

1. Record Nr.	UNINA9910620199403321
Titolo	Computer vision - ECCV 2022 . Part XXXVIII : 17th European Conference, Tel Aviv, Israel, October 23-27, 2022 : proceedings // Shai Avidan [and four others]
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
ISBN	3-031-19839-5
Descrizione fisica	1 online resource (819 pages)
Collana	Lecture Notes in Computer Science
Disciplina	006.4
Soggetti	Pattern recognition systems Computer vision
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Foreword -- Preface -- Organization -- Contents - Part XXXVIII -- Talisman: Targeted Active Learning for Object Detection with Rare Classes and Slices Using Submodular Mutual Information -- 1 Introduction -- 1.1 Our Contributions -- 2 Related Work -- 3 Background -- 3.1 Submodular Functions -- 3.2 Submodular Mutual Information (SMI) -- 3.3 Specific SMI Functions Used in TALISMAN -- 4 TALISMAN: Our Targeted Active Learning Framework for Object Detection -- 4.1 TALISMAN Framework -- 4.2 Targeted Similarity Computation -- 4.3 Using TALISMAN to Mine Rare Slices -- 4.4 Scalability of TALISMAN -- 5 Experimental Results -- 5.1 Experimental Setup -- 5.2 Baselines in All Scenarios -- 5.3 Rare Classes -- 5.4 Rare Slices -- 6 Conclusion -- References -- An Efficient Person Clustering Algorithm for Open Checkout-free Groceries -- 1 Introduction -- 2 Related Works -- 2.1 Data-stream Clustering -- 2.2 GCN on Clustering -- 3 Preliminary -- 3.1 Data-stream Clustering -- 4 Methodology -- 4.1 Crowded Sub-Graph -- 4.2 Graph Convolution Network -- 5 Experiments -- 5.1 Data and Evaluation Metric -- 5.2 Experiment Setting -- 5.3 Ablation Study -- 5.4 Comparison with Alternative Methods -- 6 Conclusion and Future Work -- References -- POP: Mining POtential Performance of New Fashion Products via Webly Cross-modal Query Expansion -- 1 Introduction -- 2 Related Literature

