Day 3 Lecture 12

Saliency

Prof. Xavier Giró, Prof. Kevin McGuinness
Student: Junting Pan

Elisa Sayrol

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Department of Signal Theory and Communications
Image Processing Group

Saliency

Saliency

What have you seen?

Saliency

Lighthouse

Saliency

Lighthouse

House

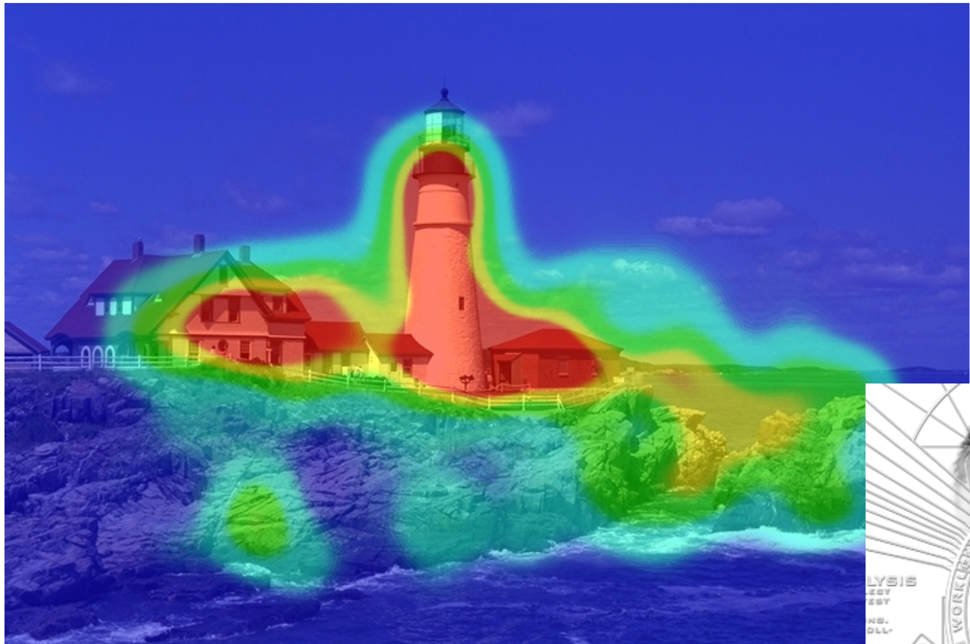
Saliency

Lighthouse

House

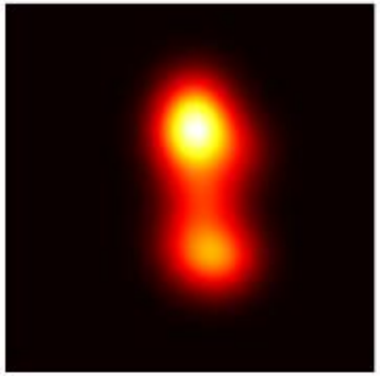
Rocks

Saliency



Saliency Map

The Goal is to obtain the Saliency Map of an Image.
Regression problem, not Classification



Original Image

Ground Truth
Saliency Map
(Eye-Fixation Map)

