1. arXiv:2103.13990 [pdf, other]

   cs.CV

   More Photos are All You Need: Semi-Supervised Learning for Fine-Grained Sketch Based Image Retrieval

   Authors: Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Aneeshan Sain, Yongxin Yang, Tao Xiang, Yi-Zhe Song

   Abstract: A fundamental challenge faced by existing Fine-Grained Sketch-Based Image Retrieval (FG-SBIR) models is the data scarcity -- model performances are largely bottlenecked by the lack of sketch-photo pairs. Whilst the number of photos can be easily scaled, each corresponding sketch still needs to be individually produced. In this paper, we aim to mitigate such an upper-bound on sketch data, and study… ▽ More

   Submitted 25 March, 2021; originally announced March 2021.

   Comments: IEEE Conf. on Computer Vision and Pattern Recognition (**CVPR**), **2021** Code : https://github.com/AyanKumarBhunia/semisupervised-FGSBIR

2. arXiv:2103.13886 [pdf, other]

   cs.CV cs.LG

   Robust and Accurate Object Detection via Adversarial Learning

   Authors: Xiangning Chen, Cihang Xie, Mingxing Tan, Li Zhang, Cho-Jui Hsieh, Boqing Gong

   Abstract: Data augmentation has become a de facto component for training high-performance deep image classifiers, but its potential is under-explored for object detection. Noting that most state-of-the-art object detectors benefit from fine-tuning a pre-trained classifier, we first study how the classifiers' gains from various data augmentations transfer to object detection. The results are discouraging; th… ▽ More

   Submitted 23 March, 2021; **originally announced** March 2021.

   Comments: **CVPR 2021**

3. arXiv:2103.13757 [pdf, other]

   cs.CV

   I^3Net: Implicit Instance-Invariant Network for Adapting One-Stage Object Detectors

   Authors: Chaoqi Chen, Zebiao Zheng, Yue Huang, Xinghao Ding, Yizhou Yu

   Abstract: Recent works on two-stage cross-domain detection have widely explored the local feature patterns to achieve more accurate adaptation results. These methods heavily rely on the region proposal mechanisms and ROI-based instance-level features to design fine-grained feature alignment modules with respect to the foreground objects. However, for one-stage detectors, it is hard or even impossible to obt… ▽ More

   Submitted 25 March, 2021; **originally announced** March 2021.

   Comments: Accepted by **CVPR 2021**

4. arXiv:2103.13716 [pdf, other]

   cs.CV

   Vectorization and Rasterization: Self-Supervised Learning for Sketch and Handwriting

   Authors: Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Yongxin Yang, Timothy M. Hospedales, Tao Xiang, Yi-Zhe Song

   Abstract: Self-supervised learning has gained prominence due to its efficacy at learning powerful representations from unlabelled data that achieve excellent performance on many challenging downstream tasks. However supervision-free pre-text tasks are challenging to design and usually modality specific. Although there is a rich literature of self-supervised methods for either spatial (such as images) or tem… ▽ More

   Submitted 25 March, 2021; originally announced March 2021.

   Comments: IEEE Conf. on Computer Vision and Pattern Recognition (**CVPR**), **2021** Code : https://github.com/AyanKumarBhunia/Self-Supervised-Learning-for-Sketch

5. arXiv:2103.13696 [pdf, other]

   cs.CV cs.LG

   SSLayout360: Semi-Supervised Indoor Layout Estimation from 360∘ Panorama

   Authors: Phi Vu Tran

   Abstract: Recent years have seen flourishing research on both semi-supervised learning and 3D room layout reconstruction. In this work, we explore the intersection of these two fields to advance the research objective of enabling more accurate 3D indoor scene modeling with less labeled data. We propose the first approach to learn representations of room corners and boundaries by using a combination of label… ▽ More

   Submitted 25 March, 2021; **originally announced** March 2021.

   Comments: **CVPR 2021**. File size 37MB

6. arXiv:2103.13660 [pdf, other]

   eess.IV cs.CV

   Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation

   Authors: Yuntong Ye, Yi Chang, Hanyu Zhou, Luxin Yan

   Abstract: Existing deep learning-based image deraining methods have achieved promising performance for synthetic rainy images, typically rely on the pairs of sharp images and simulated rainy counterparts. However, these methods suffer from significant performance drop when facing the real rain, because of the huge gap between the simplified synthetic rain and the complex real rain. In this work, we argue th… ▽ More

   Submitted 25 March, 2021; **originally announced** March 2021.

7. arXiv:2103.13558 [pdf, other]

   cs.LG cs.AI cs.CV

   Efficient Feature Transformations for Discriminative and Generative Continual Learning

   Authors: Vinay Kumar Verma, Kevin J Liang, Nikhil Mehta, Piyush Rai, Lawrence Carin

   Abstract: As neural networks are increasingly being applied to real-world applications, mechanisms to address distributional shift and sequential task learning without forgetting are critical. Methods incorporating network expansion have shown promise by naturally adding model capacity for learning new tasks while simultaneously avoiding catastrophic forgetting. However, the growth in the number of addition… ▽ More

8. arXiv:2103.13447 [pdf, other]

   cs.CV

   DRANet: Disentangling Representation and Adaptation Networks for Unsupervised Cross-Domain Adaptation

   Authors: Seunghun Lee, Sunghyun Cho, Sunghoon Im

   Abstract: In this paper, we present DRANet, a network architecture that disentangles image representations and transfers the visual attributes in a latent space for unsupervised cross-domain adaptation. Unlike the existing domain adaptation methods that learn associated features sharing a domain, DRANet preserves the distinctiveness of each domain's characteristics. Our model encodes individual representati… ▽ More

9. arXiv:2103.13425 [pdf, other]

   cs.CV cs.AI cs.LG

   Diverse Branch Block: Building a Convolution as an Inception-like Unit

   Authors: Xiaohan Ding, Xiangyu Zhang, Jungong Han, Guiguang Ding

   Abstract: We propose a universal building block of Convolutional Neural Network (ConvNet) to improve the performance without any inference-time costs. The block is named Diverse Branch Block (DBB), which enhances the representational capacity of a single convolution by combining diverse branches of different scales and complexities to enrich the feature space, including sequences of convolutions, multi-scal… ▽ More

10. arXiv:2103.13372 [pdf, other]

cs.CV cs.LG

Affective Processes: stochastic modelling of temporal context for emotion and facial expression recognition

Authors: Enrique Sanchez, Mani Kumar Tellamekala, Michel Valstar, Georgios Tzimiropoulos

Abstract: Temporal context is key to the recognition of expressions of emotion. Existing methods, that rely on recurrent or self-attention models to enforce temporal consistency, work on the feature level, ignoring the task-specific temporal dependencies, and fail to model context uncertainty. To alleviate these issues, we build upon the framework of Neural Processes to propose a method for apparent emotion… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

11. arXiv:2103.13258 [pdf, other]

cs.CV

Dynamic Slimmable Network

Authors: Changlin Li, Guangrun Wang, Bing Wang, Xiaodan Liang, Zhihui Li, Xiaojun Chang

Abstract: Current dynamic networks and dynamic pruning methods have shown their promising capability in reducing theoretical computation complexity. However, dynamic sparse patterns on convolutional filters fail to achieve actual acceleration in real-world implementation, due to the extra burden of indexing, weight-copying, or zero-masking. Here, we explore a dynamic network slimming regime, named Dynamic S… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021** as an Oral Presentation

12. arXiv:2103.13225 [pdf, other]

cs.CV

Structure-Aware Face Clustering on a Large-Scale Graph with 107 Nodes

Authors: Shuai Shen, Wanhua Li, Zheng Zhu, Guan Huang, Dalong Du, Jiwen Lu, Jie Zhou

Abstract: Face clustering is a promising method for annotating unlabeled face images. Recent supervised approaches have boosted the face clustering accuracy greatly, however their performance is still far from satisfactory. These methods can be roughly divided into global-based and local-based ones. Global-based methods suffer from the limitation of training data scale, while local-based ones are difficult… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: This work has been accepted by the **CVPR 2021**. Project: https://sstzal.github.io/STAR-FC/

13. arXiv:2103.13164 [pdf, other]

cs.CV

M3DSSD: Monocular 3D Single Stage Object Detector

Authors: Shujie Luo, Hang Dai, Ling Shao, Yong Ding

Abstract: In this paper, we propose a Monocular 3D Single Stage object Detector (M3DSSD) with feature alignment and asymmetric non-local attention. Current anchor-based monocular 3D object detection methods suffer from feature mismatching. To overcome this, we propose a two-step feature alignment approach. In the first step, the shape alignment is performed to enable the receptive field of the feature map t… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

14. arXiv:2103.13141 [pdf, other]

cs.CV

Temporal Context Aggregation Network for Temporal Action Proposal Refinement

Authors: Zhiwu Qing, Haisheng Su, Weihao Gan, Dongliang Wang, Wei Wu, Xiang Wang, Yu Qiao, Junjie Yan, Changxin Gao, Nong Sang

**Abstract**: …place in the **CVPR** 2020 - HACS challenge leaderboard on temporal action localization task. ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: Accepted by CVPR2021

15. arXiv:2103.13061 [pdf, other]

cs.CV

Revamping Cross-Modal Recipe Retrieval with Hierarchical Transformers and Self-supervised Learning

Authors: Amaia Salvador, Erhan Gundogdu, Loris Bazzani, Michael Donoser

Abstract: Cross-modal recipe retrieval has recently gained substantial attention due to the importance of food in people's lives, as well as the availability of vast amounts of digital cooking recipes and food images to train machine learning models. In this work, we revisit existing approaches for cross-modal recipe retrieval and propose a simplified end-to-end model based on well established and high perf… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

16. arXiv:2103.13021 [pdf, other]

cs.LG cs.CV

Convex Online Video Frame Subset Selection using Multiple Criteria for Data Efficient Autonomous Driving

Authors: <u>Soumi Das</u>, <u>Harikrishna Patibandla</u>, <u>Suparna Bhattacharya</u>, <u>Kshounis Bera</u>, <u>Niloy Ganguly</u>, <u>Sourangshu Bhattacharya</u>

Abstract: Training vision-based Urban Autonomous driving models is a challenging problem, which is highly researched in recent times. Training such models is a data-intensive task requiring the storage and processing of vast volumes of (possibly redundant) driving video data. In this paper, we study the problem of developing data-efficient autonomous driving systems. In this context, we study the problem of… ▽ More

Submitted 24 March, 2021; **originally announced** March 2021.

Comments: Submitted to **CVPR** 2020

ACM Class: I.2.9; G.1.6

17. <u>arXiv:2103.12944</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Scene-Intuitive Agent for Remote Embodied Visual Grounding

Authors: <u>Xiangru Lin</u>, <u>Guanbin Li</u>, <u>Yizhou Yu</u>

Abstract: Humans learn from life events to form intuitions towards the understanding of visual environments and languages. Envision that you are instructed by a high-level instruction, "Go to the bathroom in the master bedroom and replace the blue towel on the left wall", what would you possibly do to carry out the task? Intuitively, we comprehend the semantics of the instruction to form an overview of wher… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

18. <u>arXiv:2103.12934</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Efficient Regional Memory Network for Video Object Segmentation

Authors: <u>Haozhe Xie</u>, <u>Hongxun Yao</u>, <u>Shangchen Zhou</u>, <u>Shengping Zhang</u>, <u>Wenxiu Sun</u>

Abstract: Recently, several Space-Time Memory based networks have shown that the object cues (e.g. video frames as well as the segmented object masks) from the past frames are useful for segmenting objects in the current frame. However, these methods exploit the information from the memory by global-to-global matching between the current and past frames, which lead to mismatching to similar objects and high… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

19. <u>arXiv:2103.12886</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Weakly Supervised Instance Segmentation for Videos with Temporal Mask Consistency

Authors: Qing Liu, Vignesh Ramanathan, Dhruv Mahajan, Alan Yuille, Zhenheng Yang

Abstract: Weakly supervised instance segmentation reduces the cost of annotations required to train models. However, existing approaches which rely only on image-level class labels predominantly suffer from errors due to (a) partial segmentation of objects and (b) missing object predictions. We show that these issues can be better addressed by training with weakly labeled videos instead of images. In videos… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: 14 pages, 8 figures, accepted by **CVPR 2021**

20. arXiv:2103.12731 [pdf, other]

cs.CV

Scaling Local Self-Attention For Parameter Efficient Visual Backbones

Authors: Ashish Vaswani, Prajit Ramachandran, Aravind Srinivas, Niki Parmar, Blake Hechtman, Jonathon Shlens

Abstract: Self-attention has the promise of improving computer vision systems due to parameter-independent scaling of receptive fields and content-dependent interactions, in contrast to parameter-dependent scaling and content-independent interactions of convolutions. Self-attention models have recently been shown to have encouraging improvements on accuracy-parameter trade-offs compared to baseline convolut… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** Oral

21. arXiv:2103.12605 [pdf, other]

cs.CV

MonoRUn: Monocular 3D Object Detection by Reconstruction and Uncertainty Propagation

Authors: Hansheng Chen, Yuyao Huang, Wei Tian, Zhong Gao, Lu Xiong

Abstract: Object localization in 3D space is a challenging aspect in monocular 3D object detection. Recent advances in 6DoF pose estimation have shown that predicting dense 2D-3D correspondence maps between image and object 3D model and then estimating object pose via Perspective-n-Point (PnP) algorithm can achieve remarkable localization accuracy. Yet these methods rely on training with ground truth of obj… ▽ More

Submitted 24 March, 2021; v1 submitted 23 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

22. arXiv:2103.12579 [pdf, other]

cs.CV

MetaSAug: Meta Semantic Augmentation for Long-Tailed Visual Recognition

Authors: <u>Shuang Li</u>, <u>Kaixiong Gong</u>, <u>Chi Harold Liu</u>, <u>Yulin Wang</u>, <u>Feng Qiao</u>, <u>Xinjing Cheng</u>

Abstract: Real-world training data usually exhibits long-tailed distribution, where several majority classes have a significantly larger number of samples than the remaining minority classes. This imbalance degrades the performance of typical supervised learning algorithms designed for balanced training sets. In this paper, we address this issue by augmenting minority classes with a recently proposed implic… ▽ More

Submitted 25 March, 2021; v1 submitted 23 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

23. <u>arXiv:2103.12562</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Transferable Semantic Augmentation for Domain Adaptation

Authors: <u>Shuang Li</u>, <u>Mixue Xie</u>, <u>Kaixiong Gong</u>, <u>Chi Harold Liu</u>, <u>Yulin Wang</u>, <u>Wei Li</u>

Abstract: Domain adaptation has been widely explored by transferring the knowledge from a label-rich source domain to a related but unlabeled target domain. Most existing domain adaptation algorithms attend to adapting feature representations across two domains with the guidance of a shared source-supervised classifier. However, such classifier limits the generalization ability towards unlabeled target reco… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: Accepted as **CVPR 2021**. The code is publicly available at <u>https://github.com/BIT-DA/TSA</u>

24. <u>arXiv:2103.12366</u> [<u>pdf</u>, <u>ps</u>, <u>other</u>]

cs.CV

Group-aware Label Transfer for Domain Adaptive Person Re-identification

Authors: <u>Kecheng Zheng</u>, <u>Wu Liu</u>, <u>Lingxiao He</u>, <u>Tao Mei</u>, <u>Jiebo Luo</u>, <u>Zheng-Jun Zha</u>

Abstract: Unsupervised Domain Adaptive (UDA) person re-identification (ReID) aims at adapting the model trained on a labeled source-domain dataset to a target-domain dataset without any further annotations. Most successful UDA-ReID approaches combine clustering-based pseudo-label prediction with representation learning and perform the two steps in an alternating fashion. However, offline interaction between… ▽ More

