

Contents

List of Figures	xii
List of Tables	xv
List of Symbols	xvii
List of Abbreviations	xix
Preface	xxi
1 Introduction	1
1.1 Definition	2
1.2 Applications	3
1.3 Challenges	4
1.4 Problems	5
1.5 Contributions of the Thesis	6
1.6 Organization of Thesis	7
2 Related work	10
2.1 CNN-based for Action Recognition	10
2.2 Human action recognition based on RNN-LSTMs	16
2.3 Other deep architectures for human action recognition	17
2.4 Conclusion	19
3 Action Recognition in Presence of Representation Bias	20
3.1 Our Contribution	23
3.2 Organization of the Chapter	24
3.3 Literature Survey	24
3.4 Proposed Approach	26
3.4.1 Multi-scale Deformable Backbone (MDB) Network	27

3.4.2	Actor-Object-Scene Attention (AOSA) Network	32
3.4.3	Long-Range Context (LRC) Network	38
3.4.4	Class-aware Temporal Attention Pooling (CTAP) Network	42
3.4.5	Joint Training Model	44
3.5	Experimental Results	45
3.5.1	Datasets and Metrics	46
3.5.2	Implementation Details	49
3.5.3	Ablation Study	51
3.5.4	Comparison with State-of-the-Art	57
3.6	Conclusion of the Chapter	63
3.7	Publication related to the Chapter	64
4	Human Behaviour Traits aware Action Recognition	65
4.1	Our Contribution	67
4.2	Organization of the Chapter	68
4.3	Literature Survey	69
4.4	Proposed Approach	70
4.4.1	Backbone Network	71
4.4.2	Visual Attention Network (VAN)	73
4.4.3	Long-term Attention Network (LAN)	75
4.4.4	Temporal Attention Pooling (TAP)	79
4.4.5	Joint Training Mechanism	80
4.5	Experimental Results	82
4.5.1	Datasets and Metrics	82
4.5.2	Implementation Details	85
4.5.3	Ablation Study	85
4.5.4	Comparison with SOTA on benchmark datasets	89
4.6	Conclusion of the Chapter	89
4.7	Publication related to the Chapter	91
5a	Action Recognition in the Wild: Fall Action	92
5a.1	Our Contribution	95
5a.2	Organization of the Chapter	96
5a.3	Literature Survey	96
5a.4	Proposed Approach	97
5a.4.1	Feature Extraction Module	98
5a.4.2	Uncertain-Action Supervision (US) Module	99

5a.4.3 Uncertain-Certain Action (UC) Classifier	104
5a.4.4 Joint Training Model	105
5a.5 Experimental Results	106
5a.5.1 Dataset details	106
5a.5.2 Implementation Details	108
5a.5.3 Ablation Study and Qualitative Analysis	110
5a.5.4 Comparison with State-of-the-Art	117
5a.6 Conclusion of the Chapter	118
5a.7 Publication related to the Chapter	118
5b Action Recognition in the Wild: Unusual Action	119
5b.1 Challenges in this work	122
5b.2 Our Contribution	123
5b.3 Organization of the Chapter	124
5b.4 Literature Survey	125
5b.5 Proposed Approach	126
5b.5.1 Feature Extraction Module: Light-weight ShuffleNet	127
5b.5.2 Fine-grained Human-Object Relation Module	129
5b.5.3 Federated Contrastive Learning (FCL) setup	134
5b.6 Experimental Results	136
5b.6.1 Publicly available Dataset	137
5b.6.2 Our UnusualAction Dataset	137
5b.6.3 Implementation Details	138
5b.6.4 Ablation Study and Qualitative Analysis	139
5b.6.5 Results on Habitual Actions Dataset	146
5b.7 Conclusion of the Chapter	148
5b.8 Publication related to the Chapter	149
6 Multi-label Complex Action Recognition	150
6.1 Our Contribution	153
6.2 Organization of the Chapter	154
6.3 Literature Survey	154
6.4 Proposed Approach	155
6.4.1 3D ShuffleNet Backbone Network	156
6.4.2 Kronecker-based Permutation Invariant Transformer	157
6.5 Experimental Results	163
6.5.1 Datasets and Metrics	163

6.5.2	Implementation Details	166
6.5.3	Ablation Study	167
6.5.4	Comparison with State-of-the-Art on Benchmark datasets . . .	172
6.6	Conclusion of the Chapter	172
6.7	Publication related to the Chapter	174
	References	177
	List of Publications	197