PROGRAMME
GUIDE

ECCV

EUROPEAN
CONFERENCE
ON COMPUTER
VISION

2024

WORKSHOPS &
TUTORIALS

Level **0**

SPACE
2

Level **+1**

CLOAKROOM

REGISTRATION
AREA

**MAIN
ENTRANCE**

CASH BAR

Level **+2** + **Mezzanine**

| SUITE 9 | SUITE 8 | | SUITE 7 | SUITE 6 | SUITE 5 |

SUITE 3 | SUITE 4

SUITE 2 | SUITE 1

WAY TO THE LEVEL
MEZZANINE

BROWN
3

BROWN
2

BROWN
1

AMBER
7+8

AMBER
6

AMBER
5

AMBER
4

AMBER
3

AMBER
2

AMBER
1

TOWER
LOUNGE

CASH BAR

Level **+3**

PANORAMA LOUNGE

ECCV
2024
EUROPEAN
CONFERENCE
ON COMPUTER
VISION
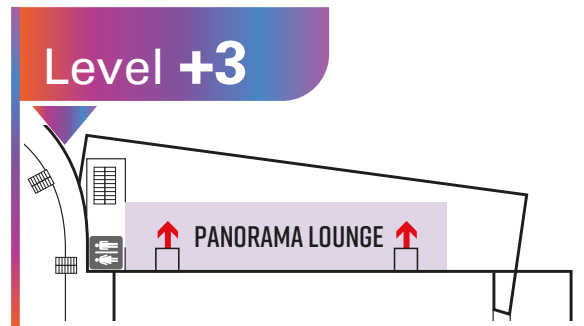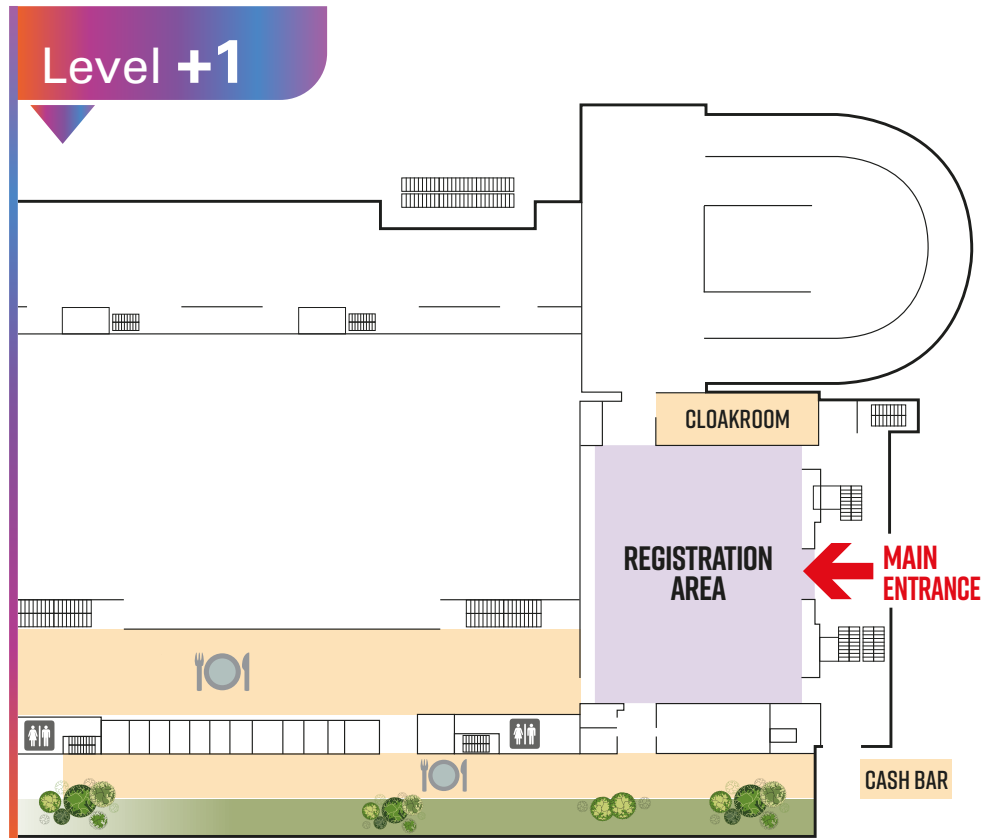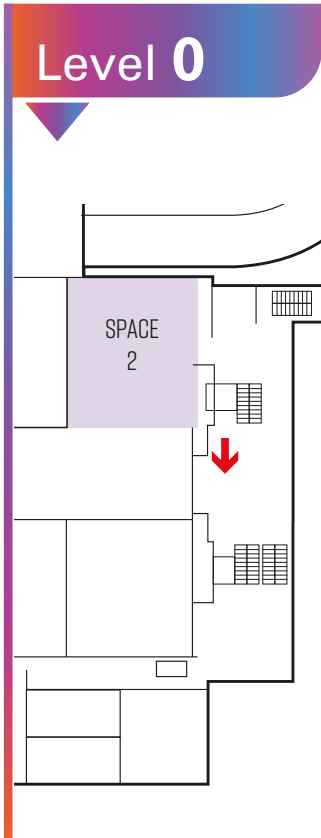
ECCV 2024 ORGANIZING COMMITTEE

General Chairs
Andrew Fitzgibbon (Graphcore)
Laura Leal-Taixé (NVIDIA)
Vittorio Murino (University of Verona
and University of Genova, Italy)

Program Chairs
Aleš Leonardis (University of Birmingham)
Elisa Ricci (University of Trento)
Gül Varol (Ecole des Ponts ParisTech)
Olga Russakovsky (Princeton University)
Stefan Roth (TU Darmstadt)
Torsten Sattler (Czech Technical
University in Prague)

Workshop & Tutorial Chairs
Alessio Del Bue (Istituto Italiano di Tecnologia)
Cristian Canton (Meta AI)
Jordi Pont-Tuset (Google DeepMind)
Tatiana Tommasi (Politecnico di Torino)

Demo Chairs
Hyung Jin Chang (University of Birmingham)
Marco Cristani (University of Verona)

Publication Chairs
Mahmoud Ali (Inria)
Francois Bremond (Inria)
Jovita Lukasik (University of Siegen)
Michael Moeller (University of Siegen)

Poster Chairs
Aljoša Ošep (Carnegie Mellon University)
Zuzana Kukelova (Czech Technical
University in Prague)

Diversity Chairs
David Fouhey (New York University)
Rita Cucchiara
(Università di Modena e Reggio Emilia)

Ethics Review Committee
Chloé Bakalar (Meta)
Kate Saenko (Boston University)
Remi Denton (Google Research)
Yisong Yue (Caltech, Asari AI & Latitude AI)

Conference Ombud
Georgia Gkioxari
(California Institute of Technology)
Greg Mori (Borealis AI / SFU)

Social Activities Chair
Giovanni Maria Farinella
(University of Catania, Italy)
Raffaella Lanzarotti
(Università degli Studi di Milano)
Simone Bianco
(University of Milano-Bicocca)

Doctoral Consortium Chairs
Cigdem Beyan (University of Verona, Italy)
Or Litany (NVIDIA / Technion)

Local Chairs
Raffaella Lanzarotti
(Università degli Studi di Milano)
Simone Bianco
(University of Milano-Bicocca)

Industry Liaison Chairs
Cees Snoek (University of Amsterdam)
Shaogang Gong
(Queen Mary University of London)

Publicity and social media chair
Konstantinos Derpanis (York University,
Samsung AI Centre Toronto)

Finance Chairs
Gerard Medioni (Amazon)
Nicole Finn (c to c events)

Web Developer
Lee Campbell (Eventhosts)

Tech Chair
Sascha Hornauer (Mines Paris – PSL)

Welcome to ECCV 2024!

It is our great pleasure to welcome you to the 18th European Conference on Computer Vision (ECCV 2024), which will takes place in the dynamic and historic city of Milan, Italy, from September 29th to October 4th, 2024.

As one of the leading global forums for computer vision, machine learning, artificial intelligence, and related fields, ECCV brings together a vibrant community of researchers and practitioners. This year, the program features an exceptional lineup, including keynotes from distinguished speakers, oral and poster sessions, workshops, tutorials, industry demonstrations, and exhibitions. These events offer a fantastic opportunity to engage with cutting-edge research and foster meaningful connections.

ECCV 2024 has attracted an unprecedented number of submissions, reflecting the ever-growing interest and advances in our field. With over 8,585 submissions, we are excited to announce that 2,387 papers have been accepted for publication, thanks to the tireless work of our Program Chairs, Area Chairs (ACs), and an incredible team of expert reviewers. Among these accepted papers, 200 have been selected for oral presentations, showcasing some of the most innovative and impactful research being conducted today.

Though the conference will take place primarily in person, we recognize the importance of inclusivity. For those unable to attend in Milan, we have designed a virtual component that will allow remote participants to access key content, including keynote talks and oral presentations, ensuring that the conference remains accessible to all. Additionally, in our commitment to inclusivity, we are offering travel grants to support attendees from low-income economies or those facing financial difficulties, helping ensure broader participation in ECCV 2024.

Milan, a city that perfectly blends tradition with modernity, is a fitting host for ECCV 2024. We encourage you to take time outside of the conference to explore its many cultural treasures—from the iconic Duomo and world-renowned museums to its vibrant culinary scenery.

The success of ECCV 2024 would not be possible without the extraordinary effort of so many individuals. We extend our deepest appreciation to the organizing committee, the reviewers, the authors, and our sponsors.

Finally, we offer our heartfelt thanks to you — the attendees. Your participation, whether in person or virtually, is what makes ECCV such a special event. We are confident that you will find the conference enriching, engaging, and inspiring.

We eagerly anticipate seeing you in Milan for what promises to be a memorable and impactful ECCV 2024!

General Chairs
Andrew Fitzgibbon (Graphcore)
Laura Leal-Taixé (NVIDIA)
Vittorio Murino
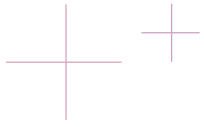(University of Verona and Genova, Italy)

Program Chairs
Aleš Leonardis (University of Birmingham)
Elisa Ricci (University of Trento)
Gül Varol (Ecole des Ponts ParisTech)
Olga Russakovsky (Princeton University)
Stefan Roth (TU Darmstadt)
Torsten Sattler (Czech Technical University in Prague)

Please visit the ECCV 2024 Virtual Conference Website:

08:00 – 18:00	Registration & Badge pickup
10:30 – 11:00	Coffee Break
13:00 – 14:00	Lunch
15:30 – 16:00	Coffee Break

TUTORIALS

Tutorial: **Large Multimodal Foundation Model Tutorial**
Organizers: Boyi Li, Sanjay Subramanian, Saining Xie, Trevor Darrell, Jitendra Malik
Website https://boyiliee.github.io/lmfm/
Room: Brown 3
Time: 09:00 - 13:00

Summary: The current discourse on technological progress underscores the interconnected roles of large multimodal foundation models. The necessity for this integration becomes evident when considering the complex dynamics of real-world environments. For instance, an autonomous vehicle in urban settings should not rely solely on visual sensors for pedestrian detection but must also proficiently interpret and respond to auditory signals, such as vocalized warnings. Similarly, the amalgamation of visual data with linguistic context in robots promises more adaptive functionalities, especially in diverse settings. Acknowledging the rapid expansion of this field, the tutorial agenda will encompass the introduction of history, applications, and future directions tailored for multimodal learning. Addressing privacy concerns related to multimodal data and equally vital safety discussions will ensure that systems adeptly interpret and act upon both visual and linguistic inputs, minimizing potential mishaps in real-world scenarios.

Through a comprehensive examination of these topics, this tutorial seeks to foster a deeper academic understanding of the intersection among vision, language as well as other modalities within the context of large multimodal foundation models. By convening experts from interdisciplinary fields, our objective is to decipher current state-of-the-art methodologies, address challenges, and chart avenues for future endeavors in large multimodal foundation model research, ensuring our findings resonate within both academic and industrial communities.

Tutorial: **Efficient Text-to-Image and Text-to-3D modeling**
Organizers: Shobhita Sundaram, Sadeep Jayasumana, Varun Jampani, Dilip Krishnan, Srikumar Ramalingam
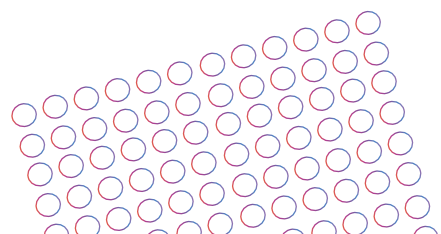**Website** https://efficient-genai.github.io/
Room: Amber 3
Time: 09:00 - 13:00

Summary: This tutorial will focus on text-to-image and text-to-3D algorithms and several efficient approaches. In particular, we will cover parallel decoding, fast image generation and efficient text-to-3D algorithms.

Tutorial: **Third Hands-on Egocentric Research Tutorial with Project Aria, from Meta**

Organizers: Richard Newcombe, Edward Miller, James Fort, Kristen Grauman
Website https://www.projectaria.com/events/eccv2024/
Room: Amber 3
Time: 14:00 - 18:00
Summary: In 2022, Meta Reality Labs Research hosted the first Project Aria Workshop at CVPR, introducing researchers to Project Aria, a research device from Reality Labs Research, worn like a regular pair of glasses, and used to accelerate research in the field of egocentric perception. Since then, the Project Aria academic program has grown to include over 100 university partners from around the world, providing researchers with devices, datasets, tools, and services, to accelerate always-on egocentric
In this third tutorial, we will introduce attendees to new features of the Aria Research Kit (ARK) and Open Science Initiatives (OSI), share novel research from academic partner program members, describe how researchers can gain access to the Project Aria academic program, and introduce new open datasets for accelerating machine perception research.

Tutorial: **Responsibly Building Generative Models**
Organizers: Gowthami Somepalli, Changhoon Kim, Tejas Gokhale, Kyle Min, Yezhou Yang, Tom Goldstein
Website https://asu-apg.github.io/rbgm/
Room: Suite 7
Time: 14:00 - 18:00
Summary: Over the past few years, generative models have evolved from simple rese arch concepts to production-ready tools, dramatically reshaping the tech landscape. Their outstanding generative capabilities have gained traction in various sectors, such as entertainment, art, journalism, and education. However, a closer look reveals that these models face several reliability and safety issues that can impact their widespread adoption. Key issues include the risk of memorizing training data, concerns about the models' accuracy in following prompts, and also the biases, misrepresentations, and hallucinations that a model may exhibit. Safety concerns also arise from issues like deepfake creation and the presence of nudity and racial biases in the models. Addres sing these concerns is more crucial than ever, given the advanced capabilities of generative models and the frequent coverage of copyright and safety issues in the media. In this half-day tutorial, we'll explore how to identify and address the failure modes of vision generative models (such as safety, fairness, memorization, etc), with an emphasis on mitigation strategies from the recent literature. We plan to include hands-on demo sessions, enabling participants to directly experience and understand some of the reliability concerns associated with these models.

