

Low Light Image Enhancement and Saliency Object Detection

by Yuanfang Zhang

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of Professor Xiangjian He

University of Technology Sydney
Faculty of Engineering and Information Technology

May 2022

Certificate of Authorship/Originality

I, Yuanfang Zhang, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology, at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Northwestern Polytechnical University.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed
prior to publication.

Signature _____

Date 5th May, 2022 _____

ABSTRACT

Low Light Image Enhancement and Saliency Object Detection

by

Yuanfang Zhang

Low light images represent a series of image types with great potential. Their research focuses on images and videos of the environment at dusk and near darkness. It can be widely used in night safety monitoring, license plate recognition, night scene shot, special target recognition at dusk, and other emergency events that occur under light scenes. After the environment is enhanced and combined with other tasks in computer vision and pattern recognition, it can bring many results, such as saliency detection and object detection under low illumination, and abnormal detection in crowded places under low-light environment. For the enhancement of low light and low light scenes, using traditional methods often results in over-exposure and halo conditions. Therefore, using deep learning network technology can fix and improve these specific shortcomings. To achieve this goal, we have done several investigations about the current state-of-art researches on low-light enhancement and the relevant computer vision tasks. For low light image enhancement, a series of qualitative and quantitative experimental comparisons conducted on a benchmark dataset demonstrate the superiority of our approach, which overcomes the drawbacks of white and colour distortion. At present, most of the research works on visual saliency have concentrated on the field of visible light, and there are few studies on night scenes. Due to insufficient lighting conditions in night scenes, and relatively lower contrasts and signal-to-noise ratios, the effectiveness of available visual features is greatly reduced. Moreover, without sufficient depth information, many features and clues are lost in the original images. Therefore, the detection of salient targets in night scenes is also difficult and it is a focus of current research in the field of computer vision. The performance leads to vague effects when the existing methods are directly con-

ducted, so we adopt a new “enhance firstly detection secondly” mechanism that firstly enhances the low-light images in order to improve the contrast and visibility, and then combines it with relevant saliency detection methods with depth information. Furthermore, we concern about the feature aggregation schemes for deep RGB-D saliency object detection and propose novel feature aggregation methods. Meanwhile, for the monocular vision, of which the depth information is hard to acquire, a novel RGB-D image saliency detection method is proposed to leverage depth cues for enhancing the saliency detection performance but without actually using depth data. Both of the extra depth cues and the proposed “enhance firstly detection secondly” mechanism can improve saliency detection abilities, according to the experimental results. The model not only outperforms the state-of-the-art RGB saliency models, but also achieves comparable or even better results compared with the state-of-the-art RGB-D saliency models

Dissertation directed by Professor Xiangjian He, Professor Michael Blumenstein and Doctor Wenjing Jia

Faculty of Engineering and Information Technology

Dedication

To my parents, and those who always love me and support me along the way.

Acknowledgements

My doctor study at UTS in the past three years has been a life-changing and priceless experience for me. Sydney is a lovely place. It has a golden light harbour with white sails, delicate and charming beaches, and a mild Mediterranean climate. The streets are filled with wild scents, lush forests, and soaring seagulls. Its natural beauty is enhanced by golden beaches and unspoiled bush lands. Sydney is a fantastic location for scientists to investigate science mysteries.

First and foremost, I would like to express my heartfelt appreciation to Professor Xiangjian He, my Principal Supervisor, who helped me tremendously by supplying me with required tools, valuable guidance, and inspiration for new ideas with exceptional patience and constant encouragement. His recommendations have drawn my attention to a variety of flaws and clarified many questions for me.

I am also grateful to Professor Michael Blumenstein and Doctor Wenjing Jia, for being my co-supervisors, especially to Dr Jia for her thoughtfulness and generosity in arranging study meetings in Professor He's Computer Vision and Pattern Recognition (CVPR) group. In research, Dr Jia is very proactive and detail-oriented, and her passion for collaboration and ability to contribute has bonded all of the research students into a family. With Dr Jia and the CVPR study community, I am really enjoying studying sophisticated deep learning techniques in image analysis and pattern recognition.

Then, I would like to express my gratitude to my classmates for their invaluable help with the original manuscript. They kindly made important remarks and sound recommendations to the paper's outline.

I would like to thank the teachers, writers, and colleagues at UTS for their time and commitment. Special thanks to Xiaochen Fan, Yue Xi, Xudong Song,

Qingqing Wang, Edward Huang, Saeed Amirgholipour, Muhammad Usman, Hesam Hesamian, Haodong Chang, Farhan Mohammed, and Chengpei Xu, who provided me with a great deal of assistance at different times.

Finally, I would like to express my gratitude to my parents for their encouragement in allowing me to study abroad. It is very beneficial to me in completing this study and I always love them.