Data Bases: Groundtruth generation

Eye Tracker

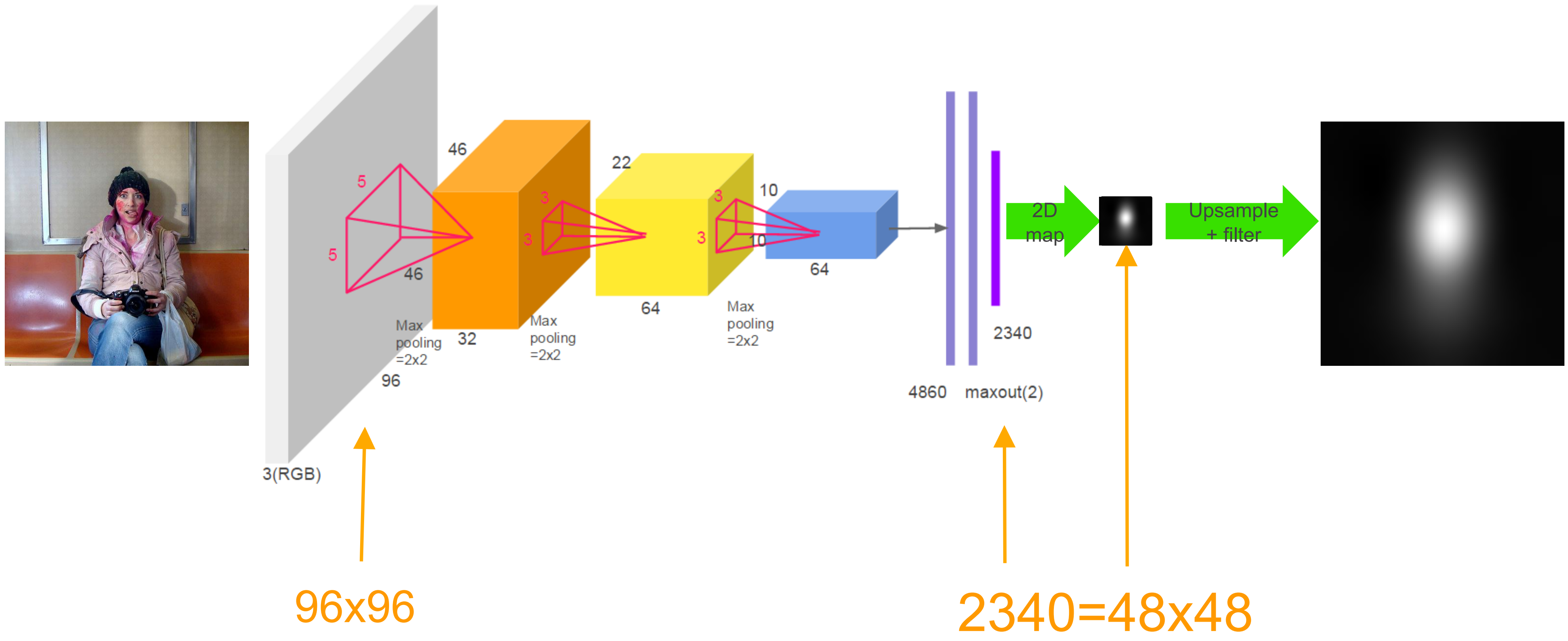
Mouse Click

DataBases

	TRAIN	VALIDATION	TEST
SALICON [Jiang'15]	10,000	5,000	5,000
iSun [Xu'15]	6,000	926	2,000
CAT2000 [Borji'15]	2,000	-	2,000
MIT300 [Judd'12]	300	-	-
Pascal-S	850		

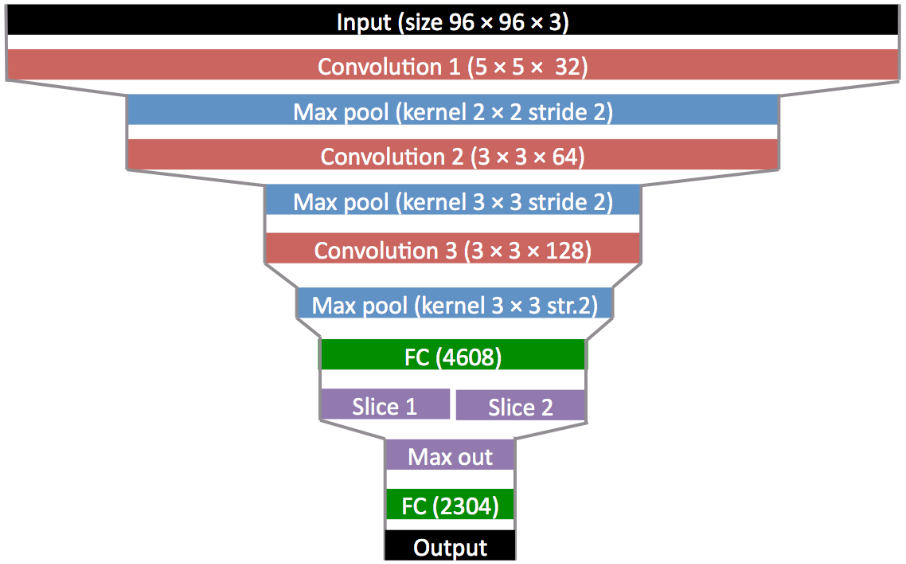
Other databases: http://saliency.mit.edu/datasets.html

Architectures: Junting Net (Shallow Network)



Architectures: Junting Net (Shallow Network)