WORKSHOPS

Topic: 3D Vision
Workshop: Recovering 6D Object Pose
Organizers: Tomas Hodan, Martin Sundermeyer, Yann Labbé, Van Nguyen Nguyen, Stephen Tyree, Andrew Guo, Mederic Fourmy, Lukas Ranftl, Jonathan Tremblay, Eric Brachmann, Bertram Drost, Sindi Shkodrani, Carsten Steger, Vincent Lepetit, Carsten Rother, Stan Birchfield, Jiri Matas
Website https://cmp.felk.cvut.cz/sixd/workshop_2024/
Room: Amber 7+8
Time: 09:00 - 13:00
Summary: This is the 9th edition of the R6D workshop that covers topics related to vision-based estimation of 6DoF object pose, which is an important task for robotic manipulation or augmented reality.
The workshop features four invited talks by experts in the field, presentation of the BOP Challenge 2024 (introducing model-free object pose estimation setup and new datasets),
and poster presentations of accepted workshop papers.

Topic: 3D Vision
Workshop: Workshop on Spatial AI
Organizers: Daniel Cremers, Xuyang Chen, Qing Cheng, Davide Scaramuzza, Xuqin Wang, Patrick Wenzel, Niclas Zeller
Website https://sites.google.com/view/spatial-ai-eccv24
Room: Brown 1
Time: 09:00 - 13:00
Summary: This workshop addresses different topics in the field of Spatial AI, where the fusion of computer vision and spatial understanding gives rise to intelligent systems capable of robust localization, enhanced perception, and nuanced scene interpretation.

Topic: 3D Vision
Workshop: 3D Vision and Modeling Challenges in eCommerce
Organizers: Kai Wang, Yiming Qian, Fenggen Yu, Loris Bazzani, Angel Chang, Chuhang Zou, Daniel Ritchie, Despoina Paschalidou, Francisca Gil-Ureta, Peter Gehler, Brian Jackson, Javier Romero, Jian Wang, Hao (Richard) Zhang, Xu Zhang
**Website** https://3dv-in-ecommerce.github.io/
Room: Brown 2
Time: 09:00 - 13:00
Summary: This workshop aims to bring together researchers working on 3D computer vision and graphics for eCommerce, with a focus on the three topics: (1) 3D shape/scene understanding and generation e.g. semantic segmentation, affordance and motion, multi-view reconstruction; (2) Digital human and fashion e.g. virtual try-ons and personalized fashion recommendation; (3) Foundation-model-assisted reasoning e.g. shape/scene synthesis from texts, language grounding in 3D models and diffusion-based 3D generative models.
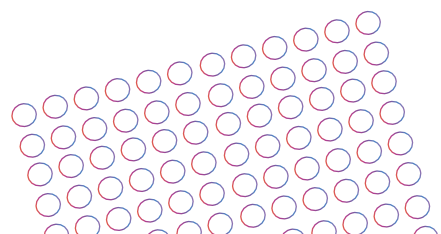
Topic:	**Applications**
Workshop:	**9th Workshop on Computer Vision in Plant Phenotyping and Agriculture (CVPPA)**
Organizers:	Valerio Giuffrida, Ribana Roscher, Feng Chen, Michael Pound, Ian Stavness, Monica Herrero-Huerta, Andrew French, Sotirios A Tsaftaris, Hanno Scharr, David Rousseau
Website	https://cvppa2024.github.io
Room:	Panorama Lounge
Time:	09:00 - 13:00
Summary:	This workshop will showcase the challenges raised by working on computer vision for plant phenotyping and agriculture. Workshop goals include demonstrating the state-of-the-art, identifying key unsolved problems, and introducing computer scientists with an interest in plant phenotyping to the field. Effective plant phenotyping is urgently needed to support the sustainability of our planet and its inhabitants: having strong community structures and computer vision scientists enter this field is more crucial now than ever.

Topic:	**Applications**
Workshop:	**3rd edition of Computer Vision for Metaverse (CV4Metaverse)**
Organizers:	Giuseppe Serra, Ali Abdari, Alex Falcon, Beatrice Portelli, Maria Pegia, Barbara Rossle, Bichen Wu, Peter Vajda, Richard Zhang
Website	https://sites.google.com/view/cv4metaverse-2024/
Room:	Amber 1
Time:	09:00 - 13:00
Summary:	In the ever-growing areas of Augmented Reality (AR), Virtual Reality (VR), and the expansive Metaverse, computer vision brings together the digital and physical worlds seamlessly. Its ability to understand and interpret visual information pushes thes e immersive technologies to new levels, enhancing user experiences, driving creative innovations, and exploring new frontiers. On the other side, Natural Language Processing (NLP) is pivotal for deciphering human language and facilitating applications like translation and summarization. Large Language Models (LLMs) are now capable of human-level conversational skills, drastically enhancing human-machine interactions. As exemplified by CLIP and other multimodal foundational models, textual information plays a significant role in understanding visual data. As a consequence, these large models may contribute significantly to improve AR, VR, and the Metaverse, enabling hands-free navigation, voice-based commands, and immersive communication between avatars. Therefore, the third edition of the CV4Metaverse workshop, acknowledging the substantial advancements language models have made in various domains, aims at integrating both language-based and pure computer vision techniques to contribute to the advancement of the field.

Topic:	**Art**
Workshop:	**AI for Visual Arts Workshop and Challenges (AI4VA)**
Organizers:	Deblina Bhattacharjee, Bahar Aydemir, Bingchen Zhao, Peter Gronquist, Tong Zhang, Mathieu Salzmann, Adam Kortylewski, Sabine Süsstrunk
Website	https://sites.google.com/view/ai4vaeccv2024
Room:	Amber 2

Time: 09:00 - 13:00

Summary: Creative fields deeply influence society, driving economic and cultural developments through art, design, museum exhibitions, photography, and films, centered on visual content's creation, use, and analysis. The rise of machine learning and computer vision has sparked significant research to advance these areas.

The AI4VA workshop and challenges aims to highlight cutting-edge trends and ideas at the intersection of computer vision and creativity. It will bring together artists, film makers, ethics advocates, and computer vision researchers, promoting vibrant idea exchanges. The workshop focuses on three main objectives:

1. Innovative Intersection of Arts and AI: This workshop showcases the latest in computer vision and creativity, featuring art manipulation competitions with a focus on art to revolutionise creative processes.

2. Comprehensive Coverage of Creative Technologies: Topics include generative models for photorealistic imagery, style transfer, understanding multilingual text in art, vision-language models for narrative understanding, neural networks for content enhancement and novel view synthesis, and robust dataset development across diverse artistic domains.

3. Diverse Participation and Accessibility: Designed for a broad audience including artists, designers, ethics advocates, and AI professionals, the workshop aims to draw creators and researchers at all levels to attend and will offer online streaming to maximise engagement and inclusivity.

More details at https://sites.google.com/view/ai4vaeccv2024.

Topic: Detection, Recognition, and Low-Level Vision
Workshop: Visual object tracking and segmentation challenge VOTS2024 workshop
Organizers: Matej Kristan, Jiři Matas, Pavel Tokmakov, Michael Felsberg, Luka Čehovin Zajc, Alan Lukežič, Gustavo Fernández, Hyung Jin Chang
Website   https://www.votchallenge.net/vots2024/
Room: Amber 4
Time: 09:00 - 13:00

Summary: The Visual Object Tracking and Segmentation challenge VOTS2024 workshop is the twelfth annual tracker benchmarking activity of the VOT initiative. The VOTS2024 challenge considers short-term, long-term, single-target and multiple-target tracking by segmentation as a single problem under the same evaluation protocol. The workshop will present results of two sub-challenges, considering holistic targets and targets undergoing topological transformations. In addition, the program features winning methods presentations, panel discussion and keynotes indicating future directions in video object segmentation, tracking of any points and broader tracking topics for scene understanding.

Topic: Human
Workshop: ACVR2024 - 12th International Workshop on Assistive Computer Vision and Robotics
Organizers: Giovanni Maria Farinella, Antonino Furnari, Marco Leo, Gerard G. Medioni, Francesco Ragusa, Mohan Trivedi
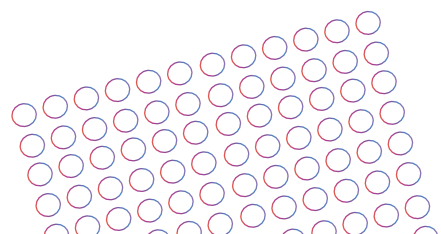**Website**   https://iplab.dmi.unict.it/acvr2024/
Room: Tower Lounge
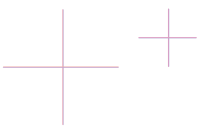**Time:**   09:00 - 13:00

Summary: Designing systems with humans in the loop able to assist the user is an active research area, with the potential for impact on society at large. Investigations in this area require a large set of innovations, tools, and evaluation criteria, even when compared to research on fully autonomous systems. Implementing such kinds of systems requires a lot of effort to reach an adequate level of reliability and introduces challenging satellite issues related to usability, privacy, and acceptability. Besides, multidisciplinary competencies are required to adapt algorithms to industrial, social, medical, and economic constraints. The goal is to provide a view of how recent findings in computer vision and robotics are changing assistive technologies, with emphasis on the related additional issues and how they have been addressed by the researchers working in the different research areas involved.

Topic: Human
Workshop: Workshop on Artificial Social Intelligence
Organizers: Leena Mathur, Evonne Ng, Fiona Ryan, Sangmin Lee, Paul Liang, Gül Varol, Shiry Ginosar, Hanbyul Joo, Justine Cassell, James Regh, Louis-Philippe Morency
Website https://sites.google.com/andrew.cmu.edu/asi-eccv-2024
Room: Suite 2
Time: 09:00 - 13:00
Summary: Humans have evolved social intelligence that enables us to interact with others through nonverbal signals (e.g, gestures) and verbal communication
(e.g., natural language), among other modalities.
Social intelligence plays a crucial role in helping humans navigate
daily interactions and has been a key factor driving the emergence of norms and culture in human society.
There has been a growing interest across computing communities
to build social intelligence competencies, such as social perception,
social memory, and social reasoning, in artificial intelligence systems.
These systems with artificial social intelligence could enable richer,
more seamless human machine interactions to support human health
and well-being in homes, hospitals, manufacturing, and other settings.
This workshop will be oriented around the following questions:
what are core technical challenges and future opportunities for cross-field collaboration to advance the science of social intelligence
and socially-intelligent AI?
Alongside benefits to society, what are ethical concerns, safeguards, and risks to building social intelligence in AI systems?
We welcome anyone interested in artificial social intelligence to discuss these questions and more.
Our ECCV 2024 workshop will feature keynote speakers
working on different facets of socially-intelligent AI systems,
from gesture generation systems to social robots.

Topic: Human
Workshop: The First Workshop on Expressive Encounters: Co-speech gestures across cultures in the wild
Organizers: Viktor Schmuck, Ariel Gjaci, Chaitanya Ahuja, Gustav Eje Henter, Rajmund Nagy, Youngwoo Yoon, Oya Celiktutan
Website https://expressive-encounters-workshop.github.io/2024/
Room: Suite 3
Time: 09:00 - 13:00

Summary: Virtual embodied agents and robots are increasingly integrated into our daily lives, serving roles such as receptionists in public services, home assistants, virtual personal trainers, and coaches for physical and mental health activities. To ensure user acceptance and trust, it is essential to design these agents to be not only functionally complex and useful but also understandable and socially appropriate.
People are more likely to accept such technologies when they perceive them as extensions of themselves. Therefore, human-human interaction, which incorporates not only language but also non-verbal communication that is richer in terms of social and cultural features, serves as a natural model for designing the behaviours of such agents.
Consequently, there has been a significant effort in recent years to generate non-verbal gestures for agents automatically in a data-driven manner.
This workshop aims to advance the development of real-world applications that involve virtual embodied agents and robots. To ensure user acceptance and trust, such real-world applications require new generative models and evaluation methods to generate socially appropriate non-verbal behaviours, considering cultural and personal factors, as well as enable real-time processing (i.e., understanding and responding to the user on the fly). The workshop aims to bring together researchers working on the abovementioned areas of non-verbal behaviour generation to discuss the future of this field. To kickstart these discussions, we invite all interested researchers to submit a paper for presentation at the workshop, to either of the tracks.

Topic: **Medical and Bio-Inspired Vision**
Workshop: **BioImage Computing**
Organizers: Alexander Krull, Martin Weigert, Qingjie Meng, Virginie Uhlmann, Peter Bajcsy, Jan Funke, Dagmar Kainmüller, Khaled Khairy
Website   https://www.bioimagecomputing.com/
Room: Amber 5
Time: 09:00 - 13:00
Summary: Bio-image computing (BIC) is a rapidly growing field at the interface of engineering, biology and computer science. Advanced light microscopy can deliver 2D and 3D image sequences of living cells with unprecedented image quality and ever increasing resolution in space and time.
The emergence of novel and diverse microscopy modalities has provided biologists with unprecedented means to explore cellular mechanisms, embryogenesis, and neural development, to mention only a few fundamental biological questions. Electron microscopy provides information on the cellular structure at nanometer resolution. Here, correlating light microscopy and electron microscopy at the subcellular level, and relating both to animal behavior at the macroscopic level, is of paramount importance.
The enormous size and complexity of these data sets, which can exceed multiple TB per volume or video, requires state-of-the-art computer vision methods.
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, event detection — important topics that appertain to the computer vision field.

Topic:	**Machine Learning**
Workshop:	**Self Supervised Learning: What is Next?**
Organizers:	Michael Dorkenwald, Mert Bulent Sariyildiz, Christian Rupprecht, Hilde Kuehne, Serge Belongie
Website	https://sslwin.org/
Room:	Space 2
Time:	09:00 - 13:00
Summary:	From GPT to DINO to diffusion models, the past years have seen major advances in self-supervised learning, with many new methods reaching astounding performances on standard benchmarks.

Still, the field of SSL is rapidly evolving with new learning paradigms coming up at an unprecedented speed.
At the same time, works on coupled data, such as image-text pairs, have shown large potential in producing even stronger models capable of zero-shot tasks and benefiting from the methodology developed in SSL.
Despite this progress, it is also apparent that there are still major unresolved challenges and it is not clear what the next step is going to be.
In this workshop, we want to highlight and provide a forum to discuss potential research directions, from radically new self-supervision tasks, data sources, and paradigms to surprising counter-intuitive results.
Through invited speakers and a poster session, 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.

Topic:	**Machine Learning**
Workshop:	**Beyond Euclidean: Hyperbolic and Hyperspherical Learning for Computer Vision**
Organizers:	Aiden Durrant, Fabio Galasso, Michael Kampffmeyer, Georgios Leontidis, Pascal Mettes, Leyla Mirvakhabova, Adín Ramírez Rivera, Indro Spinelli, Stella Yu
Website	https://sites.google.com/view/beyondeuclidean/home
Room:	Suite 8
Time:	09:00 - 13:00
Summary:	The Beyond Euclidean workshop explores the advantages of Hyperbolic and Hyperspherical learning for computer vision.

