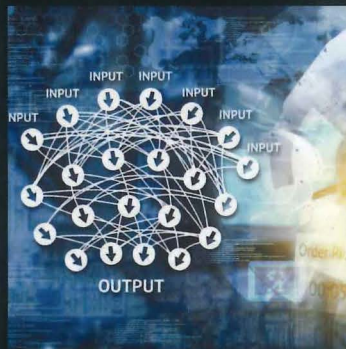
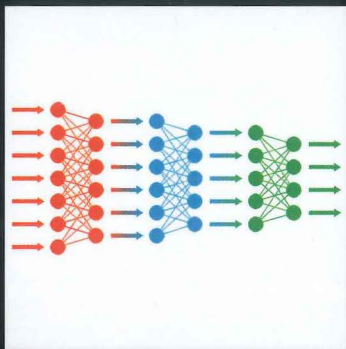


# Advanced Methods and Deep Learning in Computer Vision



Edited by  
E. R. Davies • Matthew Turk



สำนักหอสมุด มหาวิทยาลัยเชียงใหม่

๖๖58441/๙  
๑/๒๕๖๓/๑๒  
๑๒๕๖๖-๕๖๔

๖๖5  
๑/๒๕๖๓/๑๒

Computer Vision and Pattern Recognition

# ADVANCED METHODS AND DEEP LEARNING IN COMPUTER VISION



สำนักหอสมุดมหาวิทยาลัยเชียงใหม่  
CHIANG MAI UNIVERSITY LIBRARY

*Edited by*

E.R. DAVIES

MATTHEW A. TURK



ACADEMIC PRESS

An imprint of Elsevier

# Contents

---

|                      |      |
|----------------------|------|
| List of contributors | xi   |
| About the editors    | xiii |
| Preface              | xv   |

## 1. The dramatically changing face of computer vision

E.R. DAVIES

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| 1.1 Introduction – computer vision and its origins                    | 1  |
| 1.2 Part A – Understanding low-level image processing operators       | 4  |
| 1.3 Part B – 2-D object location and recognition                      | 15 |
| 1.4 Part C – 3-D object location and the importance of invariance     | 29 |
| 1.5 Part D – Tracking moving objects                                  | 55 |
| 1.6 Part E – Texture analysis                                         | 61 |
| 1.7 Part F – From artificial neural networks to deep learning methods | 68 |
| 1.8 Part G – Summary                                                  | 86 |
| References                                                            | 87 |

## 2. Advanced methods for robust object detection

ZHAOWEI CAI AND NUNO VASCONCELOS

|                                       |     |
|---------------------------------------|-----|
| 2.1 Introduction                      | 93  |
| 2.2 Preliminaries                     | 95  |
| 2.3 R-CNN                             | 96  |
| 2.4 SPP-Net                           | 97  |
| 2.5 Fast R-CNN                        | 98  |
| 2.6 Faster R-CNN                      | 101 |
| 2.7 Cascade R-CNN                     | 103 |
| 2.8 Multiscale feature representation | 106 |
| 2.9 YOLO                              | 110 |
| 2.10 SSD                              | 112 |
| 2.11 RetinaNet                        | 113 |
| 2.12 Detection performances           | 115 |

|                 |     |
|-----------------|-----|
| 2.13 Conclusion | 115 |
| References      | 116 |

## 3. Learning with limited supervision

SUJOY PAUL AND AMIT K. ROY-CHOWDHURY

|                                                                  |     |
|------------------------------------------------------------------|-----|
| 3.1 Introduction                                                 | 119 |
| 3.2 Context-aware active learning                                | 120 |
| 3.3 Weakly supervised event localization                         | 129 |
| 3.4 Domain adaptation of semantic segmentation using weak labels | 137 |
| 3.5 Weakly-supervised reinforcement learning for dynamical tasks | 144 |
| 3.6 Conclusions                                                  | 151 |
| References                                                       | 153 |

## 4. Efficient methods for deep learning

HAN CAI, JI LIN, AND SONG HAN

|                                            |     |
|--------------------------------------------|-----|
| 4.1 Model compression                      | 159 |
| 4.2 Efficient neural network architectures | 170 |
| 4.3 Conclusion                             | 185 |
| References                                 | 185 |

## 5. Deep conditional image generation

GANG HUA AND DONGDONG CHEN

|                                                |     |
|------------------------------------------------|-----|
| 5.1 Introduction                               | 191 |
| 5.2 Visual pattern learning: a brief review    | 194 |
| 5.3 Classical generative models                | 195 |
| 5.4 Deep generative models                     | 197 |
| 5.5 Deep conditional image generation          | 200 |
| 5.6 Disentanglement for controllable synthesis | 201 |
| 5.7 Conclusion and discussions                 | 216 |
| References                                     | 216 |