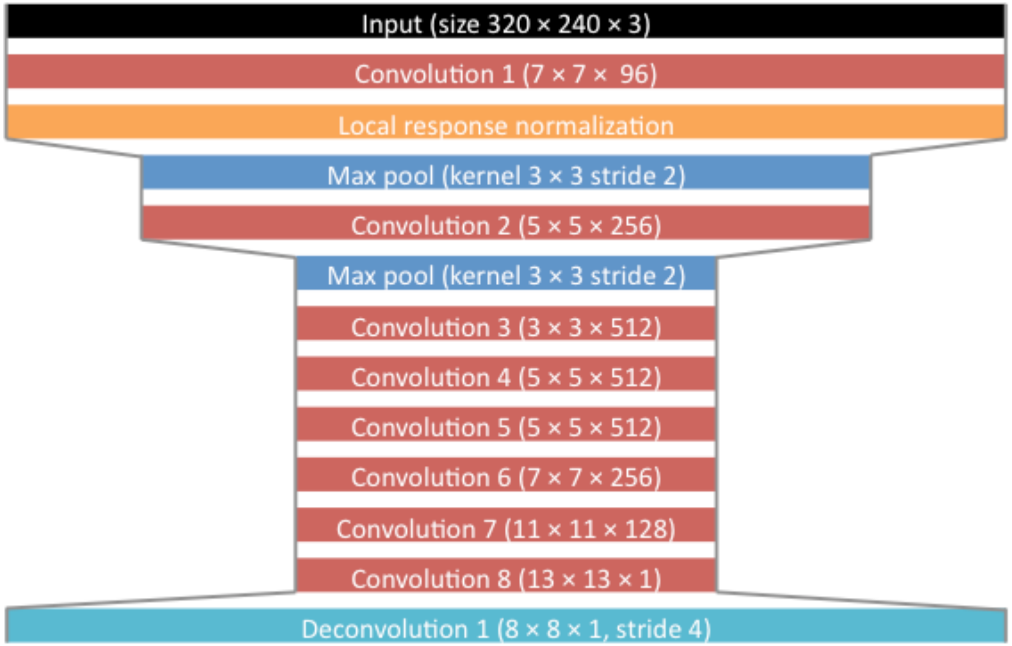
Winner of the LSUN Challenge 2015!!



Loss function	Mean Square Error (MSE)
Weight initialization	Gaussian distribution
Learning rate	0.03 to 0.0001
Mini batch size	128
Training time	7h (SALICON) / 4h (iSUN)
Acceleration	SGD+ nesterov momentum (0.9)
Regularisation	Maxout norm
GPU	NVidia GTX 980

Shallow and Deep Convolutional Networks for Saliency Prediction
Junting Pan, Kevin McGuinness, Elisa Sayrol, Noel O'Connor, Xavier Giro-i-Nieto, *CVPR 2016*

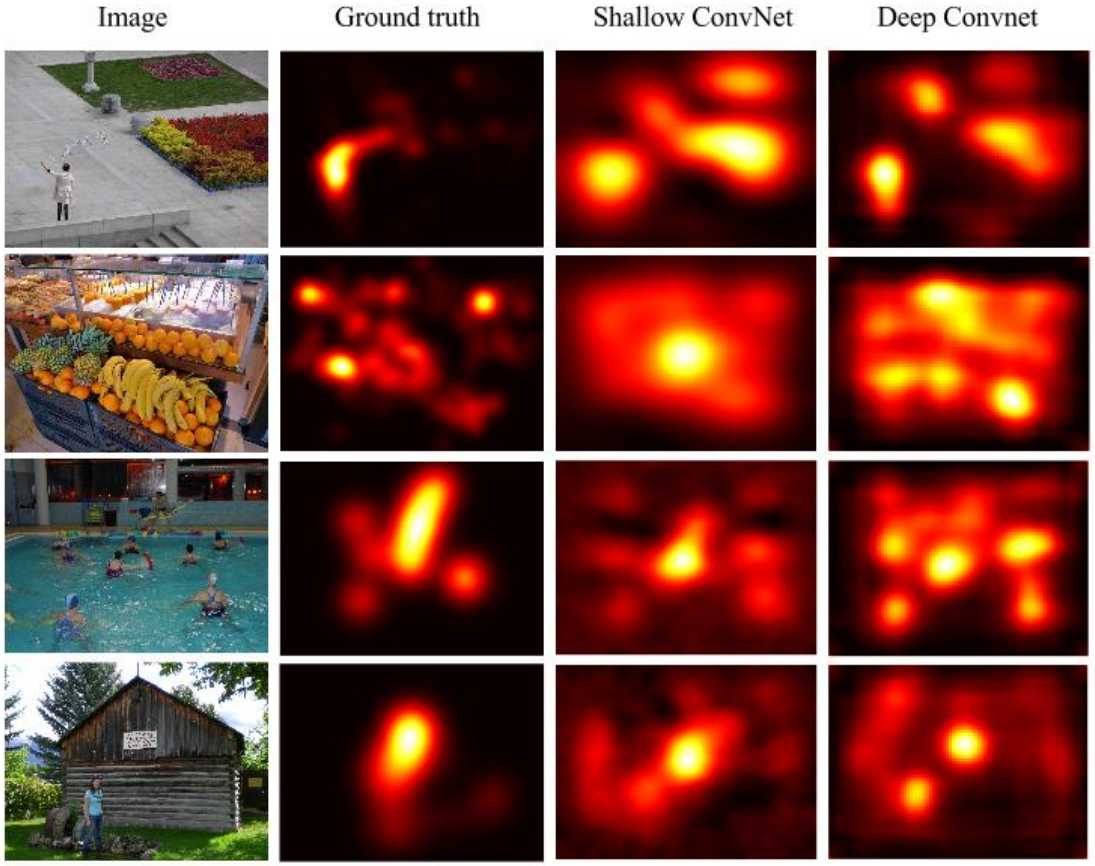
Architectures: SalNet (Deep Network)