While Euclidean space is the traditional foundation for neural networks, learning in the corresponding geometrical space when the data exhibits non-Euclidean characteristics often leads to better representations.
Hyperbolic geometry, for example, is well-suited for modeling hierarchical data and has shown promise in tasks like few-shot recognition and out-of-distribution detection.
Similarly, hyperspherical learning is omnipresent in today's contrastive learning, powered by cosine similarities. It impacts tasks such as self-supervision to long-tailed classification and learning with limited samples.
This workshop focuses on these emerging approaches, providing a platform for researchers to discuss and develop new perspectives that challenge conventional deep-learning representations.

Topic: **Multimodal**
Workshop: **The Second Perception Test Challenge**
Organizers: Viorica Patraucean, Joe Heyward, Joao Carreira,
Dima Damen, Andrew Zisserman
Website https://ptchallenge-workshop.github.io/
Room: Suite 7
Time: 09:00 - 13:00
Summary: The Second Perception Test Challenge uses the Perception Test benchmark
to comprehensively evaluate the abilities of multimodal models across video,
audio, and text modalities, in four skill areas (Memory, Abstraction, Physics,
Semantics), four types of reasoning (descriptive, explanatory, predictive,
counterfactual), and six computational tasks (multiple-choice video QA,
grounded video QA, object tracking, point tracking, action localisation,
sound localisation).
In addition, this year we probe the long context understanding of these models
by introducing a novel benchmark containing hour-long videos
and multiple-choice QAs.
We will offer prizes to top competitors across tasks, with special awards for
models that complete multiple/all tasks and/or under zero-shot evaluation
regime.
The agenda features keynotes from experts in the field and presentations from
challenge winners.
https://github.com/google-deepmind/perception_test

Topic: **Responsible AI**
Workshop: **Fairness and ethics towards transparent AI:**
facing the chalLEnge through model Debiasing (FAILED)
Organizers: Vito Paolo Pastore, Enzo Tartaglione, Vittorio Murino
Website https://failed-workshop-eccv-2024.github.io
Room: Suite 6
Time: 09:00 - 13:00
Summary: In recent decades, Artificial Intelligence (AI) has experienced
an unprecedented surge, marked by the development of intricate machine
learning models boasting vast parameter counts.
Yet, alongside this growth looms a critical concern: the presence
of biases within these models, perpetuating disparities and ethical dilemmas.
The scientific community has increasingly turned its focus towards
understanding and addressing model bias, as evidenced by a significant uptick
in research across various disciplines.
This surge coincides with the impending European Union's AI Act, emphasizing
ethics, transparency, and accountability in AI systems.
In response to these pressing challenges, we present the workshop proposal,
"FAILED: Fairness and Ethics in AI: Facing the Challenge through
Model Debiasing". This initiative aims to convene experts and practitioners
from diverse backgrounds to explore innovative strategies for rectifying biases
and promoting fairness and transparency in AI systems, particularly in the
domain of Computer Vision. By fostering interdisciplinary discussions
and sharing cutting-edge solutions, FAILED seeks to propel advancements
in building unbiased and fair AI models. Join us in this collaborative endeavor
to shape a future where AI serves as a force for equity and inclusivity.
Together, let's pave the way towards a more ethical and fair AI landscape.
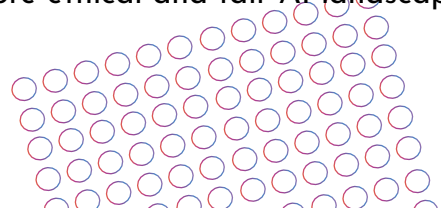
Topic: **Responsible AI**
Workshop: **2nd International Workshop on Privacy-Preserving Computer Vision**
Organizers: Vivek Sharma, Abhishek Singh, Ali Diba, Lingjuan Lyu, Lorenzo Servadei, Luc Van Gool, Ehsan Adeli, Michael Spranger, Dawn Song, Li Fei-Fei, Ramesh Raskar, Hiroaki Kitano
Website   https://privacy-preserving-computer-vision.github.io/eccv24.html
Room: Suite 4
Time: 09:00 - 13:00
Summary: In today's world, where data privacy is paramount and computer vision technologies are shaping our future, our workshop stands as a pivotal gathering of industry pioneers and academic visionaries. The convergence of these two spheres has never been more crucial, as it addresses pressing commercial and policy concerns. Our mission is clear: to foster a vibrant and collaborative community, a powerhouse of researchers committed to enhancing privacy in computer vision. Through shared insights, code, data, benchmarks, and training pipelines, we're not just driving innovation; we're redefining the very essence of privacy in the world of computer vision.
In this workshop, we come together to advance the cutting-edge of privacy-preserving machine learning for vision and imaging. Our workshop is a gateway to a brighter, more secure future, where technology and privacy coexist harmoniously.

Topic: **Responsible AI**
Workshop: **Critical Evaluation of Generative Models and their Impact on Society**
Organizers: Noa Garcia, Mayu Otani, Amelia Katirai, Kento Masui, Yankun Wu
Website   https://sites.google.com/view/cegis-workshop
Room: Suite 5
Time: 09:00 - 13:00
Summary: Visual generative models have revolutionized our ability to generate realistic images, videos, and other visual content. However, with great power comes great responsibility. While the computer vision community continues to innovate with models trained on vast datasets to improve visual quality, questions regarding the adequacy of evaluation protocols arise. Automatic measures such as CLIPScore and FID may not fully capture human perception, while human evaluation methods are costly and lack reproducibility. Alongside technical considerations, critical concerns have been raised by artists and social scientists regarding the ethical, legal, and social implications of visual generative technologies.
The democratization and accessibility of these technologies exacerbate issues such as privacy, copyright violations, and the perpetuation of social biases, necessitating urgent attention from our community.
This interdisciplinary workshop aims to convene experts from computer vision, machine learning, social sciences, digital humanities, and other relevant fields. By fostering collaboration and dialogue, we seek to address the complex challenges associated with visual generative models and their evaluation, benchmarking, and auditing.

Topic:	**Scene Understanding**
Workshop:	**Scalable 3D Scene Generation and 3D Geometric Scene Understanding**
Organizers:	Miaomiao Liu, Jose M. Alvarez, Mathieu Salzmann, Buyu Liu, Hongdong Li, Richard Hartley
Website	https://s3dsgr.github.io/
Room:	Amber 6
Time:	09:00 - 13:00
Summary:	Large-scale geometric scene understanding is one of the most critical and long-standing research directions in computer vision, with the impactful applications in autonomous driving, and robotics.

Both academia and industry have been investing heavily in pushing the research directions toward more efficiency and handling the large-scale scene. Different industries such as robotics, autonomous driving and gaming industry have distinct requirements on the quality and efficiency of the obtained 3D scene structures.
The proposed workshop will gather top researchers to discuss the future key challenges for this.

Topic:	**Sensing Devices**
Workshop:	**Eyes Of The Future: Integrating Computer Vision In Smart Eyewear (ICVSE)**
Organizers:	Francesca Palermo, Giulio Marano, Luca Merigo, Diana Trojaniello, Simone Mentasti, Matteo Matteucci
Website	https://sites.google.com/view/icvse-eccv2024
Room:	Suite 9
Time:	09:00 - 13:00
Summary:	As Smart Eyewear devices become increasingly prevalent, optimizing their functionality and user experience through sophisticated computer vision applications is crucial. These devices must not only effectively process real-time data but also operate under power and computational constraints while ensuring user privacy and ethical standards are upheld.

The "Eyes of the Future: Integrating Computer Vision in Smart Eyewear (ICVSE)" workshop, at ECCV 2024, aims to advance the field of Smart Eyewear by integrating cutting-edge computer vision technologies. This workshop addresses the need to bridge theoretical research and practical implementations in Smart Eyewear, a technology that will transform user interactions in everyday life through enhanced perception and augmented reality experiences.
The need for this workshop stems from the rapid advancements in both computer vision and wearable technology sectors, necessitating a dedicated forum where interdisciplinary insights and experiences can be shared to accelerate practical applications. Thus, ICVSE not only aims to showcase novel research but also to inspire a roadmap for future developments in Smart Eyewear technology.

Topic:	**3D Vision**
Workshop:	**Half-century of Structure-from-Motion (50SfM)**
Organizers:	Federica Arrigoni, Javier Civera, Andrea Fusiello, Luca Morelli, Francesco Nex, Rongjun Qin, Fabio Remondino
Website	https://50sfm.fbk.eu
Room:	Space 2
Time:	14:00 - 18:00
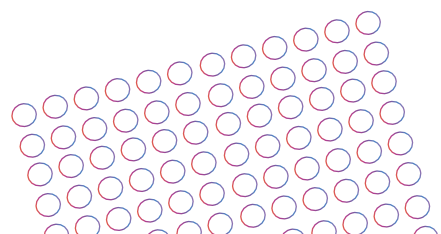
Summary: We invite you to join us in commemorating the 50th anniversary of Structure-from-Motion (SfM). As a cornerstone of computer vision, photogrammetry, and robotics, SfM has revolutionized our ability to reconstruct 3D structures from 2D images. This workshop aims to gather experts, researchers, and enthusiasts to reflect on SfM's evolution, discuss recent advancements, and explore its diverse applications. Our primary goal is to create a platform for knowledge exchange, collaboration, and the charting of future directions in this field. We will provide a comprehensive overview of SfM's historical development, explore recent breakthroughs in algorithms and methodologies, discuss challenges and solutions in current research, showcase diverse applications across industries, and foster networking and collaborations within the SfM community. We anticipate the workshop will increase awareness of SfM's historical context and evolution, enhance understanding of recent advancements in algorithms, identify challenges and potential solutions in current research, establish collaborations and networking opportunities within the SfM community, and inspire future research directions and applications of SfM.

Topic: 3D Vision
Workshop: Transparent & Reflective objects In the wild Challenges
Organizers: Jean-Baptiste Weibel, Alex Costanzino, Pierluigi Zama Ramirez, Fabio Tosi, Matteo Pogi, Luigi Di Stefano, Dominik Bauer, Doris Antensteiner, Markus Vincze

Website https://sites.google.com/view/eccv24-tricky-workshop/
Room: Suite 8
Time: 14:00 - 18:00

Summary: Exemplified by the results of the COCO, LVIS and BOP challenges, the performance of state-of-the-art methods in object detection, segmentation and pose estimation is rapidly progressing.
Their common (explicit or implicit) assumption that objects are Lambertian, i.e., only create a diffuse reflection of light, however, is an oversimplification of the actual visual world. For non-Lambertian objects, made of glass or metal, the specific scene arrangement creates variations in appearance that go beyond mere texture and occlusion changes. For example, objects are not only directly observable but also via reflection or refraction, depending on their relative location to transparent objects; whereas the appearance of specular highlights depends on light and camera location.
Depth sensing also assumes a "Lambertian world" and hence fails to correctly measure the distance to transparent objects. The performance of current approaches, independent of the input modality, therefore quickly deteriorates when faced with such tricky scenes.
The 2nd edition of the Transparent & Reflective objects In the wild Challenges (TRICKY) workshop will discuss object classification, detection, tracking, reconstruction, depth and pose estimation from imaging data for such tricky objects to highlight and identify the related challenges in these tasks and advance the state of the art. A major focus will be put on the applicability of methods in unconstrained scenarios, such as natural scene arrangement, mix of Lambertian and non-Lambertian objects, or changing illumination. This will be achieved with a depth estimation challenge as well 5 invited talks. The workshop will also include 6 spotlight talks and up to 12 posters of contributed works to encourage the discussion of novel research directions.

Topic:	**Applications**
Workshop:	**The First Workshop on: Computer Vision for Videogames (CV2)**
Organizers:	Iuri Frosio, Ekta Prashnani, Nicu Sebe, Rulon Raymond, Georgios N. Yannakakis, David Durst, Marguerite De Courcelle, Joohwan Kim
Website	https://sites.google.com/nvidia.com/cv2/
Room:	Amber 5
Time:	14:00 - 18:00

Summary: Our scope is to bring together people working in Computer Vision (CV) and, more broadly speaking, Artificial Intelligence (AI), to talk about the adoption of CV/AI methods for videogames, that represent a large capital market within creative industries and a crucial domain for AI research at the same time. Our workshop covers various aspects of videogames development and consumption, ranging from game creation, game servicing, player experience management, to bot creation, cheat detection, and human computer interaction mediated by large language models. Focusing on CV for videogames allows bringing together cohesively related works with foreseeable and practical impact on today's market, thus we have given priority to submissions specifically devoted to the application of state of the art CV/AI methods for videogames. We also favoured the presentation of novel datasets that can sparkle further research in this field.
The committee and keynotes includes multiple genders, researchers with origins from different geographical areas (USA, EU, Asia), from both industry (NVIDIA, Activision, Blockade Labs, Microsoft, Snap) and academia (Universities of Trento, Malta, Stanford), and different research experience (from PhD students to full professors and managers). We promote cross-disciplinary and diversity not only within the members of the organizing committee, but also in the list of topics covered by the workshop.
The workshop will help sharing and discussing different points of view on the future of CV in videogames in a friendly environment.

Topic:	**Art**
Workshop:	**AI4DH: Artificial Intelligence for Digital Humanities**
Organizers:	Roberto Pierdicca, Lorenzo Baraldi, Silvia Cascianelli, Marcella Cornia, Francesca Matrone, Guido Borghi, Marina Paolanti
Website	https://sites.google.com/view/ai4dh2024
Room:	Suite 6
Time:	14:00 - 18:00

Summary: Researchers have explored the benefits and applications of modern Artificial Intelligence (AI) algorithms in different scenarios. Digital Humanities (DH) is a newly-emerging field that brings together humanities, social, and computer scientists to work both on fundamental and applied research in humanities. The large-scale cultural and societal implications of these changes and the ethical questions that raise offer an important challenge as techniques in Artificial Intelligence and Data Learning have matured. Thus, there has been a wide range of computational tools, methods, and models that have enabled humanities to conduct research at a scale once thought impossible.
The goal of this workshop is to encourage and highlight novel strategies and original research in applying Artificial Intelligence techniques in digital humanities research such as data discovery, digital data creation, management, data analytics (including text mining, image mining and data visualization) in literature, linguistics, culture heritage, media, social science,

history, music and acoustics, and Artificial Intelligence for Digital Humanities in pedagogy and academic curricula.

This workshop aims not only to serve as a venue for presenting work in this area but also to build a community and share information in this new field.

Topic: Autonomous Driving and Robotics
Workshop: Third ROAD Workshop & Challenge: Event Detection for Situation Awareness in Autonomous Driving
Organizers: Fabio Cuzzolin, Salman Khan, Andrew Bradley, Reza J. Alitappeh, Eleonora Giunchiglia, Gurkirt Singh, Nadya Madjid, Bilal Hassan, Majid Khonji, Jorge Dias, Chi-Hsi Kung, Yi-Hsuan Tsai, and Yi-Ting Chen
Website https://sites.google.com/view/road-eccv2024/home
Room: Suite 5
Time: 14:00 - 18:00
Summary: The goal of this workshop is twofold: firstly, to bring to the forefront of research in autonomous driving on the topic of situation awareness. Secondly, to present ROAD++: a combination of four datasets from different domains including ROAD (UK), ROAD-Waymo (USA), ROAD-UAE (UAE), and TACO (CARLA simulation).

