

- 1, TITLE: Image Inpainting Based on Multi-frequency Probabilistic Inference Model  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413891>  
AUTHORS: Jin Wang, Chen Wang, Qingming Huang, Yunhui Shi, Jian-Feng Cai, Qing Zhu, Baocai Yin  
HIGHLIGHT: This paper handles this problem from a novel perspective of predicting low-frequency semantic structural contents and high-frequency detailed textures respectively, and proposes a multi-frequency probabilistic inference model(MPI model) to predict the multi-frequency information of missing regions by estimating the parametric distribution of multi-frequency features over the corresponding latent spaces.
  
- 2, TITLE: Dual Adversarial Network for Unsupervised Ground/Satellite-to-Aerial Scene Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413893>  
AUTHORS: Jianzhe Lin, Lichao Mou, Tianze Yu, Xiaoxiang Zhu, Z. Jane Wang  
HIGHLIGHT: Motivated by this, we propose a dual adversarial network for domain adaptation, where two adversarial learning processes are conducted iteratively, in correspondence with the feature adaptation and the classification task respectively.
  
- 3, TITLE: Adversarial Bipartite Graph Learning for Video Domain Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413897>  
AUTHORS: Yadan Luo, Zi Huang, Zijian Wang, Zheng Zhang, Mahsa Baktashmotlagh  
HIGHLIGHT: To overcome this limitation, in this paper, we learn a domain-agnostic video classifier instead of learning domain-invariant representations, and propose an Adversarial Bipartite Graph (ABG) learning framework which directly models the source-target interactions with a network topology of the bipartite graph.
  
- 4, TITLE: Give Me Something to Eat: Referring Expression Comprehension with Commonsense Knowledge  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413905>  
AUTHORS: Peng Wang, Dongyang Liu, Hui Li, Qi Wu  
HIGHLIGHT: In this paper, we collect a new referring expression dataset, called KB-Ref, containing \$43\$K expressions on 16k images.
  
- 5, TITLE: Single Image De-noising via Staged Memory Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413912>  
AUTHORS: Weijiang Yu, Jian Liang, Lu Li, Nong Xiao  
HIGHLIGHT: In this paper, we firstly propose a Staged Memory Network (SMNet) consisting of noise memory stage and image memory stage for explicitly exploring the staged memories of our network in single image de-noising with different noise levels.
  
- 6, TITLE: Self-supervised Dance Video Synthesis Conditioned on Music  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413932>  
AUTHORS: Xuanchi Ren, Haoran Li, Zijian Huang, Qifeng Chen  
HIGHLIGHT: We present a self-supervised approach with pose perceptual loss for automatic dance video generation.
  
- 7, TITLE: Dynamic GCN: Context-enriched Topology Learning for Skeleton-based Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413941>  
AUTHORS: Fanfan Ye, Shiliang Pu, Qiaoyong Zhong, Chao Li, Di Xie, Huiming Tang  
HIGHLIGHT: In this paper, we propose Dynamic GCN, in which a novel convolutional neural network named Context-encoding Network (CeN) is introduced to learn skeleton topology automatically.
  
- 8, TITLE: Meta Parsing Networks: Towards Generalized Few-shot Scene Parsing with Adaptive Metric Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413944>  
AUTHORS: Peike Li, Yunchao Wei, Yi Yang  
HIGHLIGHT: In this work, we advance this few-shot segmentation paradigm towards a more challenging yet general scenario, i.e., Generalized Few-shot Scene Parsing (GFSP).
  
- 9, TITLE: CODAN: Counting-driven Attention Network for Vehicle Detection in Congested Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413945>  
AUTHORS: Wei Li, Zhenting Wang, Xiao Wu, Ji Zhang, Qiang Peng, Hongliang Li  
HIGHLIGHT: In this paper, we explore the dense vehicle detection given the number of vehicles.
  
- 10, TITLE: Webly Supervised Image Classification with Metadata: Automatic Noisy Label Correction via Visual-Semantic Graph  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413952>  
AUTHORS: Jingkan Yang, Weirong Chen, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang  
HIGHLIGHT: In this paper, we propose an automatic label corrector VSGraph-LC based on the visual-semantic graph.

11, TITLE: CRSSC: Salvage Reusable Samples from Noisy Data for Robust Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413978>  
AUTHORS: Zeren Sun, Xian-Sheng Hua, Yazhou Yao, Xiu-Shen Wei, Guosheng Hu, Jian Zhang  
HIGHLIGHT: To this end, we propose a certainty-based reusable sample selection and correction approach, termed as CRSSC, for coping with label noise in training deep FG models with web images.

12, TITLE: Learning From Music to Visual Storytelling of Shots: A Deep Interactive Learning Mechanism  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413985>  
AUTHORS: Jen-Chun Lin, Wen-Li Wei, Yen-Yu Lin, Tyng-Luh Liu, Hong-Yuan Mark Liao  
HIGHLIGHT: In this study, we present a deep interactive learning (DIL) mechanism for building a compact yet accurate sequence-to-sequence model to accomplish the task.

13, TITLE: TextRay: Contour-based Geometric Modeling for Arbitrary-shaped Scene Text Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413819>  
AUTHORS: Fangfang Wang, Yifeng Chen, Fei Wu, Xi Li  
HIGHLIGHT: In this work, we propose an arbitrary-shaped text detection method, namely TextRay, which conducts top-down contour-based geometric modeling and geometric parameter learning within a single-shot anchor-free framework.

14, TITLE: Weakly Supervised Real-time Image Cropping based on Aesthetic Distributions  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413824>  
AUTHORS: Peng Lu, Jiahui Liu, Xujun Peng, Xiaojie Wang  
HIGHLIGHT: In order to tackle this problem, a weakly supervised cropping framework is proposed, where the distribution dissimilarity between high quality images and cropped images is used to guide the coordinate predictor's training and the ground truths of cropping windows are not required by the proposed method.

15, TITLE: Towards Unsupervised Crowd Counting via Regression-Detection Bi-knowledge Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413825>  
AUTHORS: Yuting Liu, Zheng Wang, Miaojing Shi, Shin'ichi Satoh, Qijun Zhao, Hongyu Yang  
HIGHLIGHT: In this paper, we explore it in a transfer learning setting where we learn to detect and count persons in an unlabeled target set by transferring bi-knowledge learnt from regression- and detection-based models in a labeled source set.

16, TITLE: Occluded Prohibited Items Detection: An X-ray Security Inspection Benchmark and De-occlusion Attention Module  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413828>  
AUTHORS: Yanlu Wei, Renshuai Tao, Zhangjie Wu, Yuqing Ma, Libo Zhang, Xianglong Liu  
HIGHLIGHT: In this work, we contribute the first high-quality object detection dataset for security inspection, named Occluded Prohibited Items X-ray (OPIXray) image benchmark.

17, TITLE: Temporally Guided Music-to-Body-Movement Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413848>  
AUTHORS: Hsuan-Kai Kao, Li Su  
HIGHLIGHT: This paper presents a neural network model to generate virtual violinist's 3-D skeleton movements from music audio.

18, TITLE: Compositional Few-Shot Recognition with Primitive Discovery and Enhancing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413849>  
AUTHORS: Yixiong Zou, Shanghang Zhang, Ke Chen, Yonghong Tian, Yaowei Wang, Jos&eacute; M. F. Moura  
HIGHLIGHT: Inspired by such capability of humans, to imitate humans' ability of learning visual primitives and composing primitives to recognize novel classes, we propose an approach to FSL to learn a feature representation composed of important primitives, which is jointly trained with two parts, i.e. primitive discovery and primitive enhancing.

19, TITLE: InteractGAN: Learning to Generate Human-Object Interaction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413854>  
AUTHORS: Chen Gao, Si Liu, Defa Zhu, Quan Liu, Jie Cao, Haoqian He, Ran He, Shuicheng Yan  
HIGHLIGHT: In this work, we introduce an Interact-GAN to solve this challenging task.

20, TITLE: Category-specific Semantic Coherency Learning for Fine-grained Image Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413871>  
AUTHORS: Shijie Wang, Zhihui Wang, Haojie Li, Wanli Ouyang

**HIGHLIGHT:** To address this issue, we propose an end-to-end Category-specific Semantic Coherency Network (CSC-Net) to semantically align the discriminative regions of the same subcategory.

21, **TITLE:** Scene-Aware Context Reasoning for Unsupervised Abnormal Event Detection in Videos

<https://dl.acm.org/doi/abs/10.1145/3394171.3413887>

**AUTHORS:** Che Sun, Yunde Jia, Yao Hu, Yuwei Wu

**HIGHLIGHT:** In this paper, we propose a scene-aware context reasoning method that exploits context information from visual features for unsupervised abnormal event detection in videos, which bridges the semantic gap between visual context and the meaning of abnormal events.

22, **TITLE:** Light Field Super-resolution via Attention-Guided Fusion of Hybrid Lenses

<https://dl.acm.org/doi/abs/10.1145/3394171.3413585>

**AUTHORS:** Jing Jin, Junhui Hou, Jie Chen, Sam Kwong, Jingyi Yu

**HIGHLIGHT:** To tackle this challenge, we propose a novel end-to-end learning-based approach, which can comprehensively utilize the specific characteristics of the input from two complementary and parallel perspectives.

23, **TITLE:** Trajectory Prediction in Heterogeneous Environment via Attended Ecology Embedding

<https://dl.acm.org/doi/abs/10.1145/3394171.3413602>

**AUTHORS:** Wei-Cheng Lai, Zi-Xiang Xia, Hao-Siang Lin, Lien-Feng Hsu, Hong-Han Shuai, I-Hong Jhuo, Wen-Huang Cheng

**HIGHLIGHT:** In this paper, we consider the practical environment of predicting trajectory in the heterogeneous traffic ecology.

24, **TITLE:** Text-Embedded Bilinear Model for Fine-Grained Visual Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413638>

**AUTHORS:** Liang Sun, Xiang Guan, Yang Yang, Lei Zhang

**HIGHLIGHT:** In this paper, we propose a Text-Embedded Bilinear (TEB) model which incorporates texts as extra guidance for fine-grained recognition.

25, **TITLE:** Learning Scales from Points: A Scale-aware Probabilistic Model for Crowd Counting

<https://dl.acm.org/doi/abs/10.1145/3394171.3413642>

**AUTHORS:** Zhiheng Ma, Xing Wei, Xiaopeng Hong, Yihong Gong

**HIGHLIGHT:** In this paper, we propose a scale-aware probabilistic model to handle this problem.

26, **TITLE:** Learning Global Structure Consistency for Robust Object Tracking

<https://dl.acm.org/doi/abs/10.1145/3394171.3413644>

**AUTHORS:** Bi Li, Chengquan Zhang, Zhibin Hong, Xu Tang, Jingtuo Liu, Junyu Han, Errui Ding, Wenyu Liu

**HIGHLIGHT:** Specifically, we propose an effective and efficient short-term model that learns to exploit the global structure consistency in a short time and thus can handle fast variations and distractors.

27, **TITLE:** Campus3D: A Photogrammetry Point Cloud Benchmark for Hierarchical Understanding of Outdoor Scene

<https://dl.acm.org/doi/abs/10.1145/3394171.3413661>

**AUTHORS:** Xinke Li, Chongshou Li, Zekun Tong, Andrew Lim, Junsong Yuan, Yuwei Wu, Jing Tang, Raymond Huang

**HIGHLIGHT:** To facilitate the research of this area, we present a richly-annotated 3D point cloud dataset for multiple outdoor scene understanding tasks and also an effective learning framework for its hierarchical segmentation task.

28, **TITLE:** Instability of Successive Deep Image Compression

<https://dl.acm.org/doi/abs/10.1145/3394171.3413680>

**AUTHORS:** Jun-Hyuk Kim, Soobeom Jang, Jun-Ho Choi, Jong-Seok Lee

**HIGHLIGHT:** In this paper, we conduct comprehensive analysis of successive deep image compression.

29, **TITLE:** ALANET: Adaptive Latent Attention Network for Joint Video Deblurring and Interpolation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413686>

**AUTHORS:** Akash Gupta, Abhishek Aich, Amit K. Roy-Chowdhury

**HIGHLIGHT:** We introduce a novel architecture, Adaptive Latent Attention Network (ALANET), which synthesizes sharp high frame-rate videos with no prior knowledge of input frames being blurry or not, thereby performing the task of both deblurring and interpolation.

30, **TITLE:** PCPL: Predicate-Correlation Perception Learning for Unbiased Scene Graph Generation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413722>

AUTHORS: Shaotian Yan, Chen Shen, Zhongming Jin, Jianqiang Huang, Rongxin Jiang, Yaowu Chen, Xian-Sheng Hua  
HIGHLIGHT: Therefore, we propose a novel Predicate-Correlation Perception Learning (PCPL for short) scheme to adaptively seek out appropriate loss weights by directly perceiving and utilizing the correlation among predicate classes.

31, TITLE: Discriminative Spatial Feature Learning for Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413730>  
AUTHORS: Peixi Peng, Yonghong Tian, Yangru Huang, Xiangqian Wang, HuiLong An  
HIGHLIGHT: To handle this challenge, a novel method is proposed to learn the discriminative spatial features.

32, TITLE: AdaHGNN: Adaptive Hypergraph Neural Networks for Multi-Label Image Classification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414046>  
AUTHORS: Xiangping Wu, Qingcai Chen, Wei Li, Yulun Xiao, Baotian Hu  
HIGHLIGHT: In this paper, we propose a high-order semantic learning model based on adaptive hypergraph neural networks (AdaHGNN) to boost multi-label classification performance.

33, TITLE: Reinforced Similarity Learning: Siamese Relation Networks for Robust Object Tracking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413743>  
AUTHORS: Dawei Zhang, Zhonglong Zheng, Minglu Li, Xiaowei He, Tianxiang Wang, Liyuan Chen, Riheng Jia, Feilong Lin  
HIGHLIGHT: In this paper, we pay more attention to learn an outstanding similarity measure for robust tracking.

34, TITLE: Deep Structural Contour Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413750>  
AUTHORS: Ruoxi Deng, Shengjun Liu  
HIGHLIGHT: In this work, we aim to develop a high-performance contour detection system.

35, TITLE: Cross-modal Non-linear Guided Attention and Temporal Coherence in Multi-modal Deep Video Models  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413756>  
AUTHORS: Saurabh Sahu, Palash Goyal, Shalini Ghosh, Chul Lee  
HIGHLIGHT: We show how using non-linear guided cross-modal signals and temporal coherence can improve the performance of multi-modal machine learning (ML) models for video analysis tasks like categorization.

36, TITLE: IR-GAN: Image Manipulation with Linguistic Instruction by Increment Reasoning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413777>  
AUTHORS: Zhenhuan Liu, Jincan Deng, Liang Li, Shaofei Cai, Qianqian Xu, Shuhui Wang, Qingming Huang  
HIGHLIGHT: To address this issue, we propose an Increment Reasoning Generative Adversarial Network (IR-GAN), which aims to reason the consistency between visual increment in images and semantic increment in instructions.

37, TITLE: Fine-Grained Similarity Measurement between Educational Videos and Exercises  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413783>  
AUTHORS: Xin Wang, Wei Huang, Qi Liu, Yu Yin, Zhenya Huang, Le Wu, Jianhui Ma, Xue Wang  
HIGHLIGHT: In this paper, we explore to measure the fine-grained similarity by leveraging multimodal information.

38, TITLE: One-shot Text Field labeling using Attention and Belief Propagation for Structure Information Extraction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413511>  
AUTHORS: Mengli Cheng, Minghui Qiu, Xing Shi, Jun Huang, Wei Lin  
HIGHLIGHT: To alleviate these problems, we proposed a novel deep end-to-end trainable approach for one-shot text field labeling, which makes use of attention mechanism to transfer the layout information between document images.

39, TITLE: Grad: Learning for Overhead-aware Adaptive Video Streaming with Scalable Video Coding  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413512>  
AUTHORS: Yunzhuo Liu, Bo Jiang, Tian Guo, Ramesh K. Sitaraman, Don Towsley, Xinbing Wang  
HIGHLIGHT: In this work, we propose a deep reinforcement learning method called Grad for designing ABR algorithms that take advantage of the quality upgrade mechanism of SVC.

40, TITLE: Efficient Adaptation of Neural Network Filter for Video Compression  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413536>  
AUTHORS: Yat-Hong Lam, Alireza Zare, Francesco Cricri, Jani Lainema, Miska M. Hannuksela  
HIGHLIGHT: We present an efficient finetuning methodology for neural-network filters which are applied as a postprocessing artifact-removal step in video coding pipelines.

- 41, TITLE: SonoSpace: Visual Feedback of Timbre with Unsupervised Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413542>  
AUTHORS: Naoki Kimura, Keisuke Shiro, Yota Takakura, Hiromi Nakamura, Jun Rekimoto  
HIGHLIGHT: Our goal is to develop a low-cost learning system that substitutes the teacher.
- 42, TITLE: Single Image Deraining via Scale-space Invariant Attention Neural Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413554>  
AUTHORS: Bo Pang, Deming Zhai, Junjun Jiang, Xianming Liu  
HIGHLIGHT: In this paper, we tackle the notion of scale that deals with visual changes in appearance of rain streaks with respect to the camera.
- 43, TITLE: Every Moment Matters: Detail-Aware Networks to Bring a Blurry Image Alive  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413929>  
AUTHORS: Kaihao Zhang, Wenhan Luo, Bj&ouml;m Stenger, Wenqi Ren, Lin Ma, Hongdong Li  
HIGHLIGHT: In order to alleviate this problem, we propose a detail-aware network with three consecutive stages to improve the reconstruction quality by addressing specific aspects in the recovery process.
- 44, TITLE: ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414031>  
AUTHORS: Weiqing Min, Linhu Liu, Zhiling Wang, Zhengdong Luo, Xiaoming Wei, Xiaolin Wei, Shuqiang Jiang  
HIGHLIGHT: To encourage further progress in food recognition, we introduce the dataset ISIA Food-500 with 500 categories from the list in the Wikipedia and 399,726 images, a more comprehensive food dataset that surpasses existing popular benchmark datasets by category coverage and data volume.
- 45, TITLE: An Egocentric Action Anticipation Framework via Fusing Intuition and Analysis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413964>  
AUTHORS: Tianyu Zhang, Weiqing Min, Ying Zhu, Yong Rui, Shuqiang Jiang  
HIGHLIGHT: In this paper, we focus on egocentric action anticipation from videos, which enables various applications, such as helping intelligent wearable assistants understand users' needs and enhance their capabilities in the interaction process.
- 46, TITLE: Multi-Person Action Recognition in Microwave Sensors  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413801>  
AUTHORS: Diangang Li, Jianquan Liu, Shoji Nishimura, Yuka Hayashi, Jun Suzuki, Yihong Gong  
HIGHLIGHT: To address the challenges, we propose a novel learning framework by designed original loss functions with the considerations on weakly-supervised multi-label learning and attention mechanism to improve the accuracy for action recognition.
- 47, TITLE: Coupling Deep Textural and Shape Features for Sketch Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413810>  
AUTHORS: Qi Jia, Xin Fan, Meiyu Yu, Yuqing Liu, Dingrong Wang, Longin Jan Latecki  
HIGHLIGHT: In this paper, we explicitly explore the shape properties of sketches, which has almost been neglected before in the context of deep learning, and propose a sequential dual learning strategy that combines both shape and texture features.
- 48, TITLE: Look, Read and Feel: Benchmarking Ads Understanding with Multimodal Multitask Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413582>  
AUTHORS: Huaizheng Zhang, Yong Luo, Qiming Ai, Yonggang Wen, Han Hu  
HIGHLIGHT: This motivates us to develop a novel deep multimodal multitask framework that integrates multiple modalities to achieve effective topic and sentiment prediction simultaneously for ads understanding.
- 49, TITLE: Not made for each other- Audio-Visual Dissonance-based Deepfake Detection and Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413700>  
AUTHORS: Komal Chugh, Parul Gupta, Abhinav Dhall, Ramanathan Subramanian  
HIGHLIGHT: We propose detection of deepfake videos based on the dissimilarity between the audio and visual modalities, termed as the Modality Dissonance Score (MDS).
- 50, TITLE: Hearing like Seeing: Improving Voice-Face Interactions and Associations via Adversarial Deep Semantic Matching Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413710>  
AUTHORS: Kai Cheng, Xin Liu, Yiu-ming Cheung, Rui Wang, Xing Xu, Bineng Zhong

**HIGHLIGHT:** In this paper, we present a novel adversarial deep semantic matching network for efficient voice-face interactions and associations, which can well learn the correspondence between voices and faces for various cross-modal matching and retrieval tasks.

51, **TITLE:** Multimodal Multi-Task Financial Risk Forecasting

<https://dl.acm.org/doi/abs/10.1145/3394171.3413752>

**AUTHORS:** Ramit Sawhney, Puneet Mathur, Ayush Mangal, Piyush Khanna, Rajiv Ratn Shah, Roger Zimmermann

**HIGHLIGHT:** In this work, we present a multi-task solution that utilizes domain specialized textual features and audio attentive alignment for predictive financial risk and price modeling.

52, **TITLE:** Down to the Last Detail: Virtual Try-on with Fine-grained Details

<https://dl.acm.org/doi/abs/10.1145/3394171.3413514>

**AUTHORS:** Jiahang Wang, Tong Sha, Wei Zhang, Zhoujun Li, Tao Mei

**HIGHLIGHT:** In this work, we propose a multi-stage framework to synthesize person images, where fine-grained details can be well preserved.

53, **TITLE:** Temporal Denoising Mask Synthesis Network for Learning Blind Video Temporal Consistency

<https://dl.acm.org/doi/abs/10.1145/3394171.3413788>

**AUTHORS:** Yifeng Zhou, Xing Xu, Fumin Shen, Lianli Gao, Huimin Lu, Heng Tao Shen

**HIGHLIGHT:** In this paper, we consider enforcing temporal consistency in a video as a temporal denoising problem that removing the flickering effect in given unstable pre-processed frames.

54, **TITLE:** A Lip Sync Expert Is All You Need for Speech to Lip Generation In the Wild

<https://dl.acm.org/doi/abs/10.1145/3394171.3413532>

**AUTHORS:** K R Prajwal, Rudrabha Mukhopadhyay, Vinay P. Nambodiri, C.V. Jawahar

**HIGHLIGHT:** In this work, we investigate the problem of lip-syncing a talking face video of an arbitrary identity to match a target speech segment.

55, **TITLE:** MEmoR: A Dataset for Multimodal Emotion Reasoning in Videos

<https://dl.acm.org/doi/abs/10.1145/3394171.3413909>

**AUTHORS:** Guangyao Shen, Xin Wang, Xuguang Duan, Hongzhi Li, Wenwu Zhu

**HIGHLIGHT:** In this work, we present the task of multimodal emotion reasoning in videos.

56, **TITLE:** Modeling both Intra- and Inter-modal Influence for Real-Time Emotion Detection in Conversations

<https://dl.acm.org/doi/abs/10.1145/3394171.3413949>

**AUTHORS:** Dong Zhang, Weisheng Zhang, Shoushan Li, Qiaoming Zhu, Guodong Zhou

**HIGHLIGHT:** Therefore, this paper proposes a bidirectional dynamic dual influence network for real-time emotion detection in conversations, which can simultaneously model both intra- and inter-modal influence with bidirectional information propagation for current utterance and its historical utterances.

57, **TITLE:** Transformer-based Label Set Generation for Multi-modal Multi-label Emotion Detection

<https://dl.acm.org/doi/abs/10.1145/3394171.3413577>

**AUTHORS:** Xincheng Ju, Dong Zhang, Junhui Li, Guodong Zhou

**HIGHLIGHT:** To simultaneously model above two kinds of dependency, we propose a unified approach, namely multi-modal emotion set generation network (MESGN) to generate an emotion set for an utterance.

58, **TITLE:** CM-BERT: Cross-Modal BERT for Text-Audio Sentiment Analysis

<https://dl.acm.org/doi/abs/10.1145/3394171.3413690>

**AUTHORS:** Kaicheng Yang, Hua Xu, Kai Gao

**HIGHLIGHT:** In this paper, we propose the Cross-Modal BERT (CM-BERT), which relies on the interaction of text and audio modality to fine-tune the pre-trained BERT model.

59, **TITLE:** AffectI: A Game for Diverse, Reliable, and Efficient Affective Image Annotation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413744>

**AUTHORS:** Xingkun Zuo, Jiye Li, Qili Zhou, Jianjun Li, Xiaoyang Mao

**HIGHLIGHT:** This paper proposes a novel affective image annotation technique, AffectI, for efficiently collecting diverse and reliable emotional labels with the estimate emotion distribution for images based on the concept of Game With a Purpose (GWAP).

60, **TITLE:** Attentive One-Dimensional Heatmap Regression for Facial Landmark Detection and Tracking

<https://dl.acm.org/doi/abs/10.1145/3394171.3413509>

AUTHORS: Shi Yin, Shangfei Wang, Xiaoping Chen, Enhong Chen, Cong Liang  
HIGHLIGHT: To address this, we propose a novel attentive one-dimensional heatmap regression method for facial landmark localization.

61, TITLE: Domain Adaptive Person Re-Identification via Coupling Optimization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413904>  
AUTHORS: Xiaobin Liu, Shiliang Zhang  
HIGHLIGHT: To handle those two challenges, this paper proposes a coupling optimization method including the Domain-Invariant Mapping (DIM) method and the Global-Local distance Optimization (GLO), respectively.

62, TITLE: Dual-Structure Disentangling Variational Generation for Data-Limited Face Parsing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413919>  
AUTHORS: Peipei Li, Yinglu Liu, Hailin Shi, Xiang Wu, Yibo Hu, Ran He, Zhenan Sun  
HIGHLIGHT: Since there are inaccurate pixel-level labels in synthesized parsing maps, we introduce a coarseness-tolerant learning algorithm, to effectively handle these noisy or uncertain labels.

63, TITLE: Accurate UAV Tracking with Distance-Injected Overlap Maximization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413959>  
AUTHORS: Chunhui Zhang, Shiming Ge, Kangkai Zhang, Dan Zeng  
HIGHLIGHT: In this work, we propose to alleviate this issue with distance-injected overlap maximization.

64, TITLE: PiRhDy: Learning Pitch-, Rhythm-, and Dynamics-aware Embeddings for Symbolic Music  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414032>  
AUTHORS: Hongru Liang, Wenqiang Lei, Paul Yaozhu Chan, Zhenglu Yang, Maosong Sun, Tat-Seng Chua  
HIGHLIGHT: In this work, we provide a comprehensive solution by proposing a novel framework named PiRhDy that integrates pitch, rhythm, and dynamics information seamlessly.

65, TITLE: Cloze Test Helps: Effective Video Anomaly Detection via Learning to Complete Video Events  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413973>  
AUTHORS: Guang Yu, Siqi Wang, Zhiping Cai, En Zhu, Chuanfu Xu, Jianping Yin, Marius Kloft  
HIGHLIGHT: Inspired by frequently-used cloze test in language study, we propose a brand-new VAD solution named Video Event Completion (VEC) to bridge gaps above: First, we propose a novel pipeline to achieve both precise and comprehensive enclosure of video activities.

66, TITLE: Pose-native Network Architecture Search for Multi-person Human Pose Estimation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413842>  
AUTHORS: Qian Bao, Wu Liu, Jun Hong, Lingyu Duan, Tao Mei  
HIGHLIGHT: In this work, we present the Pose-native Network Architecture Search (PoseNAS) to simultaneously design a better pose encoder and pose decoder for pose estimation.

67, TITLE: Beyond the Attention: Distinguish the Discriminative and Confusable Features For Fine-grained Image Classification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413883>  
AUTHORS: Xiruo Shi, Liutong Xu, Pengfei Wang, Yuanyuan Gao, Haifang Jian, Wu Liu  
HIGHLIGHT: In this paper, we introduce a novel classification approach, named Logical-based Feature Extraction Model (LAFE for short) to address this issue.

68, TITLE: BlockMix: Meta Regularization and Self-Calibrated Inference for Metric-Based Meta-Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413884>  
AUTHORS: Hao Tang, Zechao Li, Zhimao Peng, Jinhui Tang  
HIGHLIGHT: Toward this end, we propose new mechanisms to learn generalized and discriminative feature embeddings as well as improve the robustness of classifiers against prediction corruptions for meta-learning.

