


## CERTIFICATE

It is certified that the work contained in this thesis entitled “COMPLEX SALIENT OBJECT DETECTION USING PROBABILISTIC MODELLING AND DEEP LEARNING” by “Surya Kant Singh” has been carried out under my supervision and that it has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy and SOTA.

  
24.5.2022

Prof. Rajeev Srivastava

Professor

Department of Computer Science and Engineering

Indian Institute of Technology

(Banaras Hindu University)

Varanasi-221005

पर्यवेक्षक/Supervisor

संगणक विज्ञान एवं अभियांत्रिकी विभाग

Department of Computer Sc. & Engg

भारतीय प्रौद्योगिकी संस्थान

Indian Institute of Technology

(काशी हिन्दू विश्वविद्यालय)

(Banaras Hindu University)

वाराणसी/Varanasi-221005

## DECLARATION BY THE CANDIDATE

I, Surya Kant Singh, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of Prof. Rajeev Srivastava from JULY-2017 to MAY-2022, at the Department of Computer Science and Engineering, Indian Institute of Technology (Banaras Hindu University) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully lifted up any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and included them in this thesis and cited as my own work.

Date : 24/05/2022

Place : Varanasi

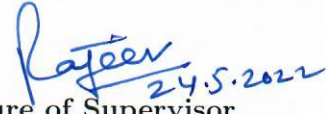


Signature of the Student

(Surya Kant Singh)

## CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.



Signature of Supervisor

(Prof. Rajeev Srivastava)

Department of Computer Science and Engineering  
Indian Institute of Technology  
(Banaras Hindu University)

Varanasi-221005

Supervisor  
समयक विज्ञान एवं अभियांत्रिकी विभाग  
Department of Computer Sc. & Engg  
भारतीय प्रौद्योगिकी संस्थान  
Indian Institute of Technology  
(काशी हिन्दू विश्वविद्यालय)  
(Banaras Hindu University)  
वाराणसी/Varanasi-221005



Signature of Head of Department

(Prof. Sanjay Kumar Singh)

Professor & Head

समयक विज्ञान एवं अभियांत्रिकी विभाग

Department of Computer Sc. & Engg

भारतीय प्रौद्योगिकी संस्थान

Indian Institute of Technology

(बनारस हिन्दू यूनिवर्सिटी)

(Banaras Hindu University)

वाराणसी-221005 / Varanasi-221005

## COPYRIGHT TRANSFER CERTIFICATE

---

Title of the Thesis: COMPLEX SALIENT OBJECT DETECTION  
USING PROBABILISTIC MODELLING AND DEEP LEARNING

Name of the Student: Surya Kant Singh

### Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi, all rights under copyright that may exist in and for the above thesis submitted for the award of the DOCTOR OF PHILOSOPHY.

*Date : 24/05/2022*

*Place : Varanasi*

  
(Surya Kant Singh)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

*To*  
*My Beloved Parents*  
*Mrs. JANKEE SINGH Mr. NANDLAL SINGH*  
*and*  
*My Better Half*  
*Mrs. NEETU SINGH*

## ACKNOWLEDGEMENTS

---

I want to take this opportunity to express my deep sense of gratitude to all who helped me directly or indirectly during this thesis work. Firstly, I would like to thank my supervisor, **Prof. Rajeev Srivastava**, for being a great mentor and the best adviser I could ever have. His advice, encouragement, and critics are a source of innovative ideas, inspiration, and causes behind the successful completion of this Thesis work. The confidence shown on me by him was the most significant source of inspiration for me. It has been a privilege working with him for several years. I am highly obliged to all the faculty members of the Computer Science and Engineering Department for their support and encouragement. I express my sincere thanks to Prof. K. K. Shukla of the Department of Computer Science and Engineering and Prof. Subir Das, Department of Mathematical Sciences, IIT (BHU), for providing continuous support, encouragement, and advice. I express my sincere thanks to all the Professors, Deans, office staff, supporting staff, and Ph.D. Research Scholars of Indian Institute of Technology (BHU) Varanasi India. I express my gratitude to Director, Registrar, Deans, Heads, Students and Alumni of the Indian Institute of Technology (BHU) Varanasi.

