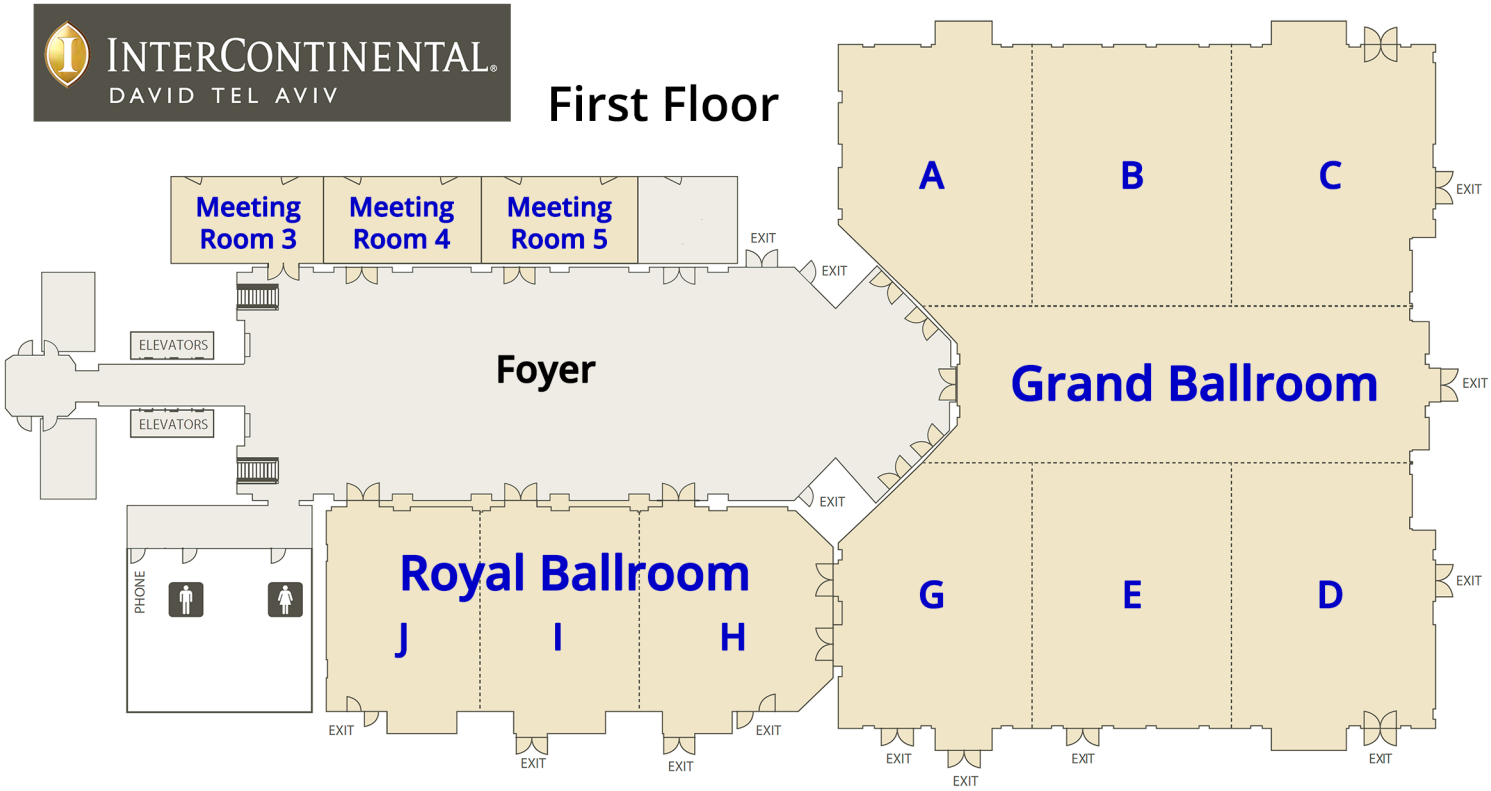
EUROPEAN CONFERENCE ON COMPUTER VISION TEL AVIV 2022

Program Guide

Workshops & Tutorials

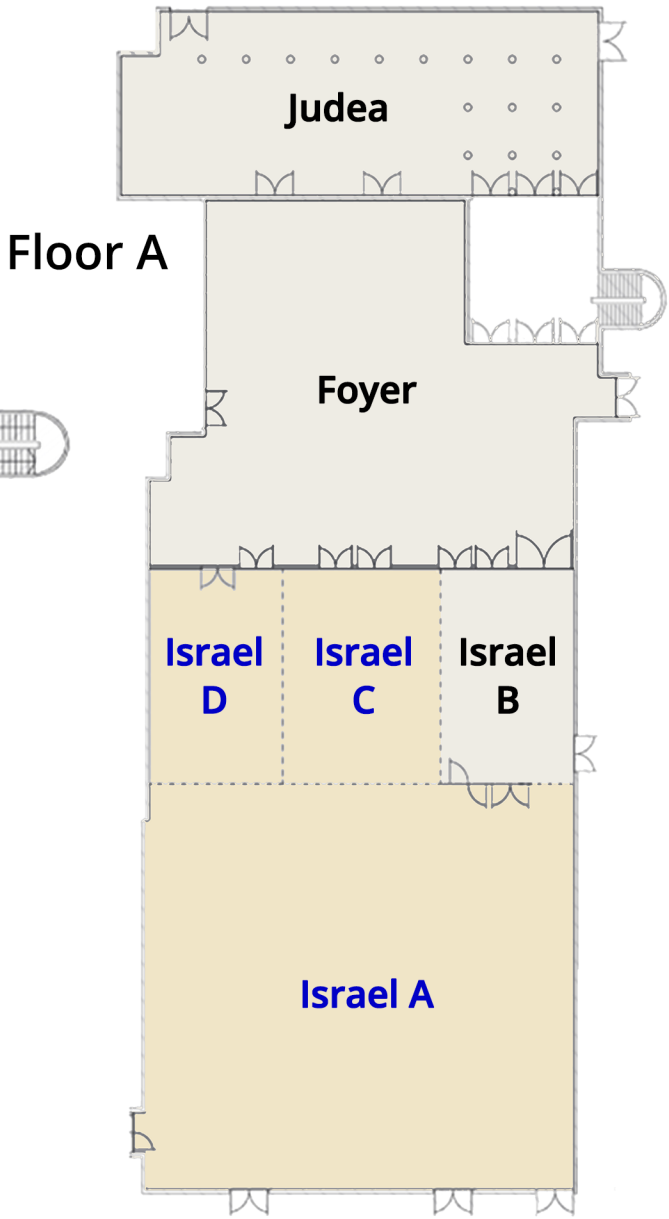
InterContinental
DAVID TEL AVIV

First Floor

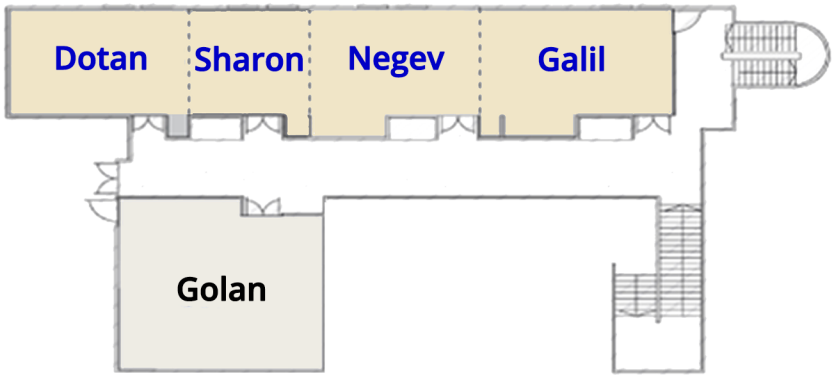


Dan Panorama

Floor A



Floor B



ECCV
TEL AVIV 2022

Message from the General and Program Chairs

Welcome to the 2022 European Conference on Computer Vision in Tel-Aviv, Israel. This is the place to see the latest research results, listen to distinguished keynote speakers, attend tutorials, workshops and demos, and have a chance to meet old friends and make new ones.

The proceedings you virtually hold in your hands are the result of a collective effort. 18310 authors submitted a total of 6773 papers that were handled by 276 area chairs (ACs) that solicited the help of 4719 reviewers. The entire process was supervised by the program chairs (PCs) with the constant support of the general chairs (GCs). In the end, we have 1645 (28%) papers that were accepted for publication, including 157 Orals (2.7%).

The double-blind reviewing process was handled by the CMT system. Authors did not know the name of the reviewers and vice versa. 846 of the submissions were desk-rejected for various reasons. Many of them because they revealed author identity, thus violating the double-blind policy. Some papers were withdrawn at different stages leaving us with a total 5804 valid submissions.

Each of these submissions received at least three reviews (except 6 papers that received only 2 reviews), totalling more than 15,000 reviews. Authors had a chance to submit a rebuttal, followed by a discussion between the area chairs (ACs) and the reviewers assigned to each paper. The final decision of each paper was taken by the AC in consultation with a buddy AC, to make sure the decisions are as fair and informative as possible. The process was monitored by the PCs, with a special emphasis on cases where the decision of the AC differs from the consensus recommendation of the reviewers.

The planning of ECCV 2022 had to deal with the uncertainties of the COVID-19 pandemic. ECCV 2022 is, still, a hybrid conference that gives researchers the ability to attend the conference either virtually or in-person. Based on past experience, we have decided that only in-person attendees can present their work on site. In addition, all attendees, in-person or virtual, can watch a 5-minute video of each of the papers on the virtual platform.

The conference runs for three days and includes two parallel oral sessions per day, as well as two poster sessions per day. In addition, there are two days of workshops and tutorials, as well as events that happen in parallel to the main conference, including mentoring sessions, industrial exhibition, academic demos, and an industrial track focused on entrepreneurs.

A separate committee was tasked with selecting the best paper award, along with the honorable mentions. The selected papers will be announced in a special session during the conference.

We have a long list of people to thank. We thank Pavel Lifshitz, our Technical Program Chair, for working tirelessly behind the scenes. We thank our demo chairs, workshop chairs, communication chairs, best-paper committee members, social activities chairs, industry track chairs, and the diversity committee members for helping us along the way. We benefited from the experience and advice of Nicole Finn regarding organizational aspects and thank her for that. A special thanks to the entire ORTRA team for organizing the conference. And last, but not least, we thank you for submitting a paper, reviewing papers and attending. We hope you enjoy ECCV'2022!

Program Chairs: Shai Avidan, Gabriel Brostow, Giovanni Maria Farinella, and Tal Hassner

General Chairs: Rita Cucchiara, Jiří Matas, Amnon Shashua, and Lihi Zelnik-Manor

ECCV 2022 Organizing Committee

General Chairs: Rita Cucchiara, Jiří Matas, Amnon Shashua, Lihi Zelnik-Manor

Program Chairs: Shai Avidan, Gabriel Brostow, Giovanni Maria Farinella, Tal Hassner

Program Technical Chair: Pavel Lifshits

Workshops Chairs: Leonid Karlinsky, Tomer Michaeli, Ko Nishino

Tutorials Chairs: Thomas Pock, Natalia Neverova

Demo Chair: Bohyung Han

Finance Chair: Gerard Medioni

Publications Chair: Eric Mortensen

Social & Student Activities Chairs: Tatiana Tommasi, Sagie Benaim

Industrial Liaison Chairs: Dimosthenis Karatzas, Chen Sagiv

Industry Track Chairs: Amir Markovitz, Yair Kittenplon

Communications Chairs: Lorenzo Baraldi, Kosta Derpanis

Diversity and Inclusion Chairs: Xi Yin, Bryan Russell

Award Papers Committee: Laura Leal-Taixé (chair), Ko Nishino, Yanxi Liu, Alessio del Bue, Todd Zickler, Jianbo Shi, Oisin Mac Aodha

Sunday, October 23

NOTE: Use the QR code for each workshop's website to find the workshop's schedule. Here's the QR code to the ECCV Workshops page. Unless otherwise noted, all times are Israel Daylight Time (UTC+3)

0700–1900 Registration (David Intercontinental Hotel)

0700–1900 Coffee & Refreshments (Hotel Foyer)

1300–1400 Lunch (Foyer)

Vision for Art

Organizers: Alessio Del Bue, Peter Bell, Leonardo L. Impett, Noa Garcia, Stuart James

Location: Royal Ballroom J

Time: Full Day (0930-1730)

Summary: The VISion for ART (VISART) workshop is the forum for the presentation, discussion and publication of Computer Vision (CV) techniques for the understanding of art. This workshop brings together leading researchers in the fields of CV, ML, IR with Art History, Visual Studies, Digital Humanities and museum curators to focus on art and Cultural Heritage problems. The potential uses of Computer Vision for cultural history and cultural analytics have created great interest in the Humanities, with large projects on applying Computer Vision in galleries and museums, including the Getty and MoMA (in collaboration with Google). A key feature of this workshop is the close collaboration between scholars of Computer Vision and the Arts and Humanities, thus both exposing new technical possibilities to the arts and humanities, as well as offering new artistic and humanistic perspectives on computer vision.