69, TITLE: Fine-grained Feature Alignment with Part Perspective Transformation for Vehicle ReID  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413573>  
AUTHORS: Dechao Meng, Liang Li, Shuhui Wang, Xingyu Gao, Zheng-Jun Zha, Qingming Huang  
HIGHLIGHT: In this paper, we propose part perspective transformation module (PPT) to map the different parts of vehicle into a unified perspective respectively.

70, TITLE: Compact Bilinear Augmented Query Structured Attention for Sport Highlights Classification

<https://dl.acm.org/doi/abs/10.1145/3394171.3413595>

AUTHORS: Yanbin Hao, Hao Zhang, Chong-Wah Ngo, Qiang Liu, Xiaojun Hu  
HIGHLIGHT: Specifically, we adapt the hierarchical attention neural networks, which contain learnable query-scheme, on the video to identify discriminative spatial/temporal visual clues within highlight clips.

71, TITLE: Semantic Image Analogy with a Conditional Single-Image GAN

<https://dl.acm.org/doi/abs/10.1145/3394171.3413601>

AUTHORS: Jiacheng Li, Zhiwei Xiong, Dong Liu, Xuejin Chen, Zheng-Jun Zha  
HIGHLIGHT: To accomplish this task, we propose a novel method to model the patch-level correspondence between semantic layout and appearance of a single image by training a single-image GAN that takes semantic labels as conditional input.

72, TITLE: A Structured Graph Attention Network for Vehicle Re-Identification

<https://dl.acm.org/doi/abs/10.1145/3394171.3413607>

AUTHORS: Yangchun Zhu, Zheng-Jun Zha, Tianzhu Zhang, Jiawei Liu, Jiebo Luo  
HIGHLIGHT: In this paper, we propose a Structured Graph Attention network (SGAT) to fully exploit these relationships and allow the message propagation to update the features of graph nodes.

73, TITLE: Contextual Multi-Scale Feature Learning for Person Re-Identification

<https://dl.acm.org/doi/abs/10.1145/3394171.3414038>

AUTHORS: Baoyu Fan, Li Wang, Runze Zhang, Zhenhua Guo, Yaqian Zhao, Rengang Li, Weifeng Gong  
HIGHLIGHT: In this paper, we proposed a novel architecture, namely contextual multi-scale network (CMSNet), for learning common and contextual multi-scale representations simultaneously.

74, TITLE: Space-Time Video Super-Resolution Using Temporal Profiles

<https://dl.acm.org/doi/abs/10.1145/3394171.3413667>

AUTHORS: Zeyu Xiao, Zhiwei Xiong, Xueyang Fu, Dong Liu, Zheng-Jun Zha  
HIGHLIGHT: In this paper, we propose a novel space-time video super-resolution method, which aims to recover a high-frame-rate and high-resolution video from its low-frame-rate and low-resolution observation.

75, TITLE: Black Re-ID: A Head-shoulder Descriptor for the Challenging Problem of Person Re-Identification

<https://dl.acm.org/doi/abs/10.1145/3394171.3414056>

AUTHORS: Boqiang Xu, Lingxiao He, Xingyu Liao, Wu Liu, Zhenan Sun, Tao Mei  
HIGHLIGHT: To solve this problem, rather than relying on the clothing information, we propose to exploit head-shoulder features to assist person Re-ID.

76, TITLE: SalGCN: Saliency Prediction for 360-Degree Images Based on Spherical Graph Convolutional Networks

<https://dl.acm.org/doi/abs/10.1145/3394171.3413733>

AUTHORS: Haoran Lv, Qin Yang, Chenglin Li, Wenrui Dai, Junni Zou, Hongkai Xiong  
HIGHLIGHT: In this paper, we propose a saliency prediction framework for 360-degree images based on graph convolutional networks (SalGCN), which directly applies to the spherical graph signals.

77, TITLE: LIGHTEN: Learning Interactions with Graph and Hierarchical Temporal Networks for HOI in videos

<https://dl.acm.org/doi/abs/10.1145/3394171.3413778>

AUTHORS: Sai Praneeth Reddy Sunkesula, Rishabh Dabral, Ganesh Ramakrishnan  
HIGHLIGHT: While traditional methods formulate the problem as inference on a sequence of video segments, we present a hierarchical approach, LIGHTEN, to learn visual features to effectively capture spatio-temporal cues at multiple granularities in a video.

78, TITLE: Concept-based Explanation for Fine-grained Images and Its Application in Infectious Keratitis Classification

<https://dl.acm.org/doi/abs/10.1145/3394171.3413557>

AUTHORS: Zhengqing Fang, Kun Kuang, Yuxiao Lin, Fei Wu, Yu-Feng Yao  
HIGHLIGHT: In this paper, we focus on the real application problem of classification of infectious keratitis and propose a visual concept mining (VCM) method to explain the fine-grained infectious keratitis images.

79, TITLE: Guided Attention Network for Object Detection and Counting on Drones

<https://dl.acm.org/doi/abs/10.1145/3394171.3413816>

AUTHORS: Cai YuanQiang, Dawei Du, Libo Zhang, Longyin Wen, Weiqiang Wang, Yanjun Wu, Siwei Lyu  
HIGHLIGHT: In this paper, we propose a new Guided Attention network (GANet) to deal with both object detection and counting tasks based on the feature pyramid.



- 80, TITLE: PIDNet: An Efficient Network for Dynamic Pedestrian Intrusion Detection  
https://dl.acm.org/doi/abs/10.1145/3394171.3413837  
AUTHORS: Jingchen Sun, Jiming Chen, Tao Chen, Jiayuan Fan, Shibo He  
HIGHLIGHT: In this paper, we propose a novel and efficient multi-task deep neural network, PIDNet, to solve this problem.
- 81, TITLE: VONAS: Network Design in Visual Odometry using Neural Architecture Search  
https://dl.acm.org/doi/abs/10.1145/3394171.3413866  
AUTHORS: Xing Cai, Lanqing Zhang, Chengyuan Li, Ge Li, Thomas H. Li  
HIGHLIGHT: Therefore, this paper explores the network design for the VO task and proposes a more general single path based one-shot NAS, named VONAS, which can model sequential information for video-related tasks.
- 82, TITLE: Learning from the Past: Meta-Continual Learning with Knowledge Embedding for Jointly Sketch, Cartoon, and Caricature Face Recognition  
https://dl.acm.org/doi/abs/10.1145/3394171.3413892  
AUTHORS: Wenbo Zheng, Lan Yan, Fei-Yue Wang, Chao Gou  
HIGHLIGHT: We propose a novel framework termed as Meta-Continual Learning with Knowledge Embedding to address the task of jointly sketch, cartoon, and caricature face recognition.
- 83, TITLE: ChoreoNet: Towards Music to Dance Synthesis with Choreographic Action Unit  
https://dl.acm.org/doi/abs/10.1145/3394171.3414005  
AUTHORS: Zijie Ye, Haozhe Wu, Jia Jia, Yaohua Bu, Wei Chen, Fanbo Meng, Yanfeng Wang  
HIGHLIGHT: Inspired by these, we systematically study such two-stage choreography approach and construct a dataset to incorporate such choreography knowledge.
- 84, TITLE: InvisibleFL: Federated Learning over Non-Informative Intermediate Updates against Multimedia Privacy Leakages  
https://dl.acm.org/doi/abs/10.1145/3394171.3413923  
AUTHORS: Qiushi Li, Wenwu Zhu, Chao Wu, Xinglin Pan, Fan Yang, Yuezhi Zhou, Yaoxue Zhang  
HIGHLIGHT: In this paper, we propose a privacy-preserving solution that avoids multimedia privacy leakages in federated learning.
- 85, TITLE: Asymmetric Deep Hashing for Efficient Hash Code Compression  
https://dl.acm.org/doi/abs/10.1145/3394171.3414033  
AUTHORS: Shu Zhao, Dayan Wu, Wanqian Zhang, Yu Zhou, Bo Li, Weiping Wang  
HIGHLIGHT: In this paper, we propose a novel deep hashing method, called Code Compression oriented Deep Hashing (CCDH), for efficiently compressing hash codes.
- 86, TITLE: A Human-Computer Duet System for Music Performance  
https://dl.acm.org/doi/abs/10.1145/3394171.3413921  
AUTHORS: Yuen-Jen Lin, Hsuan-Kai Kao, Yih-Chih Tseng, Ming Tsai, Li Su  
HIGHLIGHT: In this paper, we firstly create a virtual violinist, who can collaborate with a human pianist to perform chamber music automatically without any intervention.
- 87, TITLE: Photo Stand-Out: Photography with Virtual Character  
https://dl.acm.org/doi/abs/10.1145/3394171.3413957  
AUTHORS: Yujia Wang, Sifan Hou, Bing Ning, Wei Liang  
HIGHLIGHT: In this paper, we propose a novel optimization framework to synthesize an aesthetic pose for the virtual character with respect to the presented user's pose.
- 88, TITLE: Norm-in-Norm Loss with Faster Convergence and Better Performance for Image Quality Assessment  
https://dl.acm.org/doi/abs/10.1145/3394171.3413804  
AUTHORS: Dingquan Li, Tingting Jiang, Ming Jiang  
HIGHLIGHT: Therefore, we explore normalization in the design of loss functions for IQA.
- 89, TITLE: Context-aware Attention Network for Predicting Image Aesthetic Subjectivity  
https://dl.acm.org/doi/abs/10.1145/3394171.3413834  
AUTHORS: Munan Xu, Jia-Xing Zhong, Yurui Ren, Shan Liu, Ge Li  
HIGHLIGHT: Based on the attention model, we predict the distribution of human aesthetic ratings of images, which reflects the diversity and similarity of human subjective opinions.

- 90, TITLE: Scoring High: Analysis and Prediction of Viewer Behavior and Engagement in the Context of 2018 FIFA WC Live Streaming  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414016>  
AUTHORS: Nikolas Wehner, Michael Seufert, Sebastian Egger-Lampl, Bruno Gardlo, Pedro Casas, Raimund Schatz  
HIGHLIGHT: In this paper, we analyze a unique dataset consisting of more than a million 2018 FIFA World Cup mobile live streaming sessions, collected at a large national public broadcaster.
- 91, TITLE: Object-level Attention for Aesthetic Rating Distribution Prediction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413695>  
AUTHORS: Jingwen Hou, Sheng Yang, Weisi Lin  
HIGHLIGHT: We study the problem of image aesthetic assessment (IAA) and aim to automatically predict the image aesthetic quality in the form of discrete distribution, which is particularly important in IAA due to its nature of having possibly higher diversification of agreement for aesthetics.
- 92, TITLE: ARSketch: Sketch-Based User Interface for Augmented Reality Glasses  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413633>  
AUTHORS: Zhaohui Zhang, Haichao Zhu, Qian Zhang  
HIGHLIGHT: To tackle this problem, we introduce a sketch-based neural network-driven user interface for AR/MR glasses, called ARSketch, which enables drawing sketches freely in air to interact with the devices.
- 93, TITLE: RIRNet: Recurrent-In-Recurrent Network for Video Quality Assessment  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413717>  
AUTHORS: Pengfei Chen, Leida Li, Lei Ma, Jinjian Wu, Guangming Shi  
HIGHLIGHT: Partially inspired by psychophysical and vision science studies revealing the speed tuning property of neurons in visual cortex when performing motion perception (i.e., sensitive to different temporal frequencies), we propose a novel no-reference (NR) VQA framework named Recurrent-In-Recurrent Network (RIRNet) to incorporate this characteristic to prompt an accurate representation of motion perception in VQA task.
- 94, TITLE: Cognitive Representation Learning of Self-Media Online Article Quality  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413747>  
AUTHORS: Yiru Wang, Shen Huang, Gongfu Li, Qiang Deng, Dongliang Liao, Pengda Si, Yujun Yang, Jin Xu  
HIGHLIGHT: To solve these challenges, we establish a joint model CoQAN in combination with the layout organization, writing characteristics and text semantics, designing different representation learning subnetworks, especially for the feature learning process and interactive reading habits on mobile terminals.
- 95, TITLE: Describing Subjective Experiment Consistency by p-Value P--P Plot  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413749>  
AUTHORS: Jakub Nawala, Lucjan Janowski, Bogdan Cmiel, Krzysztof Rusek  
HIGHLIGHT: We provide a tool to classify subjective experiment (and all its results) as either consistent or inconsistent.
- 96, TITLE: Increasing Video Perceptual Quality with GANs and Semantic Coding  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413508>  
AUTHORS: Leonardo Galteri, Marco Bertini, Lorenzo Seidenari, Tiberio Uricchio, Alberto Del Bimbo  
HIGHLIGHT: In this work we show how such videos can be efficiently generated by shifting bitrate with masks derived via computer vision and how a deep generative adversarial network can be trained to restore video quality.
- 97, TITLE: Label Embedding Online Hashing for Cross-Modal Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413971>  
AUTHORS: Yongxin Wang, Xin Luo, Xin-Shun Xu  
HIGHLIGHT: To address these issues, in this paper, we propose a novel supervised online cross-modal hashing method, i.e., Label Embedding ONLINE hashing, LEMON for short.
- 98, TITLE: Quaternion-Based Knowledge Graph Network for Recommendation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413992>  
AUTHORS: Zhaopeng Li, Qianqian Xu, Yangbanyan Jiang, Xiaochun Cao, Qingming Huang  
HIGHLIGHT: In this paper, we propose Quaternion-based Knowledge Graph Network (QKGN) for recommendation, which represents users and items with quaternion embeddings in hypercomplex space, so that the latent inter-dependencies between entities and relations could be captured effectively.
- 99, TITLE: Class-Aware Modality Mix and Center-Guided Metric Learning for Visible-Thermal Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413821>

AUTHORS: Yongguo Ling, Zhun Zhong, Zhiming Luo, Paolo Rota, Shaozi Li, Nicu Sebe  
HIGHLIGHT: In this paper, we design a novel framework to jointly bridge the modality gap in pixel- and feature-level without additional parameters, as well as reduce the inter- and intra-modalities variations by a center-guided metric learning constraint.

100, TITLE: Adversarial Video Moment Retrieval by Jointly Modeling Ranking and Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413841>  
AUTHORS: Da Cao, Yawen Zeng, Xiaochi Wei, Liqiang Nie, Richang Hong, Zheng Qin  
HIGHLIGHT: Toward this end, we contribute a novel solution to thoroughly investigate the video moment retrieval issue under the adversarial learning paradigm.

101, TITLE: Beyond the Parts: Learning Multi-view Cross-part Correlation for Vehicle Re-identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413578>  
AUTHORS: Xinchun Liu, Wu Liu, Jinkai Zheng, Chenggang Yan, Tao Mei  
HIGHLIGHT: Different from existing methods, we propose a Parsing-guided Cross-part Reasoning Network, named as PCRNet, for vehicle Re-Id.

102, TITLE: Weakly-Supervised Image Hashing through Masked Visual-Semantic Graph-based Reasoning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414022>  
AUTHORS: Lu Jin, Zechao Li, Yonghua Pan, Jinhui Tang  
HIGHLIGHT: To address this issue, this work proposes a novel Masked visual-semantic Graph-based Reasoning Network, termed as MGRN, to learn joint visual-semantic representations for image hashing.

103, TITLE: Semantic Consistency Guided Instance Feature Alignment for 2D Image-Based 3D Shape Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413631>  
AUTHORS: Heyu Zhou, Weizhi Nie, Dan Song, Nian Hu, Xuanya Li, An-An Liu  
HIGHLIGHT: In this paper, we propose a novel semantic consistency guided instance feature alignment network (SC-IFA) to address these limitations.

104, TITLE: RGB2LIDAR: Towards Solving Large-Scale Cross-Modal Visual Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413647>  
AUTHORS: Niluthpol Chowdhury Mithun, Karan Sikka, Han-Pang Chiu, Supun Samarasekera, Rakesh Kumar  
HIGHLIGHT: To enable large-scale evaluation, we introduce a new dataset containing over 550K pairs (covering 143 km<sup>2</sup> area) of RGB and aerial LIDAR depth images.

105, TITLE: Performance Optimization of Federated Person Re-identification via Benchmark Analysis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413814>  
AUTHORS: Weiming Zhuang, Yonggang Wen, Xuesen Zhang, Xin Gan, Daiying Yin, Dongzhan Zhou, Shuai Zhang, Shuai Yi  
HIGHLIGHT: In this work, we implement federated learning to person re-identification (FedReID) and optimize its performance affected by statistical heterogeneity in the real-world scenario.

106, TITLE: Traffic-Aware Multi-Camera Tracking of Vehicles Based on ReID and Camera Link Model  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413863>  
AUTHORS: Hung-Min Hsu, Yizhou Wang, Jenq-Neng Hwang  
HIGHLIGHT: In this paper, we propose an effective and reliable MTMCT framework for vehicles, which consists of a traffic-aware single camera tracking (TSC) algorithm, a trajectory-based camera link model (CLM) for vehicle re-identification (ReID), and a hierarchical clustering algorithm to obtain the cross camera vehicle trajectories.

107, TITLE: Active Object Search  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413571>  
AUTHORS: Jie Wu, Tianshui Chen, Lishan Huang, Hefeng Wu, Guanbin Li, Ling Tian, Liang Lin  
HIGHLIGHT: In this work, we investigate an Active Object Search (AOS) task that is not explicitly addressed in the literature.

108, TITLE: An Analysis of Delay in Live 360deg; Video Streaming Systems  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413539>  
AUTHORS: Jun Yi, Md Reazul Islam, Shivang Aggarwal, Dimitrios Koutsonikolas, Y. Charlie Hu, Zhisheng Yan  
HIGHLIGHT: In this paper, we conduct the first in-depth measurement study of task-level time consumption for five system components in live 360deg; video streaming.

109, TITLE: DeepFacePencil: Creating Face Images from Freehand Sketches

<https://dl.acm.org/doi/abs/10.1145/3394171.3413684>

AUTHORS: Yuhang Li, Xuejin Chen, Binxin Yang, Zihan Chen, Zhihua Cheng, Zheng-Jun Zha  
HIGHLIGHT: In this paper, we explore the task of generating photo-realistic face images from hand-drawn sketches.

110, TITLE: When Bitstream Prior Meets Deep Prior: Compressed Video Super-resolution with Learning from Decoding  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413504>

AUTHORS: Peilin Chen, Wenhan Yang, Long Sun, Shiqi Wang  
HIGHLIGHT: In this paper, we systematically investigate the SR of compressed LR videos by leveraging the interactivity between decoding prior and deep prior.

111, TITLE: RL-B&eacute;l&aacute;dy: A Unified Learning Framework for Content Caching

<https://dl.acm.org/doi/abs/10.1145/3394171.3413524>

AUTHORS: Gang Yan, Jian Li  
HIGHLIGHT: This paper instead proposes a novel framework that can simultaneously learn both content admission and content eviction for caching in CDNs.

112, TITLE: ShapeCaptioner: Generative Caption Network for 3D Shapes by Learning a Mapping from Parts Detected in Multiple Views to Sentences

<https://dl.acm.org/doi/abs/10.1145/3394171.3413889>

AUTHORS: Zhizhong Han, Chao Chen, Yu-Shen Liu, Matthias Zwicker  
HIGHLIGHT: To resolve this issue, we propose ShapeCaptioner, a generative caption network, to perform 3D shape captioning from semantic parts detected in multiple views.

113, TITLE: Co-Attentive Lifting for Infrared-Visible Person Re-Identification

<https://dl.acm.org/doi/abs/10.1145/3394171.3413933>

AUTHORS: Xing Wei, Diangang Li, Xiaopeng Hong, Wei Ke, Yihong Gong  
HIGHLIGHT: This paper proposes a novel attention-based approach to handle the two difficulties in a unified framework.

114, TITLE: Multimodal Representation with Embedded Visual Guiding Objects for Named Entity Recognition in Social Media Posts

<https://dl.acm.org/doi/abs/10.1145/3394171.3413650>

AUTHORS: Zhiwei Wu, Changmeng Zheng, Yi Cai, Junying Chen, Ho-fung Leung, Qing Li  
HIGHLIGHT: In this paper, we propose a neural network which combines object-level image information and character-level text information to predict entities.

115, TITLE: Context-Aware Multi-View Summarization Network for Image-Text Matching

<https://dl.acm.org/doi/abs/10.1145/3394171.3413961>

AUTHORS: Leigang Qu, Meng Liu, Da Cao, Liqiang Nie, Qi Tian  
HIGHLIGHT: Toward this end, we present a novel context-aware multi-view summarization network to summarize context-enhanced visual region information from multiple views.

116, TITLE: Performance over Random: A Robust Evaluation Protocol for Video Summarization Methods

<https://dl.acm.org/doi/abs/10.1145/3394171.3413632>

AUTHORS: Evlampios Apostolidis, Eleni Adamantidou, Alexandros I. Metsai, Vasileios Mezaris, Ioannis Patras  
HIGHLIGHT: This paper proposes a new evaluation approach for video summarization algorithms.

117, TITLE: Concept Drift Detection for Multivariate Data Streams and Temporal Segmentation of Daylong Egocentric Videos

<https://dl.acm.org/doi/abs/10.1145/3394171.3413713>

AUTHORS: Pravin Nagar, Mansi Khemka, Chetan Arora  
HIGHLIGHT: In this paper, we present a novel unsupervised temporal segmentation technique especially suited for day-long egocentric videos.

118, TITLE: Distributed Multi-agent Video Fast-forwarding

<https://dl.acm.org/doi/abs/10.1145/3394171.3413767>

AUTHORS: Shuyue Lan, Zhilu Wang, Amit K. Roy-Chowdhury, Ermin Wei, Qi Zhu  
HIGHLIGHT: This paper presents a consensus-based distributed multi-agent video fast-forwarding framework, named DMVF, that fast-forwards multi-view video streams collaboratively and adaptively.

119, TITLE: Controllable Video Captioning with an Exemplar Sentence

- <https://dl.acm.org/doi/abs/10.1145/3394171.3413908>  
AUTHORS: Yitian Yuan, Lin Ma, Jingwen Wang, Wenwu Zhu  
HIGHLIGHT: In this paper, we investigate a novel and challenging task, namely controllable video captioning with an exemplar sentence.
- 120, TITLE: MMFL: Multimodal Fusion Learning for Text-Guided Image Inpainting  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413982>  
AUTHORS: Qing Lin, Bo Yan, Jichun Li, Weimin Tan  
HIGHLIGHT: We propose a multimodal fusion learning method for image inpainting (MMFL).
- 121, TITLE: Vision Meets Wireless Positioning: Effective Person Re-identification with Recurrent Context Propagation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413984>  
AUTHORS: Yiheng Liu, Wengang Zhou, Mao Xi, Sanjing Shen, Houqiang Li  
HIGHLIGHT: In this work, we approach person re-identification with the sensing data from both vision and wireless positioning.
- 122, TITLE: Structural Semantic Adversarial Active Learning for Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413885>  
AUTHORS: Beichen Zhang, Liang Li, Li Su, Shuhui Wang, Jincan Deng, Zheng-Jun Zha, Qingming Huang  
HIGHLIGHT: To solve this problem, we propose a structural semantic adversarial active learning (SSAAL) model that leverages both visual and textual information for deriving the most representative samples while maximizing the image captioning performance.
- 123, TITLE: MISA: Modality-Invariant and -Specific Representations for Multimodal Sentiment Analysis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413678>  
AUTHORS: Devamanyu Hazarika, Roger Zimmermann, Soujanya Poria  
HIGHLIGHT: In this paper, we aim to learn effective modality representations to aid the process of fusion.
- 124, TITLE: Multi-modal Cooking Workflow Construction for Food Recipes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413765>  
AUTHORS: Liang-Ming Pan, Jingjing Chen, Jianlong Wu, Shaoteng Liu, Chong-Wah Ngo, Min-Yen Kan, Yugang Jiang, Tat-Seng Chua  
HIGHLIGHT: In this paper, we build MM-ReS, the first large-scale dataset for cooking workflow construction, consisting of 9,850 recipes with human-labeled workflow graphs.
- 125, TITLE: Depth Guided Adaptive Meta-Fusion Network for Few-shot Video Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413502>  
AUTHORS: Yuqian Fu, Li Zhang, Junke Wang, Yanwei Fu, Yu-Gang Jiang  
HIGHLIGHT: In this paper, we propose a depth guided Adaptive Meta-Fusion Network for few-shot video recognition which is termed as AMeFu-Net.
- 126, TITLE: Adaptive Temporal Triplet-loss for Cross-modal Embedding Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413540>  
AUTHORS: David Semedo, Jo&atilde;o Magalh&atilde;es  
HIGHLIGHT: In this work, we seek for highly expressive loss functions that allow the encoding of data temporal traits into cross-modal embedding spaces.
- 127, TITLE: Scene-Aware Background Music Synthesis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413894>  
AUTHORS: Yujia Wang, Wei Liang, Wanwan Li, Dingzeyu Li, Lap-Fai Yu  
HIGHLIGHT: In this paper, we introduce an interactive background music synthesis algorithm guided by visual content.
- 128, TITLE: Deep-Modal: Real-Time Impact Sound Synthesis for Arbitrary Shapes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413572>  
AUTHORS: Xutong Jin, Sheng Li, Tianshu Qu, Dinesh Manocha, Guoping Wang  
HIGHLIGHT: We present a novel learning-based impact sound synthesis algorithm called Deep-Modal.
- 129, TITLE: Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Compositions  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413671>  
AUTHORS: Yu-Siang Huang, Yi-Hsuan Yang

**HIGHLIGHT:** In contrast with this general approach, this paper shows that Transformers can do even better for music modeling, when we improve the way a musical score is converted into the data fed to a Transformer model.

130, **TITLE:** Make Your Favorite Music Curative: Music Style Transfer for Anxiety Reduction

<https://dl.acm.org/doi/abs/10.1145/3394171.3414070>

**AUTHORS:** Zhejing Hu, Yan Liu, Gong Chen, Sheng-hua Zhong, Aiwei Zhang

**HIGHLIGHT:** This paper proposes a novel style transfer model to generate the therapeutic music according to user's preference.

131, **TITLE:** PopMAG: Pop Music Accompaniment Generation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413721>

**AUTHORS:** Yi Ren, Jinzheng He, Xu Tan, Tao Qin, Zhou Zhao, Tie-Yan Liu

**HIGHLIGHT:** To improve harmony, in this paper, we propose a novel MULTI-track MIDI representation (MuMIDI), which enables simultaneous multi-track generation in a single sequence and explicitly models the dependency of the notes from different tracks.

132, **TITLE:** DeepSonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices

<https://dl.acm.org/doi/abs/10.1145/3394171.3413716>

**AUTHORS:** Run Wang, Felix Juefei-Xu, Yihao Huang, Qing Guo, Xiaofei Xie, Lei Ma, Yang Liu

**HIGHLIGHT:** In this paper, we devise a novel approach, named DeepSonar, based on monitoring neuron behaviors of speaker recognition (SR) system, i.e., a deep neural network (DNN), to discern AI-synthesized fake voices.

133, **TITLE:** FakePolisher: Making DeepFakes More Detection-Evasive by Shallow Reconstruction

<https://dl.acm.org/doi/abs/10.1145/3394171.3413732>

**AUTHORS:** Yihao Huang, Felix Juefei-Xu, Run Wang, Qing Guo, Lei Ma, Xiaofei Xie, Jianwen Li, Weikai Miao, Yang Liu, Geguang Pu

**HIGHLIGHT:** Towards reducing the artifacts in the synthesized images, in this paper, we devise a simple yet powerful approach termed FakePolisher that performs shallow reconstruction of fake images through a learned linear dictionary, intending to effectively and efficiently reduce the artifacts introduced during image synthesis.

134, **TITLE:** Boosting Visual Question Answering with Context-aware Knowledge Aggregation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413943>

**AUTHORS:** Guohao Li, Xin Wang, Wenwu Zhu

**HIGHLIGHT:** To solve the challenging issue, we propose a Knowledge Graph Augmented (KG-Aug) model which conducts context-aware knowledge aggregation on external knowledge graphs, requiring no ground-truth knowledge facts for extra supervision.

135, **TITLE:** Memory-Augmented Relation Network for Few-Shot Learning

<https://dl.acm.org/doi/abs/10.1145/3394171.3413811>

**AUTHORS:** Jun He, Richang Hong, Xueliang Liu, Mingliang Xu, Zheng-Jun Zha, Meng Wang

**HIGHLIGHT:** We investigate a new metric-learning method to explicitly exploit these relationships.

136, **TITLE:** K-armed Bandit based Multi-Modal Network Architecture Search for Visual Question Answering

<https://dl.acm.org/doi/abs/10.1145/3394171.3413998>

**AUTHORS:** Yiyi Zhou, Rongrong Ji, Xiaoshuai Sun, Gen Luo, Xiaopeng Hong, Jinsong Su, Xinghao Ding, Ling Shao

**HIGHLIGHT:** In this paper, we propose a cross-modal network architecture search (NAS) algorithm for VQA, termed as k-Armed Bandit based NAS (KAB-NAS).