My memory of the study period at IIT (BHU) can never be complete without mentioning my fellow research scholars. Special thanks to Dr. Vibhav Prakash Singh, Dr. Ankit Jaiswal, Dr. Gargi Srivastava, Mr. Saurabh Arora and Mr. Santosh Kumar Tripathy for their great help and cooperation.

I extend special thanks to the non-teaching staff in the department, particularly the Mr. Manoj Kumar Singh, Mr. Ravi Kumar Bharti, Mr. Prakhar Kumar, Mr. Ritesh Singh, and Mr. Shubham Pandey for their consistent support.

My parents, Mrs. Jankee Singh and Mr. Nandlal Singh, who gave me the power and brain to work out on this research and their help at every level, made me see this success. I owe thanks to my wife, Mrs. Neetu Singh Solanki, for his continued and unfailing love, support, and understanding during my pursuit of Ph.D. degree that made the completion of my thesis possible. You were always around at times

I thought that it is impossible to continue, you helped me to keep things in perspective. I greatly value his contribution and sincerely appreciate his belief in me. I want to share my source of happiness during the up and down in the Ph.D. journey are the smiles of my daughter Namya and son Shresth.

I extend my thanks to my uncle Baban singh, my sister Shivangi Singh and my brother Dhan jee, whose blessings gave me the internal power to do hard work.

Words are insufficient to express my profound sense of gratitude to respected Prof. Shreesh Chaudhary and my friends Dr. Roshan Singh, Mr. Akash kumar singh, and Mr. Vipul Kumar Jha, whose encouragement gave me physical and moral strength throughout my career as well in the present research. I extend my thanks to my father in-law Dr. D.V Singh and my Mother in-law Mrs. Bindu Singh who are part of my inspiration. Finally, I would like to wind up by paying my heartfelt thanks and prayers to the Almighty for his unbound love and grace.

**- Surya Kant Singh**



# Contents

<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xxiii</b>
<b>Abbreviations</b>	<b>xx</b>
<b>Symbols</b>	<b>xxiv</b>
<b>Preface</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Motivation and Scope . . . . .	3
1.3 Problem Statement . . . . .	5
1.4 The Research Objectives of Thesis . . . . .	6
1.5 Contributions to the Thesis . . . . .	7
1.6 Outline of the Thesis . . . . .	9
<b>2 Literature Survey</b>	<b>11</b>
2.1 Literature Review . . . . .	12
2.1.1 Visual Saliency . . . . .	12
2.1.2 Visual saliency models . . . . .	13
2.1.2.1 Fixation Prediction models . . . . .	13
2.1.2.2 Salient object detection models . . . . .	14
2.1.3 Saliency computation domains: . . . . .	16
2.1.3.1 Bottom-up-search dependent statistical models: . . .	17
2.1.3.2 Top-down-task dependent deep learning based models:	17
2.1.3.3 Hybrid and Probabilistic models : . . . . .	18
2.2 Salient Object Detection . . . . .	20
2.2.1 Statistical and Probabilistic Models(2D Saliency) . . . . .	20
2.2.1.1 Biologically Inspired Models . . . . .	21
2.2.1.2 Global Contrast-Based Models . . . . .	22