Self-Supervised Learning: What Is Next?

Organizers: Yuki M. Asano, Christian Rupprecht, Diane Larlus, Andrew Zisserman

Location: Grand Ballroom C

Time: Full Day (0900-1800)

Summary: The past two years have seen major advances in self-supervised learning, with many new methods reaching astounding performances on standard benchmarks. Moreover, recent work has shown the large potential of coupled data sources such as image-text in producing even stronger models capable of zero-shot tasks, and often inspired by NLP. We have just witnessed a jump from the "default" single-modal pretraining with CNNs to transformer-based multi-modal training, and these early developments will surely mature in the coming months. However, it is also apparent that there are still major unresolved challenges and it's not clear what the next step-change is going to be. In this workshop we want to highlight and provide a forum to discuss potential research direction seeds, from radically new self-supervision tasks, data sources and paradigms to surprising counter-intuitive results. Through invited speakers and paper oral talks, our goal is to provide a forum to discuss and exchange ideas where both the leaders in this field, as well as the new, younger generation can equally contribute to discussing the future of this field.

AV4D: Visual Learning of Sounds in Spaces

Organizers: Changan Chen, Ruohan Gao, David Harwath, Chuang Gan, Andrew Owens, Andrea Vedaldi, Antonio Torralba, Kristen Grauman

Location: Royal Ballroom H

Time: Full Day (0900-1800)

Summary: We see and hear things every second of our lives. Before sounds arrive at our ears, they are first produced by some objects situated in the space, and then undergo the transformation of its surrounding space as a function of the geometry of the environment, materials, etc. Our perceived binaural sound not only tells us about the semantic property of the sound, e.g., telephone ringing, baby crying, but also helps us infer the spatial location of the sounding object. Both of these acoustic and spatial properties are captured by the visual stream, and require models to go beyond 2D understanding of images (3D with audio) and study the spatial (3D) aspect of audio in visuals (4D with audio). This is of vital importance for applications such as egocentric video understanding, robotic perception, AR/VR, etc. In support of robotic perception, where embodied agents can move around with both visual and auditory sensing, audio-visual simulations are also recently developed to facilitate research in this direction. The goal of this workshop is to share recent progress of audio-visual studies on the spatial-temporal (4D) dimensions, and also to discuss which directions the field should investigate next.

Responsible Computer Vision

Organizers: Laurens van der Maaten, Deepti Ghadiyaram, Reza Shokri, Angelina Wang, Judy Hoffman, Cristian Canton Ferrer, Been Kim, Vicente Ordonez, Aaron Adcock

Location: Royal Ballroom G

Time: Full Day (1000-1600)

Summary: Computer vision systems that rely on large-scale datasets are being used in several real-world products, including social media platforms such as Instagram, Pinterest, and Twitter; autonomous control systems such as self-driving cars; and web services such as Google Photos. Recent critical analyses of these modern computer vision systems have brought to light the existence of biases, including those trained on open-source datasets. In particular, the iterative process of deployment and refinement of an already biased system can lead to systems that do not serve underrepresented communities (such as gender minorities or people of darker skin tones) well and amplify societal biases. Following the success of the CVPR 2021 Workshop on Responsible Computer Vision, we wish to continue to provide a common platform and spark healthy conversations that address such critical socio-technical concerns that span the entire length of the computer vision pipeline: beginning from responsible data collection and moving towards robust, privacy-aware, interpretable, and fair evaluation frameworks and modeling solutions. The rapidly advancing landscape of computer vision research necessitates fairness discussions that cater specifically to computer vision methods, and can at times depart from the broader discussion on responsible AI. We believe our workshop is an important step in developing a community-wide roadmap for responsible computer vision.

Map-Based Localization for Autonomous Driving

Organizers: Patrick Wenzel, Niclas Zeller, Lukas Koestler, Daniel Cremers

Location: Meeting Room 3

Time: Full Day (0900-1700)

Summary: This is the 3rd workshop on map-based localization in the context of autonomous driving (AD). By map-based localization, we understand the problem of accurately localizing (estimating the ego-position and -orientation) an autonomous vehicle in real-time in a pre-built map. Centimeter-accurate continuous global localization is a key feature for AD as it allows to position and tracks the ego-vehicle precisely within an HD map which contains important information about the environment. Being able to accurately localize within a pre-build map using standard perceptive sensors (e.g., camera, radar, LiDAR) extends the operation to GNSS-denied environments such as urban canyons or tunnels.

This task comprises several challenges including the question on how to create maps that are compressed in size and guarantee reliable localization independent of environmental conditions (e.g. weather, lighting, the season of the year) as well as keeping them up-to-date. Another aspect is the right sensor choice (with respect to robustness, accuracy, price) for both map generation and online localization.

Besides discussing the importance of map-based localization with experts from academia and industry (Abhinav Valada, Andrew Davison, Henning Lategahn, Philipp Krähenbühl, Yuning Chai), the workshop will host the 3rd re-localization challenge for autonomous driving based on the 4Seasons dataset.

Medical Computer Vision

Organizers: Tal Arbel, Nicolas Padoy, Ayelet Akselrod-Balin, Tammy Riklin Raviv, Vasileios Belagiannis, Mathias Unberath, Qi Dou, Yuyin Zhou, Moti Freiman

Location: Grand Ballroom B

Time: Full Day (0900-1800)

Summary: The MCV workshop will provide an opportunity to students, researchers and developers in biomedical imaging companies to present, discuss and learn recent advancements in medical image analysis. The ultimate goal of the workshop is leveraging big data, deep learning and novel representation to effectively build the next generation of robust quantitative medical imaging parsing tools and products. Prominent applications include large scale cancer screening, computational heart modeling, landmark detection, neural structure and functional labeling and image-guided intervention. Computer Vision advancements and Deep learning in particular are rapidly transitioned to the medical imaging community in recent years. Additionally, there is a tremendous growth in startup activity applying medical computer vision algorithms to the healthcare industry. Collecting and accessing radiological patient images is a challenging task. Recent efforts include VISCERAL Challenge and Alzheimer's Disease Neuroimaging Initiative. The NIH and partners are working on extracting trainable anatomical and pathological semantic labels from radiology reports that are linked to patients' CT/MRI/X-ray images or volumes such as NCI's Cancer Imaging Archive. The MCV workshop aims to encourage the establishment of public medical datasets to be used as unbiased platforms to compare performances on the same set of data for various disease findings.

AI for Space

Organizers: Tat-Jun Chin, Viorela Ila, Luca Carlone, Benjamin Morrell, Djamila Aouada, Grzegorz Kakareko, Binfeng Pan

Location: Virtual Room 12

Time: Full Day (0900-2100)

Summary: The space sector is experiencing significant growth. Currently planned activities and utilisation models also greatly exceed the scope, ambition and/or commercial value of space missions in the previous century, e.g., autonomous spacecraft, space mining, and understanding the universe. Achieving these ambitious goals requires surmounting non-trivial technical obstacles. AI4Space focuses on the role of AI, particularly computer vision and machine learning, in helping to solve those technical hurdles. The workshop will highlight the space capabilities that draw from and/or overlap significantly with vision and learning research, outline the unique difficulties presented by space applications to vision and learning, and discuss recent advances towards overcoming those obstacles.

Learning With Limited and Imperfect Data

Organizers: Noel C. Codella, Xiaojuan Qi, Zsolt Kira, Sadeep Jayasumana, Shuai Zheng, Viraj Prabhu, Judy Hoffman, Yunhui Guo, Tatiana Tommasi, Ming-Ming Cheng

Location: Virtual Room 2

Time: Full Day (0900-1700)

Summary: Learning from limited or imperfect data (L²ID) refers to a variety of studies that attempt to address challenging pattern recognition tasks by learning from limited, weak, or noisy supervision. Supervised learning methods including Deep Convolutional Neural Networks have significantly improved the performance in many problems in the field of computer vision, thanks to the rise of large-scale annotated data sets and the advance in computing hardware. However, these supervised learning approaches are notoriously "data hungry", which makes them sometimes not practical in many real-world industrial applications. This issue of availability of large quantities of labeled data becomes even more severe when considering visual classes that require annotation based on expert knowledge (e.g., medical imaging), classes that rarely occur, or object detection and instance segmentation tasks where the labeling requires more effort. To address this problem, many efforts, e.g., weakly supervised learning, few-shot learning, self/semi-supervised, cross-domain few-shot learning, domain adaptation, etc., have been made to improve robustness to this scenario. The goal of this workshop, which builds on the successful CVPR 2021 L2ID workshop, is to bring together researchers across several computer vision and machine learning communities to navigate the complex landscape of methods that enable moving beyond fully supervised learning towards limited and imperfect label settings.

Adversarial Robustness in the Real World

Organizers: Angtian Wang, Yutong Bai, Adam Kortylewski, Cihang Xie, Alan Yuille, Xinyun Chen, Judy Hoffman, Wieland Brendel, Matthias Hein, Hang Su, Dawn Song, Jun Zhu, Philippe Burlina, Rama Chellappa, Yinpeng Dong, Yingwei Li, Ju He, Alexander Robey

Location: Virtual Room 8

Time: Full Day (0900-1700)

Summary: Recent deep-learning-based methods achieve great performance on various vision applications. However, insufficient robustness on adversarial cases limits real-world applications of deep-learning-based methods. AROW workshop aims to explore adversarial examples, as well as, evaluate and improve the adversarial robustness of computer vision systems. In the AROW workshop we discuss topics includes: Improving model robustness against unrestricted adversarial attacks; Improving generalization to out-of-distribution samples or unforeseen adversaries; Discovery of real-world adversarial examples; Novel architectures with robustness to occlusion, viewpoint, and other real-world domain shifts; Domain adaptation techniques for robust vision in the real world; Datasets for evaluating model robustness; Structured deep models and explainable AI.

