1. Record Nr. UNINA9910799221303321

Titolo	Pattern Recognition and Computer Vision : 6th Chinese Conference, PRCV 2023, Xiamen, China, October 13–15, 2023, Proceedings, Part I / / edited by Qingshan Liu, Hanzi Wang, Zhanyu Ma, Weishi Zheng, Hongbin Zha, Xilin Chen, Liang Wang, Rongrong Ji
Pubbl/distr/stampa	Singapore : , : Springer Nature Singapore : , : Imprint : Springer, , 2024
ISBN	9789819984299 9819984297
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (XIV, 513 p. 159 illus., 152 illus. in color.)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 14425
Disciplina	006
Soggetti	Image processing - Digital techniques Computer vision Artificial intelligence Application software Computer networks Computer systems Machine learning Computer Imaging, Vision, Pattern Recognition and Graphics Artificial Intelligence Computer and Information Systems Applications Computer Communication Networks Computer System Implementation Machine Learning
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part I -- Action Recognition -- Learning Bottleneck Transformer for Event Image-Voxel Feature Fusion Based Classification -- 1 Introduction -- 2 Related Work -- 3 Our Proposed Approach -- 3.1 Overview -- 3.2 Network Architecture -- 4 Experiment -- 4.1 Dataset and Evaluation Metric -- 4.2 Implementation Details -- 4.3 Comparison with Other SOTA Algorithms -- 4.4 Ablation Study -- 4.5 Parameter Analysis -- 5

Conclusion -- References -- Multi-scale Dilated Attention Graph Convolutional Network for Skeleton-Based Action Recognition -- 1 Introduction -- 2 Related Works -- 2.1 Attention Mechanism -- 2.2 Lightweight Models -- 3 Method -- 3.1 Multi-Branch Fusion Module -- 3.2 Semantic Information -- 3.3 Graph Convolution Module -- 3.4 Time Convolution Module -- 4 Experiment -- 4.1 Dataset -- 4.2 Experimental Details -- 4.3 Ablation Experiment -- 4.4 Comparison with State-of-the-Art -- 5 Action Visualization -- 6 Conclusion -- References -- Auto-Learning-GCN: An Ingenious Framework for Skeleton-Based Action Recognition -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 GCN-Based Skeleton Processing -- 3.2 The AL-GCN Module -- 3.3 The Attention Correction and Jump Model -- 3.4 Multi-stream Gaussian Weight Selection Algorithm -- 4 Experimental Results and Analysis -- 4.1 Datasets -- 4.2 Implementation Details -- 4.3 Compared with the State-of-the-Art Methods -- 4.4 Ablation Study -- 4.5 Visualization -- 5 Conclusion -- References -- Skeleton-Based Action Recognition with Combined Part-Wise Topology Graph Convolutional Networks -- 1 Introduction -- 2 Related Work -- 2.1 Skeleton-Based Action Recognition -- 2.2 Partial Graph Convolution in Skeleton-Based Action Recognition -- 3 Methods -- 3.1 Preliminaries -- 3.2 Part-Wise Spatial Modeling -- 3.3 Part-Wise Spatio-Temporal Modeling.
3.4 Model Architecture -- 4 Experiments -- 4.1 Datasets -- 4.2 Training Details -- 4.3 Ablation Studies -- 4.4 Comparison with the State-of-the-Art -- 5 Conclusion -- References -- Segmenting Key Clues to Induce Human-Object Interaction Detection -- 1 Introduction -- 2 Related Work -- 3 Approach -- 3.1 Key Features Segmentation-Based Module -- 3.2 Key Features Learning Encoder -- 3.3 Spatial Relationships Learning Graph-Based Module -- 3.4 Training and Inference -- 4 Experiments -- 4.1 Implementation Details -- 4.2 Implementation Results -- 4.3 Ablation Study -- 4.4 Qualitative Results -- 5 Conclusion -- References -- Lightweight Multispectral Skeleton and Multi-stream Graph Attention Networks for Enhanced Action Prediction with Multiple Modalities -- 1 Introduction -- 2 Related Work -- 2.1 Skeleton-Based Action Recognition -- 2.2 Dynamic Graph Neural Network -- 3 Methods -- 3.1 Spatial Embedding Component -- 3.2 Temporal Embedding Component -- 3.3 Action Prediction -- 4 Experiments and Discussion -- 4.1 NTU RGB+D Dataset -- 4.2 Experiments Setting -- 4.3 Evaluation of Human Action Recognition -- 4.4 Ablation Study -- 4.5 Visualization -- 5 Conclusion -- References -- Spatio-Temporal Self-supervision for Few-Shot Action Recognition -- 1 Introduction -- 2 Related Work -- 2.1 Few-Shot Action Recognition -- 2.2 Self-supervised Learning (SSL)-Based Few-Shot Learning -- 3 Method -- 3.1 Problem Definition -- 3.2 Spatio-Temporal Self-supervision Framework -- 4 Experiments -- 4.1 Experimental Settings -- 4.2 Comparison with State-of-the-Art Methods -- 4.3 Ablation Studies -- 5 Conclusions -- References -- A Fuzzy Error Based Fine-Tune Method for Spatio-Temporal Recognition Model -- 1 Introduction -- 2 Related Work -- 2.1 Spatio-Temporal (3D) Convolution Networks -- 2.2 Clips Selection and Features Aggregation -- 3 Proposed Method -- 3.1 Problem Definition.
3.2 Fuzzy Target -- 3.3 Fine Tune Loss Function -- 4 Experiment -- 4.1 Datasets and Implementation Details -- 4.2 Performance Comparison -- 4.3 Discussion -- 5 Conclusion -- References -- Temporal-Channel Topology Enhanced Network for Skeleton-Based Action Recognition -- 1 Introduction -- 2 Proposed Method -- 2.1 Network Architecture -- 2.2 Temporal-Channel Focus Module -- 2.3 Dynamic Channel Topology Attention Module -- 3 Experiments -- 3.1

