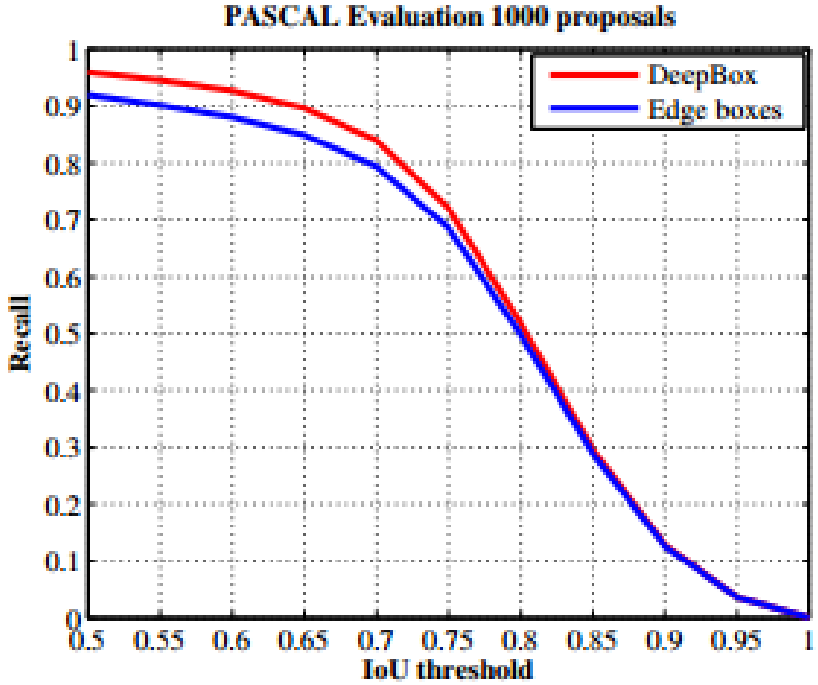
- Similar procedure to training on sliding windows, except
- Teaches the neural network to learn the notion of complete objects and learn from the errors made by the bottom-up proposer

Comparison – DeepBox (right) vs Edge box (left)

Results

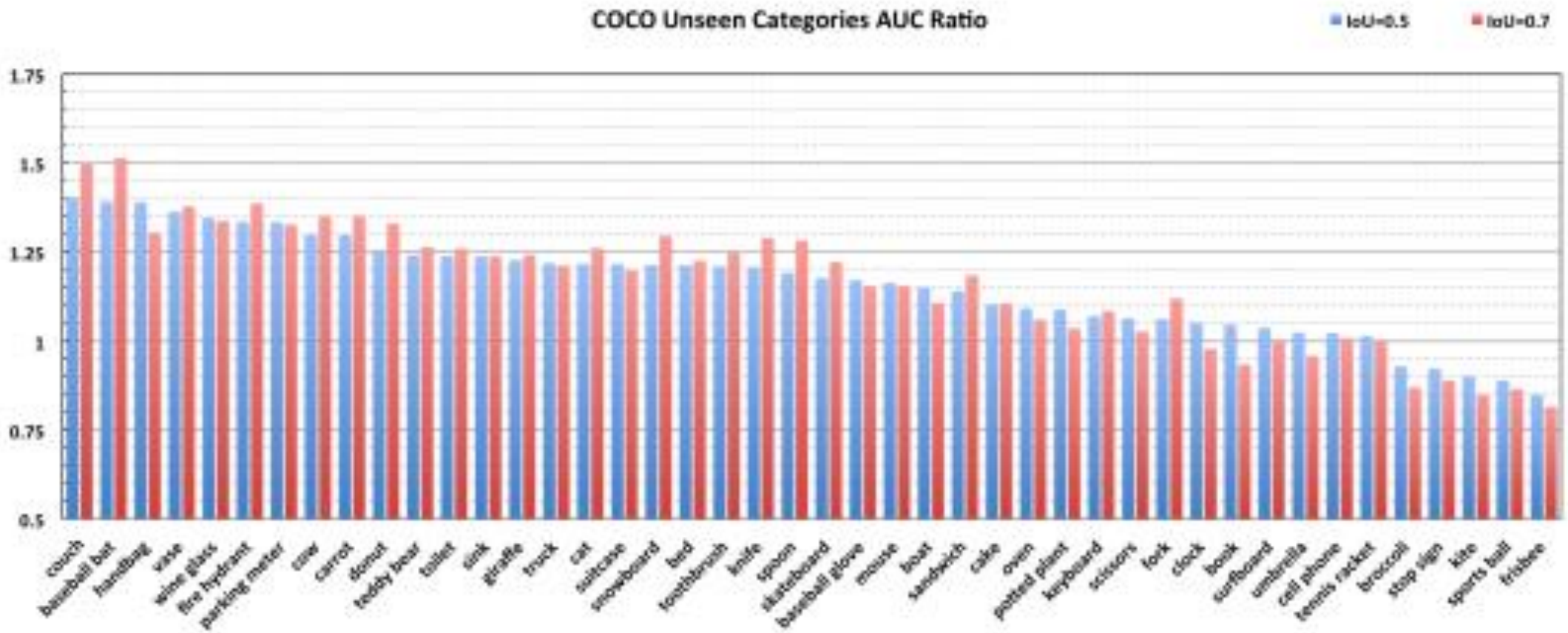






	AUC	25%	50%	75%	Recall (%)	Time (s)
BING[5]	0.20	292	-	-	29	**0.2**
Ranta[24]	0.23	184	584	-	68	10
Objectness[1]	0.27	27	-	-	39	3
Rand. P.[21]	0.35	42	349	3023	80	1
Rahtu [23]	0.37	29	307	-	70	3
SelSearch [29]	0.40	28	199	1434	**87**	10
CPMC [3]	0.41	15	111	-	65	250
MCG [2]	0.48	10	81	871	83	34
E.B [30]	0.47	12	108	800	**87**	0.25
DeepBox	**0.60**	**3**	**20**	**183**	**87**	2.5

Results for Unseen Categories



Questions?