Topic: Autonomous Driving and Robotics
Workshop: Autonomous Vehicles meet Multimodal Foundation Models
Organizers: Yan Wang, Yurong You, Kashyap Chitta, Yue Wang, Yiyi Liao, Li Erran Li, Deva Ramanan, Kilian Q. Weinberger, Laura Leal-Taixe
Website https://mllmav.github.io/
Room: Brown 2
Time: 14:00 - 18:00
Summary: Building safe and intelligent Autonomous Vehicles capable of human-like reasoning is a challenging problem, pushing the limits of computer vision. Current Autonomous Vehicle systems struggle with diverse and unseen driving scenarios, necessitating a shift in research focus. Recently, multimodal large language models have shown great promise in understanding human intent and solving complex problems.

Such models not only showcase incredible capabilities in understanding human intent and solving complex and unstructured problems, but scale gracefully with data and compute.

This workshop explores leveraging multimodal large language models to tackle key challenges in Autonomous Vehicles.

Topic: Detection, Recognition, and Low-Level Vision
Workshop: 5th Advances in Image Manipulation (AIM) Workshop and Challenges
Organizers: Radu Timofte, Andrey Ignatov, Marcos V. Conde, Dmitriy Vatolin, Eduardo Pérez-Pellitero
Website https://www.cvlai.net/aim/2024/
Room: Amber 1
Time: 14:00 - 18:00
Summary: Image restoration, enhancement and manipulation are key computer vision tasks, 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 restoration, enhancement and 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 and to gauge the state-of-the-art through the associated challenges on: sparse neural rendering, UHD blind photo quality assessment, compressed depth map super-resolution and restoration, raw burst alignment, efficient video super-resolution for AV1 compressed content, video super-resolution quality assessment, compressed video quality assessment and video saliency prediction.
Moreover, it will offer an opportunity for academic and industrial attendees to interact and explore collaborations.

Topic: **Efficiency**
Workshop: **Efficient Deep Learning for Foundation Models**
Organizers: Hongxu (Danny) Yin, Sifei Liu, Ji Lin, Maying Shen, Jason Clemons, Xin Wang, Jose M. Alvarez, Pavlo Molchanov, Xueyan Zou, Xiaolong Wang, Song Han, Jan Kautz
**Website** https://sites.google.com/view/efm24/home
Room: Brown 3
Time: 14:00 - 18:00
Summary: This year marks a sharp rise of foundation models, spurring enormous practical impacts to the society with surprising network strength across a wide range of input modality. In a new realm of foundation model that relies on an enormous amount of training data and hardware (e.g., thousands of GPUs) to yield even one network, efficient deep learning becomes unprecedentedly critical to ensure a successful development/deployment of foundation models for both academia and industry.
With tens of billions of parameters, foundation neural networks by nature urge (i) enormous amount of clean and properly labeled training data, (ii) with highly efficient and scalable training pipeline, oriented around (iii) an efficient yet strong neural network architecture, with (iv) stringent practical requirements for ubiquitous usage and fast inference.
For cloud service such as GPT-backed applications, speeding up the inference directly translates into far reduced waiting time for billions of users worldwide. In this EFM workshop, we wish to offer to our audience the most recent insights across the entire development stack to lower the barrier for large foundation models. Our topics include but are not limited to the following: Novel neural architecture designs for foundation models across modalities and tasks; Resource-efficient training/fine-tuning from both data and algorithm perspectives; New methods to improve efficiency in distributed training and inference of large models; Large model compression techniques such as quantization, architecture search, pruning, adaptation, distillation; Hardware-software co-optimization of large foundation models; Efficient and effective bench-marking efforts for large foundation models.
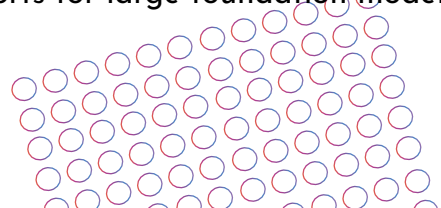
Topic: **Human**
Workshop: **T-CAP - Towards a Complete Analysis of People: Fine-grained Understanding for Real-World Applications**
Organizers: Mohamed Daoudi, Xavier Alameda-Pineda, Federico Becattini, Guido Borghi, Marcella Cornia, Claudio Ferrari, Tomaso Fontanini, Andrea Pilzer
Website   https://sites.google.com/view/t-cap-2024/home
Room: Amber 7 + 8
Time: 14:00 - 18:00
Summary: Deep learning has reached impressive achievements in recent years, especially in the field of human analysis. Indeed, human-centered data are extremely widespread and have been intensely investigated by researchers belonging to very different fields. Nevertheless, the gap between theoretical research and its application is still large due to several factors: inconsistencies between public datasets and real-world data, lack of flexibility of neural networks, and the demand of industries for almost perfect accuracy when measuring human data.
For this reason, in this workshop, our goal is to push the exploration of all the real-world applications that human analysis can make possible to bring together academic and industrial researchers. These research efforts are motivated by the several highly informative aspects of humans that are worth exploring: in our case, we focus on the fine-grained analysis of the human body from images, i.e. a precise and non-invasive understanding of the information related to humans, such as their measures, poses, detection, and so on.
The huge amount and the extreme variety of this kind of data make the analysis and the use of learning approaches extremely challenging. Furthermore, considering the crucial impact of human-centered technologies in many industrial application domains, the demand for accurate models able also to run on mobile and embedded solutions is constantly increasing.

Topic: **Medical and Bio-Inspired Vision**
Workshop: **Human-inspired Computer Vision**
Organizers: Lucia Schiatti, Mengmi Zhang, Yen-Ling Kuo, Vittorio Cuculo, Martin Schrimpf, Andrei Barbu
Website   https://sites.google.com/view/hcvworkshop2024
Room: Tower Lounge
Time: 14:00 - 18:00
Summary: The goal of the Human-inspired Computer Vision workshop is to link and disseminate parallel findings in the fields of neuroscience, psychology, cognitive science and computer vision, to inform the development of human-inspired computational models, capable of solving visual tasks in a human-like fashion. Although pure engineering approaches to computer vision can achieve high performance in standard vision datasets regardless of their resemblance to human biological mechanisms, they still fail in many visual cognitive tasks that are fundamental to human intelligence, such as zero-shot visual search, context aware object recognition, and continual learning. We believe that expanding the perspective from a merely performance-oriented approach to the development of models holding human-like properties, while still achieving high performance, can push forward the frontiers of computer vision research. Indeed, on one hand, insights from cognitive science and neuroscience can help us identifying and tackling gaps between humans and machines. Characterizing aspects where humans still surpass computer vision, such us robustness to biases, working memory tasks, and visual search, can ultimately improve models' performance as well. On the other hand, inspiration from biology can

lead to computer vision models that are useful to explain neuroscientific and cognitive observations, and to deepen our understanding of human brain and developmental mechanisms. Such a perspective can pave the way to applications of computer vision outside classic domains, for instance to augment human vision and cognition in presence of neurodevelopmental disorders.

Topic:	**Machine Learning**
Workshop:	**Traditional Computer Vision in the Age of Deep Learning (TradiCV)**
Organizers:	Federica Arrigoni, Adrien Bartoli, Andrea Fusiello, Vladislav Golyanik, Zuzana Kukelova, Luca Magri, Tomas Pajdla, Shaifali Parashar, Matteo Poggi, Fabio Tosi
Website	https://sites.google.com/view/tradicv
Room:	Amber 2
Time:	14:00 - 18:00
Summary:	Over the past decade, we have witnessed a revolutionary shift in computer vision, dominated by the rise of deep learning, which is asserting its dominance in top-tier conferences and journals. However, as we navigate this era of profound change, it is clear that certain problems and domains still favour traditional approaches over deep learning solutions. In particular, challenges that require a robust mathematical model, such as camera calibration and structure-from-motion, continue to rely on classical methods.

This workshop responds to the evolving landscape of computer vision by gathering contributions that specifically address these challenges using ``traditional" or ``classical" methods. These approaches, based on analytical and explicit models, stand in stark contrast to the dominant trend towards learned or neural solutions. We aim to shed light on scenarios where traditional methods outperform neural ones, emphasising factors such as generalisation across domains. We also look at cases where traditional methods, while perhaps falling short in terms of performance, have distinct advantages over their deep learning counterparts, such as reduced data collection effort, lower computational requirements, reduced power consumption, or improved model compactness.
We invite researchers, practitioners, and enthusiasts to contribute their insights and research findings to this workshop. Submissions should showcase the prowess of traditional computer vision approaches, emphasizing instances where they outperform or complement deep learning solutions. We encourage contributions inspiring critical discussions about the preference for traditional solutions, fostering an environment where participants explore the intricate interplay between learning and classic knowledge.
These are some objectives of the workshop: delve into the strengths and limitations of traditional computer vision methods; critically examine scenarios where traditional approaches excel, considering their computational efficiency, data requirements, and power consumption; foster discussions on bridging the gap between learned and classic knowledge, exploring strategies to integrate the best of both worlds for more robust solutions.

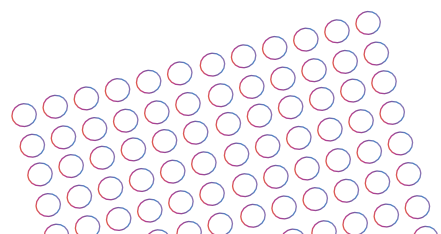|	
Topic:	**Machine Learning**
Workshop:	**Workshop on Unlearning and Model Editing (U&ME'24)**
Organizers:	Diego Garcia-Olano, Tal Hassner, Iacopo Masi, Mayank Vatsa
Website	https://sites.google.com/view/u-and-me-workshop
Room:	Suite 3
Time:	14:00 - 18:00

Summary: The 1st Workshop on Unlearning and Model Editing (U&ME'24) focuses on the growing need for new, efficient, and effective techniques for editing trained models, especially large generative models. Such models have practically unlimited functionality in the output they can generate. To provide this functionality, generative models require massive amounts of data and enormous compute costs to train, making it prohibitively expensive to retrain them whenever the need arises: when safety risks are uncovered, when deploying them to compute or storage-restricted platforms, or simply due to changing requirements. In particular, ensuring these models are safe and compliant with regulations can be difficult due to their broad range of capabilities and a continuously evolving regulatory landscape.

This workshop provides a venue for original scientific work presenting novel model editing techniques. We also encourage submissions on model compression methods, which aim to minimize the storage and computational costs of operating trained models with little impact on their performance. Additionally, papers that explore methods for adapting foundation models for efficient fine-tuning or editing are relevant.

Topic: **Machine Learning**
Workshop: **2nd Workshop on Quantum Computer Vision and Machine Learning (QCVML)**
Organizers: Jan-Nico Zaech, Tolga Birdal, Vladislav Golyanik, Michele Sasdelli
Website   https://qcvml.github.io/index.html
Room: Suite 2
Time: 14:00 - 18:00
Summary: The 2nd Workshop on Quantum Computer Vision and Machine Learning (QCVML) investigates how quantum computing can be applied to computationally hard problems in computer vision and machine learning. This workshop will explore the design of quantum algorithms specifically for image processing, the development of quantum-enhanced learning methods, and their potential applications in solving real-world vision challenges. It aims to bring together experts from computer vision and quantum computing to discuss both theoretical advancements and the practical implementation of quantum technologies within existing machine learning frameworks. An introduction to adiabatic and circuit based quantum computing will provide participants with the right resources for contributing to the research in the area. The overall goal is to foster interdisciplinary collaboration, the workshop seeks to assess the potential of quantum computing in advancing the field of computer vision and to identify especially promising areas within the community. Participants will engage in discussions on the current limitations and future directions of quantum technologies in this domain, aiming to accelerate their adoption and identify key research areas.

Topic: **Multimodal**
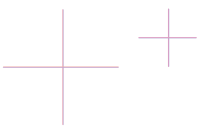**Workshop:**   **AVGenL: Audio-Visual Generation and Learning**
Organizers: Shiqi Yang, Zhixiang Wang, Tao Hu, Rodrigo Mira, Yaxing Wang, Joanna Hong, Zhi Zhong, Shusuke Takahashi, and Yuki Mitsufuji
Website   https://sites.google.com/view/avgenl
Room: Suite 9
Time: 14:00 - 18:00
Summary: In recent years, we have witnessed significant advancements in the field of visual generation which have molded the research landscape presented in computer vision conferences such as ECCV, ICCV, and CVPR. However, in a world where information is conveyed through a rich tapestry of sensory expe-

riences, the fusion of audio and visual modalities has become much more essential for understanding and replicating the intricacies of human perception and diverse real-world applications. Indeed, the integration of audio and visual information has emerged as a critical area of research in computer vision and machine learning, having numerous applications across various domains, from immersive gaming environments to lifelike simulations for medical training, such as multimedia analysis, virtual reality, advertisement and cinematic application.

Despite these strong motivations, little attention has been given to research focusing on understanding and generating audio-visual modalities compared to traditional, vision-only approaches and applications. Given the recent prominence of multi-modal foundation models, embracing the fusion of audio and visual data is expected to further advance current research efforts and practical applications within the computer vision community, which makes this workshop an encouraging addition to ECCV that will catalyze advancements in this burgeoning field.

In this workshop, we aim to shine a spotlight on this exciting yet under-investigated field by prioritizing new approaches in audio-visual generation, as well as covering a wide range of topics related to audio-visual learning, where the convergence of auditory and visual signals unlocks a plethora of opportunities for advancing creativity, understanding, and also machine perception. We hope our workshop can bring together researchers, practitioners, and enthusiasts from diverse disciplines in both academia and industry to delve into the latest developments, challenges, and breakthroughs in audio-visual generation and learning.

|	
Topic:	**Multimodal**
Workshop:	**Enabling Complex Perception Through Vision and Language Foundational Models**
Organizers:	Vijay Kumar B G, Yumin Suh, Samuel Schulter, Shiyu Zhao, Long Zhao, Dimitris N. Metaxas
Website	https://sites.google.com/view/omnilabel-workshop-eccv24/organizers
Room:	Amber 6
Time:	14:00 - 18:00
Summary:	The goal of this workshop is to foster research on the next generation of visual perception systems that reason over label spaces going beyond a list of simple category names. Modern applications require systems to understand a full spectrum of object labels in images, from plain category names (like ``person'') to complex object descriptions (``the man in the white hat walking next to the fire hydrant''). The previous edition of this workshop focused broadly on natural language to handle such complex label spaces. We introduced a challenging benchmark called Omnilabel with diverse and complex object descriptions and also proposed a comprehensive evaluation metric that handles virtually infinite label spaces. In the workshop this year, we take inspiration from the progress of foundational models like LLMs and VLMs to advance unified label spaces in perception tasks. Large vision-language models (VLMs) like CLIP, trained on massive web datasets of image-text pairs, achieve strong alignment between visual content and textual descriptions. This allows them to learn object representations usable for detection beyond predefined categories, demonstrating open vocabulary detection capabilities. Furthermore, LLMs, trained on vast text corpora, exhibit strong reasoning abilities, compositional understanding, and even a degree of world knowledge that are crucial
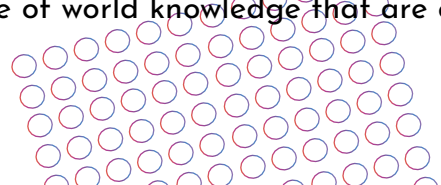
for comprehending complex, real-world object descriptions. In addition, their generative abilities can be leveraged to create diverse descriptions, potentially enhancing the training of language-based detectors. By leveraging the strengths of both VLMs and LLMs, we can not only develop next-generation perception systems that grasp complex scenes but also significantly reduce manual annotation effort, a major bottleneck in traditional approaches.