Robust Vision Challenge

Organizers: Oliver Zendel, Angela Dai, Xavier Puig, Andreas Geiger, Vladlen Koltun, Peter Kontschieder, Adam Kortylewski, Alina Kuznetsova, Tsung-Yi Lin, Torsten Sattler, Daniel Scharstein, Hendrik Schilling, Jonas Uhrig, Wulff Jonas

Location: Virtual Room 13

Time: Full Day (0900-1700)

Summary: The increasing availability of large annotated datasets such as Middlebury, PASCAL VOC, ImageNet, MS COCO, KITTI and Cityscapes has led to tremendous progress in computer vision and machine learning over the last decade. Public leaderboards make it easy to track the state-of-the-art in the field by comparing the results of dozens of methods side-by-side. While steady progress is made on each individual dataset, many of them are limited to specific domains. KITTI, for example, focuses on real-world urban driving scenarios, while Middlebury considers indoor scenes. Consequently, methods that are state-of-the-art on one dataset often perform worse on a different one or require substantial adaptation of the model parameters. The goal of this challenge is to foster the development of vision systems that are robust and consequently perform well on a variety of datasets with different characteristics. Towards this goal, we propose the Robust Vision Challenge, where performance on several tasks (eg, reconstruction, optical flow, semantic/instance segmentation, single image depth prediction) is measured across a number of challenging benchmarks with different characteristics, e.g., indoors vs. outdoors, real vs. synthetic, sunny vs. bad weather, different sensors. We encourage submissions of novel algorithms, techniques which are currently in review and methods that have already been published.

Machine Visual Common Sense: Perception, Prediction, Planning

Organizers: Yining Hong, Hsiao-Yu Tung, Kevin Smith, Zhenfang Chen, Elias Wang, Tianmin Shu, Kanishk V. Gandhi, Joshua Tenenbaum, Antonio Torralba, Daniel Yamins, Judy Fan, Chuang Gan

Location: Virtual Room TBA

Time: Full Day (0845-1700 PDT) (1845-0300 next day UTC+3)

Summary: Over the years, there have been a variety of visual reasoning tasks that evaluate machines' ability to understand and reason about visual scenes. However, these benchmarks mostly focus on classification of objects and items that exist in a scene. Common sense reasoning – an understanding of what might happen next, or what gave rise to the scene – is often absent in these benchmarks. Humans, on the other hand, are highly versatile, adept in numerous high-level cognition-related visual reasoning tasks that go beyond pattern recognition and require common sense (e.g., physics, causality, functionality, psychology, etc).

In order to design systems with human-like visual understanding of the world, we would like to emphasize benchmarks and tasks that evaluate common sense reasoning across a variety of domains, including but not limited to:

- Intuitive Physics: A general understanding and expectations about the physical world (e.g., how things support, collide, fall, contain, become unstable etc.)
- Intuitive Psychology & Social Science: A basic understanding of inter-relations and interaction of agents; An understanding of instrumental actions (e.g., assistance, imitation, speech etc.); The ability to reason about hidden mental variables that drive observable actions.
- Affordance & Functionality: What actions of agents can be applied to objects; What functions objects provide for the agents.
- Causality & Counterfactual Thinking: Understanding of causes and effects; Mental representations of alternatives to past or future events, actions, or states.

Designing and Evaluating Computer Perception Systems

Organizers: Viorica Patraucean, Joao Carreira, Dima Damen, Andrew Zisserman

Location: Virtual Room 6

Time: Full Day (1100-2000)

Summary: Much of the research in Computer Vision has focused on solving individual tasks like image or video classification, object detection, object tracking, to name only a few. These settings are suitable for narrow applications, but for complex applications like embodied intelligent assistants or self-driving cars we will need full perception models that deal with multiple modalities and have integrated scene understanding and reasoning capabilities comparable to humans. In this workshop, we propose to zoom out from Computer Vision and discuss more generally about Computer Perception, with leading thinkers from Computer Vision/Machine Learning and Cognitive Sciences. We will cover both modelling challenges and evaluation best practices.

Mobile Intelligent Photography and Imaging

Organizers: Chongyi Li, Wenxiu Sun, Shangchen Zhou, Chen Change Loy, Ruicheng Feng, Jinwei Gu, Jun Jiang

Location: Virtual Room 4

Time: Full Day (0900-1730)

Summary: This is the first workshop on mobile intelligent photography and imaging (MIPI). The workshop emphasizes the integration of novel image sensors and imaging algorithms. Together with the workshop, we organize five exciting challenge tracks, including RGB+ToF Depth Completion, Quad-Bayer Re-mosaic, RGBW Sensor Re-mosaic, RGBW Sensor Fusion, and Under-display Camera Image Restoration. The challenge attracted hundreds of participations. The workshop also received high-quality workshop papers. At this workshop, the winner teams and the authors of the best workshop papers will be invited to present their work. We also invite renowned keynote speakers from both industry and academia to share their insights and recent work. This workshop will provide a fertile ground for researchers, scientists, and engineers from around the world to disseminate their research outcomes and push forward the frontiers of knowledge within novel image sensors and imaging systems-related areas.

Computer Vision in the Wild

Organizers: Chunyuan Li, Haotian Liu, Jyoti Aneja, Haotian Zhang, Jianwei Yang, Liunian Li, Xin Wang, Aishwarya Kamath, Pengchuan Zhang

Location: Virtual Room TBA

Time: Full Day (0900-1700 PDT) (1900-0300 next day UTC+3)

Summary: Recent works show that learning from large-scale image-text data is a promising approach to building transferable visual models that can effortlessly adapt to a wide range of downstream computer vision (CV) and multimodal (MM) tasks. For example, CLIP, ALIGN and Florence for image classification, ViLD, RegionCLIP and GLIP for object detection. These vision models with language interface are naturally open-vocabulary recognition models, showing superior zero-shot and few-shot adaption performance on various real-world scenarios. We propose this "Computer Vision in the Wild" workshop, aiming to gather academic and industry communities to work on CV problems in real-world scenarios, focusing on the challenge of open-set/domain visual recognition and efficient task-level transfer. Since there are no established benchmarks to measure the progress of "CV in the Wild", we develop new benchmarks for image classification and object detection, to measure the task-level transferability of various models/methods over diverse real-world datasets, in terms of both prediction accuracy and adaption efficiency.

Advances in Image Manipulation

Organizers: Radu Timofte, Marcos V. Conde, Andrey Ignatov, Furkan Kınlı, Ren Yang

Location: Virtual Room 9

Time: Full Day (Time TBA)

Summary: Image manipulation is a key computer vision task, aiming at the restoration of degraded image content, the filling in of missing information, or the needed transformation and/or manipulation to achieve a desired target (with respect to perceptual quality, contents, or performance of apps working on such images). Recent years have witnessed an increased interest from the vision and graphics communities in these fundamental topics of research. Not only has there been a constantly growing flow of related papers, but also substantial progress has been achieved.

Each step forward eases the use of images by people or computers for the fulfillment of further tasks, as image manipulation serves as an important frontend. Not surprisingly then, there is an ever-growing range of applications in fields such as surveillance, the automotive industry, electronics, remote sensing, or medical image analysis etc. The emergence and ubiquitous use of mobile and wearable devices offer another fertile ground for additional applications and faster methods.

This workshop aims to provide an overview of the new trends and advances in those areas. Moreover, it will offer an opportunity for academic and industrial attendees to interact and explore collaborations.

Notes:

Self-Supervised Learning for Next-Generation Industry-Level Autonomous Driving

Organizers: Xiaodan Liang, Wei Zhang, Hang Xu, Michael C. Kampffmeyer, Fisher Yu, Ping Luo

Location: Grand Ballroom A

Time: Half Day - Morning (0900-1400)

Summary: Self-supervised Learning for Next-generation Industry-level Autonomous Driving refers to a variety of studies that attempt to refresh the solutions for challenging real-world perception tasks by learning from unlabeled or semi-supervised large-scale collected data to incrementally self-train powerful recognition models. Thanks to the rise of large-scale annotated data sets and advances in computing hardware, various supervised learning methods have significantly improved the performance in many problems (e.g. 2D detection, instance segmentation and 3D Lidar Detection) in the field of self-driving. However, these supervised learning approaches are notorious "data hungry", especially in the current autonomous driving fields. To facilitate an industry-level autonomous driving system in the future, the desired visual recognition model should be equipped with the ability of self-exploring, self-training and self-adapting across diverse new-appearing geographies, streets, cities, weather conditions, object labels, viewpoints or abnormal scenarios. To address this problem, many recent efforts in self-supervised learning, large-scale pretraining, weakly supervised learning and incremental/continual learning have been made to improve the perception systems to deviate from traditional paths of supervised learning for self-driving solutions. This workshop will investigate advanced ways of building next-generation industry level autonomous driving systems by resorting to self-supervised/semi-supervised learning.

Compositional and Multimodal Perception

Organizers: Kazuki Kozuka, Ranjay Krishna, Zelun Luo, Juan Carlos Niebles, Ehsan Adeli, Li Fei-Fei

Location: Royal Ballroom J

Time: Half Day - Morning (0900-1200)

Summary: The International Challenge on Compositional and Multimodal Perception (CAMP) of this ECCV2022 workshop aims at gathering researchers who work on activity/scene recgnition, compositionality, multimodal perception and its applications.

People understand the world by breaking down into parts. Events are perceived as a series of actions, objects are composed of multiple parts, and this sentence can be decomposed into a sequence of words. Although our knowledge representation is naturally compositional, most approaches to computer vision tasks generate representations that are not compositional.

We also understand that people use a variety of sensing modalities. Vision is an essential modality, but it can be noisy and requires a direct line of sight to perceive objects. Other sensors (e.g., audio, smell) can combat these shortcomings. They may allow us to detect otherwise imperceptible information about a scene. Prior workshops focused on multimodal learning have focused primarily on audio, video, and text as sensor modalities, but we found that these sensor modalities may not be inclusive enough. Both these points present interesting components that can add structure to the task of activity/scene recognition yet appear to be underexplored. To help encourage further exploration in these areas, we believe a challenge with each of these aspects is appropriate.

Computational Aspects of Deep Learning

Organizers: Iuri Frosio, Claudio Baecchi, Sophia Shao, Frederic Pariente, Lorenzo Baraldi, Giuseppe Fiameni

Location: Grand Ballroom D

Time: Half Day - Morning (0900-1300)

Summary: Deep Learning has been one the most significant breakthroughs in computer science in the last ten years. It has achieved significant progress in terms of the effectiveness of prediction models in many research topics and fields of application. This paradigm shift has radically changed the way research is conducted. AI is becoming a computational science where gigantic models with billions of parameters are trained on large-scale supercomputer. While this transition is leading to better and more accurate results by accelerating scientific discovery and technology advance, the availability of such computational power and the ability to harness it is a key success factor. In this context, optimisation and careful design of neural architectures play an increasingly important role that directly affects the pace of research, the effectiveness of state-of-the-art models, their applicability at production scale and, last but not least, the reduction of energy consumed to train and evaluate models. Architectural choices and strategies to train models, in fact, have an exceptional impact on run-time and discovery time, thus ultimately affecting the speed of progress of many research areas. The need for effective and efficient solutions is important in most research areas and essential to help researchers even in those situations where the availability of computational resources is scarce or severely restricted. This workshop will present novel research works that focus on the development of deep neural network architectures in computationally challenging domains.

Uncertainty Quantification for Computer Vision

Organizers: Andrea Pilzer, Yingzhen Li, Martin Trapp, Neill Campbell, Arno Solin

Location: Grand Ballroom E

Time: Half Day - Morning (0900-1300)

Summary: Nowadays, machine learning and deep learning approaches continually demonstrate their viability to solve vision challenges with models deployed to solve practical tasks. While performance (in terms of accuracy) is good, these models are predominately used as black boxes, and it is difficult to ascertain whether or not their outputs are reasonable. Even manual data set inspection, to discriminate between well predicted simple samples and errors on hard samples, may not be feasible. Uncertainty quantification and calibration are powerful tools that may be employed by engineers and researchers to better understand model output's reliability which is hugely beneficial to safe decision making. The machine learning community has placed great effort in developing novel techniques (e.g., Bayesian methods, post-hoc calibration and distribution-free approaches) and bench-marking them with classic research data sets. Our goal for this workshop is twofold. Firstly, we are interested in extending uncertainty quantification methods to more challenging computer vision data sets or practical use cases from an industrial perspective. Secondly, we wish to stimulate debate in the community about how to best integrate uncertainty in a community that often aims at 100% accuracy but does not always consider confidence. The workshop features three invited talks from well known experts in the field of uncertainty estimation, Prof. Yarin Gal, Prof. Sharon Yixuan Li, and Dr. Alex Kendall.

