

CONTENTS

Preface	v
Chapter 1 Introduction	1
1.1 Image Engineering	2
1.1.1 Image Technology and Image Engineering	2
1.1.2 Three Levels of Image Engineering	3
1.1.3 Disciplines and Fields Related to Image Engineering	4
1.1.4 Statistical Classification of Image Engineering Literature	6
1.2 Spatiotemporal Image Understanding	10
1.2.1 Development of Spatiotemporal Technology	12
1.2.2 Five Levels of Spatiotemporal Technology	14
1.3 Neural Networks and Deep Learning	16
1.3.1 Deep Learning in Image Understanding	18
1.3.2 Basic Concepts of Convolutional Neural Networks	19
1.3.3 Core Technologies of Deep Learning	25
1.4 The Book Structure and Chapter Introduction	27
References	30
Chapter 2 Spatiotemporal Points	33
2.1 Spatiotemporal Points of Interest	34
2.1.1 Detecting Spatial Points of Interest.....	34
2.1.2 Detecting Spatiotemporal Points of Interest	35
2.2 Key Points of Human Body	37
2.2.1 2-D Human Key Point Detection Based on Image	38
2.2.2 2-D Human Key Point Detection Based on Video	41
2.3 3-D Key Point Detection for Multiple Persons	46
2.3.1 3-D Key Point Detection Based on Multiple Perspectives	46
2.3.2 3-D Key Point Detection Based on Hierarchical Clustering	51
References	56
Chapter 3 Spatiotemporal Trajectory	59
3.1 Trajectory Learning and Analysis	59
3.1.1 Scene Modeling	60
3.1.2 Trajectory Learning	62
3.1.3 Activity Analysis	66
3.2 Trajectory Feature Clustering Tree	68
3.2.1 Flow Chart	69
3.2.2 Trajectory Feature Extraction	69
3.2.3 Feature Clustering Tree	72
3.2.4 Fischer Vector Encoding Tree	73
3.2.5 Experimental Results and Analysis.....	75
References	78

Chapter 4	Action Classification and Recognition	81
4.1	Action Classification	82
4.2	Action Recognition	84
4.2.1	Action Recognition in Various Activities	84
4.2.2	Various Networks for Action Recognition	88
4.3	Action Classification Combining Pose and Context	88
4.3.1	Action Classifier Based on Pose Model	90
4.3.2	Action Classifier Based on Context	91
4.4	Action Detection Based on Attention Mechanism	94
4.4.1	Candidate Action Generation	94
4.4.2	Temporal Action Detection	98
	References	103
Chapter 5	Activity Modeling and Recognition	107
5.1	Action and Activity Modeling	108
5.1.1	Action Modeling	108
5.1.2	Activity Modeling	114
5.2	Actor-Action Joint Modeling	119
5.2.1	Single Label Actor-Action Recognition	120
5.2.2	Multiple Label Actor-Action Recognition	121
5.2.3	Actor-Action Semantic Segmentation	122
5.3	Networks Using Skeleton Representation	125
5.3.1	Skeleton Representation	126
5.3.2	Backbone Network Architecture	129
5.3.3	Network Constructing Methods	133
	References	143
Chapter 6	Detection of Human-Object Interaction Activity	147
6.1	Two Stage and Single-Stage Methods	148
6.1.1	Two Stage Methods	148
6.1.2	Single-Stage Methods	154
6.2	Transformer-Based Approaches	158
6.2.1	Two Stage Methods Using Transformer	158
6.2.2	Single-Stage Methods Using Transformer	159
6.3	More Distinctive Methods	160
6.3.1	Zero-Shot Learning	161
6.3.2	Methods Based on Global and Local Instances	162
6.3.3	Spatiotemporal Human-Object Interaction Detection	165
6.4	Databases and Evaluation Indicators	168
6.4.1	Databases with Different Annotation Granularities	168
6.4.2	Special Databases	170
6.4.3	Evaluation Indicators	172
	References	173
Chapter 7	Behavior Recognition Networks	177
7.1	Networks Combines Motion and Context	178

7.1.1	Network Model Framework	178
7.1.2	Camera Motion Estimation and Compensation	179
7.1.3	Experimental Results and Analysis	180
7.2	Sequential Segment Network Utilizing Structural Information	185
7.2.1	Sequential Segments	185
7.2.2	Experimental Results and Analysis	187
7.3	Graph Convolutional Network for Activity and Behavior	190
7.3.1	Channel-Wise Topology Refinement Modeling	191
7.3.2	Graph Convolutional Network for Activity Recognition	194
7.3.3	Scale Space and Dynamic Time Graph Convolution	195
	References.....	198
Chapter 8	Abnormal Event Detection	201
8.1	Classification of Abnormal Event Detection Methods	202
8.2	Basic Detection Methods Based on Neural Networks	204
8.2.1	Block Learning Based on Convolutional Auto-Encoder	205
8.2.2	Detection Based on One-Class Neural Network	206
8.3	Sparse Denoising Auto-Encoder and Gradient Structure Information	207
8.3.1	SDAE-Based Method	208
8.3.2	SDAE-GS-Based Method	210
8.3.3	Comparison Between SDAE and SDAE-GS	211
8.4	LSTM Auto-Encoder Network Based on Feature Trajectory Smoothing	211
8.4.1	Overall Framework and Process	212
8.4.2	Experimental Results and Analysis	215
8.5	Depth Generative Model for Feature Space Temporal Smoothing	217
8.5.1	Generative Losses	218
8.5.2	Temporal Smoothing Loss	219
8.5.3	Training and Testing	220
8.5.4	Experimental Results and Analysis	221
8.6	A Deep Generative Model Adapted to Scene Content	223
8.6.1	Overall Workflow	224
8.6.2	Specific Characteristics of the Model	225
8.6.3	Experimental Results and Analysis	228
8.7	Master-Auxiliary Aggregation Strategy	232
8.7.1	Basic Framework and Flowchart	232
8.7.2	Master-Auxiliary Aggregation Modules.....	234
8.7.3	Experimental Results and Analysis	236
	References.....	238
Subject Index	241