In line with this topic, we are upgrading our OmniLabel benchmark dataset with more complex object descriptions, and host the next edition of the Omni-Label language-based object detection challenge.

Topic:	**Responsible AI**
Workshop:	**The Dark Side of Generative AIs and Beyond**
Organizers:	Ryuichiro Hataya, Satoshi Hara, Hiromi Arai, Han Bao, Jingfeng Zhang
Website	https://sites.google.com/view/darksideofgenaiandbeyond
Room:	Suite 4
Time:	14:00 - 18:00
Summary:	The unprecedented advances in generative AIs have substantially revolutionized computer vision. At the same time, it is essential to look at the emerging dark side of generative AIs.

This workshop aims to shed light on the shadows of generative AIs from various aspects, including, but not limited to, Ethical implications: What ethical problems are amplified by generative AIs, and what are their societal impacts? Disinformation: How are generative AIs weaponized to create and spread disinformation? Legal aspects: What are the current legal challenges surrounding generative AIs, such as copyrights? Model collapse: How does the contamination of generated data into training datasets affect future computer vision models? Remedies: How can we mitigate these problems and develop trustworthy generative AIs?

By discussing these pivotal themes, this workshop aims not only to highlight the challenges posed by generative AIs but also to discuss responsible innovations and applications in computer vision.

Topic:	**Responsible AI**
Workshop:	**Explainable Computer Vision: Where are We and Where are We Going?**
Organizers:	Robin Hesse, Sukrut Rao, Moritz Böhle, Quentin Bouniot, Simone Schaub-Meyer, Stefan Roth, Kate Saenko, Bernt Schiele
Website	https://excv-workshop.github.io
Room:	Brown 1
Time:	14:00 - 18:00
Summary:	Deep neural networks (DNNs) are an essential component in the field of computer vision and achieve state-of-the-art results in almost all of its sub-disciplines. However, they are often too complex to be understood by humans, leading to them often being referred to as "black-box models". This is of particular concern when DNNs are applied in safety-critical domains such as autonomous driving or medical applications. With this problem in mind, explainable artificial intelligence (XAI) aims to gain a better understanding of DNNs, ultimately leading to more robust, fair, and interpretable models. While this important field of research is gaining more and more traction, there is also justified criticism of the way in which the research is conducted. The goals of this workshop are thus two-fold:

1. Discussion and dissemination of ideas at the cutting-edge of XAI research ("Where are we?")
2. A critical introspection on the challenges faced by the community and the way to go forward ("Where are we going?")

|	
Topic:	**Scene Understanding**
Workshop:	**OpenSUN3D: 3rd Workshop on Open-Vocabulary 3D Scene Understanding**
Organizers:	Francis Engelmann, Ayca Takmaz, Jonas Schult, Johanna Wald, Yuanwen Yue, Despoina Paschalidou, Elisabetta Fedele, Alexandros Delitzas, Songyou Peng, Xi Wang, Or Litany, Federico Tombari, Leonidas Guibas, Marc Pollefeys
Website	https://opensun3d.github.io
Room:	Amber 4
Time:	14:00 - 18:00
Summary:	Foundation models enabling open-vocabulary 3D scene understanding is an exciting new research frontier, unveiling open problems with a potential to transform the computer vision landscape. Similar to the 1st and 2nd edition of the OpenSUN3D workshop, the third edition will consist of diverse formats including keynote talks by leading experts from the field, an exciting new set of workshop challenges, oral paper presentations, and a panel discussion bringing together the research community from academia and industry. New this year is an extended challenge for predicting interactive elements and functionalities in 3D scenes introducing exciting opportunities to infer potential interactions and utilities within complex spatial environments. The workshop will be used as a platform for researchers from different disciplines to interact and develop this emerging and rapidly developing research area. The main focus of the workshop is on the new paradigm for 3D scene understanding leveraging foundation models to go beyond the capabilities of traditional methods that operate in a closed-world setting. We invite talks and paper submissions on topics related to 3D scene understanding in an open-vocabulary setting. These include, but are not limited to: Open-vocabulary 3D semantic, instance and part segmentation; Open-vocabulary 3D semantic scene completion and generation; Language-driven and long-tail 3D object detection; Visually obscured object localization; Predicting functionality, interactions, material type, and physical properties; Object part and mobility estimation for interaction and robotic applications; Ethical implications of 3D scene understanding with foundation models; Applications of open-set 3D scene understanding in robotics and AR/VR

Topic:	**Sensing Devices**
Workshop:	**Workshop on Neuromorphic Vision (NeVi): Advantages and Applications of Event Cameras**
Organizers:	Federico Becattini (University of Siena, Italy), Gaetano Di Caterina (University of Strathclyde, UK), Yulia Sandamirskaya (ZHAW Zurich University of Applied Sciences, Switzerland), Gregory Cohen (Western Sydney University, Australia), Luca Cultrera (University of Florence, Italy), Lorenzo Berlincioni (University of Florence, Italy), Suzanne Little (Dublin City University, Ireland), Joseph Lemley (Tobii, Ireland), Chiara Bartolozzi (Italian Institute of Technology, Italy), Gaurvi Goyal (Italian Institute of Technology, Italy), Arren Glover (Italian Institute of Technology, Italy), Axel von Arnim (Fortiss, Germany)
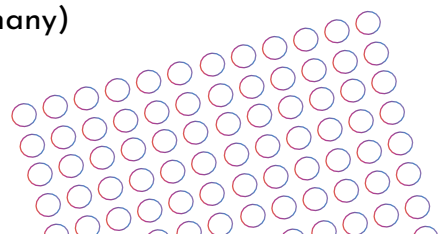
Website	https://sites.google.com/view/nevi2024
Room:	Panorama Lounge
Time:	14:00 - 18:00

Summary: Neuromorphic sensors, also known as event cameras, are a class of imaging devices mimicking biological visual systems. Unlike traditional frame-based cameras, which capture images synchronously, neuromorphic sensors continuously generate events capturing asynchronous illumination changes. Event cameras have initially gained interest in the field of robotics due to their low power consumption, extremely low latency, high dynamic range and absence of motion blur. Yet, this wide range of intriguing properties has rapidly enabled new, cutting-edge applications, especially for motion-centric tasks. The very fine temporal granularity of event cameras allows to easily capture complex temporal dynamics in a scene, so that the tackling of complex tasks can abstract from the low-level processing, and focus directly on higher-level cognition. In the past few years, we have witnessed the development of new astonishing technologies based on neuromorphic vision: low latency and low power consumption have allowed drones to effectively avoid fast-moving obstacles; high dynamic range and lack of motion blur allowed self-driving cars to detect other vehicles and pedestrians in adverse conditions such as low illumination; micro-second temporal granularity has enhanced the analysis of human micro-expressions and emotions. Many other groundbreaking applications are leveraging neuromorphic sensors, from high-speed object counting and defect detection to vibration measurement, fluid monitoring and time-to-contact estimation for spacecraft landing. Event-based processing has also been shown to provide an extra layer of privacy preservation compared to standard cameras, an important addition especially in light of the recent definition of the AI Act by the European Commission to regulate the development of artificial intelligence. This workshop aims to foster the growth of event-based research, by gathering researchers in the field and improving the communication between academia and industry, towards the discovery of new bleeding-edge neuromorphic technologies.

MONDAY, SEPTEMBER 30TH
WORKSHOPS & TUTORIALS

08:00 – 18:00	Registration & Badge pickup
10:30 – 11:00	Coffee Break
13:00 – 14:00	Lunch
15:30 – 16:00	Coffee Break

TUTORIALS

Tutorial: **A Bayesian Odyssey in Uncertainty: from Theoretical Foundations to Real-World Applications**

Organizers: Gianni Franchi, Olivier Laurent, Alexander Immer, Pavel Izmailov, Andrei Bursuc

Website https://uqtutorial.github.io/

Room: Suite 7

Time: 09:00 - 13:00

Summary: This tutorial is here to help researchers understand and handle uncertainty in their models, making them more reliable using Bayesian methods. We'll start by discussing different Bayesian approaches and then focus on Bayesian Neural Networks and how to approximate them efficiently for computer vision tasks. We will also use real-world examples and practical methods to show how to put these ideas into practice.

Tutorial: **Emerging Trends in Disentanglement and Compositionality**

Organizers: Xin Jin, Tao Yang, Yue Song, Xingyi Yang, Jingwen Ye, Wenjun (Kevin) Zeng, Nicu Sebe, Xinchao Wang, Shuicheng Yan

Website https://trend-in-disen-and-compo.github.io/

Room: Amber 4

Time: 09:00 - 13:00

Summary: Participants will learn about advanced techniques and models that allow for the disentanglement of visual factors in images and the compositionality of these factors to produce more meaningful representations. Disentanglement and Composition are believed to be one of the possible ways for AI to fundamentally understand the world, and eventually achieve Artificial General Intelligence (AGI). Specifically, we will cover the following topics:Session #1: Disentangled Representation Learning (DRL); Session #2: Latent Semantics Discovery; Session #3: Disentanglement and Equivariance; Session #4: Composition and Disentanglement for AGI.

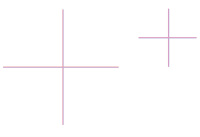**Tutorial:** **Recent Advances in Video Content Understanding and Generation**

Organizers: Yansong Tang, Ailing Zeng, Mike Zheng Shou, Josef Kittler, Philip Torr

Website https://venue-tutorial.github.io

Room: Amber 3

Time: 09:00 - 18:00

Summary: Over the last decade, tremendous interests have been attracted to this field, and great success has been achieved for various video-centric tasks (e.g., action recognition, motion capture and understanding, video understanding, temporal localization, and video generation etc.) based on conventional short videos. In recent years, with the explosion of videos and various application demands (e.g., video editing, AR/VR, human-robot interaction etc.), significantly more efforts are required to enable an intelligent system to understand

and generate video content under different scenarios within multimodal, long-term and fine-grained inputs. Moreover, with the development of recent large language models (LLMs) and large multimodal models (LMMs), there are growing new trends and challenges to be discussed and addressed. The goal of this tutorial is to foster interdisciplinary communication among researchers so that more attention from the broader community can be drawn to this field. This tutorial will discuss current progress and future directions, and new ideas and discoveries in related fields are expected to emerge.

Tutorial:	**Time is precious: Self-Supervised Learning Beyond Images**
Organizers:	Shashanka Venkataramanan, Mohammadreza Salehi, Yuki M. Asano
Website	https://shashankvkt.github.io/eccv2024-SSLBIG-tutorial.github.io/
Room:	Amber 7 + 8
Time:	09:00 - 13:00
Summary:	Self-supervised learning (SSL) is a powerful technique that's been able to reach the same levels of performance as supervised pretraining, thanks to its ability to work with billions of images. Existing image-based self-supervised methods focus on learning representations that are invariant to synthetic augmentations, addressing key image aspects like scale, color, and translation invariance. These factors represent only a fraction of the intricate variations in pose, viewpoint, and motion found naturally in videos, which contain more nuanced and informative signals. In this tutorial, we discuss an innovative shift from traditional augmentation invariance strategies prevalent in image-based SSL to a more dynamic utilization of video data. We discuss multiple techniques that leverage the rich temporal signals in videos through a dual focus on the spatial details within each frame and the changes occurring between them. Our tutorial presents a forward-looking perspective on incorporating temporal information to improve SSL algorithms, representing an paradigm shift in SSL, crucial for newcomers in the field.

Tutorial:	**Inside Plato's door: a tour in Multi-view Geometry**
Organizers:	Luca Magri, Federica Arrigoni
Website	https://sites.google.com/view/platomultiview24/home-page
Room:	Suite 7
Time:	14:00 - 18:00
Summary:	This tutorial offers a self-contained overview of multi-view geometry, an essential tool in Computer Vision that is at the core of several classical and modern algorithms involving a variety of applications, including 3D reconstruction, motion segmentation, visual SLAM and augmented reality. Aimed primarily at an audience unfamiliar with these topics, this tutorial will guide the attendees on a tour that starts from the geometric foundations of 3D vision to reach up some of the research challenges that are still open in the field. The course will introduce classical topics such as camera models, camera calibration, the epipolar geometry, and then move to multi-view relations. We will also address the problem of robust estimation and structure from motion. Special emphasis will be put on rigorous mathematical formulations and on the practical aspects that separate theory from efficient and working implementations.
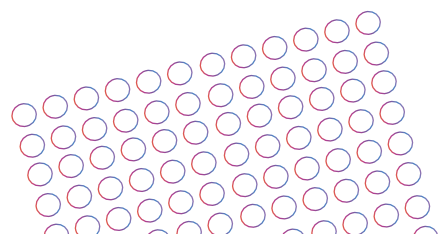
WORKSHOP

Topic:	**3D Vision**
Workshop:	**Dense Neural SLAM Workshop (NeuSLAM)**
Organizers:	Matteo Poggi, Fabio Tosi Youmin Zhang Yiyi Liao, Vladimir Yugay, Yue Li, Martin R. Oswald
Website	https://sites.google.com/view/neuslam
Room:	Amber 6
Time:	09:00 - 13:00
Summary:	The workshop aims to bring researchers together being interested in dense neural SLAM methods and provide a forum for the mutual exchange of experiences, beliefs, intuitions, ideas and future research directions. With a mix of invited keynote talks of renown experts, a poster session and a panel discussion, we invite people both from academia and industry across all seniority levels to learn from each other and engage in discussions.

Topic:	**3D Vision**
Workshop:	**Wild3D: 3D Modeling, Reconstruction, and Generation in the Wild**
Organizers:	Wei-Chiu Ma, Shenlong Wang, Lingjie Liu, Yufei Ye, Despoina Paschalidou, Natalia Neverova, David Fouhey, Shubham Tulsiani, Qixing Huang
Website	https://3d-in-the-wild.github.io/
Room:	Brown 3
Time:	09:00 - 13:00
Summary:	The goal of this workshop is to bring together researchers who are studying approaches to reconstructing and understanding 3D models in challenging, real-world environments. While significant progress has been made in the field of multi-view 3D modeling, many existing methods are limited by assumptions of controlled acquisition, such as static scenes and densely captured viewpoints. On the other hand, advances in 3D learning and the widespread availability of 2D and 3D visual data offer new opportunities to overcome these challenges and make 3D modeling more robust, widely applicable, easily accessible, and cost'effective. This workshop aims to address the challenges of 3D modeling in uncontrolled, partially-observed, and noisy environments and to explore potential future directions in the field. By fostering communication and highlighting important work in this area, we hope to inspire new research topics and breakthroughs.