Submitted 23 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

25. <u>arXiv:2103.12266</u> [<u>pdf</u>, <u>other</u>]

cs.CV cs.GR

Deep Implicit Moving Least-Squares Functions for 3D Reconstruction

Authors: <u>Shi-Lin Liu</u>, <u>Hao-Xiang Guo</u>, <u>Hao Pan</u>, <u>Peng-Shuai Wang</u>, <u>Xin Tong</u>, <u>Yang Liu</u>

Abstract: Point set is a flexible and lightweight representation widely used for 3D deep learning. However, their discrete nature prevents them from representing continuous and fine geometry, posing a major issue for learning-based shape generation. In this work, we turn the discrete point sets into smooth surfaces by introducing the well-known implicit moving least-squares (IMLS) surface formulation, which… ▽ More

Submitted 22 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**, Code: https://github.com/Andy97/DeepMLS

26. arXiv:2103.12204 [pdf, other]

cs.CV cs.AI cs.CL cs.MM

Human-like Controllable Image Captioning with Verb-specific Semantic Roles

Authors: Long Chen, Zhihong Jiang, Jun Xiao, Wei Liu

Abstract: Controllable Image Captioning (CIC) -- generating image descriptions following designated control signals -- has received unprecedented attention over the last few years. To emulate the human ability in controlling caption generation, current CIC studies focus exclusively on control signals concerning objective properties, such as contents of interest or descriptive patterns. However, we argue tha… ▽ More

Submitted 22 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**. The code is available at: https://github.com/mad-red/VSR-guided-CIC

27. arXiv:2103.11897 [pdf, other]

cs.CV

Context-Aware Layout to Image Generation with Enhanced Object Appearance

Authors: Sen He, Wentong Liao, Michael Ying Yang, Yongxin Yang, Yi-Zhe Song, Bodo Rosenhahn, Tao Xiang

Abstract: A layout to image (L2I) generation model aims to generate a complicated image containing multiple objects (things) against natural background (stuff), conditioned on a given layout. Built upon the recent advances in generative adversarial networks (GANs), existing L2I models have made great progress. However, a close inspection of their generated images reveals two major limitations: (1) the objec… ▽ More

Submitted 22 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

28. arXiv:2103.11781 [pdf, other]

cs.CV

Dynamic Metric Learning: Towards a Scalable Metric Space to Accommodate Multiple Semantic Scales

Authors: Yifan Sun, Yuke Zhu, Yuhan Zhang, Pengkun Zheng, Xi Qiu, Chi Zhang, Yichen Wei

Abstract: This paper introduces a new fundamental characteristic, \ie, the dynamic range, from real-world metric tools to deep visual recognition. In metrology, the dynamic range is a basic quality of a metric tool, indicating its flexibility to accommodate various scales. Larger dynamic range offers higher flexibility. In visual recognition, the multiple scale problem also exist. Different visual concepts… ▽ More

Submitted 22 March, 2021; **originally announced** March 2021.

Comments: 8pages, accepted by **CVPR 2021**

29. arXiv:2103.11681 [pdf, other]

cs.CV

Transformer Meets Tracker: Exploiting Temporal Context for Robust Visual Tracking

Authors: Ning Wang, Wengang Zhou, Jie Wang, Houqaing Li

Abstract: In video object tracking, there exist rich temporal contexts among successive frames, which have been largely overlooked in existing trackers. In this work, we bridge the individual video frames and explore the temporal contexts across them via a transformer architecture for robust object tracking. Different from classic usage of the transformer in natural language processing tasks, we separate it… ▽ More

Submitted 24 March, 2021; v1 submitted 22 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021** (Oral)

30. arXiv:2103.11617 [pdf, other]

cs.CV

Anchor-Free Person Search

Authors: Yichao Yan, Jingpeng Li, Jie Qin, Song Bai, Shengcai Liao, Li Liu, Fan Zhu, Ling Shao

Abstract: Person search aims to simultaneously localize and identify a query person from realistic, uncropped images, which can be regarded as the unified task of pedestrian detection and person re-identification (re-id). Most existing works employ two-stage detectors like Faster-RCNN, yielding encouraging accuracy but with high computational overhead. In this work, we present the Feature-Aligned Person Sea… ▽ More

Submitted 22 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

31. arXiv:2103.11555 [pdf, other]

cs.CV

Context-aware Biaffine Localizing Network for Temporal Sentence Grounding

Authors: Daizong Liu, Xiaoye Qu, Jianfeng Dong, Pan Zhou, Yu Cheng, Wei Wei, Zichuan Xu, Yulai Xie

Abstract: This paper addresses the problem of temporal sentence grounding (TSG), which aims to identify the temporal boundary of a specific segment from an untrimmed video by a sentence query. Previous works either compare pre-defined candidate segments with the query and select the best one by ranking, or directly regress the boundary timestamps of the target segment. In this paper, we propose a novel loca… ▽ More

Submitted 21 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

32. <u>arXiv:2103.11264</u> [<u>pdf</u>, <u>other</u>]

cs.CV cs.AI cs.LG

Temporally-Weighted Hierarchical Clustering for Unsupervised Action Segmentation

Authors: <u>M. Saquib Sarfraz</u>, <u>Naila Murray</u>, <u>Vivek Sharma</u>, <u>Ali Diba</u>, <u>Luc Van Gool</u>, <u>Rainer Stiefelhagen</u>

Abstract: Action segmentation refers to inferring boundaries of semantically consistent visual concepts in videos and is an important requirement for many video understanding tasks. For this and other video understanding tasks, supervised approaches have achieved encouraging performance but require a high volume of detailed frame-level annotations. We present a fully automatic and unsupervised approach for… ▽ More

Submitted 25 March, 2021; v1 submitted 20 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

33. <u>arXiv:2103.10895</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Sewer-ML: A Multi-Label Sewer Defect Classification Dataset and Benchmark

Authors: <u>Joakim Bruslund Haurum</u>, <u>Thomas B. Moeslund</u>

Abstract: Perhaps surprisingly sewerage infrastructure is one of the most costly infrastructures in modern society. Sewer pipes are manually inspected to determine whether the pipes are defective. However, this process is limited by the number of qualified inspectors and the time it takes to inspect a pipe. Automatization of this process is therefore of high interest. So far, the success of computer vision… ▽ More

Submitted 19 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**. Project webpage: <u>https://vap.aau.dk/sewer-ml/</u>

34. <u>arXiv:2103.10814</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Skeleton Merger: an Unsupervised Aligned Keypoint Detector

Authors: <u>Ruoxi Shi</u>, <u>Zhengrong Xue</u>, <u>Yang You</u>, <u>Cewu Lu</u>

Abstract: Detecting aligned 3D keypoints is essential under many scenarios such as object tracking, shape retrieval and robotics. However, it is generally hard to prepare a high-quality dataset for all types of objects due to the ambiguity of keypoint itself. Meanwhile, current unsupervised detectors are unable to generate aligned keypoints with good coverage. In this paper, we propose an unsupervised align… ▽ More

Submitted 19 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

35. arXiv:2103.10583 [pdf, other]

cs.CV

Dynamic Transfer for Multi-Source Domain Adaptation

Authors: Yunsheng Li, Lu Yuan, Yinpeng Chen, Pei Wang, Nuno Vasconcelos

Abstract: Recent works of multi-source domain adaptation focus on learning a domain-agnostic model, of which the parameters are static. However, such a static model is difficult to handle conflicts across multiple domains, and suffers from a performance degradation in both source domains and target domain. In this paper, we present dynamic transfer to address domain conflicts, where the model parameters are… ▽ More

Submitted 18 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

36. arXiv:2103.10571 [pdf, other]

cs.CV

Generic Perceptual Loss for Modeling Structured Output Dependencies

Authors: Yifan Liu, Hao Chen, Yu Chen, Wei Yin, Chunhua Shen

Abstract: The perceptual loss has been widely used as an effective loss term in image synthesis tasks including image super-resolution, and style transfer. It was believed that the success lies in the high-level perceptual feature representations extracted from CNNs pretrained with a large set of images. Here we reveal that, what matters is the network structure instead of the trained weights. Without any l… ▽ More

Submitted 18 March, 2021; originally announced March 2021.

Comments: Accepted to Proc. IEEE Conf. Computer Vision and Pattern Recognition (**CVPR**), **2021**

37. arXiv:2103.10559 [pdf, other]

cs.CV

CDFI: Compression-Driven Network Design for Frame Interpolation

Authors: Tianyu Ding, Luming Liang, Zhihui Zhu, Ilya Zharkov

Abstract: DNN-based frame interpolation--that generates the intermediate frames given two consecutive frames--typically relies on heavy model architectures with a huge number of features, preventing them from being deployed on systems with limited resources, e.g., mobile devices. We propose a compression-driven network design for frame interpolation (CDFI), that leverages model pruning through sparsity-indu… ▽ More

Submitted 18 March, 2021; originally announced March 2021.

Comments: To be published in the proceedings of **2021** IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**)

38. arXiv:2103.10429 [pdf, other]

cs.CV

Neural Parts: Learning Expressive 3D Shape Abstractions with Invertible Neural Networks

Authors: Despoina Paschalidou, Angelos Katharopoulos, Andreas Geiger, Sanja Fidler

Abstract: Impressive progress in 3D shape extraction led to representations that can capture object geometries with high fidelity. In parallel, primitive-based methods seek to represent objects as semantically consistent part arrangements. However, due to the simplicity of existing primitive representations, these methods fail to accurately reconstruct 3D shapes using a small number of primitives/parts. We… ▽ More

Submitted 18 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

39. arXiv:2103.10391 [pdf, other]

cs.CV

Learning to Recommend Frame for Interactive Video Object Segmentation in the Wild

Authors: Zhaoyuan Yin, Jia Zheng, Weixin Luo, Shenhan Qian, Hanling Zhang, Shenghua Gao

Abstract: This paper proposes a framework for the interactive video object segmentation (VOS) in the wild where users can choose some frames for annotations iteratively. Then, based on the user annotations, a segmentation algorithm refines the masks. The previous interactive VOS paradigm selects the frame with some worst evaluation metric, and the ground truth is required for calculating the evaluation metr… ▽ More

Submitted 18 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

40. arXiv:2103.10095 [pdf, other]

cs.CV

On Semantic Similarity in Video Retrieval

Authors: Michael Wray, Hazel Doughty, Dima Damen

Abstract: Current video retrieval efforts all found their evaluation on an instance-based assumption, that only a single caption is relevant to a query video and vice versa. We demonstrate that this assumption results in performance comparisons often not indicative of models' retrieval capabilities. We propose a move to semantic similarity video retrieval, where (i) multiple videos/captions can be deemed eq… ▽ More

Submitted 18 March, 2021; **originally announced** March 2021.

Comments: Accepted in **CVPR 2021**. Project Page: https://mwray.github.io/SSVR/

41. arXiv:2103.10022 [pdf, other]

cs.CV

Generating Diverse Structure for Image Inpainting With Hierarchical VQ-VAE

Authors: Jialun Peng, Dong Liu, Songcen Xu, Houqiang Li

Abstract: Given an incomplete image without additional constraint, image inpainting natively allows for multiple solutions as long as they appear plausible. Recently, multiplesolution inpainting methods have been proposed and shown the potential of generating diverse results. However, these methods have difficulty in ensuring the quality of each solution, e.g. they produce distorted structure and/or blurry… ▽ More

Submitted 18 March, 2021; **originally announced** March 2021.

Comments: Accepted in **CVPR 2021**

42. arXiv:2103.09458 [pdf, other]

cs.CV

Learning Discriminative Prototypes with Dynamic Time Warping

Authors: Xiaobin Chang, Frederick Tung, Greg Mori

Abstract: Dynamic Time Warping (DTW) is widely used for temporal data processing. However, existing methods can neither learn the discriminative prototypes of different classes nor exploit such prototypes for further analysis. We propose Discriminative Prototype DTW (DP-DTW), a novel method to learn class-specific discriminative prototypes for temporal recognition tasks. DP-DTW shows superior performance co… ▽ More

Submitted 17 March, 2021; originally announced March 2021.

Comments: **CVPR**'21 preview, 10 pages, 8 figures

43. arXiv:2103.09213 [pdf, other]

cs.CV

Back to the Feature: Learning Robust Camera Localization from Pixels to Pose

Authors: Paul-Edouard Sarlin, Ajaykumar Unagar, Måns Larsson, Hugo Germain, Carl Toft, Viktor Larsson, Marc Pollefeys, Vincent Lepetit, Lars Hammarstrand, Fredrik Kahl, Torsten Sattler

Abstract: Camera pose estimation in known scenes is a 3D geometry task recently tackled by multiple learning algorithms. Many regress precise geometric quantities, like poses or 3D points, from an input image. This either fails to generalize to new viewpoints or ties the model parameters to a specific scene. In this paper, we go Back to the Feature: we argue that deep networks should focus on learning robus… ▽ More

Submitted 16 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

44. arXiv:2103.09096 [pdf, other]

cs.CV

Frequency-aware Discriminative Feature Learning Supervised by Single-Center Loss for Face Forgery Detection

Authors: Jiaming Li, Hongtao Xie, Jiahong Li, Zhongyuan Wang, Yongdong Zhang

Abstract: Face forgery detection is raising ever-increasing interest in computer vision since facial manipulation technologies cause serious worries. Though recent works have reached sound achievements, there are still unignorable problems: a) learned features supervised by softmax loss are separable but not discriminative enough, since softmax loss does not explicitly encourage intra-class compactness and… ▽ More

Submitted 16 March, 2021; originally announced March 2021.

Comments: 10 pages,6 figures;**cvpr** accept

45. arXiv:2103.08907 [pdf, other]

cs.CV

BBAM: Bounding Box Attribution Map for Weakly Supervised Semantic and Instance Segmentation

Authors: Jungbeom Lee, Jihun Yi, Chaehun Shin, Sungroh Yoon

Abstract: Weakly supervised segmentation methods using bounding box annotations focus on obtaining a pixel-level mask from each box containing an object. Existing methods typically depend on a class-agnostic mask generator, which operates on the low-level information intrinsic to an image. In this work, we utilize higher-level information from the behavior of a trained object detector, by seeking the smalle… ▽ More

Submitted 16 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

46. arXiv:2103.08896 [pdf, other]

cs.CV

Anti-Adversarially Manipulated Attributions for Weakly and Semi-Supervised Semantic Segmentation

Authors: Jungbeom Lee, Eunji Kim, Sungroh Yoon

Abstract: Weakly supervised semantic segmentation produces a pixel-level localization from a classifier, but it is likely to restrict its focus to a small discriminative region of the target object. AdvCAM is an attribution map of an image that is manipulated to increase the classification score. This manipulation is realized in an anti-adversarial manner, which perturbs the images along pixel gradients in… ▽ More

Submitted 16 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

47. arXiv:2103.08808 [pdf, other]

cs.CV

Track to Detect and Segment: An Online Multi-Object Tracker

Authors: Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, Junsong Yuan

Abstract: Most online multi-object trackers perform object detection stand-alone in a neural net without any input from tracking. In this paper, we present a new online joint detection and tracking model, TraDeS (TRAck to DEtect and Segment), exploiting tracking clues to assist detection end-to-end. TraDeS infers object tracking offset by a cost volume, which is used to propagate previous object features fo… ▽ More

Submitted 15 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

48. arXiv:2103.08468 [pdf, other]

cs.CV cs.SD

Beyond Image to Depth: Improving Depth Prediction using Echoes

Authors: Kranti Kumar Parida, Siddharth Srivastava, Gaurav Sharma

Abstract: We address the problem of estimating depth with multi modal audio visual data. Inspired by the ability of animals, such as bats and dolphins, to infer distance of objects with echolocation, some recent methods have utilized echoes for depth estimation. We propose an end-to-end deep learning based pipeline utilizing RGB images, binaural echoes and estimated material properties of various objects wi… ▽ More

Submitted 15 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

49. arXiv:2103.08292 [pdf, other]

cs.CV

Rotation Coordinate Descent for Fast Globally Optimal Rotation Averaging

Authors: Álvaro Parra, Shin-Fang Chng, Tat-Jun Chin, Anders Eriksson, Ian Reid

Abstract: Under mild conditions on the noise level of the measurements, rotation averaging satisfies strong duality, which enables global solutions to be obtained via semidefinite programming (SDP) relaxation. However, generic solvers for SDP are rather slow in practice, even on rotation averaging instances of moderate size, thus developing specialised algorithms is vital. In this paper, we present a fast a… ▽ More

Submitted 15 March, 2021; v1 submitted 15 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021** as an oral presentation

50. arXiv:2103.08273 [pdf, other]

cs.CV

Refine Myself by Teaching Myself: Feature Refinement via Self-Knowledge Distillation

Authors: Mingi Ji, Seungjae Shin, Seunghyun Hwang, Gibeom Park, Il-Chul Moon

Abstract: Knowledge distillation is a method of transferring the knowledge from a pretrained complex teacher model to a student model, so a smaller network can replace a large teacher network at the deployment stage. To reduce the necessity of training a large teacher model, the recent literatures introduced a self-knowledge distillation, which trains a student network progressively to distill its own knowl… ▽ More

Submitted 15 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

1. arXiv:2103.08204 [pdf, other]

   cs.CV

   3DCaricShop: A Dataset and A Baseline Method for Single-view 3D Caricature Face Reconstruction

   Authors: Yuda Qiu, Xiaojie Xu, Lingteng Qiu, Yan Pan, Yushuang Wu, Weikai Chen, Xiaoguang Han

   Abstract: Caricature is an artistic representation that deliberately exaggerates the distinctive features of a human face to convey humor or sarcasm. However, reconstructing a 3D caricature from a 2D caricature image remains a challenging task, mostly due to the lack of data. We propose to fill this gap by introducing 3DCaricShop, the first large-scale 3D caricature dataset that contains 2000 high-quality d… ▽ More

   Submitted 15 March, 2021; **originally announced** March 2021.