In one word, I would like to express my thanks to all who have assisted me in finishing this study.

List of Publications

Journal Papers

- J-1. Zhang Y, Zheng J, Li L, Nian L, Wenjing Jia, Xiaochen Fan, Chengpei Xu, Xiangjian He. Rethinking feature aggregation for deep RGB-D salient object detection [J]. Neurocomputing, 2021, 423: 463-473.
- J-2. Zhang Y F , Zheng J , Jia W , et al. Deep RGB-D Saliency Detection without Depth[J]. IEEE Transactions on Multimedia, 2021, PP(99):1-1. Doi: 0.1109/TMM.2021.3058788
- J-3. Zhang Y, Zheng J, Fei Li, Wenjing Jia, Wenfeng Huang, Xiangjian He. Low Light Image Dedarking via Deep Semantic Fusion[J]. IEEE Signal Processing Letter (Under Review)

Contents

Certificate	ii
Abstract	iii
Dedication	v
Acknowledgments	vi
List of Publications	viii
List of Figures	xii
List of Tables	xvii
Abbreviation	xix
1 Introduction	1
1.1 Related Works	3
1.1.1 Introduction to Low Light Image Enhancement	4
1.1.2 Introduction to Saliency Object Detection	5
1.1.3 Introduction to Low light Saliency Object Detection	8
1.2 Research Objectives	9
1.3 Main Contributions	11
1.4 Thesis Organization	12
1.4.1 Chapter 2	12
1.4.2 Chapter 3	13
1.4.3 Chapter 4	13

2	Low Light Image Enhancement via CNN-based Models	15
2.1	Introduction	15
2.2	Cognitive Perception Retinex Theory	18
2.3	Loss Functions	22
2.4	Experiments	24
2.4.1	Dataset	25
2.4.2	Implementation Details	25
2.4.3	Results and Analysis	26
2.4.4	Object Theme Enhancement Analysis	28
2.4.5	Comparison on Other Datasets	31
2.4.6	Ablation study	32
2.5	Summary	33
3	Saliency Detection from Low Light RGB-D Images	35
3.1	Introduction	36
3.2	Related Work	41
3.3	Proposed Method	43
3.3.1	Encoder Network	44
3.3.2	Decoder Networks	46
3.3.3	Factorized Gated Attention	48
3.4	Experiments	51
3.4.1	Datasets	51
3.4.2	Implementation Details	52
3.4.3	Evaluation Metrics	53
3.4.4	Component Analysis	54

3.4.5	Comparison with State-of-the-art Models	60
3.4.6	Failure Analysis	64
3.5	Summary	65
4	Low-Light Saliency Detection via Deep CNN without Depth	66
4.1	Introduction	66
4.2	Related Work	69
4.3	RGB-D Saliency Detection	71
4.3.1	Encoder Network	73
4.3.2	Decoder Networks	73
4.3.3	Loss Functions	79
4.4	Experiments	81
4.4.1	Datasets and Evaluation Metrics	81
4.4.2	Implementation Details	82
4.4.3	Comparison with State-of-the-art Models	83
4.4.4	Ablation Study	90
4.4.5	Discussion	93
4.5	Summary	95
5	Conclusion and Future Work	97
5.1	Conclusion	97
5.2	Future Work	99
	Bibliography	101

List of Figures

2.1	The Retinex model	18
2.2	The overall view of our proposed network architecture.	20
2.3	The Inception Module in our proposed network	21
2.4	Example indoor and outdoor images of the ExDARK dataset [69] . .	24
2.5	Examples of image enhancement results obtained on the synthetic dataset with benchmark approaches and our proposed approach. (a) Ground truth. (b) low light images. (c) MSR [40] results. (d) LIME [28] results. (e) LECARM [126] results. (f) Our results.	27
2.6	Examples of image enhancement results obtained on natural low light dataset. (a) low light images. (b) MSR [40] results. (c) LIME [28] results. (d) LECARM [126] results. (e) Our results	28
2.7	Comparison of the NIQE results of the enhanced images obtained with benchmark and our approaches.	29
2.8	Image enhancement results of the two example images from the ExDARK Dataset[69] obtained with the comparison methods. Note the details shown in the red bounding boxes. The first column represents the source images without enhancement. The second to fifth columns represent the image enhancement results obtained with MSR [40], LIME [28], LECARM [126] and our approach, respectively.	30

