ICCV HONOLULU HAWAII

OCT 19-23, 2025

International Conference
on Computer Vision



PROGRAM - WORKSHOPS & TUTORIALS

ULTIMATE SPONSORS

Google

Meta

Tencent

PLATINUM SPONSORS

amazon

GOLD SPONSORS

Adobe

ANT RESEARCH

MINIMAX

Baidu 百度

ByteDance

Figma

ORACLE

SILVER SPONSORS

abaka·ai

Applied Intuition

ENCORD

kitware

SONY

STACK

WAYMO

ZOOX

MESSAGE FROM THE GENERAL AND PROGRAM CHAIRS

Welcome to the 2025 IEEE/CVF International Conference on Computer Vision (ICCV) in Honolulu, Hawai'i! ICCV is the premier biennial conference in computer vision, gathering researchers and practitioners from academia, industry, and government to present and discuss the latest advances in computer vision. Our program features invited keynotes, oral and poster presentations, workshops, tutorials, demonstrations, exhibitions, and numerous networking opportunities. ICCV 2025 is primarily an in-person conference, but a virtual component will also be available, providing access to conference papers, posters, videos, and talks for those unable to attend physically.

ICCV 2025 received 11,239 valid paper submissions, reflecting the community's continued growth and engagement. The review process was overseen by six program chairs, 510 area chairs, and 11,859 reviewers. Each submission underwent a rigorous evaluation, including at least three independent reviews and an author rebuttal period, culminating in final recommendations by the area chairs. A total of 2,701 papers were accepted, resulting in an acceptance rate of 24%, a figure that was not pre-determined and is consistent with historical ICCV rates. All accepted papers will be presented in one of six poster sessions, with 64 selected for oral presentations and 263 designated as highlights in recognition of their exceptional quality and potential impact. A subset of accepted papers has been named award candidates; the best papers, as selected by a committee of distinguished community members, will be announced during the conference's opening session.

For ICCV 2025, we introduced several major changes, some of which have been adopted at other computer vision and AI conferences. To partially address the explosive growth of submissions in the field and the impact on the reviewing process, paper submissions per author were capped at 25. In addition, every qualified author (unless serving in another organizational capacity) was required to serve as a reviewer. Reviewers who failed to submit reviews by the deadline, or whose reviews were flagged as highly irresponsible, faced desk rejection of all papers on which they were authors. These changes were meant to improve accountability, reduce review delays, and preserve the high review standards of the computer vision community.

We extend our deepest appreciation to everyone who has contributed to ICCV 2025. Special thanks go to our area chairs, reviewers, authors, demo and tutorial participants, and our generous sponsors and exhibitors. We also gratefully acknowledge Nicole Finn and the C to C Events team for conference logistics, as well as all the committee members whose tireless efforts have made this event possible.

Thank you for joining us for ICCV 2025 in beautiful Honolulu. We encourage you to meaningfully engage with our vibrant community and also explore the scenic beauty and rich culture of Hawai'i.

Enjoy ICCV 2025!

ICCV 2025 General Chairs
Hilde Kuehne, *University of Tübingen*
Gerard Medioni, *Amazon*
Dimitris Samaras, *Stony Brook University*
Jingyi Yu, *ShanghaiTech University*
Ramin Zabih, *Cornell University*

ICCV 2025 Program Chairs
Bohyung Han, *Seoul National University*
Aniruddha Kembhavi, *Wayve*
Richard Souvenir, *Temple University*
Deqing Sun, *Google DeepMind*
Ayellet Tal, *Technion*
Jingdong Wang, *Baidu*

ICCV 2025 QR CODES

ICCV 2025 Virtual Platform

Access schedules, papers, workshops, tutorials, etc

Exhibit Hall Floorplan

Access location of exhibitors on the main floor

General Chairs
Hilde Kuehne, *University of Tuebingen*
Gerard Medioni, *Amazon*
Dimitris Samaras, *Stony Brook University*
Jingyi Yu, *ShanghaiTech University*
Ramin Zabih, *Cornell*

Program Chairs
Bohyung Han, *Seoul National University*
Aniruddha Kembhavi, *Wayve*
Richard Souvenir, *Temple University*
Deqing Sun, *Google DeepMind*
Ayellet Tal, *Technion*
Jingdong Wang, *Baidu*

Advisor to the Program Committee
David Forsyth, *University of Illinois at Urbana-Champaign*

Tutorial Chairs
Lijie Fan, *Massachusetts Institute of Technology*
Tejas Gokhale, *University of Maryland, Baltimore County*
Jun-Yan Zhu, *Carnegie Mellon University*

Workshop Chairs
Vitor Albiero, *Meta*
Boqing Gong, *Boston University*
Seunghoon Hong, *KAIST*
James Tompkin, *Brown University*
Maria Vakalopoulou, *CentraleSupelec*

Demonstration Chairs
Huaizu Jiang, *Northeastern University*
Luca Weihs, *Allen Institute for Artificial Intelligence*

Broadening Participation Chairs
Paola Cascante-Bonilla, *Stony Brook University*
Mahdi Hosseini, *Concordia University*
Vicky Kalogeiton, *Ecole polytechnique, IP Paris*
Lingjie Liu, *Saarland Informatics Campus, Max-Planck Institute*
Si Liu, *Beihang University*
Yael Moses, *Interdisciplinary Center*
Tamar Rott Shaham, *MIT*
Saining Xie, *Courant Institute of Mathematical Sciences*

Technical Chair
Yoshitomo Matsubara, *Yahoo!*

Publications Coordinator
Patrick Kellenberger, *IEEE Computer Society*

Ombud
David Forsyth, *University of Illinois at Urbana-Champaign*

Finance Chair
Walter Scheirer, *University of Notre Dame*

Doctoral Consortium Chairs
Tanmay Gupta, *Allen Institute for Artificial Intelligence*
Anna Kukleva, *MPII*

Local Chair
Huaijin Chen, *University of Hawaii*

Publicity Chairs
Kosta Derpanis, *York University*
Antonio Furnari, *University of Catania*
Nataniel Ruiz, *Google*
Siyu Zhu, *Fudan University*

Web Developers
Lee Campbell, *ICCV Staff*
Tony Manzo, *ICCV Staff*

Conference Producer
Nicole Finn, *c to c events*

MESSAGE FROM THE WORKSHOP CHAIRS

From all the chairs, welcome to the workshops program of ICCV 2025! Workshops are a key part of the computer vision community, showing the breadth, vibrancy, and impact of our expanding field. Workshops also show the ambitious efforts of the organizers and participants, who create great speaker lineups, run and compete in complex competitions, and produce challenge reports and papers that document the cutting edge of our community's progress. We express our gratitude to the many organizers, program committees, speakers, authors, and all participants for their contributions to our diverse and interesting workshop program. To make it easier to navigate all this, we have grouped the workshops into thematic "tracks" that are closely related, and we've tried our best to avoid scheduling conflicts between them. We hope you enjoy the workshops!

Workshop Chairs

Vitor Albiero, Boqing Gong, Seunghoon Hong, James Tompkin, Maria Vakalopoulou

TABLE OF CONTENTS

Tracks and Workshops	Date	Time	Location	Page
Community and Data				
The 4th DataCV Workshop and Challenge	19-Oct	8:00 - 12:00	304 A	14
Computer Vision for Developing Countries	19-Oct	8:00 - 17:00	302 A	16
Women in Computer Vision	19-Oct	13:00 - 17:00	301 B	23
Findings of the ICCV	20-Oct	8:45 - 17:00	310	32
Perception and Cognition				
Memory and Vision	19-Oct	8:00 - 12:00	304 B	13
The Third Perception Test Challenge	19-Oct	8:00 - 17:00	Ballroom B	19
Human-inspired Computer Vision	20-Oct	9:00 - 12:00	309	29
Low-level and Physics-based Imaging, Cameras, Materials, Light				
Workshop on Computer Vision with Single-Photon Cameras	19-Oct	8:15 - 12:00	301 B	15
Multispectral Imaging for Robotics and Automation	19-Oct	9:00 - 17:00	310	18
Visual Quality Assessment Competition	19-Oct	13:00 - 17:00	303 B	23
Transparent & Reflective objects In the wild Challenges	19-Oct	13:00 - 17:00	302 B	22
Mobile Intelligent Photography and Imaging	20-Oct	8:00 - 12:00	327	27
Workshop on Neuromorphic Vision (NeVi): Advantages and Applications of Event Cameras	20-Oct	8:00 - 12:00	303 B	29
Computer Vision for Materials Science	20-Oct	8:00 - 12:00	305 A	26
Advances in Image Manipulation Workshop and Challenges	20-Oct	8:00 - 17:00	311	30
Responsible Imaging	20-Oct	13:00 - 17:00	305 A	36
UniLight: Unifying Evaluation Metrics for Image-based Lighting; Relighting and Compositing	20-Oct	13:00 - 17:00	318 A	36
Pose, Localization, Mapping, 3D Space				
Camera Calibration and Pose Estimation	19-Oct	8:00 - 12:00	301 A	12
Multi-modal Localization and Mapping	19-Oct	8:00 - 17:00	311	18
10th International Workshop on Recovering 6D Object Pose	19-Oct	13:00 - 17:00	313 A	20
Neural SLAM Workshop	19-Oct	13:00 - 17:00	304 A	21
From street to space: 3D Vision AcrosS altiTudes	20-Oct	8:00 - 12:00	313 B	26
Workshop & Competition on Computationally Optimal Gaussian Splatting	20-Oct	8:30 - 12:00	325 B	29
Large Scale Cross Device Localization	20-Oct	13:00 - 17:00	301 B	36

