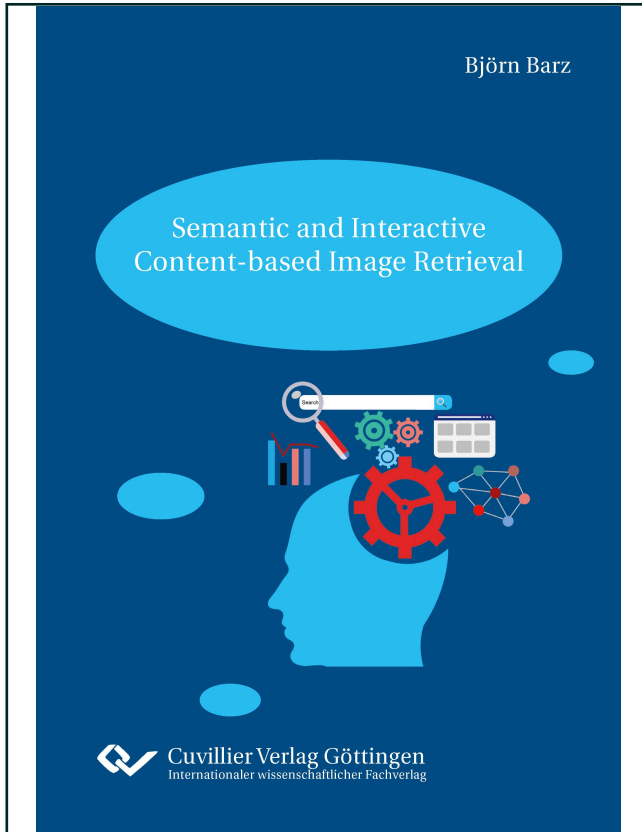




Björn Barz (Autor)

# Semantic and Interactive Content-based Image Retrieval



<https://cuvillier.de/de/shop/publications/8365>

Copyright:

Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen,  
Germany

Telefon: +49 (0)551 54724-0, E-Mail: [info@cuvillier.de](mailto:info@cuvillier.de), Website: <https://cuvillier.de>

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Content-based image retrieval . . . . .	1
1.2	Instance vs. category retrieval . . . . .	5
1.3	Challenges . . . . .	7
1.4	Interactive image retrieval . . . . .	9
1.5	Semantic image retrieval . . . . .	12
1.6	Contributions of this thesis . . . . .	14
<b>2</b>	<b>Methodical Background</b>	<b>17</b>
2.1	Fundamental concepts and definitions . . . . .	17
2.2	Classification . . . . .	20
2.2.1	Problem setting . . . . .	20
2.2.2	Support vector machines . . . . .	21
2.2.3	Linear discriminant analysis . . . . .	25
2.2.4	Nearest neighbor classification . . . . .	26
2.2.5	Gaussian processes . . . . .	27
2.2.6	Neural networks . . . . .	30
2.2.7	Active learning . . . . .	39
2.3	Clustering . . . . .	41
2.3.1	k-means . . . . .	42
2.3.2	Gaussian Mixture Models . . . . .	43
2.4	Metric Learning . . . . .	44
2.4.1	Problem setting and definitions . . . . .	44
2.4.2	Duality between metric and feature learning . . . . .	45
2.4.3	Learning metrics for fixed features . . . . .	46

## CONTENTS

2.4.4	Deep metric learning . . . . .	49
2.5	Information retrieval . . . . .	53
2.5.1	Problem description . . . . .	53
2.5.2	Evaluation metrics . . . . .	54
2.5.3	Learning to rank . . . . .	56
2.5.4	System architecture . . . . .	58
2.5.5	Spatial verification and re-ranking . . . . .	60
2.5.6	Query expansion and diffusion . . . . .	61
2.5.7	Cross- and multi-modal retrieval . . . . .	65
2.6	Image representations for CBIR . . . . .	66
2.6.1	Hand-crafted local features . . . . .	67
2.6.2	Hand-crafted transformations and aggregations . . . . .	69
2.6.3	PCA and whitening . . . . .	71
2.6.4	Off-the-shelf CNN features . . . . .	73
2.6.5	End-to-end learning for image retrieval . . . . .	74
2.7	Relevance feedback . . . . .	78
<b>3</b>	<b>The Cosine Loss</b>	<b>81</b>
3.1	Introduction and motivation . . . . .	81
3.1.1	The problem of small data . . . . .	82
3.1.2	Weakly supervised localization . . . . .	84
3.2	Related work . . . . .	86
3.2.1	Cosine Loss . . . . .	86
3.2.2	Learning from small data . . . . .	87
3.2.3	Weakly supervised localization . . . . .	89
3.3	The cosine loss . . . . .	91
3.3.1	Objective and notation . . . . .	91
3.3.2	Comparison with other loss functions . . . . .	92
3.4	Dense classification and scene understanding . . . . .	96
<b>4</b>	<b>Hierarchy-based Semantic Image Embeddings</b>	<b>101</b>
4.1	In the need of prior knowledge . . . . .	101
4.1.1	Semantic image retrieval . . . . .	102
4.1.2	Explaining classification decisions . . . . .	104
4.2	Related work . . . . .	105
4.3	Knowledge in trees: class taxonomies . . . . .	107
4.3.1	Hierarchy-based semantic similarity . . . . .	108