Recovering 6D Object Pose

Organizers: Martin Sundermeyer, Tomas Hodan, Yann Labbé, Gu Wang, Lingni Ma, Eric Brachmann, Bertram Drost, Sindi Shkodrani, Rigas Kouskouridas, Ales Leonardis, Carsten Steger, Vincent Lepetit, Jiří Matas

Location: Royal Ballroom J

Time: Half Day - Morning (0900-1300)

Summary: This workshop covers topics related to 6DoF object pose estimation, which is of major importance to many higher-level applications such as robotic manipulation and augmented/virtual reality. The workshop features four invited talks by experts in the field, discussion on open problems, and presentations of accepted workshop papers and of relevant papers invited from the main conference. In conjunction with the workshop, we organize the BOP Challenge 2022 to determine state-of-the-art object pose estimation methods.

Affective Behavior Analysis In-the-Wild

Organizers: Dimitrios Kollias, Stefanos Zafeiriou, Elnar Hajiyev, Viktoriia Sharmanska

Location: Virtual Room 3

Time: Half Day - Morning (0900-1300)

Summary: The ABAW Workshop has a unique aspect of fostering cross-pollination of different disciplines, bringing together experts and researchers of computer vision and pattern recognition, artificial intelligence, machine learning, HCI and multimedia. The diversity of human behavior, the richness of multi-modal data that arises from its analysis, and the multitude of applications that demand rapid progress in this area ensure that our event provides a timely and relevant discussion and dissemination platform.

The Workshop tackles the problem of affective behavior analysis in-the-wild, that is a major targeted characteristic of HCI systems used in real life applications. The target is to create machines and robots that are capable of understanding people's feelings, emotions and behaviors; thus, being able to interact in a `human-centered' and engaging manner with them, and effectively serving them as their digital assistants.

The Workshop also hosts a Competition (a continuation of the ones held at CVPR 2022 & 2017, ICCV 2021, IEEE FG 2020), which encompasses two Challenges: i) the Multi-Task Learning Challenge, which uses a static version of the Aff-Wild2 database, i.e., a large scale in-the-wild database and the first one to be annotated in terms of valence-arousal, basic expression & action units; ii) the Learning from Synthetic Data Challenge which uses synthetic images generated from the Aff-Wild2 database. Many novel, creative and interesting approaches -with significant results- have been developed by participating teams; they will be presented and discussed in the Workshop.

Large-Scale Point Clouds Analysis for Urban Scenes Understanding

Organizers: Qingyong Hu, Meida Chen, Ta-Ying Cheng, Bo Yang, Sheikh Khalid, Yulan Guo, Ronald Clark, Ales Leonardis, Niki Trigoni, Andrew Markham

Location: Virtual Room 1

Time: Half Day - Morning (0900-1300)

Summary: The 2nd Challenge on Large Scale Point-cloud Analysis for Urban Scenes Understanding (Urban3D) at ECCV 2022 aims to establish new benchmarks for 3D semantic and instance segmentation on urban-scale point clouds. In particular, we prime the challenge with both SensatUrban and STPLS3D datasets. SensatUrban consists of large-scale subsections of multiple urban areas in the UK. With the high quality of per-point annotations and the diverse distribution of semantic categories. STPLS3D is composed of both real-world and synthetic environments which cover more than 17 km2 of the city landscape in the U.S. with up to 18 fine-grained semantic classes and 14 instance classes. These two datasets are complementary to each other and allow us to explore a number of key research problems and directions for 3D semantic and instance learning in this workshop. We aspire to highlight the challenges faced in 3D segmentation on extremely large and dense point clouds of urban environments, sparking innovation in applications such as smart cities, digital twins, autonomous vehicles, automated asset management of large national infrastructures, and intelligent construction sites. We hope that our datasets, and this workshop could inspire the community to explore the next level of 3D learning.

We will be hosting 2 invited speakers and holding 2 parallel challenges (i.e., semantic and instance segmentation) for the topic of point cloud segmentation.

People Analysis: From Face, Body and Fashion to 3D Virtual Avatars

Organizers: Alberto Del Bimbo, Mohamed Daoudi, Roberto Vezzani, Xavier Alameda-Pineda, Marcella Cornia, Guido Borghi, Claudio Ferrari, Federico Becattini, Andrea Pilzer, Zhiwen Chen, Xiangyu Zhu, Ye Pan, Xiaoming Liu

Location: Virtual Room 10

Time: Half Day - Morning (0900-1300)

Summary: In the workshop and challenge on people analysis we address human-centered data analysis. These data are extremely widespread and have been intensely investigated by researchers belonging to even very different fields, including Computer Vision, Machine learning, and Artificial Intelligence. These research efforts are motivated by the several highly-informative aspects of humans that can be investigated, ranging from corporal elements (e.g., bodies, faces, hands, anthropometric measurements) to emotions and outward appearance (e.g. human garments and accessories). The huge amount and the extreme variety of this kind of data make the analysis and the use of learning approaches extremely challenging. The workshop will present novel research in the areas of human understanding and announce the winners of the 3D human body and 3D face reconstruction challenge. It will also feature two invited talks by experts in the field.

Observing and Understanding Hands in Action

Organizers: Antonis Argyros, Anil Armagan, Guillermo Garcia-Hernando, Shreyas Hampali, Otmar Hilliges, Tae-Kyun Kim, Vincent Lepetit, Iason Oikonomidis, Linlin Yang, Angela Yao

Location: Grand Ballroom E

Time: Half Day – Afternoon (1400-1800)

Summary: The sixth edition of this ECCV2022 workshop aims at gathering researchers who work on 2D/3D hand detection, segmentation, pose estimation, and tracking problems and its applications. This edition will emphasize reduced ground truth labels and focus on topics such as semi-supervised or self-supervised learning for training hand pose estimation systems. Development of RGB-D sensors and camera miniaturization (wearable cameras, smart phones, ubiquitous computing) have opened the door to a whole new range of technologies and applications which require detecting hands and recognizing hand poses in a variety of scenarios, including AR/VR, assisted car driving, robot grasping, and health care. However, labelling accurate real-world hand poses is still non-trivial. Most existing hand pose methods fail to generalize well to the real-world scenarios, especially when considering hand-object or hand-hand interaction scenarios. As new multiview video benchmarks have been proposed for the hand-object or hand-hand interaction, our goal is to encourage semi-/self-supervised learning for hand poses to utilize spatial-temporal information and reduce reliance on annotations. We will also cover up a "breadth of application" including sign language recognition, desktop interaction, egocentric views, object manipulations, far range and over-the-shoulder driver footage.

ISIC Skin Image Analysis

Organizers: M. Emre Celebi, Catarina Barata, Allan Halpern, Philipp Tschandl, Marc Combalia, Yuan Liu

Location: Virtual Room 1

Time: Half Day – Afternoon (1400-1829)

Summary: Skin is the largest organ of the human body, and is the first area of a patient assessed by clinical staff. The skin delivers numerous insights into a patient's underlying health: for example, pale or blue skin suggests respiratory issues, unusually yellowish skin can signal hepatic issues, or certain rashes can be indicative of autoimmune issues. In addition, dermatological complaints are also among the most prevalent in primary care. Images of the skin are the most easily captured form of medical image in healthcare, and the domain shares qualities to standard computer vision datasets, serving as a natural bridge between standard computer vision tasks and medical applications. However, significant and unique challenges still exist in this domain. For example, there is remarkable visual similarity across disease conditions, and compared to other medical imaging domains, varying genetics, disease states, imaging equipment, and imaging conditions can significantly change the appearance of the skin, making localization and classification in this domain unsolved tasks. This workshop will serve as a venue to facilitate advancements and knowledge dissemination in the field of skin image analysis, raising awareness and interest for these socially valuable tasks. Invited speakers include major influencers in computer vision and skin imaging, and authors of accepted papers.

Visual Perception for Navigation in Human Environments: The JackRabbot Human Body Pose Dataset and Benchmark

Organizers: Hamid Rezatofighi, Edward Vendrow, Ian Reid, Silvio Savarese

Location: Grand Ballroom D

Time: Half Day - Afternoon (1400-1800)

Summary: Recently, computer vision and robotics communities have proposed several centralized benchmarks to evaluate and compare different machine visual perception solutions. With the rising popularity of LiDAR-based 3D sensory data systems, some benchmarks have begun to provide both 2D and 3D sensor data, and to define new scene understanding tasks on this geometric information. In this workshop, we target a unique visual domain tailored to the perceptual tasks related to navigation in human environments, both indoors and outdoors. In the first workshop in ICCV19, we presented the JackRabbot social navigation dataset, and several visual benchmarks associated to it including 2D and 3D person detection and tracking. In our second workshop in CVPR21, we released additional annotations for human social group formation and identification, as well as individual and social activity labels for the humans in the scene. In this workshop, we release additional annotations for our captured data including human body pose annotations. Using both the existing and recent annotations, we will provide several new standardized benchmarks for different new visual perception tasks, including human pose estimation, human pose tracking and human motion forecasting and the perception of individual people, their group formation and their social activities. These new annotations increase the scope of use of the dataset between ECCV audiences, especially those who are researching on different computer vision tasks for robot perception in dynamic, human-centered environment.

Learning To Generate 3D Shapes and Scenes

Organizers: Angel X. Chang, Akshay Gadi Patil, Paul Guerrero, Daniel Ritchie, Manolis Savva, Kai Wang

Location: Virtual Room 3

Time: Half Day – Afternoon (1500-1900)

Summary: The past several years has seen an explosion of interest in generative modeling: models which learn to synthesize new elements from the training data domain. The representations learned by these models can also prove powerful when used as feature sets for other learning tasks. As the vision community turns from passive internet-images based vision toward more embodied vision tasks, such generative models become increasingly important for 3D data: as unsupervised feature learners, as training data synthesizers, as platforms to study 3D representations for vision tasks, and as a way of equipping embodied agents with a 3D `imagination' about the kinds of objects and scenes it might encounter. With this workshop, we aim to bring together researchers working on generative models of 3D shapes and scenes with researchers and practitioners who can use these generative models to improve vision tasks. This workshop focuses broadly on conditional (e.g., from sensory inputs, languages, other high-level specification, etc.) and unconditional generation of 3D shapes and scenes, and their application for vision, robotics, graphics and AI. Examples of such applications include scene classification and segmentation, 3D reconstruction, human activity recognition, robotic navigation, question answering, and more.

Sketching for Human Expressivity

Organizers: Qian Yu, Yulia Gryaditskaya, Yonggang Qi, Stella X. Yu, Giorgos Tolias, Mikhail Bessmeltsev, Xiaoguang Han

Location: Virtual Room TBA

Time: Half Day – Afternoon (1400-1820)

Summary: Sketches are created by humans through an iterative process and reflect one's sketching skills, taste, world perception, and even personality in just a set of sparse lines. Being the result of semantic, perceptual, or conceptual processing, sketches are distinctive from photos. While the CV and ML communities have firmly invested in reasoning with photos, sketch data just recently got into the spotlight. This shift of focus on using sketch data has already started to cause a profound impact on many facets of research on CV, CG, ML, HCI, and AI at large. Sketch has not only been used for image retrieval, 3D modeling, user interface design, but also as a key enabler in our fundamental understanding of visual abstraction, creativity, and expressivity. This series of workshop aims to bring together researchers of diverse background to consolidate cross-discipline insights, identify and encourage new directions, and ultimately foster the growth of the sketch research community.