Scene Understanding, Spatial Reasoning, and World Models

Instance-Level Recognition and Generation Workshop ... 19-Oct ... 8:30 - 12:00 ... 306 A ... 13
Binocular Egocentric-360 Multi-modal Scene Understanding in the Wild ... 19-Oct ... 8:00 - 12:00 ... 306 B ... 12
Learning to See: Advancing Spatial Understanding for Embodied Intelligence ... 19-Oct ... 8:00 - 17:00 ... 312 ... 18
5th Workshop and Challenge on Open-World 3D Scene Understanding ... 19-Oct ... 13:30 - 17:00 ... 306 B ... 20
End-to-End 3D Learning ... 19-Oct ... 13:00 - 17:00 ... 304 B ... 21
Embodied Spatial Reasoning ... 20-Oct ... 8:00 - 12:00 ... Ballroom A ... 26
Robust and Interactable World Models in Computer Vision ... 20-Oct ... 8:00 - 17:00 ... 316 A ... 32
Multimodal Spatial Intelligence ... 20-Oct ... 13:00 - 17:00 ... 306 A ... 36

Video: Tracking, Sementation, and Understanding Long and Short

Workshop on Benchmarking Multi-Target Tracking: Towards Spatiotemporal Action Grounding in Videos ... 19-Oct ... 8:00 - 12:00 ... 313 B ... 15
Story-Level Movie Understanding and Audio Description ... 19-Oct ... 8:50 - 12:30 ... 327 ... 13
Short-Form Video Understanding: The Next Frontier in Video Intelligence ... 19-Oct ... 13:00 - 17:00 ... 313 B ... 22
Large-scale Video Object Segmentation ... 20-Oct ... 8:00 - 12:00 ... 318 A ... 27
Visual object tracking and segmentation challenge workshop ... 20-Oct ... 13:00 - 17:00 ... 302 A ... 37
1st Workshop on Long Multi-Scene Video Foundations: Generation; Understanding and Evaluation ... 20-Oct ... 13:00 - 17:00 ... 303 B ... 34

Reasoning, Representations, and Learning

The 1st International Workshop and Challenge on Disentangled Representation Learning for Real-world Applications ... 19-Oct ... 8:00 - 12:00 ... 307 B ... 14
2nd Beyond Euclidean Workshop: Hyperbolic and Hyperspherical Learning for Computer Vision ... 19-Oct ... 8:00 - 17:00 ... 305 B ... 15
Structural Priors for Vision ... 19-Oct ... 8:00 - 17:00 ... 323 A ... 19
Workshop on Knowledge-Intensive Multimodal Reasoning ... 19-Oct ... 13:00 - 17:00 ... 313 C ... 23
Multimodal Continual Learning ... 19-Oct ... 13:00 - 17:00 ... 301 A ... 21
Multi-Modal Reasoning for Agentic Intelligence ... 20-Oct ... 8:00 - 17:00 ... 301 A ... 32
Multimodal reasoning and slow thinking in large model era: towards system2 and beyond ... 20-Oct ... 8:00 - 17:00 ... Ballroom B ... 32
6th Workshop on Continual Learning in Computer Vision ... 20-Oct ... 8:00 - 17:00 ... 304 B ... 30
Workshop on Scene Graphs and Graph Representation Learning ... 20-Oct ... 13:30 - 17:30 ... 314 ... 37

Vision, Language, and Multimodal

Workshop on Safe and Trustworthy Multimodal AI Systems ... 19-Oct ... 8:00 - 17:00 ... 308 B ... 19
The Second Workshop on Multimodal Representation and Retrieval ... 20-Oct ... 8:30 - 12:30 ... 308 B ... 28
What is Next in Multimodal Foundation Models? ... 20-Oct ... 8:00 - 12:00 ... 326 B ... 28
Closing the Loop Between Vision and Language (Decade Mark) ... 20-Oct ... 8:00 - 17:00 ... 313 C ... 31
Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities ... 20-Oct ... 8:00 - 17:00 ... 318 B ... 33

Generation

Generative AI for Audio-Visual Content Creation ... 19-Oct ... 8:00 - 12:00 ... 323 B ... 13
Human-Interactive Generation and Editing ... 19-Oct ... 9:00 - 18:00 ... 309 ... 17
Geometry-Free Novel View Synthesis and Controllable Video Models ... 19-Oct ... 8:00 - 17:00 ... 323 C ... 17
Generative Scene Completion for Immersive Worlds ... 20-Oct ... 8:00 - 12:00 ... 301 B ... 27
The Third Workshop on AI for 3D Content Creation ... 20-Oct ... 8:00 - 12:00 ... 315 ... 28
2nd AI for Content Generation; Quality Enhancement and Streaming ... 20-Oct ... 8:00 - 17:00 ... 317 B ... 30
Generating Digital Twins from Images and Videos ... 20-Oct ... 8:00 - 17:00 ... 319 A ... 32
Wild3D: 3D Modeling; Reconstruction; and Generation in the Wild ... 20-Oct ... 8:00 - 17:00 ... 312 ... 33
Personalization in Generative AI Workshop ... 20-Oct ... 8:00 - 17:00 ... 317 A ... 34
2nd Workshop on Audio-Visual Generation and Learning ... 20-Oct ... 13:00 - 17:00 ... 327 ... 34
2nd Workshop on Scalable 3D Generation and 3D Geometric Scene Understanding ... 20-Oct ... 13:00 - 17:00 ... 325 B ... 34

Efficient CV: Power, Data, Compute, Learning

Embedded Vision Workshop ... 19-Oct ... 8:00 - 17:00 ... 316 A ... 17
Representation Learning with Very Limited Resources: When Data; Modalities; Labels; and Computing Resources are Scarce ... 19-Oct ... 13:00 - 17:00 ... 306 A ... 22
The 2nd Workshop on Efficient Computing under Limited Resources: Visual Computing ... 19-Oct ... 13:00 - 17:00 ... 327 ... 22
The 3rd workshop on Binary and Extreme Quantization for Computer Vision ... 20-Oct ... 8:00 - 12:00 ... 308 A ... 27
Workshop on Curated Data for Efficient Learning ... 20-Oct ... 8:00 - 17:00 ... 304 A ... 34

Trust, Explainabilty, Fairness

Systematic Trust in AI Models: Ensuring Fairness; Reliability; Explainability; and Accountability in Machine Learning Frameworks ... 19-Oct ... 9:00 - 12:00 ... 325 B ... 13
2nd Workshop on Explainable Computer Vision: Quo Vadis? ... 19-Oct ... 8:00 - 17:00 ... Ballroom A ... 15
2nd Workshop and Challenge on Unlearning and Model Editing ... 19-Oct ... 13:00 - 18:00 ... 308 A ... 20
TrustFM: Workshop on Trustworthy Foundation Models ... 20-Oct ... 13:00 - 17:00 ... 308 B ... 38
Critical Evaluation of Generative Models and their Impact on Society ... 20-Oct ... 13:00 - 17:00 ... 309 ... 37
Fairness and Ethics in AI: facing the ChalLEnge through Model Debiasing ... 20-Oct ... 8:00 - 12:00 ... 302 B ... 38

Authenticity, Biometrics, and Forgeries

The Challenge of Detecting Synthetic Manipulations in ID Documents ... 19-Oct ... 8:50 - 12:10 ... 303 B ... 14
Authenticity & Provenance in the age of Generative AI ... 19-Oct ... 8:00 - 17:00 ... 315 ... 16
The 6th Face Anti-Spoofing Workshop: Unified Physical-Digital Attacks Detection ... 19-Oct ... 13:00 - 17:00 ... 323 B ... 22
Foundation & Generative Models in Biometrics ... 20-Oct ... 8:00 - 12:00 ... 305 B ... 26
Computer Vision for Biometrics; Identity & Behavior Science ... 20-Oct ... 8:00 - 17:00 ... 307 B ... 31

Humans, Bodies, Hands, Faces, Gestures, and Social

Affective & Behavior Analysis in-the-wild ... 19-Oct ... 8:00 - 12:00 ... 313 C ... 12
Interactive Human-centric Foundation Models ... 19-Oct ... 8:00 - 17:00 ... 317 A ... 18
Artificial Social Intelligence Workshop ... 19-Oct ... 8:00 - 17:00 ... 317 B ... 16
First Workshop on Skilled Activity Understanding; Assessment and Feedback Generation ... 19-Oct ... 14:00 - 17:00 ... 318 A ... 21
The 12th IEEE International Workshop on Analysis and Modeling of Faces and Gestures ... 20-Oct ... 8:00 - 17:00 ... 320 ... 33
1st Workshop on Multimodal Sign Language Recognition ... 20-Oct ... 8:00 - 17:00 ... 319 B ... 29
Egocentric Body Motion Tracking; Synthesis and Action Recognition ... 20-Oct ... 13:00 - 17:00 ... Ballroom C ... 35
International Workshop on Observing and Understanding Hands in Action ... 20-Oct ... 13:00 - 17:00 ... 305 B ... 35

Art, Fashion, Design, and Documents

The 2nd AI for Visual Arts Workshop and Challenges ... 19-Oct ... 9:00 - 12:00 ... 313 A ... 14
Computer Vision for Fashion; Art; and Design: Bridging Creativity and Responsible AI ... 19-Oct ... 8:00 - 12:00 ... 316 B ... 12
Comic Intelligence Quotient: Advances and Challenges in AI-driven Comic Analysis ... 19-Oct ... 13:00 - 17:00 ... 305 A ... 20
Workshop on Graphic Design Understanding and Generation ... 19-Oct ... 13:00 - 17:00 ... 328 ... 23
Generative AI for Storytelling ... 20-Oct ... 8:00 - 12:00 ... 302 A ... 27
Workshop on Cultural Continuity of Artists: Leveraging Artistic Legacies for AI-Driven Cultural Heritage ... 20-Oct ... 8:00 - 12:00 ... 307 A ... 28
Biometrics for Art ... 20-Oct ... 13:00 - 17:00 ... 307 A ... 35
Workshop on Computer Vision Systems for Document Analysis and Recognition ... 20-Oct ... 13:00 - 17:00 ... 308 A ... 37