   Comments: **CVPR 2021**. Project page:https://qiuyuda.github.io/3DCaricShop/

2. arXiv:2103.07969 [pdf, other]

   cs.CV cs.AI cs.LG

   Monte Carlo Scene Search for 3D Scene Understanding

   Authors: Shreyas Hampali, Sinisa Stekovic, Sayan Deb Sarkar, Chetan Srinivasa Kumar, Friedrich Fraundorfer, Vincent Lepetit

   Abstract: We explore how a general AI algorithm can be used for 3D scene understanding in order to reduce the need for training data. More exactly, we propose a modification of the Monte Carlo Tree Search (MCTS) algorithm to retrieve objects and room layouts from noisy RGB-D scans. While MCTS was developed as a game-playing algorithm, we show it can also be used for complex perception problems. It has few e… ▽ More

   Submitted 14 March, 2021; **originally announced** March 2021.

   Comments: Accepted to **CVPR 2021**. We will upload the final paper version after March 29th

3. arXiv:2103.07941 [pdf, other]

   cs.CV

   Modular Interactive Video Object Segmentation: Interaction-to-Mask, Propagation and Difference-Aware Fusion

Authors: Ho Kei Cheng, Yu-Wing Tai, Chi-Keung Tang

Abstract: We present Modular interactive VOS (MiVOS) framework which decouples interaction-to-mask and mask propagation, allowing for higher generalizability and better performance. Trained separately, the interaction module converts user interactions to an object mask, which is then temporally propagated by our propagation module using a novel top-k filtering strategy in reading the space-time memory. To… ▽ More

Submitted 21 March, 2021; v1 submitted 14 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**. Project page: https://hkchengrex.github.io/MiVOS/

4. arXiv:2103.07939 [pdf, other]

cs.CV

Semi-Supervised Video Deraining with Dynamic Rain Generator

Authors: Zongsheng Yue, Jianwen Xie, Qian Zhao, Deyu Meng

Abstract: While deep learning (DL)-based video deraining methods have achieved significant success recently, they still exist two major drawbacks. Firstly, most of them do not sufficiently model the characteristics of rain layers of rainy videos. In fact, the rain layers exhibit strong physical properties (e.g., direction, scale and thickness) in spatial dimension and natural continuities in temporal dimens… ▽ More

Submitted 14 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

ACM Class: I.4.3

5. arXiv:2103.07894 [pdf, other]

cs.CV

Refer-it-in-RGBD: A Bottom-up Approach for 3D Visual Grounding in RGBD Images

Authors: Haolin Liu, Anran Lin, Xiaoguang Han, Lei Yang, Yizhou Yu, Shuguang Cui

Abstract: Grounding referring expressions in RGBD image has been an emerging field. We present a novel task of 3D visual grounding in single-view RGBD image where the referred objects are often only partially scanned due to occlusion. In contrast to previous works that directly generate object proposals for grounding in the 3D scenes, we propose a bottom-up approach to gradually aggregate context-aware info… ▽ More

Submitted 17 March, 2021; v1 submitted 14 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**, project page: https://unclemedm.github.io/Refer-it-in-RGBD/

6. arXiv:2103.07889 [pdf, other]

cs.CV

Learning a Proposal Classifier for Multiple Object Tracking

Authors: Peng Dai, Renliang Weng, Wongun Choi, Changshui Zhang, Zhangping He, Wei Ding

Abstract: The recent trend in multiple object tracking (MOT) is heading towards leveraging deep learning to boost the tracking performance. However, it is not trivial to solve the data-association problem in an end-to-end fashion. In this paper, we propose a novel proposal-based learnable framework, which models MOT as a proposal generation, proposal scoring and trajectory inference paradigm on an affinity… ▽ More

Submitted 23 March, 2021; v1 submitted 14 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**, Poster, EEE/CVF Conference on Computer Vision and Pattern Recognition

7. arXiv:2103.07838 [pdf, other]

cs.CV

Cycle4Completion: Unpaired Point Cloud Completion using Cycle Transformation with Missing Region Coding

Authors: Xin Wen, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, Yu-Shen Liu

Abstract: In this paper, we present a novel unpaired point cloud completion network, named Cycle4Completion, to infer the complete geometries from a partial 3D object. Previous unpaired completion methods merely focus on the learning of geometric correspondence from incomplete shapes to complete shapes, and ignore the learning in the reverse direction, which makes them suffer from low completion accuracy du… ▽ More

Submitted 13 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

8. arXiv:2103.07738 [pdf, other]

cs.CV cs.LG

Reconsidering Representation Alignment for Multi-view Clustering

Authors: Daniel J. Trosten, Sigurd Løkse, Robert Jenssen, Michael Kampffmeyer

Abstract: Aligning distributions of view representations is a core component of today's state of the art models for deep multi-view clustering. However, we identify several drawbacks with naïvely aligning representation distributions. We demonstrate that these drawbacks both lead to less separable clusters in the representation space, and inhibit the model's ability to prioritize views. Based on these obser… ▽ More

Submitted 13 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**. Code available at https://github.com/DanielTrosten/mvc

9. arXiv:2103.07600 [pdf, other]

cs.LG cs.CV stat.ML

Student-Teacher Learning from Clean Inputs to Noisy Inputs

Authors: Guanzhe Hong, Zhiyuan Mao, Xiaojun Lin, Stanley H. Chan

Abstract: Feature-based student-teacher learning, a training method that encourages the student's hidden features to mimic those of the teacher network, is empirically successful in transferring the knowledge from a pre-trained teacher network to the student network. Furthermore, recent empirical results demonstrate that, the teacher's features can boost the student network's generalization even when the st… ▽ More

Submitted 12 March, 2021; originally announced March 2021.

Comments: Published at the Conference on Computer Vision and Pattern Recognition (**CVPR 2021**)

10. arXiv:2103.07531 [pdf, other]

cs.CV

Uncertainty-guided Model Generalization to Unseen Domains

Authors: Fengchun Qiao, Xi Peng

Abstract: We study a worst-case scenario in generalization: Out-of-domain generalization from a single source. The goal is to learn a robust model from a single source and expect it to generalize over many unknown distributions. This challenging problem has been seldom investigated while existing solutions suffer from various limitations. In this paper, we propose a new solution. The key idea is to augment… ▽ More

Submitted 12 March, 2021; **originally announced** March 2021.

Comments: In **CVPR 2021** (13 pages including supplementary material)

11. arXiv:2103.07503 [pdf, other]

cs.CV cs.LG

Cross-Domain Similarity Learning for Face Recognition in Unseen Domains

Authors: Masoud Faraki, Xiang Yu, Yi-Hsuan Tsai, Yumin Suh, Manmohan Chandraker

Abstract: Face recognition models trained under the assumption of identical training and test distributions often suffer from poor generalization when faced with unknown variations, such as a novel ethnicity or unpredictable individual make-ups during test time. In this paper, we introduce a novel cross-domain metric learning loss, which we dub Cross-Domain Triplet (CDT) loss, to improve face recognition in… ▽ More

Submitted 12 March, 2021; originally announced March 2021.

Comments: Accepted to **CVPR**'21

12. arXiv:2103.07372 [pdf, other]

cs.CV

ACTION-Net: Multipath Excitation for Action Recognition

Authors: Zhengwei Wang, Qi She, Aljosa Smolic

Abstract: Spatial-temporal, channel-wise, and motion patterns are three complementary and crucial types of information for video action recognition. Conventional 2D CNNs are computationally cheap but cannot catch temporal relationships; 3D CNNs can achieve good performance but are computationally intensive. In this work, we tackle this dilemma by designing a generic and effective module that can be embedded… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

13. arXiv:2103.07289 [pdf, other]

cs.CV

Searching by Generating: Flexible and Efficient One-Shot NAS with Architecture Generator

Authors: Sian-Yao Huang, Wei-Ta Chu

Abstract: In one-shot NAS, sub-networks need to be searched from the supernet to meet different hardware constraints. However, the search cost is high and N times of searches are needed for N different constraints. In this work, we propose a novel search strategy called architecture generator to search sub-networks by generating them, so that the search process can be much more efficient and flexible. W… ▽ More

Submitted 12 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

14. arXiv:2103.07254 [pdf, other]

cs.CV

Deep Dual Consecutive Network for Human Pose Estimation

Authors: Zhenguang Liu, Haoming Chen, Runyang Feng, Shuang Wu, Shouling Ji, Bailin Yang, Xun Wang

Abstract: Multi-frame human pose estimation in complicated situations is challenging. Although state-of-the-art human joints detectors have demonstrated remarkable results for static images, their performances come short when we apply these models to video sequences. Prevalent shortcomings include the failure to handle motion blur, video defocus, or pose occlusions, arising from the inability in capturing t… ▽ More

Submitted 19 March, 2021; v1 submitted 12 March, 2021; **originally announced** March 2021.

Comments: This paper is accepted by **CVPR 2021**

15. arXiv:2103.07156 [pdf, other]

cs.CV cs.LG

Learnable Companding Quantization for Accurate Low-bit Neural Networks

Authors: Kohei Yamamoto

Abstract: Quantizing deep neural networks is an effective method for reducing memory consumption and improving inference speed, and is thus useful for implementation in resource-constrained devices. However, it is still hard for extremely low-bit models to achieve accuracy comparable with that of full-precision models. To address this issue, we propose learnable companding quantization (LCQ) as a novel non-… ▽

More

Submitted 12 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

16. arXiv:2103.07152 [pdf, other]

eess.IV cs.CV

Deep Gaussian Scale Mixture Prior for Spectral Compressive Imaging

Authors: Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, Guangming Shi

Abstract: In coded aperture snapshot spectral imaging (CASSI) system, the real-world hyperspectral image (HSI) can be reconstructed from the captured compressive image in a snapshot. Model-based HSI reconstruction methods employed hand-crafted priors to solve the reconstruction problem, but most of which achieved limited success due to the poor representation capability of these hand-crafted priors. Deep le… ▽ More

Submitted 12 March, 2021; **originally announced** March 2021.

Comments: 10 pages, 8 figures, **CVPR 2021**

17. arXiv:2103.07017 [pdf, other]

cs.CV

CRFace: Confidence Ranker for Model-Agnostic Face Detection Refinement

Authors: Noranart Vesdapunt, Baoyuan Wang

Abstract: Face detection is a fundamental problem for many downstream face applications, and there is a rising demand for faster, more accurate yet support for higher resolution face detectors. Recent smartphones can record a video in 8K resolution, but many of the existing face detectors still fail due to the anchor size and training data. We analyze the failure cases and observe a large number of correct… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

18. arXiv:2103.06937 [pdf, other]

cs.CV cs.LG

The Semi-Supervised iNaturalist-Aves Challenge at FGVC7 Workshop

Authors: Jong-Chyi Su, Subhransu Maji

**Abstract**: This document describes the details and the motivation behind a new dataset we collected for the semi-supervised recognition challenge~\cite{semi-aves} at the FGVC7 workshop at **CVPR** 2020. The dataset contains 1000 species of birds sampled from the iNat-2018 dataset for a total of nearly 150k images. From this collection, we sample a subset of classes and the… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: Tech report for Semi-iNat 2020 challenge, please see http://github.com/cvl-umass/semi-inat-2020

19. arXiv:2103.06879 [pdf, other]

cs.CV cs.AI cs.LG

CoMoGAN: continuous model-guided image-to-image translation

Authors: Fabio Pizzati, Pietro Cerri, Raoul de Charette

Abstract: CoMoGAN is a continuous GAN relying on the unsupervised reorganization of the target data on a functional manifold. To that matter, we introduce a new Functional Instance Normalization layer and residual mechanism, which together disentangle image content from position on target manifold. We rely on naive physics-inspired models to guide the training while allowing private model/translations featu… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** oral

20. arXiv:2103.06878 [pdf, other]

cs.CV cs.GR

Diverse Semantic Image Synthesis via Probability Distribution Modeling

Authors: Zhentao Tan, Menglei Chai, Dongdong Chen, Jing Liao, Qi Chu, Bin Liu, Gang Hua, Nenghai Yu

Abstract: Semantic image synthesis, translating semantic layouts to photo-realistic images, is a one-to-many mapping problem. Though impressive progress has been recently made, diverse semantic synthesis that can efficiently produce semantic-level multimodal results, still remains a challenge. In this paper, we propose a novel diverse semantic image synthesis framework from the perspective of semantic class… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: Accepted By **CVPR 2021**

21. arXiv:2103.06877 [pdf, other]

cs.CV cs.LG

Fast and Accurate Model Scaling

Authors: Piotr Dollár, Mannat Singh, Ross Girshick

Abstract: In this work we analyze strategies for convolutional neural network scaling; that is, the process of scaling a base convolutional network to endow it with greater computational complexity and consequently representational power. Example scaling strategies may include increasing model width, depth, resolution, etc. While various scaling strategies exist, their tradeoffs are not fully understood. Ex… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

22. arXiv:2103.06871 [pdf, other]

cs.CV

SMPLicit: Topology-aware Generative Model for Clothed People

Authors: Enric Corona, Albert Pumarola, Guillem Alenyà, Gerard Pons-Moll, Francesc Moreno-Noguer

Abstract: In this paper we introduce SMPLicit, a novel generative model to jointly represent body pose, shape and clothing geometry. In contrast to existing learning-based approaches that require training specific models for each type of garment, SMPLicit can represent in a unified manner different garment topologies (e.g. from sleeveless tops to hoodies and to open jackets), while controlling other propert… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

23. arXiv:2103.06669 [pdf, other]

cs.CV

Temporal Action Segmentation from Timestamp Supervision

Authors: Zhe Li, Yazan Abu Farha, Juergen Gall

Abstract: Temporal action segmentation approaches have been very successful recently. However, annotating videos with frame-wise labels to train such models is very expensive and time consuming. While weakly supervised methods trained using only ordered action lists require much less annotation effort, the performance is still much worse than fully supervised approaches. In this paper, we introduce timestam… ▽ More

Submitted 15 March, 2021; v1 submitted 11 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

24. arXiv:2103.06627 [pdf, other]

cs.CV

MagFace: A Universal Representation for Face Recognition and Quality Assessment

Authors: Qiang Meng, Shichao Zhao, Zhida Huang, Feng Zhou

Abstract: The performance of face recognition system degrades when the variability of the acquired faces increases. Prior work alleviates this issue by either monitoring the face quality in pre-processing or predicting the data uncertainty along with the face feature. This paper proposes MagFace, a category of losses that learn a universal feature embedding whose magnitude can measure the quality of the giv… ▽ More

Submitted 15 March, 2021; v1 submitted 11 March, 2021; **originally announced** March 2021.