2.2.1.3	Regional Contrast-Based Models . . . . .	23
2.2.1.4	Background Approximation Based Models . . . . .	29
2.2.1.5	Psychology Inspired Models . . . . .	31
2.2.1.6	Summary . . . . .	35
2.2.2	RGBD(3D) Statistical Models . . . . .	37
2.2.2.1	Summary . . . . .	43
2.2.3	RGBD(3D) Deep learning Models . . . . .	43
2.2.3.1	Multi-Layer Perceptron (MLP) Models . . . . .	44
2.2.3.2	Fully Convolutional Network (FCN) Models . . . . .	45
2.2.3.3	RGBD 3D SOD Models . . . . .	46
2.2.3.4	Single-stream Models . . . . .	47
2.2.3.5	Multi-stream Models . . . . .	47
2.2.3.6	Fusion Model . . . . .	48
2.2.3.7	Skip-Connection based models . . . . .	53
2.2.3.8	Attention mechanism . . . . .	54
2.2.3.9	Summary . . . . .	56
2.3	Research Gaps . . . . .	57
2.4	Dataset . . . . .	61
2.4.1	RGB(2D) Dataset . . . . .	62
2.4.2	RGBD(3D) Dataset . . . . .	65
2.5	Performance Evaluation Metrics . . . . .	69
2.6	Conclusion . . . . .	73
<b>3</b>	<b>Probabilistic Contrast and Edge Enhanced Global Topographical Surface based Complex RGB Salient Object Detection</b>	<b>75</b>
3.1	Introduction . . . . .	77
3.2	The Proposed Models . . . . .	83
3.2.1	Probabilistic contrast based complex salient object detection .	84
3.2.1.1	Initialization through Poisson probabilistic contrast (PC) . . . . .	84
3.2.1.2	Poisson probabilistic distribution . . . . .	86
3.2.1.3	Regional contrast integration with global-PC . . . . .	87
3.2.1.4	Background suppression model . . . . .	88
3.2.1.5	Saliency Refinement Model . . . . .	89
3.2.1.6	Theoretic foundation . . . . .	90
3.2.2	Experiments and Results of Probabilistic Contrast based Model	94
3.2.2.1	Evaluating Parameters Setting . . . . .	95
3.2.2.2	Ablation study . . . . .	98
3.2.2.3	Comparative study . . . . .	100
3.2.2.4	Comparison with Deep Learning Based Models . . . . .	106
3.2.3	Edge Enhanced Global Topographical Saliency . . . . .	108

3.2.3.1	Initial global topographical surface(GTS) through iterative Laplacian of Gaussian (ILG) . . . . .	108
3.2.3.2	Regional contrast integration within GTS . . . . .	109
3.2.3.3	Saliency enhancement . . . . .	112
3.2.4	Result Analysis of Edge Enhanced Global Topographical Saliency	114
3.2.4.1	Evaluating parameters setting . . . . .	115
3.2.4.2	Successive steps validation . . . . .	116
3.2.4.3	Comparative analysis . . . . .	116
3.2.4.4	Comparison of proposed method GTS with Deep Learning Based Models: . . . . .	122
3.2.4.5	Comparison of average computational speed : . . . . .	122
3.3	Conclusion . . . . .	123
<b>4</b>	<b>RGBD Complex salient object detection with improved probabilistic contrast and global concave topographical saliency</b>	<b>125</b>
4.1	Introduction . . . . .	126
4.2	The proposed method . . . . .	131
4.2.1	Initialization through global concave surface (GCS) . . . . .	131
4.2.1.1	Improved Poisson probabilistic contrast (IPC) . . . . .	132
4.2.1.2	Contour based global surface . . . . .	133
4.2.2	Regional contrast integration into GCS . . . . .	134
4.2.2.1	Regional saliency integration with GCS . . . . .	135
4.2.2.2	Background estimation model . . . . .	136
4.2.2.3	Saliency enhancement . . . . .	137
4.2.3	Theoretic foundation . . . . .	138
4.2.4	The proposed algorithm . . . . .	140
4.3	Experimental and Result Analysis . . . . .	141
4.3.1	Data-Set . . . . .	141
4.3.2	Performance Measures . . . . .	142
4.3.3	Parameters and Constraints Selection . . . . .	142
4.3.4	Successive steps validation . . . . .	143
4.3.5	Comparative Analysis . . . . .	144
4.3.6	Comparison with other Own proposed methods . . . . .	148
4.4	Conclusion . . . . .	149
<b>5</b>	<b>CSA-Net:Deep Cross Complementary Self Attention and Modality-Specific Preservation for Saliency Detection</b>	<b>151</b>
5.1	Introduction . . . . .	152
5.2	The Proposed Model . . . . .	157
5.2.1	Modality-specific Saliency Fusion Model(MSF) . . . . .	159
5.2.2	Cross -Complementary Self-Attention -CSA . . . . .	160
5.2.3	Complementary Features Fusion Model . . . . .	163