Health and biomedical

The Eighth International Workshop on Computer Vision for Physiological Measurement (CVPM)	19-Oct	8:00 - 12:00	328	14
Vision-based AI for Digital Health: From Pixels to Practics	19-Oct	8:00 - 17:00	303 A	19
3rd Workshop on Computer Vision for Automated Medical Diagnosis	19-Oct	8:00 - 17:00	314	16
Multi-Modal Foundation Models for Cancer Detection and Prevention	19-Oct	13:00 - 17:00	326 B	21
Workshop on Advanced Perception for Autonomous Healthcare	20-Oct	8:00 - 12:30	306 A	28
Generative AI for Biomedical Image Analysis: Opportunities; Challenges and Futures	20-Oct	8:00 - 12:00	316 B	27
BioImage Computing	20-Oct	9:00 - 17:00	325 A	31
Vision-Language Modeling in 3D Medical Imaging	20-Oct	8:00 - 17:00	323 A	33

Health and biomedical - Fairness and Explainability

Workshop on Biomedical Image and Signal Computing for Unbiasedness; Interpretability; and Trustworthiness	19-Oct	8:00 - 12:00	302 B	15
PHAROS - Adaptation; Fairness; Explainability in AI Medical Imaging	20-Oct	13:00 - 17:00	Ballroom A	36

Ecology and Marined

Joint Workshop on Marine Vision	19-Oct	9:00 - 18:00	318 B	18
SEA: 1st workshop on Sustainability with Earth observation and AI	19-Oct	13:00 - 17:00	307 A	22
2nd Workshop on Computer Vision for Ecology	20-Oct	8:00 - 17:00	323 B	30
Computer Vision in Plant Phenotyping and Agriculture	20-Oct	13:00 - 17:00	316 B	35

Autonomous Driving, Mobility, and Cities

Driving Simulation from Real-World Data: How Well Can We Render and Drive?	19-Oct	8:00 - 12:00	305 A	12
2nd Workshop on the Challenge Of Out-of-Label Hazard Detection in Autonomous Driving	19-Oct	8:00 - 17:00	319 A	16
Foundations Models for V2X-Based Cooperative Autonomous Driving	19-Oct	8:00 - 17:00	Ballroom C	17
Workshop on Distillation of Foundation Models for Autonomous Driving	20-Oct	8:00 - 12:00	316 C	28
Ego-Exo Sensing for Smart Mobility	20-Oct	8:00 - 17:00	323 C	31
9th AI City Challenge	20-Oct	8:00 - 17:00	313 A	30

Robotics

13th International Workshop on Assistive Computer Vision and Robotics	19-Oct	13:00 - 17:00	316 C	20
1st Workshop and Challenge on Category-Level Object Pose Estimation for Robotic Manipulation	20-Oct	8:00 - 17:00	324	29
Human-Robot-Scene Interaction and Collaboration	20-Oct	13:00 - 17:00	315	35

Industrial and Retail

Foundation Data for Industrial Tech Transfer	19-Oct	8:00 - 12:00	316 C	12
3rd Workshop on Vision-based InduStrial InspectiON	19-Oct	8:00 - 17:00	319 B	16
RetailVision6 - Revolutionizing the World of Retail	19-Oct	8:45 - 17:00	325 A	19
Anomaly Detection with Foundation Models	20-Oct	8:25 - 12:15	314	26
Computer Vision in Advertising and Marketing	20-Oct	8:00 - 17:00	303 A	31

Saturday, October 18

13:00 - 20:00 **Registration / Badge Pickup**

Sunday, October 19

NOTE: Tutorial rooms are subject to change. Refer to the online site for up-to-date locations. Use the QR code for each tutorial to see its schedule. Here is the QR code for the ICCV 2025 Tutorials page.

7:00 - 17:00 **Registration/Badge Pickup** (HCC Main Lobby)

7:00 - 17:00 **Nursing Pod**
(Located on lobby level by Exhibit Hall I)

7:00 - 17:00 **Prayer / Quiet Room / Mother's Room** (326 A)

7:00 - 9:00 **Breakfast** (Exhibit Hall II)

8:00 - 18:00 **Workshop & Tutorials**

10:00 - 11:00 **Coffee Break & Posters** (Exhibit Hall II)

12:00 - 13:45 **Lunch** (Exhibit Hall II / Overflow Seating Rooftop)

15:00 - 16:00 **Coffee Break & Posters** (Exhibit Hall II)

8:00 - 17:00 **POSTERS** (Exhibit Hall II)

TUTORIALS

A Tour Through AI-powered Photography and Imaging

Organizer: Marcos Conde
Date: 10/19
Time: 8:00 - 11:00
Location 326 B

Summary: Computational Photography and low-level vision are pivotal research areas within Computer Vision, significantly impacting both academia and industry. Despite their importance, progress in these fields often lags behind areas like generative AI, primarily due to the scarcity of standardized datasets, clear benchmarks, and limited transparency from camera manufacturers. This tutorial bridges the gap between academic research and industry applications by providing an in-depth, hands-on exploration of computational photography and imaging using deep learning. Collaboratively presented by leading academic researchers and prominent industry experts from Sony, this tutorial systematically covers learned Image Signal Processors (ISPs), cutting-edge transformer and convolutional neural network architectures for image restoration and enhancement, and the development of realistic synthetic data generation pipelines. Attendees will acquire practical skills in dataset creation, realistic pipeline simulation, and evaluation protocols, empowering them with the tools and insights needed to accelerate innovation in this field.

Beyond Self-Driving: Exploring Three Levels of Driving Automation

Organizer: Zhiyu Huang
Date: 10/19
Time: 8:00 - 11:00
Location 308 A

Summary: Self-driving technologies have demonstrated significant potential to transform human mobility. However, single-agent systems face inherent limitations in perception and decision-making capabilities. Transitioning from self-driving vehicles to cooperative multi-vehicle systems and large-scale intelligent transportation systems is essential to enable safer and more efficient mobility. Realizing such sophisticated mobility systems introduces significant challenges, requiring comprehensive tools and models, simulation environments, real-world datasets, and deployment frameworks. This tutorial will delve into key areas of driving automation, beginning with advanced end-to-end self-driving techniques such as vision-language-action (VLA) models, interactive prediction and planning, and scenario generation. The tutorial emphasizes V2X communication and cooperative perception in real-world settings, as well as datasets including V2X-Real and V2XPnP. It also covers simulation and deployment frameworks for urban mobility, such as MetaDrive, MetaUrban, and UrbanSim. By bridging foundational research with real-world deployment, this tutorial offers practical insights into developing future-ready autonomous mobility systems.

Towards Comprehensive Reasoning in Vision-Language Models

Organizer: Yujun Cai
Date: 10/19
Time: 8:00 - 12:00
Location 318 A

Summary: Vision-Language Models (VLMs) have achieved remarkable progress in image captioning and visual question answering, yet developing genuine reasoning capabilities remains an open challenge. Unlike recent breakthroughs in reasoning-focused LLMs, many VLMs still rely primarily on pattern recognition and struggle with compositional logic. This tutorial provides a comprehensive overview of reasoning capabilities in VLMs, focusing on the transition from basic perception to complex inference. We will explore reasoning-oriented prompting and training techniques in multimodal contexts, reasoning-focused benchmarks, and architectural innovations for visual-textual fusion. Through lectures and hands-on demonstrations, participants will gain insights into current capabilities, persistent challenges in compositional generalization and explainability, and practical guidance for implementing reasoning mechanisms. This tutorial uniquely bridges advances in LLM reasoning with the visual domain, addressing the distinct challenges of spatial information processing and providing a roadmap toward more cognitively capable vision-language systems.

Foundation Models in Visual Anomaly Detection: Advances, Challenges, and Applications

Organizer: Jiawen Zhu
Date: 10/19
Time: 9:00 - 12:00
Location 307 A

Summary: In recent years, foundation models have emerged as transformative tools in computer vision, offering powerful zero-shot and few-shot learning capabilities across a wide range of tasks. Their integration into visual anomaly detection—a critical and high-stakes field spanning healthcare, industrial inspection, security, and autonomous systems—has opened new frontiers in both research and real-world applications. This tutorial aims to deliver a comprehensive and timely overview of the role of foundation models in visual anomaly detection. We will cover multiple visual modalities, including 2D images, 3D images, and videos—each presenting unique challenges and necessitating modality-specific solutions. Specifically, we will delve into the entire pipeline, from data (pre-)training and prompt engineering to methodological innovations, inference strategies, and deployment in real-world environments. Key topics include zero- and few-shot learning, pseudo-labeling, anomaly generation, and multi-modal alignment between vision and language. To facilitate a deep and practical understanding of these areas, the tutorial will bring together leading experts from both academia and industry. Through in-depth technical presentations and discussions, participants will gain valuable insights into the latest advances, real-world applications, and open challenges shaping this rapidly evolving field.

3D Human Motion Generation and Simulation

Organizer: Huaizu Jiang
Date: 10/19
Time: 8:00 - 16:00
Location 320

Summary: 3D human motion generation and simulation is an important area of research with applications in virtual reality, gaming, animation, robotics, and AI-driven content creation. Generating realistic and controllable human motion is essential for creating interactive digital environments, improving character animation, and enhancing human-computer interaction. Recent advances in deep learning have made it possible to automate motion generation, reducing the need for expensive motion capture and manual animation. Techniques such as diffusion models, generative masking, and variational autoencoders (VAEs) have been used to synthesize diverse and realistic human motion. Transformer-based models have improved the ability to capture temporal dependencies, leading to smoother and more natural movement. In addition, reinforcement learning and physics-based methods have helped create physically consistent and responsive motion, which is useful for applications like robotics and virtual avatars. This tutorial will bridge the gap between computer vision, graphics, and robotics, providing a comprehensive guide to the latest methods, practical applications, and future challenges. This tutorial will be organized into six core parts, guiding you from foundational knowledge to advanced research frontiers: (1) Human Motion Generation Basics: introducing fundamentals, key concepts and data representations; (2) Kinematic-Based Generation Methods: explore popular data-driven techniques that learn from motion capture datasets to produce lifelike animations; (3) Physics-Based Generation Methods: dive into methods that use reinforcement learning and physics simulations to create physically consistent and responsive motion; (4) Controllability of Human Motion Generation: learn how to direct and control motion synthesis using inputs like text, audio, or specific goals; (5) Human-Object/Human/Scene Interactions: cover advanced scenarios involving complex interactions with objects, other people, and the surrounding environment, and (6) Open Research Problems: discussing the major unsolved challenges and exciting opportunities for future work in the field.