Loss function	Mean Square Error (MSE)
Weight initialization	First 3 layers pre-trained with VGG, the rest of the layers random distribution
Learning rate	0,01(halved every 100 iterations)
Mini batch size	2 images for 24.000 iterations
Training time	15h
Acceleration	SGD+ nesterov momentum (0.9)
Regularisation	L2 weight
GPU	NVidia GTX Titan

Shallow and Deep Convolutional Networks for Saliency Prediction
Junting Pan, Kevin McGuinness, Elisa Sayrol, Noel O'Connor, Xavier Giro-i-Nieto, *CVPR 2016*

Quality Results



Architectures: Junting Net (Shallow Network) Winner of the LSUN Challenge 2015!!

Results from CVPR LSUN Challenge 2015 (iSUN Database)

Method	Similarity	CC	AUC_shuffled	AUC_Borji	AUC_Judd
UPC	0.6833	0.8230	0.6650	0.8463	0.8693
Xidian	0.5713	0.6167	0.6484	0.7949	0.8207
WHU_IIP	0.5593	0.6263	0.6307	0.7960	0.8197
LCYLab	0.5474	0.5699	0.6259	0.7921	0.8133
Rare 2012 Improved	0.5199	0.5199	0.6283	0.7582	0.7846
Baseline: BMS[1]	0.5026	0.3465	0.5885	0.6560	0.6914
Baseline: GBVS[2]	0.4798	0.5087	0.6208	0.7913	0.8115
Baseline: Itti[3]	0.4251	0.3728	0.6024	0.7262	0.7489