5.2.3.1	Cross-Complementary Features Fusion(CFF)	163
5.2.3.2	Intra-Complementary Features Fusion(IFF)	165
5.2.4	Optimal Selective Saliency Fusion Model-OSS	166
5.2.5	Loss Function	168
5.3	Experimental Details:	169
5.3.1	Data-Set	169
5.3.2	Implementation Details	170
5.3.3	Training Details	171
5.3.4	Evaluation Metrics	172
5.3.4.1	S-Measure	173
5.3.4.2	F-Measure	173
5.3.4.3	Mean Absolute Error(MAE)	174
5.3.4.4	E-Measure( $E_\psi$ )	174
5.4	Comparison and Result Analysis	176
5.4.1	Ablation Analysis	178
5.4.1.1	Validation of Network Structure and Fusion model	178
5.4.1.2	Effectiveness of CSA-Net Model	180
5.4.1.3	Failure Cases and Analyses	181
5.5	Conclusion	182
<b>6</b>	<b>SL-Net: Self-Learning and Mutual Attention based Distinguished Window for RGBD Complex Salient Object Detection</b>	<b>185</b>
6.1	Introduction	186
6.2	The proposed method	192
6.2.1	Overview	192
6.2.2	Enhanced Encoded Feature through Composite Backbone	194
6.2.3	Mutual Attention Based Distinguished Window-MADW	194
6.2.4	Self-Learning based Dense Decoding-SDD	198
6.2.4.1	Global Localized Feature	198
6.2.4.2	Cross-complementary fusion(CF)	199
6.2.4.3	Self Learning Aggregation model	200
6.2.5	Loss Function	201
6.3	Experiment and Result Analysis	202
6.3.1	Data-Set	202
6.3.2	Evaluation metrics	203
6.3.3	Implementation Details	203
6.3.4	Comparison and Result Analysis	204
6.3.5	Ablation Analysis	208
6.3.5.1	Effectiveness of Mutual attention based distinguished window-MADW	209
6.3.5.2	Effectiveness of SDD and Fusion model	210
6.3.5.3	Effectiveness of Proposed Composite model	211

6.3.6	Discussion of Failure case and limitations . . . . .	212
6.3.7	Comparison of own proposed methods . . . . .	213
6.4	Conclusion . . . . .	214
<b>7</b>	<b>Conclusion and Future Work</b>	<b>215</b>
7.1	Conclusions . . . . .	215
7.2	Future Work . . . . .	219
	<b>References</b>	<b>221</b>
	<b>List of Publications</b>	<b>245</b>