Learning Deep Low-Dimensional Models from High-Dimensional Data: From Theory to Practice

Organizer: Qing Qu
Date: 10/19
Time: 9:00 - 17:00
Location 324

Summary: Over the past decade, the advent of deep learning and large-scale computing has immeasurably changed the ways we process, interpret, and predict with data in imaging and computer vision. The ``traditional" approach to algorithm design, based around parametric models for specific structures of signals and measurements---say sparse and low-rank models---and the associated optimization toolkit, is now significantly enriched with data-driven learning-based techniques, where large-scale networks are pre-trained and then adapted to a variety of specific tasks. Nevertheless, the successes of both modern data-driven and classic model-based paradigms rely crucially on correctly identifying the low-dimensional structures present in real-world data, to the extent that we see the roles of learning and compression of data processing algorithms---whether explicit or implicit, as with deep networks---as inextricably linked. As such, this tutorial provides a timely tutorial that uniquely bridges low-dimensional models with deep learning in imaging and vision. This tutorial will show how (i) these low-dimensional models and principles provide a valuable lens for formulating problems and understanding the behavior of modern deep models in imaging and computer vision, and (ii) how ideas from low-dimensional models can provide valuable guidance for designing new parameter efficient, robust, and interpretable deep learning models for computer vision problems in practice. The tutorial will start by introducing fundamental low-dimensional models (e.g., basic sparse and low-rank models) with motivating computer vision applications. Based on these developments, we will discuss strong conceptual, algorithmic, and theoretical connections between low-dimensional structures and deep models, providing new perspectives to understand state-of-the-art deep models in terms of learned representations and generative models. Finally, we will demonstrate that these connections can lead to new principles for designing deep networks and learning low-dimensional structures in computer vision, with both clear interpretability and practical benefits.

Benchmarking Egocentric Visual-Inertial SLAM at City Scale

Organizer: Shaohui Liu
Date: 10/19
Time: 13:00 - 17:00
Location 325 B

Summary: "Simultaneous localization and mapping (SLAM) is a fundamental technique with applications spanning robotics, spatial AI, and autonomous navigation. It addresses two tightly coupled challenges: localizing the device while incrementally building a coherent map of the surroundings. Localization, or positioning, involves estimating a 6 Degrees-of-Freedom (6-DoF) pose for each image in a continuous sequence, typically aided by other sensor data, while mapping involves constructing an evolving representation of the surrounding environment. This tutorial specifically addresses the task of accurate positioning for large-scale egocentric data using visual-inertial SLAM and odometry (VIO). It offers a comprehensive overview of the challenges faced by VIO/SLAM methods on egocentric data and introduces a new dataset and benchmark that can serve as a robust testbed for benchmarking these systems. With the help of well-positioned speakers, this tutorial explores the new benchmarking approach by analyzing failure cases,

identifying limitations, and highlighting open problems in open-source academic VIO/SLAM systems. Additionally, it provides hands-on experience using the dataset and evaluation tools for researchers to get started with their own SLAM evaluations.

Foundation Models for 3D Asset Synthesis

Organizer: Yangguang Li
Date: 10/19
Time: 13:00 - 17:00
Location 307 B

Summary: In recent years, thanks to the continuous innovation and progress of diffusion technology, significant advancements have been made in image and video generation. By inputting textual descriptions or images, we can generate high-quality images or videos, which greatly enhance creative efficiency and imagination. However, progress in the 3D generation field has been relatively slow. Initially, optimization routes, represented by DreamFusion, were explored. This was followed by the exploration of reconstruction routes, such as LRM. It was only later that diffusion based on 3D generation techniques, similar to those in image and video generation, were gradually developed.In addition, based on the token-by-token prediction form similar to LLM, 3D generation based on autoregressive method has gradually made significant progress. Therefore, this tutorial focuses on the topic of 3D asset generation using diffusion and autoregression, specifically including: (1) Geometry generation modeling based on the diffusion paradigm; (2) Geometry generation modeling based on the autoregression paradigm; (3) Texture generation modeling based on the diffusion paradigm.

Responsible Vision-Language Generative Models

Organizer: Changhoon Kim
Date: 10/19
Time: 13:00 - 17:00
Location 316 B

Summary: Vision-language generative models, such as text-to-image and image-to-text systems, have rapidly transitioned from research prototypes to widely deployed tools across domains like education, journalism, and design. However, their real-world adoption has introduced critical challenges surrounding robustness, controllability, and ethical risks—including issues like prompt misalignment, unauthorized content generation, adversarial attacks, and data memorization. This tutorial provides a comprehensive overview of these concerns and emerging solutions by covering recent advances and failure modes in state-of-the-art models, robust concept erasure techniques in diffusion models, and adversarial vulnerabilities and defenses in image-to-text systems. Through a blend of theoretical foundations, participants will examine failure scenarios, explore attack and defense strategies, and gain practical insights into enhancing the trustworthiness of multimodal generative models. Designed for researchers and practitioners in vision, language, and AI safety, this tutorial uniquely focuses on the responsible deployment of these models—bridging technical rigor with societal impact and offering guidance for future research directions in secure and reliable generative AI.

Notes:

WORKSHOPS

Affective & Behavior Analysis in-the-wild

Organizers: Dimitrios Kollias, Stefanos Zafeiriou, Irene Kotsia, Greg Slabaugh
Date: 19-Oct
Time: 8:00 - 12:00
Location: 313 C

Summary: The ABAW Workshop is a premier platform highlighting the latest advancements in multimodal analysis, generation, modeling, and understanding of human behavior in unconstrained environments. It emphasizes cutting-edge systems that integrate facial expressions, body movements, gestures, natural language, voice to enable impactful research and practical applications. The workshop fosters interdisciplinary collaboration across fields (e.g. computer vision, AI, HCI, psychology, robotics, ethics & healthcare) and is a vital forum for building equitable, generalizable & human-centered AI systems. Finally, the Workshop also includes 3 challenges (Valence-Arousal Estimation, Compound Expression Recognition and Fine-Grained Violence Detection).

Binocular Egocentric-360 Multi-modal Scene Understanding in the Wild

Organizers: Jianbo Jiao, Shangzhe Wu, , Dylan Campbell, Yunchao Wei, Lu Qi, Yasmine Mellah, Aleš Leonardis, Chenyuan Qu, Han Hu, Qiming Huang, Hao Chen
Date: 19-Oct
Time: 8:00 - 12:00
Location: 306 B

Summary: This workshop mainly looks at multi-modal scene understanding and perception in a human-like manner. Specifically, we will focus on binocular/stereo egocentric and 360° panoramic perspectives, which measure both first-person views and third-person panoptic views, mimicking a human in the scene, by combining with multi-modal cues such as spatial audio, textual descriptions, and geo-metadata. This workshop will cover but not be limited to the following topics: Embodied 360° scene understanding & egocentric visual reasoning; multi-modal scene understanding; stereo vision; open-world learning & domain adaptation.

Camera Calibration and Pose Estimation

Organizers: Zuzana Kukelova, Gabrielle Flood, Viktor Larsson, Torsten Sattler, Akihiro Sugimoto
Date: 19-Oct
Time: 8:00 - 12:00
Location: 301 A

Summary: This workshop focuses on the closely related problems of camera calibration and pose estimation. These are essential for many advanced 3D computer vision methods, including NeRFs, 3D Gaussian splatting, and scene understanding. The quality of these estimates greatly affects performance, yet many researchers treat them as black boxes. This workshop offers an opportunity for those using calibration and pose algorithms to learn about the latest methods and open challenges. It also provides a forum for researchers working on traditional and learning-based solutions to share ideas, improve methods, and expand the possibilities of 3D vision through better calibration and pose estimation.

Computer Vision for Fashion; Art; and Design: Bridging Creativity and Responsible AI

Organizers: Leonidas Lefakis. Ziad Al-Halah, Negar Rostamzadeh, Thomas Boquet, Julia Lasserre, Loris Bazzani, Ibtihel Amara, Sahar Mbarek, Reza Shirvany
Date: 19-Oct
Time: 8:00 - 12:00
Location: 316 B

Summary: The Computer Vision for Fashion, Art, and Design workshop series aims to to foster interdisciplinary discussions among researchers and practitioners in computer vision and machine learning, as well as artists, designers, sociotechnical researchers, policymakers, social scientists, and other cultural stakeholders. By creating a collaborative space, it aims to address complex challenges that arise at the intersection of generative AI, creativity, and ethics. This year the workshop includes, in addition to multiple invited talks by scientists working in the field, an Art Gallery, and a related Panel discussion.

Driving Simulation from Real-World Data: How Well Can We Render and Drive?

Organizers: Yiyi Liao, Hongyu Zhou, Yichong Lu, Bingbing Liu, Hongbo Zhang, Jiansheng Wei, Ziqian Ni, Yiming Li, Andreas Geiger
Date: 19-Oct
Time: 8:00 - 12:00
Location: 305 A

Summary: This workshop brings together researchers in autonomous driving, computer vision, and graphics to advance the development of real-world data-driven driving simulators, as well as the autonomous driving algorithms in these photorealistic simulation environments. By tackling novel view synthesis and closed-loop autonomy in photorealistic simulations, we aim to push scalable, high-fidelity simulation forward. To promote community engagement and benchmarking, we also host two challenges: extrapolated novel view synthesis for urban scenes and closed-loop evaluation in photorealistic simulators.

Foundation Data for Industrial Tech Transfer

Organizers: Yoshihiro Fukuhara, Hirokatsu Kataoka, Püren Güler, Shunsuke Kitada, Xavier Boix, Dan Hendrycks, Keisuke Tateno, Shinichi Mae, Tatsuya Komatsu, Nishant Rai, Ryo Nakamura, Risa Shinoda, Takahiro Itazuri, Yoshiki Kubotani, Guarin Flück, Wadim Kehl, Kazuki Kozuka, Philipp Wirth
Date: 19-Oct
Time: 8:00 - 12:00
Location: 316 C

Summary: Recently, transformer-based foundation models have excelled across a wide range of recognition and generation benchmarks, yet real industrial impact requires robust tech transfer. Adapting them to heterogeneous industries demands domain-specific fine-tuning, reliable MLOps, and abundant, high-quality data. Conventional IID benchmarks are increasingly saturated, prompting evaluations that probe out-of-distribution and long-tail behavior. Both challenges hinge on curating and exploiting broad-

er, deeper — "Foundation Data." This workshop gathers academia and industry to examine methods for constructing high-quality datasets, refine model-adaptation pipelines, and design novel evaluation tasks grounded in Foundation Data, aiming to unlock new horizons in AI research and application.