Autonomous Vehicle Vision

Organizers: Rui Fan, Nemanja Djuric, Wenshuo Wang, Peter Ondruska, Jie Li

Location: Virtual Room 11

Time: Half Day – Afternoon (1400-1800)

Summary: The 3rd AVVision workshop aims to bring together industry professionals and academics to brainstorm and exchange ideas on the advancement of computer vision techniques for autonomous driving. In this half-day workshop, we will have four keynote talks and several paper presentations to discuss the state-of-the-art approaches and existing challenges in the field of autonomous driving.

Cross-Modal Human-Robot Interaction

Organizers: Fengda Zhu, Yi Zhu, Xiaodan Liang, Liwei Wang, Xiaojun Chang, Nicu Sebe

Location: Virtual Room 5

Time: Half Day - Afternoon (1400-1800)

Summary: A long-term goal of AI research is to build intelligent agents that can see the rich visual environment around us, interact with humans in multiple modalities, and act in a physical or embodied environment. As one of the most promising directions, cross-modal human-robot interaction has increasingly attracted attention from both academic and industry fields. The community has developed numerous methods to address the problems in cross-modal human-robot interaction. Visual recognition methods like detection and segmentation enable the robot to understand the semantics in an environment. Large-scale pretraining methods and cross-modal representation learning aim at effective cross-modal alignment. Reinforcement learning methods are applied to learn human-robotic interaction policy. Moreover, the community requires the agent to have other abilities such as life-long/incremental learning or active learning, which broadens the application of real-world human-robot interaction.

Many research works have been devoted to related topics, leading to rapid growth of related publications in the top-tier conferences and journals. We believe this workshop will be a very successful one and it will indeed benefit the progress of human-robot interaction significantly.

Language for 3D Scenes

Organizers: Angel X. Chang, Panos Achlioptas, Zhenyu Chen, Ahmed Abdelreheem, Matthias Niessner, Leonidas Guibas

Location: Virtual Room 7

Time: Half Day - Afternoon (1400-1800)

Summary: This is the second workshop on natural language and 3D-oriented object understanding of real-world scenes. Our primary goal is to spark research interest in this emerging area, and we set two objectives to achieve this. Our first objective is to bring together researchers interested in natural language and object representations of the physical world. This way, we hope to foster a multidisciplinary and broad discussion on how humans use language to communicate about different aspects of objects present in their surrounding 3D environments. The second objective is to benchmark progress in connecting language to 3D to identify and localize 3D objects with natural language. Tapping on the recently introduced large-scale datasets of ScanRefer and ReferIt3D, we host two benchmark challenges on language-assisted 3D localization and identification tasks. The workshop consists of presentations by experts in the field and short talks regarding methods addressing the benchmark challenges designed to highlight the emerging open problems in this area.

Sensing, Understanding and Synthesizing Humans

Organizers: Ziwei Liu, Ziqi Huang, Yuanhan Zhang, Jingkang Yang, Jiawei Ren, Lingdong Kong, Zhongang Cai, Chen Change Loy

Location: Virtual Room 10

Time: Half Day - Afternoon (1500-1900)

Summary: Great progress has been achieved in human sensing, understanding and synthesis. We further identify three key issues of the future directions:

1. We should take a holistic view on the integral pipeline of human sensing/understanding/synthesis, and explore the frontier problems emerged from their intersections.
2. The lessons, practices and foresights from both academia and industry should be shared and discussed together.
3. These topics lay the foundation for human-centric AI and will play a greater role in the age of intelligent well-being.

We also hold three challenges along with the workshop this year:

1. Point Cloud Robustness (PointCloud-C) Challenge 2022.
2. OmniBenchmark Challenge 2022.
3. Panoptic Scene Graph (PSG) Generation Challenge 2022.

We hope this workshop can foster the inter-disciplinary research in these fields that could profoundly advance our society.

Monday, October 24

NOTE: Use the QR code for each workshop's website to find the workshop's schedule. Here's the QR code to the ECCV Workshops page. Unless otherwise noted, all times are Israel Daylight Time (UTC+3)

0700–1900 Registration (David Intercontinental Hotel)

0700–1900 Coffee & Refreshments (Hotel Foyer)

1300–1400 Lunch (Foyer)

Drawings and Abstract Imagery: Representation and Analysis

Organizers: Diane Oyen, Kushal Kafle, Michal Kucer, Pradyumna Reddy, Cory Scott
Location: Royal Ballroom I
Time: Full Day (0900-1700)

Summary: The DIRA workshop aims to bring together researchers from both industry and academia who are working in diverse areas of research in representation, analysis and applications of abstract images, such as illustrations, drawings, technical diagrams, charts, and plots. The goal of the workshop is to explore and highlight technical challenges, insights and solutions that exist in representation, information extraction, semantic understanding, content-based retrieval, question-answering and various other topics pertaining to representation and analysis of abstract images, which has traditionally lagged behind progress in their natural image counterparts.

Real-World Surveillance: Applications & Challenges

Organizers: Kamal Nasrollahi, Sergio Escalera, Radu Tudor Ionescu, Fahad Shahbaz Khan, Thomas B. Moeslund, Anthony Hoogs, Shmuel Peleg, Mubarak Shah
Location: Nomi
Time: Full Day (Time TBA)

Summary: This is the second edition of RWS focusing on identifying and dealing with challenges of deploying machine learning models in real-world application. The workshop also included a challenge on thermal object detection on the largest thermal dataset that was annotated for this challenge.

What Is Motion For?

Organizers: Deqing Sun, Fatma Guney, Huaizu Jiang, Fitsum Reda, Laura Sevilla, Charles Herrmann, Pia Bideau, Jonas Wulff
Location: Royal Ballroom H
Time: Full Day (0900-1645)

Summary: This workshop will explore various ways of representing and extracting motion information, and provide a venue to exchange ideas about the use of motion in computer vision.

Safe Artificial Intelligence for Automated Driving

Organizers: Timo Saemann, Oliver Wasenmüller, Markus Enzweiler, Peter Schlicht, Joachim Sicking, Stefan Milz, Fabian Hüger, Seyed Ghobadi, Ruby Moritz, Oliver Grau, Frederik Blank, Thomas Stauner
Location: Meeting Room 4
Time: Full Day (0930-1800)

Summary: The realization of highly automated driving relies heavily on the safety of AI. Demonstrations of current systems that are showcased on appropriate portals can give the impression that AI has already achieved sufficient performance and is safe. However, this by no means represents statistically significant evidence that AI is safe. A changed environment in which the system is deployed quickly leads to significantly reduced performance of DNNs. The occurrence of natural or adversarial perturbations to the input data has fatal consequences for the safety of DNNs. In addition, DNNs have an insufficient explainability of their behavior, which drastically complicates the detection of mispredictions as well as the proof that AI is safe. The workshop addresses all topics related to the safety of AI in the context of highly automated driving.

Ego4D: First-Person Multi-Modal Video Understanding

Organizers: Rohit Girdhar, Andrew Westbury, Michael Wray, Antonino Furnari, Siddhant Bansal, Kristen Grauman, Jitendra Malik, Dima Damen, Giovanni Maria Farinella, James Rehg, David Crandall, Hyun Soo Park, Mike Zheng Shou, C.V. Jawahar, Kris Kitani, Bernard Ghanem, Jianbo Shi, Yoichi Sato, Pablo Arbelaez, Aude Oliva, Antonio Torralba
Location: Royal Ballroom J
Time: Full Day (0900-1815)

Summary: Worn cameras, smart glasses and headsets are becoming increasingly important as research test cases and off-the-shelf products. They capture the wearer's interactions with the world through image/video as well as gaze, audio, geolocation, and IMU data. Combined with head-mounted displays, they provide new forms of interaction and visualization, such as augmented reality. Egocentric vision is also a promising avenue for further work at the boundary of robotics and perception, where robotic agents learn how to act in human-centric environments by watching people.

Earlier this year, we introduced the massive Ego4D dataset as a first step towards catalyzing progress in this field. It contains more than 3,000 hours of around-the-world egocentric videos and 17 benchmark tasks, including episodic memory understanding, audio-visual reasoning, object interaction recognition, and forecasting. In the 2nd Ego4D workshop, building upon our success in the first EPIC + Ego4D workshop at CVPR, we will share the results of the 15-track Ego4D challenges. We have invited eminent speakers from a variety of research areas, including Robotics, Computer Vision and Multimodal understanding for keynote talks, along with papers from the conference relevant to the field of egocentric understanding. Finally, Meta's Project Aria technical team will present how academic partners can capture and consume egocentric data with Aria hardware and services.

BioImage Computing

Organizers: Jan Funke, Alexander Krull, Dagmar Kainmueller, Florian Jug, Anna Kreshuk, Martin Weigert, Virginie Uhlmann, Peter Bajcsy, Erik Meijering

Location: Meeting Room 5

Time: Full Day (0900-1825)

Summary: The seventh edition of the BioImage Computing workshop. This workshop will bring the latest challenges in bio-image computing to the computer vision community. It will showcase the specificities of bio-image computing and its current achievements, including issues related to image modeling, denoising, super-resolution, multi-scale instance- and semantic segmentation, motion estimation, image registration, tracking, classification, and event detection.

3D Perception for Autonomous Driving

Organizers: Raja Giryes, Yoni Kasten, Danfei Xu, Omri Harish, Eyal Gil-Ad

Location: Grand Ballroom E

Time: Full Day (0900-1815)

Summary: The 3DAD workshop will discuss the challenges and advantages in performing 3D perception for autonomous driving, and the recent trends in the field. Autonomous driving relies heavily on computer vision to guarantee safe driving. It involves solving many important tasks such as object detection, scene segmentation, motion prediction, and ego-motion calculation – all are important for safe planning in the autonomous driving task. While many academic works have focused on using 2D images to perform perception, it is widely agreed that adding other modalities, such as 3D LiDAR data, can improve scene understanding and safety.

Using 3D information for autonomous driving has its unique challenges. A LiDAR reacts differently than a camera to different weather conditions, and there are challenges relating to its data annotation. The data is not represented on a grid as is the case with 2D images, therefore, a dedicated effort is required for processing the 3D data. The workshop included a challenge and several talks on perception using 3D data for autonomous driving.

Neural Geometry and Rendering: Advances and the Common Objects in 3D Challenge

Organizers: David Novotny, Shangzhe Wu, Roman Shapovalov, Samarth Sinha, Natalia Neverova, Andrea Vedaldi, Jitendra Malik

Location: Grand Ballroom A

Time: Full Day (0900-1845)

Summary: The success of neural geometry and rendering has created enormous interest in 3D reconstruction and its potential in graphics and image understanding. The goal of this full-day workshop is to bring together the neural geometry and graphics communities to discuss recent advances and future challenges, such as reconstructing dynamic events from limited views. We will have invited talks from an incredible line-up of distinguished speakers in this field. Alongside the workshop, we are hosting a new CO3D Challenge, based on the Common Objects in 3D dataset (CO3Dv2).

Computer Vision for Metaverse

Organizers: Bichen Wu, Peizhao Zhang, Xiaoliang Dai, Tao Xu, Hang Zhang, Peter Vajda, Fernando de la Torre, Angela Dai, Bryan Catanzaro

Location: Grand Ballroom G

Time: Full Day (0900-1815)

Summary: Computer Vision (CV) research plays an essential role in enabling the future applications of Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR), which are nowadays referred to as the Metaverse. Building the Metaverse requires CV technologies to better understand people, objects, scenes, the world around us, and better render contents in more immersive and realistic ways. This brings new problems to CV research and inspires us to look at existing CV problems from new perspectives. As the general public grows interest and industry put more efforts in Metaverse, we think it is a good opportunity to organize a workshop for the computer vision community to get together to showcase our latest research, discuss new directions and problems, and influence the future trajectory of Metaverse research and applications.

