1. Record Nr. UNINA9910847092103321

Titolo	Computational Visual Media : 12th International Conference, CVM 2024, Wellington, New Zealand, April 10–12, 2024, Proceedings, Part II / / edited by Fang-Lue Zhang, Andrei Sharf
Pubbl/distr/stampa	Singapore : , : Springer Nature Singapore : , : Imprint : Springer, , 2024
ISBN	981-9720-92-3
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (384 pages)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 14593
Disciplina	006.6
Soggetti	Computer vision Pattern recognition systems Application software Computer graphics Artificial intelligence Algorithms Computer Vision Automated Pattern Recognition Computer and Information Systems Applications Computer Graphics Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part II -- Contents - Part I -- Facial Images -- Zero-Shot Real Facial Attribute Separation and Transfer at Novel Views -- 1 Introduction -- 2 Related Works -- 2.1 Explicit Face Morphable Models -- 2.2 3D-Aware Implicit Models -- 2.3 Disentanglement Representation Learning -- 3 Method -- 3.1 Model Architecture -- 3.2 EM-Like Alternating Training Procedure -- 3.3 Model Parameters Initialization -- 3.4 Rendering Refinement with Blind Face Restoration -- 4 Experiment -- 4.1 Implementation Details -- 4.2 Zero-Shot Attribute Separation from Single Image -- 4.3 Comparisons -- 4.4 Ablation Study -- 5 Conclusion -- 5.1 Limitation -- References -- Explore and Enhance the Generalization of Anomaly

DeepFake Detection -- 1 Introduction -- 2 Related Work -- 2.1 Conventional DeepFake Detection -- 2.2 Anomaly DeepFake Detection -- 3 Approach -- 3.1 Overview -- 3.2 Review and Exploration of ADFD -- 3.3 Boundary Blur Mask Generator -- 3.4 Noise Refinement Strategy -- 3.5 Algorithm -- 4 Experiments -- 4.1 Experiments Setting -- 4.2 Exploration Experiments of ADFD Methods -- 4.3 Comparison Experiments -- 5 Conclusion -- References -- Deep Tiny Network for Recognition-Oriented Face Image Quality Assessment -- 1 Introduction -- 2 Related Work -- 2.1 Image Quality Assessment -- 2.2 Face Image Quality Assessment -- 3 Method -- 3.1 Recognition-Oriented Non-reference Quality Measurement -- 3.2 Tiny Face Quality Network -- 3.3 Generating Training Dataset with Quality Labels -- 3.4 Data Sampling and Augmentation Strategy for Balancing the Distribution of Scores -- 4 Experimental Results -- 4.1 Experimental Setup -- 4.2 Datasets and Protocols -- 4.3 Visualization of Different FIQA Methods -- 4.4 Memory and Computation Costs -- 4.5 Quantitative Evaluation on IJB-B and IJB-C Datasets -- 4.6 Quantitative Evaluation on YTF Dataset.
4.7 Ablation Studies -- 5 Conclusion -- References -- Face Expression Recognition via Product-Cross Dual Attention and Neutral-Aware Anchor Loss -- 1 Introduction -- 2 Related Work -- 2.1 Landmark -- 2.2 Transformer in FER -- 2.3 Losses Used in FER -- 3 Our Method -- 3.1 Product-Cross Dual Attention Module -- 3.2 Neutral Expression Aware Anchor Loss -- 3.3 Total Loss Function -- 4 Experiments -- 4.1 Datasets -- 4.2 Implementation Details -- 4.3 Ablation Study -- 4.4 Comparison with the State-of-the-Art Methods -- 4.5 Comparison on Number of Parameters and Running Performance -- 5 Conclusion -- References -- Image Generation and Enhancement -- Deformable CNN with Position Encoding for Arbitrary-Scale Super-Resolution -- 1 Introduction -- 2 Related Work -- 2.1 Implicit Neural Representation -- 2.2 Single Image Super-Resolution (SISR) -- 2.3 Arbitrary-Scale Super-Resolution -- 3 Methods -- 3.1 Deformable Feature Unfolding (DFU) -- 3.2 Fusion with Learned Position Encoding (FPE) -- 3.3 Deep ResMLP -- 4 Experiments -- 4.1 Datasets and Metrics -- 4.2 Implementation Detail -- 4.3 Evaluation -- 4.4 Ablation Study -- 5 Conclusion -- References -- Single-Video Temporal Consistency Enhancement with Rolling Guidance -- 1 Introduction -- 2 Related Work -- 2.1 Temporal Consistency for Specific Tasks -- 2.2 Blind Video Temporal Consistency -- 2.3 Spatial Smoothing Filters and Rolling Guidance -- 3 Method -- 3.1 Overview -- 3.2 Constructing Coarse Guidance Video -- 3.3 Recovering Image Details -- 3.4 Global Refinement -- 3.5 Comparison with the Deflickering Algorithm -- 4 Experiment -- 4.1 Dataset -- 4.2 Quality Assessment -- 4.3 Comparison to State-of-the-Art Methods -- 4.4 Ablation Study -- 5 Discussion and Conclusion -- References -- GTLayout: Learning General Trees for Structured Grid Layout Generation -- 1 Introduction -- 2 Related Work -- 3 Method.
3.1 Structural Layout Representation -- 3.2 Generative Model for Structured Grid Layouts -- 3.3 Training -- 4 Evaluation -- 4.1 Layout Generation -- 4.2 Layout Reconstruction -- 4.3 Layout Interpolation -- 5 Conclusion -- References -- Image Understanding -- Silhouette-Based 6D Object Pose Estimation -- 1 Introduction -- 2 Related Work -- 2.1 Traditional Methods -- 2.2 Methods with Deep Learning -- 3 The Method -- 3.1 Problem Formulation and Notation -- 3.2 Dimensionality Reduction -- 3.3 Optimized Particle Swarm Optimization -- 4 Experiments -- 4.1 Experiments Setup -- 4.2 Comparison to State of the Art -- 4.3 Performance on YCB-V-NT and TR-RW -- 4.4 Silhouette Stability Experiments -- 4.5 Ablation Study on YCB-V -- 5 Conclusion and Outlook -- References -- Robust Light Field Depth Estimation over Occluded and Specular Regions -- 1