Generative AI for Audio-Visual Content Creation

Organizers: Masato Ishii, Takashi Shibuya, Yuki Mitsufuji, Ho Kei Cheng, Alexander Schwing, Prem Seetharaman, Oriol Nieto, Justin Salamon, David Bourgin, Bryan Russell, Ziyang Chen, Sanjoy Chowdhury

Date: 19-Oct

Time: 8:00 - 12:00

Location: 323 B

Summary: Seamless integration of audio and visual elements is crucial for creating immersive and engaging content. Audio-visual generation, involving the synthesis of one modality from the other or both jointly, has become a key research area. This capability holds significant potential for applications like virtual reality, gaming, film production, and interactive media, using advanced generative models to enhance multimedia quality and realism. This workshop highlights the growing importance of audio-visual generation in modern content creation, bringing together researchers and practitioners from academia and industry to explore the latest advances, challenges, and emerging opportunities in this dynamic field.

Instance-Level Recognition and Generation Workshop

Organizers: Andre Araujo, Bingyi Cao, Kaifeng Chen, Ondrej Chum, Noa Garcia, Guangxing Han, Giorgos Kordopatis-Zilos, Giorgos Tolias, Hao Yang, Nikolaos-Antonios Ypsilantis, Xu Zhang

Date: 19-Oct

Time: 8:30 - 12:00

Location: 306 A

Summary: The Instance-Level Recognition and Generation (ILR+G) Workshop focuses on computer vision tasks that operate at instance-level granularity, covering both recognition (ILR) and generation (ILG). Unlike category-level, ILR identifies and compares specific objects, scenes, or events, enabling open-world applications with a vast number of distinct classes. ILG, or personalized generation, aims to create content while preserving the identity of particular instances. This 7th edition explores potential synergies between ILR and ILG. The workshop features keynote talks by renowned speakers, invited papers, and a call for papers, aiming to bring together researchers working on instance-level tasks and inspire new research and collaborations.

Memory and Vision

Organizers: Zexue He, Jovana Kondic, Dmitry Krotov, Rogerio Feris

Date: 19-Oct

Time: 8:00 - 12:00

Location: 304 B

Summary: Memory is a core aspect of human intelligence, and artificial memory systems have recently seen a resurgence through foundational breakthroughs. At the same time, advances in computer vision, especially through generative AI, have enabled models to synthesize realistic imagery and understand complex scenes with remarkable generalization. Despite their shared relevance to cognition, memory and vision have evolved largely as separate fields. MemVis is the dedicated platform organized around the growing need to unify memory and vision, in the development of intelligent AI systems that can process, store, and recall visual information in a more human-like manner.

Story-Level Movie Understanding and Audio Description

Organizers: Junyu Xie, Ridouane Ghermi, Tengda Han, Max Bain, Arsha Nagrani, Gül Varol, Weidi Xie, Vicky Kalogeiton, Ivan Laptev, Andrew Zisserman

Date: 19-Oct

Time: 8:50 - 12:30

Location: 327

Summary: The SLoMO workshop brings together researchers focused on the understanding of long-form, edited videos—such as movies and TV episodes. We spotlight two central research directions: (i) Audio Description (AD) Generation: This track explores the generation of concise and coherent descriptions that complement the original audio for blind and visually impaired (BVI) audiences. We have invited four leading experts in movie understanding and AD generation to share their insights and recent advancements in the field. (ii) Movie Question Answering: This track evaluates models' capabilities in narrative comprehension, emphasizing story-level understanding. As part of this effort, we host the Short-Films 20K (SF20K) Competition, which aims to drive progress in story-level video understanding using the newly introduced SF20K dataset.

Systematic Trust in AI Models: Ensuring Fairness; Reliability; Explainability; and Accountability in Machine Learning Frameworks

Organizers: Arun George Zachariah, Michael Boone, Ryo Hachiuma, Nikki Pope, Shivika Prasanna, Khulud Alsultan

Date: 19-Oct

Time: 9:00 - 12:00

Location: 325 B

Summary: The STREAM Workshop aims to bring together researchers and practitioners working at the intersection of systems design and trustworthy AI. As AI technologies are increasingly deployed in critical domains such as healthcare, finance, and mobility, STREAM focuses on system-level approaches to embedding trustworthiness across the full pipeline, from data collection and architecture design to training, deployment, and evaluation.

The 1st International Workshop and Challenge on Disentangled Representation Learning for Real-world Applications

Organizers: Xin Jin, Qiuyu Chen, Yue Song, Xihui Liu, Shuai Yang, Tao Yang, Ziqiang Li, Jianguo Huang, Yuntao Wei, Ba'ao Xie, Nicu Sebe, Wenjun (Kevin) Zeng

Date: 19-Oct

Time: 8:00 - 12:00

Location: 307 B

Summary: Disentangled Representation Learning (DRL) shows promise for enhancing AI's fundamental understanding of the world, potentially addressing hallucination issues in language models and improving controllability in generative systems. Despite significant academic interest, DRL research remains confined to synthetic scenarios due to a lack of realistic benchmarks and unified evaluation metrics. DRL4Real Workshop aims to bridge this gap by introducing novel, realistic datasets and comprehensive benchmarks for evaluating DRL methods in practical applications. We will focus on key areas including controllable generation and autonomous driving, exploring how DRL can advance model robustness, interpretability, and generalization capabilities.

The 2nd AI for Visual Arts Workshop and Challenges

Organizers: Deblina Bhattacharjee, Bingchen Zhao, Rahul Raja

Date: 19-Oct

Time: 9:00 - 12:00

Location: 313 A

Summary: The AI4VA workshop at ICCV explores the intersection of artificial intelligence and the visual arts, including art, design, exhibitions, photography, and film. It brings together artists, art historians, ethicists, and researchers to foster cross-disciplinary innovation. Topics include generative art, AI for art history, 3D reconstruction from artworks, human pose estimation in art, VQA and captioning for artworks, multimodal interaction, AR/VR for art, and media content analysis. A key aim is fostering participation across diverse creators and researchers. A special focus is AI for Cultural and Artistic Heritage, highlighting advances in analysing, restoring, and interpreting artefacts using multimodal AI across visual, textual, and historical data.

The 4th DataCV Workshop and Challenge

Organizers: Fatemeh Saleh, Liang Zheng, Qiang Qiu, José Lezama, Xin Zhao, Qiuhong Ke, Manmohan Chandraker, Xiaoxiao Sun, Yue Yao, Kevin W. Bowyer, Haiyu Wu

Date: 19-Oct

Time: 8:00 - 12:00

Location: 304 A

Summary: The 4th DataCV Workshop focuses on advancing data-centric perspectives in computer vision, shifting attention from algorithm-centric research to the analysis and understanding of vision datasets. We aim to explore dataset-level properties, representations, and similarities, as well as challenges in bias, fairness, and generalization. Topics include evaluating vision-language models, improving dataset quality through simulation, and reducing reliance on labeled data. The workshop encourages research on how dataset insights can guide model development, performance prediction, and ethical considerations. By fostering discussion and innovation in dataset analysis, DataCV promotes more robust, generalizable, and responsible vision systems.

The Challenge of Detecting Synthetic Manipulations in ID Documents

Organizers: Pavel Korshunov, Nevena Shamoska, Magdalena Połać, Vedrana Krivokuca, Vidit, Amir Mohammadi, Christophe Ecabert, Sébastien Marcel

Date: 19-Oct

Time: 8:50 - 12:10

Location: 303 B

Summary: An increase in AI based manipulations of ID document images threatens KYC systems widely used in online banking and other digital authentication services. DeepID challenge aimed to advance the research in the methods for detecting synthetic manipulations in ID documents. For that purpose a FantasyID dataset of both bona fide and manipulated fantasy ID cards was provided to the participants for training and tuning of their systems. Participating submissions were evaluated on a test set of FantasyID card was created with both seen and unseen attacks and on an out of domain private dataset of 20K real ID documents with both genuine bona fide and manipulated samples. The challenge included two tracks 1) a binary detection track to detect whether an ID document is manipulated or not and 2) a localization track, where the goal was to identify the manipulated regions of an ID document. With more than 100 registrations in the challenge, 26 teams have participated and 6 of them managed to beat the provided TruFor baseline method in detection track and 4 teams in the localization track.

The Eighth International Workshop on Computer Vision for Physiological Measurement (CVPM)

Organizers: Daniel McDuff, Wenjin Wang, Sander Stuijk, Tim Marks, Hassan Mansour, Vineet R. Shenoy

Date: 19-Oct

Time: 8:00 - 12:00

Location: 328

Summary: The Eighth International Workshop on Computer Vision for Physiological Measurement (CVPM) is the top venue for research on computer vision methods for measuring and modeling physiological processes. The goal of the workshop is to bridge the disciplines of computer vision and biomedical science and help effectively translate advances in AI into practice.

Workshop on Benchmarking Multi-Target Tracking: Towards Spatiotemporal Action Grounding in Videos

Organizers: Tanveer Hannan, Shuaicong Wu, Mark Weber, Suprosanna Shit, Rajat Koner, Jindong Gu, Aljosa Osep, Prof. Dr. Thomas Seidl, Prof. Dr. Laura Leal-Taixé

Date: 19-Oct

Time: 8:00 - 12:00

Location: 313 B

Summary: The 8th BMTT Workshop focuses on action-aware multi-object tracking, aiming to unify temporal action localization and object tracking through natural language queries. While existing benchmarks often address these tasks separately, this workshop presents unified challenges to evaluate both capabilities. Participants are encouraged to develop models that can understand complex actions, follow detailed language instructions, and track multiple objects across time. The workshop aims to close the gap between vision and language, advancing multimodal video understanding and supporting research on scalable, real-world systems capable of fine-grained, action-driven reasoning in dynamic scenes.