Topic:	**Applications**
Workshop:	**CV For Ecology Workshop (CV4E)**
Organizers:	Sara Beery, Julia Chae, Faizan Khan, David Russell, Andrew Temple, Edward Vendrow, Derek Young, Mohamed Elhoseiny
Website	https://cv4e.netlify.app/
Room:	Suite 8
Time:	09:00 - 13:00
Summary:	As the biodiversity crisis mounts, scientists are collecting biologically and ecologically relevant data at unprecedented scales and speeds. Yet, their ability to operationalize these data and apply them to tackling the many far-reaching challenges that face biodiversity in a modern world has become among the greatest bottlenecks constraining progress in this field. Computer vision holds the potential to unlock this bottleneck and in doing so help to

avert a mass extinction in the Anthropocene. The Computer Vision for Ecology workshop aims to bring together experts to foster discussion on the automation of ecological data collection, collation, and analyses. Our goal is to establish a hub for the broader computer vision and ecology community at ECCV. We hope that this workshop will bring attention to this important and growing field, encouraging more researchers to work on these topics, creating lasting collaborations, and moving forward research in this area.

The workshop will encompass applications of computer vision across a wide variety of ecological systems, spanning both terrestrial and aquatic systems, ranging from the tropics to the Arctic, and from urban to rural settings. The workshop will also make specific efforts to encompass applications of computer vision which can be deployed across both low- and high-income nations. The topics we aim to address include, but are not limited to Remote sensing, Bioacoustics, Video and image-based monitoring, Citizen science, Long-tailed recognition, Zero-shot learning, expert AI systems, and Robust model deployment.

Topic:	**Applications**
Workshop:	**2nd workshop on Vision-based InduStrial InspectiON (VISION)**
Organizers:	Shiyu Li, Shancong Mou, Hao Yan, Gokberk Cinbis, Yuhao Chen, Zhengfeng Lai, Ziyue Li, Tatiana Likhomanenko, Javad Shafiee, C Thomas, Alexander Wong, Carrie Yu, Chen Zhang
Website	https://vision-based-industrial-inspection.github.io/eccv-24/
Room:	Tower Lounge
Time:	09:00 - 13:00
Summary:	The VISION 24 workshop will serve as a dynamic forum for the exchange of scholarly innovations and practical challenges in Vision-based Industrial Inspection. The workshop will feature keynote talks, technical presentations, and challenge competitions, fostering collaboration between interdisciplinary research communities and industry practitioners. Topics of interest include the development and deployment of vision-based inspection technology, meta-learning, domain adaptation, multi-modality data fusion, and robust machine learning for high-stake applications. A key highlight of the workshop is the VISION 24 Data Challenge, focusing on One Shot Industrial Defect Segmentation. Detailed information of the workshop and the challenge are available on the official website at https://vision-based-industrial-inspection.github.io/eccv-24/

Topic:	**Art**
Workshop:	**Vision for Art (VISART) VII Workshop**
Organizers:	Amanda Wasielewski, Nanne van Noord, Noa Garcia, Peter Bell, Stuart James
Website	https://visarts.eu/
Room:	Amber 1
Time:	09:00 - 13:00
Summary:	VISART continues its role as a forum for the presentation, discussion and publication of Computer Vision (CV) techniques for the understanding of art and visual culture. The explosion in the generative art, digitization of artworks and digitally-born art highlights the importance of application in the overlap between CV and art; such as ways to reason, connecting language, structuring data (i.e. databases) for Art to Cultural Heritage.
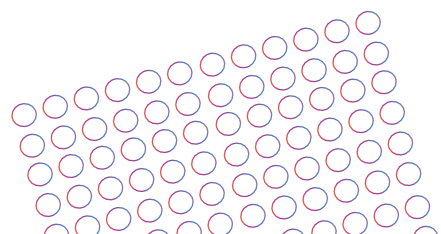
Topic:	**Autonomous Driving and Robotics**
Workshop:	**Vision-Centric Autonomous Driving (VCAD) Workshop**
Organizers:	Yue Wang, Hang Zhao, Yiming Li, Vitor Guizilini, Xinshuo Weng, Boyi Li, Marco Pavone
Website	https://vcad-workshop.github.io/
Room:	Suite 9
Time:	09:00 - 13:00
Summary:	With the commercialization of autonomous driving and assisted driving systems, the demand for high-performance, efficient, and scalable machine learning solutions is becoming more urgent than ever before. Visual perception is a key research area of self-driving that is always attracting a lot of attention since (1) visual data provides much richer information than other sensors; (2) cameras are affordable and pervasive on vehicles as well as other robotic systems; (3) visual foundation models are trending directions in machine learning. This workshop embraces topics around vision-centric and data-driven autonomous driving, including but not limited to the following topics: Visual perception for autonomous driving, Vision-language models for autonomous driving, Neural rendering of driving scenes, Visual world models for autonomous driving, Vision-centric end-to-end driving, Visual representation learning for autonomous driving, Multi-sensory fusion, Few-shot/semi-supervised/self-supervised learning, Motion prediction, planning, and simulation, New datasets and metrics for autonomous driving, Privacy concerns on visual data.

Topic:	**Autonomous Driving and Robotics**
Workshop:	**Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving**
Organizers:	Hong Kong University of Science and Technology, The Chinese University of Hong Kong, Dalian University of Technology, Huawei, TU Delft
Website	https://coda-dataset.github.io/w-coda2024/
Room:	Space 2
Time:	09:00 - 13:00
Summary:	This workshop aims to bridge the gap between state-of-the-art autonomous driving techniques and fully intelligent, reliable self-driving agents, particularly when confronted with corner cases, rare but critical situations that challenge limits of reliable autonomous driving. The advent of Multimodal Large Language Models (MLLMs), represented by GPT-4V, demonstrates the unprecedented capabilities in multimodal perception and comprehension even under dynamic street scenes. However, leveraging MLLMs to tackle the nuanced challenges of self-driving still remains an open field. This workshop seeks to foster innovative research in multimodal perception and comprehension, end-to-end driving systems, and the application of advanced AIGC techniques to autonomous systems. We conduct a challenge comprising two tracks: corner case scene understanding and corner case scene generation. The dual-track challenge is designed to advance reliability and interpretability of autonomous systems in both typical and extreme corner cases.

Topic:	**Detection, Recognition, and Low-Level Vision**
Workshop:	**Instance-Level Recognition**
Organizers:	Andre Araujo, Bingyi Cao, Kaifeng Chen, Ondrej Chum, Noa Garcia, Bohyung Han, Guangxing Han, Giorgos Tolias, Hao Yang, Nikolaos-Antonios Ypsilantis, Xu Zhang

Website	https://ilr-workshop.github.io/ECCVW2024/
Room:	Amber 5
Time:	09:00 - 13:00
Summary:	Our workshop is focused on visual Instance-Level Recognition (ILR), with a primary objective of identifying, comparing, or synthesizing images related to specific objects, scenes, or events. Unlike the broad categorization found in category-level recognition, where classes are defined semantically (e.g., "a chair"), ILR delves into tasks with the utmost granularity in class definition, such as identifying "the chair of my desk".

This year, we expand the scope of our workshop by introducing a call for papers, in addition to hosting keynote talks by renowned speakers and invited paper talks from the main conference.

The 2024 Instance-Level Recognition (ILR) Workshop is a follow-up of five successful editions of our previous workshops — the first two having focused only on landmark recognition (CVPRW18, CVPRW19), the following ones expanding to the domains of artworks and products (ECCVW20, ICCVW21), and the latest one introducing the universal image embedding problem (ECCVW22).

Topic:	**Efficiency**
Workshop:	**Computational Aspects of Deep Learning**
Organizers:	Giuseppe Fiameni, Lorenzo Baraldi, Iuri Frosio, Claudio Baecchi
Website	http://www.cadl.it
Room:	Amber 2
Time:	09:00 - 13:00
Summary:	The fourth edition of the International Workshop on Computational Aspects of Deep Learning (CADL), held in conjunction with the ECCV conference, addresses the growing computational demands of deep learning (DL) in different fields. The workshop aims to bring together experts to discuss and develop solutions for optimizing DL models, making them scalable and energy-efficient. Key topics include distributed training, novel architectures, energy reduction, and efficient algorithms. The workshop also emphasizes inclusivity by encouraging participation from researchers with limited computational resources. The workshop's relevance lies in the rapid expansion of DL, which requires significant computational resources, especially for training large-scale neural networks. By promoting high-performance computing (HPC) systems and efficient DL architectures, CADL seeks to democratize access to advanced DL research, enabling broader participation in the field. The event features tutorials, talks, and keynotes, with a program designed to address architectural design, vision-language models, and transformers for vision. Organized by experienced professionals from NVIDIA and various universities, the workshop fosters diversity and inclusion, aiming to increase participation from underrepresented groups and regions. The CADL workshop continues to build on its success from previous editions, maintaining its focus on computational efficiency and inclusivity in DL and CV research.

Topic:	**Machine Learning**
Workshop:	**Uncertainty Quantification for Computer Vision**
Organizers:	Andrea Pilzer (NVIDIA), Gianni Franchi (ENSTA Paris), Andrei Bursuc (valeo.ai), Arno Solin (Aalto University), Martin Trapp (Aalto University), Rui Li (Aalto University), Marcus Klasson (Aalto University), Gaëtan Brison (Hi!Paris), Angela Yao (National University of Singapore), Ivor Simpson (University of Sussex), Neill D. F. Campbell (University of Bath)

Website	https://uncertainty-cv.github.io/2024/
Room:	Brown 2
Time:	09:00 - 13:00
Summary:	In the last decade, substantial progress has been made w.r.t.

the performance of computer vision systems, a significant part of it thanks to deep learning. These advancements prompted sharp community growth and a rise in industrial investment. However, most current models lack the ability to reason about the confidence of their predictions; integrating uncertainty quantification into vision systems will help recognize failure scenarios and enable robust applications.

In addition to advances in Bayesian deep learning, providing practical approaches for vision problems, the workshop will provide a forum for discussing promising research directions, which have received less attention, as well as advancing current practices to drive future research. Examples include: the development of new metrics that reflect the real-world need for uncertainty when using vision systems with down-stream tasks; and moving beyond point-estimates to address the multi-modal ambiguities inherent in many vision tasks.

This years UNcertainty quantification for Computer Vision (UNCV) Workshop aims to raise awareness and generate discussion regarding how predictive uncertainty can, and should, be effectively incorporated into models within the vision community. The workshop will bring together experts from machine learning and computer vision to create a new generation of well-calibrated and effective methods that know when they do not know.

Topic:	**Machine Learning**
Workshop:	**The 3rd Workshop for Out-of-Distribution Generalization in Computer Vision Foundation Models**
Organizers:	Bingchen Zhao; Wufei Ma; Artur Jesslen; Jiahao Wang; Siwei Yang; Sagar Vaze; Haoqin Tu; Zijun Wang; Xinyu Yang; Huaxiu Yao; Cihang Xie; Oliver Zendel; Kai Han; Alan Yuille; Adam Kortylewski
Website	https://www.ood-cv.org/
Room:	Brown 1
Time:	09:00 - 13:00
Summary:	Deep learning models are usually developed and tested under the implicit assumption that the training and test data are drawn independently and identically distributed (IID) from the same distribution. Overlooking out-of-distribution (OOD) images can result in poor performance in unseen or adverse viewing conditions, which is common in real-world scenarios.

In this workshop, we are interested in discussing the performance of computer vision models on OOD images which follows a different distribution than the training images. Also given the current trend of web-scale pretrained computer vision models, it is of interest to better understand their performance in the OOD or rare scenarios. Our workshop will be featuring three competitions, OOD generalization on the OOD-CV dataset, open-set recognition and generalized category discovery on the Semantic-shift benchmark.

Topic:	**Machine Learning**
Workshop:	**2nd Workshop on More Exploration, Less Exploitation (MELEX)**
Organizers:	Vasileios Belagiannis, Azade Farshad, Angela Yao, Gustavo Carneiro
Website	https://sites.google.com/view/melex2024
Room:	Suite 2
Time:	09:00 - 13:00

Summary: Computer vision methods are constantly being optimised with waves of exploration and exploitation. During exploration, exciting innovative ideas are proposed, but they have a hard time to get noticed because they do not necessarily beat the state-of-the-art (SOTA) results. During exploitation, researchers tend to fine-tune high-performing approaches to beat the SOTA, where large efforts are needed to obtain relatively small gains. With the second CVF/ECCV More Exploration, Less Exploitation (MELEX) workshop we aim to publish papers in all areas of computer vision that propose explorative papers with innovative models, algorithms, and ideas that show competitive, but not necessarily the SOTA results in benchmarks.

Topic: **Responsible AI**

Workshop: TWYN: Trust What You Learn. 1st Workshop on Trustworthiness in Computer Vision

Organizers: Marco Cotogni, Jacopo Bonato, Luigi Sabetta, Lorenzo Baraldi, Sara Sarto, Samuele Poppi

Website   https://twyn.unimore.it/

Room: Suite 5

Time: 09:00 - 13:00

Summary: In an era of rapid advancements in Artificial Intelligence, fostering Trustworthy AI has become more crucial than ever. The inaugural "Trust What You learN (TWYN)" workshop aims to establish a dynamic platform for researchers, practitioners, and industry experts to explore and advance the convergence of Trustworthy AI and DeepFake Analysis within the field of Computer Vision. This workshop will address the complex task of developing AI systems that are not only technically advanced but also ethical, transparent, and accountable. Participants will have the opportunity to either examine the intricate balance between learning and unlearning, along with the vital role of privacy in shaping the future of Computer Vision, or gain a comprehensive understanding of the latest methods and advancements in detecting and combating the challenges posed by deepfake technologies.
The workshop's dual focus on Trustworthy AI and DeepFake Analysis underscores its commitment to tackling the growing threats posed by deepfake technologies while promoting the development of responsible AI practices.

Topic: **Responsible AI**

Workshop: xAI4Biometrics at ECCV 2024 - 4th Workshop on Explainable & Interpretable Artificial Intelligence for Biometrics

Organizers: Ana F. Sequeira, Helena Montenegro, Isabel Rio-Torto, Leonardo G. Capozzi, Pedro C. Neto, Rafael C. Maia, Tiago Gonçalves, Wilson Silva

Website   https://vcmi.inesctec.pt/xai4biom_eccv2024/index.html

Room: Suite 3

Time: 09:00 - 13:00

Summary: The ultimate goal of the proposed xAI4Biometrics 2024 workshop is to motivate the better understanding, through explainable artificial intelligence (xAI), of several aspects of biometrics and, with this understanding, to foster trust and reliability in current common accepted practices in biometrics' applications. These applications can be found in scenarios comprising e.g. identity verification for access/border control, detection of persons of interest using watch lists of suspects, forensic applications (like iris-based corpse identification or comparison of suspects' and crime scene evidence such as fingerprints), that affect the daily life of an ever-growing population. Understanding the ML
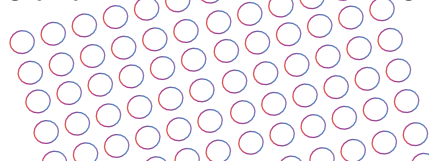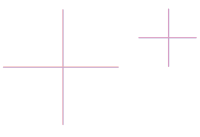
models' decisions and processes leading to decision making, will allow us to understand the rules used for making decisions, which, in turn, can improve our own AI-assisted (human) decision making. Furthermore, the deep knowledge on the inner workings of ML models may enable the prevention of unexpected situations and elimination of vulnerabilities.