# List of Figures

1.1	A set of images to demonstrate the challenge to extract salient objects in complex and cluttered backgrounds. . . . .	4
2.1	Visual saliency and Computational domains. . . . .	16
2.2	Computational domains and associated outstanding models in Salient Object Detection. . . . .	18
2.3	Regional saliency and Superpixel based salient object detection. . . . .	24
2.4	The 3D saliency and Depth Information. . . . .	38
2.5	A set of RGB(2D) Datasets. . . . .	63
2.6	A set of RGBD(3D) Datasets. . . . .	67
3.1	The motivation for proposing probabilistic models to minimize discrepancies in interior and exterior saliency. . . . .	78
3.2	The importance of Global Topographical Surface in salient objects from complex and clutter backgrounds . . . . .	82
3.3	Block diagram of proposed probabilistic model. . . . .	83
3.4	Enhancement and suppression of luminance (ESL) with relative variance in chrominance planes, where (a) Input image (b) Input image luminance plane (c) Normalized luminance plane. . . . .	86
3.5	Successive saliency computation (a) Input image (b)Initial $S_{pc}$ through, (c) Final $S_{pc}$ through, (d) $S_{PR}$ regional color, spatial, depth enhance saliency (e) $S_c$ central saliency, (f) Final saliency $S$ through . . . . .	90
3.6	PR-Curves on MSRA dataset for demonstrating the evaluating parameters. (a) Angular threshold, (b) Complex image indicator( $\lambda$ ). Blue line represents the without ESL. . . . .	95
3.7	Visual comparison of saliency maps related with contrast based method. . . . .	96
3.8	Comparison of purely contrast-based saliency on (1) MSRA (2) PASCAL (3) DUTOMRON, dataset respectively . . . . .	97
3.9	Comparison of purely contrast-based saliency on (1) ECSSD and (2) IMGSA dataset respectively . . . . .	98
3.10	Steps wise validation of proposed method using F-Measure. . . . .	99
3.11	Visual comparison of saliency maps of a complex image having cluttered background. . . . .	101

3.12	Quantitative comparison of saliency Maps on (1)ECSSD (2) IMGSA (3) MSRA dataset respectively . . . . .	102
3.13	Quantitative comparison of saliency maps on (1) PASCAL and (2) DUTOMRON dataset respectively . . . . .	103
3.14	Quantitative comparison of saliency maps using F-Measure. . . . .	104
3.15	Qualitative comparison of proposed method with deep learning based methods. . . . .	106
3.16	Demonstration of successive steps of proposed method GTS (a) Input image, (b) Initialization through iterative ILG, (c) $S_{GTS}$ , (d) Regional saliencies integration $S_{FS}$ (e) $S_{Central}$ central saliency, (f) Gaussian weighted background suppression and central saliency enhancement $S$	112
3.17	Qualitative visual comparison among Various salient object detection methods. . . . .	115
3.18	Comparison of Global-Contrast-based saliency using (a) PR-Curve (b) ROC-Curve on (1) MSRA (2) PASCAL dataset respectively. . . . .	117
3.19	Comparison of Global-Contrast-based saliency using (a) PR-Curve (b) ROC-Curve on (1) DUTOMRON and (2) ECSSD dataset respectively.	118
3.20	Comparison of Global-Contrast-based saliency using F-Measurer on (1) MSRA (2) PASCAL (3) DUTOMRON and (4) ECSSD dataset respectively. . . . .	120
3.21	Comparison of deep learning based methods with GTS on PASCAL dataset using PR-Curve and F-Measure. . . . .	121
4.1	Saliency detection in complex and clutter background with minimize discrepancies in interior, exterior and border regions saliency. . . . .	129
4.2	Block diagram of proposed Improved Probabilistic Model. . . . .	132
4.3	Visual demonstration of Contribution of each step in GCS . . . . .	143
4.4	Visual comparison of saliency of proposed method with other state- of-art methods. . . . .	144
4.5	Quantitative comparison of proposed method GCS with PR-Curve and ROC-Curve. . . . .	146
4.6	Quantitative comparison of proposed method with F-Measure . . . . .	147
4.7	Comparison of GCS with other proposed own probabilistic contrast based method using PR-Curve . . . . .	148
5.1	The $2 \times 3$ encoder and decoder streams utilize, deeply guided proposed attention map $CSA - Net$ to find the exact salient object. . . . .	153
5.2	The proposed two-stage <b>Cross -complementary Self-Attention- CSA</b> model is based on the Non-Local Network. . . . .	161
5.3	The proposed architecture of $2 \times 3$ encoder and decoder, $CSA - Net$ models. . . . .	165
5.4	The block diagram of the Intra-complementary features Fusion model(IFF).	167