4.3.2	Tree-shaped taxonomies . . . . .	109
4.4	Hierarchy-based semantic embeddings . . . . .	112
4.4.1	Exact solution . . . . .	112
4.4.2	Low-dimensional approximation . . . . .	114
4.5	Learning semantic image embeddings . . . . .	117
4.6	Subsequent works on semantic embeddings . . . . .	118
<b>5</b>	<b>Experiments: Cosine Loss and Semantic Embeddings</b>	<b>123</b>
5.1	Datasets . . . . .	123
5.1.1	Visual classification datasets . . . . .	123
5.1.2	FGVC datasets . . . . .	125
5.1.3	Extreme Weather dataset . . . . .	127
5.1.4	AG News dataset . . . . .	128
5.1.5	MS COCO . . . . .	129
5.2	Training details . . . . .	129
5.3	Semantic image retrieval . . . . .	132
5.3.1	Performance metrics . . . . .	132
5.3.2	Competitors . . . . .	133
5.3.3	Semantic image retrieval performance . . . . .	134
5.3.4	Low-dimensional approximation . . . . .	138
5.4	Learning from small data . . . . .	139
5.4.1	Classification performance . . . . .	139
5.4.2	Effect of semantic information . . . . .	142
5.4.3	Effect of dataset size . . . . .	143
5.5	Learned feature space . . . . .	145
5.6	Dense classification . . . . .	147
5.6.1	Weakly supervised localization . . . . .	148
5.6.2	Explaining classifier decisions . . . . .	154
<b>6</b>	<b>Interactive Image Retrieval</b>	<b>159</b>
6.1	Introduction . . . . .	159
6.2	Related work . . . . .	160
6.2.1	Multiple instance feedback . . . . .	160
6.2.2	Active learning for image retrieval . . . . .	162
6.2.3	Important image regions . . . . .	164
6.3	Automatic query image disambiguation . . . . .	166
6.3.1	Identification of image senses . . . . .	168

## CONTENTS

6.3.2	Refinement of results . . . . .	170
6.4	Information-theoretic active learning . . . . .	171
6.4.1	Idea and ideal objective . . . . .	174
6.4.2	Relevance model . . . . .	176
6.4.3	User model . . . . .	178
6.4.4	Approximation of mutual information . . . . .	179
6.4.5	Greedy batch construction . . . . .	180
6.5	Adaptive pooling . . . . .	181
6.5.1	Learning the Exemplar-LDA model . . . . .	183
6.5.2	Modifying the representations of database images . . . . .	184
<b>7</b>	<b>Experimental Evaluation of Interactive Retrieval</b>	<b>187</b>
7.1	Datasets and features . . . . .	187
7.1.1	Natural image datasets . . . . .	188
7.1.2	Domain-specific datasets . . . . .	190
7.2	Experiments for AID . . . . .	191
7.2.1	Simulation of the user . . . . .	191
7.2.2	Evaluation metrics . . . . .	192
7.2.3	Competitors . . . . .	192
7.2.4	Quantitative results . . . . .	193
7.2.5	Qualitative examples . . . . .	195
7.3	Experiments for ITAL . . . . .	196
7.3.1	Datasets and simulation of user feedback . . . . .	197
7.3.2	Competitors . . . . .	198
7.3.3	Hyper-parameter tuning . . . . .	200
7.3.4	Quantitative performance evaluation . . . . .	201
7.3.5	Comparison with MCMC[min] and AdaptAL . . . . .	203
7.3.6	Effect of the user model . . . . .	204
7.3.7	Sensitivity regarding feature dimensionality . . . . .	204
7.3.8	Qualitative results . . . . .	206
7.4	Experiments for adaptive pooling . . . . .	210
7.4.1	Quantitative results . . . . .	210
7.4.2	Qualitative examples . . . . .	211

<b>8</b>	<b>Conclusions</b>	<b>217</b>
8.1	Summary and thesis contributions . . . . .	217
8.2	Future work . . . . .	223
8.2.1	Semantic embeddings . . . . .	224
8.2.2	Learning from small data . . . . .	225
8.2.3	Weakly supervised localization . . . . .	226
8.2.4	Interactive image retrieval . . . . .	227
<b>A</b>	<b>Appendix</b>	<b>231</b>
A.1	Derivation of Eq. (6.5) . . . . .	231
A.2	Derivation of Eq. (6.6) . . . . .	232
A.3	Additional metrics for semantic image retrieval . . . . .	233
A.4	Further semantic retrieval results . . . . .	234
A.5	Setup for instance and category retrieval comparison	235
A.6	ITAL failure cases . . . . .	236
A.7	Taxonomy used for MS COCO . . . . .	238
A.8	Taxonomy used for CIFAR-100 . . . . .	240
	<b>Bibliography</b>	<b>241</b>
	<b>List of Own Publications</b>	<b>277</b>
	<b>List of Figures</b>	<b>281</b>
	<b>List of Tables</b>	<b>285</b>
	<b>List of Definitions, Theorems, and Lemmas</b>	<b>287</b>
	<b>Notations</b>	<b>289</b>
	<b>Acronyms</b>	<b>297</b>
	<b>Index</b>	<b>301</b>