Text in Everything

Organizers: Ron Litman, Aviad Aberdam, Shai Mazor, Hadar Averbuch-Elor, Dimosthenis Karatzas, R. Manmatha

Location: Gallery

Time: Full Day (0900-1730)

Summary: Understanding written communication through vision is a key aspect of human civilization and should also be an important capacity of intelligent agents aspiring to function in man-made environments. For example, interpreting written information in natural environments is essential in order to perform most everyday tasks like making a purchase, using public transportation, finding a place in the city, getting an appointment, or checking whether a store is open or not, to mention just a few. As such, the analysis of written communication in images and videos has recently gained an increased interest, as well as significant progress in a variety of text-based vision tasks. While in earlier years the main focus of this discipline was on OCR and the ability to read business documents, today this field contains various applications that require going beyond just text recognition, onto additionally reasoning over multiple modalities such as the structure and layout of documents. Recent advances in this field have been a result of a multi-disciplinary perspective spanning not only computer vision, but also natural language processing, document and layout understanding, knowledge representation and reasoning, data mining, information retrieval, and more. The goal of this workshop is to raise awareness about the aforementioned topics in the broader computer vision community, and gather vision, NLP and other researchers together to drive a new wave of progress by cross pollinating more ideas between text/documents and non-vision related fields.

Computer Vision for Civil and Infrastructure Engineering

Organizers: Joakim Bruslund Haurum, Ajmal Mian, Mingzhu Wang, Thomas B. Moeslund

Location: Meeting Room 3

Time: Full Day (0900-1715)

Summary: Civil and infrastructure engineering are corner stones in modern society, and as the world population continues to grow, the infrastructure and built environment has to keep up. This has led to a large interest in utilizing computer vision to assist and contribute with the inspection processes and contributing to the built environment, both during and after construction. There is huge potential for computer vision in many aspects of the civil and infrastructure domain which has yet to be realized, and this workshop aims at bringing practitioners and researchers from both domains together to realize this potential.

Distributed Smart Cameras

Organizers: Niki Martinel, Yue Gao, Ehsan Adeli, Christian Micheloni, Rita Pucci, Hamid Aghajan, Animashree Anandkumar, Li Fei-Fei, Caifeng Shan

Location: Virtual Room 1

Time: Full Day (1000-1800)

Summary: This is the second edition of the International Workshop on Distributed Smart Cameras (IWDSC) that followed after the 13th editions of the Intl. Conf. on Distributed Smart Cameras (ICDSC), since 2007. Smart camera networks are of paramount importance for our intelligent cities where a huge number of interconnected devices are actively collaborating to improve and ease our everyday life. This is achieved through advanced image chips and intelligent computer vision systems. In this workshop we present and encourage a discussion on the latest technologies and developments of these two heavily intertwined fundamental players. This workshop brings together the different communities that are relevant to distributed smart cameras (DSCs) to facilitate the interaction between researchers from different areas by discussing ongoing and recent ideas, demos, and applications in support of human performance through DSCs.

Causality in Vision

Organizers: Yulei Niu, Qianru Sun, Hanwang Zhang, Mike Zheng Shou, Peng Cui, Kaihua Tang, Song-Chun Zhu

Location: Virtual Room 3

Time: Full Day (0900-1700)

Summary: This is the second edition of Causality in Vision workshop. Causality is a new science of data generation, model training, and inference. Only by understanding the data causality, we can remove the spurious bias, disentangle the desired model effects, and modularize reusable features that generalize well. We deeply feel that it is a pressing demand for our CV community to adopt causality and use it as a new mind to re-think the hype of feeding big data into gigantic deep models. The goal of this workshop is to provide a comprehensive yet accessible overview of existing causality research and to help CV researchers to know why and how to apply causality in their own work. We aim to invite speakers from this area to present their latest works and propose new challenges.

LatinX in CV

Organizers: Matias Valdenegro Toro, Laura Montoya, Gilberto Ochoa-Ruiz, Carlos Hinojosa, Estefania Talavera, Fernando Wario, Maria Luisa Santiago, Fabian Caba, Francisco López-Tiro, Gildardo Sánchez, Eduardo Ulises Moya Sánchez, Rodolfo Valiente, Giancarlo Yuvini Pablo De Leon, Wayner Barrios, Lidia Talavera Martínez, Miguel Gonzalez, Daniel Flores-Araiza, Abraham Ramos

Location: Virtual Room TBA

Time: Full Day (0900-1730 EDT) (1600-0030 next day UTC+3)

Summary: The workshop is a one-day event with invited speakers, oral presentations, and posters. The event brings together faculty, graduate students, research scientists, and engineers for an opportunity to connect and exchange ideas. There will be a panel discussion and a mentoring session to discuss current research trends and career choices in computer vision. While all presenters will identify primarily as LatinX, all are invited to attend.

AI-Enabled Medical Image Analysis: Digital Pathology & Radiology/COVID19

Organizers: Jaime S. Cardoso, Isabel M. Pinto, Stefanos Kollias, Pedro C. Neto, Sara P. Oliveira, Xujiong Ye, Mattias Rantalainen, Luc Bidaut, Jeroen van der Laak, Francesco Rundo, Cameron Po-Hsuan Chen, Dimitrios Kollias, Diana Felizardo, Giuseppe Banna, Ana Monteiro

Location: Virtual Room 8

Time: Full Day (1000-1745)

Summary: Recently, Deep Learning has made rapid advances in the performance of medical image analysis challenging physicians in their traditional fields. In the pathology and radiology fields, in particular, automated procedures can help to reduce the workload of pathologists and radiologists and increase the accuracy and precision of medical image assessment, which is often considered subjective and not optimally reproducible. In addition, Deep Learning and Computer Vision demonstrate the ability/potential to extract more clinically relevant information from medical images than what is possible in current routine clinical practice by human assessors. Nevertheless, considerable development and validation work lie ahead before AI-based methods can be fully ready for integrated into medical departments.

The workshop on AI-enabled medical image analysis (AIMIA) at ECCV 2022 aims to foster discussion and presentation of ideas to tackle the challenges of whole slide image and CT/MRI/X-ray analysis/processing and identify research opportunities in the context of Digital Pathology and Radiology/COVID19.

High-quality original contributions should be targeted in several contexts such as, using self-supervised and unsupervised methods to enforce shared patterns emerging directly from data, developing strategies to leverage few (or partial) annotations, promoting interpretability in both model development and/or the results obtained, or ensuring generalizability to support medical staff in their analysis of data coming from multi-centres, multi-modalities or multi-diseases.

Visual Object-Oriented Learning Meets Interaction: Discovery, Representations, and Applications

Organizers: Kaichun Mo, Yanchao Yang, Jiayuan Gu, Shubham Tulsiani, Hongjing Lu, Leonidas Guibas

Location: Virtual Room 7

Time: Full Day (0900-1800)

Summary: Objects, as the most basic and composable units in visual data, exist in specific visual appearances and geometrical forms, carrying rich semantic, functional, dynamic, and relational information. One may discover objects by watching passive videos or actively interacting with the world to find them. Once detected, it is also an open research problem how to interact with the objects to extract and represent such object-oriented semantics. Furthermore, such representations need to be designed to be easily useful for various downstream perception and robotic interaction tasks. In this workshop, we will invite experts from related fields but different backgrounds, e.g., vision, robotics, and cognitive psychology, and foster the community to discuss how to define, discover and represent objects in visual data from/for interaction for various downstream applications.

Sign Language Understanding

Organizers: Liliane Momeni, Gul Varol, Hannah Bull, Prajwal K R, Neil Fox, Ben Saunders, Necat Cihan Camgöz, Richard Bowden, Andrew Zisserman, Bencie Woll, Sergio Escalera, Jose L. Alba-Castro, Thomas B. Moeslund, Julio C. S. Jacques Junior, Manuel Vazquez-Enriquez

Location: Virtual Room 9

Time: Full Day (0900-1800)

Summary: Sign languages are spatial-temporal languages and constitute a key form of communication for Deaf communities. Recent progress in fine-grained gesture and action classification, machine translation and image captioning, point to the possibility of automatic sign language understanding becoming a reality. The Sign Language Understanding Workshop is a full-day event which brings together two workshops: firstly, Open Challenges in Continuous Sign Language Recognition (morning session), and secondly, Sign Language Recognition, Translation & Production (afternoon session - virtual).

The Open Challenges in Continuous Sign Language Recognition workshop will focus on advances and new challenges on the topic of SLR. To advance and motivate the research on the field, this workshop will have an associated challenge on fine-grain sign spotting for continuous SLR. The aim of this workshop is to put a spotlight on the strengths and limitations of the existing approaches, and define the future directions of the field.

The Sign Language Recognition, Translation & Production workshop will bring together computer vision researchers, sign language linguists and members of the Deaf community. The workshop will consist of invited talks and also a challenge with three tracks: individual sign recognition; English sentence to sign sequence alignment; and sign spotting. The focus of this workshop is to engage with members of the Deaf community, broaden participation in sign language research, and cultivate collaborations.

A Challenge for Out-of-Distribution Generalization in Computer Vision

Organizers: Adam Kortylewski, Bingchen Zhao, Jiahao Wang, Shaozuo Yu, Siwei Yang, Dan Hendrycks, Oliver Zendel, Dawn Song, Alan Yuille

Location: Virtual Room 6

Time: Full Day (0900-1715)

Summary: Deep learning sparked a tremendous increase in the performance of computer vision systems over the past decade, under the implicit assumption that the training and test data are drawn independently and identically distributed (IID) from the same distribution. However, Deep Neural Networks (DNNs) are still far from reaching human-level performance at visual recognition tasks in real-world environments. The most important limitation of DNNs is that they fail to give reliable predictions in unseen or adverse viewing conditions, which would not fool a human observer, such as when objects have an unusual pose, texture, shape, or when objects occur in an unusual context or in challenging weather conditions. The lack of robustness of DNNs in such out-of-distribution (OOD) scenarios is generally acknowledged but largely remains unsolved, however, it needs to be overcome to make computer vision a reliable component of AI.

Notes:

Instance-Level Recognition

Organizers: Andre Araujo, Bingyi Cao, Shih-Fu Chang, Ondrej Chum, Noa Garcia, Bohyung Han, Guangxing Han, Pradeep Natarajan, Torsten Sattler, Giorgos Tolias, Tobias Weyand, Xu Zhang, Sanqiang Zhao

Location: Grand Ballroom B

Time: Half Day - Morning (0900-1300)

Summary: Visual instance-level recognition and retrieval are fundamental tasks in computer vision. Despite the recent advances in this field, many techniques have been evaluated on a limited number of domains, with a small number of classes. We believe that the research community can benefit from a new suite of datasets and associated challenges, to improve the understanding about the limitations of current technology, and with an opportunity to introduce new techniques. This year, we propose the first Universal Image Embedding Challenge, where the goal is to develop image representations that work well across several domains combined. The Instance-Level Recognition (ILR) Workshop is a follow-up of four successful editions of our previous workshops — the first two having focused only on landmark recognition (CVPRW18, CVPRW19), and the latest two expanded to two extra domains (artworks and products) (ECCVW20, ICCVW21).

Visual Inductive Priors for Data-Efficient Deep Learning

Organizers: Jan C. van Gemert, Nergis Tömen, Ekin D. Cubuk, Robert-Jan Bruintjes, Attila Lengyel, Osman Semih Kayhan, Marcos Baptista Rios, Lorenzo Brigato

Location: Grand Ballroom D

Time: Half Day - Morning (0900-1300)

Summary: Save data by adding visual knowledge priors to Deep Learning! Data is fueling deep learning, yet it is costly to gather and to annotate. Training on massive datasets has a huge energy consumption adding to our carbon footprint. In addition, there are only a select few deep learning behemoths which have billions of data points and thousands of expensive deep learning hardware GPUs at their disposal. This workshop focuses on how to pre-wire deep networks with generic visual inductive innate knowledge structures, which allows to incorporate hard won existing generic knowledge. Visual inductive priors are data efficient: what is built-in no longer has to be learned, saving valuable training data.

