Deep learning for visual understanding

Guo, Y.; Guo Y.

Deep Learning for Visual Understanding

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 donderdag 5 oktober 2017
klokke 10.00 uur

door

Yanming Guo

geboren te Hebei, China
in 1989

Promotiecommissie

Promotor:	Prof. Dr. J.N. Kok	
Co-promotor:	Dr. M.S. Lew	
Overige leden:	Prof. Dr. A. Plaat	
	Prof. Dr. W. Kraaij	
	Prof. Dr. T.H.W. Bäck	
	Prof. Dr. H. Trautmann	(University of Münster, Germany)
	Prof. Dr. S. Rüger	(Open University, United Kingdom)

Advanced School for Computing and Imaging

This work was carried out in the ASCI graduate school.
ASCI dissertation series number: 378

ISBN: 978-94-6299-729-5

Printed by: Ridderprint, the Netherlands

This research is financially supported by the China Scholarship Council (CSC), Grant No. 201306110026.

Contents

1 Introduction 1

1.1 Background 1

1.2 Research Goals and Contributions 3

1.3 Thesis Overview 3

2 A Comprehensive Review of Deep Learning Methods and Applications 7

2.1 Introduction 8

2.2 Methods and recent developments 9

2.2.1 Convolutional Neural Networks (CNNs) 10

2.2.1.1 Types of layers 11

2.2.1.2 Training Strategy 15

2.2.1.3 CNN architecture 17

2.2.2 Restricted Boltzmann Machines (RBMs) 23

2.2.2.1 Deep Belief Networks (DBNs) 25

2.2.2.2 Deep Boltzmann Machines (DBMs) 26

2.2.2.3 Deep Energy Models (DEMs) 27

2.2.3 Autoencoder 28

2.2.3.1 Sparse Autoencoder 29

2.2.3.2 Denoising Autoencoder 30

2.2.3.3 Contractive Autoencoder 30

2.2.4 Sparse Coding 31

2.2.4.1 Solving the sparse coding equation 31

2.2.4.2 Developments 33

2.2.5 Discussion 36

2.3 Applications and Results 37

2.3.1 Image Classification 38

2.3.2 Object Detection 41

2.3.3 Image Retrieval 45

2.3.4 Semantic Segmentation 47

2.3.5 Human Pose Estimation 49

2.4 Trends and Challenges 53

2.4.1 Theoretical Understanding 53

2.4.2 Human-level Vision 54

2.4.3 Training with limited data 55

2.4.4 Time complexity 56

2.4.5 More Powerful Models 56

2.5 Conclusion 58

3 Convolutional Neural Networks Features: Principal Pyramidal Convolution 59

3.1 Introduction 60

3.2 Principal Pyramidal Convolution 61

3.3 Experiment 62

3.3.1 Datasets 63

3.3.2 Comparisons on different networks 63

3.3.3 Comparisons on different dimensions 65

3.4 Conclusion 67

4 Bag of Surrogate Parts Feature for Visual Recognition 69

4.1 Introduction 70

4.2 Related Work 72

4.3 Bag of Surrogate Parts Feature 74

4.3.1 Bag of Surrogate Parts Feature 74

4.3.2 Interpretation of the Surrogate Part 76

4.4 Enhancement schemes 79

4.4.1 Spatial BoSP 79

4.4.2 Scale Pooling 80

4.4.3 Global-Part Prediction 83
4.5 Experiments 84
4.5.1 Analysis of our method 86
4.5.1.1 Which classifier to use? 86
4.5.1.2 The comparison of BoSP from different layers . . 87
4.5.1.3 Evaluation of the Scale Pooling 89
4.5.1.4 Evaluation of the global-part prediction 91
4.5.2 Comparison with the state-of-the-art 91
4.6 Discussion 93
4.7 Conclusion and Future Work 95

5 CNN-RNN: A Large-scale Hierarchical Image Classification Framework 97
5.1 Introduction 99
5.2 Related Work 102
5.2.1 Usage of CNN-RNN framework 102
5.2.2 Hierarchical models for image classification 103
5.3 Hierarchical Image Classification 104
5.3.1 CNN-based generator 104
5.3.2 CNN-RNN generator 105
5.4 Experiments 108
5.4.1 Hierarchical predictions 108
5.4.1.1 CIFAR-100 108
5.4.1.2 ImageNet 2012 113
5.4.2 From coarse categories to fine categories 115
5.5 Conclusion 119

6 What Convnets Make for Image Captioning? 121
6.1 Introduction 122
6.2 Related Work 124
6.3 Proposed Approach 125
6.3.1 Convnets for Image Captioning 125
6.3.2 Multi-Convnet Aggregation 128
6.3.3 Multi-scale Testing 129

6.4 Experiments 130
6.4.1 Evaluation Configuration 130
6.4.2 Results on Caption Generation 131
6.4.3 Results on Image-sentence Retrieval 134
6.5 Conclusion 134

7 Conclusions **135**
7.1 Conclusions 135
7.2 Research Limitations 137
7.3 Future Work 138

Bibliography 141

English Summary 175

Nederlandse Samenvatting 177

Acknowledgements 179

Curriculum Vitae 181