Datasets and Implementation Details -- 3.2 Ablation Study -- 3.3 Comparison with the State-of-the-Art -- 4 Conclusion -- References -- HFGCN-Based Action Recognition System for Figure Skating -- 1 Introduction -- 2 Figure Skating Hierarchical Dataset -- 3 Figure Skating Action Recognition System -- 3.1 Data Preprocessing -- 3.2 Multi-stream Generation -- 3.3 Hierarchical Fine-Grained Graph Convolutional Neural Network (HFGCN) -- 3.4 Decision Fusion Module -- 4 Experiments and Results -- 4.1 Experimental Environment -- 4.2 Experiment Results and Analysis -- 5 Conclusion -- References -- Multi-modal Information Processing -- Image Priors Assisted Pre-training for Point Cloud Shape Analysis -- 1 Introduction -- 2 Proposed Method -- 2.1 Problem Setting -- 2.2 Overview Framework -- 2.3 Multi-task Cross-Modal SSL -- 2.4 Objective Function -- 3 Experiments and Analysis -- 3.1 Pre-training Setup -- 3.2 Downstream Tasks -- 3.3 Ablation Study -- 4 Conclusion -- References -- AMM-GAN: Attribute-Matching Memory for Person Text-to-Image Generation -- 1 Introduction -- 2 Related Work -- 2.1 Text-to-image Generative Adversarial Network -- 2.2 GANs for Person Image -- 3 Method -- 3.1 Feature Extraction -- 3.2 Multi-scale Feature Fusion Generator -- 3.3 Real-Result-Driven Discriminator -- 3.4 Objective Functions -- 4 Experiment -- 4.1 Dataset -- 4.2 Implementation -- 4.3 Evaluation Metrics -- 4.4 Quantitative Evaluation.
4.5 Qualitative Evaluation -- 4.6 Ablation Study -- 5 Conclusion -- References -- RecFormer: Recurrent Multi-modal Transformer with History-Aware Contrastive Learning for Visual Dialog -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Preliminaries -- 3.2 Model Architecture -- 3.3 Training Objectives -- 4 Experimental Setup -- 4.1 Dataset -- 4.2 Baselines -- 4.3 Evaluation Metric -- 4.4 Implementation Details -- 5 Results and Analysis -- 5.1 Main Results -- 5.2 Ablation Study -- 5.3 Attention Visualization -- 6 Conclusion -- References -- KV Inversion: KV Embeddings Learning for Text-Conditioned Real Image Action Editing -- 1 Introduction -- 2 Background -- 2.1 Text-to-Image Generation and Editing -- 2.2 Stable Diffusion Model -- 3 KV Inversion: Training-Free KV Embeddings Learning -- 3.1 Task Setting and Reason of Existing Problem -- 3.2 KV Inversion Overview -- 4 Experiments -- 4.1 Comparisons with Other Concurrent Works -- 4.2 Ablation Study -- 5 Limitations and Conclusion -- References -- Enhancing Text-Image Person Retrieval Through Nuances Varied Sample -- 1 Introduction -- 2 Relataed Work -- 2.1 Text-Image Retrieval -- 2.2 Text-Image Person Retrieval -- 3 Method -- 3.1 Feature Extraction and Alignment -- 3.2 Nuanced Variation Module -- 3.3 Image Text Matching Loss -- 3.4 Hard Negative Metric Loss -- 4 Experiment -- 4.1 Datasets and Evaluation Setting -- 4.2 Comparison with State-of-the-Art Methods -- 4.3 Ablation Study -- 5 Conclusion -- References -- Unsupervised Prototype Adapter for Vision-Language Models -- 1 Introduction -- 2 Related Work -- 2.1 Large-Scale Pre-trained Vision-Language Models -- 2.2 Adaptation Methods for Vision-Language Models -- 2.3 Self-training with Pseudo-Labeling -- 3 Method -- 3.1 Background -- 3.2 Unsupervised Prototype Adapter -- 4 Experiments -- 4.1 Image Recognition -- 4.2 Domain Generalization.
4.3 Ablation Study -- 5 Conclusion -- References -- Multimodal Causal Relations Enhanced CLIP for Image-to-Text Retrieval -- 1 Introduction -- 2 Related Works -- 3 Method -- 3.1 Overview -- 3.2 MCD: Multimodal Causal Discovery -- 3.3 MMC-CLIP -- 3.4 Image-Text Alignment -- 4 Experiments -- 4.1 Datasets and Settings -- 4.2 Results on MSCOCO -- 4.3 Results on Flickr30K -- 4.4 Ablation Studies -- 5 Conclusion -- References -- Exploring Cross-Modal Inconsistency

