



Universiteit
Leiden
The Netherlands

Exploring images with deep learning for classification, retrieval and synthesis

Liu, Y.

Citation

Liu, Y. (2018, October 24). *Exploring images with deep learning for classification, retrieval and synthesis*. *ASCI dissertation series*. Retrieved from <https://hdl.handle.net/1887/66480>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/66480>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/66480> holds various files of this Leiden University dissertation.

Author: Liu, Y.

Title: Exploring images with deep learning for classification, retrieval and synthesis

Issue Date: 2018-10-24

Exploring Images with Deep Learning for Classification, Retrieval and Synthesis

Yu Liu

Copyright © 2018 Yu Liu, All Rights Reserved

ISBN 978-94-6375-139-1

Printed by Ridderprint BV, The Netherlands

An electronic version of this dissertation is available at
Link <https://openaccess.leidenuniv.nl/handle/1887/9744>

Cover design: Wei Liu, Yu Liu

Exploring Images with Deep Learning for Classification, Retrieval and Synthesis

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 24 oktober 2018
klokke 11.15 uur

door

Yu Liu

geboren te Heilongjiang, China
in 1988

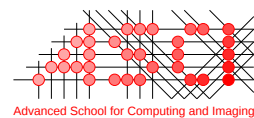
Promotiecommissie

Promotors: Prof. dr. J.N. Kok
Dr. M.S. Lew

Overige leden: Prof. dr. A. Plaat
Prof. dr. T.H.W. Bäck
Prof. dr. W. Kraaij
Prof. dr. H. Trautmann (University of Münster)
Prof. dr. A. Hanjalic (Delft University of Technology)
Prof. dr. ir. B.P.F. Lelieveldt
Dr. ir. R. Poppe (Utrecht University)



Yu Liu was financially supported through the China Scholarship Council (CSC) to participate in the PhD programme of Leiden University. Grant number 201406060010.



This work was carried out in the ASCI graduate school.
ASCI dissertation series number: 387

The research in this thesis was performed at the LIACS Media Lab, Leiden University, The Netherlands, and we would like to thank the NVIDIA Corporation for the donation of GPU cards.

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Background and Related Work	2
1.2.1	Classification	3
1.2.2	Retrieval	6
1.2.3	Synthesis	8
1.3	Thesis Outline and Research Questions	10
1.4	Main Contributions	15
1.4.1	Models and algorithms	15
1.4.2	Practical scenarios	16
1.4.3	Empirical analysis	17
2	Convolutional Fusion Networks for Image Classification	19
2.1	Introduction	20
2.2	Convolutional Fusion Networks	22
2.2.1	Network architecture	22
2.2.2	Training procedure	26
2.2.3	Comparisons with other models	27
2.3	Fully Convolutional Fusion Networks	27
2.3.1	Semantic segmentation	28
2.3.2	Edge detection	29
2.4	Experiments	30
2.4.1	Image classification on CIFAR	30
2.4.2	Image classification on ImageNet	34
2.4.3	Transferring deep fused features	37
2.4.4	Semantic segmentation on PASCAL VOC	39
2.4.5	Edge detection on BSDS500	40
2.5	Chapter Conclusions	42
3	Recognizing Image Edges	43
3.1	Introduction	44
3.2	Relaxed Deep Supervision	46
3.2.1	Network details	46
3.2.2	Loss formulation	49

3.3	Pre-training Procedure	51
3.4	Experiments	53
3.4.1	Implementation details	53
3.4.2	Ablation study on BSDS500	53
3.4.3	Cross-dataset generalization	56
3.4.4	Computational cost	58
3.5	Chapter Conclusions	58
4	DeepIndex for Image Retrieval	59
4.1	Introduction	60
4.2	Bag of Deep Features	61
4.2.1	Spatial patches	61
4.2.2	Feature extraction and quantization	63
4.3	DeepIndex	63
4.3.1	Single DeepIndex	63
4.3.2	Multiple DeepIndex	65
4.3.3	Global image signature	66
4.4	Experiments	67
4.4.1	Datasets and metrics	68
4.4.2	Results and discussion	68
4.4.3	Comparison with other methods	71
4.5	Chapter Conclusions	72
5	Image-Text Matching for Cross-modal Retrieval	73
5.1	Introduction	74
5.2	Recurrent Residual Fusion	75
5.3	Matching Network	79
5.3.1	Feature extractor	79
5.3.2	Feature embedding	80
5.3.3	Bi-rank loss	80
5.4	Experiments	82
5.4.1	Results and discussion	82
5.4.2	Comparison with other approaches	84
5.4.3	Model ensemble	85
5.5	Chapter Conclusions	86
6	Cycle-consistent Embeddings for Cross-modal Retrieval	87
6.1	Introduction	88
6.2	Related Work	90
6.3	Cycle-consistent Embeddings	91
6.3.1	System architecture	92
6.3.2	Formulation	93
6.3.3	Full objective	94
6.3.4	Late-fusion inference	95

6.4	Experiments	98
6.4.1	Experimental setup	98
6.4.2	Comparisons with baseline methods	100
6.4.3	Analysis of late-fusion inference	101
6.4.4	Comparisons with state-of-the-art approaches	103
6.4.5	Effect of feature encoders	105
6.5	Chapter Conclusions	106
7	Joint Matching and Classification	107
7.1	Introduction	108
7.2	Joint Matching and Classification Network	110
7.2.1	Multi-modal input	111
7.2.2	Multi-modal matching	111
7.2.3	Multi-modal classification	113
7.3	Training and Inference	117
7.4	Experiments	119
7.4.1	Experimental setup	119
7.4.2	Results on multi-modal retrieval	121
7.4.3	Results on multi-modal classification	122
7.4.4	Parameter analysis	124
7.4.5	Component analysis	127
7.4.6	Comparison with other approaches	130
7.4.7	Computational cost	132
7.5	Chapter Conclusions	132
8	Applications of Image Synthesis	133
8.1	Image-to-Image Translation	134
8.1.1	Methodology	135
8.1.2	Instantiation network	138
8.1.3	Experiment setup	140
8.1.4	Results on photo $\leftrightarrow$ label	140
8.1.5	Results on photo $\leftrightarrow$ sketch	141
8.2	Fashion Style Transfer	143
8.2.1	Methodology	145
8.2.2	Network architecture	150
8.2.3	Experiment setup	152
8.2.4	Results and discussion	154
8.2.5	Ablation study	156
8.2.6	Limitations and discussion	158
8.3	Chapter Conclusions	158

CONTENTS

9	Conclusions	159
9.1	Main Findings	160
9.2	Limitations and Possible Solutions	162
9.3	Future Research Directions	163
	Bibliography	167
	List of Abbreviations	179
	English Summary	181
	Nederlandse Samenvatting	183
	Curriculum Vitae	185