Excellent recent research investigates data efficiency in deep networks by exploiting other data sources through unsupervised learning, re-using existing datasets, or synthesizing artificial training data. However, not enough attention is given on how to overcome the data dependency by adding prior knowledge to deep nets. As a consequence, all knowledge has to be (re-)learned implicitly from data, making deep networks hard to understand black boxes which are susceptible to dataset bias requiring huge datasets and compute resources. This workshop aims to remedy this gap by investigating how to flexibly pre-wire deep networks with generic visual innate knowledge structures, which allows to incorporate hard won existing knowledge from physics such as light reflection or geometry.

Multiple Object Tracking and Segmentation in Complex Environments

Organizers: Peize Sun, Yi Jiang, Jinkun Cao, Yifu Zhang, Song Bai, Xinchao Wang, Linjie Yang, Ning Xu, Yang Fu, Yuchen Fan, Weiyao Wang, Tarun Kalluri, Heng Wang, Du Tran, Xinggang Wang, Ping Luo, Kris Kitani, Philip H.S. Torr

Location: Virtual Room 5

Time: Half Day - Morning (0900-1300)

Summary: Multiple-object tracking and segmentation aims to localize and associate objects of interest along time, and serve as fundamental technologies in many practical applications, such as visual surveillance, public security, video analysis, and human-computer interaction. Computer vision systems nowadays have achieved great performance in simple scenes, but are not as robust as the human vision system, especially in complex environments. To advance current vision systems performance in complex environments, our workshop explores four settings of multi-object tracking and segmentation: (a) long video (b) occluded object (c) diverse motion (d) open-world. Our four challenges consist of: (a) 4th YouTubeVIS and Long Video Instance Segmentation Challenge (b) 2nd Occluded Video Instance Segmentation Challenge (c) 1st Multiple People Tracking in Group Dance Challenge (d) 2nd Open-World Video Object Detection and Segmentation Challenge.

DeeperAction: Detailed Video Action Understanding and Anomaly Recognition

Organizers: Limin Wang, Yali Wang, Jing Shao, Yu Qiao

Location: Virtual Room 4

Time: Half Day - Morning (0900-1300)

Summary: This workshop aims to advance the area of video understanding with a shift from traditional action recognition to deeper understanding tasks of action, with a focus on detailed understanding of human action and anomaly recognition from videos in the wild. Specifically, we benchmark five related tasks on detailed action understanding by introducing newly-annotated and high-quality datasets, and organize the video action understanding challenge on these benchmarks.

Notes:

Human Body, Hands, and Activities from Egocentric and Multi-View Cameras

Organizers: Siwei Zhang, Marc Pollefeys, Taein Kwon, Dibyadip Chatterjee, Francis Engelmann, Tomas Hodan, Qianli Ma, Jun Liu, Yan Zhang, Stan Sclaroff, Bugra Tekin, Shugao Ma, Federica Bogo, Fadime Sener, Alexander Ilic, Robert Wang, Siyu Tang, Angela Yao

Location: Grand Ballroom D

Time: Half Day – Afternoon (1400-1800)

Summary: Egocentric perception of humans is a key feature for rapidly developing AR/VR glasses, human-computer interactions, and other assistive robotics. Deploying these applications requires considering the human in a 3D world. Indeed, perceiving and understanding humans from a third-person view is a long-standing research topic. Human understanding from an egocentric (or first-person) perspective, however, is significantly less explored and encompasses its own unique challenges, such as analysing and predicting the motion, body/hand reconstruction and action categories with other humans and the environment, all from strongly truncated and/or motion-blurred first-person views. The first edition of this workshop aims to gather researchers working on egocentric body pose, hand pose and 3D activity recognition and associated applications at ECCV 2022.

Assistive Computer Vision and Robotics

Organizers: Marco Leo, Mohan Trivedi, Giovanni Maria Farinella, Gerard Medioni, Antonino Furnari

Location: Virtual Room 4

Time: Half Day – Afternoon (1400-1800)

Summary: With the pervasive successes of Computer Vision and Robotics and the advent of industry 4.0, it has become paramount to design systems that can truly assist humans and augment their abilities to tackle both physical and intellectual tasks. We broadly refer to such systems as "assistive technologies". Examples of these technologies include approaches to assist visually impaired people to navigate and perceive the world, wearable devices which make use of artificial intelligence, mixed and augmented reality to improve perception and bring computation directly to the user, and systems designed to aid industrial processes and improve the safety of workers. These technologies need to consider an operational paradigm in which the user is central and can both influence and be influenced by the system. Despite some examples of this approach exist, implementing applications according to this "human-in-the-loop" scenario still requires a lot of effort to reach an adequate level of reliability and introduces challenging satellite issues related to usability, privacy, and acceptability. The main scope of ACVR 2022 is to bring together researchers from the diverse fields of engineering, computer science, social and biomedical sciences who work in the context of technologies involving Computer Vision and Robotics related to real-time continuous assistance and support of humans while performing any task.

Frontiers of Monocular 3D Perception: Implicit X Explicit

Organizers: Vitor Guizilini, Rareș A. Ambruș, Adrien Gaidon, Igor Vasiljevic, Greg Shakhnarovich, Sergey Zakharov, Matthew Walter

Location: Grand Ballroom C

Time: Half Day – Afternoon (1430-1830)

Summary: Recent advances in neural implicit representations introduced a new powerful way to think about scene understanding. As opposed to discrete explicit representations, learned implicit representations can effectively encode both geometry and appearance by representing them as continuous mappings over vector spaces. In this workshop we aim to discuss the benefits and shortcomings of both explicit and implicit approaches for the problem of monocular 3D understanding, as well as ways they could be merged, combining the benefits of each.

Visual Object Tracking Challenge

Organizers: Matej Kristan, Luka Čehovin Zajc, Aleš Leonardis, Alan Lukežič, Jiři Matas, Gustavo Fernández, Hyung Jin Chang, Michael Felsberg, Joni-Kristian Kämäräinen, Martin Danelljan, Roman Pflugfelder

Location: Grand Ballroom B

Time: Half Day – Afternoon (1400-1800)

Summary: The VOT challenges provide the tracking community with a precisely defined and repeatable way of comparing short-term trackers and long-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking. Following nine highly successful VOT challenges, the 10th Visual Object Tracking Challenge VOT2022 was held in spring of 2022 (challenge closed) hosting 7 subchallenges. This workshop includes results presentations, winning tracker talks, a keynote and contributed papers talks.

Vision With Biased or Scarce Data

Organizers: Kuan-Chuan Peng, Ziyan Wu

Location: Virtual Room 2

Time: Half Day – Afternoon (1400-1800)

Summary: With the increasing appetite for data in data-driven methods, the issues of biased and scarce data have become a major bottleneck in developing generalizable and scalable computer vision solutions, as well as effective deployment of these solutions in real-world scenarios. To tackle these challenges, researchers from both academia and industry must collaborate and make progress in fundamental research and applied technologies. The organizing committee and keynote speakers of VBSD 2022 consist of experts from both academia and industry with rich experiences in designing and developing robust computer vision algorithms and tranferring them to real-world solutions. VBSD 2022 provides a focused venue to discuss and disseminate research related to bias and scarcity topics in computer vision.

In-Vehicle Sensing and Monitorization

Organizers: Jaime Cardoso, Pedro Carvalho, João R. Pinto, Paula Viana, Christer Ahlström, Carolina Pinto

Location: Virtual Room 5

Time: Half Day – Afternoon (1500-1900)

Summary: Driver assistance and autonomous driving technologies have made significant progress over the past decade. Much of the research has been devoted to monitoring the external environment, while not nearly as much attention has been paid to the interior. Interior monitoring increases safety, comfort, and convenience for all vehicle occupants, especially in the case of autonomous shared vehicles. The In-vehicle Sensing and Monitorization workshop at ECCV 2022 targets the processing of data collected inside the vehicle for monitoring and event detection. It covers topics such as activity detection, emotional monitoring, identification of undesired behavior, damage detection, and many others related to the automatic supervision of the interior of shared vehicles and its occupants.

Notes:

Monday, October 24

NOTE: Use the QR code for each tutorials's website for more information on that tutorial. Here's the QR code to the CVPR Tutorials page. All times are Israel Daylight Time (UTC+3)

0700–1900 Registration (David Intercontinental Hotel)

0700–1900 Coffee & Refreshments (Hotel Foyer)

1300–1400 Lunch (Foyer)

Tutorial: Hyperbolic Representation Learning for Computer Vision

Organizers: Pascal Mettes, Jeffrey Gu, Mina Ghadimi Atigh, Serena Yeung, Martin Keller-Ressel

Location: Israel A

Time: Half Day - Morning (0900-1300)

Description: Learning in computer vision is all about deep networks and such networks operate on Euclidean manifolds by design. But is Euclidean geometry the best choice for deep learning or simply a practical option? Recent literature in machine learning and computer vision has shown that hyperbolic geometry provides a strong alternative, with an improved ability to embed hierarchies, graphs, text, images, and videos. In light of recent advancements in hyperbolic representation learning for computer vision, this tutorial seeks to advocate hyperbolic geometry and its strong potential for computer vision to a broader audience. The tutorial provides a theoretical and practical starting point for the field. At the conference, we will provide an easy-going introduction to hyperbolic geometry for non-mathematicians, where we focus on intuition and high-level understanding. We then outline the current state of hyperbolic geometry for vision from supervised and unsupervised perspectives. At the end, we dive into open research problems and future potential for hyperbolic geometry and visual understanding. Unique for this tutorial is that we do not stop at a theoretical foundation. The tutorial website will also host a series of notebook-style code snippets with foundational works on hyperbolic geometry, to get a better understanding of its workings and lower the barrier to start your dive into this exciting new research direction in computer vision.

Tutorial: Action Localization and Segmentation in Untrimmed Videos

Organizers: Angela Yao, Fadime Sener, Junsong Yuan, Guodong Ding

Location: Israel C

Time: Half Day - Morning (0900-1300)

Description: The majority of research in action understanding focuses on designing methods to encode a few seconds of short, trimmed clips and classify these with single action labels. Such methods, however, are rarely applicable for temporally localizing and/or classifying actions from longer, untrimmed streams of video. In this tutorial, we would like to focus on research on understanding actions in untrimmed, long videos up to tens of minutes. Compared to action recognition from trimmed video clips, untrimmed long video understanding tasks pose more challenges due to the long span of videos and complex temporal relations between occurring actions. Such challenges include: "What are the actions and when do these actions happen in the untrimmed long video sequences?" Our main focus for this tutorial is two tasks that aim to find human actions in videos, i.e., Temporal Action Localization/Detection (TAL/D) and Temporal Action Segmentation (TAS).

Tutorial: Implicit Rendering for Novel View Synthesis using Implicitron and PyTorch3D

Organizers: Jeremy Reizenstein, Roman Shapovalov, Nikhila Ravi, Patrick Labatut, David Novotny

Location: Israel D

Time: Half Day - Morning (0900-1300)

Description: The field of 3D computer vision is undergoing a paradigm shift. For the past few decades, it relied on explicit representations for reconstruction, such as point clouds obtained by Structure from Motion. Advances in deep learning enabled modeling 3D representation implicitly in the form of occupancy or radiance fields, or signed distance functions. Those methods are able to re-render images from new views with a level of quality unseen before. The research was mostly focused on one of two task formulations: single-scene and multiple-scene. The first formulation was popularized by the paper NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis published at the previous ECCV. These methods overfit the model to several dozen images of a single scene, where camera parameters are assumed known; it is well suited for high-fidelity rendering. Multiple-scene methods on the other hand, learn shape and appearance priors from a collection of videos of objects belonging to a semantic category. They allow for fast re-rendering from sparse (e.g., fewer than 10) views without fine-tuning a neural network on the new scene. Most recent research in both areas has been focused on improving quality of the rendering, speed and memory efficiency, and generalizing to non-rigid scenes. In this tutorial, we propose a quick survey of the subject, and introduce a new framework, Implicitron, directly available within the widely used open-source PyTorch3D library (~6k stars on GitHub).