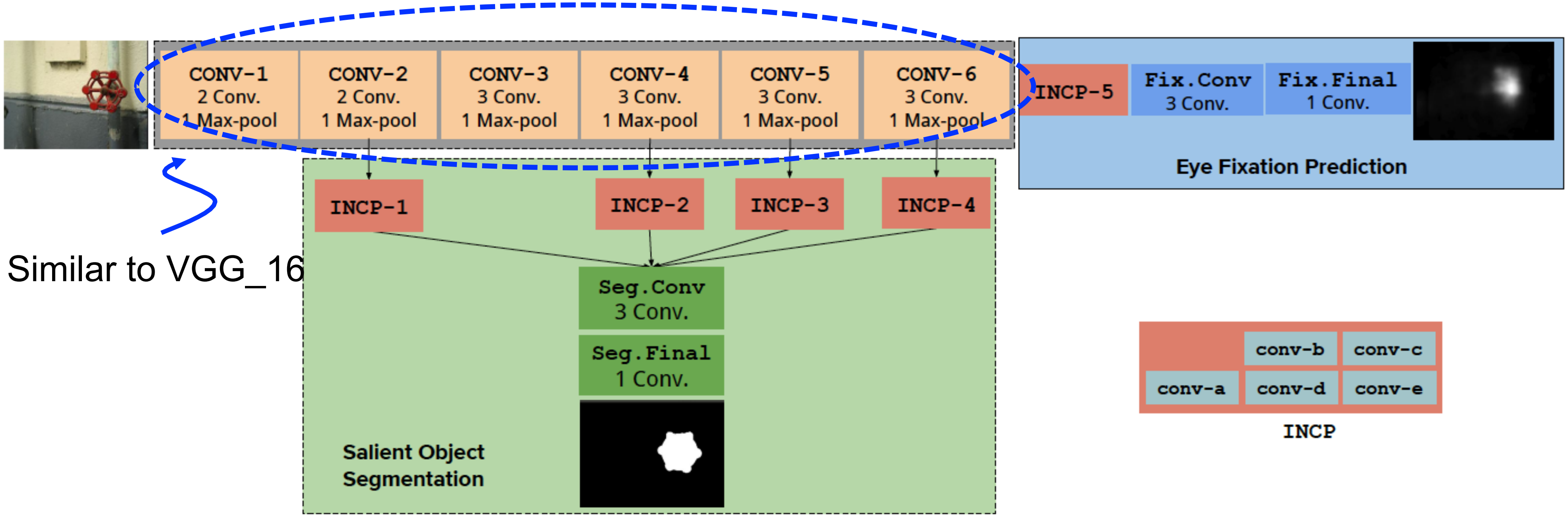
Quantitative Results

	Similarity	CC	AUC shuffled	AUC Borji	AUC Judd
Baseline: Infinite Humans	1.00	1.00	0.80	0.87	0.91
SALICON [11]	0.60	0.74	0.74	0.85	0.87
DeepFix	0.67	0.78	0.71	0.80	0.87
Deep Gaze 1 [17]	0.39	0.48	0.66	0.83	0.84
Deep Convnet	**0.52**	**0.58**	**0.69**	**0.82**	**0.83**
BMS [31]	0.51	0.55	0.65	0.82	0.83
eDN [27]	0.41	0.45	0.62	0.81	0.82
GBVS [9]	0.48	0.48	0.63	0.80	0.81
Judd [15]	0.42	0.47	0.60	0.80	0.81
Shallow Convnet	**0.46**	**0.53**	**0.64**	**0.78**	**0.80**
Mr-CNN [20]	0.48	0.48	0.69	0.75	0.79
Rare 2012 Improved [22]	0.46	0.42	0.67	0.75	0.77
Baseline: One human	0.38 – 0.46	0.52 – 0.65	0.63 – 0.67	0.66 – 0.71	0.80 – 0.83

Table 6. Results of the MIT300 dataset.

Metrics: Saliency and Human Fixations: State-of-the-art and Study of Comparison Metrics
Nicolas Riche, Matthieu Duvinage, Matei Mancas, Bernard Gosselin and Thierry Dutoit, iccv 2013

Architectures: Saliency Unified (Very Deep Network)



Saliency Unified: A Deep Architecture for simultaneous Eye Fixation Prediction and Salient Object Segmentation
Srinivas S S Kruthiventi, Vennela Gudisa, Jaley H Dholakiya and R. Venkatesh Babu, *CVPR 2016*

Quantitative Results

Dataset	Metric	ITTI [2]	GBVS [13]	AWS [46]	BMS [47]	eDN [48]	MrCNN [26]	JuntingNet [49]	**Proposed**
PASCAL-S [9]	s-AUC ↑	0.64	0.65	0.67	0.67	0.65	–	0.69	0.72
	EMD ↓	1.21	1.16	1.38	1.32	1.29	–	1.03	0.73
	NSS ↑	1.30	1.36	1.12	1.28	1.42	–	1.90	2.22
DUT-OMRON [40]	s-AUC ↑	0.78	0.81	0.78	0.79	0.80	–	0.83	0.83
	EMD ↓	1.47	1.32	1.62	1.58	1.56	–	1.37	1.03
	NSS ↑	1.54	1.71	1.51	1.66	1.33	–	2.03	3.02
MIT1003 [14]	s-AUC ↑	0.66	0.66	0.69	0.69	0.66	0.71	0.68	0.73
	EMD ↓	2.33	2.19	2.54	2.40	2.39	2.30	1.91	1.49
	NSS ↑	1.06	1.17	1.07	1.19	1.24	1.28	1.60	2.08
IS [43]	s-AUC ↑	0.66	0.67	0.72	0.71	0.61	–	0.65	0.70
	EMD ↓	1.30	1.22	1.49	1.43	1.49	–	1.11	0.77
	NSS ↑	1.50	1.58	1.58	1.74	1.27	–	1.72	2.30

Table 3. Quantitative results of our approach on eye fixation prediction compared against other state-of-the art methods on PASCAL-S, DUT-OMRON, MIT1003 and IS datasets. The best results are shown in red and the second best in blue.

Saliency Unified: A Deep Architecture for simultaneous Eye Fixation Prediction and Salient Object Segmentation
Srinivas S S Kruthiventi, Vennela Gudisa, Jaley H Dholakiya and R. Venkatesh Babu, *CVPR 2016*