Comments: accepted at **CVPR 2021**, Oral

Journal ref: IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), **2021**

25. arXiv:2103.06541 [pdf, other]

cs.CV cs.AI cs.MM

Affect2MM: Affective Analysis of Multimedia Content Using Emotion Causality

Authors: Trisha Mittal, Puneet Mathur, Aniket Bera, Dinesh Manocha

Abstract: We present Affect2MM, a learning method for time-series emotion prediction for multimedia content. Our goal is to automatically capture the varying emotions depicted by characters in real-life human-centric situations and behaviors. We use the ideas from emotion causation theories to computationally model and determine the emotional state evoked in clips of movies. Affect2MM explicitly models the… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

26. arXiv:2103.06495 [pdf, other]

cs.CV

Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition

Authors: Shancheng Fang, Hongtao Xie, Yuxin Wang, Zhendong Mao, Yongdong Zhang

Abstract: Linguistic knowledge is of great benefit to scene text recognition. However, how to effectively model linguistic rules in end-to-end deep networks remains a research challenge. In this paper, we argue that the limited capacity of language models comes from: 1) implicitly language modeling; 2) unidirectional feature representation; and 3) language model with noise input. Correspondingly, we propose… ▽ More

Submitted 11 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

27. arXiv:2103.06342 [pdf, other]

cs.CV cs.AI cs.LG cs.MM

Continual Semantic Segmentation via Repulsion-Attraction of Sparse and Disentangled Latent Representations

Authors: Umberto Michieli, Pietro Zanuttigh

Abstract: Deep neural networks suffer from the major limitation of catastrophic forgetting old tasks when learning new ones. In this paper we focus on class incremental continual learning in semantic segmentation, where new categories are made available over time while previous training data is not retained. The proposed continual learning scheme shapes the latent space to reduce forgetting whilst improving… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**. 22 pages, 10 figures, 11 tables

28. arXiv:2103.06255 [pdf, other]

cs.CV

Involution: Inverting the Inherence of Convolution for Visual Recognition

Authors: Duo Li, Jie Hu, Changhu Wang, Xiangtai Li, Qi She, Lei Zhu, Tong Zhang, Qifeng Chen

Abstract: Convolution has been the core ingredient of modern neural networks, triggering the surge of deep learning in vision. In this work, we rethink the inherent principles of standard convolution for vision tasks, specifically spatial-agnostic and channel-specific. Instead, we present a novel atomic operation for deep neural networks by inverting the aforementioned design principles of convolution, coin… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**. Code and models are available at https://github.com/d-li14/involution

29. arXiv:2103.06122 [pdf, other]

cs.CV cs.LG

Spatially Consistent Representation Learning

Authors: Byungseok Roh, Wuhyun Shin, Ildoo Kim, Sungwoong Kim

Abstract: Self-supervised learning has been widely used to obtain transferrable representations from unlabeled images. Especially, recent contrastive learning methods have shown impressive performances on downstream image classification tasks. While these contrastive methods mainly focus on generating invariant global representations at the image-level under semantic-preserving transformations, they are pro… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

30. arXiv:2103.06030 [pdf, other]

cs.CV

FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space

Authors: Quande Liu, Cheng Chen, Jing Qin, Qi Dou, Pheng-Ann Heng

Abstract: Federated learning allows distributed medical institutions to collaboratively learn a shared prediction model with privacy protection. While at clinical deployment, the models trained in federated learning can still suffer from performance drop when applied to completely unseen hospitals outside the federation. In this paper, we point out and solve a novel problem setting of federated domain gener… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

31. arXiv:2103.05983 [pdf, other]

cs.CV

Reformulating HOI Detection as Adaptive Set Prediction

Authors: Mingfei Chen, Yue Liao, Si Liu, Zhiyuan Chen, Fei Wang, Chen Qian

Abstract: Determining which image regions to concentrate on is critical for Human-Object Interaction (HOI) detection. Conventional HOI detectors focus on either detected human and object pairs or pre-defined interaction locations, which limits learning of the effective features. In this paper, we reformulate HOI detection as an adaptive set prediction problem, with this novel formulation, we propose an Adap… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

32. arXiv:2103.05977 [pdf, other]

cs.CV

SDD-FIQA: Unsupervised Face Image Quality Assessment with Similarity Distribution Distance

Authors: Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, Yuan-Gen Wang

Abstract: In recent years, Face Image Quality Assessment (FIQA) has become an indispensable part of the face recognition system to guarantee the stability and reliability of recognition performance in an unconstrained scenario. For this purpose, the FIQA method should consider both the intrinsic property and the recognizability of the face image. Most previous works aim to estimate the sample-wise embedding… ▽ More

Submitted 10 March, 2021; originally announced March 2021.

Journal ref: IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), **2021**

33. arXiv:2103.05950 [pdf, other]

cs.CV

FSCE: Few-Shot Object Detection via Contrastive Proposal Encoding

Authors: Bo Sun, Banghuai Li, Shengcai Cai, Ye Yuan, Chi Zhang

Abstract: Emerging interests have been brought to recognize previously unseen objects given very few training examples, known as few-shot object detection (FSOD). Recent researches demonstrate that good feature embedding is the key to reach favorable few-shot learning performance. We observe object proposals with different Intersection-of-Union (IoU) scores are analogous to the intra-image augmentation used… ▽ More

Submitted 13 March, 2021; v1 submitted 10 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** Accepted

34. arXiv:2103.05905 [pdf, other]

cs.CV cs.LG cs.MM

VideoMoCo: Contrastive Video Representation Learning with Temporally Adversarial Examples

Authors: Tian Pan, Yibing Song, Tianyu Yang, Wenhao Jiang, Wei Liu

Abstract: MoCo is effective for unsupervised image representation learning. In this paper, we propose VideoMoCo for unsupervised video representation learning. Given a video sequence as an input sample, we improve the temporal feature representations of MoCo from two perspectives. First, we introduce a generator to drop out several frames from this sample temporally. The discriminator is then learned to enc… ▽ More

Submitted 16 March, 2021; v1 submitted 10 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

35. arXiv:2103.05898 [pdf, other]

cs.CV cs.AI cs.LG

Limitations of Post-Hoc Feature Alignment for Robustness

Authors: Collin Burns, Jacob Steinhardt

Abstract: Feature alignment is an approach to improving robustness to distribution shift that matches the distribution of feature activations between the training distribution and test distribution. A particularly simple but effective approach to feature alignment involves aligning the batch normalization statistics between the two distributions in a trained neural network. This technique has received renew… ▽ More

Submitted 10 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

36. arXiv:2103.05863 [pdf, other]

cs.CV cs.AI cs.LG

AutoDO: Robust AutoAugment for Biased Data with Label Noise via Scalable Probabilistic Implicit Differentiation

Authors: Denis Gudovskiy, Luca Rigazio, Shun Ishizaka, Kazuki Kozuka, Sotaro Tsukizawa

Abstract: AutoAugment has sparked an interest in automated augmentation methods for deep learning models. These methods estimate image transformation policies for train data that improve generalization to test data. While recent papers evolved in the direction of decreasing policy search complexity, we show that those methods are not robust when applied to biased and noisy data. To overcome these limitation… ▽ More

Submitted 11 March, 2021; v1 submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**. Preprint

37. arXiv:2103.05861 [pdf, other]

cs.CV

Manifold Regularized Dynamic Network Pruning

Authors: <u>Yehui Tang</u>, <u>Yunhe Wang</u>, <u>Yixing Xu</u>, <u>Yiping Deng</u>, <u>Chao Xu</u>, <u>Dacheng Tao</u>, <u>Chang Xu</u>

Abstract: Neural network pruning is an essential approach for reducing the computational complexity of deep models so that they can be well deployed on resource-limited devices. Compared with conventional methods, the recently developed dynamic pruning methods determine redundant filters variant to each input instance which achieves higher acceleration. Most of the existing methods discover effective sub-ne… ▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: This paper is accepted by **CVPR 2021**. Key words: Filter pruning, Dynamic network, Network compression, Manifold regularization

38. <u>arXiv:2103.05630</u> [pdf, other]

cs.CV cs.LG

ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis

Authors: <u>Yinan He</u>, <u>Bei Gan</u>, <u>Siyu Chen</u>, <u>Yichun Zhou</u>, <u>Guojun Yin</u>, <u>Luchuan Song</u>, <u>Lu Sheng</u>, <u>Jing Shao</u>, <u>Ziwei Liu</u>

Abstract: The rapid progress of photorealistic synthesis techniques has reached at a critical point where the boundary between real and manipulated images starts to blur. Thus, benchmarking and advancing digital forgery analysis have become a pressing issue. However, existing face forgery datasets either have limited diversity or only support coarse-grained analysis. To counter this emerging threat, we cons… ▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: 17 pages, 11 figures, Accepted to **CVPR 2021** (Oral), project webpage: <u>https://yinanhe.github.io/projects/forgerynet.html</u>

39. <u>arXiv:2103.05606</u> [pdf, other]

cs.CV cs.GR cs.LG

NeX: Real-time View Synthesis with Neural Basis Expansion

Authors: <u>Suttisak Wizadwongsa</u>, <u>Pakkapon Phongthawee</u>, <u>Jiraphon Yenphraphai</u>, <u>Supasorn Suwajanakorn</u>

Abstract: We present NeX, a new approach to novel view synthesis based on enhancements of multiplane image (MPI) that can reproduce next-level view-dependent effects -- in real time. Unlike traditional MPI that uses a set of simple RGB$\alpha$ planes, our technique models view-dependent effects by instead parameterizing each pixel as a linear combination of basis functions learned from a neural network. Moreover… ▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** (Oral)

40. <u>arXiv:2103.05471</u> [pdf, other]

cs.CV

Contrastive Neural Architecture Search with Neural Architecture Comparators

Authors: <u>Yaofo Chen</u>, <u>Yong Guo</u>, <u>Qi Chen</u>, <u>Minli Li</u>, <u>Yaowei Wang</u>, <u>Wei Zeng</u>, <u>Mingkui Tan</u>

Abstract: One of the key steps in Neural Architecture Search (NAS) is to estimate the performance of candidate architectures. Existing methods either directly use the validation performance or learn a predictor to estimate the performance. However, these methods can be either computationally expensive or very inaccurate, which may severely affect the search efficiency and performance. Moreover, as it is ver… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: Accpeted by **CVPR 2021**. The code is available at <u>https://github.com/chenyaofo/CTNAS</u>

41.  <u>arXiv:2103.05465</u> [<u>pdf</u>, <u>other</u>]

cs.CV

PointDSC: Robust Point Cloud Registration using Deep Spatial Consistency

Authors: <u>Xuyang Bai</u>, <u>Zixin Luo</u>, <u>Lei Zhou</u>, <u>Hongkai Chen</u>, <u>Lei Li</u>, <u>Zeyu Hu</u>, <u>Hongbo Fu</u>, <u>Chiew-Lan Tai</u>

Abstract: Removing outlier correspondences is one of the critical steps for successful feature-based point cloud registration. Despite the increasing popularity of introducing deep learning methods in this field, spatial consistency, which is essentially established by a Euclidean transformation between point clouds, has received almost no individual attention in existing learning frameworks. In this paper,… ▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**, supplementary materials included

42.  <u>arXiv:2103.05347</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Understanding the Robustness of Skeleton-based Action Recognition under Adversarial Attack

Authors: <u>He Wang</u>, <u>Feixiang He</u>, <u>Zhexi Peng</u>, <u>Tianjia Shao</u>, <u>Yong-Liang Yang</u>, <u>Kun Zhou</u>, <u>David Hogg</u>

Abstract: Action recognition has been heavily employed in many applications such as autonomous vehicles, surveillance, etc, where its robustness is a primary concern. In this paper, we examine the robustness of state-of-the-art action recognizers against adversarial attack, which has been rarely investigated so far. To this end, we propose a new method to attack action recognizers that rely on 3D skeletal m… ▽ More

Submitted 18 March, 2021; v1 submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted in **CVPR 2021**. arXiv admin note: substantial text overlap with arXiv:1911.07107

43.  <u>arXiv:2103.05284</u> [<u>pdf</u>, <u>other</u>]

cs.CV cs.CL

Open-book Video Captioning with Retrieve-Copy-Generate Network

Authors: <u>Ziqi Zhang</u>, <u>Zhongang Qi</u>, <u>Chunfeng Yuan</u>, <u>Ying Shan</u>, <u>Bing Li</u>, <u>Ying Deng</u>, <u>Weiming Hu</u>

Abstract: Due to the rapid emergence of short videos and the requirement for content understanding and creation, the video captioning task has received increasing attention in recent years. In this paper, we convert traditional video captioning task into a new paradigm, \ie, Open-book Video Captioning, which generates natural language under the prompts of video-content-relevant sentences, not limited to the… ▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

44. arXiv:2103.05271 [pdf, other]

cs.CV

Probabilistic Modeling of Semantic Ambiguity for Scene Graph Generation

Authors: Gengcong Yang, Jingyi Zhang, Yong Zhang, Baoyuan Wu, Yujiu Yang

Abstract: To generate "accurate" scene graphs, almost all existing methods predict pairwise relationships in a deterministic manner. However, we argue that visual relationships are often semantically ambiguous. Specifically, inspired by linguistic knowledge, we classify the ambiguity into three types: Synonymy Ambiguity, Hyponymy Ambiguity, and Multi-view Ambiguity. The ambiguity naturally leads to the issu… ▽ More

Submitted 10 March, 2021; v1 submitted 9 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** poster

45. arXiv:2103.05266 [pdf, other]

cs.CV cs.AI

BASAR:Black-box Attack on Skeletal Action Recognition

Authors: Yunfeng Diao, Tianjia Shao, Yong-Liang Yang, Kun Zhou, He Wang

Abstract: Skeletal motion plays a vital role in human activity recognition as either an independent data source or a complement. The robustness of skeleton-based activity recognizers has been questioned recently, which shows that they are vulnerable to adversarial attacks when the full-knowledge of the recognizer is accessible to the attacker. However, this white-box requirement is overly restrictive in mos… ▽ More

Submitted 19 March, 2021; v1 submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted in **CVPR 2021**

46. arXiv:2103.05254 [pdf, other]

cs.CV

MetaCorrection: Domain-aware Meta Loss Correction for Unsupervised Domain Adaptation in Semantic Segmentation

Authors: Xiaoqing Guo, Chen Yang, Baopu Li, Yixuan Yuan

Abstract: Unsupervised domain adaptation (UDA) aims to transfer the knowledge from the labeled source domain to the unlabeled target domain. Existing self-training based UDA approaches assign pseudo labels for target data and treat them as ground truth labels to fully leverage unlabeled target data for model adaptation. However, the generated pseudo labels from the model optimized on the source domain inevi…
▽ More

Submitted 9 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

47. arXiv:2103.05152 [pdf, other]

cs.CV cs.AI cs.LG

Knowledge Evolution in Neural Networks

Authors: Ahmed Taha, Abhinav Shrivastava, Larry Davis

Abstract: Deep learning relies on the availability of a large corpus of data (labeled or unlabeled). Thus, one challenging unsettled question is: how to train a deep network on a relatively small dataset? To tackle this question, we propose an evolution-inspired training approach to boost performance on relatively small datasets. The knowledge evolution (KE) approach splits a deep network into two hypothese… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: **CVPR** Oral **2021**

48. arXiv:2103.05121 [pdf, other]

cs.CV cs.AI

Multiple Instance Captioning: Learning Representations from Histopathology Textbooks and Articles

Authors: Jevgenij Gamper, Nasir Rajpoot

Abstract: We present ARCH, a computational pathology (CP) multiple instance captioning dataset to facilitate dense supervision of CP tasks. Existing CP datasets focus on narrow tasks; ARCH on the other hand contains dense diagnostic and morphological descriptions for a range of stains, tissue types and pathologies. Using intrinsic dimensionality estimation, we show that ARCH is the only CP dataset to (ARCH-… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

49. arXiv:2103.05086 [pdf, other]

cs.CV

How Privacy-Preserving are Line Clouds? Recovering Scene Details from 3D Lines

Authors: Kunal Chelani, Fredrik Kahl, Torsten Sattler

Abstract: Visual localization is the problem of estimating the camera pose of a given image with respect to a known scene. Visual localization algorithms are a fundamental building block in advanced computer vision applications, including Mixed and Virtual Reality systems. Many algorithms used in practice represent the scene through a Structure-from-Motion (SfM) point cloud and use 2D-3D matches between a q… ▽ More

Submitted 8 March, 2021; originally announced March 2021.