Tutorial: New Frontiers in Efficient Neural Architecture Search

Organizers: Cho-Jui Hsieh, Ruochen Wang

Location: Virtual Room 13

Time: Half Day - Morning (0900-1300)

Description: Neural Architecture Search (NAS) has become increasingly important for many computer vision systems to automate the design of neural network architectures. However, due to the exponential size of the search space, many classical NAS algorithms require hundreds of GPU days which is not practical for standard users. Recently, significant progress has been made to improve the efficiency of NAS and make running NAS possible even for regular users on standard GPU machines. This tutorial aims to introduce and summarize recent progress in efficient NAS which enables the application of NAS to a diverse set of tasks.

Tutorial: Video Synthesis: Early Days and New Developments

Organizers: Sergey Tulyakov, Jian Ren, Stéphane Lathuilière, Aliaksandr Siarohin

Location: Negev & Galil

Time: Half Day - Morning (0900-1300)

Description: The introduction of generative adversarial networks in 2014 had a profound impact on video synthesis. Initial works generated videos with plain backgrounds and simple motions. Image synthesis advanced quite rapidly over the years. Multiple works in video synthesis capitalized on this success. Various subfields of video synthesis were introduced: prediction, animation, retargeting, manipulation, and stylization. Many of them led to a number of practical applications, democratizing video editing for non-experienced users and sparking start-ups. With the introduction of language-based models, image-based diffusion and large-scale datasets, video synthesis is seeing substantial improvement, with students, researchers and practitioners wanting to enter and contribute to the domain. Our tutorial will help them get the necessary knowledge, understand challenges and benchmarks, and choose a promising research direction. For practitioners, our tutorial will provide a detailed overview of the domain. We expect an attendee to have intermediate knowledge of CV & ML.

Tutorial: Self-Supervision on Wheels: Advances in Self-Supervised Learning from Autonomous Driving Data

Organizers: Spyros Gidaris, Andrei Bursuc, Patrick Pérez, Vitor Guizilini, Adrien Gaidon, Daniele de Martini, Paul Newman, Adam Harley, Katerina Fragkiadaki

Location: Virtual Room 11

Time: Half Day - Morning (0900-1300)

Description: The tremendous progress of deep-learning-based approaches to image understanding problems has inspired new advanced perception functionalities for autonomous systems. However, real-world vision applications often require models that can learn from large bulks of unlabeled and uncurated data with few labeled samples, usually costly to select and annotate. In contrast, typical supervised methods require extensive collections of carefully selected labeled data, a condition that is seldom fulfilled in practical applications. Self-supervised learning (SSL) arises as a promising line of research to mitigate this gap by training models using various supervision signals extracted from the data itself, without any human-generated labels. SSL has seen a lot of exciting progress in the last two years, with many new SSL methods managing to match or even surpass the performance of fully supervised techniques. While most popular SSL methods revolve around web image datasets, e.g., ImageNet, new diverse forms of self-supervision are investigated for autonomous driving (AD). AD represents a unique sandbox for SSL methods as it brings among the largest public data collections in the community and provides some of the most challenging computer vision tasks: object detection, depth estimation, image-based odometry and localization, etc. Here, the canonical SSL pipeline (i.e., self-supervised pre-training a model and fine-tuning it on a downstream task) is revisited and extended to learn tasks for which ground-truth annotations are difficult to compute (e.g., dense depth) leading to utterly new SSL approaches for computer vision and robotics. In this tutorial we will provide an in-depth coverage of the various paradigms for self-supervised learning (old and new) through the lens of essential perception tasks for AD.

Notes:

Tutorial: Neural Volumetric Rendering for Computer Vision

Organizers: Matt Tancik, Jon Barron, Ben Mildenhall, Angjoo Kanazawa, Pratul Srinivasan

Location: Israel D

Time: Half Day – Afternoon (1400-1800)

Description: Neural Radiance Fields (NeRFs), presented in ECCV 2020, demonstrated exciting potential for photo-realistic and immersive 3D scene reconstruction from a set of calibrated images. It was followed by a surge of works that explore the potential of using Neural Volumetric Rendering as a technique for enabling many exciting applications and addressing fundamental problems in Computer Vision, Graphics, Robotics and more. This tutorial approaches Neural Volumetric Rendering from the first principles, including its relation to the history of image based rendering and inverse graphics, its core components and their derivations, common practices, future challenges, and hands-on coding examples. The goal of this half-day tutorial is not to present a series of talks on recent papers in this area, but to provide pedagogical building blocks for novice and intermediate researchers to deeply understand the material by abstracting away the recent developments in the area of Neural Volumetric Rendering.

Tutorial: Self-Supervised Representation Learning in Computer Vision

Organizers: Xinlei Chen, Kaiming He, Christoph Feichtenhofer

Location: Israel A

Time: Half Day - Afternoon (1400-1900)

Description: This tutorial covers popular approaches and recent advancements in the field of self-supervised visual representation learning. We will cover topics such as Masked Autoencoders and Contrastive Learning. We will show how such frameworks are successfully learning from 2D static image and dynamic video information. Finally, we will also discuss self-supervised learning from a machine learning perspective. Overall, we will show connections and distinctions between different techniques for self-supervised learning, and provide insights about popular approaches in the community.

Tutorial: Localization and Mapping for Augmented Reality

Organizers: Paul-Edouard Sarlin, Viktor Larsson, Mihai Dusmanu, Ondrej Miksik, Johannes Schönberger, Marc Pollefeys

Location: Galil & Negev

Time: Half Day - Afternoon (1400-1800)

Description: This tutorial covers the task of large-scale localization and mapping for Augmented Reality (AR). Placing virtual content in the physical 3D world, persisting it over time, and sharing it with other users are typical scenarios for AR. In order to reliably overlay virtual content in the real world with pixel-level precision, these scenarios require AR devices to accurately determine their pose (3D position and orientation), at any point in time. While visual localization and mapping is one of the most studied problems in computer vision, its use for AR entails specific challenges and opportunities: devices capture temporal streams from multiple sensors besides cameras, they exhibit specific motion patterns, and they provide data crowdsourced from multiple users and device types, which can be mined for building large-scale maps.

Tutorial: Benchmarking Embodied AI Solutions in Natural Tasks

Organizers: Ruohan Zhang, Roberto Martín-Martín, Cem Gokmen, Chengshu Li, Sanjana Srivastava, Michael Lingelbach, Chen Wang, Josiah Wong, Jiajun Wu, Fei-Fei Li

Location: Virtual Room 11

Time: Half Day - Afternoon (1400-1800)

Description: Embodied artificial intelligence (EAI) has recently become a significant element in computer vision (CV). Researchers in CV are evaluating their algorithms by controlling and enabling intelligent behavior in autonomous agents (e.g., robots). This demonstrates that the agents are able to extract task-relevant information from images, understand their surroundings, and make decisions. An important enabler of this synergy between vision and EAI is the availability of simulators that allowed CV researchers to start tackling problems such as visual navigation and visual Q&A. We would like to extend the repertoire of EAI problems the vision community can study with our recently presented BEHAVIOR, Benchmark for Everyday Household Activities in Virtual, Interactive, and ecOlogical enviRonments). BEHAVIOR is a simulation benchmark to evaluate EAI agents with the physical challenges humans solve in their everyday life, i.e., 1000 household activities such as cooking food, picking up toys, setting the table, or cleaning houses. BEHAVIOR is simulator independent and has been implemented in several of them: iGibson, OmniGibson, and Habitat 2. This tutorial aims at providing a starting guide for researchers in computer vision, EAI, and general machine learning, interested in using BEHAVIOR in their own research, so that they know how to use BEHAVIOR in these simulators.

Tutorial: Outline and Shape Reconstruction in 2D

Organizers: Stefan Ohrhallinger, Jiju Peethambaran, Amal Dev Parakkat

Location: Virtual Room 12

Time: Half Day – Afternoon (1400-1800)

Description: Outline and shape reconstruction from unstructured points in a plane is a fundamental problem with many applications that has generated research interest for decades. Involved aspects like handling open, sharp, multiple and non-manifold outlines, run-time and provability as well as potential extension to 3D for surface reconstruction have led to many different algorithms. This multitude of reconstruction methods with quite different strengths and focus makes it a difficult task for users to choose a suitable algorithm for their specific problem. In this tutorial, we present proximity graphs, graph-based algorithms, algorithms with sampling guarantees, all in detail. Then, we show algorithms targeted at specific problem classes, such as reconstructing from noise, outliers, or sharp corners. Examples of the evaluation will show how its results can guide users to select an appropriate algorithm for their input data. As a special application, we show reconstruction of lines from sketches that can intersect themselves. Shape characterization of dot patterns will be shown as an additional field closely related to boundary reconstruction.

Tutorial: Semantic Segmentation of Point Clouds: A Deep Learning Framework for Cultural Heritage

Organizers: Marina Paolanti, Emanuele Frontoni, Francesca Matrone, Andrea Lingua, Eleonora Grilli, Yue Wang

Location: Virtual Room 14

Time: Half Day – Afternoon (1400-1800)

Description: The use of 3D point clouds for Cultural Heritage (CH) assets is becoming paramount, since they allow metric and morphological analyses impossible with the use of 2D data, better interpretation of phenomena, valorization and visualization, innovative management, development of conservation strategies, etc.In the last few decades, the increasing adoption of geomatics techniques for data collection and processing, contributed to the massive 3D-metric documentation of the built heritage. However, the complexity of CH assets makes the exploitation of 3D point clouds very impervious. Moreover, the provision of massive 3D geometric information with only color attributes could hamper their full exploitation due to the lack of semantic information. Therefore, the need of efficient, reliable and automated solutions for heritage point cloud classification is necessary in order to widespread the use of such data among heritage conservators, restorators, managers and HBIM (Historical Building Information Modeling) experts. Towards this end, the development of Deep Learning frameworks for point clouds classification is filling this gap, proving to be a very promising, but complex, field of research. These frameworks are designed to semantically enrich point clouds based on some specific classes, quite often case-dependent. These frameworks may facilitate the recognition of historic architectural elements at an appropriate level of detail, thus speeding up the process of reconstruction of geometries in the HBIM environment, or they can automatically identify degraded areas to speed up restoration processes. In this tutorial, we cover different Deep Learning methods for semantic segmentation tasks, providing the audience with a wide outlook upon the most recent Neural Networks.

Tutorial: Deep Energy-Based Learning in Computer Vision

Organizer: Jianwen Xie

Location: Virtual Room 13

Time: Half Day – Afternoon (1400-1800)

Description: There has been growing interest and advancement in deep energy-based learning. The deep energy-based model specifies an explicit probability density up to a normalization by using a modern bottom-up neural network to parameterize the energy function. The model can be trained by Langevin dynamics-based maximum likelihood estimation. It unifies the bottom-up representation and top-down generation into a single framework, which makes it different from the other generative models, such as generative adversarial net (GAN) and variational auto-encoder (VAE). This tutorial provides a quick introduction of current deep energy-based modeling and learning methodologies. It starts from the background of energy-based models from the perspective of computer vision, and then presents three categories of deep energy-based frameworks, including deep energy-based models in data space, energy-based cooperative learning frameworks, and energy-based models in latent space. This tutorial aims to enable researchers to learn about the current advance of deep energy-based learning and apply the knowledge to other domains.

Notes:

Gold Donors

Silver Donors

Exhibitors

Startup Exhibitors

Media Sponsors

Diamond Donors

APPLIED MATERIALS®
make possible

gm

Meta

mobileye™
An Intel Company

Platinum Donors

amazon | science

BOSCH

Google Research

HUAWEI