-- 3 Methodology -- 3.1 Image Tagging -- 3.2 Time-dependent Query Expansion -- 3.3 Image Web Search -- 3.4 Learning from Noisy Labels -- 3.5 Signal Forming -- 4 Experiments -- 4.1 Task 1: New Fashion Product Sales Curve Prediction -- 4.2 Task 2: Popularity Prediction of Fashion Styles -- 5 Conclusion -- References -- Pose Forecasting in Industrial Human-Robot Collaboration -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Background. 3.2 Separable and Sparse Graph Convolutional Networks (SeS-GCN) -- 3.3 Decoder Forecasting -- 4 The CHICO dataset -- 5 Experiments on Human3.6M -- 5.1 Modelling Choices of SeS-GCN -- 5.2 Comparison with the State-of-the-art (SoA) -- 6 Experiments on CHICO -- 6.1 Pose Forecasting Benchmark -- 6.2 Collision Detection Experiments -- 7 Conclusions -- References -- Actor-Centered Representations for Action Localization in Streaming Videos -- 1 Introduction -- 2 Related Work -- 3 Actor-Centered Action Localization -- 3.1 Extracting Perceptual Features -- 3.2 Event-Centric Perception -- 3.3 Contextualization: Actor-Centered Features -- 3.4 Hierarchical Predictive Learning -- 3.5 Attention-Based Action Localization -- 4 Experimental Evaluation -- 4.1 Data, Metrics and Baselines -- 4.2 Quantitative Analysis -- 4.3 Multi-Actor Group Activity Localization -- 4.4 Ablation Studies -- 4.5 Qualitative Analysis -- 5 Conclusion -- References -- Bandwidth-Aware Adaptive Codec for DNN Inference Offloading in IoT -- 1 Introduction -- 2 Related Work -- 3 AutoJPEG Design -- 3.1 Mathematical Modeling of JPEG Codec Workflow -- 3.2 Modeling Compressed Image Size of JPEG Encoder -- 3.3 AutoJPEG Makes JPEG Codec End-to-End Trainable -- 3.4 Solving the Optimization Problem Using ADMM -- 4 Evaluation -- 4.1 Experiment Setup -- 4.2 AutoJPEG in Semantic Segmentation -- 4.3 AutoJPEG in Image Classification -- 5 Limitations and Future Works -- 6 Conclusion -- References -- Domain Knowledge-Informed Self-supervised Representations for Workout Form Assessment -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Self-supervised Pose Contrastive Learning -- 3.2 Self-supervised Motion Disentangling -- 4 Fitness-AQA Dataset -- 5 Experiments -- 5.1 Case Study 1: Simple Conditions -- 5.2 Case Study 2: In-the-Wild Conditions -- 5.3 Cross-Exercise Transfer. 5.4 Applications to Other Domains -- 6 Conclusion -- References -- Responsive Listening Head Generation: A Benchmark Dataset and Baseline -- 1 Introduction -- 2 Related Works -- 3 Task Overview -- 4 Dataset Construction -- 5 Responsive Listening Head Generation -- 5.1 Model Architecture -- 5.2 Implementation Details -- 5.3 Experimental Results -- 6 Conclusion -- References -- Towards Scale-Aware, Robust, and Generalizable Unsupervised Monocular Depth Estimation by Integrating IMU Motion Dynamics -- 1 Introduction -- 2 Related Work -- 2.1 Unsupervised Monocular Depth Estimation -- 2.2 Scale-Aware Depth Learning -- 2.3 Visual-Inertial SLAM Systems -- 3 Methodology -- 3.1 IMU Motion Dynamics -- 3.2 The DynaDepth Framework -- 4 Experiment -- 4.1 Implementation -- 4.2 Scale-Aware Depth Estimation on KITTI -- 4.3 Generalizability on Make3D -- 4.4 Ablation Studies -- 5 Conclusion -- References -- TIPS: Text-Induced Pose Synthesis -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Text to Keypoints Generation -- 3.2 Facial Keypoints Refinement -- 3.3 Pose Rendering -- 4 Dataset and Training -- 5 Results -- 5.1 Evaluation -- 5.2 Ablation -- 6 Limitations -- 7 Conclusion -- References -- Addressing Heterogeneity in Federated Learning via Distributional Transformation -- 1 Introduction -- 2 Related Work -- 3 Motivation -- 4 Method -- 4.1 Double-Input-Channel Model Architecture -- 4.2 Joint Optimization -- 4.3 Model and Offset

Aggregation -- 4.4 Offset Aggregation Methods -- 5 Experiments -- 5.1 Results Under Different Data Distributions -- 5.2 Comparing with Data Transformation -- 5.3 Ablation Studies -- 5.4 Scalability -- 5.5 Communication Overhead -- 5.6 Prediction Visualization -- 6 Conclusion -- References -- Where in the World Is This Image? Transformer-Based Geo-localization in the Wild -- 1 Introduction -- 2 Related Works.