Comments: Computer Vision and Pattern Recognition (**CVPR**) **2021**

50. arXiv:2103.04677 [pdf, other]

cs.CV

Behavior-Driven Synthesis of Human Dynamics

Authors: Andreas Blattmann, Timo Milbich, Michael Dorkenwald, Björn Ommer

Abstract: Generating and representing human behavior are of major importance for various computer vision applications. Commonly, human video synthesis represents behavior as sequences of postures while directly predicting their likely progressions or merely changing the appearance of the depicted persons, thus not being able to exercise control over their actual behavior during the synthesis process. In con… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021** as Poster

1. arXiv:2103.04618 [pdf, other]

cs.CV

Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification

Authors: Fengxiang Yang, Zhun Zhong, Zhiming Luo, Yuanzheng Cai, Yaojin Lin, Shaozi Li, Nicu Sebe

Abstract: This paper considers the problem of unsupervised person re-identification (re-ID), which aims to learn discriminative models with unlabeled data. One popular method is to obtain pseudo-label by clustering and use them to optimize the model. Although this kind of approach has shown promising accuracy, it is hampered by 1) noisy labels produced by clustering and 2) feature variations caused by camer… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

2. arXiv:2103.04570 [pdf, other]

cs.CV

Differentiable Multi-Granularity Human Representation Learning for Instance-Aware Human Semantic Parsing

Authors: Tianfei Zhou, Wenguan Wang, Si Liu, Yi Yang, Luc Van Gool

Abstract: To address the challenging task of instance-aware human part parsing, a new bottom-up regime is proposed to learn category-level human semantic segmentation as well as multi-person pose estimation in a joint and end-to-end manner. It is a compact, efficient and powerful framework that exploits structural information over different human granularities and eases the difficulty of person partitioning… ▽ More

Submitted 8 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** (Oral). Code: https://github.com/tfzhou/MG-HumanParsing

3. arXiv:2103.04507 [pdf, other]

cs.CV

OPANAS: One-Shot Path Aggregation Network Architecture Search for Object Detection

Authors: Tingting Liang, Yongtao Wang, Zhi Tang, Guosheng Hu, Haibin Ling

Abstract: Recently, neural architecture search (NAS) has been exploited to design feature pyramid networks (FPNs) and achieved promising results for visual object detection. Encouraged by the success, we propose a novel One-Shot Path Aggregation Network Architecture Search (OPANAS) algorithm, which significantly improves both searching efficiency and detection accuracy. Specifically, we first introduce six… ▽ More

Submitted 11 March, 2021; v1 submitted 7 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

4. arXiv:2103.04400 [pdf, other]

cs.CV

What If We Only Use Real Datasets for Scene Text Recognition? Toward Scene Text Recognition With Fewer Labels

Authors: Jeonghun Baek, Yusuke Matsui, Kiyoharu Aizawa

Abstract: Scene text recognition (STR) task has a common practice: All state-of-the-art STR models are trained on large synthetic data. In contrast to this practice, training STR models only on fewer real labels (STR with fewer labels) is important when we have to train STR models without synthetic data: for handwritten or artistic texts that are difficult to generate synthetically and for languages other t… ▽ More

Submitted 7 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

5. arXiv:2103.04379 [pdf, other]

cs.CV cs.LG

Repurposing GANs for One-shot Semantic Part Segmentation

Authors: Nontawat Tritrong, Pitchaporn Rewatbowornwong, Supasorn Suwajanakorn

Abstract: While GANs have shown success in realistic image generation, the idea of using GANs for other tasks unrelated to synthesis is underexplored. Do GANs learn meaningful structural parts of objects during their attempt to reproduce those objects? In this work, we test this hypothesis and propose a simple and effective approach based on GANs for semantic part segmentation that requires as few as one la… ▽ More

Submitted 24 March, 2021; v1 submitted 7 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** (Oral)

6. arXiv:2103.04260 [pdf, other]

cs.CV cs.AI

ARVo: Learning All-Range Volumetric Correspondence for Video Deblurring

Authors: Dongxu Li, Chenchen Xu, Kaihao Zhang, Xin Yu, Yiran Zhong, Wenqi Ren, Hanna Suominen, Hongdong Li

Abstract: Video deblurring models exploit consecutive frames to remove blurs from camera shakes and object motions. In order to utilize neighboring sharp patches, typical methods rely mainly on homography or optical flows to spatially align neighboring blurry frames. However, such explicit approaches are less effective in the presence of fast motions with large pixel displacements. In this work, we propose… ▽ More

Submitted 6 March, 2021; **originally announced** March 2021.

Comments: Preprint for **CVPR 2021** Poster

7. arXiv:2103.04256 [pdf, other]

cs.CV

Robust Point Cloud Registration Framework Based on Deep Graph Matching

Authors: Kexue Fu, Shaolei Liu, Xiaoyuan Luo, Manning Wang

Abstract: 3D point cloud registration is a fundamental problem in computer vision and robotics. There has been extensive research in this area, but existing methods meet great challenges in situations with a large proportion of outliers and time constraints, but without good transformation initialization. Recently, a series of learning-based algorithms have been introduced and show advantages in speed. Many… ▽ More

Submitted 6 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**. The code will be made publicly available at https://github.com/fukexue/RGM

8. arXiv:2103.04224 [pdf, other]

cs.CV

MeGA-CDA: Memory Guided Attention for Category-Aware Unsupervised Domain Adaptive Object Detection

Authors: Vibashan VS, Poojan Oza, Vishwanath A. Sindagi, Vikram Gupta, Vishal M. Patel

Abstract: Existing approaches for unsupervised domain adaptive object detection perform feature alignment via adversarial training. While these methods achieve reasonable improvements in performance, they typically perform category-agnostic domain alignment, thereby resulting in negative transfer of features. To overcome this issue, in this work, we attempt to incorporate category information into the domai… ▽ More

9. arXiv:2103.04200 [pdf, other]

cs.CV cs.AI

Consensus Maximisation Using Influences of Monotone Boolean Functions

Authors: Ruwan Tennakoon, David Suter, Erchuan Zhang, Tat-Jun Chin, Alireza Bab-Hadiashar

Abstract: Consensus maximisation (MaxCon), which is widely used for robust fitting in computer vision, aims to find the largest subset of data that fits the model within some tolerance level. In this paper, we outline the connection between MaxCon problem and the abstract problem of finding the maximum upper zero of a Monotone Boolean Function (MBF) defined over the Boolean Cube. Then, we link the concept o… ▽ More

10. arXiv:2103.04133 [pdf, other]

cs.CV

Learning Statistical Texture for Semantic Segmentation

Authors: Lanyun Zhu, Deyi Ji, Shiping Zhu, Weihao Gan, Wei Wu, Junjie Yan

Abstract: Existing semantic segmentation works mainly focus on learning the contextual information in high-level semantic features with CNNs. In order to maintain a precise boundary, low-level texture features are directly skip-connected into the deeper layers. Nevertheless, texture features are not only about local structure, but also include global statistical knowledge of the input image. In this paper,… ▽ More

11. arXiv:2103.04059 [pdf, other]

cs.CV

Semantic-aware Knowledge Distillation for Few-Shot Class-Incremental Learning

Authors: Ali Cheraghian, Shafin Rahman, Pengfei Fang, Soumava Kumar Roy, Lars Petersson, Mehrtash Harandi

Abstract: Few-shot class incremental learning (FSCIL) portrays the problem of learning new concepts gradually, where only a few examples per concept are available to the learner. Due to the limited number of examples for training, the techniques developed for standard incremental learning cannot be applied verbatim to FSCIL. In this work, we introduce a distillation algorithm to address the problem of FSCIL… ▽ More

12. arXiv:2103.04011 [pdf, other]

cs.CV

Simultaneously Localize, Segment and Rank the Camouflaged Objects

Authors: Yunqiu Lyu, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, Deng-Ping Fan

Abstract: Camouflage is a key defence mechanism across species that is critical to survival. Common strategies for camouflage include background matching, imitating the color and pattern of the environment, and disruptive coloration, disguising body outlines [35]. Camouflaged object detection (COD) aims to segment camouflaged objects hiding in their surroundings. Existing COD models are built upon binary gr… ▽ More

13. arXiv:2103.03891 [pdf, other]

cs.CV cs.LG

LOHO: Latent Optimization of Hairstyles via Orthogonalization

Authors: Rohit Saha, Brendan Duke, Florian Shkurti, Graham W. Taylor, Parham Aarabi

Abstract: Hairstyle transfer is challenging due to hair structure differences in the source and target hair. Therefore, we propose Latent Optimization of Hairstyles via Orthogonalization (LOHO), an optimization-based approach using GAN inversion to infill missing hair structure details in latent space during hairstyle transfer. Our approach decomposes hair into three attributes: perceptual structure, appear… ▽ More

14. arXiv:2103.03501 [pdf, other]

cs.CV cs.AI

Unsupervised Learning for Robust Fitting:A Reinforcement Learning Approach

Authors: Giang Truong, Huu Le, David Suter, Erchuan Zhang, Syed Zulqarnain Gilani

Abstract: Robust model fitting is a core algorithm in a large number of computer vision applications. Solving this problem efficiently for datasets highly contaminated with outliers is, however, still challenging due to the underlying computational complexity. Recent literature has focused on learning-based algorithms. However, most approaches are supervised which require a large amount of labelled training… ▽ More

Submitted 5 March, 2021; **originally announced** March 2021.

Comments: The preprint of paper accepted to **CVPR 2021**

15. arXiv:2103.03467 [pdf, other]

cs.CV

Teachers Do More Than Teach: Compressing Image-to-Image Models

Authors: Qing Jin, Jian Ren, Oliver J. Woodford, Jiazhuo Wang, Geng Yuan, Yanzhi Wang, Sergey Tulyakov

Abstract: Generative Adversarial Networks (GANs) have achieved huge success in generating high-fidelity images, however, they suffer from low efficiency due to tremendous computational cost and bulky memory usage. Recent efforts on compression GANs show noticeable progress in obtaining smaller generators by sacrificing image quality or involving a time-consuming searching process. In this work, we aim to ad… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: 18 pages, 10 figures, accepted by **CVPR 2021**

16. arXiv:2103.03375 [pdf, other]

cs.CV cs.LG

Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food

Authors: Quin Thames, Arjun Karpur, Wade Norris, Fangting Xia, Liviu Panait, Tobias Weyand, Jack Sim

Abstract: Understanding the nutritional content of food from visual data is a challenging computer vision problem, with the potential to have a positive and widespread impact on public health. Studies in this area are limited to existing datasets in the field that lack sufficient diversity or labels required for training models with nutritional understanding capability. We introduce Nutrition5k, a novel dat… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: 8 pages, 3 of appendices. **CVPR 2021**

17. arXiv:2103.03243 [pdf, other]

cs.CV

Anycost GANs for Interactive Image Synthesis and Editing

Authors: Ji Lin, Richard Zhang, Frieder Ganz, Song Han, Jun-Yan Zhu

Abstract: Generative adversarial networks (GANs) have enabled photorealistic image synthesis and editing. However, due to the high computational cost of large-scale generators (e.g., StyleGAN2), it usually takes seconds to see the results of a single edit on edge devices, prohibiting interactive user experience. In this paper, we take inspirations from modern rendering software and propose Anycost GAN for i… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**. The code and demo are available: https://github.com/mit-han-lab/anycost-gan

18. arXiv:2103.03114 [pdf, other]

cs.CV cs.LG cs.RO

Self-supervised Geometric Perception

Authors: Heng Yang, Wei Dong, Luca Carlone, Vladlen Koltun

Abstract: We present self-supervised geometric perception (SGP), the first general framework to learn a feature descriptor for correspondence matching without any ground-truth geometric model labels (e.g., camera poses, rigid transformations). Our first contribution is to formulate geometric perception as an optimization problem that jointly optimizes the feature descriptor and the geometric models given a… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**, Oral presentation. 8 pages main results, 19 pages in total, including references and supplementary

19. arXiv:2103.03067 [pdf, other]

cs.CV cs.RO

TPCN: Temporal Point Cloud Networks for Motion Forecasting

Authors: Maosheng Ye, Tongyi Cao, Qifeng Chen

Abstract: We propose the Temporal Point Cloud Networks (TPCN), a novel and flexible framework with joint spatial and temporal learning for trajectory prediction. Unlike existing approaches that rasterize agents and map information as 2D images or operate in a graph representation, our approach extends ideas from point cloud learning with dynamic temporal learning to capture both spatial and temporal informa… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: accepted to **CVPR 2021**

20. arXiv:2103.03046 [pdf, other]

cs.CR cs.CV cs.LG

PointGuard: Provably Robust 3D Point Cloud Classification

Authors: Hongbin Liu, Jinyuan Jia, Neil Zhenqiang Gong

Abstract: 3D point cloud classification has many safety-critical applications such as autonomous driving and robotic grasping. However, several studies showed that it is vulnerable to adversarial attacks. In particular, an attacker can make a classifier predict an incorrect label for a 3D point cloud via carefully modifying, adding, and/or deleting a small number of its points. Randomized smoothing is state… ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021**

21. arXiv:2103.02884 [pdf, other]

eess.IV cs.CV

A Cross Channel Context Model for Latents in Deep Image Compression

Authors: Changyue Ma, Zhao Wang, Ruling Liao, Yan Ye

**Abstract**: …BD-rate reductions of 6.30% and 6.31% over the baseline entropy model, and 2.50% and 2.20% over the latest video coding standard Versatile Video Coding (VVC) for the Kodak and **CVPR** CLIC2020 professional dataset, respectively. In addition, when optimized for the MS-SSIM metric, our approach generates visually more pleasant reconstructed images. ▽ More

Submitted 4 March, 2021; **originally announced** March 2021.