## 6. Deep face recognition using full and partial face images

HASSAN UGAIL

- 6.1 Introduction 221
- 6.2 Components of deep face recognition 227
- 6.3 Face recognition using full face images 231
- 6.4 Deep face recognition using partial face data 233
- 6.5 Specific model training for full and partial faces 237
- 6.6 Discussion and conclusions 239
- References 240

## 7. Unsupervised domain adaptation using shallow and deep representations

YOGESH BALAJI, HIEN NGUYEN, AND RAMA CHELLAPPA

- 7.1 Introduction 243
- 7.2 Unsupervised domain adaptation using manifolds 244
- 7.3 Unsupervised domain adaptation using dictionaries 247
- 7.4 Unsupervised domain adaptation using deep networks 258
- 7.5 Summary 270
- References 270

## 8. Domain adaptation and continual learning in semantic segmentation

UMBERTO MICHIeli, MARCO TOLDO, AND PIETRO ZANUTTIGH

- 8.1 Introduction 275
- 8.2 Unsupervised domain adaptation 277
- 8.3 Continual learning 291
- 8.4 Conclusion 298
- References 299

## 9. Visual tracking

MICHAEL FELSBERG

- 9.1 Introduction 305
- 9.2 Template-based methods 308

- 9.3 Online-learning-based methods 314
- 9.4 Deep learning-based methods 323
- 9.5 The transition from tracking to segmentation 327
- 9.6 Conclusions 331
- References 332

## 10. Long-term deep object tracking

EFSTRATIOS GAVVES AND DEEPAK GUPTA

- 10.1 Introduction 337
- 10.2 Short-term visual object tracking 341
- 10.3 Long-term visual object tracking 345
- 10.4 Discussion 367
- References 368

## 11. Learning for action-based scene understanding

CORNELIA FERMÜLLER AND MICHAEL MAYNORD

- 11.1 Introduction 373
- 11.2 Affordances of objects 375
- 11.3 Functional parsing of manipulation actions 383
- 11.4 Functional scene understanding through deep learning with language and vision 390
- 11.5 Future directions 397
- 11.6 Conclusions 399
- References 399

## 12. Self-supervised temporal event segmentation inspired by cognitive theories

RAMY MOUNIR, SATHYANARAYANAN AAKUR, AND SUDEEP SARKAR

- 12.1 Introduction 406
- 12.2 The event segmentation theory from cognitive science 408
- 12.3 Version 1: single-pass temporal segmentation using prediction 410
- 12.4 Version 2: segmentation using attention-based event models 421
- 12.5 Version 3: spatio-temporal localization using prediction loss map 428
- 12.6 Other event segmentation approaches in computer vision 440

|      |             |     |
|------|-------------|-----|
| 12.7 | Conclusions | 443 |
|      | References  | 444 |

### 13. Probabilistic anomaly detection methods using learned models from time-series data for multimedia self-aware systems

CARLO REGAZZONI, ALI KRAYANI, GIULIA SLAVIC, AND LUCIO MARCENARO

|      |                                                                                      |     |
|------|--------------------------------------------------------------------------------------|-----|
| 13.1 | Introduction                                                                         | 450 |
| 13.2 | Base concepts and state of the art                                                   | 451 |
| 13.3 | Framework for computing anomaly in self-aware systems                                | 458 |
| 13.4 | Case study results: anomaly detection on multisensory data from a self-aware vehicle | 467 |
| 13.5 | Conclusions                                                                          | 476 |
|      | References                                                                           | 477 |

### 14. Deep plug-and-play and deep unfolding methods for image restoration

KAI ZHANG AND RADU TIMOFTE

|      |                                          |     |
|------|------------------------------------------|-----|
| 14.1 | Introduction                             | 481 |
| 14.2 | Half quadratic splitting (HQS) algorithm | 484 |

|      |                                      |     |
|------|--------------------------------------|-----|
| 14.3 | Deep plug-and-play image restoration | 485 |
| 14.4 | Deep unfolding image restoration     | 492 |
| 14.5 | Experiments                          | 495 |
| 14.6 | Discussion and conclusions           | 504 |
|      | References                           | 505 |

### 15. Visual adversarial attacks and defenses

CHANGJAE OH, ALESSIO XOMPERO, AND ANDREA CAVALLARO

|       |                                            |     |
|-------|--------------------------------------------|-----|
| 15.1  | Introduction                               | 511 |
| 15.2  | Problem definition                         | 512 |
| 15.3  | Properties of an adversarial attack        | 514 |
| 15.4  | Types of perturbations                     | 515 |
| 15.5  | Attack scenarios                           | 515 |
| 15.6  | Image processing                           | 522 |
| 15.7  | Image classification                       | 523 |
| 15.8  | Semantic segmentation and object detection | 529 |
| 15.9  | Object tracking                            | 529 |
| 15.10 | Video classification                       | 531 |
| 15.11 | Defenses against adversarial attacks       | 533 |
| 15.12 | Conclusions                                | 537 |
|       | References                                 | 538 |

### Index 545