2.1 Single-image Geo-Localization -- 2.2 Vision Transformer -- 3 Proposed System - TransLocator -- 3.1 Global Context with Vision Transformer -- 3.2 Semantic Segmentation for Robustness to Appearance Variation -- 3.3 Single Model with Multi-Task Learning -- 3.4 Training Objective -- 4 Experiments -- 4.1 Datasets -- 4.2 Baselines -- 4.3 Evaluation Metrics -- 4.4 Implementation Details -- 5 Results, Discussions and Analysis -- 5.1 Comparison with Baselines -- 5.2 Ablation Study -- 5.3 Interpretability of TransLocator -- 5.4 Error Analysis -- 6 Conclusion -- References -- Colorization for in situ Marine Plankton Images -- 1 Introduction -- 2 Related Work -- 2.1 Underwater Plankton Cameras for in situ Plankton Color Imaging -- 2.2 Deep Learning-Based Image Colorization -- 2.3 Metrics for Colorization Evaluation -- 3 Methodology -- 3.1 Palette Customization -- 3.2 Colorization Network -- 3.3 Loss Function -- 3.4 Evaluation Metric -- 4 Experiments -- 4.1 Dataset -- 4.2 Comparisons with Previous Works -- 4.3 Ablation Experiments -- 4.4 CDSIM Metric -- 5 Conclusion -- References -- Efficient Deep Visual and Inertial Odometry with Adaptive Visual Modality Selection -- 1 Introduction -- 2 Related Works -- 2.1 Visual-Inertial Odometry -- 2.2 Adaptive Inference -- 3 Method -- 3.1 End-to-End Neural Visual-Inertial Odometry -- 3.2 Deep VIO with Visual Modality Selection -- 3.3 Training with Gumbel-Softmax -- 3.4 Loss Function -- 4 Experiments -- 4.1 Experiment Setup -- 4.2 Main Results -- 5 Conclusion -- References -- A Sketch is Worth a Thousand Words: Image Retrieval with Text and Sketch -- 1 Introduction -- 2 Related Work -- 3 Proposal: TASK-former -- 3.1 Model and Training Pipeline -- 3.2 Objective Function -- 3.3 Sketch Generation and Data Augmentation -- 3.4 Data Collection: Sketching from Memory -- 4 Results and Discussion -- 4.1 Ablation Study.

4.2 Robustness to Missing Input -- 4.3 Sketch Complexity and Retrieval Performance -- 4.4 On the Effect of Text Completeness -- 5 Limitations and Future Work -- References -- A Cloud 3D Dataset and Application-Specific Learned Image Compression in Cloud 3D -- 1 Introduction -- 2 Related Work -- 2.1 Traditional Image Compression -- 2.2 Learned Image Compression -- 2.3 Formulation of Learned Image Compression -- 3 Proposed Research -- 3.1 Application-Specific Learned Image Compression in Cloud 3D: A Practical Use Case -- 3.2 Cloud 3D Image Dataset -- 3.3 A Slim Framework: How Slim the Framework Can Be? -- 3.4 Model-Task Balance on GPU -- 4 Evaluation -- 4.1 Performance Comparison -- 4.2 Ablation Study -- 4.3 Visualization Study -- 5 Discussion and Future Work -- 6 Conclusions -- References -- AutoTransition: Learning to Recommend Video Transition Effects -- 1 Introduction -- 2 Related Work -- 3 Task Definition -- 4 Video Transition Dataset -- 4.1 Raw Data Collection -- 4.2 Data Filtering -- 5 Video Transition Recommendation -- 5.1 Pre-training Transition Embedding -- 5.2 Multi-modal Transformer -- 5.3 Transition Recommendation -- 6 Experiments -- 6.1 Implementation Details -- 6.2 Extracting Transition Embedding -- 6.3 Ablation Studies and Comparisons -- 6.4 User Study -- 7 Conclusions and Future Works -- References -- Online Segmentation of LiDAR Sequences: Dataset and Algorithm -- 1 Introduction -- 2 Related Work -- 3 HelixNet: A

Dataset for Online LiDAR Segmentation -- 4 Helix4D: Fast LiDAR Segmentation with Transformers -- 4.1 Temporal Slicing -- 4.2 Cylindrical U-Net -- 4.3 Spatio-Temporal Transformer -- 5 Evaluating Online Semantic Segmentation -- 6 Conclusion -- References -- Open-world Semantic Segmentation for LIDAR Point Clouds -- 1 Introduction -- 2 Related Work -- 3 Open-World Semantic Segmentation -- 4 Methodology.
4.1 Redundancy Classifier Framework (REAL).