Topic: Scene Understanding
Workshop: Map-free Visual Relocalization Workshop
Organizers: Áron Monszpart, Eric Brachmann, Filipe Gaspar, Guillermo Garcia-Hernando, Axel Barroso-Laguna, Daniyar Turmukhambetov, Gabriel J. Brostow, Victor Adrian Prisacariu
Website https://nianticlabs.github.io/map-free-workshop/2024
Room: Suite 6
Time: 09:00 - 13:00
Summary: The workshop addresses the challenge of metric visual relocalization using a single reference image instead of a map, a key issue for applications such as AR/MR, SLAM and 3D reconstruction. As debates continue regarding the necessity of HD maps, the workshop seeks to reduce reliance on them. The workshop hosts the first Map-free Visual Relocalization Challenge 2024, featuring two tracks: one for pose estimation from a single image and another from a sequence of images, with the latter providing a more practical AR-like scenario. Accepted workshop papers and challenge winners will be presented at the poster session.

Topic: Sensing Devices
Workshop: Neural Fields Beyond Conventional Cameras
Organizers: Tzofi Klinghoffer, Shengyu Huang, Daniel Gilo, Kushagra Tiwary, Akshat Dave, Lingjie Liu, James Tompkin, Or Litany, Ramesh Raskar
Website https://neural-fields-beyond-cams.github.io/
Room: Panorama Lounge
Time: 09:00 - 13:00
Summary: Neural fields have been widely adopted for learning novel view synthesis and 3D reconstruction from RGB images by modeling transport of light in the visible spectrum. This workshop focuses on neural fields beyond conventional cameras, including (1) learning neural fields from data from different sensors across the electromagnetic spectrum and beyond, such as lidar, cryo-electron microscopy (cryoEM), thermal, event cameras, acoustic, and more, and (2) modeling associated physics-based differentiable forward models and/or the physics of complex light transport (reflections, shadows, polarization, diffraction limits, optics, scattering in fog or water, etc.). Our goal is to bring together a diverse group of researchers using neural fields across sensor domains to foster learning and discussion in this growing area.

Topic: 3D Vision
Workshop: Geometry in Large Model Era
Organizers: Yichen Li, Congyue Deng, Katie Luo, Yonglong Tian, Yue Wang, Jiajun Wu, Minhyuk Sung, Wojciech Matusik, Leonidas Guibas
Website https://geo-lme.github.io
Room: Brown 3
Time: 14:00 - 18:00
Summary: Geometric computing, which is crucial for extracting topological and geometric information from shapes, scenes, and sequences, has advanced

significantly through deep learning. Recent breakthroughs in areas such as 2D-to-3D content generation and open-vocabulary segmentation have opened up new opportunities for 3D content creation, perception, and analysis. We aim to bring together researchers from 3D computer vision, computational geometry, computer graphics, and deep learning to discuss the integration of geometric computing techniques with deep learning, the development of geometric losses and regularizations, feature learning from large-scale data, and solving geometric problems with large models. The workshop will also address geometry's role in open-vocabulary perception, and the future of geometric learning with large models, as the community continues to explore new possibilities with large-scale data and multimodal models.

Topic:	**3D Vision**
Workshop:	**AI3DCC: The Second Workshop of AI for 3D Content Creation**
Organizers:	Despoina Paschalidou, Georgios Pavlakos, Davis Rempe ,Angel Xuan Chang, Kai Wang, Amlan Kar, Daniel Ritchie, Kaihcun Mo, Manolis Savva, Paul Guerrero,Siyu Tang, Leonidas
Website	https://ai3dcc.github.io/
Room:	Brown 1
Time:	14:00 - 18:00
Summary:	Developing algorithms capable of generating realistic, high quality 3D content at scale has been a long standing problem in Computer Vision and Graphics. We anticipate that having generative models that can reliably synthesize meaningful 3D content will completely revolutionize the workflow of artists and content creators, and will also enable new levels of creativity through ``generative art". Although recently there has been considerable success in generating photorealistic images, the quality and generality of 3D generative models has lagged behind their 2D counterparts. Additionally, efficiently controlling what needs to be generated and scaling these approaches to complex scenes with several static and dynamic objects still remains an open challenge. In this workshop, we seek to bring together researchers working on generative models for 3D shapes, humans, and scenes to discuss the latest advances, existing limitations and next steps towards developing generative pipelines capable of producing fully controllable 3D environments with multiple humans interacting with each other or with objects in the scene. In the last few years, there has been significant progress in generating 3D objects, humans, and scenes independently, but only recently has the research community shifted their attention towards generating meaningful dynamics and interactions between humans or humans and other scene elements. To this end, in our workshop we look forward to cover the following topics: (i) representations: What is the most appropriate representation for generating meaningful high-quality 3D textured objects? What is the best representation that can enable intuitive control over the generated objects? How can we effectively represent interactions between humans and objects?, (ii) modelling: How can we construct powerful models that can reliably generate humans performing plausible real life activities? How can we construct an accurate 3D model of our world that could allow us to manipulate both the appearance of the scene elements as well as their spatial composition? (iii) common sense reasoning and knowledge: How can we incorporate common sense knowledge about 3D objects and scenes, such as part structures, arrangements of objects from Large Language Models and Vision Language Models to enable training with fewer data? (iv) applications: How can we leverage 2D priors to enable photo-realistic 3D content

creation? What are the creative tools that designers and product designers really need make their workflow more efficient? Are there new fields that could benefit from generative AI, such as construction and agriculture? (iv) risks and ethical considerations: What are the ethical implications that arise from artificially-generated 3D content and how we can address them?

Topic:	**Applications**
Workshop:	**FashionAI: Exploring the intersection of Fashion and Artificial Intelligence for reshaping the Industry**
Organizers:	Rita Cucchiara, Emanuele Frontoni, Marcella Cornia, Marina Paolanti
Website	https://sites.google.com/view/fashionai2024
Room:	Amber 4
Time:	14:00 - 18:00
Summary:	The fashion domain is entering a transformative era marked by both opportunities and challenges that are closely linked to the integration of generative AI and computer vision solutions, with a strong focus on both research and technology transfer. These challenges, which range from the automation of complex design processes to the personalization of customer experiences and the optimization of supply chains, are not unique to any single company but are shared by the entire spectrum of companies operating in the luxury fashion sector. They also resonate deeply within the computer vision community, where ongoing research and development are pushing the limits of what is possible with fashion AI.
	This workshop is designed to be a central platform for presenting the latest advances in addressing these common challenges. More importantly, it aims to establish a dedicated fashion AI community. Such a community is conceived as a collaborative network of AI researchers, fashion designers, industry practitioners, and technology innovators. Its purpose is to foster ongoing dialogue, exchange of ideas, and collaborative projects that not only address current challenges but also anticipate and prepare for future trends. By encouraging open communication and collaboration, the workshop aims to create a synergy that leverages the strengths of both the AI and fashion sectors.

Topic:	**Autonomous Driving and Robotics**
Workshop:	**ROAM: Robust, Out-of-Distribution And Multi-Modal models for Autonomous Driving**
Organizers:	Arturo Deza, Marcelo Contreras, Victor Flores, Dunant Cusipuma, David Ortega
Website	https://sites.google.com/view/roameccv2024/
Room:	Amber 6
Time:	14:00 - 18:00
Summary:	Workshop related to the discussion, analysis and creation of new Multi-Modal models that must be robust to out-of-distribution data across several modalities such as Vision, LIDAR, and Language applied to Autonomous Driving. This workshop in particular will also try to stress the generalization capacity of these models across different parts of the world including emerging economies.
Topic:	**Autonomous Driving and Robotics**
Workshop:	**Multi-Agent Autonomous Systems Meet Foundation Models: Challenges and Futures**
Organizers:	Haibao Yu, Jianing Qiu, Jiankai Sun, Li Chen, Mac Schwager, Ping Luo, Ruigang Yang, Si Liu, Zaiqing Nie
Website	https://coop-intelligence.github.io/
Room:	Amber 1

Time: 14:00 - 18:00

Summary: This workshop focuses on cooperative intelligence within multi-agent autonomous systems. The first session will primarily discuss "Multi-Agent Autonomous Systems Meet Foundation Models: Challenges and Futures". The progress in artificial intelligence has propelled the development of Embodied AI, particularly in autonomous driving and robotics. However, achieving autonomy in complex and open environments remains challenging for individual agents. This is where cooperative intelligence comes in, a paradigm where agents collaborate and interact with infrastructure equipment to handle diverse tasks effectively. In autonomous driving, the availability of datasets and breakthrough algorithms has spurred research interest in cooperative autonomous driving. Through Vehicle-to-Everything (V2X) interactions including Vehicleto-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I), autonomous vehicles can improve their perception and safety capabilities, surpassing the limitations of single-vehicle autonomy and paving the way for widespread adoption. Similarly, in robotics, recent advancements in multi-agent systems offer a cost-effective solution for exploring unknown environments. These breakthroughs empower robots to assist humans in navigating challenging tasks within open environments. Despite notable progress, the challenges of coordinating multi-agent systems remain insufficiently explored. Issues like determining what information to transmit, how to transmit it, and how to fuse information across different task levels (perception, prediction, planning, etc.) pose practical deployment hurdles. Recent breakthroughs in large language models and foundational models offer a promising avenue for addressing these challenges. This workshop, complementing existing workshops focused on individual intelligence in autonomous driving and robotics, introduces a fresh perspective. It fosters discussions on a comprehensive, system-level approach to cooperative framework design for autonomous agents across diverse domains.

Topic: **Detection, Recognition, and Low-Level Vision**
Workshop: **Large-scale Video Object Segmentation**
Organizers: Henghui Ding, Lingyi Hong, Chang Liu, Ning Xu, Linjie Yang, Yuchen Fan
**Website** https://lsvos.github.io/
Room: Suite 6
Time: 14:00 - 18:00

Summary: The 6th LSVOS challenge is held in conjunction with ECCV 2024 in MiCo Milano. In this workshop and challenge, we replace the classic YouTube-VOS benchmark with MOSE and LVOS to study the VOS under more challenging complex environments. MOSE focuses on complex scenes, including the disappearance-reappearance of objects, inconspicuous small objects, heavy occlusions, crowded environments, etc. LVOS focuses on long-term videos, with complex object motion and long-term reappearance. Besides, we also replace the origin YouTube-RVOS benchmark with MeViS. MeViS focuses on referring the target object in a video through its motion descriptions instead of static attributes, which breaks the basic design principles behind existing RVOS methods and boosts the rethinking of motion modeling. In addition, we will hold a series of talks by the leading experts in video understating.
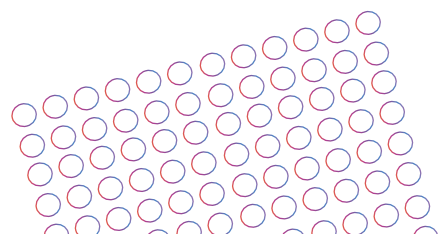
Topic: **Human**
Workshop: **Foundation Models for 3D Humans**
Organizers: Yao Feng, Yan Zhang, Naureen Mahmood, Yuliang Xiu, Weiyang Liu,
                    Rafael Wampfler, Georgios Pavlakos, Gul Varol, and Michael J. Black
Website   https://human-foundation.github.io/workshop-eccv-2024/
Room: Tower Lounge
Time: 14:00 - 18:00
Summary: Foundation models (*e.g.,* large language models (LLMs), large vision models,
                    text-to-image generative models) are impacting nearly every area of computer
                    vision and represent a paradigm shift from end-to-end learning of task-speci-
                    fic models to pretraining-finetuninng generalist models. Such a paradigm shift
                    urgently calls for rethinking the new challenges and opportunities in the study
                    of digital humans. This workshop is particularly focused on exploring two key
                    questions: (1) how foundation models can help the study of digital humans; (2)
                    how we can build foundation models for digital humans. The study of digital
                    humans encompasses a wide breadth of research areas, from high-fidelity
                    digital avatar modeling and lifelike human motion generation to human ani-
                    mation and behavior analysis. This workshop offers an in-depth exploration of
                    the vibrant intersection between 3D digital human modeling and foundation
                    models. We plan to bring experts in general foundation models and experts in
                    3D digital humans together to exchange ideas and discuss open challenges.

Topic: **Human**
Workshop: **Observing and Understanding Hands in Action**
Organizers: Hyung Jin Chang, Rongyu Chen, Zicong Fan, Otmar Hilliges, Kun He,
                    Take Ohkawa, Yoichi Sato, Elden Tse, Linlin Yang, Lixin Yang, Angela Yao,
                    Linguang Zhang
Website   https://hands-workshop.org
Room: Suite 8
Time: 14:00 - 18:00
Summary: Our HANDS workshop will gather vision researchers working on perceiving
                    hands performing actions, including 2D and 3D hand detection, segmentation,
                    pose/shape estimation, tracking, etc. We will also cover related applications
                    including gesture recognition, hand-object manipulation analysis, hand activity
                    understanding, and interactive interfaces. The eighth edition of this workshop
                    will emphasize the use of large foundation models (*e.g.* CLIP, Point-E, Segment
                    Anything, Latent Diffusion Models) for hand-related tasks. These models have
                    revolutionized the perceptions of AI, and demonstrate groundbreaking contri-
                    butions to multimodal understanding, zero-shot learning, and transfer learning.
                    However, there remains an untapped potential for exploring their applications
                    in hand-related tasks.

Topic: **Human**
Workshop: **7th Workshop and Competition on Affective Behavior Analysis in-the-wild**
Organizers: Dimitrios Kollias, Stefanos Zafeiriou, Irene Kotsia, Abhinav Dhall,
                    Shreya Ghosh
Website   https://affective-behavior-analysis-in-the-wild.github.io/7th/
Room: Suite 5
Time: 14:00 - 18:00

Summary: The diversity of human behavior, coupled with the richness of multi-modal data emerging from its analysis, and the multitude of applications requiring rapid advancements in this domain, ensure that our Workshop serves as a timely and relevant platform for discussion and dissemination. The Workshop will showcase novel contributions on recent progress in the recognition, analysis, generation-synthesis, and modeling of face, body, gesture, speech, audio, text, and language, while incorporating the most advanced systems available for in-the-wild analysis (i.e., in unconstrained environments) and across various modalities.

The Competition features two Challenges: one focused on Multi-Task Learning of widely used affect representations, and the other on Compound Expression Recognition. Multiple teams have participated in these Challenges and will present their results and findings.