22. arXiv:2103.02772 [pdf, other]

cs.CV

DeepTag: An Unsupervised Deep Learning Method for Motion Tracking on Cardiac Tagging Magnetic Resonance Images

Authors: Meng Ye, Mikael Kanski, Dong Yang, Qi Chang, Zhennan Yan, Qiaoying Huang, Leon Axel, Dimitris Metaxas

Abstract: Cardiac tagging magnetic resonance imaging (t-MRI) is the gold standard for regional myocardium deformation and cardiac strain estimation. However, this technique has not been widely used in clinical diagnosis, as a result of the difficulty of motion tracking encountered with t-MRI images. In this paper, we propose a novel deep learning-based fully unsupervised method for in vivo motion tracking o… ▽ More

Submitted 13 March, 2021; v1 submitted 3 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021** Oral

23. arXiv:2103.02758 [pdf, other]

cs.CV

Learning Asynchronous and Sparse Human-Object Interaction in Videos

Authors: Romero Morais, Vuong Le, Svetha Venkatesh, Truyen Tran

Abstract: Human activities can be learned from video. With effective modeling it is possible to discover not only the action labels but also the temporal structures of the activities such as the progression of the sub-activities. Automatically recognizing such structure from raw video signal is a new capability that promises authentic modeling and successful recognition of human-object interactions. Toward… ▽ More

Submitted 3 March, 2021; originally announced March 2021.

Comments: Accepted for publication in **CVPR**'21

24. arXiv:2103.02603 [pdf, other]

cs.CV cs.AI cs.LG

Towards Open World Object Detection

Authors: K J Joseph, Salman Khan, Fahad Shahbaz Khan, Vineeth N Balasubramanian

Abstract: Humans have a natural instinct to identify unknown object instances in their environments. The intrinsic curiosity about these unknown instances aids in learning about them, when the corresponding knowledge is eventually available. This motivates us to propose a novel computer vision problem called: `Open World Object Detection', where a model is tasked to: 1) identify objects that have not been i… ▽ More

Submitted 3 March, 2021; **originally announced** March 2021.

Comments: To appear in **CVPR 2021** as an ORAL paper. Code is available in https://github.com/JosephKJ/OWOD

25. arXiv:2103.02584 [pdf, other]

cs.CV

Cross-View Regularization for Domain Adaptive Panoptic Segmentation

Authors: Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu

Abstract: Panoptic segmentation unifies semantic segmentation and instance segmentation which has been attracting increasing attention in recent years. However, most existing research was conducted under a supervised learning setup whereas unsupervised domain adaptive panoptic segmentation which is critical in different tasks and applications is largely neglected. We design a domain adaptive panoptic segmen… ▽ More

Submitted 3 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021** as an Oral Presentation

26. arXiv:2103.02535 [pdf, other]

cs.CV

Style-based Point Generator with Adversarial Rendering for Point Cloud Completion

Authors: Chulin Xie, Chuxin Wang, Bo Zhang, Hao Yang, Dong Chen, Fang Wen

Abstract: In this paper, we proposed a novel Style-based Point Generator with Adversarial Rendering (SpareNet) for point cloud completion. Firstly, we present the channel-attentive EdgeConv to fully exploit the local structures as well as the global shape in point features. Secondly, we observe that the concatenation manner used by vanilla foldings limits its potential of generating a complex and faithful s… ▽ More

Submitted 3 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

27. arXiv:2103.02396 [pdf, other]

cs.CV

S3: Learnable Sparse Signal Superdensity for Guided Depth Estimation

Authors: Yu-Kai Huang, Yueh-Cheng Liu, Tsung-Han Wu, Hung-Ting Su, Yu-Cheng Chang, Tsung-Lin Tsou, Yu-An Wang, Winston H. Hsu

Abstract: Dense depth estimation plays a key role in multiple applications such as robotics, 3D reconstruction, and augmented reality. While sparse signal, e.g., LiDAR and Radar, has been leveraged as guidance for enhancing dense depth estimation, the improvement is limited due to its low density and imbalanced distribution. To maximize the utility from the sparse source, we propose S3 technique, which e… ▽ More

Submitted 22 March, 2021; v1 submitted 3 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

28. arXiv:2103.02340 [pdf, other]

cs.CV

General Instance Distillation for Object Detection

Authors: Xing Dai, Zeren Jiang, Zhao Wu, Yiping Bao, Zhicheng Wang, Si Liu, Erjin Zhou

Abstract: In recent years, knowledge distillation has been proved to be an effective solution for model compression. This approach can make lightweight student models acquire the knowledge extracted from cumbersome teacher models. However, previous distillation methods of detection have weak generalization for different detection frameworks and rely heavily on ground truth (GT), ignoring the valuable relati… ▽ More

Submitted 3 March, 2021; **originally announced** March 2021.

Comments: 10 pages (including 2 pages of References), 5 figures, 7 tables. Accepted by **CVPR 2021**

29. arXiv:2103.02193 [pdf, other]

cs.CV cs.LG

Adaptive Consistency Regularization for Semi-Supervised Transfer Learning

Authors: Abulikemu Abuduweili, Xingjian Li, Humphrey Shi, Cheng-Zhong Xu, Dejing Dou

Abstract: While recent studies on semi-supervised learning have shown remarkable progress in leveraging both labeled and unlabeled data, most of them presume a basic setting of the model is randomly initialized. In this work, we consider semi-supervised learning and transfer learning jointly, leading to a more practical and competitive paradigm that can utilize both powerful pre-trained models from source d… ▽ More

Submitted 3 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

30. arXiv:2103.02148 [pdf, other]

eess.IV cs.CV

Multi-institutional Collaborations for Improving Deep Learning-based Magnetic Resonance Image Reconstruction Using Federated Learning

Authors: Pengfei Guo, Puyang Wang, Jinyuan Zhou, Shanshan Jiang, Vishal M. Patel

Abstract: Fast and accurate reconstruction of magnetic resonance (MR) images from under-sampled data is important in many clinical applications. In recent years, deep learning-based methods have been shown to produce superior performance on MR image reconstruction. However, these methods require large amounts of data which is difficult to collect and share due to the high cost of acquisition and medical dat… ▽ More

Submitted 10 March, 2021; v1 submitted 2 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**

31. arXiv:2103.01903 [pdf, other]

cs.CV

Semantic Relation Reasoning for Shot-Stable Few-Shot Object Detection

Authors: Chenchen Zhu, Fangyi Chen, Uzair Ahmed, Zhiqiang Shen, Marios Savvides

Abstract: Few-shot object detection is an imperative and long-lasting problem due to the inherent long-tail distribution of real-world data. Its performance is largely affected by the data scarcity of novel classes. But the semantic relation between the novel classes and the base classes is constant regardless of the data availability. In this work, we investigate utilizing this semantic relation together w… ▽ More

Submitted 19 March, 2021; v1 submitted 2 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

32. arXiv:2103.01856 [pdf, other]

cs.CV

Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain

Authors: Honggu Liu, Xiaodan Li, Wenbo Zhou, Yuefeng Chen, Yuan He, Hui Xue, Weiming Zhang, Nenghai Yu

Abstract: The remarkable success in face forgery techniques has received considerable attention in computer vision due to security concerns. We observe that up-sampling is a necessary step of most face forgery techniques, and cumulative up-sampling will result in obvious changes in the frequency domain, especially in the phase spectrum. According to the property of natural images, the phase spectrum preserv… ▽ More

Submitted 10 March, 2021; v1 submitted 2 March, 2021; originally announced March 2021.

Comments: Accepted by IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), **2021**

33. arXiv:2103.01786 [pdf, other]

eess.IV cs.CV

MetaSCI: Scalable and Adaptive Reconstruction for Video Compressive Sensing

Authors: Zhengjue Wang, Hao Zhang, Ziheng Cheng, Bo Chen, Xin Yuan

Abstract: To capture high-speed videos using a two-dimensional detector, video snapshot compressive imaging (SCI) is a promising system, where the video frames are coded by different masks and then compressed to a snapshot measurement. Following this, efficient algorithms are desired to reconstruct the high-speed frames, where the state-of-the-art results are achieved by deep learning networks. However, the… ▽ More

Submitted 2 March, 2021; **originally announced** March 2021.

Comments: 12 pages, 6 figures, **CVPR 2021**

34. arXiv:2103.01520 [pdf, other]

cs.CV

When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework

Authors: Zhizhong Huang, Junping Zhang, Hongming Shan

Abstract: To minimize the effects of age variation in face recognition, previous work either extracts identity-related discriminative features by minimizing the correlation between identity- and age-related features, called age-invariant face recognition (AIFR), or removes age variation by transforming the faces of different age groups into the same age group, called face age synthesis (FAS); however, the f… ▽ More

Submitted 2 March, 2021; v1 submitted 2 March, 2021; **originally announced** March 2021.

Comments: accepted by **CVPR 2021**

35. arXiv:2103.01486 [pdf, other]

cs.CV

Patch-NetVLAD: Multi-Scale Fusion of Locally-Global Descriptors for Place Recognition

Authors: Stephen Hausler, Sourav Garg, Ming Xu, Michael Milford, Tobias Fischer

Abstract: Visual Place Recognition is a challenging task for robotics and autonomous systems, which must deal with the twin problems of appearance and viewpoint change in an always changing world. This paper introduces Patch-NetVLAD, which provides a novel formulation for combining the advantages of both local and global descriptor methods by deriving patch-level features from NetVLAD residuals. Unlike the… ▽ More

Submitted 2 March, 2021; originally announced March 2021.

Comments: Accepted to IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2021**)

36. arXiv:2103.01468 [pdf, other]

cs.CV cs.RO

Depth from Camera Motion and Object Detection

Authors: Brent A. Griffin, Jason J. Corso

Abstract: This paper addresses the problem of learning to estimate the depth of detected objects given some measurement of camera motion (e.g., from robot kinematics or vehicle odometry). We achieve this by 1) designing a recurrent neural network (DBox) that estimates the depth of objects using a generalized representation of bounding boxes and uncalibrated camera movement and 2) introducing the Object Dept… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

37. arXiv:2103.01458 [pdf, other]

cs.CV

Diffusion Probabilistic Models for 3D Point Cloud Generation

Authors: Shitong Luo, Wei Hu

Abstract: We present a probabilistic model for point cloud generation, which is critical for various 3D vision tasks such as shape completion, upsampling, synthesis and data augmentation. Inspired by the diffusion process in non-equilibrium thermodynamics, we view points in point clouds as particles in a thermodynamic system in contact with a heat bath, which diffuse from the original distribution to a nois… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

38. arXiv:2103.01456 [pdf, other]

cs.CV

Image-to-image Translation via Hierarchical Style Disentanglement

Authors: Xinyang Li, Shengchuan Zhang, Jie Hu, Liujuan Cao, Xiaopeng Hong, Xudong Mao, Feiyue Huang, Yongjian Wu, Rongrong Ji

Abstract: Recently, image-to-image translation has made significant progress in achieving both multi-label (\ie, translation conditioned on different labels) and multi-style (\ie, generation with diverse styles) tasks. However, due to the unexplored independence and exclusiveness in the labels, existing endeavors are defeated by involving uncontrolled manipulations to the translation results. In this paper,… ▽ More

39. arXiv:2103.01353 [pdf, other]

cs.CV cs.LG cs.RO

There is More than Meets the Eye: Self-Supervised Multi-Object Detection and Tracking with Sound by Distilling Multimodal Knowledge

Authors: Francisco Rivera Valverde, Juana Valeria Hurtado, Abhinav Valada

Abstract: Attributes of sound inherent to objects can provide valuable cues to learn rich representations for object detection and tracking. Furthermore, the co-occurrence of audiovisual events in videos can be exploited to localize objects over the image field by solely monitoring the sound in the environment. Thus far, this has only been feasible in scenarios where the camera is static and for single obje… ▽ More

40. arXiv:2103.01315 [pdf, other]

cs.CV cs.LG

Exploring Complementary Strengths of Invariant and Equivariant Representations for Few-Shot Learning

Authors: Mamshad Nayeem Rizve, Salman Khan, Fahad Shahbaz Khan, Mubarak Shah

Abstract: In many real-world problems, collecting a large number of labeled samples is infeasible. Few-shot learning (FSL) is the dominant approach to address this issue, where the objective is to quickly adapt to novel categories in presence of a limited number of samples. FSL tasks have been predominantly solved by leveraging the ideas from gradient-based meta-learning and metric learning approaches. Howe… ▽ More

41. arXiv:2103.01302 [pdf, other]

cs.CV

Coarse-Fine Networks for Temporal Activity Detection in Videos

Authors: Kumara Kahatapitiya, Michael S. Ryoo

Abstract: In this paper, we introduce 'Coarse-Fine Networks', a two-stream architecture which benefits from different abstractions of temporal resolution to learn better video representations for long-term motion. Traditional Video models process inputs at one (or few) fixed temporal resolution without any dynamic frame selection. However, we argue that, processing multiple temporal resolutions of the input… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: Accepted to be published at **CVPR 2021**

42. arXiv:2103.01261 [pdf, other]

cs.CV cs.GR

A Deep Emulator for Secondary Motion of 3D Characters

Authors: Mianlun Zheng, Yi Zhou, Duygu Ceylan, Jernej Barbič

Abstract: Fast and light-weight methods for animating 3D characters are desirable in various applications such as computer games. We present a learning-based approach to enhance skinning-based animations of 3D characters with vivid secondary motion effects. We design a neural network that encodes each local patch of a character simulation mesh where the edges implicitly encode the internal forces between th… ▽ More

Submitted 8 March, 2021; v1 submitted 1 March, 2021; **originally announced** March 2021.

Comments: Accepted at **CVPR 2021**, oral presentation

43. arXiv:2103.01255 [pdf, other]

cs.CV

Auto-Exposure Fusion for Single-Image Shadow Removal

Authors: Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, Song Wang

Abstract: Shadow removal is still a challenging task due to its inherent background-dependent and spatial-variant properties, leading to unknown and diverse shadow patterns. Even powerful state-of-the-art deep neural networks could hardly recover traceless shadow-removed background. This paper proposes a new solution for this task by formulating it as an exposure fusion problem to address the challenges. In… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: accept to cvpr2021

Journal ref: **CVPR 2021**

44. arXiv:2103.01134 [pdf, other]

cs.LG cs.CV

Domain Generalization via Inference-time Label-Preserving Target Projections

Authors: Prashant Pandey, Mrigank Raman, Sumanth Varambally, Prathosh AP

Abstract: Generalization of machine learning models trained on a set of source domains on unseen target domains with different statistics, is a challenging problem. While many approaches have been proposed to solve this problem, they only utilize source data during training but do not take advantage of the fact that a single target example is available at the time of inference. Motivated by this, we propose… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: **CVPR 2021**

45. arXiv:2103.01100 [pdf, other]

cs.CV

Categorical Depth Distribution Network for Monocular 3D Object Detection

Authors: Cody Reading, Ali Harakeh, Julia Chae, Steven L. Waslander

Abstract: Monocular 3D object detection is a key problem for autonomous vehicles, as it provides a solution with simple configuration compared to typical multi-sensor systems. The main challenge in monocular 3D detection lies in accurately predicting object depth, which must be inferred from object and scene cues due to the lack of direct range measurement. Many methods attempt to directly estimate depth to… ▽ More

Submitted 23 March, 2021; v1 submitted 1 March, 2021; **originally announced** March 2021.

Comments: Accepted in **CVPR 2021**

46. arXiv:2103.00887 [pdf, other]

cs.CV cs.AI

Counterfactual Zero-Shot and Open-Set Visual Recognition

Authors: Zhongqi Yue, Tan Wang, Hanwang Zhang, Qianru Sun, Xian-Sheng Hua

Abstract: We present a novel counterfactual framework for both Zero-Shot Learning (ZSL) and Open-Set Recognition (OSR), whose common challenge is generalizing to the unseen-classes by only training on the seen-classes. Our idea stems from the observation that the generated samples for unseen-classes are often out of the true distribution, which causes severe recognition rate imbalance between the seen-class… ▽ More

Submitted 1 March, 2021; **originally announced** March 2021.