in Entities and Emotions for Multimodal Fake News Detection -- 1 Introduction -- 2 Related Works -- 2.1 Single-Modality Fake News Detection -- 2.2 Multimodal Fake News Detection -- 3 Methodology -- 3.1 Feature Extraction -- 3.2 Cross-Modal Contrastive Learning -- 3.3 Entity Consistency Learning -- 3.4 Emotional Consistency Learning -- 3.5 Multimodal Fake News Detector -- 4 Experiments -- 4.1 Experimental Configurations -- 4.2 Overall Performance -- 4.3 Ablation Studies -- 5 Conclusion -- References -- Deep Consistency Preserving Network for Unsupervised Cross-Modal Hashing -- 1 Introduction -- 2 The Proposed Method -- 2.1 Problem Definition -- 2.2 Deep Feature Extraction and Hashing Learning -- 2.3 Features Fusion and Similarity Matrix Construction -- 2.4 Hash Code Fusion and Reconstruction -- 2.5 Objective Function -- 3 Experiments -- 3.1 Datasets and Baselines -- 3.2 Implementation Details -- 3.3 Results and Analysis -- 4 Conclusion -- References -- Learning Adapters for Text-Guided Portrait Stylization with Pretrained Diffusion Models -- 1 Introduction -- 2 Related Work -- 2.1 Text-to-Image Diffusion Models -- 2.2 Control of Pretrained Diffusion Model -- 2.3 Text-Guided Portrait Stylizing -- 3 Method -- 3.1 Background and Preliminaries -- 3.2 Overview of Our Method -- 3.3 Portrait Stylization with Text Prompt -- 3.4 Convolution Adapter -- 3.5 Adapter Optimization -- 4 Experiments.
4.1 Implementation Settings.

Sommario/riassunto

The 13-volume set LNCS 14425-14437 constitutes the refereed proceedings of the 6th Chinese Conference on Pattern Recognition and Computer Vision, PRCV 2023, held in Xiamen, China, during October 13–15, 2023. The 532 full papers presented in these volumes were selected from 1420 submissions. The papers have been organized in the following topical sections: Action Recognition, Multi-Modal Information Processing, 3D Vision and Reconstruction, Character Recognition, Fundamental Theory of Computer Vision, Machine Learning, Vision Problems in Robotics, Autonomous Driving, Pattern Classification and Cluster Analysis, Performance Evaluation and Benchmarks, Remote Sensing Image Interpretation, Biometric Recognition, Face Recognition and Pose Recognition, Structural Pattern Recognition, Computational Photography, Sensing and Display Technology, Video Analysis and Understanding, Vision Applications andSystems, Document Analysis and Recognition, Feature Extraction and Feature Selection, Multimedia Analysis and Reasoning, Optimization and Learning methods, Neural Network and Deep Learning, Low-Level Vision and Image Processing, Object Detection, Tracking and Identification, Medical Image Processing and Analysis. .