Workshop on Biomedical Image and Signal Computing for Unbiasedness; Interpretability; and Trustworthiness

Organizers: Romeo Lanzino, Bardh Prenkaj, Joanna Materzynska, Ananya Joshi, Silvia Zottin, Axel De Nardin, Tsui-Wei (Lily) Weng, Fabio Galasso, Gian Luca Foresti, Luigi Cinque, Roberto Cipolla

Date: 19-Oct

Time: 8:00 - 12:00

Location: 302 B

Summary: This workshop focuses on the foundational challenges of building AI systems that are unbiased, interpretable, and trustworthy. It aims to uncover the origins of algorithmic and data bias, advance the science of interpretability, and explore rigorous evaluation methods to ensure AI reliability. By bringing together researchers across biomedical imaging and signal processing, the workshop highlights novel methodologies and theoretical insights, emphasizing UIT as a scientific discipline rather than just an application concern. The event will showcase recent advances and foster discussions on future directions for inherently fair and transparent AI systems.

Workshop on Computer Vision with Single-Photon Cameras

Organizers: Atul Ingle, Sotiris Nousias, Mian Wei, Mel White

Date: 19-Oct

Time: 8:15 - 12:00

Location: 301 B

Summary: Single-photon cameras are an emerging class of camera technology with the potential to revolutionize the way today's computer vision systems capture and process scene information, thanks to their extreme sensitivity, high speed capabilities, and increasing commercial availability. These cameras can be used for a wide range of applications: self-driving cars and autonomous robots, high-sensitivity cameras for night photography and fluorescence-guided surgeries, and high dynamic range cameras for industrial machine vision and biomedical imaging applications. This workshop will showcase the myriad ways in which single-photon cameras are used today in computer vision and inspire new unexplored applications.

2nd Beyond Euclidean Workshop: Hyperbolic and Hyperspherical Learning for Computer Vision

Organizers: Georgios Leontidis, Aiden Durrant, Fabio Galasso, Michael Kampffmeyer, Pascal Mettes, Leyla Mirvakhabova, Adín Ramírez Rivera, Indro Spinelli, Stella Yu

Date: 19-Oct

Time: 8:00 - 17:00

Location: 305 B

Summary: Within deep learning, Euclidean geometry is the default basis for deep neural networks, yet the naive assumption that such a topology is optimal for all data types and tasks does not necessarily hold. A growing body of evidence suggests that data and the representations we aim to learn can be better captured through learning in corresponding geometries that exhibit non-Euclidean structures. Interest in non-Euclidean deep learning has grown dramatically in recent years, driven by advancing methodologies, libraries, and applications. The 2nd Beyond Euclidean workshop brings together computer vision researchers and keynote speakers who share an interest in exploring non-Euclidean geometry.

2nd Workshop on Explainable Computer Vision: Quo Vadis?

Organizers: Sukrut Rao, Robin Hesse, Quentin Bouniot, Sweta Mahajan, Amin Parchami-Araghi, Jayneel Parekh, Simone Schaub-Meyer, Florence d'Alché-Buc, Zeynep Akata, Stefan Roth, Bernt Schiele

Date: 19-Oct

Time: 8:00 - 17:00

Location: Ballroom A

Summary: This workshop aims to examine the state of the field of explainable AI (XAI) for computer vision, with the following goals: (1) discussion and dissemination of ideas at the cutting-edge of XAI research, and (2) a critical introspection on the challenges faced by the community and the way forward. The workshop includes papers, talks on recent advances, and a formal debate among invited speakers on the field's core issues. We hope to encourage brainstorming in the community to bridge the gap from theory to practice and address challenges brought forth by the rise of large-scale foundation models, such as fundamentally rethinking what one wants from an explanation, obtaining it, performing appropriate evaluations, complying with regulatory requirements, and maintaining model performance.

2nd Workshop on the Challenge Of Out-of-Label Hazard Detection in Autonomous Driving

Organizers: Ali K. AlShami, Ryan Rabinowitz, Maged Shoman, Jianwu Fang, Lukáš Picek, Shao-yuan Lo, Steve Cruz, Khang Lam, Jugal Kalita, Terrance E. Boult

Date: 19-Oct

Time: 8:00 - 17:00

Location: 319 A

Summary: The 2nd Workshop on the Challenge Of Out-Of-Label Hazards in Autonomous Driving (2COOOL) focuses on enhancing safety and robustness in autonomous systems by tackling challenges posed by unknown or out-of-distribution objects and behaviors. The workshop brings together researchers and practitioners from academia and industry, including our diverse team of co-organizers, to explore state-of-the-art solutions and techniques across various domains for real-world driving environments. It features expert keynotes, paper presentations, and a Kaggle challenge on generating hazard and accident reports from dashboard cameras. 2COOOL aims to advance the frontier of autonomous driving by fostering innovation to handle unexpected scenarios and build more robust ADAS systems.

3rd Workshop on Computer Vision for Automated Medical Diagnosis

Organizers: Fuying Wang, Sheng Liu, Qingyue Wei, Yi Lin, Lequan Yu, Yuyin Zhou, Yuzhe Yang, Angelica Aviles-Rivero, Hao Chen, Tingying Peng, Yifan Peng, Atlas Wang

Date: 19-Oct

Time: 8:00 - 17:00

Location: 314

Summary: The striding advances of computer vision techniques are revolutionizing many long-standing automatic medical diagnosis tasks. Emerging trends—such as Large Language Models (LLMs), Foundation Models (FMs), advanced learning paradigms (e.g., un-/semi-/self-supervised learning), and considerations of fairness and generalization—remain underexplored for secure and reliable automated medical diagnosis. Distinctly, this workshop emphasizes integrating insights from clinicians and radiologists alongside technical discussions to better advance the field.

3rd Workshop on Vision-based InduStrial InspectiON

Organizers: Shancong Mou, Hao Yan, Zirui Liu, Juan Du, Gokberk Cinbis, Wan Wang

Date: 19-Oct

Time: 8:00 - 17:00

Location: 319 B

Summary: The VISION workshop will provide a platform for the exchange of scholarly innovations and emerging practical challenges in Vision-based Industrial Inspection. Through a series of keynote talks, technical presentations, and challenge competitions, this workshop aims to (i) bring together researchers from the interdisciplinary research communities related to computer vision-based inspection; and (ii) connect researchers and industry practitioners to synergize recent research progress and current needs in industrial practice.

Artificial Social Intelligence Workshop

Organizers: Fiona Ryan, Leena Mathur, Anshul Gupta, Evonne Ng, Shiry Ginosar, Sangmin Lee, Paul Liang, Judy Hoffman, James M. Rehg, Louis-Philippe Morency

Date: 19-Oct

Time: 8:00 - 17:00

Location: 317 B

Summary: Humans use social intelligence to interpret and navigate interactions with other people and agents in our shared world. As AI systems become pervasive in human social situations, it is crucial to improve the social intelligence of these systems in order for them to seamlessly work with, for, and around humans. This workshop aims to bring together researchers from computer vision and other communities to collaborate towards building computational foundations for core social intelligence abilities. This edition of our workshop centers discussions, keynotes, and paper presentations around the topics of reasoning, multimodality, and embodiment in socially-intelligent AI.

Authenticity & Provenance in the age of Generative AI

Organizers: Shruti Agarwal, Sarah Barrington, Maty Bohacek, Cristian Canton, Laura Cassani, Hany Farid, Luisa Verdoliva

Date: 19-Oct

Time: 8:00 - 17:00

Location: 315

Summary: Generative AI allows for the rapid and automatic generation of highly realistic audio, images, and videos (so-called deepfakes). The field of media forensics and digital provenance focus on detection and authentication of this content, thus helping in mitigating the potential risks. This workshop aims at bringing a heterogeneous group of specialists from academia, industry, and civil society together to discuss emerging threats, technologies, and mitigation strategies. The workshop will focus on the application of tools from computer vision, pattern recognition, and machine learning, as well as the development of novel approaches for verifying the integrity and tracing the origins of digital media, the creation of novel datasets for evaluation, large-scale evaluations of existing forensic techniques, and ethical/policy considerations around generative AI and forensic techniques.

Computer Vision for Developing Countries

Organizers: Cuong Dao, Du Tran, Tuan-Anh Vu, Williem, Siddhartha Gairola, Ujjwal Verma, Vannkinh Nom

Date: 19-Oct

Time: 8:00 - 17:00

Location: 302 A

Summary: The Computer Vision for Developing Countries (CV4DC) workshop aims to create a supportive environment where students/researchers in the field of computer vision and related areas in AI can connect with each other, share their latest work, and expand their network for potential future collaborations and mentorships. This workshop empowers students and researchers from underrepresented, developing countries by providing opportunities to network, learn from the field experts, and share their work. We believe giving opportunities to students and researchers from lesser-known countries will foster diversity in computer research, leading to richer and more innovative contributions to the field.

Embedded Vision Workshop

Organizers: Tse-Wei Chen, Branislav Kisacanin, Ahmed Nabil Belbachir, Marius Leordeanu

Date: 19-Oct

Time: 8:00 - 17:00

Location: 316 A

Summary: Embedded vision is an active field of research, bringing together efficient learning models with fast computer vision and pattern recognition algorithms, to tackle many areas of robotics and intelligent systems that are enjoying an impressive growth today. Such strong impact comes with many challenges that stem from the difficulty of understanding complex visual scenes under the tight computational constraints required by real-time solutions on embedded devices. The Embedded Vision Workshop will provide a venue for discussing these challenges by bringing together researchers and practitioners from the different fields outlined above.