5.5	The Visual comparison of the proposed model with other State-of-the-art-methods. . . . .	172
5.6	The successive validation of three-stream networks in Complex Images with inferior and low depth images. . . . .	178
5.7	The visual demonstration and validation of attention map learned through proposed deeply guided, two-stage additive, Cross-complementary Self-Attention(CSA-Net). . . . .	180
5.8	The visual demonstration of failure cases. . . . .	181
6.1	The importance of mutual attention mechanism to distinguish the salient object in complex and clutter background. . . . .	187
6.2	The illustration of the proposed framework <i>SL-Net</i> . . . . .	193
6.3	The illustration of the proposed Mutual Attention Based Distinguished Window-MADW. . . . .	196
6.4	The design and process of Self-Learning based Dense Decoder- SDD .	197
6.5	Visual Demonstration of proposed method SL-Net with other closely and recent State-of-art-methods. . . . .	205
6.6	Visual illustration of the global localized features. . . . .	210
6.7	Visual illustration of failure case in incomplete depth maps and complex and cluttered background. . . . .	212





# List of Tables

2.1	Comprehensive details of exemplary models for salient object detection using statistics and probabilistic approaches. . . . .	27
2.2	Comprehensive details of Top Performing conventional 3D model of salient object detection. . . . .	41
2.3	Comprehensive Survey of some Deep learning based RGBD Salient object detection model . . . . .	50
2.4	RGB Dataset used in various saliency computations . . . . .	62
2.5	RGBD Dataset for Salient Object Detection . . . . .	66
3.1	Steps wise MAE in proposed Algorithm . . . . .	100
3.2	Steps wise Mean Absolute Error-MAE in proposed Method-GTS . . .	116
3.3	Mean Absolute Error-MAE of different state-of-the-art methods . . .	121
3.4	Mean Absolute Error-MAE of different state-of-the-art methods . . .	121
3.5	Comparison of average computational run time. . . . .	123
3.6	Comparison of average computational run time('*' = GPU time). . . .	123
4.1	Steps wise Mean Absolute Error-MAE in proposed Method-GTS . . .	143
4.2	Mean Absolute Error-MAE of different state-of-the-art methods . . .	147
5.1	The quantitative comparison of proposed framework on seven benchmark RGBD datasets with four recent evaluation parameters. . . . .	175
5.2	The validation of the effectiveness of Network Architecture and Streams-wise saliency using Mean Absolute Error-MAE. . . . .	179
5.3	The ablation study of each component in the CSA-Net module. . . . .	181
6.1	The quantitative comparison of proposed framework on seven benchmark RGBD datasets with four recent evaluation parameters. . . . .	207
6.2	The ablation study of each component in the <i>SL - Net</i> . . . . .	208
6.3	Validation of stage-wise improvements using proposed Mutual Attention based Distinguished Window-MADW . . . . .	211
6.4	The ablation analysis of Cross-complementary fusion using <i>SDD</i> module and basic fusion strategy. . . . .	212
6.5	The comparison of own proposed deep learning based models. . . . .	213