The ABAW Workshop and Competition is distinguished by its unique ability to foster cross-pollination among different disciplines, bringing together experts from academia, industry, and government, as well as researchers in mobile and ubiquitous computing, computer vision and pattern recognition, AI and ML/DL, multimedia, robotics, HCI, ambient intelligence, and psychology. The ABAW Workshop and Competition builds on the legacy of its predecessors held at IEEE CVPR 2024, 2023, and 2022, ECCV 2022, ICCV 2021, IEEE FG 2020, and IEEE CVPR 2017 Conferences.

Topic: Machine Learning
Workshop: Knowledge in Generative Models
Organizers: Anand Bhattad, Xiaodan Du, Nick Kolkin, Shuang Li, Grace Luo, Greg Shakhnarovich

Website https://sites.google.com/ttic.edu/knowledge-in-generative-models/home
Room: Brown 2
Time: 14:00 - 18:00
Summary: Recent advances in generative modeling have been successful in creating rich, diverse and convincing photorealistic and stylized images. In this workshop, we aim to understand how generative models internally represent and process visual information, and whether they bring us closer to fulfillment of the known mantra "vision is inverse graphics". We seek to understand how well generative models comprehend semantic constructs that are commonly used to convey visual understanding, such as object recognition, scene understanding, spatial awareness, intrinsic image decomposition, and so on. Can this understanding be leveraged to solve inverse (recognition) problems? Can it be improved to further enhance generative models' abilities? Is something important still missing from how our large models represent the visual world? We believe this workshop is essential to answer these questions and deepen our understanding of how generative models represent and process visual information, bridging the gap between current capabilities and the potential to mimic human-like vision.

Topic: Machine Learning
Workshop: Emergent Visual Abilities and Limits of Foundation Models (EVAL-FoMo)
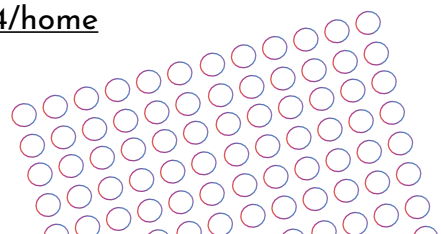**Organizers:** Ashkan Khakzar, Ameya Prabhu, Almut Sophia Koepke, Boyi Li, Sarah Schwettmann, Francesco Pinto, Neil Chowdhury, Jindong Gu, Aida Nematzadeh, Bo Li, Trevor Darrell, Philip Torr

Website https://sites.google.com/view/eval-fomo-24/home
Room: Amber 5

Time: 14:00 - 18:00

Summary: TLDR: This workshop focuses on analysis and evaluations to understand and identify emerging visual capabilities and pinpoint visual limits in foundation models.

Abstract: The computer vision landscape is dramatically evolving due to the capabilities of foundation models. These models exhibit capabilities they were not explicitly trained for, i.e. emergent abilities. Foundation models surpass standard bench-marks, highlighting the need for new evaluations. The inadequacy of current benchmarks inhibits the understanding and characterization of the emergent capabilities of these models (are they a mirage?). Moreover, limits in the visual abilities of these models have recently come into the spotlight, pointing to the need for innovative evaluations to identify these shortcomings.

Topic: **Machine Learning**
Workshop: **Workshop on Visual Concepts**
Organizers: Joy Hsu, Yunzhi Zhang, Jiayuan Mao, Shangzhe Wu, Daniel Cohen-Or, Jiajun Wu
Website https://sites.google.com/cs.stanford.edu/visualconcepts
Room: Suite 4
Time: 14:00 - 18:00

Summary: Concept learning and reasoning aim to extract compact and structured representations of the physical world, including sensing and actuation, and recompose them to tackle novel intricate problems. This paradigm suggests a unified view and solution to different recognition and generative tasks across different domains and modalities. An important research question is to under-stand and design concept representations that facilitate better learning from various datasets and compositional reasoning for novel tasks. As an endeavor to answering this question, in this workshop, we gather together researchers in computer vision, robotics, multi-modal learning, machine learning, and cogniti-ve science to discuss various topics on visual concept learning.

Topic: **Machine Learning**
Workshop: **Sometimes Less is More: The First Dataset Distillation Challenge**
Organizers: Ahmad Sajedi, Kai Wang
**Website** https://dd-challenge-main.vercel.app/
Room: Amber 2
Time: 14:00 - 18:00

Summary: Dataset distillation (DD) has emerged in recent years, aiming at significantly reducing a large dataset into a more compact one. In this workshop, we orga-nize the first dataset distillation challenge (https://www.dd-challenge.com) to provide a fair comparison of various approaches and facilitate the develop-ment of this area. This challenge has three tracks: fixed mage-per-class (IPC) track, generative track, and open track. In the fixed IPC track, all methods are evaluated with a small IPC, to demonstrate the effectiveness under the extreme compression ratio. Meanwhile, approaches in the generative track are permitted to produce any number of images within a certain time, evaluating the effectiveness and efficiency simultaneously. Finally, we do not set restri-ctions in the open track and encourage any insightful explorations. We hope this challenge could provide a standard benchmark for dataset distillation and facilitate future research.

Topic: **Machine Learning**
Workshop: **Synthetic Data for Computer Vision**
Organizers: Lucia Cascone, Zilong Huang, Michele Nappi, Xinggang Wang
Website https://syntheticdata4cv.wordpress.com/
Room: Space 2
Time: 14:00 - 18:00
Summary: In the rapidly evolving field of computer vision, synthetic data is gaining prominence as a critical tool to address data limitations and enhance the accuracy and reliability of systems. This workshop seeks to delve into the positive implications of synthetic data, such as improved data accessibility and privacy protection, while also addressing the inherent challenges and risks, including the potential misuse of synthetic data in security contexts. The exploitation of synthetic data to bypass recognition systems poses significant threats, highlighting the need for high-quality data generation and the development of robust detection algorithms to safeguard against unauthorized access and fraudulent activities.

Topic: **Multimodal**
Workshop: **Multimodal Agents Workshop**
Organizers: Zane Durante, Ehsan Adeli, Juan Carlos Niebles, Naoki Wake, Bidipta Sarkar, Ran Gong, Jae Sung Park, Yejin Choi, Fei-Fei Li, Qiuyuan Huang
Website https://multimodalagents.github.io/
Room: Amber 7 + 8
Time: 14:00 - 18:00
Summary: The field of artificial intelligence is experiencing a surge in the effectiveness of multimodal agents, signaling their potential to become ubiquitous in our everyday lives. Our proposed workshop is dedicated to fostering the advancement of multimodal AI systems by focusing on their development as agents in distinct environments. Our workshop strives to extend traditional agent frameworks of next-action prediction and reinforcement learning by emphasizing critical elements such as the integration of natural language, dynamic multimodal interactions, and interactive feedback with humans and the environment. In an effort to further advance research on agent-based multimodal intelligence, we propose our new workshop on multimodal agents. Our workshop is designed to create a platform for publishing research on the multifaceted aspects of agentic, multimodal interactions.

Topic: **Responsible AI**
Workshop: **FOundation models Creators meet USers (FOCUS)**
Organizers: Antonio Alliegro, Francesca Pistilli, Songyou Peng, Biplab Banerjee, Gabriela Csurka, Giuseppe Averta
Website https://focus-workshop.github.io/
Room: Suite 2
Time: 14:00 - 18:00
Summary: Over the last few years, the field of Artificial Intelligence has witnessed significant growth, largely fueled by the development of large-scale machine learning models. The foundational models are characterized by extensive training on diverse datasets that encompass various input modalities (e.g. images, text, audio, 3D data), showing excellent flexibility and effectiveness across a
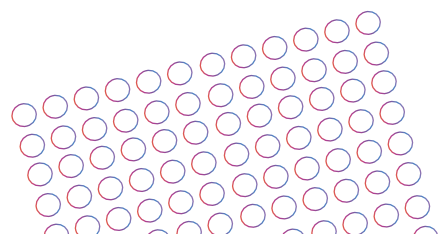
wide range of standard NLP and Computer Vision tasks. Such general-purpose solutions often reveal potentials that go beyond what their creators originally envisioned, motivating users to adopt these models for a broad spectrum of applications. Nevertheless, the knowledge embedded in these models may not be enough when the final goal goes beyond perception benchmarks. These considerations spark important questions that can only be answered with a collaborative dialogue between researchers developing these models (creators) and those employing them in downstream tasks (users). Each group brings a unique perspective that will be crucial in shaping the future of this technology. The goal of this workshop is to identify and discuss strategies to assess both positive and negative (possibly unexpected) behaviors in the development and use of foundation models.

Particular attention will be given to applications that diverge significantly from the scenarios encountered during the training phase of foundational models. These include application-specific visual understanding, uncertainty evaluation, goal-conditioned reasoning, human habits learning, task and motion planning, scene navigation, vision-based manipulation, etc. Our purpose is to foster an open discussion between foundation model creators and users, targeting the analysis of the most pressing open questions for the two communities and fostering new fruitful collaborations.

Topic:	**Responsible AI**
Workshop:	**Women In Computer Vision**
Organizers:	Mamatha Thota, Patsorn Sangkloy, Nirat Saini, Naina Dhingra, Evin Pınar Ornek, Elena Izzo, Xi Wang, Yujiao Shi
Website	https://sites.google.com/view/wicveccv2024/home
Room:	Panorama Lounge
Time:	14:00 - 18:00
Summary:	Women in Computer Vision (WiCV) workshop at ECCV warmly invites researchers of all genders to participate in and support female-identifying researchers within the field. This workshop features poster sessions, invited talks, short oral presentations, and panel discussions addressing career challenges, success strategies, and more. Additionally, it provides valuable networking opportunities, sponsor interactions, and a mentoring dinner designed for junior researchers. WiCV is dedicated to enhancing the visibility of women in computer vision, with a strong commitment to inclusivity, particularly for individuals from underrepresented backgrounds.

Topic:	**Responsible AI**
Workshop:	**Workshop on Green Foundation Models**
Organizers:	Yiming Wang, Subhankar Roy, Massimiliano Mancini, Davide Talon, Kaiyang Zhou, Enzo Tartaglione, Aishwarya Agrawal, Girmaw Abebe Tadesse, Marco Cristani, Zeynep Akata
Website	https://green-fomo.github.io/ECCV2024/index.html
Room:	Suite 9
Time:	14:00 - 18:00
Summary:	Foundation models generalize well to various downstream tasks, thanks to their web-scale pre-training, and have become a de-facto tool in pushing the frontiers of computer vision research. Despite the exciting progress, developing foundation models requires large compute resources, incurring heavy environmental costs. For instance, CLIP reports the use of hundreds of GPUs and LLaMA 2 family led to an emission of 539 tCO2eq, requiring about 27,000

trees in one year to capture the emissions.

To reduce computation and improve data efficiency, the computer vision community has explored efficient ways of adapting large foundation models for downstream tasks. For example, recent methods tune only a small set of learnable tokens (or prompts) while keeping the weights of the pre-trained model frozen. Compression methods also play a crucial role in optimizing the efficiency of foundation models by pruning or distillation. Very recent methods have pioneered in a new training-free paradigm by leveraging existing foundation models or an assembly of them, to address computer vision tasks such as image classification, 3D scene understanding, visual reasoning without fine-tuning. Such research advancement is exciting and encouraging as it enables faster adaptation to downstream tasks without excessive computation, while mitigating the carbon footprint associated with resource-intensive processes.

The Green FOundation MOdels (GreenFOMO) workshop aims to accelerate momentum around these emerging research topics, foster an inclusive research and innovation ecosystem involving small/medium sized practitioners in both academia and industry, and collectively making a green impact to society. GreenFOMO promotes novel methodologies for efficient exploitation of foundation models and encourage applications of FOMOs in domains that induce green impacts, such as biodiversity, agricultural, food security, among others.

|	
Topic:	**Sensing Devices**
Workshop:	**GigaVision: When Gigapixel Videography Meets Computer Vision**
Organizers:	Lu Fang, David Brady, Haozhe Lin, Feng Yang, Sebastiano Battiato, Mario Valerio Giuffrida, Mattia Litrico
	Website https://gigavision.cn/data/news?nav=ECCV%202024%20Workshop&type=nav&t=1715068056315
Room:	Suite 3
Time:	14:00 - 18:00
Summary:	In the past decades, digital cameras have witnessed the emergence of smart imaging solutions driven by AI and CV algorithms. While the development of deep learning has radically improved the capacity of computational imaging, how to explore the benefit of high-performance imaging to inspire and promote more capable computer vision solutions has not been investigated extensively. E.g., although the array cameras can realize gigapixel videography for large-scale scenes full of informative details, existing computer vision tasks remain to investigate the small-scale, low-resolution, and less-informative visual data. Given the popularity of high-performance imaging even with mobile phones, what are the opportunities and challenges for computer vision research? How can computer vision algorithms further help with the super camera to achieve REAL smart? They are open questions demanding prompt solutions. GigaVision Workshop aims to bring the attention of both the computer vision and computational imaging communities, especially toward the mutually unified solutions for joint perception, reconstruction, and understanding. In particular, the GigaVision workshop 2024 plans to discuss the interaction between 3D vision and computational photography, e.g., when the sensing reaches the gigapixel-level, what are the opportunities and challenges for 3D reconstruction, rendering, and understanding? The invited talks will cover novel 3D representation, large-scale 3D reconstruction, sparse reconstruction, 3D instance segmentation, etc. Meanwhile, GigaTracking and GigaRendering challenges will be held to call for attention and solutions for GigaVision. We will also select the best GigaVision-related publication of the year.
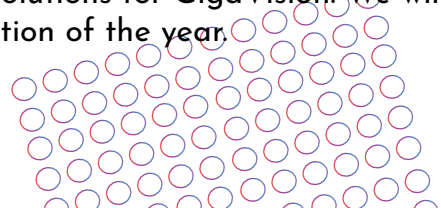
DIAMOND SPONSORS

BENDING SPOONS

Google Research

HPC-AI.COM
PREMIUM GPU, ANYTIME

PLATINUM SPONSORS

Lightning AI
Creators of PyTorch Lightning

Meta

GOLD SPONSORS

ANT RESEARCH

Baidu 百度

ByteDance

HUAWEI

iit ISTITUTO ITALIANO DI TECNOLOGIA

SILVER SPONSORS

TENCENT ROBOTICS X

Weights & Biases

BRONZE SPONSORS

CHEN TIANQIAO & CHRISSY INSTITUTE

INSAIT | Institute for Computer Science, Artificial Intelligence and Technology

EXHIBITORS

3dMD

Advex

covision media

encord

evs

IO INDUSTRIES

Labelbox

LatticeFlow

LIGHTLY

MDPI Academic Open Access Publishing since 1996

meshcapade

MOVE 4D

ONESOURCE CLOUD

parallel domain

SuperAnnotate

THREE LINES OF CODE

VOXEL51

STARTUP EXHIBITORS

abaka·ai

CL CINEMERSIVELABS

ITALAI

KEYLABS

Living Optics

TENYKS

IET The Institution of Engineering and Technology

人工智能与数据科学学院

Visual Layer