137, **TITLE:** Adversarial Graph Representation Adaptation for Cross-Domain Facial Expression Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413822>

**AUTHORS:** Yuan Xie, Tianshui Chen, Tao Pu, Hefeng Wu, Liang Lin

**HIGHLIGHT:** In this work, we propose a novel Adversarial Graph Representation Adaptation (AGRA) framework that unifies graph representation propagation with adversarial learning for cross-domain holistic-local feature co-adaptation.

138, **TITLE:** KBGN: Knowledge-Bridge Graph Network for Adaptive Vision-Text Reasoning in Visual Dialogue

<https://dl.acm.org/doi/abs/10.1145/3394171.3413826>

**AUTHORS:** Xiaoze Jiang, Siyi Du, Zengchang Qin, Yajing Sun, Jing Yu

**HIGHLIGHT:** In this paper, we propose a novel Knowledge-Bridge Graph Network (KBGN) model by using graph to bridge the cross-modal semantic relations between vision and text knowledge in fine granularity, as well as retrieving required knowledge via an adaptive information selection mode.

- 139, TITLE: Cascade Grouped Attention Network for Referring Expression Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414006>  
AUTHORS: Gen Luo, Yiyi Zhou, Rongrong Ji, Xiaoshuai Sun, Jinsong Su, Chia-Wen Lin, Qi Tian  
HIGHLIGHT: In this paper, we focus on addressing this issue by proposing a Cascade Grouped Attention Network (CGAN) with two innovative designs: Cascade Grouped Attention (CGA) and Instance-level Attention (ILA) loss.
- 140, TITLE: Reinforcement Learning for Weakly Supervised Temporal Grounding of Natural Language in Untrimmed Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413862>  
AUTHORS: Jie Wu, Guanbin Li, Xiaoguang Han, Liang Lin  
HIGHLIGHT: In this paper, we propose a Boundary Adaptive Refinement (BAR) framework that resorts to reinforcement learning (RL) to guide the process of progressively refining the temporal boundary.
- 141, TITLE: Poet: Product-oriented Video Captioner for E-commerce  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413880>  
AUTHORS: Shengyu Zhang, Ziqi Tan, Jin Yu, Zhou Zhao, Kun Kuang, Jie Liu, Jingren Zhou, Hongxia Yang, Fei Wu  
HIGHLIGHT: To address this problem, we propose a product-oriented video captioner framework, abbreviated as Poet.
- 142, TITLE: Text-Guided Neural Image Inpainting  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414017>  
AUTHORS: Lisai Zhang, Qingcai Chen, Baotian Hu, Shuoran Jiang  
HIGHLIGHT: The goal of this paper is to fill the semantic information in corrupted images according to the provided descriptive text.
- 143, TITLE: Single-Shot Two-Pronged Detector with Rectified IoU Loss  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413691>  
AUTHORS: Keyang Wang, Lei Zhang  
HIGHLIGHT: In this paper, we introduce a novel two-pronged transductive idea to explore the relationship among different layers in both backward and forward directions, which can enrich the semantic information of low-level features and detailed information of high-level features at the same time.
- 144, TITLE: Dynamic Context-guided Capsule Network for Multimodal Machine Translation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413715>  
AUTHORS: Huan Lin, Fandong Meng, Jinsong Su, Yongjing Yin, Zhengyuan Yang, Yubin Ge, Jie Zhou, Jiebo Luo  
HIGHLIGHT: To address the above issues, in this paper, we propose a novel Dynamic Context-guided Capsule Network (DCCN) for MMT.
- 145, TITLE: Differentiable Manifold Reconstruction for Point Cloud Denoising  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413727>  
AUTHORS: Shitong Luo, Wei Hu  
HIGHLIGHT: To this end, we propose to learn the underlying manifold of a noisy point cloud from differentially subsampled points with trivial noise perturbation and their embedded neighborhood feature, aiming to capture intrinsic structures in point clouds.
- 146, TITLE: BS-MCVR: Binary-sensing based Mobile-cloud Visual Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413500>  
AUTHORS: Hongyi Zheng, Wangmeng Zuo, Lei Zhang  
HIGHLIGHT: In this work, we present a machine-perception-oriented MCVR system, called BS-MCVR, where the mobile end is designed to efficiently sense highly compact and discriminative features directly from the scene, and the sensed features are analyzed on the cloud for recognition.
- 147, TITLE: Learning Modality-Invariant Latent Representations for Generalized Zero-shot Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413503>  
AUTHORS: Jingjing Li, Mengmeng Jing, Lei Zhu, Zhengming Ding, Ke Lu, Yang Yang  
HIGHLIGHT: In this paper, therefore, we present a new method which can simultaneously generate both visual representations and semantic representations so that the essential multi-modal information associated with unseen classes can be captured.
- 148, TITLE: Describe What to Change: A Text-guided Unsupervised Image-to-image Translation Approach  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413505>  
AUTHORS: Yahui Liu, Marco De Nadai, Deng Cai, Huayang Li, Xavier Alameda-Pineda, Nicu Sebe, Bruno Lepri

**HIGHLIGHT:** In this work, we propose a novel unsupervised approach, based on image-to-image translation, that alters the attributes of a given image through a command-like sentence such as “change the hair color to black”.

149, **TITLE:** INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413528>  
**AUTHORS:** Advaith Sridhar, Rohith Gandhi Ganesan, Pratyush Kumar, Mitesh Khapra  
**HIGHLIGHT:** In this work, we present the Indian Lexicon Sign Language Dataset - INCLUDE - an ISL dataset that contains 0.27 million frames across 4,287 videos over 263 word signs from 15 different word categories.

150, **TITLE:** Amora: Black-box Adversarial Morphing Attack  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413544>  
**AUTHORS:** Run Wang, Felix Juefei-Xu, Qing Guo, Yihao Huang, Xiaofei Xie, Lei Ma, Yang Liu  
**HIGHLIGHT:** In this paper, we investigate and introduce a new type of adversarial attack to evade FR systems by manipulating facial content, called adversarial morphing attack (a.k.a. Amora).

151, **TITLE:** Visual Relation of Interest Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413566>  
**AUTHORS:** Fan Yu, Haonan Wang, Tongwei Ren, Jinhui Tang, Gangshan Wu  
**HIGHLIGHT:** In this paper, we propose a novel Visual Relation of Interest Detection (VROID) task, which aims to detect visual relations that are important for conveying the main content of an image, motivated from the intuition that not all correctly detected relations are really “interesting” in semantics and only a fraction of them really make sense for representing the image main content.

152, **TITLE:** University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413896>  
**AUTHORS:** Zhedong Zheng, Yunchao Wei, Yi Yang  
**HIGHLIGHT:** To verify the effectiveness of the drone platform, we introduce a new multi-view multi-source benchmark for drone-based geo-localization, named University-1652.

153, **TITLE:** DIPDefend: Deep Image Prior Driven Defense against Adversarial Examples  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413898>  
**AUTHORS:** Tao Dai, Yan Feng, Dongxian Wu, Bin Chen, Jian Lu, Yong Jiang, Shu-Tao Xia  
**HIGHLIGHT:** Motivated by deep image prior that can capture rich image statistics from a single image, we propose an effective Deep Image Prior Driven Defense (DIPDefend) method against adversarial examples.

154, **TITLE:** TRIE: End-to-End Text Reading and Information Extraction for Document Understanding  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413900>  
**AUTHORS:** Peng Zhang, Yunlu Xu, Zhanzhan Cheng, Shiliang Pu, Jing Lu, Liang Qiao, Yi Niu, Fei Wu  
**HIGHLIGHT:** In this paper, we propose a unified end-to-end text reading and information extraction network, where the two tasks can reinforce each other.

155, **TITLE:** Adversarial Privacy-preserving Filter  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413906>  
**AUTHORS:** Jiaming Zhang, Jitao Sang, Xian Zhao, Xiaowen Huang, Yanfeng Sun, Yongli Hu  
**HIGHLIGHT:** This work aims to develop a privacy-preserving solution, called Adversarial Privacy-preserving Filter (APF), to protect the online shared face images from being maliciously used.

156, **TITLE:** Mix Dimension in Poincaré Geometry for 3D Skeleton-based Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413910>  
**AUTHORS:** Wei Peng, Jingang Shi, Zhaoqiang Xia, Guoying Zhao  
**HIGHLIGHT:** In this paper, we provide an orthogonal way to explore the underlying connections.

157, **TITLE:** Dynamic Extension Nets for Few-shot Semantic Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413915>  
**AUTHORS:** Lizhao Liu, Junyi Cao, Minqian Liu, Yong Guo, Qi Chen, Mingkui Tan  
**HIGHLIGHT:** To address the above issues, we propose a Dynamic Extension Network (DENet) in which we dynamically construct and maintain a classifier for the novel class by leveraging the knowledge from the base classes and the information from novel data.

158, **TITLE:** Fast Enhancement for Non-Uniform Illumination Images using Light-weight CNNs



<https://dl.acm.org/doi/abs/10.1145/3394171.3413925>

AUTHORS: Feifan Lv, Bo Liu, Feng Lu

HIGHLIGHT: This paper proposes a new light-weight convolutional neural network (~5k params) for non-uniform illumination image enhancement to handle color, exposure, contrast, noise and artifacts, etc., simultaneously and effectively.

159, TITLE: Animating Through Warping: An Efficient Method for High-Quality Facial Expression Animation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413926>

AUTHORS: Zili Yi, Qiang Tang, Vishnu Sanjay Ramiya Srinivasan, Zhan Xu

HIGHLIGHT: Motivated by the idea that HD images can be generated by adding high-frequency residuals to low-resolution results produced by a neural network, we propose a novel framework known as Animating Through Warping (ATW) to enable efficient animation of HD images.

160, TITLE: Exploiting Better Feature Aggregation for Video Object Detection

<https://dl.acm.org/doi/abs/10.1145/3394171.3413927>

AUTHORS: Liang Han, Pichao Wang, Zhaozheng Yin, Fan Wang, Hao Li

HIGHLIGHT: To exploit better feature aggregation, in this paper, we propose two improvements over previous works: a class-constrained spatial-temporal relation network and a correlation-based feature alignment module.

161, TITLE: NuI-Go: Recursive Non-Local Encoder-Decoder Network for Retinal Image Non-Uniform Illumination Removal

<https://dl.acm.org/doi/abs/10.1145/3394171.3413928>

AUTHORS: Chongyi Li, Huazhu Fu, Runmin Cong, Zechao Li, Qianqian Xu

HIGHLIGHT: To address this issue, we propose a non-uniform illumination removal network for retinal image, called NuI-Go, which consists of three Recursive Non-local Encoder-Decoder Residual Blocks (NEDRBs) for enhancing the degraded retinal images in a progressive manner.

162, TITLE: Online Filtering Training Samples for Robust Visual Tracking

<https://dl.acm.org/doi/abs/10.1145/3394171.3413930>

AUTHORS: Jie Zhao, Kenan Dai, Dong Wang, Huchuan Lu, Xiaoyun Yang

HIGHLIGHT: In this paper, we propose an optimization module named MetricNet for online filtering training samples.

163, TITLE: Boosting Continuous Sign Language Recognition via Cross Modality Augmentation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413931>

AUTHORS: Junfu Pu, Wengang Zhou, Hezhen Hu, Houqiang Li

HIGHLIGHT: To tackle this issue, we propose a novel architecture with cross modality augmentation.

164, TITLE: ThumbNet: One Thumbnail Image Contains All You Need for Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413937>

AUTHORS: Chen Zhao, Bernard Ghanem

HIGHLIGHT: Based on the fact that input images of a CNN contain substantial redundancy, in this paper, we propose a unified framework, dubbed as ThumbNet, to simultaneously accelerate and compress CNN models by enabling them to infer on one thumbnail image.

165, TITLE: Dual Temporal Memory Network for Efficient Video Object Segmentation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413942>

AUTHORS: Kaihua Zhang, Long Wang, Dong Liu, Bo Liu, Qingshan Liu, Zhu Li

HIGHLIGHT: We present an end-to-end network which stores short- and long-term video sequence information preceding the current frame as the temporal memories to address the temporal modeling in VOS.

166, TITLE: Cooperative Bi-path Metric for Few-shot Learning

<https://dl.acm.org/doi/abs/10.1145/3394171.3413946>

AUTHORS: Zeyuan Wang, Yifan Zhao, Jia Li, Yonghong Tian

HIGHLIGHT: In this paper, we make two contributions to investigate the few-shot classification problem.

167, TITLE: From Design Draft to Real Attire: Unaligned Fashion Image Translation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413953>

AUTHORS: Yu Han, Shuai Yang, Wenjing Wang, Jiaying Liu

HIGHLIGHT: In this paper, we study a new unaligned translation problem between design drafts and real fashion items, whose main challenge lies in the huge misalignment between the two modalities.

- 168, TITLE: Siamese Attentive Graph Tracking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413956>  
AUTHORS: Fei Zhao, Ting Zhang, Chao Ma, Ming Tang, Jinqiao Wang, Xiaobo Wang  
HIGHLIGHT: In this paper, we propose to advance Siamese trackers with graph convolutional networks, which pay more attention to the structural layout of target objects, to learn features robust to large appearance changes over time.
- 169, TITLE: HiFaceGAN: Face Renovation via Collaborative Suppression and Replenishment  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413965>  
AUTHORS: Lingbo Yang, Shanshe Wang, Siwei Ma, Wen Gao, Chang Liu, Pan Wang, Peiran Ren  
HIGHLIGHT: In this paper, we investigate a more challenging and practical &quot;dual-blind&quot; version of the problem by lifting the requirements on both types of prior, termed as &quot;Face Renovation&quot;(FR).
- 170, TITLE: Discernible Image Compression  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413968>  
AUTHORS: Zhaohui Yang, Yunhe Wang, Chang Xu, Peng Du, Chao Xu, Chunjing Xu, Qi Tian  
HIGHLIGHT: Based on the encoder-decoder framework, we propose using a pre-trained CNN to extract features of the original and compressed images, and making them similar.
- 171, TITLE: Forest R-CNN: Large-Vocabulary Long-Tailed Object Detection and Instance Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413970>  
AUTHORS: Jialian Wu, Liangchen Song, Tiancai Wang, Qian Zhang, Junsong Yuan  
HIGHLIGHT: To alleviate the imbalanced learning caused by the long-tail phenomena, we propose a simple yet effective resampling method, NMS Resampling, to re-balance the data distribution.
- 172, TITLE: Adv-watermark: A Novel Watermark Perturbation for Adversarial Examples  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413976>  
AUTHORS: Xiaojun Jia, Xingxing Wei, Xiaochun Cao, Xiaoguang Han  
HIGHLIGHT: In this paper, we propose a novel watermark perturbation for adversarial examples (Adv-watermark) which combines image watermarking techniques and adversarial example algorithms.
- 173, TITLE: Dual In-painting Model for Unsupervised Gaze Correction and Animation in the Wild  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413981>  
AUTHORS: Jichao Zhang, Jingjing Chen, Hao Tang, Wei Wang, Yan Yan, Enver Sangineto, Nicu Sebe  
HIGHLIGHT: We address the problem of unsupervised gaze correction in the wild, presenting a solution that works without the need of precise annotations of the gaze angle and the head pose.
- 174, TITLE: Learning Hierarchical Graph for Occluded Pedestrian Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413983>  
AUTHORS: Gang Li, Jian Li, Shanshan Zhang, Jian Yang  
HIGHLIGHT: In this paper, we propose a novel Hierarchical Graph Pedestrian Detector (HGPD), which integrates semantic and spatial relation information to construct two graphs named intra-proposal graph and inter-proposal graph, without relying on extra cues w.r.t visible regions.
- 175, TITLE: Adaptively-Accumulated Knowledge Transfer for Partial Domain Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413986>  
AUTHORS: Taotao Jing, Haifeng Xia, Zhengming Ding  
HIGHLIGHT: In this work, we propose an Adaptively-Accumulated Knowledge Transfer framework (A<sup>2</sup>SKT) to align the relevant categories across two domains for effective domain adaptation.
- 176, TITLE: Box Guided Convolution for Pedestrian Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413989>  
AUTHORS: Jinpeng Li, Shengcai Liao, Hangzhi Jiang, Ling Shao  
HIGHLIGHT: In particular, we proposed a box guided convolution (BGC) that can dynamically adjust the sizes of convolution kernels guided by the predicted bounding boxes.
- 177, TITLE: Stronger, Faster and More Explainable: A Graph Convolutional Baseline for Skeleton-based Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413802>  
AUTHORS: Yi-Fan Song, Zhang Zhang, Caifeng Shan, Liang Wang

**HIGHLIGHT:** In this work, we propose an efficient but strong baseline based on Graph Convolutional Network (GCN), where three main improvements are aggregated, i.e., early fused Multiple Input Branches (MIB), Residual GCN (ResGCN) with bottleneck structure and Part-wise Attention (PartAtt) block.

178, **TITLE:** Adversarial Image Attacks Using Multi-Sample and Most-Likely Ensemble Methods  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413808>

**AUTHORS:** Xia Du, Chi-Man Pun

**HIGHLIGHT:** In this paper, we propose the multi-sample ensemble method (MSEM) and most-likely ensemble method (MLEM) to generate adversarial attacks that successfully fool the classifier for images in both the digital and real worlds.

179, **TITLE:** DCSFN: Deep Cross-scale Fusion Network for Single Image Rain Removal  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413820>

**AUTHORS:** Cong Wang, Xiaoying Xing, Yutong Wu, Zhixun Su, Junyang Chen

**HIGHLIGHT:** In this paper, we explore the cross-scale manner between networks and inner-scale fusion operation to solve the image rain removal task.

180, **TITLE:** Self-Paced Video Data Augmentation by Generative Adversarial Networks with Insufficient Samples  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414003>

**AUTHORS:** Yumeng Zhang, Gaoguo Jia, Li Chen, Mingrui Zhang, Junhai Yong

**HIGHLIGHT:** In this study, we propose a generative data augmentation method for video classification using dynamic images.

181, **TITLE:** CF-SIS: Semantic-Instance Segmentation of 3D Point Clouds by Context Fusion with Self-Attention  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413829>

**AUTHORS:** Xin Wen, Zhizhong Han, Geunhyuk Youk, Yu-Shen Liu

**HIGHLIGHT:** To address the above two issues, we propose a novel network of feature context fusion for SIS task, named CF-SIS.

182, **TITLE:** Hybrid Resolution Network Using Edge Guided Region Mutual Information Loss for Human Parsing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413831>

**AUTHORS:** Yunan Liu, Liang Zhao, Shanshan Zhang, Jian Yang

**HIGHLIGHT:** In this paper, we propose a new method for human parsing, which effectively maintains high-resolution representations and leverages body edge details to improve the performance.

183, **TITLE:** Meta-RCNN: Meta Learning for Few-Shot Object Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413832>

**AUTHORS:** Xiongwei Wu, Doyen Sahoo, Steven Hoi

**HIGHLIGHT:** In this paper, we investigate the problem of few-shot object detection, where a detector has access to only limited amounts of annotated data.

184, **TITLE:** Objectness Consistent Representation for Weakly Supervised Object Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413835>

**AUTHORS:** Ke Yang, Peng Zhang, Peng Qiao, Zhiyuan Wang, Dongsheng Li, Yong Dou

**HIGHLIGHT:** In this paper, we propose a novel object representation named Objectness Consistent Representation (OCRepr) to meet the consistency criterion of objectness.

185, **TITLE:** Unpaired Image Enhancement with Quality-Attention Generative Adversarial Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413839>

**AUTHORS:** Zhangkai Ni, Wenhan Yang, Shiqi Wang, Lin Ma, Sam Kwong

**HIGHLIGHT:** In this work, we aim to learn an unpaired image enhancement model, which can enrich low-quality images with the characteristics of high-quality images provided by users.

186, **TITLE:** ASTA-Net: Adaptive Spatio-Temporal Attention Network for Person Re-Identification in Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413843>

**AUTHORS:** Xierong Zhu, Jiawei Liu, Haoze Wu, Meng Wang, Zheng-Jun Zha

**HIGHLIGHT:** In this work, we propose a novel Adaptive Spatio-Temporal Attention Network (ASTA-Net) to adaptively aggregate the spatial and temporal attention features into discriminative pedestrian representation for person re-identification in videos.

187, **TITLE:** Talking Face Generation with Expression-Tailored Generative Adversarial Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413844>

- AUTHORS: Dan Zeng, Han Liu, Hui Lin, Shiming Ge  
HIGHLIGHT: In this paper, we propose an end-to-end Expression-Tailored Generative Adversarial Network (ET-GAN) to generate an expression enriched talking face video of arbitrary identity.
- 188, TITLE: Cross-Modal Omni Interaction Modeling for Phrase Grounding  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413846>  
AUTHORS: Tianyu Yu, Tianrui Hui, Zhihao Yu, Yue Liao, Sansi Yu, Faxi Zhang, Si Liu  
HIGHLIGHT: In this paper, we propose a novel Cross-Modal Omni Interaction network (COI Net) composed of a neighboring interaction module, a global interaction module, a cross-modal interaction module and a multilevel alignment module.
- 189, TITLE: Bridging the Web Data and Fine-Grained Visual Recognition via Alleviating Label Noise and Domain Mismatch  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413851>  
AUTHORS: Yazhou Yao, Xiansheng Hua, Guanyu Gao, Zeren Sun, Zhibin Li, Jian Zhang  
HIGHLIGHT: Specifically, we propose an end-to-end deep denoising network (DDN) model to jointly solve these problems in the process of web images selection.
- 190, TITLE: Is Depth Really Necessary for Salient Object Detection?  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413855>  
AUTHORS: Jiawei Zhao, Yifan Zhao, Jia Li, Xiaowu Chen  
HIGHLIGHT: Taking the advantages of RGB and RGBD methods, we propose a novel depth-aware salient object detection framework, which has following superior designs: 1) It does not rely on depth data in the testing phase.
- 191, TITLE: Self-Play Reinforcement Learning for Fast Image Retargeting  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413857>  
AUTHORS: Nobukatsu Kajiura, Satoshi Kosugi, Xueting Wang, Toshihiko Yamasaki  
HIGHLIGHT: In this study, we address image retargeting, which is a task that adjusts input images to arbitrary sizes.
- 192, TITLE: Brain-media: A Dual Conditioned and Lateralization Supported GAN (DCLS-GAN) towards Visualization of Image-evoked Brain Activities  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413858>  
AUTHORS: Ahmed Fares, Sheng-hua Zhong, Jianmin Jiang  
HIGHLIGHT: To ensure that such extracted multimedia elements remain meaningful, we introduce a dually conditioned learning technique in the proposed deep framework, where one condition is analyzing EEGs through deep learning to extract a class-dependent and more compact brain feature space utilizing the distinctive characteristics of hemispheric lateralization and brain stimulation, and the other is to extract expressive visual features assisting our automated analysis of brain activities as well as their visualizations aided by artificial intelligence.
- 193, TITLE: Mesh Guided One-shot Face Reenactment Using Graph Convolutional Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413865>  
AUTHORS: Guangming Yao, Yi Yuan, Tianjia Shao, Kun Zhou  
HIGHLIGHT: In this paper, we introduce a method for one-shot face reenactment, which uses the reconstructed 3D meshes (i.e., the source mesh and driving mesh) as guidance to learn the optical flow needed for the reenacted face synthesis.
- 194, TITLE: Controllable Continuous Gaze Redirection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413868>  
AUTHORS: Weihao Xia, Yujie Yang, Jing-Hao Xue, Wensen Feng  
HIGHLIGHT: In this work, we present interpGaze, a novel framework for controllable gaze redirection that achieves both precise redirection and continuous interpolation.
- 195, TITLE: Preserving Global and Local Temporal Consistency for Arbitrary Video Style Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413872>  
AUTHORS: Xinxiao Wu, Jialu Chen  
HIGHLIGHT: In this paper, we propose a novel fast method that explores both global and local temporal consistency for video style transfer without estimating optical flow.
- 196, TITLE: Deep Shapely Portraits  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413873>  
AUTHORS: Qinjie Xiao, Xiangjun Tang, You Wu, Leyang Jin, Yong-Liang Yang, Xiaogang Jin  
HIGHLIGHT: We present deep shapely portraits, a novel method based on deep learning, to automatically reshape an input portrait to be better proportioned and more shapely while keeping personal facial characteristics.

- 197, TITLE: Depth Super-Resolution via Deep Controllable Slicing Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413874>  
AUTHORS: Xinchun Ye, Baoli Sun, Zhihui Wang, Jingyu Yang, Rui Xu, Haojie Li, Baopu Li  
HIGHLIGHT: To alleviate these problems, we propose a deep controllable slicing network from a novel perspective.
- 198, TITLE: Efficient Joint Gradient Based Attack Against SOR Defense for 3D Point Cloud Classification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413875>  
AUTHORS: Chengcheng Ma, Weiliang Meng, Baoyuan Wu, Shibiao Xu, Xiaopeng Zhang  
HIGHLIGHT: In this paper, we propose a novel white-box attack method, Joint Gradient Based Attack (JGBA), aiming to break the SOR defense.
- 199, TITLE: Discrete Haze Level Dehazing Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413876>  
AUTHORS: Xiaofeng Cong, Jie Gui, Kai-Chao Miao, Jun Zhang, Bing Wang, Peng Chen  
HIGHLIGHT: In this paper, a Discrete Haze Level Dehazing network (DHL-Dehaze), a very effective method to dehaze multiple different haze level images, is proposed.
- 200, TITLE: Deep Heterogeneous Multi-Task Metric Learning for Visual Recognition and Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413574>  
AUTHORS: Shikang Gan, Yong Luo, Yonggang Wen, Tongliang Liu, Han Hu  
HIGHLIGHT: To overcome this drawback, we propose a deep heterogeneous MTML (DHMTML) method, in which a nonlinear mapping is learned for each task by using a deep neural network.
- 201, TITLE: HOSE-Net: Higher Order Structure Embedded Network for Scene Graph Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413575>  
AUTHORS: Meng Wei, Chun Yuan, Xiaoyu Yue, Kuo Zhong  
HIGHLIGHT: Accordingly, in this paper, we propose a Higher Order Structure Embedded Network (HOSE-Net) to mitigate this issue.
- 202, TITLE: Dual Semantic Fusion Network for Video Object Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413583>  
AUTHORS: Lijian Lin, Haosheng Chen, Honglun Zhang, Jun Liang, Yu Li, Ying Shan, Hanzi Wang  
HIGHLIGHT: In this work, we propose a dual semantic fusion network (abbreviated as DSFNet) to fully exploit both frame-level and instance-level semantics in a unified fusion framework without external guidance.
- 203, TITLE: Sharp Multiple Instance Learning for DeepFake Video Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414034>  
AUTHORS: Xiaodan Li, Yining Lang, Yuefeng Chen, Xiaofeng Mao, Yuan He, Shuhui Wang, Hui Xue, Quan Lu  
HIGHLIGHT: In this paper, we introduce a new problem of partial face attack in DeepFake video, where only video-level labels are provided but not all the faces in the fake videos are manipulated.
- 204, TITLE: Learning to Detect Specular Highlights from Real-world Images  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413586>  
AUTHORS: Gang Fu, Qing Zhang, Qifeng Lin, Lei Zhu, Chunxia Xiao  
HIGHLIGHT: In this paper, we present a large-scale real-world highlight dataset containing a rich variety of material categories, with diverse highlight shapes and appearances, in which each image is with an annotated ground-truth mask.
- 205, TITLE: Video Super-Resolution using Multi-scale Pyramid 3D Convolutional Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413587>  
AUTHORS: Jianping Luo, Shaofei Huang, Yuan Yuan  
HIGHLIGHT: In this paper, we propose a multi-scale pyramid 3D convolutional (MP3D) network for video SR, where 3D convolution can explore temporal correlation directly without explicit motion compensation.
- 206, TITLE: PCA-SRGAN: Incremental Orthogonal Projection Discrimination for Face Super-resolution  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413590>  
AUTHORS: Hao Dou, Chen Chen, Xiyuan Hu, Zuxing Xuan, Zhisen Hu, Silong Peng  
HIGHLIGHT: To further improve the performance of GAN-based models on super-resolving face images, we propose PCA-SRGAN which pays attention to the cumulative discrimination in the orthogonal projection space spanned by PCA projection matrix of face data.