# Abbreviations

<b>GR</b>	<b>G</b> raphical <b>R</b> anking
<b>ASNet</b>	<b>A</b> ttentive <b>S</b> aliency <b>N</b> etwork
<b>AUC</b>	<b>A</b> rea <b>U</b> nder <b>C</b> urve
<b>MCDL</b>	<b>M</b> ulti <b>c</b> ontext <b>D</b> eep <i>L</i> earning
<b>CNN</b>	<b>C</b> onvolutional <b>N</b> eural <b>N</b> etwork
<b>ESL</b>	<b>E</b> nhance and <b>S</b> uppress <b>L</b> uminance
<b>ECSSD</b>	<b>E</b> xtended <b>C</b> omplex <b>S</b> cene <b>S</b> aliency <b>D</b> ataset
<b>FCN</b>	<b>F</b> ully <b>C</b> onvolutional <b>N</b> eural <b>N</b> etworks
<b>FN</b>	<b>F</b> alse <b>N</b> egative
<b>FP</b>	<b>F</b> ixation <b>P</b> rediction
<b>GAN</b>	<b>G</b> enerative <b>A</b> dversarial <b>N</b> etwork
<b>HED</b>	<b>H</b> olistically-Nested <b>E</b> dge <b>I</b> oU
<b>Intersection over Union</b>	
<b>MAE</b>	<b>M</b> ean <b>A</b> bsolute <b>E</b> rror
<b>MDB</b>	<b>M</b> inimum <b>D</b> irectional <b>B</b> ackgroundness
<b>MDC</b>	<b>M</b> inimum <b>D</b> irectional <b>C</b> ontrast
<b>MSRA</b>	<b>M</b> icrosoft <b>R</b> esearch <b>A</b> sia
<b>MSRC</b>	<b>M</b> icrosoft <b>R</b> esearch <b>C</b> ambridge
<b>PR</b>	<b>P</b> recision <b>R</b> ecall
<b>PC</b>	<b>P</b> oisson <b>P</b> robabilistic <b>C</b> ontrast <b>R</b> BD
<b>Robust Background Detection</b>	
<b>R-CNN</b>	<b>R</b> egion- <b>C</b> onvolutional <b>N</b> eural <b>N</b> etwork

textbfRGB	<b>Red Green Blue</b>
<b>RGBD</b>	<b>Red Green Blue Depth</b>
<b>ROC</b>	<b>Receiver Operating Characteristics</b>
<b>RPN</b>	<b>Region Proposal Network</b>
<b>SOD</b>	<b>Salient Object Detection</b>
<b>VGG</b>	<b>Visual Geometry Group</b>
<b>GU</b>	<b>Global Uniqueness</b>
<b>HC</b>	<b>Histogram-based global Contrast</b>
<b>RC</b>	<b>Regional Contrast</b>
<b>DoG</b>	<b>Difference of Gaussian</b>
<b>MSS</b>	<b>Maximum Symmetric Surround</b>
<b>MBD</b>	<b>Minimum Barrier Distance</b>
<b>AMC</b>	<b>Absorbing Markov Chain</b>
<b>MST</b>	<b>Minimum Spanning Tree</b>
<b>FPR</b>	<b>False Positive Rate</b>
<b>TPR</b>	<b>TruePositive Rate</b>
<b>ILG</b>	<b>Iterative Laplacian of Gaussian</b>
<b>FIT</b>	<b>Feature Integration Theory</b>
<b>DHS</b>	<b>Deep Hierarchical Saliency</b>
<b>GTS</b>	<b>Global Topographical Surface</b>
<b>MADW</b>	<b>Mutual Attention based Distinguished Window</b>
<b>SDD</b>	<b>Self Learning-Based Dense Decoder</b>
<b>JL-DCF</b>	<b>Joint Learning and Densely Cooperative Fusion</b>
<b>S2MA</b>	<b>Self-Selective Mutual Attention</b>
<b>CFF</b>	<b>Coss-Complementary Fusion</b>
<b>CSA-Net</b>	<b>Deep Cross-Complementary Self Attention</b>
<b>MSF</b>	<b>Modality-Specific Saliency Fusion Model</b>
<b>CFF</b>	<b>Cross-Complementary Features Fusion</b>

*Abbreviations*

---

<b>IFF</b>	<b>I</b> ntra- <b>C</b> omplementary <b>F</b> eatures <b>F</b> usion
<b>OSS</b>	<b>O</b> ptimal <b>S</b> elective <b>S</b> aliency <b>F</b> usion <b>M</b> odel