Foundations Models for V2X-Based Cooperative Autonomous Driving

Organizers: Walter Zimmer, Ross Greer, Max Ronecker, Lars Ullrich, Arpita Vats, Chuheng Wei, Haibao Yu, Rui Song, Jiajie Zhang, Julie Stephany Berrio Perez, Zewei Zhou, Tianhui Cai, Yifan Liu, Haoxuan Ma, Xingcheng Zhou, Rahul Raja, Zhengzhong Tu, Holger Caesar, Alina Roitberg, Guoyuan Wu, Jiaqi Ma, Daniel Watzenig, Mohan Trivedi, Alois Knoll

Date: 19-Oct

Time: 8:00 - 17:00

Location: Ballroom C

Summary: DriveX explores the integration of foundation models and V2X-based cooperative systems to improve perception, planning, and decision-making in autonomous vehicles. While traditional single-vehicle systems have advanced tasks like 3D object detection, emerging challenges like holistic scene understanding and 3D occupancy prediction require more comprehensive solutions. Collaborative driving systems, utilizing V2X communication and roadside infrastructure, extend sensory range, provide hazard warnings, and improve decision-making through shared data. Simultaneously, foundation models like Vision-Language Models (VLMs) offer generalization abilities, enabling zero-shot learning, open-vocabulary recognition, and scene explanation for novel scenarios. Recent advancements in end-to-end systems and foundation models like DriveLLM further enhance autonomous systems. The workshop aims to bring together experts to explore these technologies, address challenges, and advance road safety.

Geometry-Free Novel View Synthesis and Controllable Video Models

Organizers: Andrea Tagliasacchi, Sherwin Bahmani, Despoina Paschalidou, David Lindell, Konstantinos Derpanis, Marcus Brubaker, Boyang Deng, Haven (Haiwen) Feng, Qianqian Wang, Siyu Tang, Leonidas Guibas

Date: 19-Oct

Time: 8:00 - 17:00

Location: 323 C

Summary: This workshop focuses on recent advances in video generative models and their applications in 3D and 4D generation and reconstruction. Topics include camera- and motion-controlled video synthesis, large-scale 3D/4D reconstruction, neural rendering, and generative model-guided pipelines. A central focus is geometry-free novel view synthesis with video diffusion models, enabling spatial control without explicit 3D geometry. The program also covers the distillation of temporal models into spatial representations. By highlighting these developments, the workshop aims to chart the path toward more controllable, photorealistic, and efficient generative pipelines that unify video generation with 3D and 4D reconstruction.

Human-Interactive Generation and Editing

Organizers: Xi Chen, Shaoteng Liu, Jinbo Xing, Xin Yu, Yuanhao Cai, Tianyu Wang, Xiaojuan Qi, Hengshuang Zhao, Scott Cohen, Radu Timofte, Alan Yuille, Zhe Lin

Date: 19-Oct

Time: 9:00 - 18:00

Location: 309

Summary: The rapid evolution of generative AI has reshaped content creation across images, video, and 3D/4D visuals. This workshop focuses on cutting-edge methodologies, practical applications, and open challenges in image/video/3D/4D generation and related editing tasks with an emphasis on flexible and friendly human interactions and multi-modal control signals. This workshop will serve as a platform for researchers and practitioners to discuss key topics related to visual content creation and editing with versatile interactions.

Interactive Human-centric Foundation Models

Organizers: Shixiang Tang, Yizhou Wang, Xin Chen, Wanli Ouyang, Shiyao Xu, Jing Liu, Emily Kim, Xiaowei Zhou, Taku Komura, Gül Varol, Nicu Sebe, Wampfler Rafael

Date: 19-Oct

Time: 8:00 - 17:00

Location: 317 A

Summary: While Human-Centric Foundation Models (HFM) excel at perceiving and generating human data, they remain passive, struggling with real-time interaction and adaptation. This limits real-world deployment. The emerging field of Interactive HFM (I-HFM) addresses this by enabling bidirectional engagement. I-HFMs operate across three critical dimensions: (a) interacting with users for intuitive content creation/refinement, (b) interacting with environments to learn and adapt like humans, and (c) interacting with other agents for collaborative task-solving. This interactivity transforms AI from passive models into proactive, human-like agents, bridging the gap towards responsive, socially intelligent AGI that integrates seamlessly into human societies.

Joint Workshop on Marine Vision

Organizers: David Nakath, Malte Pedersen, Alexandra Branzan Albu, Anthony Hoogs, Derya Akkaynak, Maia Hoeberechts, Kevin Köser, Thomas B. Moeslund, Joakim B. Haurum, Justin Kay, Rupa Kurinchi-Vendhan

Date: 19-Oct

Time: 9:00 - 18:00

Location: 318 B

Summary: This workshop is organized as a collaboration between the 6th Workshop on Computer Vision for Analysis of Underwater Imagery(CVAUI) and the 3rd Automated Analysis of Marine Visual Data for Environmental Monitoring (AAMVEM). Visually monitoring marine environments poses a vastly different task compared to monitoring terrestrial environments. It is physically challenging to acquire underwater data, and the data typically have low signal-to-noise ratios due to the scattering nature of the water body. The aim of this workshop is to deepen the understanding of the challenges related to marine monitoring and to advance computer vision techniques to address them.

Learning to See: Advancing Spatial Understanding for Embodied Intelligence

Organizers: Hongyang Li, Philipp Krähenbühl, Kashyap Chitta, Eric Jang, Andrei Bursuc, Huijie Wang

Date: 19-Oct

Time: 8:00 - 17:00

Location: 312

Summary: The world is three-dimensional. This fact was first seen by trilobites, the first organisms capable of sensing light. From that moment, nervous systems began to evolve, gradually transforming mere sight into insight, understanding, and action. All these combined gives rise to intelligence. Despite remarkable technological advancements in recent decades, modern embodied systems remain far from achieving full intelligence. They fall short in several key aspects: (i) contain information necessary for physical interaction, such as temporal dynamics of the scene; (ii) have a prior over semantic relevance, and should focus on task-relevant features like objects and their relationships; and (iii) be compact, avoiding the inclusion of irrelevant details, such as background elements. Attempts have been made, including integrating foundational models and utilizing large-scale data. Yet, the path to true intelligence remains long, with significant progress still required.

Multi-modal Localization and Mapping

Organizers: Timothy D Barfoot, Luca Carlone, Daniel Cremers, Frank Dellaert, Ayoung Kim, Yan Xia, Niclas Zeller

Date: 19-Oct

Time: 8:00 - 17:00

Location: 311

Summary: Multi-modal Localization and Mapping is an essential component of computer vision, with diverse applications in fields such as autonomous robotics, augmented reality, and beyond. This workshop aims to unite researchers, practitioners, and enthusiasts to explore the latest advancements, challenges, and innovations in multi-modal localization and mapping. By leveraging information from various sensors (e.g. camera, IMU, LiDAR, radar, and language), multi-modal approaches can significantly enhance localization and mapping accuracy in complex environments.

Multispectral Imaging for Robotics and Automation

Organizers: Shiho Kim, Yagiz Nalcakan, Rui Fan, Kailun Yang, Yalın Baştanlar, Ömer Şahin Taş, Jun Won Choi, Ukcheol Shin, Michal Kovac

Date: 19-Oct

Time: 9:00 - 17:00

Location: 310

Summary: The Multispectral Imaging for Robotics and Automation (MIRA) workshop brings together researchers and practitioners at the intersection of multispectral imaging, computer vision, and robotics. By leveraging data beyond the visible spectrum, multispectral imaging enables robust perception in challenging conditions, supporting applications from autonomous driving and industrial inspection to agricultural automation and search and rescue. MIRA aims to foster interdisciplinary collaboration across academia and industry, highlighting advances in sensor technology, spectral image processing, and downstream tasks like detection, segmentation, and decision-making. We welcome contributions exploring novel methods, applications, and datasets that advance the state of multispectral robotics.

Structural Priors for Vision

Organizers: Sangwoo Mo, Congyue Deng, Hila Chefer, Daniel Zoran, Kaichun Mo, Leonidas Guibas, Stella Yu

Date: 19-Oct

Time: 8:00 - 17:00

Location: 323 A

Summary: In recent years, there has been a growing trend toward training data-centric, large-scale foundation models that reduce reliance on structural priors. However, is simply scaling up Transformers truly the ultimate solution for computer vision? In this workshop, we aim to reintroduce structural priors and explore how they can further push the boundaries of foundation models. Our workshop provides an interdisciplinary space for sharing ideas across domains. For example, scene-aware 2D perception can enhance 3D modeling and robotic manipulation, while geometric reasoning can enhance the visual grounding of 2D perception and multimodal models. Through these interactions, we aim to better define the role of priors in vision foundation models.

The Third Perception Test Challenge

Organizers: Joe Heyward, Nikhil Parthasarathy, Joao Carreira, Dima Damen, Andrew Zisserman, Viorica Patraucean, Eunice Yiu, Shiry Ginosar, Saman Motamed, Priyank Jaini

Date: 19-Oct

Time: 8:00 - 17:00

Location: Ballroom B

Summary: The 3rd Perception Test challenge comprehensively evaluates the perception capabilities of large multimodal models using the Perception Test benchmark. This year, novel tracks unify diverse tasks under common interfaces: joint object/point tracking, joint action/sound localisation, and unified multiple-choice videoQA (integrating non-semantic tasks via inpainted queries). A new VLM interpretability track is included to investigate model strengths and failures. Guest tracks cover image understanding (KiVA) and video generation (Physics-IQ). Our workshop provides a venue to evaluate all foundation vision models—discriminative, generative, image- or video-based. Prizes up to 50k EUR are available.

Vision-based AI for Digital Health: From Pixels to Practics

Organizers: Hui Zhang, Bojian Ho, Yuanfang Guan, Guangchen Ruan

Date: 19-Oct

Time: 8:00 - 17:00

Location: 303 A

Summary: The workshop aims to unite researchers and practitioners at the intersection of vision-based AI and large language models (LLMs) to advance digital health innovation. By showcasing deep-learning applications on high-resolution imaging modalities (e.g., MRI, CT, retinal photography), we'll explore how early disease detection and automated image review can boost diagnostic accuracy and streamline clinical workflows. We'll also delve into emerging "Vision + LLM" systems that fuse visual understanding with natural-language capabilities for automated report generation, intelligent literature retrieval, and interactive decision support. Through presentations, discussions, and hands-on demos, participants will identify challenges, exchange best practices, and chart pathways toward more personalized, data-driven care.