- 207, TITLE: Exploring Font-independent Features for Scene Text Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413592>  
AUTHORS: Yizhi Wang, Zhouhui Lian  
HIGHLIGHT: Specifically, we introduce trainable font embeddings to shape the font styles of generated glyphs, with the image feature of scene text only representing its essential patterns.
- 208, TITLE: Context-aware Feature Generation For Zero-shot Semantic Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413593>  
AUTHORS: Zhangxuan Gu, Siyuan Zhou, Li Niu, Zihan Zhao, Liqing Zhang  
HIGHLIGHT: In this paper, we propose a novel context-aware feature generation method for zero-shot segmentation named CaGNet.
- 209, TITLE: Defending Adversarial Examples via DNN Bottleneck Reinforcement  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413604>  
AUTHORS: Wenqing Liu, Miaoqing Shi, Teddy Furon, Li Li  
HIGHLIGHT: In order to reinforce the information bottleneck, we introduce the multi-scale low-pass objective and multi-scale high-frequency communication for better frequency steering in the network.
- 210, TITLE: Weakly-Supervised Video Object Grounding by Exploring Spatio-Temporal Contexts  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413610>  
AUTHORS: Xun Yang, Xueliang liu, Meng Jian, Xinjian Gao, Meng Wang  
HIGHLIGHT: To fill the research gap, this paper presents a weakly-supervised framework for linking objects mentioned in a sentence with the corresponding regions in videos.
- 211, TITLE: S2SiamFC: Self-supervised Fully Convolutional Siamese Network for Visual Tracking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413611>  
AUTHORS: Chon Hou Sio, Yu-Jen Ma, Hong-Han Shuai, Jun-Cheng Chen, Wen-Huang Cheng  
HIGHLIGHT: To exploit rich information from unlabeled data, in this work, we propose a novel self-supervised framework for visual tracking which can easily adapt the state-of-the-art supervised Siamese-based trackers into unsupervised ones by utilizing the fact that an image and any cropped region of it can form a natural pair for self-training.
- 212, TITLE: Learnable Optimal Sequential Grouping for Video Scene Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413612>  
AUTHORS: Daniel Rotman, Yevgeny Yaroker, Elad Amrani, Udi Barzelay, Rami Ben-Ari  
HIGHLIGHT: In this work, we extend the capabilities of OSG to the learning regime.
- 213, TITLE: NOH-NMS: Improving Pedestrian Detection by Nearby Objects Hallucination  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413617>  
AUTHORS: Penghao Zhou, Chong Zhou, Pai Peng, Junlong Du, Xing Sun, Xiaowei Guo, Feiyue Huang  
HIGHLIGHT: Thus, we propose 'heatmapname (\heatmapnameshort ), which pinpoints the objects nearby each proposal with a Gaussian distribution, together with \nmsname, which dynamically eases the suppression for the space that might contain other objects with a high likelihood.
- 214, TITLE: Dual-Gradients Localization Framework for Weakly Supervised Object Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413622>  
AUTHORS: Chuangchuang Tan, Guanghua Gu, Tao Ruan, Shikui Wei, Yao Zhao  
HIGHLIGHT: In this work, we propose an offline framework to achieve precise localization on any convolutional layer of a classification model by exploiting two kinds of gradients, called Dual-Gradients Localization (DGL) framework.
- 215, TITLE: DualLip: A System for Joint Lip Reading and Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413623>  
AUTHORS: Weicong Chen, Xu Tan, Yingce Xia, Tao Qin, Yu Wang, Tie-Yan Liu  
HIGHLIGHT: In this paper, we develop DualLip, a system that jointly improves lip reading and generation by leveraging the task duality and using unlabeled text and lip video data.
- 216, TITLE: Dual Attention GANs for Semantic Image Synthesis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416270>  
AUTHORS: Hao Tang, Song Bai, Nicu Sebe

**HIGHLIGHT:** In this paper, we focus on the semantic image synthesis task that aims at transferring semantic label maps to photo-realistic images.

217, **TITLE:** SimSwap: An Efficient Framework For High Fidelity Face Swapping  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413630>  
**AUTHORS:** Renwang Chen, Xuanhong Chen, Bingbing Ni, Yanhao Ge  
**HIGHLIGHT:** We propose an efficient framework, called Simple Swap (SimSwap), aiming for generalized and high fidelity face swapping.

218, **TITLE:** Self-Mimic Learning for Small-scale Pedestrian Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413634>  
**AUTHORS:** Jialian Wu, Chunluan Zhou, Qian Zhang, Ming Yang, Junsong Yuan  
**HIGHLIGHT:** In this paper, we conduct an in-depth analysis of the small-scale pedestrian detection problem, which reveals that weak representations of small-scale pedestrians are the main cause for a classifier to miss them.

219, **TITLE:** Action2Motion: Conditioned Generation of 3D Human Motions  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413635>  
**AUTHORS:** Chuan Guo, Xinxin Zuo, Sen Wang, Shihao Zou, Qingyao Sun, Annan Deng, Minglun Gong, Li Cheng  
**HIGHLIGHT:** This paper, on the other hand, considers a relatively new problem, which could be thought of as an inverse of action recognition: given a prescribed action type, we aim to generate plausible human motion sequences in 3D.

220, **TITLE:** Skin Textural Generation via Blue-noise Gabor Filtering based Generative Adversarial Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413637>  
**AUTHORS:** Hui Zhang, Chuan Wang, Nenglu Chen, Jue Wang, Wenping Wang  
**HIGHLIGHT:** To this end, we propose a new facial noise generation method.

221, **TITLE:** A Slow-I-Fast-P Architecture for Compressed Video Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413641>  
**AUTHORS:** Jiapeng Li, Ping Wei, Yongchi Zhang, Nanning Zheng  
**HIGHLIGHT:** In this paper, we propose a novel Slow-I-Fast-P (SIFP) neural network model for compressed video action recognition.

222, **TITLE:** DMVOS: Discriminative Matching for Real-time Video Object Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414035>  
**AUTHORS:** Peisong Wen, Ruolin Yang, Qianqian Xu, Chen Qian, Qingming Huang, Runmin Cong, Jianlou Si  
**HIGHLIGHT:** In this work, we propose Discriminative Matching for real-time Video Object Segmentation (DMVOS), a real-time VOS framework with high-accuracy to fill this gap.

223, **TITLE:** Multi-Group Multi-Attention: Towards Discriminative Spatiotemporal Representation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413646>  
**AUTHORS:** Zhensheng Shi, Liangjie Cao, Cheng Guan, Ju Liang, Qianqian Li, Zhaorui Gu, Haiyong Zheng, Bing Zheng  
**HIGHLIGHT:** In this paper, we propose Multi-Group Multi-Attention, dubbed MGMA, paying more attention to “where and when”; the action happens, for learning discriminative spatiotemporal representation in videos.

224, **TITLE:** Vaccine-style-net: Point Cloud Completion in Implicit Continuous Function Space  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413648>  
**AUTHORS:** Wei Yan, Ruonan Zhang, Jing Wang, Shan Liu, Thomas H. Li, Ge Li  
**HIGHLIGHT:** In this paper, we propose Vaccine-Style-Net, a new point cloud completion method that can produce high resolution 3D shapes with complete smooth surface.

225, **TITLE:** Adaptive Wasserstein Hourglass for Weakly Supervised RGB 3D Hand Pose Estimation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413651>  
**AUTHORS:** Yumeng Zhang, Li Chen, Yufeng Liu, Wen Zheng, Junhai Yong  
**HIGHLIGHT:** In this paper, we propose a domain adaptation method called Adaptive Wasserstein Hourglass for weakly-supervised 3D hand pose estimation to close the large gap between synthetic and real-world datasets flexibly.

226, **TITLE:** Weakly Supervised Segmentation with Maximum Bipartite Graph Matching  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413652>  
**AUTHORS:** Weide Liu, Chi Zhang, Guosheng Lin, Tzu-Yi HUNG, Chunyan Miao  
**HIGHLIGHT:** We propose to improve the CAMs from a novel graph perspective.

- 227, TITLE: Recognizing Camera Wearer from Hand Gestures in Egocentric Videos: <https://egocentricbiometric.github.io/https://dl.acm.org/doi/abs/10.1145/3394171.3413654>  
AUTHORS: Daksh Thapar, Aditya Nigam, Chetan Arora  
HIGHLIGHT: In this work, we take the privacy threat a notch higher and show that even the wearer's hand gestures, as seen through an egocentric video, leak wearer's identity.
- 228, TITLE: Prototype-Matching Graph Network for Heterogeneous Domain Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413662>  
AUTHORS: Zijian Wang, Yadan Luo, Zi Huang, Mahsa Baktashmotlagh  
HIGHLIGHT: To tackle this problem, in this paper, we propose the Prototype-Matching Graph Network (PMGN), which gradually explores the domain-invariant class prototype representations.
- 229, TITLE: Towards Lighter and Faster: Learning Wavelets Progressively for Image Super-Resolution  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413664>  
AUTHORS: Huanrong Zhang, Zhi Jin, Xiaojun Tan, Xiying Li  
HIGHLIGHT: To address this trade-off issue between reconstruction performance, the number of network parameters, and inference time, we propose a lightweight and fast network (WSR) to learn wavelet coefficients of the target image progressively for single image super-resolution.
- 230, TITLE: Spatio-Temporal Inception Graph Convolutional Networks for Skeleton-Based Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413666>  
AUTHORS: Zhen Huang, Xu Shen, Xinmei Tian, Houqiang Li, Jianqiang Huang, Xian-Sheng Hua  
HIGHLIGHT: Specifically, we design a simple and highly modularized graph convolutional network architecture for skeleton-based action recognition.
- 231, TITLE: Dynamic Future Net: Diversified Human Motion Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413669>  
AUTHORS: Wenheng Chen, He Wang, Yi Yuan, Tianjia Shao, Kun Zhou  
HIGHLIGHT: In this paper, we present Dynamic Future Net, a new deep learning model where we explicitly focuses on the aforementioned motion stochasticity by constructing a generative model with non-trivial modelling capacity in temporal stochasticity.
- 232, TITLE: ATF: Towards Robust Face Alignment via Leveraging Similarity and Diversity across Different Datasets  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414037>  
AUTHORS: Xing Lan, Qinghao Hu, Fangzhou Xiong, Cong Leng, Jian Cheng  
HIGHLIGHT: To address the above problems, we proposed a novel Alternating Training Framework (ATF), which leverages similarity and diversity across multi-media sources for a more robust detector.
- 233, TITLE: Dual Gaussian-based Variational Subspace Disentanglement for Visible-Infrared Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413673>  
AUTHORS: Nan Pu, Wei Chen, Yu Liu, Erwin M. Bakker, Michael S. Lew  
HIGHLIGHT: To solve the problem, we present a carefully designed dual Gaussian-based variational auto-encoder (DG-VAE), which disentangles an identity-discriminable and an identity-ambiguous cross-modality feature subspace, following a mixture-of-Gaussians (MoG) prior and a standard Gaussian distribution prior, respectively.
- 234, TITLE: Attention Based Dual Branches Fingertip Detection Network and Virtual Key System  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413685>  
AUTHORS: Chong Mou, Xin Zhang  
HIGHLIGHT: To rectify these problems, this paper proposes an attention-based dual branches network that can efficiently fulfill both fingertip detection and gesture recognition tasks.
- 235, TITLE: Action Completeness Modeling with Background Aware Networks for Weakly-Supervised Temporal Action Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413687>  
AUTHORS: Md Moniruzzaman, Zhaozheng Yin, Zhihai He, Ruwen Qin, Ming C. Leu  
HIGHLIGHT: To solve these problems, we introduce a novel weakly-supervised Action Completeness Modeling with Background Aware Networks (ACM-BANets).
- 236, TITLE: Adversarial Knowledge Transfer from Unlabeled Data  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413688>



AUTHORS: Akash Gupta, Rameswar Panda, Sujoy Paul, Jianming Zhang, Amit K. Roy-Chowdhury  
HIGHLIGHT: In this paper, we present a novel Adversarial Knowledge Transfer (AKT) framework for transferring knowledge from internet-scale unlabeled data to improve the performance of a classifier on a given visual recognition task.

237, TITLE: Task Decoupled Knowledge Distillation For Lightweight Face Detectors  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414069>  
AUTHORS: Xiaoqing Liang, Xu Zhao, Chaoyang Zhao, Nanfei Jiang, Ming Tang, Jinqiao Wang  
HIGHLIGHT: In this paper, we propose a task decoupled knowledge distillation method, which decouples the detection distillation task into two subtasks and uses different samples in distilling the features of different subtasks.

238, TITLE: Self-supervised Video Representation Learning Using Inter-intra Contrastive Framework  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413694>  
AUTHORS: Li Tao, Xueting Wang, Toshihiko Yamasaki  
HIGHLIGHT: We propose a self-supervised method to learn feature representations from videos.

239, TITLE: Memory Recursive Network for Single Image Super-Resolution  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413696>  
AUTHORS: Jie Liu, Minqiang Zou, Jie Tang, Gangshan Wu  
HIGHLIGHT: To address these issues, we propose the memory recursive network (MRNet) to make full use of the output features at each stage.

240, TITLE: Scale-aware Progressive Optimization Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413698>  
AUTHORS: Ying Chen, Lifeng Huang, Chengying Gao, Ning Liu  
HIGHLIGHT: To this end, we propose a scale-aware progressive optimization network (SPO-Net) for crowd counting, which trains a scale adaptive network to achieve high-quality density map estimation and overcome the variable scale dilemma in highly congested scenes.

241, TITLE: Resource Efficient Domain Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413701>  
AUTHORS: Jinguang Jiang, Ximei Wang, Mingsheng Long, Jianmin Wang  
HIGHLIGHT: In this paper, we propose Resource Efficient Domain Adaptation (REDA), a general framework that can adaptively adjust computation re- sources across 'easier' and 'harder' inputs.

242, TITLE: MGAAttack: Toward More Query-efficient Black-box Attack by Microbial Genetic Algorithm  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413703>  
AUTHORS: Lina Wang, Kang Yang, Wenqi Wang, Run Wang, Aoshuang Ye  
HIGHLIGHT: To address the efficiency of querying in black-box attack, we propose a novel attack, called MGAAttack, which is a query-efficient and gradient-free black-box attack without obtaining any knowledge of the target model.

243, TITLE: A Novel Graph-TCN with a Graph Structured Representation for Micro-expression Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413714>  
AUTHORS: Ling Lei, Jianfeng Li, Tong Chen, Shigang Li  
HIGHLIGHT: To the best of our knowledge, we are the first to use the learning-based video motion magnification method to extract the features of shape representations from the intermediate layer while magnifying MEs.

244, TITLE: Masked Face Recognition with Generative Data Augmentation and Domain Constrained Ranking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413723>  
AUTHORS: Mengyue Geng, Peixi Peng, Yangru Huang, Yonghong Tian  
HIGHLIGHT: To obtain sufficient training data, based on the MFSR, we introduce a novel Identity Aware Mask GAN (IAMGAN) with segmentation guided multi-level identity preserve module to generate the synthetic masked face images from the full face images.

245, TITLE: Occlusion Detection for Automatic Video Editing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413725>  
AUTHORS: Junhua Liao, Haihan Duan, Xin Li, Haoran Xu, Yanbing Yang, Wei Cai, Yanru Chen, Liangyin Chen  
HIGHLIGHT: In this paper, for releasing the burden of video editors, a frame-level video occlusion detection method is proposed, which is a fundamental component of automatic video editing.

246, TITLE: Cartoon Face Recognition: A Benchmark Dataset

- <https://dl.acm.org/doi/abs/10.1145/3394171.3413726>  
AUTHORS: Yi Zheng, Yifan Zhao, Mengyuan Ren, He Yan, Xiangju Lu, Junhui Liu, Jia Li  
HIGHLIGHT: To further investigate this challenging dataset, we propose a multi-task domain adaptation approach that jointly utilizes the human and cartoon domain knowledge with three discriminative regularizations.
- 247, TITLE: Reversible Watermarking in Deep Convolutional Neural Networks for Integrity Authentication  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413729>  
AUTHORS: Xiquan Guan, Huamin Feng, Weiming Zhang, Hang Zhou, Jie Zhang, Nenghai Yu  
HIGHLIGHT: In this paper, we propose a reversible watermarking algorithm for integrity authentication.
- 248, TITLE: Masked Face Recognition with Latent Part Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413731>  
AUTHORS: Feifei Ding, Peixi Peng, Yangru Huang, Mengyue Geng, Yonghong Tian  
HIGHLIGHT: This paper focuses on a novel task named masked faces recognition (MFR), which aims to match masked faces with common faces and is important especially during the global outbreak of COVID-19.
- 249, TITLE: PanelNet: A Novel Deep Neural Network for Predicting Collective Diagnostic Ratings by a Panel of Radiologists for Pulmonary Nodules  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413735>  
AUTHORS: Chunyan Zhang, Songhua Xu, Zongfang Li  
HIGHLIGHT: To fill the overlooked gap, this study introduces a novel deep neural network, titled PanelNet, that is able to computationally model and reproduce the aforesaid collective diagnosis capability demonstrated by a group of medical experts.
- 250, TITLE: Privacy-Preserving Visual Content Tagging using Graph Transformer Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414047>  
AUTHORS: Xuan-Son Vu, Duc-Trong Le, Christoffer Edlund, Lili Jiang, Hoang D. Nguyen  
HIGHLIGHT: Therefore, this paper proposes an end-to-end framework (SGTN) using Graph Transformer and Convolutional Networks to significantly improve classification and privacy preservation of visual data.
- 251, TITLE: Rotationally-Consistent Novel View Synthesis for Humans  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413754>  
AUTHORS: Youngjoong Kwon, Stefano Petrangeli, Dahun Kim, Haoliang Wang, Henry Fuchs, Viswanathan Swaminathan  
HIGHLIGHT: To solve these problems, we present in this paper a learning framework for the novel view synthesis of human subjects, which explicitly enforces consistency across different generated views of the subject.
- 252, TITLE: Integrating Semantic Segmentation and Retinex Model for Low-Light Image Enhancement  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413757>  
AUTHORS: Minhao Fan, Wenjing Wang, Wenhan Yang, Jiaying Liu  
HIGHLIGHT: We propose an enhancement pipeline with three parts that effectively utilize the semantic layer information.
- 253, TITLE: Alleviating Human-level Shift: A Robust Domain Adaptation Method for Multi-person Pose Estimation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414040>  
AUTHORS: Xixia Xu, Qi Zou, Xue Lin  
HIGHLIGHT: Therefore, we propose a novel domain adaptation method for multi-person pose estimation to conduct the human-level topological structure alignment and fine-grained feature alignment.
- 254, TITLE: SpatialGAN: Progressive Image Generation Based on Spatial Recursive Adversarial Expansion  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413760>  
AUTHORS: Lei Zhao, Sihuan Lin, Ailin Li, Huaizhong Lin, Wei Xing, Dongming Lu  
HIGHLIGHT: In this paper, we propose a progressive spatial recursive adversarial expansion model(called SpatialGAN) capable of producing high-quality samples of the natural image.
- 255, TITLE: Medical Visual Question Answering via Conditional Reasoning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413761>  
AUTHORS: Li-Ming Zhan, Bo Liu, Lu Fan, Jiaxin Chen, Xiao-Ming Wu  
HIGHLIGHT: In this paper, we propose a novel conditional reasoning framework for Med-VQA, aiming to automatically learn effective reasoning skills for various Med-VQA tasks.
- 256, TITLE: Nighttime Dehazing with a Synthetic Benchmark  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413763>

AUTHORS: Jing Zhang, Yang Cao, Zheng-Jun Zha, Dacheng Tao  
HIGHLIGHT: To address this issue, we propose a novel synthetic method called 3R to simulate nighttime hazy images from daytime clear images, which first reconstructs the scene geometry, then simulates the light rays and object reflectance, and finally renders the haze effects.

257, TITLE: Pay Attention Selectively and Comprehensively: Pyramid Gating Network for Human Pose Estimation without Pre-training

<https://dl.acm.org/doi/abs/10.1145/3394171.3414041>

AUTHORS: Chenru Jiang, Kaizhu Huang, Shufei Zhang, Xinheng Wang, Jimin Xiao

HIGHLIGHT: To mitigate these problems, we propose a novel comprehensive recalibration model called Pyramid Gating Network (PGA-Net) that is capable of distilling, selecting, and fusing the discriminative and attention-aware features at different scales and different levels (i.e., both semantic and natural levels).

258, TITLE: Data-driven Meta-set Based Fine-Grained Visual Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3414044>

AUTHORS: Chuanyi Zhang, Yazhou Yao, Xiangbo Shu, Zechao Li, Zhenmin Tang, Qi Wu

HIGHLIGHT: To this end, we propose a data-driven meta-set based approach to deal with noisy web images for fine-grained recognition.

259, TITLE: WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection

<https://dl.acm.org/doi/abs/10.1145/3394171.3413769>

AUTHORS: Bojia Zi, Minghao Chang, Jingjing Chen, Xingjun Ma, Yu-Gang Jiang

HIGHLIGHT: To better support detection against real-world deepfakes, in this paper, we introduce a new dataset WildDeepfake, which consists of 7,314 face sequences extracted from 707 deepfake videos collected completely from the internet.

260, TITLE: LodoNet: A Deep Neural Network with 2D Keypoint Matching for 3D LiDAR Odometry Estimation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413771>

AUTHORS: Ce Zheng, Yecheng Lyu, Ming Li, Ziming Zhang

HIGHLIGHT: In contrast, motivated by the success of image based feature extractors, we propose to transfer the LiDAR frames to image space and reformulate the problem as image feature extraction.

261, TITLE: Memory-Based Network for Scene Graph with Unbalanced Relations

<https://dl.acm.org/doi/abs/10.1145/3394171.3413507>

AUTHORS: Weitao Wang, Ruyang Liu, Meng Wang, Sen Wang, Xiaojun Chang, Yang Chen

HIGHLIGHT: For these reasons, we propose a novel scene graph generation model that can effectively improve the detection of low-frequency relations.

262, TITLE: Pairwise Similarity Regularization for Adversarial Domain Adaptation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413516>

AUTHORS: Haotian Wang, Wenjing Yang, Ji Wang, Ruxin Wang, Long Lan, Mingyang Geng

HIGHLIGHT: To resolve this issue, we propose a Pairwise Similarity Regularization (PSR) approach that exploits cluster structures of the target domain data and minimizes the divergence between the pairwise similarity of clustering partition and that of pseudo predictions.

263, TITLE: Generalized Zero-Shot Video Classification via Generative Adversarial Networks

<https://dl.acm.org/doi/abs/10.1145/3394171.3413517>

AUTHORS: Mingyao Hong, Guorong Li, Xinfeng Zhang, Qingming Huang

HIGHLIGHT: In order to solve this problem, we propose a description text dataset based on the UCF101 action recognition dataset.

264, TITLE: Drum Synthesis and Rhythmic Transformation with Adversarial Autoencoders

<https://dl.acm.org/doi/abs/10.1145/3394171.3413519>

AUTHORS: Maciej Tomczak, Masataka Goto, Jason Hockman

HIGHLIGHT: This paper presents a method for joint synthesis and rhythm transformation of drum sounds through the use of adversarial autoencoders (AAE).

265, TITLE: MMNet: Multi-Stage and Multi-Scale Fusion Network for RGB-D Salient Object Detection

<https://dl.acm.org/doi/abs/10.1145/3394171.3413523>

AUTHORS: Guibiao Liao, Wei Gao, Qiuping Jiang, Ronggang Wang, Ge Li

**HIGHLIGHT:** To effectively capture multi-scale cross-modal fusion features, this paper proposes a novel Multi-stage and Multi-Scale Fusion Network (MMNet), which consists of a cross-modal multi-stage fusion module (CMFM) and a bi-directional multi-scale decoder (BMD).

266, **TITLE:** Stable Video Style Transfer Based on Partial Convolution with Depth-Aware Supervision  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413526>  
**AUTHORS:** Songhua Liu, Hao Wu, Shoutong Luo, Zhengxing Sun  
**HIGHLIGHT:** presents a novel training framework for video style transfer without dependency on video dataset of target style; (2).

267, **TITLE:** Video Synthesis via Transform-Based Tensor Neural Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413527>  
**AUTHORS:** Yimeng Zhang, Xiao-Yang Liu, Bo Wu, Anwar Walid  
**HIGHLIGHT:** In this paper, we propose a novel multi-phase deep neural network Transform-Based Tensor-Net that exploits the low-rank structure of video data in a learned transform domain, which unfolds an Iterative Shrinkage-Thresholding Algorithm (ISTA) for tensor signal recovery.

268, **TITLE:** Cluster Attention Contrast for Video Anomaly Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413529>  
**AUTHORS:** Ziming Wang, Yuexian Zou, Zeming Zhang  
**HIGHLIGHT:** To avoid these problems, we introduce a novel contrastive representation learning task, Cluster Attention Contrast, to establish subcategories of normality as clusters.

269, **TITLE:** Automatic Interest Recognition from Posture and Behaviour  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413530>  
**AUTHORS:** Wolmer Bigi, Claudio Baccchi, Alberto Del Bimbo  
**HIGHLIGHT:** To address all these aspects, we propose an automatic system that aims to recognize the user's interest towards a garment by just looking at body posture and behaviour.

270, **TITLE:** Referenceless Rate-Distortion Modeling with Learning from Bitstream and Pixel Features  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413545>  
**AUTHORS:** Yangfan Sun, Li Li, Zhu Li, Shan Liu, none none  
**HIGHLIGHT:** Therefore, to improve the fidelity of prediction, we propose a referenceless prediction-based R-QP modeling (PmR-QP) method to estimate bitrate by leveraging a deep learning algorithm with only one-pass coding.

271, **TITLE:** MS2L: Multi-Task Self-Supervised Learning for Skeleton Based Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413548>  
**AUTHORS:** Lilang Lin, Sijie Song, Wenhan Yang, Jiaying Liu  
**HIGHLIGHT:** In this paper, we address self-supervised representation learning from human skeletons for action recognition.

272, **TITLE:** Domain-Adaptive Object Detection via Uncertainty-Aware Distribution Alignment  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413553>  
**AUTHORS:** Dang-Khoa Nguyen, Wei-Lun Tseng, Hong-Han Shuai  
**HIGHLIGHT:** Specifically, we propose a Multi-level Entropy Attention Alignment (MEAA) method that consists of two main components: (1) Local Uncertainty Attentional Alignment (LUAA) module to accelerate the model better perceiving structure-invariant objects of interest by utilizing information theory to measure the uncertainty of each local region via the entropy of the pixel-wise domain classifier and (2) Multi-level Uncertainty-Aware Context Alignment (MUCA) module to enrich domain-invariant information of relevant objects based on the entropy of multi-level domain classifiers.

273, **TITLE:** MM-Hand: 3D-Aware Multi-Modal Guided Hand Generation for 3D Hand Pose Synthesis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413555>  
**AUTHORS:** Zhenyu Wu, Duc Hoang, Shih-Yao Lin, Yusheng Xie, Liangjian Chen, Yen-Yu Lin, Zhangyang Wang, Wei Fan  
**HIGHLIGHT:** We propose a 3D-aware multi-modal guided hand generative network (MM-Hand), together with a novel geometry-based curriculum learning strategy.

274, **TITLE:** Joint Self-Attention and Scale-Aggregation for Self-Calibrated Deraining Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413559>  
**AUTHORS:** Cong Wang, Yutong Wu, Zhixun Su, Junyang Chen  
**HIGHLIGHT:** In this paper, we propose an effective algorithm, called JDNet, to solve the single image deraining problem and conduct the segmentation and detection task for applications.