2.9	Comparison of the dedarking results of the 12-object themes from the ExDARK Dataset [69] obtained with the four comparison benchmark methods. (a) Original images. (b) MSR [40] results. (c) LIME [28] results. (d) LECARM [126] results. (e) Our results.	31
3.1	Comparison of the performance of different SOD methods on low-light images.	36
3.2	Comparison of different network architectures. (a) Two-stream FCN [70]. (b) Two-stream UNet [85]. (c) Our proposed network. We cascade both top-down and bottom-up feature aggregation for deep RGB-D SOD to further leverage improved low-level features for promoting high-level features. We also propose to holistically aggregate features across all levels to learn plentiful multi-level feature interactions. Early aggregation paths are also presented to aggregate and propagate cross-modal encoder features.	38
3.3	Network architecture of the proposed RGB-D SOD model. We first use two encoder branches for the RGB and depth inputs to extract multi-level encoder features (\mathbf{F}_*^R and \mathbf{F}_*^D). Within the two-stream encoders, we adopt early aggregation paths (\mathbf{F}_*^{EA}) to propagate cross-model information from the very beginning. Here, the early aggregation path for the two Conv5_3 layers is not shown. Then, we successively adopt a top-down decoder network (\mathbf{D}_*^\downarrow) and a bottom-up one (\mathbf{D}_*^\uparrow) to aggregate multi-level features. We also use holistic aggregation paths to directly aggregate features across all levels. The size of each feature map is also given and denoted by <i>channel</i> \times <i>height</i> \times <i>width</i> . \odot denotes concatenation and \oplus means element-wise summation.	44

3.4	Architecture of the proposed factorized gated attention module. We factorize the gated attention of the feature map \mathbf{X} as the multiplication of multi-factored channel-wise gate weights \mathbf{G}^c and spatial gate weights \mathbf{G}^s to reduce computation and memory costs and introduce attention ensemble. AAP: adaptive average pooling. \odot : element-wise multiplication. \otimes : matrix multiplication. Sizes of some crucial features are marked by gray font.	50
3.5	Visual comparison of different model settings. We compare the results of the baseline Two-stream UNet (d), adding the holistic aggregation paths and the bottom-up aggregation (e), and further adding the factorized gated attention (f).	55
3.6	Visualization of two learned two spatial attention factors for \mathbf{D}_2^\uparrow . “Att 1” and “Att 2” denote the two spatial attention maps, respectively.	56
3.7	Visualization of the saliency maps of our SOD model and other state-of-the-art RGB-D SOD models.	60
3.8	Comparison of different SOD methods conducted on enhanced low light images from ExDARK Dataset [69] with the CNN-based enhanced method.	61
3.9	Visualization of common failure patterns.	64
4.1	Comparison of the saliency detection results without (“w/o_Depth”) and with (“w_Depth”) using depth cues. (a) and (b) show two example images and their ground truth (GT) saliency maps. (c) shows the saliency detection results of a baseline deep saliency model without using depth cues. (d) shows our predicted depth maps. (e) shows our predicted saliency maps with using depth cues.	67

4.2	Network architecture of the proposed model. The whole model has an encoder network (green cubes) and two decoder networks (white and gray cubes). The encoder network is used to extract multi-level encoder features, while the two decoder networks are used for predicting the depth map and the saliency map, respectively. We use the VGG-16 network [88] as our encoder, and its multi-level features are marked on the cubes. Each decoder progressively fuses the multi-level features by using skip-connections. The depth features are also fused with the RGB features via fusion connections for enhancing the saliency detection performance. The channel numbers of the decoding modules are also marked under the cubes.	74
4.3	Network architecture of the DASPP model and the proposed DMSF model.	75
4.4	Network architecture of the decoding module and the fusion decoding module. “BR” means BN [36] and ReLU, “CBR” means Conv, BN and ReLU. “UP” means upsampling.	78
4.5	Comparison with state-of-the-art RGB saliency models in terms of PR curves on four datasets. The compared models are Amulet [132], DSS [33], BMP [129], PiCANet [62], R3Net [16], CPD [113], EGNNet [134], MINet [78], and ITSD [136].	84
4.6	Visual comparison between our model and state-of-the-art RGB and RGBD saliency models. Our model outperforms SOTA RGB saliency models and surprisingly achieve comparable or even better results than SOTA RGB-D saliency models.	87
4.7	Visualization comparison of different SOD methods conducted on enhanced low light images from ExDARK Dataset [69] with using CNN-based enhanced method.	89
4.8	Visual comparison between “RGB U-Net” and the “+Depth” setting. The GT depth maps and our predicted ones are also given.	90

4.9	Visual comparison between the “+DMSF_w/o_NL” and the “+Depth” setting.	91
4.10	Visual comparison between the “+NL” and the “+DMSF_w/o_NL” setting.	92
4.11	Failure case analysis.	95