Introduction -- 2 Related Work -- 3 The Depth Estimation -- 3.1 Consistency Data and Confidence -- 3.2 NPCR Depth Estimation -- 3.3 Depth Refinement -- 4 Experiment -- 4.1 Occlusion Processing Comparisons -- 4.2 Specular Regions Processing -- 4.3 Depth Map -- 4.4 Computational Time -- 5 Conclusion and Limitation -- References -- Foreground and Background Separate Adaptive Equilibrium Gradients Loss for Long-Tail Object Detection -- 1 Introduction -- 2 Related Works -- 2.1 General Object Detection -- 2.2 Long-Tail Object Detection -- 3 Methodology -- 3.1 Revisiting Sigmoid Cross-Entropy Loss -- 3.2 Foreground and Background Separate Adaptive Equilibrium Gradients Loss -- 4 Experiments on LVIS -- 4.1 Datasets and Evaluation Metric -- 4.2 Implementation Details -- 4.3 Ablation Studies -- 4.4 Generalization on Stronger Models -- 4.5 Performance Analysis -- 4.6 Comparison with State-of-the-Art Methods -- 4.7 Evaluation on COCO-LT -- 4.8 Result Visualization -- 5 Conclusion -- References -- Stylization.
Multi-level Patch Transformer for Style Transfer with Single Reference Image -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Multi-level Patch Transformer Encoder -- 3.2 Dynamic Filter-Based Decoder -- 3.3 Loss Functions -- 4 Experiments and Evaluations -- 4.1 Implementation Details -- 4.2 Qualitative Evaluation -- 4.3 Ablation Study -- 4.4 User Study -- 4.5 Quantitative Evaluations -- 4.6 Discussion CycleTransformer vs CycleGAN -- 5 Conclusion and Future Work -- References -- Palette-Based Content-Aware Image Recoloring -- 1 Introduction -- 2 Related Works -- 2.1 Palette-Based Image Recoloring -- 2.2 Edit Propagation (Stroke-Based Image Recoloring) -- 2.3 Style Transfer (Example-Based Image Recoloring) -- 3 Method -- 3.1 Overview -- 3.2 Palette Extraction -- 3.3 Content-Aware Recoloring -- 4 Experiments -- 4.1 Results -- 4.2 Evaluation -- 4.3 Comparisons -- 5 Conclusion, Limitation and Future Work -- References -- FreeStyler: A Free-Form Stylization Method via Multimodal Vector Quantization -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Vector Quantization Framework -- 3.2 Pseudo-Paired Token Predictor -- 4 Experiments -- 4.1 Implementation Details -- 4.2 Qualitative Results -- 4.3 Quantitative Results -- 4.4 Ablation Study -- 4.5 Applications -- 5 Limitations and Future Work -- 6 Conclusion -- References -- Vision Meets Graphics -- Denoised Dual-Level Contrastive Network for Weakly-Supervised Temporal Sentence Grounding -- 1 Introduction -- 2 Related Work -- 2.1 Weakly-Supervised Temporal Sentence Grounding -- 2.2 Contrastive Representation Learning -- 3 The Proposed Method -- 3.1 Problem Formulation -- 3.2 Visual-Text Feature Extraction -- 3.3 Gaussian-Based Proposal Generation -- 3.4 Intra-video Contrastive Learning -- 3.5 Inter-video Contrastive Learning -- 3.6 Pseudo-Label Noise Removal -- 3.7 Training and Inference.
4 Experiments -- 4.1 Datasets -- 4.2 Evaluation Metric -- 4.3 Implementation Details -- 4.4 Comparisons with State-of-the-Art Methods -- 4.5 Ablation Study and Analysis -- 4.6 Qualitative Results -- 5 Conclusion -- References -- Isolation and Integration: A Strong Pre-trained Model-Based Paradigm for Class-Incremental Learning -- 1 Introduction -- 2 Realeated Work -- 3 Method -- 3.1 Problem Setting -- 3.2 A Simple Baseline -- 3.3 Dynamically Adaption and Aggregation -- 4 Experiments -- 4.1 Experimental Setups -- 4.2 Comparison with State of the Art -- 4.3 Ablation Study -- 5 Conclusion -- References -- Object Category-Based Visual Dialog for Effective Question Generation -- 1 Introduction -- 2 Related Work -- 3 Model -- 3.1 Object Information Extraction -- 3.2 Category Selection -- 3.3 Object Fusion Feature Update -- 3.4 Object-Self Difference Attention Module -- 3.5