- 275, TITLE: Hybrid Dynamic-static Context-aware Attention Network for Action Assessment in Long Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413560>  
AUTHORS: Ling-An Zeng, Fa-Ting Hong, Wei-Shi Zheng, Qi-Zhi Yu, Wei Zeng, Yao-Wei Wang, Jian-Huang Lai  
HIGHLIGHT: In this work, we present a novel hybrid dynamic-static Context-aware attention NETWORK (ACTION-NET) for action assessment in long videos.
- 276, TITLE: F2GAN: Fusing-and-Filling GAN for Few-shot Image Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413561>  
AUTHORS: Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, Liqing Zhang  
HIGHLIGHT: In this paper, we propose a Fusing-and-Filling Generative Adversarial Network (F2GAN) to generate realistic and diverse images for a new category with only a few images.
- 277, TITLE: JAFPro: Joint Appearance Fusion and Propagation for Human Video Motion Transfer from Multiple Reference Images  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414001>  
AUTHORS: Xianggang Yu, Haolin Liu, Xiaoguang Han, Zhen Li, Zixiang Xiong, Shuguang Cui  
HIGHLIGHT: We present a novel framework for human video motion transfer.
- 278, TITLE: A W2VV++ Case Study with Automated and Interactive Text-to-Video Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414002>  
AUTHORS: Jakub Lokoš, Tomáš Souček, Patrik Veselý, František Mejzlík, Jiaqi Ji, Chaoxi Xu, Xirong Li  
HIGHLIGHT: To report on this challenging problem, we present two orthogonal task-based performance studies centered around the state-of-the-art W2VV++ query representation learning model for video retrieval.
- 279, TITLE: Attention Cube Network for Image Restoration  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413564>  
AUTHORS: Yucheng Hang, Qingmin Liao, Wenming Yang, Yupeng Chen, Jie Zhou  
HIGHLIGHT: To address these issues, we propose an attention cube network (A-CubeNet) for image restoration for more powerful feature expression and feature correlation learning.
- 280, TITLE: CRNet: A Center-aware Representation for Detecting Text of Arbitrary Shapes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413565>  
AUTHORS: Yu Zhou, Hongtao Xie, Shancheng Fang, Yan Li, Yongdong Zhang  
HIGHLIGHT: To tackle these problems, we propose an anchor-free scene text detector leveraging Center-aware Representation to achieve accurate arbitrary-shaped scene text detection namely CRNet.
- 281, TITLE: Expressional Region Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413567>  
AUTHORS: Xiaoqian Guo, Xiangyang Li, Shuqiang Jiang  
HIGHLIGHT: In this paper, we introduce a new task called expressional region retrieval, in which the query is formulated as a region of image with the associated description.
- 282, TITLE: ATRW: A Benchmark for Amur Tiger Re-identification in the Wild  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413569>  
AUTHORS: Shuyuan Li, Jianguo Li, Hanlin Tang, Rui Qian, Weiyao Lin  
HIGHLIGHT: This paper tries to fill the gap by introducing a novel large-scale dataset, the Amur Tiger Re-identification in the Wild (ATRW) dataset.
- 283, TITLE: VideoIC: A Video Interactive Comments Dataset and Multimodal Multitask Learning for Comments Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413890>  
AUTHORS: Weiyang Wang, Jieting Chen, Qin Jin  
HIGHLIGHT: In order to support various related research, we build a large scale video interactive comments dataset called VideoIC, which consists of 4951 videos spanning 557 hours and 5 million comments.
- 284, TITLE: Human Identification and Interaction Detection in Cross-View Multi-Person Videos with Wearable Cameras  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413903>  
AUTHORS: Jiewen Zhao, Ruize Han, Yiyang Gan, Liang Wan, Wei Feng, Song Wang

**HIGHLIGHT:** By focusing on two wearable cameras and the interactive activities that involve only two people, in this paper we develop a new approach that can simultaneously: (i) identify the same persons across the two videos, (ii) detect the interactive activities of interest, including their occurrence intervals and involved people, and (iii) recognize the category of each interactive activity.

285, **TITLE:** Surface Reconstruction with Unconnected Normal Maps: An Efficient Mesh-based Approach  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413920>

**AUTHORS:** Miaohui Wang, Wuyuan Xie, Maolin Cui

**HIGHLIGHT:** For the first time, this paper presents an efficient approach to address the fundamental problem of surface reconstruction from unconnected normal maps (denoted as  $\text{SfN}^+$ ) using discrete geometry.

286, **TITLE:** MOR-UAV: A Benchmark Dataset and Baselines for Moving Object Recognition in UAV Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413934>

**AUTHORS:** Murari Mandal, Lav Kush Kumar, Santosh Kumar Vipparthi

**HIGHLIGHT:** Therefore, in this paper, we introduce MOR-UAV, a large-scale video dataset for MOR in aerial videos.

287, **TITLE:** Learning Tuple Compatibility for Conditional Outfit Recommendation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413936>

**AUTHORS:** Xuewen Yang, Dongliang Xie, Xin Wang, Jiangbo Yuan, Wanying Ding, Pengyun Yan

**HIGHLIGHT:** To better define the fashion compatibility and more flexibly meet different needs, we propose a novel problem of learning compatibility among multiple tuples (each consisting of an item and category pair), and recommending fashion items following the category choices from customers.

288, **TITLE:** Efficient Crowd Counting via Structured Knowledge Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413938>

**AUTHORS:** Lingbo Liu, Jiaqi Chen, Hefeng Wu, Tianshui Chen, Guanbin Li, Liang Lin

**HIGHLIGHT:** To liberate these crowd counting models, we propose a novel Structured Knowledge Transfer (SKT) framework, which fully exploits the structured knowledge of a well-trained teacher network to generate a lightweight but still highly effective student network.

289, **TITLE:** DeSmoothGAN: Recovering Details of Smoothed Images via Spatial Feature-wise Transformation and Full Attention  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413958>

**AUTHORS:** Yifei Huang, Chenhui Li, Xiaohu Guo, Jing Liao, Chenxu Zhang, Changbo Wang

**HIGHLIGHT:** In this work, we propose DeSmoothGAN to utilize both characteristics specifically.

290, **TITLE:** PatchMatch based Multiview Stereo with Local Quadric Window  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413966>

**AUTHORS:** Hyewon Song, Jaeseong Park, Suwoong Heo, Jiwoo Kang, Sanghoon Lee

**HIGHLIGHT:** In this paper, we propose an accurate PatchMatch based multiview stereo matching method with a quadric support window that efficiently captures the surface of a complex structured object.

291, **TITLE:** Expert Performance in the Examination of Interior Surfaces in an Automobile: Virtual Reality vs. Reality  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413980>

**AUTHORS:** Alexander Tesch, Ralf Dörmr

**HIGHLIGHT:** In this paper, we evaluate the applicability of VR using head mounted displays (HMDs) in an experiment where we had experts examine the design quality of an interior in VR and compared the results with the examination on a powerwall as well as in reality.

292, **TITLE:** Uncertainty-based Traffic Accident Anticipation with Spatio-Temporal Relational Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413827>

**AUTHORS:** Wentao Bao, Qi Yu, Yu Kong

**HIGHLIGHT:** In this paper, we propose an uncertainty-based accident anticipation model with spatio-temporal relational learning.

293, **TITLE:** A Tightly-coupled Semantic SLAM System with Visual, Inertial and Surround-view Sensors for Autonomous Indoor Parking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413867>

**AUTHORS:** Xuan Shao, Lin Zhang, Tianjun Zhang, Ying Shen, Hongyu Li, Yicong Zhou

**HIGHLIGHT:** To this end, this paper proposes a novel tightly-coupled semantic SLAM system by integrating Visual, Inertial, and Surround-view sensors, VIS SLAM for short, for autonomous indoor parking.

- 294, TITLE: Searching Privately by Imperceptible Lying: A Novel Private Hashing Method with Differential Privacy  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413882>  
AUTHORS: Yimu Wang, Shiyin Lu, Lijun Zhang  
HIGHLIGHT: In this paper, we tackle this valuable yet challenging problem and formulate a task termed as private hashing, which takes into account both searching performance and privacy protection.
- 295, TITLE: Leverage Social Media for Personalized Stress Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413596>  
AUTHORS: Xin Wang, Huijun Zhang, Lei Cao, Ling Feng  
HIGHLIGHT: We construct a three-leveled framework, aiming at personalized stress detection based on social media.
- 296, TITLE: Arbitrary Style Transfer via Multi-Adaptation Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414015>  
AUTHORS: Yingying Deng, Fan Tang, Weiming Dong, Wen Sun, Feiyue Huang, Changsheng Xu  
HIGHLIGHT: In this paper, we propose the multi-adaptation network which involves two self-adaptation (SA) modules and one co-adaptation (CA) module: the SA modules adaptively disentangle the content and style representations, i.e., content SA module uses position-wise self-attention to enhance content representation and style SA module uses channel-wise self-attention to enhance style representation; the CA module rearranges the distribution of style representation based on content representation distribution by calculating the local similarity between the disentangled content and style features in a non-local fashion.
- 297, TITLE: Dual-view Attention Networks for Single Image Super-Resolution  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413613>  
AUTHORS: Jingcai Guo, Shiheng Ma, Jie Zhang, Qihua Zhou, Song Guo  
HIGHLIGHT: In this paper, we propose the Dual-view Attention Networks to alleviate these problems for SISR.
- 298, TITLE: MRI Measurement Matrix Learning via Correlation Reweighting  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413629>  
AUTHORS: Zhongnian Li, Tao Zhang, Ruoyu Chen, Daoqiang Zhang  
HIGHLIGHT: In this paper, we propose a novel Measurement Matrix Learning via Correlation Reweighting (MML-CR) approach for exploring and solving this problem by optimizing reweighted model.
- 299, TITLE: Complementary-View Co-Interest Person Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413659>  
AUTHORS: Ruize Han, Jiewen Zhao, Wei Feng, Yiyang Gan, Liang Wan, Song Wang  
HIGHLIGHT: In this paper, we study a much more realistic and challenging problem, namely co-interest person~(CIP) detection from multiple temporally-synchronized videos taken by the complementary and time-varying views.
- 300, TITLE: Multimodal Dialogue Systems via Capturing Context-aware Dependencies of Semantic Elements  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413679>  
AUTHORS: Weidong He, Zhi Li, Dongcai Lu, Enhong Chen, Tong Xu, Baoxing Huai, Jing Yuan  
HIGHLIGHT: To address these issues, we propose a Multimodal dialogue systems with semantic Elements, MATE for short.
- 301, TITLE: EyeShopper: Estimating Shoppers' Gaze using CCTV Cameras  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413683>  
AUTHORS: Carlos Bermejo, Dimitris Chatzopoulos, Pan Hui  
HIGHLIGHT: In this paper, we present EyeShopper, an innovative system that tracks the gaze of shoppers when facing away from the camera and provides insights about their behavior in physical stores.
- 302, TITLE: Exploiting Active Learning in Novel Refractive Error Detection with Smartphones  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413748>  
AUTHORS: Eugene Yujun Fu, Zhongqi Yang, Hong Va Leong, Grace Ngai, Chi-wai Do, Lily Chan  
HIGHLIGHT: To address these challenges, this paper exploits active learning methods with a set of Convolutional Neural Network features encoding information of human eyes from pre-trained gaze estimation model.
- 303, TITLE: Price Suggestion for Online Second-hand Items with Texts and Images  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413759>  
AUTHORS: Liang Han, Zhaozheng Yin, Zhurong Xia, Minqian Tang, Rong Jin  
HIGHLIGHT: This paper presents an intelligent price suggestion system for online second-hand listings based on their uploaded images and text descriptions.

- 304, TITLE: An Advanced LiDAR Point Cloud Sequence Coding Scheme for Autonomous Driving  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413537>  
AUTHORS: Xuebin Sun, Sukai Wang, Miaohui Wang, Shing Shin Cheng, Ming Liu  
HIGHLIGHT: Learning from the high efficiency video coding (HEVC) coding framework, we propose an advanced coding scheme for large-scale LiDAR point cloud sequences, in which several techniques have been developed to remove the spatial and temporal redundancy.
- 305, TITLE: Learning Optimization-based Adversarial Perturbations for Attacking Sequential Recognition Models  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413543>  
AUTHORS: Xing Xu, Jiefu Chen, Jinhui Xiao, Zheng Wang, Yang Yang, Heng Tao Shen  
HIGHLIGHT: In this paper, we study the adversarial attack on the general and popular DNN structure of CNN+RNN, i.e., the combination of convolutional neural network (CNN) and recurrent neural network (RNN), which has been widely used in various SR tasks.
- 306, TITLE: Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413570>  
AUTHORS: Trisha Mittal, Uttaran Bhattacharya, Rohan Chandra, Aniket Bera, Dinesh Manocha  
HIGHLIGHT: We present a learning-based method for detecting real and fake deepfake multimedia content.
- 307, TITLE: Deep Disturbance-Disentangled Learning for Facial Expression Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413907>  
AUTHORS: Delian Ruan, Yan Yan, Si Chen, Jing-Hao Xue, Hanzi Wang  
HIGHLIGHT: In this paper, we propose a novel Deep Disturbance-disentangled Learning (DDL) method for FER.
- 308, TITLE: Unsupervised Learning Facial Parameter Regressor for Action Unit Intensity Estimation via Differentiable Renderer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413955>  
AUTHORS: Xinhui Song, Tianyang Shi, Zunlei Feng, Mingli Song, Jackie Lin, Chuanjie Lin, Changjie Fan, Yi Yuan  
HIGHLIGHT: In this paper, we present a framework to predict the facial parameters (including identity parameters and AU parameters) based on a bone-driven face model (BDFM) under different views.
- 309, TITLE: Semi-supervised Multi-modal Emotion Recognition with Cross-Modal Distribution Matching  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413579>  
AUTHORS: Jingjun Liang, Ruichen Li, Qin Jin  
HIGHLIGHT: In this work, we propose a novel semi-supervised multi-modal emotion recognition model based on cross-modality distribution matching, which leverages abundant unlabeled data to enhance the model training under the assumption that the inner emotional status is consistent at the utterance level across modalities.
- 310, TITLE: PersonalitySensing: A Multi-View Multi-Task Learning Approach for Personality Detection based on Smartphone Usage  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413591>  
AUTHORS: Songcheng Gao, Wenzhong Li, Lynda J. Song, Xiao Zhang, Mingkai Lin, Sanglu Lu  
HIGHLIGHT: In this paper, we propose a deep learning approach to infer people's Big Five personality traits based on smartphone data.
- 311, TITLE: AU-assisted Graph Attention Convolutional Network for Micro-Expression Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414012>  
AUTHORS: Hong-Xia Xie, Ling Lo, Hong-Han Shuai, Wen-Huang Cheng  
HIGHLIGHT: In this paper, we propose a novel micro-expression recognition approach by combining Action Units (AUs) and emotion category labels.
- 312, TITLE: DFEW: A Large-Scale Database for Recognizing Dynamic Facial Expressions in the Wild  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413620>  
AUTHORS: Xingxun Jiang, Yuan Zong, Wenming Zheng, Chuangao Tang, Wanchuang Xia, Cheng Lu, Jiateng Liu  
HIGHLIGHT: In this paper, we focus on this challenging but interesting topic and make contributions from three aspects.
- 313, TITLE: Region of Interest Based Graph Convolution: A Heatmap Regression Approach for Action Unit Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413674>  
AUTHORS: Zheng Zhang, Taoyue Wang, Lijun Yin



**HIGHLIGHT:** In this paper, we first extend the heatmaps to ROI maps, encoding the location of both positive and negative occurred AUs, then employ a well-designed backbone network to regress it.

314, **TITLE:** IExpressNet: Facial Expression Recognition with Incremental Classes

<https://dl.acm.org/doi/abs/10.1145/3394171.3413718>

**AUTHORS:** Junjie Zhu, Bingjun Luo, Sicheng Zhao, Shihui Ying, Xibin Zhao, Yue Gao

**HIGHLIGHT:** To address these problems, we develop an Incremental Facial Expression Recognition Network (IExpressNet), which can learn a competitive multi-class classifier at any time with a lower requirement of computing resources.

315, **TITLE:** SST-EmotionNet: Spatial-Spectral-Temporal based Attention 3D Dense Network for EEG Emotion Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413724>

**AUTHORS:** Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, Jing Wang

**HIGHLIGHT:** In this paper, we propose a novel spatial-spectral-temporal based attention 3D dense network, named SST-EmotionNet, for EEG emotion recognition.

316, **TITLE:** Language Models as Emotional Classifiers for Textual Conversation

<https://dl.acm.org/doi/abs/10.1145/3394171.3413755>

**AUTHORS:** Connor T. Heaton, David M. Schwartz

**HIGHLIGHT:** In this study we present a novel methodology for classifying emotion in a conversation.

317, **TITLE:** Occluded Facial Expression Recognition with Step-Wise Assistance from Unpaired Non-Occluded Images

<https://dl.acm.org/doi/abs/10.1145/3394171.3413773>

**AUTHORS:** Bin Xia, Shangfei Wang

**HIGHLIGHT:** Considering facial images without occlusions usually provide more information for facial expression recognition compared to occluded facial images, we propose a step-wise learning strategy for occluded facial expression recognition that utilizes unpaired non-occluded images as guidance in the feature and label space.

318, **TITLE:** Learning from Macro-expression: a Micro-expression Recognition Framework

<https://dl.acm.org/doi/abs/10.1145/3394171.3413774>

**AUTHORS:** Bin Xia, Weikang Wang, Shangfei Wang, Enhong Chen

**HIGHLIGHT:** Since micro-expression and macro-expression share some similarities in facial muscle movements and texture changes, in this paper we propose a micro-expression recognition framework that leverages macro-expression samples as guidance.

319, **TITLE:** Emotion-Based End-to-End Matching Between Image and Music in Valence-Arousal Space

<https://dl.acm.org/doi/abs/10.1145/3394171.3413776>

**AUTHORS:** Sicheng Zhao, Yaxian Li, Xingxu Yao, Weizhi Nie, Pengfei Xu, Jufeng Yang, Kurt Keutzer

**HIGHLIGHT:** In this paper, we study end-to-end matching between image and music based on emotions in the continuous valence-arousal (VA) space.

320, **TITLE:** Exploiting Multi-Emotion Relations at Feature and Label Levels for Emotion Tagging

<https://dl.acm.org/doi/abs/10.1145/3394171.3413506>

**AUTHORS:** Zhiwei Xu, Shangfei Wang, Can Wang

**HIGHLIGHT:** In this paper, we propose a novel emotion tagging method, that thoroughly explores emotion relations from both the feature and label levels.

321, **TITLE:** Uncertainty-aware Cross-dataset Facial Expression Recognition via Regularized Conditional Alignment

<https://dl.acm.org/doi/abs/10.1145/3394171.3413515>

**AUTHORS:** Linyi Zhou, Xijian Fan, Yingjie Ma, Tardi Tjahjadi, Qiaolin Ye

**HIGHLIGHT:** To mitigate this problem, this paper proposes an unsupervised domain adaptation method via regularized conditional alignment for FER, which adversarially reduces domain- and class-wise discrepancies while explicitly dealing with uncertainties within and across domain.

322, **TITLE:** Fonts Like This but Happier: A New Way to Discover Fonts

<https://dl.acm.org/doi/abs/10.1145/3394171.3413534>

**AUTHORS:** Tugba Kulahcioglu, Gerard de Melo

**HIGHLIGHT:** In this study, we propose a new multimodal font discovery method in which users provide a reference font together with the changes they wish to obtain in order to get closer to their ideal font.

323, **TITLE:** Adaptive Multimodal Fusion for Facial Action Units Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413538>  
AUTHORS: Huiyuan Yang, Taoyue Wang, Lijun Yin  
HIGHLIGHT: In this paper, we propose a novel Adaptive Multimodal Fusion method (AMF) for AU detection, which learns to select the most relevant feature representations from different modalities by a re-sampling procedure conditioned on a feature scoring module.

324, TITLE: Exploiting Self-Supervised and Semi-Supervised Learning for Facial Landmark Tracking with Unlabeled Data  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413547>  
AUTHORS: Shi Yin, Shangfei Wang, Xiaoping Chen, Enhong Chen  
HIGHLIGHT: To relieve the burden of manual annotations, we propose a novel facial landmark tracking method that makes full use of unlabeled facial videos by exploiting both self-supervised and semi-supervised learning mechanisms.

325, TITLE: Cross Corpus Physiological-based Emotion Recognition Using a Learnable Visual Semantic Graph Convolutional Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413552>  
AUTHORS: Woan-Shiuan Chien, Hao-Chun Yang, Chi-Chun Lee  
HIGHLIGHT: In this study, we aim to develop a network learning strategy for robust cross-corpus emotion recognition using physiological features jointly with affective video content.

326, TITLE: Few-Shot Ensemble Learning for Video Classification with SlowFast Memory Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416269>  
AUTHORS: Mengshi Qi, Jie Qin, Xiantong Zhen, Di Huang, Yi Yang, Jiebo Luo  
HIGHLIGHT: In this paper, we address few-shot video classification by learning an ensemble of SlowFast networks augmented with memory units.

327, TITLE: Look Through Masks: Towards Masked Face Recognition with De-Occlusion Distillation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413960>  
AUTHORS: Chenyu Li, Shiming Ge, Daichi Zhang, Jia Li  
HIGHLIGHT: Inspired by recent progress on amodal perception, we propose to migrate the mechanism of amodal completion for the task of masked face recognition with an end-to-end de-occlusion distillation framework, which consists of two modules.

328, TITLE: Privacy-sensitive Objects Pixelation for Live Video Streaming  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413972>  
AUTHORS: Jizhe Zhou, Chi-Man Pun, Yu Tong  
HIGHLIGHT: To cope with the inevitable but impacting detection issue, we propose a novel Privacy-sensitive Objects Pixelation (PsOP) framework for automatic personal privacy filtering during live video streaming.

329, TITLE: Deep Local Binary Coding for Person Re-Identification by Delving into the Details  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413979>  
AUTHORS: Jiabin Chen, Jie Qin, Yichao Yan, Lei Huang, Li Liu, Fan Zhu, Ling Shao  
HIGHLIGHT: In this work, we present a novel binary representation learning framework for efficient person ReID, namely Deep Local Binary Coding (DLBC).

330, TITLE: March on Data Imperfections: Domain Division and Domain Generalization for Semantic Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413852>  
AUTHORS: Hai Xu, Hongtao Xie, Zheng-Jun Zha, Sun-ao Liu, Yongdong Zhang  
HIGHLIGHT: In contrast to previous works, we present a novel model-agnostic training optimization algorithm which has two prominent components: Domain Division and Domain Generalization.

331, TITLE: Gait Recognition with Multiple-Temporal-Scale 3D Convolutional Neural Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413861>  
AUTHORS: Beibei Lin, Shunli Zhang, Feng Bao  
HIGHLIGHT: To address the above issues, we propose a novel multiple-temporal-scale gait recognition framework which integrates the temporal information in multiple temporal scales, making use of both the frame and interval fusion information.

332, TITLE: SRHEN: Stepwise-Refining Homography Estimation Network via Parsing Geometric Correspondences in Deep Latent Space  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413870>  
AUTHORS: Yi Li, Wenjie Pei, Zhenyu He  
HIGHLIGHT: In this paper, we propose to parse the geometric correspondences between related images explicitly to bridge the gap between deep appearance features and the homography.

333, TITLE: Tactile Sketch Saliency  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413588>  
AUTHORS: Jianbo Jiao, Ying Cao, Manfred Lau, Rynson Lau  
HIGHLIGHT: In this paper, we aim to understand the functionality of 2D sketches by predicting how humans would interact with the objects depicted by sketches in real life.

334, TITLE: Towards Clustering-friendly Representations: Subspace Clustering via Graph Filtering  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413597>  
AUTHORS: Zhengrui Ma, Zhao Kang, Guangchun Luo, Ling Tian, Wenyu Chen  
HIGHLIGHT: To recover the "clustering-friendly" representation and facilitate the subsequent clustering, we propose a graph filtering approach by which a smooth representation is achieved.

335, TITLE: One-shot Scene Graph Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414025>  
AUTHORS: Yuyu Guo, Jingkuan Song, Lianli Gao, Heng Tao Shen  
HIGHLIGHT: In this paper, we propose Multiple Structured Knowledge (Relational Knowledge and Commonsense Knowledge) for the one-shot scene graph generation task.

336, TITLE: Cross-Granularity Learning for Multi-Domain Image-to-Image Translation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413656>  
AUTHORS: Huiyuan Fu, Ting Yu, Xin Wang, Huadong Ma  
HIGHLIGHT: To ensure the important instance to be more realistically translated, we propose a cross-granularity learning model for multi-domain image-to-image translation.

337, TITLE: Enhancing Self-supervised Monocular Depth Estimation via Incorporating Robust Constraints  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413706>  
AUTHORS: Rui Li, Xiantuo He, Yu Zhu, Xianjun Li, Jinqiu Sun, Yanning Zhang  
HIGHLIGHT: In this paper, we address this issue by enhancing the robustness of the self-supervised paradigm using a set of image-based and geometry-based constraints.

338, TITLE: A Novel Object Re-Track Framework for 3D Point Clouds  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413742>  
AUTHORS: Tuo Feng, Licheng Jiao, Hao Zhu, Long Sun  
HIGHLIGHT: In this paper, we propose a 3D object two-stage re-track framework directly utilizing point clouds as the input, without using the ground truth as the reference box.

339, TITLE: Video Relation Detection via Multiple Hypothesis Association  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413764>  
AUTHORS: Zixuan Su, Xindi Shang, Jingjing Chen, Yu-Gang Jiang, Zhiyong Qiu, Tat-Seng Chua  
HIGHLIGHT: In this paper, we propose a novel relation association method called Multiple Hypothesis Association (MHA).

340, TITLE: HOT-Net: Non-Autoregressive Transformer for 3D Hand-Object Pose Estimation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413775>  
AUTHORS: Lin Huang, Jianchao Tan, Jingjing Meng, Ji Liu, Junsong Yuan  
HIGHLIGHT: To overcome these issues, we propose to fully utilize the structural correlations among hand joints and object corners in order to obtain more reliable poses.

341, TITLE: Multi-Features Fusion and Decomposition for Age-Invariant Face Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413499>  
AUTHORS: Lixuan Meng, Chenggang Yan, Jun Li, Jian Yin, Wu Liu, Hongtao Xie, Liang Li  
HIGHLIGHT: To address this issue, in this work we propose a novel Multi-Features Fusion and Decomposition (MFFD) framework to learn more discriminative feature representations and alleviate the intra-class variations for AIFR.

342, TITLE: Part-Aware Interactive Learning for Scene Graph Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413501>  
AUTHORS: Hongshuo Tian, Ning Xu, An-An Liu, Yongdong Zhang  
HIGHLIGHT: In this paper, we propose a part-aware interactive learning method, which are divided into the intra-object and inter-object scenarios.

- 343, TITLE: Retrieval Guided Unsupervised Multi-domain Image to Image Translation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413785>  
AUTHORS: Raul Gomez, Yahui Liu, Marco De Nadai, Dimosthenis Karatzas, Bruno Lepri, Nicu Sebe  
HIGHLIGHT: In this paper we propose the use of an image retrieval system to assist the image-to-image translation task.
- 344, TITLE: GangSweep: Sweep out Neural Backdoors by GAN  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413546>  
AUTHORS: Liwan Zhu, Rui Ning, Cong Wang, Chunsheng Xin, Hongyi Wu  
HIGHLIGHT: This work proposes GangSweep, a new backdoor detection framework that leverages the super reconstructive power of Generative Adversarial Networks (GAN) to detect and "sweep out" neural backdoors.
- 345, TITLE: Iterative Back Modification for Faster Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413901>  
AUTHORS: Zhengcong Fei  
HIGHLIGHT: In this paper, we propose a non-autoregressive approach for faster image caption generation.
- 346, TITLE: VIMES: A Wearable Memory Assistance System for Automatic Information Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413663>  
AUTHORS: Carlos Bermejo, Tristan Braud, Ji Yang, Shayan Mirjafari, Bowen Shi, Yu Xiao, Pan Hui  
HIGHLIGHT: In this work, we propose VIMES, an augmented reality-based memory assistance system that helps recall declarative memory, such as whom the user meets and what they chat.
- 347, TITLE: Neutral Face Game Character Auto-Creation via PokerFace-GAN  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413806>  
AUTHORS: Tianyang Shi, Zhengxia Zou, Xinhui Song, Zheng Song, Changjian Gu, Changjie Fan, Yi Yuan  
HIGHLIGHT: In this paper, considering the above problems, we propose a novel method named "PokerFace-GAN" for neutral face game character auto-creation.
- 348, TITLE: Gray2ColorNet: Transfer More Colors from Reference Image  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413594>  
AUTHORS: Peng Lu, Jinbei Yu, Xujun Peng, Zhaoran Zhao, Xiaojie Wang  
HIGHLIGHT: Thus, an end-to-end colorization network Gray2ColorNet is proposed in this work, where an attention gating mechanism based color fusion network is designed to accomplish the colorization tasks.
- 349, TITLE: Crossing You in Style: Cross-modal Style Transfer from Music to Visual Arts  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413624>  
AUTHORS: Cheng-Che Lee, Wan-Yi Lin, Yen-Ting Shih, Pei-Yi (Patricia) Kuo, Li Su  
HIGHLIGHT: Assuming that musical features can be properly mapped to visual contents through semantic links between the two domains, we solve the music-to-visual style transfer problem in two steps: music visualization and style transfer.
- 350, TITLE: Modeling Caricature Expressions by 3D Blendshape and Dynamic Texture  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413643>  
AUTHORS: Keyu Chen, Jianmin Zheng, Jianfei Cai, Juyong Zhang  
HIGHLIGHT: This paper presents a solution to the problem, with an emphasis on enhancing the ability to create desired expressions and meanwhile preserve the identity exaggeration style of the caricature, which imposes challenges due to the complicated nature of caricatures.
- 351, TITLE: SketchMan: Learning to Create Professional Sketches  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413720>  
AUTHORS: Jia Li, Nan Gao, Tong Shen, Wei Zhang, Tao Mei, Hui Ren  
HIGHLIGHT: We propose a new challenging task sketch enhancement (SE) defined in an ill-posed space, i.e. enhancing a non-professional sketch (NPS) to a professional sketch (PS), which is a creative generation task different from sketch abstraction, sketch completion and sketch variation.
- 352, TITLE: Anisotropic Stroke Control for Multiple Artists Style Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413770>  
AUTHORS: Xuanhong Chen, Xirui Yan, Naiyuan Liu, Ting Qiu, Bingbing Ni  
HIGHLIGHT: To circumvent this issue, we propose a Stroke Control Multi-Artist Style Transfer framework.

