Table of Contents

Preface **vii**

1 Multimedia Services, 5G Technology, and XR: An In-depth Exploration of Technologies, Standards, and Challenges **1**
Tiziana Cattai, Stefania Colonnese
1.1 Introduction . . . 1
1.2 Conversational services . . . 2
1.3 Streaming services . . . 7
1.4 5G networks provisioning for multimedia services . . . 8
1.5 Extended reality services in 5G networks . . . 12
1.6 Unmanned vehicles for XR . . . 16
References . . . 18

2 Visual Compression: New Paradigms? **21**
Nicola Adami, Marco Dalai, Alessandro Gnutti, Fabrizio Guerrini, Riccardo Leonardi, Pierangelo Migliorati, Alberto Signoroni
2.1 Introduction . . . 21
2.2 Background . . . 23
2.2.1 General concepts . . . 23
2.2.2 Principles of image compression . . . 24
2.2.3 Principles of video compression . . . 27
2.3 History of Image and Video Compression . . . 29
2.3.1 Evolution of Image and Video Compression Standards . . . 29
2.3.2 Distributed Video Coding . . . 30
2.3.3 Wavelet-based (scalable) video coding . . . 33
2.4 Traditional methods: novel investigative approaches . . . 38
2.4.1 Multiple transforms . . . 38
2.4.2 Graph domain . . . 40
2.4.3 SBGFT . . . 41
2.5 Artificial Intelligence (AI) methods: new paradigms . . . 42
2.5.1 General principles . . . 42
2.5.2 Image compression: seminal works . . . 43
2.5.3 Video compression: seminal works . . . 44
2.5.4 Conditional Augmented Normalizing Flows for Video Compression 45
2.6 Conclusions . . . 46
References . . . 46

3 Gaussian Class-Conditional Training for Secure and Robust Deep Neural Networks **55**
Tiziano Bianchi, Andrea Migliorati, Enrico Magli
3.1 Introduction . . . 55
3.1.1 Related works . . . 56
3.2 Gaussian Class-Conditional training . . . 58
3.3 Multiclass classification . . . 59
3.3.1 Target Distributions Parameters . . . 60
3.3.2 Adversarial Robustness Evaluation . . . 62
3.4 Metric learning . . . 68
3.4.1 Network design . . . 68
3.4.2 Face verification . . . 70
3.5 Conclusions . . . 71
References . . . 71

4 Graph Neural Networks for Inverse Problems in Multimedia **79**
Diego Valsesia, Emanuele Aiello
4.1 Introduction . . . 79
4.2 Graph neural networks and Transformers . . . 80
4.3 Point Cloud Generation . . . 82
4.4 Point Cloud Denoising . . . 84
4.5 Point Cloud Completion . . . 86
4.5.1 Local and Global Pretraining of GNNs for Point Cloud Completion 87
4.5.2 Point Cloud and RGB image fusion . . . 90
4.6 Image Denoising . . . 93
4.7 Image Restoration with Transformers . . . 94
4.8 Conclusions . . . 96
References . . . 96

5 Distributed Learning for Adaptive Multimedia Processing **101**
Marco Carpentiero, Vincenzo Matta, Ali H. Sayed
5.1 Introduction . . . 101
5.1.1 Adaptation and Communication Constraints . . . 102
5.1.2 Network Learning Architectures . . . 103
5.2 Related Work . . . 106
5.3 Background . . . 107
5.3.1 Network . . . 107
5.3.2 Optimization Problem . . . 108
5.3.3 Classical ATC Diffusion Strategy . . . 109
5.4 Adaptive Compressed Learning . . . 110
5.4.1 Performance . . . 112
5.5 Distributed Image Classification Example . . . 115
5.6 Conclusion . . . 117
References . . . 117

6 Multimodal Deepfake Detection Using Audio-Visual Cues **121**
Davide Salvi, Sara Mandelli, Honggu Liu, Paolo Bestagini, Stefano Tubaro

6.1 Introduction . . . 121
6.2 Deepfake Detection . . . 123
6.2.1 Visual-only deepfake detection . . . 123
6.2.2 Audio-only deepfake detection . . . 124
6.2.3 Audio-visual deepfake detection . . . 125
6.3 Problem Formulation and Proposed Methodology . . . 125
6.3.1 Problem Formulation . . . 126
6.3.2 Proposed Methodology . . . 126
6.4 Experimental Setup . . . 129
6.4.1 Considered datasets . . . 130
6.4.2 Processing pipeline . . . 132
6.4.3 Training Strategy . . . 132
6.5 Results . . . 133
6.5.1 Evaluation metrics . . . 133
6.5.2 Monomodal results . . . 133
6.5.3 Multimodal results . . . 133
6.6 Conclusions and Future Works . . . 138
References . . . 139

7 Uncertainty-driven detection and localization of image forgeries 145
Fabrizio Guillaro, Davide Cozzolino, Giovanni Poggi, Luisa Verdoliva
7.1 Introduction . . . 145
7.2 Proposed uncertainty-driven approach . . . 148
7.2.1 Noiseprint++ . . . 149
7.2.2 Anomaly localization map . . . 150
7.2.3 Confidence map and integrity score . . . 151
7.3 Datasets and metrics . . . 152
7.3.1 Noiseprint++ training . . . 153
7.3.2 Localization and detection training . . . 153
7.3.3 Testing . . . 154
7.3.4 Performance evaluation . . . 154
7.4 Results . . . 154
7.4.1 Ablation study . . . 154
7.4.2 State-of-the-art comparison . . . 155
7.4.3 Robustness analysis . . . 158
7.4.4 Qualitative comparisons . . . 159
7.5 Conclusions . . . 160
References . . . 161

8 On-the-move Multimodal Biometric Recognition 165
Emanuele Maiorana, Lorenzo Giusti, Patrizio Campisi
8.1 Introduction . . . 165
8.2 State of the art . . . 167
8.3 Design of the Proposed Device . . . 170
8.4 Data Processing . . . 173
8.4.1 Face Processing . . . 174
8.4.2 Vein Pattern Processing . . . 176

8.5 Experimental Tests . . . 179
8.6 Conclusions . . . 180
8.7 Acknowledgements . . . 180
References . . . 181

9 Digital Twins for Data Generation and Crowd Understanding 185
Nicola Conci, Niccolò Bisagno, Nicola Garau, Antonio Luigi Stefani
9.1 Introduction . . . 185
9.2 Related Work . . . 187
9.3 Synthetic Data and Visual Fidelity . . . 189
9.3.1 People Segmentation . . . 190
9.3.2 Crowd counting . . . 192
9.4 Synthetic Data and Behavioral Fidelity . . . 195
9.5 Conclusions . . . 199
9.6 Acknowledgement . . . 201
References . . . 201

10 Affective-based Modelling Approaches for Quality of Experience-based Management Systems 207
Alessandro Floris, Simone Porcu, Matteo Anedda, Daniele Giusto
10.1 Introduction . . . 207
10.2 QoE estimation based on facial expression and gaze direction for video streaming services . . . 209
10.2.1 Framework . . . 209
10.2.2 Data Collection . . . 209
10.2.3 Data Preprocessing . . . 212
10.2.4 QoE Estimation Model and Performance Results . . . 214
10.3 QoE estimation based on facial expressions for WebRTC services . . . 215
10.3.1 Methodology . . . 215
10.3.2 Data collection . . . 216
10.3.3 Data Preprocessing . . . 217
10.3.4 QoE Estimation Model and Performance Results . . . 219
10.4 Conclusions . . . 221
References . . . 221