Question Decoder -- 3.6 Object-Level Attention Update -- 4 Experiments -- 4.1 Dataset -- 4.2 Evaluation Metrics -- 4.3 Experiment Settings -- 4.4 Results -- 5 Conclusions -- References -- AST: An Attention-Guided Segment Transformer for Drone-Based Cross-View Geo-Localization -- 1 Introduction -- 2 Related Work -- 2.1 Image-Based Cross-View Geo-Localization -- 2.2 Vision Transformer -- 3 Proposed Method -- 3.1 Problem Formulation -- 3.2 Vision Transformer for Cross-View Geo-Localization -- 3.3 Attention-Guided Segment Tokens -- 3.4 Loss Function and Training Strategy -- 4 Experiment -- 4.1 Datasets and Evaluation Metrics -- 4.2 Implementation Details -- 4.3 Comparison with Existing Methods -- 4.4 Ablation Study -- 4.5 Visualization -- 5 Conclusion -- References -- Improved YOLOv5 Algorithm for Small Object Detection in Drone Images -- 1 Introduction -- 2 Related Work -- 2.1 Object Detection -- 2.2 Small Object Detection -- 2.3 YOLOv5 -- 3 HTH-YOLOv5 -- 3.1 Hybrid Transformer Head.
3.2 Convolutional Attention Feature Fusion Module.

Sommario/riassunto

This book constitutes the refereed proceedings of CVM 2024, the 12th International Conference on Computational Visual Media, held in Wellington, New Zealand, in April 2024. The 34 full papers were carefully reviewed and selected from 212 submissions. The papers are organized in topical sections as follows: Part I: Reconstruction and Modelling, Point Cloud, Rendering and Animation, User Interations. Part II: Facial Images, Image Generation and Enhancement, Image Understanding, Stylization, Vision Meets Graphics.