- 353, TITLE: A Multi-update Deep Reinforcement Learning Algorithm for Edge Computing Service Offloading  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413702>  
AUTHORS: Hao Hao, Changqiao Xu, Lujie Zhong, Gabriel-Miro Muntean  
HIGHLIGHT: Instead, this paper proposes an innovative deep reinforcement learning method to solve it.
- 354, TITLE: Identity-Aware Attribute Recognition via Real-Time Distributed Inference in Mobile Edge Clouds  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414048>  
AUTHORS: Zichuan Xu, Jiangkai Wu, Qiufen Xia, Pan Zhou, Jiankang Ren, Huizhi Liang  
HIGHLIGHT: In this paper, we design novel models for pedestrian attribute recognition with re-ID in an MEC-enabled camera monitoring system.
- 355, TITLE: Deep Unsupervised Hybrid-similarity Hadamard Hashing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414028>  
AUTHORS: Wanqian Zhang, Dayan Wu, Yu Zhou, Bo Li, Weiping Wang, Dan Meng  
HIGHLIGHT: In this paper, we propose a simple yet effective unsupervised hashing method, dubbed Deep Unsupervised Hybrid-similarity Hadamard Hashing (DU3H), which tackles these issues in an end-to-end deep hashing framework.
- 356, TITLE: Incomplete Cross-modal Retrieval with Dual-Aligned Variational Autoencoders  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413676>  
AUTHORS: Mengmeng Jing, Jingjing Li, Lei Zhu, Ke Lu, Yang Yang, Zi Huang  
HIGHLIGHT: In this paper, we propose a Dual-Aligned Variational Autoencoders (DAVAE) to address the incomplete CMR problem.
- 357, TITLE: MRS-Net: Multi-Scale Recurrent Scalable Network for Face Quality Enhancement of Compressed Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413899>  
AUTHORS: Tie Liu, Mai Xu, Shengxi Li, Rui Ding, Huaida Liu  
HIGHLIGHT: Motivated by scalable video coding, we propose a multi-scale recurrent scalable network (MRS-Net) to enhance the quality of multi-scale faces in compressed videos.
- 358, TITLE: Panoptic Image Annotation with a Collaborative Assistant  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413812>  
AUTHORS: Jasper R.R. Uijlings, Mykhaylo Andriluka, Vittorio Ferrari  
HIGHLIGHT: This paper aims to reduce the time to annotate images for panoptic segmentation, which requires annotating segmentation masks and class labels for all object instances and stuff regions.
- 359, TITLE: Blind Natural Video Quality Prediction via Statistical Temporal Features and Deep Spatial Features  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413845>  
AUTHORS: Jari Korhonen, Yicheng Su, Junyong You  
HIGHLIGHT: In this study, we combine the hand-crafted statistical temporal features used in a state-of-the-art video quality model and spatial features obtained from convolutional neural network trained for image quality assessment via transfer learning.
- 360, TITLE: Aesthetic-Aware Image Style Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413853>  
AUTHORS: Zhiyuan Hu, Jia Jia, Bei Liu, Yaohua Bu, Jianlong Fu  
HIGHLIGHT: In this paper, we propose a novel problem called Aesthetic-Aware Image Style Transfer task, which aims to transfer colour and texture separately and independently to manipulate the aesthetic effect of an image.
- 361, TITLE: Building Movie Map - A Tool for Exploring Areas in a City - and its Evaluations  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413881>  
AUTHORS: Naoki Sugimoto, Yoshihito Ebine, Kiyoharu Aizawa  
HIGHLIGHT: We propose a new Movie Map, which will enable users to explore a given city area using omnidirectional videos.
- 362, TITLE: A Probabilistic Graphical Model for Analyzing the Subjective Visual Quality Assessment Data from Crowdsourcing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413619>  
AUTHORS: Jing Li, Suiyi Ling, Junle Wang, Patrick Le Callet  
HIGHLIGHT: In this paper, we propose a probabilistic graphical annotation model to infer the underlying ground truth and discovering the annotator's behavior.

- 363, TITLE: DroidCloud: Scalable High Density AndroidTM Cloud Rendering  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413675>  
AUTHORS: Linsheng Li, Bin Yang, Cathy Bao, Shuo Liu, Randy Xu, Yong Yao, Mohammad R. Haghghat, Jerry W. Hu, Shoumeng Yan, Zhengwei Qi  
HIGHLIGHT: This paper presents DroidCloud, the first open-source Android\footnoteAndroid is a trademark of Google LLC.
- 364, TITLE: Interpretable Embedding for Ad-Hoc Video Search  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413916>  
AUTHORS: Jiaxin Wu, Chong-Wah Ngo  
HIGHLIGHT: This paper integrates feature embedding and concept interpretation into a neural network for unified dual-task learning.
- 365, TITLE: Joint Attribute Manipulation and Modality Alignment Learning for Composing Text and Image to Image Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413917>  
AUTHORS: Feifei Zhang, Mingliang Xu, Qirong Mao, Changsheng Xu  
HIGHLIGHT: In this paper, we devote to an emerging task in cross-modal retrieval, Composing Text and Image to Image Retrieval (CTI-IR), which aims at retrieving images relevant to a query image with text describing desired modifications to the query image.
- 366, TITLE: Semi-supervised Online Multi-Task Metric Learning for Visual Recognition and Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413948>  
AUTHORS: Yangxi Li, Han Hu, Jin Li, Yong Luo, Yonggang Wen  
HIGHLIGHT: In this paper, we proposed a novel semi-supervised online multi-task DML method termed SOMTML, which enables the models describing different tasks to help each other during the metric learning procedure and thus improving their respective performance.
- 367, TITLE: Supervised Hierarchical Deep Hashing for Cross-Modal Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413962>  
AUTHORS: Yu-Wei Zhan, Xin Luo, Yongxin Wang, Xin-Shun Xu  
HIGHLIGHT: In this paper, we propose an effective cross-modal hashing method, named Supervised Hierarchical Deep Cross-modal Hashing, SHDCH for short, to learn hash codes by explicitly delving into the hierarchical labels.
- 368, TITLE: Multi-graph Convolutional Network for Unsupervised 3D Shape Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413987>  
AUTHORS: Weizhi Nie, Yue Zhao, An-An Liu, Zan Gao, Yuting Su  
HIGHLIGHT: To solve these problems, we propose a novel multi-graph network (MGN) for unsupervised 3D shape retrieval, which utilizes the correlations among modalities and structural similarity between two models to guide the shape representation learning process without category information.
- 369, TITLE: Bottom-Up Foreground-Aware Feature Fusion for Person Search  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413991>  
AUTHORS: Wenjie Yang, Dangwei Li, Xiaotang Chen, Kaiqi Huang  
HIGHLIGHT: In this work, we propose a subnet to fuse the bounding box features that pooled from multiple ConvNet stages in a bottom-up manner, termed bottom-up fusion (BUF) network.
- 370, TITLE: Rethinking Generative Zero-Shot Learning: An Ensemble Learning Perspective for Recognising Visual Patches  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413813>  
AUTHORS: Zhi Chen, Sen Wang, Jingjing Li, Zi Huang  
HIGHLIGHT: To address these issues, we propose a novel framework called multi-patch generative adversarial nets (MPGAN) that synthesises local patch features and labels unseen classes with a novel weighted voting strategy.
- 371, TITLE: Surpassing Real-World Source Training Data: Random 3D Characters for Generalizable Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413815>  
AUTHORS: Yanan Wang, Shengcai Liao, Ling Shao  
HIGHLIGHT: To address this, we propose to automatically synthesize a large-scale person re-identification dataset following a set-up similar to real surveillance but with virtual environments, and then use the synthesized person images to train a generalizable person re-identification model.
- 372, TITLE: Zero-Shot Multi-View Indoor Localization via Graph Location Networks

- <https://dl.acm.org/doi/abs/10.1145/3394171.3413856>  
AUTHORS: Meng-Jiun Chiou, Zhenguang Liu, Yifang Yin, An-An Liu, Roger Zimmermann  
HIGHLIGHT: In this paper, we propose a novel neural network based architecture Graph Location Networks (GLN) to perform infrastructure-free, multi-view image based indoor localization.
- 373, TITLE: Hierarchical Gumbel Attention Network for Text-based Person Search  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413864>  
AUTHORS: Kecheng Zheng, Wu Liu, Jiawei Liu, Zheng-Jun Zha, Tao Mei  
HIGHLIGHT: In this work, we propose a novel hierarchical Gumbel attention network for text-based person search via Gumbel top-k re-parameterization algorithm.
- 374, TITLE: Dual Context-Aware Refinement Network for Person Search  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413878>  
AUTHORS: Jiawei Liu, Zheng-Jun Zha, Richang Hong, Meng Wang, Yongdong Zhang  
HIGHLIGHT: In this work, we propose a novel dual context-aware refinement network (DCRNet) for person search, which jointly explores two kinds of contexts including intra-instance context and inter-instance context to learn discriminative representation.
- 375, TITLE: Heterogeneous Fusion of Semantic and Collaborative Information for Visually-Aware Food Recommendation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413598>  
AUTHORS: Lei Meng, Fuli Feng, Xiangnan He, Xiaoyan Gao, Tat-Seng Chua  
HIGHLIGHT: To address this problem, this paper presents a heterogeneous multi-task learning framework, termed privileged-channel infused network (PiNet).
- 376, TITLE: How to Learn Item Representation for Cold-Start Multimedia Recommendation?  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413628>  
AUTHORS: Xiaoyu Du, Xiang Wang, Xiangnan He, Zechao Li, Jinhui Tang, Tat-Seng Chua  
HIGHLIGHT: In this work, we pay special attention to cold items in multimedia recommender training.
- 377, TITLE: Personalized Item Recommendation for Second-hand Trading Platform  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413640>  
AUTHORS: Xuzheng Yu, Tian Gan, Yinwei Wei, Zhiyong Cheng, Liqiang Nie  
HIGHLIGHT: Accordingly, we proposed a method to simultaneously learn representations of items and users from coarse-grained and fine-grained features, and a multi-task learning strategy is designed to address the issue of data sparsity.
- 378, TITLE: What Aspect Do You Like: Multi-scale Time-aware User Interest Modeling for Micro-video Recommendation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413653>  
AUTHORS: Hao Jiang, Wenjie Wang, Yinwei Wei, Zan Gao, Yinglong Wang, Liqiang Nie  
HIGHLIGHT: In view of this, we propose an end-to-end Multi-scale Time-aware user Interest modeling Network (MTIN).
- 379, TITLE: Domain-Specific Alignment Network for Multi-Domain Image-Based 3D Object Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413655>  
AUTHORS: Yuting Su, Yuqian Li, Dan Song, Zhendong Mao, Xuanya Li, An-An Liu  
HIGHLIGHT: To address these issues, we propose an unsupervised Domain-Specific Alignment Network (DSAN) for multi-domain image-based 3D object retrieval.
- 380, TITLE: Multi-modal Attentive Graph Pooling Model for Community Question Answer Matching  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413711>  
AUTHORS: Jun Hu, Quan Fang, Shengsheng Qian, Changsheng Xu  
HIGHLIGHT: In this paper, we propose a multi-modal attentive graph pooling approach (MMAGP) to model the multi-modal content of questions and answers with GNNs in a unified framework, which explores the multi-modal and redundant properties of CQA systems.
- 381, TITLE: Task-distribution-aware Meta-learning for Cold-start CTR Prediction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413739>  
AUTHORS: Tianwei Cao, Qianqian Xu, Zhiyong Yang, Qingming Huang  
HIGHLIGHT: In this paper, we propose an adaptive loss that ensures the consistency between the task weight and difficulty.
- 382, TITLE: CFVMNet: A Multi-branch Network for Vehicle Re-identification Based on Common Field of View  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413541>  
AUTHORS: Ziruo Sun, Xiushan Nie, Xiaoming Xi, Yilong Yin

**HIGHLIGHT:** In this study, we proposed a multi-branch network based on common field of view (CFVMNet) to address these issues.

383, **TITLE:** Exploiting Heterogeneous Artist and Listener Preference Graph for Music Genre Classification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414000>  
**AUTHORS:** Chunyuan Yuan, Qianwen Ma, Junyang Chen, Wei Zhou, Xiaodan Zhang, Xuehai Tang, Jizhong Han, Songlin Hu  
**HIGHLIGHT:** In this paper, we make use of both artist-music and listener-music preference relations to construct a heterogeneous preference graph.

384, **TITLE:** Graph-Refined Convolutional Network for Multimedia Recommendation with Implicit Feedback  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413556>  
**AUTHORS:** Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Tat-Seng Chua  
**HIGHLIGHT:** In this work, we focus on adaptively refining the structure of interaction graph to discover and prune potential false-positive edges.

385, **TITLE:** Visually Precise Query  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413558>  
**AUTHORS:** Riddhiman Dasgupta, Francis Tom, Sudhir Kumar, Mithun Das Gupta, Yokesh Kumar, Badri N. Patro, Vinay P. Nambodiri  
**HIGHLIGHT:** In this paper we introduce the task of VPQ generation which takes a product image and its title as its input and provides a word level extractive summary of the title, containing a list of salient attributes, which can now be used as a query to search for similar products.

386, **TITLE:** All-in-depth via Cross-baseline Light Field Camera  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413974>  
**AUTHORS:** Dingjian Jin, Anke Zhang, Jiamin Wu, Gaochang Wu, Haoqian Wang, Lu Fang  
**HIGHLIGHT:** Aiming for all-in-depth solution, we propose a cross-baseline LF camera using a commercial LF camera and a monocular camera, which naturally form a 'stereo camera' enabling compensated baseline for LF camera.

387, **TITLE:** Revealing True Identity: Detecting Makeup Attacks in Face-based Biometric Systems  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413606>  
**AUTHORS:** Mohammad Amin Arab, Puria Azadi Moghadam, Mohamed Hussein, Wael Abd-Elmageed, Mohamed Hefeeda  
**HIGHLIGHT:** In this paper, we propose a novel solution to address makeup attacks, which are the hardest to detect in such systems because makeup can substantially alter the facial features of a person, including making them appear older/younger by adding/hiding wrinkles, modifying the shape of eyebrows, beard, and moustache, and changing the color of lips and cheeks.

388, **TITLE:** Relevance-Based Compression of Cataract Surgery Videos Using Convolutional Neural Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413658>  
**AUTHORS:** Negin Ghamsarian, Hadi Amirpourazarian, Christian Timmerer, Mario Taschwer, Klaus Schömann  
**HIGHLIGHT:** To address this problem, we propose a relevance-based compression technique consisting of two modules: (i) relevance detection, which uses neural networks for semantic segmentation and classification of the videos to detect relevant spatio-temporal information, and (ii) content-adaptive compression, which restricts the amount of distortion applied to the relevant content while allocating less bitrate to irrelevant content.

389, **TITLE:** A Modular Approach for Synchronized Wireless Multimodal Multisensor Data Acquisition in Highly Dynamic Social Settings  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413697>  
**AUTHORS:** Chirag Raman, Stephanie Tan, Hayley Hung  
**HIGHLIGHT:** In this work, we propose a modular and cost-effective wireless approach for synchronized multisensor data acquisition of social human behavior.

390, **TITLE:** SphericRTC: A System for Content-Adaptive Real-Time 360-Degree Video Communication  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413999>  
**AUTHORS:** Shuoqian Wang, Xiaoyang Zhang, Mengbai Xiao, Kenneth Chiu, Yao Liu  
**HIGHLIGHT:** We present the SphericRTC system for real-time 360-degree video communication.

391, **TITLE:** Single Image Shape-from-Silhouettes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413625>  
**AUTHORS:** Yawen Lu, Yuxing Wang, Guoyu Lu



**HIGHLIGHT:** In this work, we present a novel shape-from-silhouette method based on just a single image, which is an end-to-end learning framework relying on view synthesis and shape-from-silhouette methodology to reconstruct a 3D shape.

392, **TITLE:** VVSec: Securing Volumetric Video Streaming via Benign Use of Adversarial Perturbation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413639>

**AUTHORS:** Zhongze Tang, Xianglong Feng, Yi Xie, Huy Phan, Tian Guo, Bo Yuan, Sheng Wei

**HIGHLIGHT:** We for the first time identify an effective threat model that extracts 3D face models from volumetric videos and compromises face ID-based authentications To defend against such attack, we develop a novel volumetric video security mechanism, namely VVSec, which makes benign use of adversarial perturbations to obfuscate the security and privacy-sensitive 3D face models.

393, **TITLE:** Bitrate Requirements of Non-Panoramic VR Remote Rendering

<https://dl.acm.org/doi/abs/10.1145/3394171.3413681>

**AUTHORS:** Viktor Kelkkanen, Markus Fiedler, David Lindero

**HIGHLIGHT:** This paper shows the impact of bitrate settings on objective quality measures when streaming non-panoramic remote-rendered Virtual Reality (VR) images.

394, **TITLE:** Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality

<https://dl.acm.org/doi/abs/10.1145/3394171.3413699>

**AUTHORS:** Serhan G&uuml;l, Sebastian Bosse, Dimitri Podborski, Thomas Schierl, Cornelius Hellge

**HIGHLIGHT:** In this paper, we design a Kalman filter for head motion prediction in our cloud-based volumetric video streaming system.

395, **TITLE:** Perception-Lossless Codec of Haptic Data with Low Delay

<https://dl.acm.org/doi/abs/10.1145/3394171.3413728>

**AUTHORS:** Chaoyang Zeng, Tiesong Zhao, Qian Liu, Yiwen Xu, Kai Wang

**HIGHLIGHT:** In this paper, we propose an end-to-end haptic codec for high-efficiency, low-delay and perception-lossless compression of kinesthetic signal, one of two major components of haptic signals.

396, **TITLE:** Neural3D: Light-weight Neural Portrait Scanning via Context-aware Correspondence Learning

<https://dl.acm.org/doi/abs/10.1145/3394171.3413734>

**AUTHORS:** Xin Suo, Minye Wu, Yanshun Zhang, Yingliang Zhang, Lan Xu, Qiang Hu, Jingyi Yu

**HIGHLIGHT:** Aiming at light-weight and realistic human portrait reconstruction, in this paper we propose Neural3D: a novel neural human portrait scanning system using only a single RGB camera.

397, **TITLE:** Presence, Embodied Interaction and Motivation: Distinct Learning Phenomena in an Immersive Virtual Environment

<https://dl.acm.org/doi/abs/10.1145/3394171.3413520>

**AUTHORS:** Jack Ratcliffe, Laurissa Tokarchuk

**HIGHLIGHT:** This paper describes an experiment designed to interrogate these approaches, and provides evidence that embodied controls and presence encourage learning in immersive virtual environments, but for distinct,non-interacting reasons, which are also not explained by motivational benefits.

398, **TITLE:** User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling

<https://dl.acm.org/doi/abs/10.1145/3394171.3413535>

**AUTHORS:** Shishir Subramanyam, Irene Viola, Alan Hanjalic, Pablo Cesar

**HIGHLIGHT:** In this paper, we present a low-complexity tiling approach to perform adaptive streaming of point cloud content.

399, **TITLE:** Leveraging QoE Heterogeneity for Large-Scale Liveicast Scheduling

<https://dl.acm.org/doi/abs/10.1145/3394171.3413918>

**AUTHORS:** Rui-Xiao Zhang, Ming Ma, Tianchi Huang, Hanyu Li, Jiangchuan Liu, Lifeng Sun

**HIGHLIGHT:** In this paper, we conduct measurement studies over large-scale data provided by a top livecast platform in China.

400, **TITLE:** Towards Viewport-dependent 6DoF 360 Video Tiled Streaming for Virtual Reality Systems

<https://dl.acm.org/doi/abs/10.1145/3394171.3413712>

**AUTHORS:** Jong-Beom Jeong, Soonbin Lee, Il-Woong Ryu, Tuan Thanh Le, Eun-Seok Ryu

**HIGHLIGHT:** Therefore, this paper proposes a viewport-dependent high-efficiency video coding (HEVC)-compliant tiled streaming system on test model for immersive video (TMIV), MPEG-Immersive multiview compression reference software.

401, **TITLE:** Low-latency FoV-adaptive Coding and Streaming for Interactive 360&deg; Video Streaming

<https://dl.acm.org/doi/abs/10.1145/3394171.3413751>

AUTHORS: Yixiang Mao, Liyang Sun, Yong Liu, Yao Wang

HIGHLIGHT: This work focuses on developing low-latency and FoV-adaptive coding and streaming strategies for interactive 360-degree video streaming.

402, TITLE: Towards Modality Transferable Visual Information Representation with Optimal Model Compression

<https://dl.acm.org/doi/abs/10.1145/3394171.3413762>

AUTHORS: Rongqun Lin, Linwei Zhu, Shiqi Wang, Sam Kwong

HIGHLIGHT: In this paper, we propose a new scheme for visual signal representation that leverages the philosophy of transferable modality.

403, TITLE: AdaP-360: User-Adaptive Area-of-Focus Projections for Bandwidth-Efficient 360-Degree Video Streaming

<https://dl.acm.org/doi/abs/10.1145/3394171.3413521>

AUTHORS: Chao Zhou, Shuoqian Wang, Mengbai Xiao, Sheng Wei, Yao Liu

HIGHLIGHT: In this work, we motivate a user-adaptive approach to address inefficiencies in 360-degree streaming through an analysis of user-viewing traces.

404, TITLE: Tile Rate Allocation for 360-Degree Tiled Adaptive Video Streaming

<https://dl.acm.org/doi/abs/10.1145/3394171.3413550>

AUTHORS: Praveen Kumar Yadav, Wei Tsang Ooi

HIGHLIGHT: In this paper, we model the tile rate allocation problem as a multiclass knapsack problem with a dynamic profit function that is a function of the FoV and the buffer occupancy.

405, TITLE: Lab2Pix: Label-Adaptive Generative Adversarial Network for Unsupervised Image Synthesis

<https://dl.acm.org/doi/abs/10.1145/3394171.3414027>

AUTHORS: Lianli Gao, Junchen Zhu, Jingkuan Song, Feng Zheng, Heng Tao Shen

HIGHLIGHT: Therefore, we propose an unsupervised framework named Lab2Pix to adaptively synthesize images from labels by elegantly considering the particular properties of label to image synthesis task.

406, TITLE: Deep Multimodal Neural Architecture Search

<https://dl.acm.org/doi/abs/10.1145/3394171.3413977>

AUTHORS: Zhou Yu, Yuhao Cui, Jun Yu, Meng Wang, Dacheng Tao, Qi Tian

HIGHLIGHT: In this paper, we devise a generalized deep multimodal neural architecture search (MMnas) framework for various multimodal learning tasks.

407, TITLE: DIMC-net: Deep Incomplete Multi-view Clustering Network

<https://dl.acm.org/doi/abs/10.1145/3394171.3413807>

AUTHORS: Jie Wen, Zheng Zhang, Zhao Zhang, Zhihao Wu, Lunke Fei, Yong Xu, Bob Zhang

HIGHLIGHT: In this paper, a new deep incomplete multi-view clustering network, called DIMC-net, is proposed to address the challenge of multi-view clustering on missing views.

408, TITLE: Cross-domain Cross-modal Food Transfer

<https://dl.acm.org/doi/abs/10.1145/3394171.3413809>

AUTHORS: Bin Zhu, Chong-Wah Ngo, Jing-jing Chen

HIGHLIGHT: This paper addresses the challenge of resource scarcity in the scenario that only partial data instead of a complete view of data is accessible for model transfer.

409, TITLE: Texture Semantically Aligned with Visibility-aware for Partial Person Re-identification

<https://dl.acm.org/doi/abs/10.1145/3394171.3413833>

AUTHORS: Lishuai Gao, Hua Zhang, Zan Gao, Weili Guan, Zhiyong Cheng, Meng Wang

HIGHLIGHT: In this work, we propose a novel texture semantic alignment (TSA) approach with the visibility-aware for partial person ReID task where the occlusion issue and changes in poses are simultaneously explored in an end-to-end unified framework.

410, TITLE: KTN: Knowledge Transfer Network for Multi-person DensePose Estimation

<https://dl.acm.org/doi/abs/10.1145/3394171.3414014>

AUTHORS: Xuanhan Wang, Lianli Gao, Jingkuan Song, Heng Tao Shen

HIGHLIGHT: In this paper, we address the multi-person densepose estimation problem, which aims at learning dense correspondences between 2D pixels of human body and 3D surface.

- 411, TITLE: Activity-driven Weakly-Supervised Spatio-Temporal Grounding from Untrimmed Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413614>  
AUTHORS: Junwen Chen, Wentao Bao, Yu Kong  
HIGHLIGHT: In this paper, we study the problem of weakly-supervised spatio-temporal grounding from raw untrimmed video streams.
- 412, TITLE: Modeling Temporal Concept Receptive Field Dynamically for Untrimmed Video Analysis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413618>  
AUTHORS: Zhaobo Qi, Shuhui Wang, Chi Su, Li Su, Weigang Zhang, Qingming Huang  
HIGHLIGHT: Accordingly, we introduce temporal dynamic convolution (TDC) to give stronger flexibility to concept-based event analytics.
- 413, TITLE: Relational Graph Learning for Grounded Video Description Generation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413746>  
AUTHORS: Wenqiao Zhang, Xin Eric Wang, Siliang Tang, Haizhou Shi, Haochen Shi, Jun Xiao, Yueting Zhuang, William Yang Wang  
HIGHLIGHT: To tackle the above limitations, we design a novel relational graph learning framework for GVD, in which a language-refined scene graph representation is designed to explore fine-grained visual concepts.
- 414, TITLE: Finding Achilles' Heel: Adversarial Attack on Multi-modal Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413531>  
AUTHORS: Deepak Kumar, Chetan Kumar, Chun Wei Seah, Siyu Xia, Ming Shao  
HIGHLIGHT: Unfortunately, frame selection is usually computationally expensive in run-time, and adding noises to all frames is unrealistic, either. In this paper, we present a novel yet efficient approach to address this issue.
- 415, TITLE: Online Multi-view Subspace Learning with Mixed Noise  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413911>  
AUTHORS: Jinxing Li, Hongwei Yong, Feng Wu, Mu Li  
HIGHLIGHT: To tackle these problems, a novel online multi-view subspace learning is proposed in this paper.
- 416, TITLE: LSOTB-TIR: A Large-Scale High-Diversity Thermal Infrared Object Tracking Benchmark  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413922>  
AUTHORS: Qiao Liu, Xin Li, Zhenyu He, Chenglong Li, Jun Li, Zikun Zhou, Di Yuan, Jing Li, Kai Yang, Nana Fan, Feng Zheng  
HIGHLIGHT: In this paper, we present a Large-Scale and high-diversity general Thermal InfraRed (TIR) Object Tracking Benchmark, called LSOTB-TIR, which consists of an evaluation dataset and a training dataset with a total of 1,400 TIR sequences and more than 600K frames.
- 417, TITLE: Towards More Explainability: Concept Knowledge Mining Network for Event Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413954>  
AUTHORS: Zhaobo Qi, Shuhui Wang, Chi Su, Li Su, Qingming Huang, Qi Tian  
HIGHLIGHT: To address the above issues, we propose a concept knowledge mining network (CKMN) for event recognition.
- 418, TITLE: Simultaneous Semantic Alignment Network for Heterogeneous Domain Adaptation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413995>  
AUTHORS: Shuang Li, Binhui Xie, Jiashu Wu, Ying Zhao, Chi Harold Liu, Zhengming Ding  
HIGHLIGHT: In this paper, we propose a Simultaneous Semantic Alignment Network (SSAN) to simultaneously exploit correlations among categories and align the centroids for each category across domains.
- 419, TITLE: Diverter-Guider Recurrent Network for Diverse Poems Generation from Image  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413830>  
AUTHORS: Liang Li, Shijie Yang, Li Su, Shuhui Wang, Chenggang Yan, Zheng-jun Zha, Qingming Huang  
HIGHLIGHT: This paper proposes the paradigm of multiple poems generation from one image, which is closer to human poetizing but more challenging.
- 420, TITLE: Look, Listen, and Attend: Co-Attention Network for Self-Supervised Audio-Visual Representation Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413869>  
AUTHORS: Ying Cheng, Ruize Wang, Zhihao Pan, Rui Feng, Yuejie Zhang  
HIGHLIGHT: In this paper, we propose a novel self-supervised framework with co-attention mechanism to learn generic cross-modal representations from unlabelled videos in the wild, and further benefit downstream tasks.