Comments: Accepted by **CVPR 2021**

47. arXiv:2103.00430 [pdf, other]

cs.CV cs.LG eess.IV

Training Generative Adversarial Networks in One Stage

Authors: Chengchao Shen, Youtan Yin, Xinchao Wang, Xubin LI, Jie Song, Mingli Song

Abstract: Generative Adversarial Networks (GANs) have demonstrated unprecedented success in various image generation tasks. The encouraging results, however, come at the price of a cumbersome training process, during which the generator and discriminator are alternately updated in two stages. In this paper, we investigate a general training scheme that enables training GANs efficiently in only one stage. Ba… ▽ More

Submitted 3 March, 2021; v1 submitted 28 February, 2021; **originally announced** March 2021.

Comments: Accepted to **CVPR 2021**

48. arXiv:2102.12145 [pdf, other]

cs.CV cs.RO

GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation

Authors: Gu Wang, Fabian Manhardt, Federico Tombari, Xiangyang Ji

Abstract: 6D pose estimation from a single RGB image is a fundamental task in computer vision. The current top-performing deep learning-based methods rely on an indirect strategy, i.e., first establishing 2D-3D correspondences between the coordinates in the image plane and object coordinate system, and then applying a variant of the PnP/RANSAC algorithm. However, this two-stage pipeline is not end-to-end… ▽ More

Submitted 9 March, 2021; v1 submitted 24 February, 2021; **originally announced** February 2021.

Comments: **CVPR 2021** camera ready, typo fixed

49. arXiv:2102.09334 [pdf, other]

cs.CV

StablePose: Learning 6D Object Poses from Geometrically Stable Patches

Authors: Yifei Shi, Junwen Huang, Xin Xu, Yifan Zhang, Kai Xu

Abstract: We introduce the concept of geometric stability to the problem of 6D object pose estimation and propose to learn pose inference based on geometrically stable patches extracted from observed 3D point clouds. According to the theory of geometric stability analysis, a minimal set of three planar/cylindrical patches are geometrically stable and determine the full 6DoFs of the object pose. We train a d… ▽ More

Submitted 19 March, 2021; v1 submitted 18 February, 2021; **originally announced** February 2021.

Comments: **CVPR 2021**, 10 pages

50. arXiv:2102.02808 [pdf, other]

cs.CV

Multi-Stage Progressive Image Restoration

Authors: Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, Ling Shao

Abstract: Image restoration tasks demand a complex balance between spatial details and high-level contextualized information while recovering images. In this paper, we propose a novel synergistic design that can optimally balance these competing goals. Our main proposal is a multi-stage architecture, that progressively learns restoration functions for the degraded inputs, thereby breaking down the overall r… ▽ More

Submitted 16 March, 2021; v1 submitted 4 February, 2021; **originally announced** February 2021.

Comments: Accepted at **CVPR 2021**

1. arXiv:2101.10837 [pdf, other]

   cs.CV

   Ikshana: A Theory of Human Scene Understanding Mechanism

   Authors: Venkata Satya Sai Ajay Daliparthi

   Abstract: In recent years, deep neural networks achieved state-of-the-art performance on many computer vision tasks. The two most commonly observed drawbacks of these deep neural networks are: the requirement of a massive amount of labeled data and a vast number of parameters. In this work, we propose a theory named Ikshana, to explain the functioning of the human brain, while humans understand a natural sc… ▽ More

   Submitted 21 January, 2021; **originally announced** January 2021.

   Comments: 12 pages, 3 figures; Submitted to **CVPR 2021**

2. arXiv:2101.10511 [pdf, other]

   cs.CV

   Generic Event Boundary Detection: A Benchmark for Event Segmentation

   Authors: Mike Zheng Shou, Stan Lei, Weiyao Wang, Deepti Ghadiyaram, Matt Feiszli

   **Abstract**: …approaches on the TAPOS dataset and our Kinetics-GEBD, together with method design explorations that suggest future directions. We release our annotations and baseline codes at **CVPR**'21 LOVEU Challenge: https://sites.google.com/view/loveucvpr21 ▽ More

   Submitted 24 March, 2021; v1 submitted 25 January, 2021; originally announced January 2021.

3. arXiv:2101.09650 [pdf, other]

   cs.LG

   GST: Group-Sparse Training for Accelerating Deep Reinforcement Learning

   Authors: Juhyoung Lee, Sangyeob Kim, Sangjin Kim, Wooyoung Jo, Hoi-Jun Yoo

Abstract: Deep reinforcement learning (DRL) has shown remarkable success in sequential decision-making problems but suffers from a long training time to obtain such good performance. Many parallel and distributed DRL training approaches have been proposed to solve this problem, but it is difficult to utilize them on resource-limited devices. In order to accelerate DRL in real-world edge devices, memory band… ▽ More

Submitted 24 January, 2021; **originally announced** January 2021.

Comments: 10 pages, 10 figures, **CVPR 2021** Submitted

4. arXiv:2101.06742 [pdf, other]

cs.CV cs.AI cs.LG cs.RO stat.ML

doi10.1109/CVPR.2018.00274

Deep Parametric Continuous Convolutional Neural Networks

Authors: Shenlong Wang, Simon Suo, Wei-Chiu Ma, Andrei Pokrovsky, Raquel Urtasun

Abstract: Standard convolutional neural networks assume a grid structured input is available and exploit discrete convolutions as their fundamental building blocks. This limits their applicability to many real-world applications. In this paper we propose Parametric Continuous Convolution, a new learnable operator that operates over non-grid structured data. The key idea is to exploit parameterized kernel fu… ▽ More

Submitted 17 January, 2021; **originally announced** January 2021.

Comments: Accepted by **CVPR** 2018

5. arXiv:2101.06679 [pdf, other]

cs.CV cs.RO

End-to-end Interpretable Neural Motion Planner

Authors: Wenyuan Zeng, Wenjie Luo, Simon Suo, Abbas Sadat, Bin Yang, Sergio Casas, Raquel Urtasun

Abstract: In this paper, we propose a neural motion planner (NMP) for learning to drive autonomously in complex urban scenarios that include traffic-light handling, yielding, and interactions with multiple road-users. Towards this goal, we design a holistic model that takes as input raw LIDAR data and a HD map and produces interpretable intermediate representations in the form of 3D detections and their fut… ▽ More

Submitted 17 January, 2021; **originally announced** January 2021.

Comments: **CVPR** 2019 (Oral)

6. arXiv:2101.06184 [pdf, other]

cs.CV

Temporal-Relational CrossTransformers for Few-Shot Action Recognition

Authors: Toby Perrett, Alessandro Masullo, Tilo Burghardt, Majid Mirmehdi, Dima Damen

Abstract: We propose a novel approach to few-shot action recognition, finding temporally-corresponding frame tuples between the query and videos in the support set. Distinct from previous few-shot works, we construct class prototypes using the CrossTransformer attention mechanism to observe relevant sub-sequences of all support videos, rather than using class averages or single best matches. Video represent…
▽ More

Submitted 18 March, 2021; v1 submitted 15 January, 2021; **originally announced** January 2021.

Comments: Accepted in **CVPR 2021**

7. arXiv:2101.03787 [pdf, other]

cs.CV

WiCV 2020: The Seventh Women In Computer Vision Workshop

Authors: Hazel Doughty, Nour Karessli, Kathryn Leonard, Boyi Li, Carianne Martinez, Azadeh Mobasher, Arsha Nagrani, Srishti Yadav

**Abstract**: In this paper we present the details of Women in Computer Vision Workshop - WiCV 2020, organized in alongside virtual **CVPR** 2020. This event aims at encouraging the women researchers in the field of computer vision. It provides a voice to a minority (female) group in computer vision community and focuses on increasingly the visibility of these researchers, bo… ▽ More

Submitted 11 January, 2021; **originally announced** January 2021.

8. arXiv:2101.00588 [pdf, other]

cs.CV

Style Normalization and Restitution for DomainGeneralization and Adaptation

Authors: Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen

Abstract: For many practical computer vision applications, the learned models usually have high performance on the datasets used for training but suffer from significant performance degradation when deployed in new environments, where there are usually style differences between the training images and the testing images. An effective domain generalizable model is expected to be able to learn feature represe… ▽ More

Submitted 3 January, 2021; originally announced January 2021.

Comments: Conference version: Style Normalization and Restitution for Generalizable Person Re-identification, published on **CVPR**-2020

9. arXiv:2101.00529 [pdf, other]

cs.CV cs.AI cs.CL cs.LG

VinVL: Revisiting Visual Representations in Vision-Language Models

Authors: Pengchuan Zhang, Xiujun Li, Xiaowei Hu, Jianwei Yang, Lei Zhang, Lijuan Wang, Yejin Choi, Jianfeng Gao

Abstract: This paper presents a detailed study of improving visual representations for vision language (VL) tasks and develops an improved object detection model to provide object-centric representations of images. Compared to the most widely used \emph{bottom-up and top-down} model \cite{anderson2018bottom}, the new model is bigger, better-designed for VL tasks, and pre-trained on much larger training corp… ▽ More

Submitted 9 March, 2021; v1 submitted 2 January, 2021; **originally announced** January 2021.

Journal ref: **CVPR 2021**

10. arXiv:2012.13253 [pdf, other]

cs.LG cs.AI cs.CV

Soft-IntroVAE: Analyzing and Improving the Introspective Variational Autoencoder

Authors: Tal Daniel, Aviv Tamar

Abstract: The recently introduced introspective variational autoencoder (IntroVAE) exhibits outstanding image generations, and allows for amortized inference using an image encoder. The main idea in IntroVAE is to train a VAE adversarially, using the VAE encoder to discriminate between generated and real data samples. However, the original IntroVAE loss function relied on a particular hinge-loss formulation… ▽ More

Submitted 25 March, 2021; v1 submitted 24 December, 2020; originally announced December 2020.

Comments: **CVPR 2021**, Extended version. Code and additional information is available on the project website - https://taldatech.github.io/soft-intro-vae-web

11. arXiv:2012.08323 [pdf, other]

cs.CV

Improved Image Matting via Real-time User Clicks and Uncertainty Estimation

Authors: Tianyi Wei, Dongdong Chen, Wenbo Zhou, Jing Liao, Hanqing Zhao, Weiming Zhang, Nenghai Yu

Abstract: Image matting is a fundamental and challenging problem in computer vision and graphics. Most existing matting methods leverage a user-supplied trimap as an auxiliary input to produce good alpha matte. However, obtaining high-quality trimap itself is arduous, thus restricting the application of these methods. Recently, some trimap-free methods have emerged, however, the matting quality is still far… ▽ More

Submitted 7 March, 2021; v1 submitted 15 December, 2020; originally announced December 2020.

Comments: Accepted by IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), **2021**

12. arXiv:2012.08103 [pdf, other]

cs.CV

KOALAnet: Blind Super-Resolution using Kernel-Oriented Adaptive Local Adjustment

Authors: Soo Ye Kim, Hyeonjun Sim, Munchurl Kim

Abstract: Blind super-resolution (SR) methods aim to generate a high quality high resolution image from a low resolution image containing unknown degradations. However, natural images contain various types and amounts of blur: some may be due to the inherent degradation characteristics of the camera, but some may even be intentional, for aesthetic purposes (eg. Bokeh effect). In the case of the latter, it b… ▽ More

Submitted 5 March, 2021; v1 submitted 15 December, 2020; originally announced December 2020.

Comments: Accepted to **CVPR 2021**. The first two authors contributed equally to this work

13. arXiv:2012.04462 [pdf, other]

cs.CV

Multi-Objective Interpolation Training for Robustness to Label Noise

Authors: Diego Ortego, Eric Arazo, Paul Albert, Noel E. O'Connor, Kevin McGuinness

Abstract: Deep neural networks trained with standard cross-entropy loss memorize noisy labels, which degrades their performance. Most research to mitigate this memorization proposes new robust classification loss functions. Conversely, we propose a Multi-Objective Interpolation Training (MOIT) approach that jointly exploits contrastive learning and classification to mutually help each other and boost perfor… ▽ More

Submitted 18 March, 2021; v1 submitted 8 December, 2020; originally announced December 2020.

Comments: Accepted to **CVPR 2021**. 10 pages, 1 figure, and 9 tables

14. arXiv:2012.03408 [pdf, other]

cs.CV

PMP-Net: Point Cloud Completion by Learning Multi-step Point Moving Paths

Authors: Xin Wen, Peng Xiang, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, Yu-Shen Liu

Abstract: The task of point cloud completion aims to predict the missing part for an incomplete 3D shape. A widely used strategy is to generate a complete point cloud from the incomplete one. However, the unordered nature of point clouds will degrade the generation of high-quality 3D shapes, as the detailed topology and structure of discrete points are hard to be captured by the generative process only usin… ▽ More

Submitted 7 March, 2021; v1 submitted 6 December, 2020; originally announced December 2020.

Comments: Accepted by **CVPR 2021**

15. arXiv:2012.03308 [pdf, other]

cs.CV cs.AI cs.MM

TediGAN: Text-Guided Diverse Face Image Generation and Manipulation

Authors: Weihao Xia, Yujiu Yang, Jing-Hao Xue, Baoyuan Wu

Abstract: In this work, we propose TediGAN, a novel framework for multi-modal image generation and manipulation with textual descriptions. The proposed method consists of three components: StyleGAN inversion module, visual-linguistic similarity learning, and instance-level optimization. The inversion module maps real images to the latent space of a well-trained StyleGAN. The visual-linguistic similarity lea… ▽ More

Submitted 17 March, 2021; v1 submitted 6 December, 2020; originally announced December 2020.

Comments: **CVPR 2021**. Code: https://github.com/weihaox/TediGAN Data: https://github.com/weihaox/Multi-Modal-CelebA-HQ Video: https://youtu.be/L8Na2f5viAM

16. arXiv:2012.02107 [pdf, other]

cs.CV

Robust Instance Segmentation through Reasoning about Multi-Object Occlusion

Authors: Xiaoding Yuan, Adam Kortylewski, Yihong Sun, Alan Yuille

Abstract: Analyzing complex scenes with Deep Neural Networks is a challenging task, particularly when images contain multiple objects that partially occlude each other. Existing approaches to image analysis mostly process objects independently and do not take into account the relative occlusion of nearby objects. In this paper, we propose a deep network for multi-object instance segmentation that is robust… ▽ More

Submitted 1 March, 2021; v1 submitted 3 December, 2020; originally announced December 2020.

Comments: Accepted by **CVPR 2021**

17. arXiv:2012.00321 [pdf, other]

cs.CV cs.LG

Disentangling Label Distribution for Long-tailed Visual Recognition

Authors: Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, Buru Chang

Abstract: The current evaluation protocol of long-tailed visual recognition trains the classification model on the long-tailed source label distribution and evaluates its performance on the uniform target label distribution. Such protocol has questionable practicality since the target may also be long-tailed. Therefore, we formulate long-tailed visual recognition as a label shift problem where the target an… ▽ More

Submitted 20 March, 2021; v1 submitted 1 December, 2020; originally announced December 2020.

Comments: **CVPR 2021**

18. arXiv:2012.00313 [pdf, other]

cs.CV

Unsupervised Part Discovery via Feature Alignment

Authors: Mengqi Guo, Yutong Bai, Zhishuai Zhang, Adam Kortylewski, Alan Yuille

Abstract: Understanding objects in terms of their individual parts is important, because it enables a precise understanding of the objects' geometrical structure, and enhances object recognition when the object is seen in a novel pose or under partial occlusion. However, the manual annotation of parts in large scale datasets is time consuming and expensive. In this paper, we aim at discovering object parts… ▽ More

Submitted 1 December, 2020; originally announced December 2020.