List of Tables

2.1	The NIQE results obtained on the ExDARK dataset [69] using and without using our proposed Perceptual Loss and SSIM Loss.	33
3.1	Ablation study on the effectiveness of the holistic aggregation paths (HA), the bottom-up aggregation (BU), the factorized gated attention (FGA), and the early aggregation (EA). Bold indicates the best performance.	54
3.2	Comparison between FGA and existing attention models, including convolutional gated attention (CGA), spatial attention (SA), and the Convolutional Block Attention Module (CBAM). We report both RGB-D SOD performance and computational costs, which include both memory costs and running times during testing. Here we only test the network forwarding time and ignore the time for reading and writing images for rigorous comparisons. Bold indicates the best performance.	58

3.3 Quantitative comparison of our proposed model with state-of-the-art RGB-D SOD methods. We report comparison results under two settings, i.e., training with 2 datasets (NJUD and NLPR) and training with 3 datasets (NJUD, NLPR, and DUT-RGBD). Underline and **Bold** indicate the best and the second best performance under each setting, respectively. **Underline** means the best performance under both settings. Note that, for fair comparisons, we show the results of the JL-DCF [26] model with the VGG backbone, whose results are only reported on 6 datasets in their paper. 62

4.1 Quantitative comparison between our proposed model and state-of-the-art RGB and RGB-D salient object detection models. We compare our model with nine state-of-the-art (SOTA) CNN-based RGB saliency models and twelve SOTA deep learning based RGB-D saliency models on seven datasets in terms of four evaluation metrics. “Train w D” means training with depth while “Test w D” means test with depth. The number in **bold** indicates the best performance in each group (i.e., RGB and RGB-D). The number in *italic* indicates the cases where our model outperforms RGB SOTA models, while * indicates the cases where our model outperforms the A2dele model. “-” means the results are unavailable since the authors did not release them 85

4.2 Quantitative comparison among our proposed model, baseline RGB U-Net, and state-of-the-art RGB salient object detection models on six RGB saliency datasets. The number in **bold** indicates the best performance in each group. 93

Abbreviation

ASPP - Atrous Spatial Pyramid Pooling
DASPP - Dense Atrous Spatial Pyramid Pooling
SOD - Saliency Object Detection
CGA - Convolutional Gated Attention
SA - Spatial Attention
CBAM- Convolutional Block Attention Module
DMSF- Dense MultiScale Fusion
NIQE- Natural image quality evaluator
PSNR- Peak Signal to Noise Ratio
SSIM- Structural Similarity
CNN - Convolutional Neural Networks
ExDARK - A Dataset for low light image enhancement
MSR - Multi-scale Retinex
LIME - A Dataset for low light image enhancement
LEARM - A Dataset for low light image enhancement
AAP - Adaptive Average Pooling
BN - Batch Normalization
ReLU - Rectified Linear Unit
UP - Upsampling
PCC - Pearson Correlation Coefficient
NL - Non Local
GT - Ground Truth
HA - Holistic Aggregation
EA - Early Aggregation
BU - Bottom-up

FGA - Factorized Gated Attention
SSF - A model for Saliency Object Detection
UCNet - A model for Saliency Object Detection
JLDCF - A model for Saliency Object Detection
NJUD - A dataset for Saliency Object Detection
NLPR - A dataset for Saliency Object Detection
SSD - A dataset for Saliency Object Detection
RGBD135 - A dataset for Saliency Object Detection
STEREO - A dataset for Saliency Object Detection
DUT-RGBD - A dataset for Saliency Object Detection
A2dele - A model for Saliency Object Detection
SOTA - State-of-the-art
Amulet - A model for Saliency Object Detection
DSS - A model for Saliency Object Detection
BMP - A model for Saliency Object Detection
PiCANet - A model for Saliency Object Detection
R3Net - A model for Saliency Object Detection
CPD - A model for Saliency Object Detection
EGNet - A model for Saliency Object Detection
PoolNet - A model for Saliency Object Detection
BASNet - A model for Saliency Object Detection
MINet - A model for Saliency Object Detection
ITSD - A model for Saliency Object Detection
DF - A model for Saliency Object Detection
AFNet - A model for Saliency Object Detection
CTMF - A model for Saliency Object Detection
MMCI - A model for Saliency Object Detection
PCF - A model for Saliency Object Detection
TANet - A model for Saliency Object Detection
CPFP - A model for Saliency Object Detection
DMRA - A model for Saliency Object Detection

S^2 MA - A model for Saliency Object Detection

RGB-D - RGB and Depth

LFSD - A dataset for Saliency Object Detection

maxF - A performance index for Saliency Object Detection

STERE - A dataset for Saliency Object Detection

SIP - A dataset for Saliency Object Detection

SSD - A dataset for Saliency Object Detection

MAE - A performance index for Saliency Object Detection

PR Curve - A performance index for Saliency Object Detection