- 421, TITLE: Cross-Modal Relation-Aware Networks for Audio-Visual Event Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413581>  
AUTHORS: Haoming Xu, Runhao Zeng, Qingyao Wu, Mingkui Tan, Chuang Gan  
HIGHLIGHT: Motivated by these, in this paper, we propose a relation-aware network to leverage both audio and visual information for accurate event localization.
- 422, TITLE: Learning Deep Multimodal Feature Representation with Asymmetric Multi-layer Fusion  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413621>  
AUTHORS: Yikai Wang, Fuchun Sun, Ming Lu, Anbang Yao  
HIGHLIGHT: We propose a compact and effective framework to fuse multimodal features at multiple layers in a single network.
- 423, TITLE: Look, Listen and Infer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414023>  
AUTHORS: Ruijian Jia, Xinsheng Wang, Shanmin Pang, Jihua Zhu, Jianru Xue  
HIGHLIGHT: In this work, for the first time, a Look, Listen and Infer Network (LLINet) is proposed to learn a zero-shot model that can infer the relations of visual scenes and sounds from novel categories never appeared before.
- 424, TITLE: DCNet: Dense Correspondence Neural Network for 6DoF Object Pose Estimation in Occluded Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413672>  
AUTHORS: Zhi Chen, Wei Yang, Zhenbo Xu, Xike Xie, Liusheng Huang, null null  
HIGHLIGHT: In this work, we propose DCNet, an end-to-end framework for estimating 6DoF object poses.
- 425, TITLE: Transferrable Referring Expression Grounding with Concept Transfer and Context Inheritance  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413677>  
AUTHORS: Xuejing Liu, Liang Li, Shuhui Wang, Zheng-Jun Zha, Dechao Meng, Qingming Huang  
HIGHLIGHT: In this paper, we explore REG in a new scenario, where the REG model can ground novel objects out of REG training data.
- 426, TITLE: Deep Multi-modality Soft-decoding of Very Low Bit-rate Face Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413709>  
AUTHORS: Yanhui Guo, Xi Zhang, Xiaolin Wu  
HIGHLIGHT: We propose a novel deep multi-modality neural network for restoring very low bit rate videos of talking heads.
- 427, TITLE: Multi-modal Multi-relational Feature Aggregation Network for Medical Knowledge Representation Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413736>  
AUTHORS: Yingying Zhang, Quan Fang, Shengsheng Qian, Changsheng Xu  
HIGHLIGHT: In this paper, we propose a Multi-modal Multi-Relational Feature Aggregation Network (MMRFAN) for medical knowledge representation learning.
- 428, TITLE: Photo Stream Question Answer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413745>  
AUTHORS: Wenqiao Zhang, Siliang Tang, Yanpeng Cao, Jun Xiao, Shiliang Pu, Fei Wu, Yueting Zhuang  
HIGHLIGHT: In this paper, we present a new visual question answering (VQA) task -- Photo Stream QA, which aims to answer the open-ended questions about a narrative photo stream.
- 429, TITLE: Generalized Zero-shot Learning with Multi-source Semantic Embeddings for Scene Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413568>  
AUTHORS: Xinhang Song, Haitao Zeng, Sixian Zhang, Luis Herranz, Shuqiang Jiang  
HIGHLIGHT: In this paper we focus on zero-shot scene recognition, a more challenging setting with hundreds of categories where their differences can be subtle and often localized in certain objects or regions.
- 430, TITLE: A Unified Framework for Detecting Audio Adversarial Examples  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413603>  
AUTHORS: Xia Du, Chi-Man Pun, Zheng Zhang  
HIGHLIGHT: In this paper, we propose a unified adversarial detection framework for detecting adaptive audio adversarial examples, which combines noise padding with sound reverberation.
- 431, TITLE: Emerging Topic Detection on the Meta-data of Images from Fashion Social Media

<https://dl.acm.org/doi/abs/10.1145/3394171.3413914>

AUTHORS: Kunihiro Miyazaki, Takayuki Uchiba, Scarlett Young, Yuichi Sasaki, Kenji Tanaka  
HIGHLIGHT: Therefore, in this research, we propose a novel framework for capturing changes in people's tastes in terms of coordination rather than individual items.

432, TITLE: Deep Concept-wise Temporal Convolutional Networks for Action Localization

<https://dl.acm.org/doi/abs/10.1145/3394171.3413860>

AUTHORS: Xin Li, Tianwei Lin, Xiao Liu, Wangmeng Zuo, Chao Li, Xiang Long, Dongliang He, Fu Li, Shilei Wen, Chuang Gan

HIGHLIGHT: In this paper, we empirically find that stacking more conventional temporal convolution layers actually deteriorates action classification performance, possibly ascribing to that all channels of 1D feature map, which generally are highly abstract and can be regarded as latent concepts, are excessively recombined in temporal convolution.

433, TITLE: Who You Are Decides How You Tell

<https://dl.acm.org/doi/abs/10.1145/3394171.3413589>

AUTHORS: Shuang Wu, Shaojing Fan, Zhiqi Shen, Mohan Kankanhalli, Anthony K.H. Tung

HIGHLIGHT: In this paper, we focus on human-centered automatic image captioning.

434, TITLE: Query Twice: Dual Mixture Attention Meta Learning for Video Summarization

<https://dl.acm.org/doi/abs/10.1145/3394171.3414064>

AUTHORS: Junyan Wang, Yang Bai, Yang Long, Bingzhang Hu, Zhenhua Chai, Yu Guan, Xiaolin Wei

HIGHLIGHT: In this paper, we propose a novel framework named Dual Mixture Attention (DMASum) model with Meta Learning for video summarization that tackles the softmax bottleneck problem, where the Mixture of Attention layer (MoA) effectively increases the model capacity by employing twice self-query attention that can capture the second-order changes in addition to the initial query-key attention, and a novel Single Frame Meta Learning rule is then introduced to achieve more generalization to small datasets with limited training sources.

435, TITLE: Textual Dependency Embedding for Person Search by Language

<https://dl.acm.org/doi/abs/10.1145/3394171.3413895>

AUTHORS: Kai Niu, Yan Huang, Liang Wang

HIGHLIGHT: In this work, we focus on the long-distance dependencies in a sentence for better textual encoding, and accordingly propose the Textual Dependency Embedding (TDE) method.

436, TITLE: Visual-Semantic Graph Matching for Visual Grounding

<https://dl.acm.org/doi/abs/10.1145/3394171.3413902>

AUTHORS: Chenchen Jing, Yuwei Wu, Mingtao Pei, Yao Hu, Yunde Jia, Qi Wu

HIGHLIGHT: In this paper, we formulate visual grounding as a graph matching problem to find node correspondences between a visual scene graph and a language scene graph.

437, TITLE: LAL: Linguistically Aware Learning for Scene Text Recognition

<https://dl.acm.org/doi/abs/10.1145/3394171.3413913>

AUTHORS: Yi Zheng, Wenda Qin, Derry Wijaya, Margrit Betke

HIGHLIGHT: In this work, we propose a bimodal framework that simultaneously utilizes visual and linguistic information to enhance recognition performance.

438, TITLE: Cascade Reasoning Network for Text-based Visual Question Answering

<https://dl.acm.org/doi/abs/10.1145/3394171.3413924>

AUTHORS: Fen Liu, Guanghui Xu, Qi Wu, Qing Du, Wei Jia, Minghui Tan

HIGHLIGHT: We study the problem of text-based visual question answering (T-VQA) in this paper.

439, TITLE: Jointly Cross- and Self-Modal Graph Attention Network for Query-Based Moment Localization

<https://dl.acm.org/doi/abs/10.1145/3394171.3414026>

AUTHORS: Daizong Liu, Xiaoye Qu, Xiao-Yang Liu, Jianfeng Dong, Pan Zhou, Zichuan Xu

HIGHLIGHT: To this end, we propose a novel Cross- and Self-Modal Graph Attention Network (CSMGAN) that recasts this task as a process of iterative messages passing over a joint graph.

440, TITLE: Text-Guided Image Inpainting

<https://dl.acm.org/doi/abs/10.1145/3394171.3413939>

AUTHORS: Zijian Zhang, Zhou Zhao, Zhu Zhang, Baoxing Huai, Jing Yuan

HIGHLIGHT: Based on these observations, we propose a new inpainting problem that introduces text as a kind of guidance to direct and control the inpainting process.

- 441, TITLE: RT-VENet: A Convolutional Network for Real-time Video Enhancement  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413951>  
AUTHORS: Mohan Zhang, Qiqi Gao, Jinglu Wang, Henrik Turbell, David Zhao, Jinhui Yu, Yan Lu  
HIGHLIGHT: We present a novel convolutional network that can perform high-quality enhancement on 1080p videos at 45 FPS with a single CPU, which has high potential for real-world deployment.
- 442, TITLE: Regularized Two-Branch Proposal Networks for Weakly-Supervised Moment Retrieval in Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413967>  
AUTHORS: Zhu Zhang, Zhijie Lin, Zhou Zhao, Jieming Zhu, Xiuqiang He  
HIGHLIGHT: In this paper, we propose a novel Regularized Two-Branch Proposal Network to simultaneously consider the inter-sample and intra-sample confrontments.
- 443, TITLE: Feature Reintegration over Differential Treatment: A Top-down and Adaptive Fusion Network for RGB-D Salient Object Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413969>  
AUTHORS: Miao Zhang, Yu Zhang, Yongri Piao, Beiqi Hu, Huchuan Lu  
HIGHLIGHT: In this paper, we propose a novel top-down multi-level fusion structure where different fusion strategies are utilized to effectively explore the low-level and high-level features.
- 444, TITLE: Dual Path Interaction Network for Video Moment Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413975>  
AUTHORS: Hao Wang, Zheng-Jun Zha, Xuejin Chen, Zhiwei Xiong, Jiebo Luo  
HIGHLIGHT: In this paper, we propose a unified top-down and bottom-up approach called Dual Path Interaction Network (DPIN), where the alignment and discrimination information are closely connected to jointly make the prediction.
- 445, TITLE: Cap2Seg: Inferring Semantic and Spatial Context from Captions for Zero-Shot Image Segmentation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413990>  
AUTHORS: Guiyu Tian, Shuai Wang, Jie Feng, Li Zhou, Yadong Mu  
HIGHLIGHT: In this work we describe Cap2Seg, a novel solution of zero-shot image segmentation that harnesses accompanying image captions for intelligently inferring spatial and semantic context for the zero-shot image segmentation task.
- 446, TITLE: Spatial-Temporal Knowledge Integration: Robust Self-Supervised Facial Landmark Tracking  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413993>  
AUTHORS: Congcong Zhu, Xiaoqiang Li, Jide Li, Guangtai Ding, Weiqin Tong  
HIGHLIGHT: To address these problems, we propose a Spatial-Temporal Knowledge Integration (STKI) approach.
- 447, TITLE: Weakly Supervised 3D Object Detection from Point Clouds  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413805>  
AUTHORS: Zengyi Qin, Jinglu Wang, Yan Lu  
HIGHLIGHT: In this work, we propose VS3D, a framework for weakly supervised 3D object detection from point clouds without using any ground truth 3D bounding box for training.
- 448, TITLE: Bridging the Gap between Vision and Language Domains for Improved Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414004>  
AUTHORS: Fenglin Liu, Xian Wu, Shen Ge, Xiaoyu Zhang, Wei Fan, Yuexian Zou  
HIGHLIGHT: In this paper, we propose to bridge the gap between the vision and language domains in the encoder, by enriching visual information with textual concepts, to achieve deep image understandings.
- 449, TITLE: STRONG: Spatio-Temporal Reinforcement Learning for Cross-Modal Video Moment Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413840>  
AUTHORS: Da Cao, Yawen Zeng, Meng Liu, Xiangnan He, Meng Wang, Zheng Qin  
HIGHLIGHT: In this article, we tackle the cross-modal video moment localization issue, namely, localizing the most relevant video moment in an untrimmed video given a sentence as the query.
- 450, TITLE: Language-Aware Fine-Grained Object Representation for Referring Expression Comprehension  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413850>  
AUTHORS: Heqian Qiu, Hongliang Li, Qingbo Wu, Fanman Meng, Hengcan Shi, Taijin Zhao, King Nghi Ngan  
HIGHLIGHT: To address these problems, we propose a language-aware deformable convolution model (LDC) to learn language-aware fine-grained object representations.

451, TITLE: Hierarchical Scene Graph Encoder-Decoder for Image Paragraph Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413859>  
AUTHORS: Xu Yang, Chongyang Gao, Hanwang Zhang, Jianfei Cai  
HIGHLIGHT: We propose irredundant attention in SSG-RNN to improve the possibility of abstracting topics from rarely described sub-graphs and inheriting attention in WSG-RNN to generate more grounded sentences with the abstracted topics, both of which give rise to more distinctive paragraphs.

452, TITLE: Improving Intra- and Inter-Modality Visual Relation for Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413877>  
AUTHORS: Yong Wang, WenKai Zhang, Qing Liu, Zhengyuan Zhang, Xin Gao, Xian Sun  
HIGHLIGHT: In this paper, we present a novel Intra- and Inter-modality visual Relation Transformer to improve connections among visual features, termed I2RT.

453, TITLE: Exploring Language Prior for Mode-Sensitive Visual Attention Modeling  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414008>  
AUTHORS: Xiaoshuai Sun, Xuying Zhang, Liujuan Cao, Yongjian Wu, Feiyue Huang, Rongrong Ji  
HIGHLIGHT: In this paper, we propose a new probabilistic framework for attention, and introduce the concept of mode to model the flexibility and adaptability of attention modulation in complex environments.

454, TITLE: Topic Adaptation and Prototype Encoding for Few-Shot Visual Storytelling  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413886>  
AUTHORS: Jiacheng Li, Siliang Tang, Juncheng Li, Jun Xiao, Fei Wu, Shiliang Pu, Yueting Zhuang  
HIGHLIGHT: In this paper, we focus on enhancing the generalization ability of the VIST model by considering the few-shot setting.

455, TITLE: ICECAP: Information Concentrated Entity-aware Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413576>  
AUTHORS: Anwen Hu, Shizhe Chen, Qin Jin  
HIGHLIGHT: In this work, we focus on the entity-aware news image captioning task which aims to generate informative captions by leveraging the associated news articles to provide background knowledge about the target image.

456, TITLE: Attacking Image Captioning Towards Accuracy-Preserving Target Words Removal  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414009>  
AUTHORS: Jiayi Ji, Xiaoshuai Sun, Yiyi Zhou, Rongrong Ji, Fuhai Chen, Jianzhuang Liu, Qi Tian  
HIGHLIGHT: In this paper, we investigate the fragility of deep image captioning models against adversarial attacks.

457, TITLE: ConsNet: Learning Consistency Graph for Zero-Shot Human-Object Interaction Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413600>  
AUTHORS: Ye Liu, Junsong Yuan, Chang Wen Chen  
HIGHLIGHT: Leveraging the compositional and relational peculiarities of HOI labels, we propose ConsNet, a knowledge-aware framework that explicitly encodes the relations among objects, actions and interactions into an undirected graph called consistency graph, and exploits Graph Attention Networks (GATs) to propagate knowledge among HOI categories as well as their constituents.

458, TITLE: ChefGAN: Food Image Generation from Recipes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413636>  
AUTHORS: Siyuan Pan, Ling Dai, Xuhong Hou, Huating Li, Bin Sheng  
HIGHLIGHT: To achieve this, we propose a GANs based method termed ChefGAN.

459, TITLE: Dual Hierarchical Temporal Convolutional Network with QA-Aware Dynamic Normalization for Video Story Question Answering  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413649>  
AUTHORS: Fei Liu, Jing Liu, Xinxin Zhu, Richang Hong, Hanqing Lu  
HIGHLIGHT: In this paper, we propose a novel framework named Dual Hierarchical Temporal Convolutional Network (DHTCN) to address the aforementioned defects together.

460, TITLE: Generalized Zero-Shot Learning using Generated Proxy Unseen Samples and Entropy Separation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413657>  
AUTHORS: Omkar Gune, Biplob Banerjee, Subhasis Chaudhuri, Fabio Cuzzolin

**HIGHLIGHT:** In this work, we propose to use a generative model (GAN) for synthesizing the visual proxy samples while strictly adhering to the standard assumptions of the GZSL.

461, **TITLE:** Answer-Driven Visual State Estimator for Goal-Oriented Visual Dialogue  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413668>  
**AUTHORS:** Zipeng Xu, Fangxiang Feng, Xiaojie Wang, Yushu Yang, Huixing Jiang, Zhongyuan Wang  
**HIGHLIGHT:** In this paper, we propose an Answer-Driven Visual State Estimator (ADVSE) to impose the effects of different answers on visual states.

462, **TITLE:** Fine-grained Iterative Attention Network for Temporal Language Localization in Videos  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414053>  
**AUTHORS:** Xiaoye Qu, Pengwei Tang, Zhikang Zou, Yu Cheng, Jianfeng Dong, Pan Zhou, Zichuan Xu  
**HIGHLIGHT:** In this paper, we propose a Fine-grained Iterative Attention Network (FIAN) that consists of an iterative attention module for bilateral query-video in-formation extraction.

463, **TITLE:** Hierarchical Bi-Directional Feature Perception Network for Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413689>  
**AUTHORS:** Zhipu Liu, Lei Zhang, Yang Yang  
**HIGHLIGHT:** To solve this issue, we propose a novel model named Hierarchical Bi-directional Feature Perception Network (HBFP-Net) to correlate multi-level information and reinforce each other.

464, **TITLE:** Hard Negative Samples Emphasis Tracker without Anchors  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413692>  
**AUTHORS:** Zhongzhou Zhang, Lei Zhang  
**HIGHLIGHT:** To address this issue, we propose a simple yet efficient hard negative samples emphasis method, which constrains Siamese network to learn features that are aware of hard negative samples and enhance the discrimination of embedding features.

465, **TITLE:** JointFontGAN: Joint Geometry-Content GAN for Font Generation via Few-Shot Learning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413705>  
**AUTHORS:** Yankun Xi, Guoli Yan, Jing Hua, Zichun Zhong  
**HIGHLIGHT:** In this paper, we propose a novel model, JointFontGAN, to derive fonts, including both geometric structures and shape contents in correctness and consistency with very few font samples available.

466, **TITLE:** DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413707>  
**AUTHORS:** Hua Qi, Qing Guo, Felix Juefei-Xu, Xiaofei Xie, Lei Ma, Wei Feng, Yang Liu, Jianjun Zhao  
**HIGHLIGHT:** In this work, we propose DeepRhythm, a DeepFake detection technique that exposes DeepFakes by monitoring the heartbeat rhythms.

467, **TITLE:** FastLR: Non-Autoregressive Lipreading Model with Integrate-and-Fire  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413740>  
**AUTHORS:** Jinglin Liu, Yi Ren, Zhou Zhao, Chen Zhang, Baoxing Huai, Jing Yuan  
**HIGHLIGHT:** To breakthrough this constraint, we propose FastLR, a non-autoregressive (NAR) lipreading model which generates all target tokens simultaneously.

468, **TITLE:** Multimodal Attention with Image Text Spatial Relationship for OCR-Based Image Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413753>  
**AUTHORS:** Jing Wang, Jinhui Tang, Jiebo Luo  
**HIGHLIGHT:** In this paper, we present a novel design - Multimodal Attention Captioner with OCR Spatial Relationship (dubbed as MMA-SR) architecture, which manages information from different modalities with a multimodal attention network and explores spatial relationships between text tokens for OCR-based image captioning.

469, **TITLE:** Towards Accuracy-Fairness Paradox: Adversarial Example-based Data Augmentation for Visual Debiasing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3413772>  
**AUTHORS:** Yi Zhang, Jitao Sang  
**HIGHLIGHT:** Specifically, to ensure the adversarial generalization as well as cross-task transferability, we propose to couple the operations of target task classifier training, bias task classifier training, and adversarial example generation.

470, **TITLE:** Learning Semantic Concepts and Temporal Alignment for Narrated Video Procedural Captioning



<https://dl.acm.org/doi/abs/10.1145/3394171.3413498>

AUTHORS: Botian Shi, Lei Ji, Zhendong Niu, Nan Duan, Ming Zhou, Xilin Chen  
HIGHLIGHT: In this paper, we start with an encoder-decoder backbone using transformer models.

471, TITLE: LGNN: A Context-aware Line Segment Detector

<https://dl.acm.org/doi/abs/10.1145/3394171.3413784>

AUTHORS: Quan Meng, Jiakai Zhang, Qiang Hu, Xuming He, Jingyi Yu  
HIGHLIGHT: When a 3D point cloud is accessible, we present a multi-modal line segment classification technique for extracting a 3D wireframe of the environment robustly and efficiently.

472, TITLE: DeVLBert: Learning Deconfounded Visio-Linguistic Representations

<https://dl.acm.org/doi/abs/10.1145/3394171.3413518>

AUTHORS: Shengyu Zhang, Tan Jiang, Tan Wang, Kun Kuang, Zhou Zhao, Jianke Zhu, Jin Yu, Hongxia Yang, Fei Wu  
HIGHLIGHT: In this paper, we propose to investigate the problem of out-of-domain visio-linguistic pretraining, where the pretraining data distribution differs from that of downstream data on which the pretrained model will be fine-tuned.

473, TITLE: Sequential Attention GAN for Interactive Image Editing

<https://dl.acm.org/doi/abs/10.1145/3394171.3413551>

AUTHORS: Yu Cheng, Zhe Gan, Yitong Li, Jingjing Liu, Jianfeng Gao  
HIGHLIGHT: To explore more practical and interactive real-life applications, we introduce a new task - Interactive Image Editing, where users can guide an agent to edit images via multi-turn textual commands on-the-fly.

474, TITLE: Portraits of No One: An Internet Artwork

<https://dl.acm.org/doi/abs/10.1145/3394171.3416337>

AUTHORS: Tiago Martins, Joãõ Correia, Sãõ Rebelo, Joãõ Bicker, Penousal Machado  
HIGHLIGHT: Portraits of No One: An Internet Artwork

475, TITLE: MaLiang: An Emotion-driven Chinese Calligraphy Artwork Composition System

<https://dl.acm.org/doi/abs/10.1145/3394171.3416338>

AUTHORS: Ruixue Liu, Shaozu Yuan, Meng Chen, Baoyang Chen, Zhijie Qiu, Xiaodong He  
HIGHLIGHT: We present a novel Chinese calligraphy artwork composition system (MaLiang) which can generate aesthetic, stylistic and diverse calligraphy images based on the emotion status from the input text.

476, TITLE: First Impression: AI Understands Personality

<https://dl.acm.org/doi/abs/10.1145/3394171.3416339>

AUTHORS: Xiaohui Wang, Xia Liang, Miao Lu, Jingyan Qin  
HIGHLIGHT: First impression, an interactive art, is proposed to let AI understand human personality at first glance.

477, TITLE: Draw Portraits by Music: A Music based Image Style Transformation

<https://dl.acm.org/doi/abs/10.1145/3394171.3416340>

AUTHORS: Siyu Jin, Jingyan Qin, Wenfa Li  
HIGHLIGHT: Draw Portraits by Music: A Music based Image Style Transformation

478, TITLE: Little World: Virtual Humans Accompany Children on Dramatic Performance

<https://dl.acm.org/doi/abs/10.1145/3394171.3417225>

AUTHORS: Xiaohui Wang, Xiaoxue Ding, Jinke Li, Jingyan Qin  
HIGHLIGHT: To help them achieve performance, an interactive art called 'little world' is proposed to let virtual humans accompany children on drama performance.

479, TITLE: Keep Running - AI Paintings of Horse Figure and Portrait

<https://dl.acm.org/doi/abs/10.1145/3394171.3416342>

AUTHORS: James She, Carmen Ng, Wadia Sheng  
HIGHLIGHT: "Keep Running" is a collection of human and machine generated paintings using a generative adversarial network technology.

480, TITLE: AI Mirror: Visualize AI's Self-knowledge

<https://dl.acm.org/doi/abs/10.1145/3394171.3416343>

AUTHORS: Siyu Hu, Bo Shui, Siyu Jin, Xiaohui Wang  
HIGHLIGHT: "AI mirror", an interactive art, tends to visualize the self-knowledge mechanism from the AI's perspective, and arouses people's reflection on artificial intelligence.