# Symbols

$PDF$	Poisson probability distribution
$T_f$	Enhance and Suppress Luminance-(ESL) coefficient
$dd$	Luminance contrast normalization coefficient
$d$	Angular threshold
$\lambda$	Represents the complex image indicator parameter
$AD_j$	Regional areal density
$Dis_0(r_k, r_j)$	Spatial distance weight
$\sigma$	Controlling parameter
$IF_w$	Gaussian weighted background suppression
$\mu$	Mean
$Areal_k(r)$	Spatial regional saliency
$S_{RC}(r_k)$	Reginol distance based saliency
$S_{RCC}(r_k)$	Reginol color saliency
$IFactor_w(r_k)$	Integrating factor
$S_{Center}$	Central saliency
$\psi(\dots)$	Up-sampling function
$\varphi(\dots)$	Convolution operation
$Sig(\dots)$	Sigmoid function
$Con(i, j)$	Convolution operation
$Max - pool(i, j)$	Max-pooling operation
$E_\psi$	E-measure



## *Symbols*

---

$M_{att}$	Mutual attention feature maps
$f_g$	Global Localized feature
$f_h^i$	High level semantic features
$f_c^i$	Cross-complementary features

## PREFACE

---

The phenomenal use of digital media in our day-to-day life and society brings us to the corner of another world. This transformation appeals and inspires me to innovate and add something to this revolution. Images play an essential role in online shopping and purchasing to online education. It empowers us in various ways, like photo and information sharing, medical imaging, simulation, and military handling of the pandemic. Together with the empowerment, it is necessary to organize these images and use them for various computer vision purposes. One of the most critical computer vision tasks is salient object detection. Salient object detection is to identify the most prominent and relevant object in images. Humans can do this task automatically. The computer can learn to do this by understanding the human visual attention mechanism. It has diverse applications in image captioning, image segmentation, object recognition, content-based image retrieval, image and video compression, and video summarization. This thesis aims to develop computer algorithms, methods, and models for salient object detection, which can detect salient objects similar to human beings.

The main focus of the thesis is to generate probabilistic models for salient object detection to simulate structure, shape, size, and background uncertainty. While designing these models, an attempt is made to discover the current research gaps in existing algorithms, methods, and models. For this, a comprehensive literature

survey is performed. Research gaps of these fields are identified. After that, statistical, probabilistic, and deep learning-based models are developed to address the identified research gaps.

The five models have been developed for salient object detection in 2D and 3D modalities. Among these models, three are based on probabilistic approaches and rest are based on deep learning based approaches. The first two models propose probabilistic contrast and edge enhanced Global Topographical Surface. This surface encloses the prominent object with all its structural and spatial information. Then it is used as a reference plane for regional depth, color, and spatial saliency integration. Most saliency methods generate prominent objects from 2D information, while human attention systems are 3D perception mechanisms. Inspired by this perception, the following models are proposed based on 3D saliency.

The third proposed models is utilized additional depth information from RGBD to robustly and correctly detect the salient object in a complex and cluttered background. To distinguish the salient object in complex and cluttered background, the Poisson probabilistic contrast space is proposed. This process produces a global concave reference surface. The various regional saliencies are integrated into this global concave reference surface to detect the salient object correctly. Background estimation and central saliency integration will thoroughly remove the background. This algorithm generates a robust conspicuous object.

The thesis is presented in a way so that the conventional, as well as newer models like

deep learning, are covered. The exiting models incur inconsistency or distribution loss of salient points and regions. These drawbacks are targets to design two deep network models. The fourth model is CSA-Net, which produces essential features: non-complementary, cross-complementary, intra-complementary, and deep localized improved high-level features. The designed  $2 \times 3$  encoder and decoder streams produce these essential features and assure modality-specific saliency preservation. The cross and intra- complementary fusion are deeply guided by proposed novel, cross-complementary self-attention to produce fused saliency. The attention map is computed by two-stage additive fusion based on a Non-Local network.

Finally, in fifth model, a composite backbone network-based deep CNN framework is designed to produce more acute saliency in a challenging and complex scenario. This composite backbone produces enhanced features that are fused with a self-learning-based dense decoder. A comparative analysis of the proposed models are summarized and discussed. The models are compared against state-of-the-art algorithms and evaluated on benchmark databases. Possible future works and research directions are also included in the last chapter.