Comments: 10 pages, 9 figures, submitted to **CVPR 2021**

19. arXiv:2011.13328 [pdf, other]

cs.CV

DyCo3D: Robust Instance Segmentation of 3D Point Clouds through Dynamic Convolution

Authors: Tong He, Chunhua Shen, Anton van den Hengel

Abstract: Previous top-performing approaches for point cloud instance segmentation involve a bottom-up strategy, which often includes inefficient operations or complex pipelines, such as grouping over-segmented components, introducing additional steps for refining, or designing complicated loss functions. The inevitable variation in the instance scales can lead bottom-up methods to become particularly sensi… ▽ More

Submitted 5 March, 2021; v1 submitted 26 November, 2020; originally announced November 2020.

Comments: Appearing in IEEE Conf. Computer Vision and Pattern Recognition (**CVPR**), **2021**

20. arXiv:2011.13005 [pdf, other]

cs.CV eess.IV

PREDATOR: Registration of 3D Point Clouds with Low Overlap

Authors: Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, Konrad Schindler

Abstract: We introduce PREDATOR, a model for pairwise point-cloud registration with deep attention to the overlap region. Different from previous work, our model is specifically designed to handle (also) point-cloud pairs with low overlap. Its key novelty is an overlap-attention block for early information exchange between the latent encodings of the two point clouds. In this way the subsequent decoding of… ▽ More

Submitted 19 March, 2021; v1 submitted 25 November, 2020; originally announced November 2020.

Comments: **CVPR 2021** (Oral) - Camera ready. The first two authors contributed equally

21. arXiv:2011.12172 [pdf, other]

cs.CV cs.AI

VIGOR: Cross-View Image Geo-localization beyond One-to-one Retrieval

Authors: Sijie Zhu, Taojiannan Yang, Chen Chen

Abstract: Cross-view image geo-localization aims to determine the locations of street-view query images by matching with GPS-tagged reference images from aerial view. Recent works have achieved surprisingly high retrieval accuracy on city-scale datasets. However, these results rely on the assumption that there exists a reference image exactly centered at the location of any query image, which is not applica… ▽ More

Submitted 22 March, 2021; v1 submitted 24 November, 2020; originally announced November 2020.

Comments: **CVPR 2021**

22. arXiv:2011.11390 [pdf, other]

cs.CV

PLOP: Learning without Forgetting for Continual Semantic Segmentation

Authors: Arthur Douillard, Yifu Chen, Arnaud Dapogny, Matthieu Cord

Abstract: Deep learning approaches are nowadays ubiquitously used to tackle computer vision tasks such as semantic segmentation, requiring large datasets and substantial computational power. Continual learning for semantic segmentation (CSS) is an emerging trend that consists in updating an old model by sequentially adding new classes. However, continual learning methods are usually prone to catastrophic fo… ▽ More

Submitted 11 March, 2021; v1 submitted 23 November, 2020; originally announced November 2020.

Comments: Accepted at **CVPR 2021**, code: https://github.com/arthurdouillard/CVPR2021_PLOP

23. arXiv:2011.10678 [pdf, other]

cs.CV cs.AI cs.LG

Open-Vocabulary Object Detection Using Captions

Authors: Alireza Zareian, Kevin Dela Rosa, Derek Hao Hu, Shih-Fu Chang

Abstract: Despite the remarkable accuracy of deep neural networks in object detection, they are costly to train and scale due to supervision requirements. Particularly, learning more object categories typically requires proportionally more bounding box annotations. Weakly supervised and zero-shot learning techniques have been explored to scale object detectors to more categories with less supervision, but t… ▽ More

Submitted 14 March, 2021; v1 submitted 20 November, 2020; originally announced November 2020.

Comments: To be presented at **CVPR 2021** (oral paper)

24. arXiv:2011.09230 [pdf, other]

cs.CV

FixBi: Bridging Domain Spaces for Unsupervised Domain Adaptation

Authors: Jaemin Na, Heechul Jung, Hyung Jin Chang, Wonjun Hwang

Abstract: Unsupervised domain adaptation (UDA) methods for learning domain invariant representations have achieved remarkable progress. However, most of the studies were based on direct adaptation from the source domain to the target domain and have suffered from large domain discrepancies. In this paper, we propose a UDA method that effectively handles such large domain discrepancies. We introduce a fixed… ▽ More

Submitted 25 March, 2021; v1 submitted 18 November, 2020; originally announced November 2020.

Comments: Accepted to **CVPR 2021**

25. arXiv:2011.08464 [pdf, other]

cs.CV cs.LG cs.RO

Exploring intermediate representation for monocular vehicle pose estimation

Authors: Shichao Li, Zengqiang Yan, Hongyang Li, Kwang-Ting Cheng

Abstract: We present a new learning-based approach to recover egocentric 3D vehicle pose from a single RGB image. In contrast to previous works that directly map from local appearance to 3D angles, we explore a progressive approach by extracting meaningful Intermediate Geometrical Representations (IGRs) for 3D pose estimation. We design a deep model that transforms perceived intensities to IGRs, which are m… ▽ More

Submitted 16 March, 2021; v1 submitted 17 November, 2020; originally announced November 2020.

Comments: Accepted to **CVPR 2021**

26. arXiv:2011.08435 [pdf, other]

cs.LG cs.AI cs.CV

AdCo: Adversarial Contrast for Efficient Learning of Unsupervised Representations from Self-Trained Negative Adversaries

Authors: Qianjiang Hu, Xiao Wang, Wei Hu, Guo-Jun Qi

Abstract: Contrastive learning relies on constructing a collection of negative examples that are sufficiently hard to discriminate against positive queries when their representations are self-trained. Existing contrastive learning methods either maintain a queue of negative samples over minibatches while only a small portion of them are updated in an iteration, or only use the other examples from the curren… ▽ More

Submitted 5 March, 2021; v1 submitted 17 November, 2020; originally announced November 2020.

Comments: Appendices with more results on symmetric loss, different numbers of negative samples, computing costs is presented. We also discuss "whether we still need negative examples" in Appendix C, a question emerging from the comparison with the BYOL. The source code is also available at https://github.com/maple-research-lab/AdCo/

Journal ref: IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), June 19th - June 25th, **2021**

27. arXiv:2011.07491 [pdf, other]

cs.CV cs.LG eess.IV

Anomaly Detection in Video via Self-Supervised and Multi-Task Learning

Authors: Mariana-Iuliana Georgescu, Antonio Barbalau, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, Mubarak Shah

Abstract: Anomaly detection in video is a challenging computer vision problem. Due to the lack of anomalous events at training time, anomaly detection requires the design of learning methods without full supervision. In this paper, we approach anomalous event detection in video through self-supervised and multi-task learning at the object level. We first utilize a pre-trained detector to detect objects. The… ▽ More

Submitted 6 March, 2021; v1 submitted 15 November, 2020; originally announced November 2020.

Comments: Accepted at **CVPR 2021**. Main paper and supplementary are both included

28. arXiv:2008.13367 [pdf, other]

cs.CV

VarifocalNet: An IoU-aware Dense Object Detector

Authors: Haoyang Zhang, Ying Wang, Feras Dayoub, Niko Sünderhauf

Abstract: Accurately ranking the vast number of candidate detections is crucial for dense object detectors to achieve high performance. Prior work uses the classification score or a combination of classification and predicted localization scores to rank candidates. However, neither option results in a reliable ranking, thus degrading detection performance. In this paper, we propose to learn an Iou-aware Cla… ▽ More

Submitted 4 March, 2021; v1 submitted 31 August, 2020; originally announced August 2020.

Comments: Accepted to **CVPR 2021** as an oral

29. arXiv:2007.08558 [pdf, other]

cs.CV cs.LG

On Robustness and Transferability of Convolutional Neural Networks

Authors: Josip Djolonga, Jessica Yung, Michael Tschannen, Rob Romijnders, Lucas Beyer, Alexander Kolesnikov, Joan Puigcerver, Matthias Minderer, Alexander D'Amour, Dan Moldovan, Sylvain Gelly, Neil Houlsby, Xiaohua Zhai, Mario Lucic

Abstract: Modern deep convolutional networks (CNNs) are often criticized for not generalizing under distributional shifts. However, several recent breakthroughs in transfer learning suggest that these networks can cope with severe distribution shifts and successfully adapt to new tasks from a few training examples. In this work we study the interplay between out-of-distribution and transfer performance of m… ▽ More

Submitted 23 March, 2021; v1 submitted 16 July, 2020; originally announced July 2020.

Comments: Accepted at **CVPR 2021**

30. arXiv:2007.04171 [pdf, other]

cs.CV cs.LG

Domain Adaptation with Auxiliary Target Domain-Oriented Classifier

Authors: Jian Liang, Dapeng Hu, Jiashi Feng

Abstract: Domain adaptation (DA) aims to transfer knowledge from a label-rich but heterogeneous domain to a label-scare domain, which alleviates the labeling efforts and attracts considerable attention. Different from previous methods focusing on learning domain-invariant feature representations, some recent methods present generic semi-supervised learning (SSL) techniques and directly apply them to DA task… ▽ More

Submitted 25 March, 2021; v1 submitted 8 July, 2020; originally announced July 2020.

Comments: **CVPR 2021** camera ready. Code is available at https://github.com/tim-learn/ATDOC

31. arXiv:2006.07229 [pdf, other]

cs.CV

A Sliced Wasserstein Loss for Neural Texture Synthesis

Authors: Eric Heitz, Kenneth Vanhoey, Thomas Chambon, Laurent Belcour

Abstract: We address the problem of computing a textural loss based on the statistics extracted from the feature activations of a convolutional neural network optimized for object recognition (e.g. VGG-19). The underlying mathematical problem is the measure of the distance between two distributions in feature space. The Gram-matrix loss is the ubiquitous approximation for this problem but it is subject to s… ▽ More

Submitted 11 March, 2021; v1 submitted 12 June, 2020; originally announced June 2020.

Comments: 9 pages, 14 figures, accepted at **CVPR 2021**

Journal ref: IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**) **2021**

32. arXiv:2006.02635 [pdf, other]

cs.CL cs.CV

M3P: Learning Universal Representations via Multitask Multilingual Multimodal Pre-training

Authors: Minheng Ni, Haoyang Huang, Lin Su, Edward Cui, Taroon Bharti, Lijuan Wang, Dongdong Zhang, Nan Duan

Abstract: We present M3P, a Multitask Multilingual Multimodal Pre-trained model that combines multilingual pre-training and multimodal pre-training into a unified framework via multitask pre-training. Our goal is to learn universal representations that can map objects occurred in different modalities or texts expressed in different languages into a common semantic space. In addition, to explicitly encourage… ▽ More

Submitted 22 March, 2021; v1 submitted 3 June, 2020; originally announced June 2020.

Comments: Accepted to **CVPR 2021**

33. arXiv:2003.12243 [pdf, other]

cs.CV

Dynamic Region-Aware Convolution

Authors: <u>Jin Chen</u>, <u>Xijun Wang</u>, <u>Zichao Guo</u>, <u>Xiangyu Zhang</u>, <u>Jian Sun</u>

Abstract: We propose a new convolution called Dynamic Region-Aware Convolution (DRConv), which can automatically assign multiple filters to corresponding spatial regions where features have similar representation. In this way, DRConv outperforms standard convolution in modeling semantic variations. Standard convolutional layer can increase the number of filers to extract more visual elements but results in… ▽ More

Submitted 15 March, 2021; v1 submitted 27 March, 2020; originally announced March 2020.

Comments: Accepted at **CVPR 2021**

34. <u>arXiv:2003.04857</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Image Restoration for Under-Display Camera

Authors: <u>Yuqian Zhou</u>, <u>David Ren</u>, <u>Neil Emerton</u>, <u>Sehoon Lim</u>, <u>Timothy Large</u>

Abstract: The new trend of full-screen devices encourages us to position a camera behind a screen. Removing the bezel and centralizing the camera under the screen brings larger display-to-body ratio and enhances eye contact in video chat, but also causes image degradation. In this paper, we focus on a newly-defined Under-Display Camera (UDC), as a novel real-world single image restoration problem. First, we… ▽ More

Submitted 14 March, 2021; v1 submitted 10 March, 2020; originally announced March 2020.

Comments: Accepted by **CVPR 2021**

35. <u>arXiv:2003.04279</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Restore from Restored: Video Restoration with Pseudo Clean Video

Authors: <u>Seunghwan Lee</u>, <u>Donghyeon Cho</u>, <u>Jiwon Kim</u>, <u>Tae Hyun Kim</u>

Abstract: In this study, we propose a self-supervised video denoising method called "restore-from-restored." This method fine-tunes a pre-trained network by using a pseudo clean video during the test phase. The pseudo clean video is obtained by applying a noisy video to the baseline network. By adopting a fully convolutional neural network (FCN) as the baseline, we can improve video denoising performance wi… ▽ More

Submitted 15 March, 2021; v1 submitted 9 March, 2020; originally announced March 2020.

Comments: **CVPR 2021**

36. <u>arXiv:2003.03711</u> [<u>pdf</u>, <u>other</u>]

cs.CV

Single-View 3D Object Reconstruction from Shape Priors in Memory

Authors: <u>Shuo Yang</u>, <u>Min Xu</u>, <u>Haozhe Xie</u>, <u>Stuart Perry</u>, <u>Jiahao Xia</u>

Abstract: Existing methods for single-view 3D object reconstruction directly learn to transform image features into 3D representations. However, these methods are vulnerable to images containing noisy backgrounds and heavy occlusions because the extracted image features do not contain enough information to reconstruct high-quality 3D shapes. Humans routinely use incomplete or noisy visual cues from an image… ▽ More

Submitted 4 March, 2021; v1 submitted 7 March, 2020; originally announced March 2020.

Comments: **CVPR 2021**

37. arXiv:2002.05712 [pdf, other]

cs.LG cs.CV stat.ML

Cross-Iteration Batch Normalization

Authors: Zhuliang Yao, Yue Cao, Shuxin Zheng, Gao Huang, Stephen Lin

Abstract: A well-known issue of Batch Normalization is its significantly reduced effectiveness in the case of small mini-batch sizes. When a mini-batch contains few examples, the statistics upon which the normalization is defined cannot be reliably estimated from it during a training iteration. To address this problem, we present Cross-Iteration Batch Normalization (CBN), in which examples from multiple rec… ▽ More

Submitted 25 March, 2021; v1 submitted 13 February, 2020; originally announced February 2020.

Comments: Accepted to **CVPR 2021**

38. arXiv:1910.01523 [pdf, other]

cs.LG cs.NE

ReNAS:Relativistic Evaluation of Neural Architecture Search

Authors: Yixing Xu, Yunhe Wang, Kai Han, Yehui Tang, Shangling Jui, Chunjing Xu, Chang Xu

Abstract: An effective and efficient architecture performance evaluation scheme is essential for the success of Neural Architecture Search (NAS). To save computational cost, most of existing NAS algorithms often train and evaluate intermediate neural architectures on a small proxy dataset with limited training epochs. But it is difficult to expect an accurate performance estimation of an architecture in suc… ▽ More

Submitted 22 March, 2021; v1 submitted 29 September, 2019; originally announced October 2019.

Comments: **CVPR 2021**, Oral

39. arXiv:1907.07174 [pdf, other]

cs.LG cs.CV stat.ML

Natural Adversarial Examples

Authors: Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, Dawn Song

Abstract: We introduce two challenging datasets that reliably cause machine learning model performance to substantially degrade. The datasets are collected with a simple adversarial filtration technique to create datasets with limited spurious cues. Our datasets' real-world, unmodified examples transfer to various unseen models reliably, demonstrating that computer vision models have shared weaknesses. The… ▽ More