- 481, TITLE: Image Sentiment Transfer  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414690>  
AUTHORS: Tianlang Chen, Wei Xiong, Haitian Zheng, Jiebo Luo  
HIGHLIGHT: In this work, we introduce an important but still unexplored research task -- image sentiment transfer.
- 482, TITLE: Personal Food Model  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414691>  
AUTHORS: Ali Rostami, Vaibhav Pandey, Nitish Nag, Vesper Wang, Ramesh Jain  
HIGHLIGHT: In this paper, we adopt a person-centric multimedia and multimodal perspective on food computing and show how multimedia and food computing are synergistic and complementary.
- 483, TITLE: Helping Users Tackle Algorithmic Threats on Social Media: A Multimedia Research Agenda  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414692>  
AUTHORS: Christian von der Weth, Ashraf Abdul, Shaojing Fan, Mohan Kankanhalli  
HIGHLIGHT: We investigate how multimedia researchers can help tackle these problems to level the playing field for social media users.
- 484, TITLE: Reproducibility Companion Paper: Instance of Interest Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414811>  
AUTHORS: Fan Yu, Dandan Wang, Haonan Wang, Tongwei Ren, Jinhui Tang, Gangshan Wu, Jingjing Chen, Michael Riegler  
HIGHLIGHT: In this paper, we explain the file structure of the source code and publish the details of our IOID dataset, which can be used to retrain the model with custom parameters.
- 485, TITLE: Reproducibility Companion Paper: Outfit Compatibility Prediction and Diagnosis with Multi-Layered Comparison Network  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414812>  
AUTHORS: Xin Wang, Bo Wu, Yueqi Zhong, Wei Hu, Jan Zah&acute;lka  
HIGHLIGHT: We provide the software package for replicating the implementation of Multi-Layered Comparison Network (MCN), as well as the Polyvore-T dataset and baseline methods compared in the original paper.
- 486, TITLE: Reproducibility Companion Paper: Visual Sentiment Analysis for Review Images with Item-Oriented and User-Oriented CNN  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414813>  
AUTHORS: Quoc-Tuan Truong, Hady W. Lauw, Martin Aum&uuml;ller, Naoko Nitta  
HIGHLIGHT: We revisit our contributions on visual sentiment analysis for online review images published at ACM Multimedia 2017, where we develop item-oriented and user-oriented convolutional neural networks that better capture the interaction of image features with specific expressions of users or items.
- 487, TITLE: Reproducibility Companion Paper: Selective Deep Convolutional Features for Image Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414814>  
AUTHORS: Tuan Hoang, Thanh-Toan Do, Ngai-Man Cheung, Michael Riegler, Jan Zah&acute;lka  
HIGHLIGHT: In this companion paper, firstly, we briefly summarize the contributions of our main manuscript: Selective Deep Convolutional Features for Image Retrieval, published in ACM MultiMedia 2017.
- 488, TITLE: MLModelCI: An Automatic Cloud Platform for Efficient MLaaS  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414535>  
AUTHORS: Huaizheng Zhang, Yuanming Li, Yizheng Huang, Yonggang Wen, Jianxiong Yin, Kyle Guan  
HIGHLIGHT: MLModelCI provides multimedia researchers and developers with a one-stop platform for efficient machine learning (ML) services.
- 489, TITLE: Hysia: Serving DNN-Based Video-to-Retail Applications in Cloud  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414536>  
AUTHORS: Huaizheng Zhang, Yuanming Li, Qiming Ai, Yong Luo, Yonggang Wen, Yichao Jin, Nguyen Binh Duong Ta  
HIGHLIGHT: In this paper, we provide practitioners and researchers in multimedia with a cloud-based platform named Hysia for easy development and deployment of V2R applications.
- 490, TITLE: PyRetri: A PyTorch-based Library for Unsupervised Image Retrieval by Deep Convolutional Neural Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414537>

- AUTHORS: Benyi Hu, Ren-Jie Song, Xiu-Shen Wei, Yazhou Yao, Xian-Sheng Hua, Yuehu Liu  
HIGHLIGHT: In order to fill this gap, we introduce PyRetri, an open source library for deep learning based unsupervised image retrieval.
- 491, TITLE: Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414538>  
AUTHORS: Ralph Gasser, Luca Rossetto, Silvan Heller, Heiko Schuldt  
HIGHLIGHT: In this paper we introduce Cottontail DB, an open source database management system that integrates support for scalar and vector attributes in a unified data and query model that allows for both Boolean retrieval and nearest neighbour search.
- 492, TITLE: BMXNet 2: An Open Source Framework for Low-bit Networks - Reproducing, Understanding, Designing and Showcasing  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414539>  
AUTHORS: Joseph Bethge, Christian Bartz, Haojin Yang, Christoph Meinel  
HIGHLIGHT: BMXNet 2 is an open-source framework that provides a broad basis for academia and industry.
- 493, TITLE: PyAnomaly: A Pytorch-based Toolkit for Video Anomaly Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414540>  
AUTHORS: Yuhao Cheng, Wu Liu, Pengrui Duan, Jingen Liu, Tao Mei  
HIGHLIGHT: In this paper, we present a PyTorch-based video anomaly detection toolbox, namely PyAnomaly that contains high modular and extensible components, comprehensive and impartial evaluation platforms, a friendly manageable system configuration, and the abundant engineering deployment functions.
- 494, TITLE: TAPAS-360deg: A Tool for the Design and Experimental Evaluation of 360deg Video Streaming Systems  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414541>  
AUTHORS: Giuseppe Ribezzo, Luca De Cicco, Vittorio Palmisano, Saverio Mascolo  
HIGHLIGHT: In this paper, we present TAPAS-360deg, an open-source tool that enables designing and experimenting all the components required to build omnidirectional video streaming systems.
- 495, TITLE: SOMHunter: Lightweight Video Search System with SOM-Guided Relevance Feedback  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414542>  
AUTHORS: Miroslav Kratochvil, František Mejzlík, Patrik Veselý, Tomáš Souček, Jakub Lokoš  
HIGHLIGHT: To partially alleviate this difficulty, we provide an open-source version of the lightweight known-item search system SOMHunter that competed successfully at VBS 2020.
- 496, TITLE: Text-to-Image Synthesis via Aesthetic Layout  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414357>  
AUTHORS: Samah Saeed Baraheem, Trung-Nghia Le, Tam V. Nguyen  
HIGHLIGHT: In this work, we introduce a practical system which synthesizes an appealing image from natural language descriptions such that the generated image should maintain the aesthetic level of photographs.
- 497, TITLE: Progressive Domain Adaptation for Robot Vision Person Re-identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414358>  
AUTHORS: Zijun Sha, Zelong Zeng, Zheng Wang, Yoichi Natori, Yasuhiro Taniguchi, Shin'ichi Satoh  
HIGHLIGHT: In this paper, we demonstrate a guiding robot with person followers system, which recognizes the follower using a person re-identification technology.
- 498, TITLE: Semantic Storytelling Automation: A Context-Aware and Metadata-Driven Approach  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416528>  
AUTHORS: Paula Viana, Pedro Carvalho, Maria Teresa Andrade, Pieter P. Jonker, Vasileios Papanikolaou, Inês N. Teixeira, Luis Vilaçães, José Pinto, Tiago Costa  
HIGHLIGHT: We propose an innovative approach that uses context and content information to transform a still photo into an appealing context-aware video clip.
- 499, TITLE: ADHD Intelligent Auxiliary Diagnosis System Based on Multimodal Information Fusion  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414359>  
AUTHORS: Yanyi Zhang, Ming Kong, Tianqi Zhao, Wenchen Hong, Qiang Zhu, Fei Wu  
HIGHLIGHT: We have designed and developed the ADHD intelligent auxiliary diagnosis system with software and hardware cooperation.

- 500, TITLE: Video 360 Content Navigation for Mobile HMD Devices  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414389>  
AUTHORS: Jounsup Park, Mingyuan Wu, Eric Lee, Klara Nahrstedt, Yash Shah, Arielle Rosenthal, John Murray, Kevin Spiteri, Michael Zink, Ramesh Sitaraman  
HIGHLIGHT: We demonstrate a video 360 navigation and streaming system for Mobile HMD devices.
- 501, TITLE: GoldenRetriever: A Speech Recognition System Powered by Modern Information Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414392>  
AUTHORS: Yuanfeng Song, Di Jiang, Xiaoling Huang, Yawen Li, Qian Xu, Raymond Chi-Wing Wong, Qiang Yang  
HIGHLIGHT: Exploiting their commonality, this demonstration proposes a novel system named GoldenRetriever that marries IR with ASR.
- 502, TITLE: Integrating Event Camera Sensor Emulator  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414394>  
AUTHORS: Andrew C. Freeman, Ketan Mayer-Patel  
HIGHLIGHT: Arguing the potential usefulness of this sensor, this paper introduces a system for simulating the sensor's event outputs and pixel firing rate control from 3D-rendered input images.
- 503, TITLE: Scene-segmented Video Information Annotation System V2.0  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414396>  
AUTHORS: Alex Lee, Chang-Uk Kwak, Jeong-Woo Son, Gyeong-June Hahm, Min-Ho Han, Sun-Joong Kim  
HIGHLIGHT: We have built the scene-segmented video information annotation system and upgraded it to version 2.0.
- 504, TITLE: SmartShots: Enabling Automatic Generation of Videos with Data Visualizations Embedded  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414356>  
AUTHORS: Tan Tang, Junxiu Tang, Jiewen Lai, Lu Ying, Peiran Ren, Lingyun Yu, Yingcai Wu  
HIGHLIGHT: Specifically, we propose a computational framework that integrates non-verbal video clips, images, a melody, and a data table to create a video with data visualizations embedded.
- 505, TITLE: A Smart-Site-Survey System using Image-based 3D Metric Reconstruction and Interactive Panorama Visualization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414425>  
AUTHORS: Sha Yu, Kevin Mcguinness, Patricia Moore, David Azcona, Noel O'Connor  
HIGHLIGHT: This work presents a so-called Smart Site Survey (SSS) system that provides an efficient, web-based platform for virtual inspection of remote sites with absolute 3D metrics.
- 506, TITLE: AI-SAS: Automated In-match Soccer Analysis System  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414426>  
AUTHORS: Ning Zhang, Tong Shen, Yue Chen, Wei Zhang, Dan Zeng, Jingen Liu, Tao Mei  
HIGHLIGHT: In this work, we present an Automated In-match Soccer Analysis System (AI-SAS), using a domain-knowledge-based multi-view global tracking.
- 507, TITLE: Detecting Urban Issues With the Object Detection Kit  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414427>  
AUTHORS: Maarten Sukel, Stevan Rudinac, Marcel Worring  
HIGHLIGHT: In the Object Detection Kit demo we will demonstrate how the framework can be used to detect urban issues and showcase the capabilities of the system.
- 508, TITLE: Visual-speech Synthesis of Exaggerated Corrective Feedback  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414444>  
AUTHORS: Yaohua Bu, Weijun Li, Tianyi Ma, Shengqi Chen, Jia Jia, Kun Li, Xiaobo Lu  
HIGHLIGHT: To provide more discriminative feedback for the second language (L2) learners to better identify their mispronunciation, we propose a method for exaggerated visual-speech feedback in computer-assisted pronunciation training (CAPT).
- 509, TITLE: TindART: A Personal Visual Arts Recommender  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414445>  
AUTHORS: Gjorgji Strezoski, Lucas Fijen, Jonathan Mitnik, D&#x00e9;niel L&#x00e9;szl&#x00e9;acute;, Pieter de Marez Oyens, Yoni Schirris, Marcel Worring  
HIGHLIGHT: We present TindART - a comprehensive visual arts recommender system.

- 510, TITLE: Fashionist: Personalising Outfit Recommendation for Cold-Start Scenarios  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414446>  
AUTHORS: Dhruv Verma, Kshitij Gulati, Vasu Goel, Rajiv Ratn Shah  
HIGHLIGHT: We attempt to address the cold-start problem for new users, by leveraging a novel visual preference modelling approach on a small set of input images.
- 511, TITLE: EmotionTracker: A Mobile Real-time Facial Expression Tracking System with the Assistant of Public AI-as-a-Service  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414447>  
AUTHORS: Xuncheng Liu, Jingyi Wang, Weizhan Zhang, Qinghu Zheng, Xuanya Li  
HIGHLIGHT: In this demonstration, we present EmotionTracker, a real-time mobile facial expression tracking system combining AlaaS and mobile local auxiliary computing, including facial expression tracking and the corresponding task offloading.
- 512, TITLE: AvatarMeeting: An Augmented Reality Remote Interaction System With Personalized Avatars  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414449>  
AUTHORS: Xuanyu Wang, Yang Wang, Yan Shi, Weizhan Zhang, Qinghua Zheng  
HIGHLIGHT: Specifically, we propose a novel framework including a consumer-grade set-up, a complete transmission scheme and a processing pipeline, which consists of prescan modeling, pose detection and action reconstruction.
- 513, TITLE: An Interactive Design for Visualizable Person Re-Identification  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414450>  
AUTHORS: Haolin Ren, Zheng Wang, Zhixiang Wang, Lixiong Chen, Shin'ichi Satoh, Daning Hu  
HIGHLIGHT: As system operators need a comfortable access to these important elements, we introduce an interactive design of a person ReID system to visualize these quantities.
- 514, TITLE: Image and Video Restoration and Compression Artefact Removal Using a NoGAN Approach  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414451>  
AUTHORS: Filippo Marni, Marco Bertini, Leonardo Galteri, Alberto Del Bimbo  
HIGHLIGHT: In this work, we report results obtained using the NoGAN training approach and adapting the popular DeOldify architecture used for colorization, for image and video compression artefact removal and restoration.
- 515, TITLE: Beautify As You Like  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414452>  
AUTHORS: Wentao Jiang, Si Liu, Chen Gao, Ran He, Bo Li, Shuicheng Yan  
HIGHLIGHT: In this demo, we present the first fast makeup transfer system named as Fast Pose and expression robust Spatial-Aware GAN (FPSGAN).
- 516, TITLE: iDirector: An Intelligent Directing System for Live Broadcast  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414453>  
AUTHORS: Jiawei Zuo, Yue Chen, Linfang Wang, Yingwei Pan, Ting Yao, Ke Wang, Tao Mei  
HIGHLIGHT: In this paper, we demonstrate an end-to-end intelligent system for live sports broadcasting, namely iDirector, which aims to mimic the human-in-loop live broadcasting process by aggregating the input multi-camera video streaming into the final output program video (PGM video) for audience.
- 517, TITLE: Multimedia Food Logger  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414454>  
AUTHORS: Ali Rostami, Bihao Xu, Ramesh Jain  
HIGHLIGHT: We will demonstrate the complete functionality of such a system in this demo.
- 518, TITLE: A Cross-modality and Progressive Person Search System  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414455>  
AUTHORS: Xiaodong Chen, Wu Liu, Xinchun Liu, Yongdong Zhang, Tao Mei  
HIGHLIGHT: This demonstration presents an instant and progressive cross-modality person search system, called 'CMPS'.
- 519, TITLE: Binocular Multi-CNN System for Real-Time 3D Pose Estimation  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414456>  
AUTHORS: Teo T. Niemirepo, Marko Viitanen, Jarno Vanne  
HIGHLIGHT: This paper introduces the first open-source algorithm for binocular 3D pose estimation.

- 520, TITLE: An Interaction-based Video Viewing Support System using Geographical Relationships  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414457>  
AUTHORS: Itsuki Hashimoto, Yuanyuan Wang, Yukiko Kawai, Kazutoshi Sumiya  
HIGHLIGHT: Therefore, in this paper, we propose a video viewing support system to recommend supplementary information using geographical relationships based on user interaction.
- 521, TITLE: Infinity Battle: A Glance at How Blockchain Techniques Serve in a Serverless Gaming System  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414458>  
AUTHORS: Feijie Wu, Ho Yin Yuen, Henry C.B. Chan, Victor C.M. Leung, Wei Cai  
HIGHLIGHT: In this work, we present the Infinity Battle, a serverless turn-based strategy game supported by a novel Proof-of-Play consensus model.
- 522, TITLE: ConfFlow: A Tool to Encourage New Diverse Collaborations  
<https://dl.acm.org/doi/abs/10.1145/3394171.3414459>  
AUTHORS: Ekin Gedik, Hayley Hung  
HIGHLIGHT: ConfFlow is an interactive web application that allows conference participants to inspect other attendees through a visualized similarity space.
- 523, TITLE: HyFea: Winning Solution to Social Media Popularity Prediction for Multimedia Grand Challenge 2020  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416273>  
AUTHORS: Xin Lai, Yihong Zhang, Wei Zhang  
HIGHLIGHT: In this paper, we present HyFea, our winning solution to the Social Media Prediction (SMP) Challenge for multimedia grand challenge of ACM Multimedia 2020.
- 524, TITLE: A Feature Generalization Framework for Social Media Popularity Prediction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416294>  
AUTHORS: Kai Wang, Penghui Wang, Xin Chen, Qiushi Huang, Zhendong Mao, Yongdong Zhang  
HIGHLIGHT: In this paper, we propose a novel combined framework for social media popularity prediction, which accomplishes feature generalization and temporal modeling based on multi-modal feature extraction.
- 525, TITLE: Curriculum Learning for Wide Multimedia-Based Transformer with Graph Target Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416275>  
AUTHORS: Weilong Chen, Feng Hong, Chenghao Huang, Shaoliang Zhang, Rui Wang, Ruobing Xie, Feng Xia, Leyu Lin, Yanru Zhang, Yan Wang  
HIGHLIGHT: In this paper, we propose a novel approach named curriculum learning for wide multimedia-based transformer with graph target detection (CL-WMTG).
- 526, TITLE: Multimodal Deep Learning for Social Media Popularity Prediction With Attention Mechanism  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416274>  
AUTHORS: Kele Xu, Zhimin Lin, Jianqiao Zhao, Peicang Shi, Wei Deng, Huaimin Wang  
HIGHLIGHT: Inspired by the recent success of multimodal learning, we propose a novel multimodal deep learning framework for the popularity prediction task, which aims to leverage the complementary knowledge from different modalities.
- 527, TITLE: Rethinking Relation between Model Stacking and Recurrent Neural Networks for Social Media Prediction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3417332>  
AUTHORS: Chih-Chung Hsu, Wen-Hai Tseng, Hao-Ting Yang, Chia-Hsiang Lin, Chi-Hung Kao  
HIGHLIGHT: In this paper, we discover a more dominant feature representation of text information, as well as propose a single ensemble learning model to obtain the popularity scores, for social media prediction challenge.
- 528, TITLE: Video Relation Detection with Trajectory-aware Multi-modal Features  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416284>  
AUTHORS: Wentao Xie, Guanghui Ren, Si Liu  
HIGHLIGHT: In this paper, we present video relation detection with trajectory-aware multi-modal features to solve this task.
- 529, TITLE: A Strong Baseline for Multiple Object Tracking on VidOR Dataset  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416304>  
AUTHORS: Zhipeng Luo, Zhiguang Zhang, Yuehan Yao  
HIGHLIGHT: According to the above characteristics, we design a robust detection model, proposed a new deep metric learning method, and explored some useful tracking algorithms to help complete the video object detection task.

- 530, TITLE: XlanV Model with Adaptively Multi-Modality Feature Fusing for Video Captioning  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416290>  
AUTHORS: Yiqing Huang, Qiuyu Cai, Siyu Xu, Jiansheng Chen  
HIGHLIGHT: We adaptively fuse these two kinds of features in the X-Linear Attention Network Video and propose XlanV model for video captioning.
- 531, TITLE: VideoTRM: Pre-training for Video Captioning Challenge 2020  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416291>  
AUTHORS: Jingwen Chen, Hongyang Chao  
HIGHLIGHT: As a part of the submission to this challenge, we propose a Transformer based framework named VideoTRM, which consists of four modules: a textual encoder for encoding the linguistic relationship among words in the input sentence, a visual encoder for capturing the temporal dynamics in the input video, a cross-modal encoder for modeling the interactions between the two modalities (i.e., textual and visual) and a decoder for sentence generation conditioned on the input video and words generated previously.
- 532, TITLE: Multi-stage Tag Guidance Network in Video Caption  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416288>  
AUTHORS: Lanxiao Wang, Chao Shang, Heqian Qiu, Taijin Zhao, Benliu Qiu, Hongliang Li  
HIGHLIGHT: In this work, we propose a tag guidance module to learn a representation which can better build the interaction in cross-modal between visual content and textual sentences.
- 533, TITLE: Dense Scene Multiple Object Tracking with Box-Plane Matching  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416283>  
AUTHORS: Jinlong Peng, Yueyang Gu, Yabiao Wang, Chengjie Wang, Jilin Li, Feiyue Huang  
HIGHLIGHT: Following the tracking-by-detection framework, we propose the Box-Plane Matching (BPM) method to improve the MOT performance in dense scenes.
- 534, TITLE: Transductive Multi-Object Tracking in Complex Events by Interactive Self-Training  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416282>  
AUTHORS: Ancong Wu, Chengzhi Lin, Bogao Chen, Weihao Huang, Zeyu Huang, Wei-Shi Zheng  
HIGHLIGHT: We propose a transductive interactive self-training method to adapt the tracking model to unseen crowded scenes with unlabeled testing data by means of teacher-student interactive learning.
- 535, TITLE: Application of Multi-Object Tracking with Siamese Track-RCNN to the Human in Events Dataset  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416297>  
AUTHORS: Bing Shuai, Andrew Berneshawi, Manchen Wang, Chunhui Liu, Davide Modolo, Xinyu Li, Joseph Tighe  
HIGHLIGHT: Towards this, we propose Siamese Track-RCNN, a two stage detect-and-track framework which consists of three functional branches: (1) the detection branch localizes object instances; (2) the Siamese-based track branch estimates the object motion and (3) the object re-identification branch re-activates the previously terminated tracks when they re-emerge.
- 536, TITLE: Towards Accurate Human Pose Estimation in Videos of Crowded Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416299>  
AUTHORS: Shuning Chang, Li Yuan, Xuecheng Nie, Ziyuan Huang, Yichen Zhou, Yupeng Chen, Jiashi Feng, Shuicheng Yan  
HIGHLIGHT: In this paper, we focus on improving human pose estimation in videos of crowded scenes from the perspectives of exploiting temporal context and collecting new data.
- 537, TITLE: Combined Distillation Pose  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416278>  
AUTHORS: Lei Yuan, Shu Zhang, Feng Fubiao, Naike Wei, Huadong Pan  
HIGHLIGHT: In this paper, the knowledge distillation method is used to make the network predictions results act as supervision information, then a multi-stage supervision training framework is designed from shallow to deep layers.
- 538, TITLE: Deep Relationship Analysis in Video with Multimodal Feature Fusion  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416303>  
AUTHORS: Fan Yu, DanDan Wang, Beibei Zhang, Tongwei Ren  
HIGHLIGHT: In this paper, we propose a novel multimodal feature fusion method based on scene segmentation to detect the relationships between entities in a long duration video.
- 539, TITLE: Towards Using Semantic-Web Technologies for Multi-Modal Knowledge Graph Construction  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416292>

AUTHORS: Matthias Baumgartner, Luca Rossetto, Abraham Bernstein  
HIGHLIGHT: In this paper we present our approaches used in the first instance of the Deep Video Understanding Challenge, using a combination of several multi-modal detectors and an integration scheme informed by methods from the semantic web context in order to determine the capabilities limitations of currently available methods for the extraction of semantic relations between the characters and locations relevant to the narrative of a movie.

540, TITLE: Story Semantic Relationships from Multimodal Cognitions  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416305>  
AUTHORS: Vishal Anand, Raksha Ramesh, Ziyin Wang, Yijing Feng, Jiana Feng, Wenfeng Lyu, Tianle Zhu, Serena Yuan, Ching-Yung Lin  
HIGHLIGHT: We consider the problem of building semantic relationship of unseen entities from free-form multi-modal sources.

541, TITLE: ACM Multimedia BioMedia 2020 Grand Challenge Overview  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416287>  
AUTHORS: Steven A. Hicks, Vajira Thambawita, Hugo L. Hammer, Trine B. Haugen, Jorunn M. Andersen, Oliwia Witzak, P&aring;l Halvorsen, Michael A. Riegler  
HIGHLIGHT: In this year's challenge, participants are asked to develop algorithms that automatically predict the quality of a given human semen sample using a combination of visual, patient-related, and laboratory-analysis-related data.

542, TITLE: A Quantitative Comparison of Different Machine Learning Approaches for Human Spermatozoa Quality Prediction Using Multimodal Datasets  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416285>  
AUTHORS: Ming Feng, Kele Xu, Yin Wang  
HIGHLIGHT: In this paper, we make a quantitative comparison of different machine learning approaches for the human spermatozoa quality prediction task, leveraging multiple modalities dataset.

543, TITLE: Enhancing Anomaly Detection in Surveillance Videos with Transfer Learning from Action Recognition  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416298>  
AUTHORS: Kun Liu, Minzhi Zhu, Huiyuan Fu, Huadong Ma, Tat-Seng Chua  
HIGHLIGHT: In this paper, we propose to utilize transfer learning to leverage the good results from action recognition for anomaly detection in surveillance videos.

544, TITLE: Modularized Framework with Category-Sensitive Abnormal Filter for City Anomaly Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416279>  
AUTHORS: Jie Wu, Yingying Li, Wei Zhang, Yi Wu, Xiao Tan, Hongwu Zhang, Shilei Wen, Errui Ding, Guanbin Li  
HIGHLIGHT: In this paper, we propose a modularized framework to perform general and specific anomaly detection.

545, TITLE: Large Scale Hierarchical Anomaly Detection and Temporal Localization  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416302>  
AUTHORS: Soumil Kanwal, Vineet Mehta, Abhinav Dhall  
HIGHLIGHT: In this paper, we propose a multiple feature-based approach for CitySCENE challenge-based anomaly detection.

546, TITLE: Global Information Guided Video Anomaly Detection  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416277>  
AUTHORS: Hui Lv, Chunyan Xu, Zhen Cui  
HIGHLIGHT: In this paper, we propose an end-to-end Global Information Guided (GIG) anomaly detection framework for anomaly detection using the video-level annotations (i.e., weak labels).

547, TITLE: A Simple Baseline for Pose Tracking in Videos of Crowded Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416300>  
AUTHORS: Li Yuan, Shuning Chang, Ziyuan Huang, Yichen Zhou, Yupeng Chen, Xuecheng Nie, Francis E.H. Tay, Jiashi Feng, Shuicheng Yan  
HIGHLIGHT: This paper presents our solution to ACM MM challenge: Large-scale Human-centric Video Analysis in Complex Events[13]; specifically, here we focus on Track3: Crowd Pose Tracking in Complex Events.

548, TITLE: HiEve ACM MM Grand Challenge 2020: Pose Tracking in Crowded Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416295>  
AUTHORS: Lumin Xu, Ruihan Xu, Sheng Jin  
HIGHLIGHT: We propose a simple yet effective top-down crowd pose tracking algorithm.



549, TITLE: Toward Accurate Person-level Action Recognition in Videos of Crowded Scenes  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416301>  
AUTHORS: Li Yuan, Yichen Zhou, Shuning Chang, Ziyuan Huang, Yupeng Chen, Xuecheng Nie, Tao Wang, Jiashi Feng, Shuicheng Yan  
HIGHLIGHT: In this paper, we focus on improving spatio-temporal action recognition by fully-utilizing the information of scenes and collecting new data.

550, TITLE: Person-level Action Recognition in Complex Events via TSD-TSM Networks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416276>  
AUTHORS: Yanbin Hao, Zi-Niu Liu, Hao Zhang, Bin Zhu, Jingjing Chen, Yu-Gang Jiang, Chong-Wah Ngo  
HIGHLIGHT: In this paper, we present a simple yet efficient pipeline for this task, referred to as TSD-TSM networks.

551, TITLE: Group-Skeleton-Based Human Action Recognition in Complex Events  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416280>  
AUTHORS: Tingtian Li, Zixun Sun, Xiao Chen  
HIGHLIGHT: In this paper, we propose a novel group-skeleton-based human action recognition method in complex events.

552, TITLE: Attention Based Beauty Product Retrieval Using Global and Local Descriptors  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416289>  
AUTHORS: Jun Yu, Guochen Xie, Mengyan Li, Haonian Xie, Xinlong Hao, Fang Gao, Feng Shuang  
HIGHLIGHT: In this paper, we first introduce attention mechanism into a global image descriptor, i.e., Maximum Activation of Convolutions (MAC), and propose Attention-based MAC (AMAC).

553, TITLE: Multi-Feature Fusion Method Based on Salient Object Detection for Beauty Product Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416272>  
AUTHORS: Runming Yan, Yongchun Lin, Zhichao Deng, Liang Lei, Chudong Xu  
HIGHLIGHT: In this paper, we propose a multi-feature fusion method based on salient object detection to improve retrieval performance.

554, TITLE: Attention-driven Unsupervised Image Retrieval for Beauty Products with Visual and Textual Clues  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416271>  
AUTHORS: Jingwen Hou, Sijie Ji, Annan Wang  
HIGHLIGHT: Therefore, we propose a search method utilizing both images and product descriptions in this work.

555, TITLE: Learning Visual Features from Product Title for Image Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416296>  
AUTHORS: Fangxiang Feng, Tianrui Niu, Ruifan Li, Xiaojie Wang, Huixing Jiang  
HIGHLIGHT: In this paper, we utilize the easily accessible text information, that is, the product title, as a supervised signal to learn the features of the product image.

556, TITLE: Learning to Remember Beauty Products  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416281>  
AUTHORS: Toan H. Vu, An Dang, Jia-Ching Wang  
HIGHLIGHT: This paper develops a deep learning model for the beauty product image retrieval problem.

557, TITLE: Multi-Scale Generalized Attention-Based Regional Maximum Activation of Convolutions for Beauty Product Retrieval  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416293>  
AUTHORS: Kele Xu, Yuzhong Liu, Ming Feng, Jianqiao Zhao, Huaimin Wang, Hengxing Cai  
HIGHLIGHT: In this paper, we propose a novel descriptors, named Multi-Scale Generalized Attention-Based Regional Maximum Activation of Convolutions (MS-GRMAC).

558, TITLE: Low-level Optimizations for Faster Mobile Deep Learning Inference Frameworks  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416516>  
AUTHORS: Mathieu Febvay  
HIGHLIGHT: In this paper, we present the performance benchmark of four popular open-source deep learning inference frameworks used on mobile devices on three different convolutional neural network models.

559, TITLE: Deep Neural Networks for Predicting Affective Responses from Movies  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416517>  
AUTHORS: Ha Thi Phuong Thao  
HIGHLIGHT: In this work, we develop deep neural networks for predicting affective responses from movies taking both audio and video streams into account.

560, TITLE: Learning Self-Supervised Multimodal Representations of Human Behaviour  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416518>  
AUTHORS: Abhinav Shukla  
HIGHLIGHT: In this extended abstract, I present the direction of research that I have followed during the first half of my PhD, along with ideas and work in progress for the second half.

561, TITLE: Multi-person Pose Estimation in Complex Physical Interactions  
<https://dl.acm.org/doi/abs/10.1145/3394171.3416519>  
AUTHORS: Wen Guo  
HIGHLIGHT: In this work we input the initial pose along with its interactees into a recurrent network to refine the pose of the person-of-interest, and we demonstrate the effectiveness of our method in the MuPoTS dataset, setting the new state-of-the-art on it.