Workshop on Safe and Trustworthy Multimodal AI Systems

Organizers: Carlos Hinojosa, Yinpeng Dong, Adel Bibi, Jindong Gu, Yichi Zhang, Wenxuan Zhang, Lama Alssum, Andres Villa, Juan Carlos L. Alcazar, Chen Zhao, Lingjuan Lyu, Mohamed Elhoseiny, Bernard Ghanem, Philip Torr

Date: 19-Oct

Time: 8:00 - 17:00

Location: 308 B

Summary: Multimodal systems are transforming AI by enabling models to understand and act across language, vision, and other modalities, driving advances in robotics, autonomous driving, and scientific discovery. However, these capabilities raise serious safety and trustworthiness concerns, as traditional safeguards often fall short in multimodal contexts. The Workshop on Safe and Trustworthy Multimodal AI Systems (SaFeMM-AI) at ICCV 2025 brings together the computer vision community to address challenges including hallucinations, privacy leakage, and jailbreak vulnerabilities, and to promote the development of safer, more robust, and reliable multimodal models that can handle unsafe or adversarial inputs and consistently produce trustworthy outputs.

RetailVision6 - Revolutionizing the World of Retail

Organizers: Ehud Barnea, Yosi Keller, Marina Paolanti, Sean Ma, Austen Groener, Weijian Li, Quanfu Fan, Rocco Pietrini

Date: 19-Oct

Time: 8:45 - 17:00

Location: 325 A

Summary: Recent advances in computer vision have significantly impacted the retail sector, introducing new opportunities and challenges across both physical and online domains. This workshop explores key problems such as shopper-product interaction, fine-grained recognition of visually similar and frequently changing products, and large-scale visual search across over 100,000 product classes. It also showcases advancements in generative models for tasks like product image synthesis and virtual try-on. These are just some of the challenges in the retail domain. By highlighting recent progress and open research directions, the workshop aims to bring together researchers and practitioners to advance the state of computer vision in retail.

10th International Workshop on Recovering 6D Object Pose

Organizers: Martin Sundermeyer, Tomáš Hodaň, Médéric Fourmy, Van Nguyen Nguyen, Junwen Huang, Stephen Tyree, Jonathan Tremblay, Eric Brachmann, Sindi Shkodrani, Bertram Drost, Carsten Steger, Vincent Lepetit, Carsten Rother, Stan Birchfield, Jiří Matas

Date: 19-Oct

Time: 13:00 - 17:00

Location: 313 A

Summary: The R6D workshop discusses topics related to model-based and model-free 6D object pose estimation which are relevant for applications such as robotic manipulation and augmented reality. The 10th workshop edition is organized in conjunction with the BOP Challenge 2025 that benchmarks the latest pose estimation methods in challenging settings including the new BOP-Industrial datasets. Find out about the latest trends and remaining challenges in object-centric 3D vision and learn how the latest methods perform in the wild on real robots.

13th International Workshop on Assistive Computer Vision and Robotics

Organizers: Giovanni Maria Farinella, Antonino Furnari, Marco Leo, Gerard G. Medioni, Francesco Ragusa, Mohan Trivedi

Date: 19-Oct

Time: 13:00 - 17:00

Location: 316 C

Summary: Designing systems with humans in the loop to assist users is an active research area with potential societal impact. Investigations require many innovations, tools, and evaluation criteria, even compared to fully autonomous systems. Implementing such systems demands significant effort to achieve reliability and raises issues related to usability, privacy, and acceptability. Moreover, multidisciplinary competencies are needed 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, emphasizing related issues and how researchers in various fields have addressed them.

2nd Workshop and Challenge on Unlearning and Model Editing

Organizers: Kartik Thakral, Diego Garcia-Olano, Tal Hassner, Iacopo Masi, Mayank Vatsa

Date: 19-Oct

Time: 13:00 - 18:00

Location: 308 A

Summary: The 2nd Workshop and Challenge on Unlearning and Model Editing (U&ME) is a half-day event at ICCV 2025 in Hawaii on October 19, 2025, in the afternoon, and 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.

5th Workshop and Challenge on Open-World 3D Scene Understanding

Organizers: Francis Engelmann, Ayca Takmaz, Alex Delitzas, Elisabetta Fedele, Anna-Maria Halacheva, Katerina Adam, Yang Miao, Jan-Nico Zaech, Zuria Bauer, Johanna Wald, Danda Pani Paudel, Or Litany, Federico Tombari, Marc Pollefeys, Leonidas Guibas

Date: 19-Oct

Time: 13:30 - 17:00

Location: 306 B

Summary: The ability to perceive, understand, and interact with 3D scenes is crucial for applications in AR/VR, robotics, healthcare, and beyond. Current 3D scene understanding models are largely limited to low-level recognition tasks such as object detection or semantic segmentation, and struggle to generalize beyond predefined training labels. Recently, large VLMs such as LLAVA have demonstrated impressive capabilities. Initial works have shown their potential to extend 3D scene understanding not only to open vocabulary recognition, but also reasoning about affordances, activities, and properties of unseen environments. This workshop aims to define tasks, metrics, and benchmarks to advance this emerging direction.

Comic Intelligence Quotient: Advances and Challenges in AI-driven Comic Analysis

Organizers: Ragav Sachdeva, Emanuele Vivoli, Artemis Llabrés, Deblina Bhattacharjee, Dimosthenis Karatzas, Andrew Zisserman

Date: 19-Oct

Time: 13:00 - 17:00

Location: 305 A

Summary: Comics are a uniquely compelling visual storytelling medium, blending images and text, but they present significant challenges for Artificial Intelligence. Unlike natural images, comics rely on abstract, stylized panels and implicit transitions that demand complex inference, causing even state-of-the-art vision-language models to struggle with tasks like panel sequencing and cross-panel reasoning. This workshop brings together researchers from computer vision, cognitive science, and multimedia analysis to advance AI-driven comic understanding. Through talks and discussions, we will explore new methodologies for multimodal reasoning.

End-to-End 3D Learning

Organizers: Zhiwen Fan, Qianqian Wang, Yuanbo Xiangli, Wenyan Cong, Yiqing Liang, Jiachen Li, Zhengzhong Tu, Georgios Pavlakos, Yan Wang, Achuta Kadambi

Date: 19-Oct

Time: 13:00 - 17:00

Location: 304 B

Summary: Recent advances in foundation models for language and 2D vision have revolutionized data processing, but real-world applications in robotics, XR, and autonomous systems demand robust 3D understanding. The End-to-End 3D Learning (E2E3D) workshop explores unified, differentiable frameworks that transform raw imagery directly into 3D representations, replacing slow, error-prone multistage pipelines. We will examine unified architectures, pretraining on large unannotated datasets, and efficient real-time inference. Discussions will address data challenges—leveraging massive multi-view sources and multimodal datasets—and real-world impacts in autonomous driving, AR/VR, and scientific imaging. By unifying modeling, inference, and optimization, E2E3D aims to spark innovations in scalable, generalizable 3D AI.

First Workshop on Skilled Activity Understanding; Assessment and Feedback Generation

Organizers: Paritosh Parmar, Angela Yao, Brendan Morris, Basura Fernando

Date: 19-Oct

Time: 14:00 - 17:00

Location: 318 A

Summary: Imagine a world where computer vision-based systems can analyze a video of an athlete, a surgeon, a patient, or a factory worker and instantly provide expert-level actionable feedback---correcting techniques, identifying inefficiencies, and helping people refine their skills in real time. Thanks to rapid progress in video understanding, this vision is becoming reality. AI-powered systems can now analyze complex human activities, assess performance, and generate intelligent feedback, unlocking new possibilities in sports, healthcare, manufacturing, education, rehabilitation, and beyond. Through Expert Keynotes and Invited Contributions, this workshop will explore the cutting edge of skilled activity understanding, assessment, and feedback generation, bridging research and real-world applications.

Multi-Modal Foundation Models for Cancer Detection and Prevention

Organizers: Ali Diba, Biagio Brattoli, Thijs Kooi, Tae Soo Kim, Sergio Pereira, Donggeun Yoo, Kayhan Batmanghelich, Yun Liu, Daniel Golden, Pranav Rajpurkar, Eun Kyoung Hong, Zelda Mariet, Shekoofeh Azizi

Date: 19-Oct

Time: 13:00 - 17:00

Location: 326 B

Summary: This workshop explores how multi-modal foundation models can revolutionize cancer care by integrating AI, computer vision, and machine learning. By leveraging diverse data types—such as medical imaging, genomics, and EHRs—these models enable earlier detection, personalized treatment, and better outcome prediction. Pre-trained on large datasets and fine-tuned for specific tasks, they offer adaptability across cancer types and clinical settings. The event brings together experts from academia, industry, and healthcare to share research, tackle challenges in data integration and model interpretability, and promote clinical translation. The goal is to advance cancer research and accelerate the real-world impact of AI in oncology.

Multimodal Continual Learning

Organizers: Yunhui Guo, Yapeng Tian, Mingrui Liu, Sayna Ebrahimi, Henry Gouk, Sarthak Maharana

Date: 19-Oct

Time: 13:00 - 17:00

Location: 301 A

Summary: In recent years, advances in machine learning and computer vision have driven continual learning (CL), allowing models to learn new tasks incrementally while retaining prior knowledge without full retraining. Early CL focused on unimodal data like images for classification, but powerful multimodal models now unify images, videos, text, and audio. Multimodal continual learning (MCL) must tackle unique challenges, including modality-specific forgetting, imbalance, and maintaining cross-modal links. This MCL workshop will address these issues, highlight new research directions, and promote collaboration among researchers, practitioners, and industry, advancing inclusive, efficient continual learning for modern AI systems.

Neural SLAM Workshop

Organizers: Martin R. Oswald, Matteo Poggi, Fabio Tosi, Youmin Zhang, Yiyi Liao, Vladimir Yugay, Yue Li

Date: 19-Oct

Time: 13:00 - 17:00

Location: 304 A

Summary: Over the past two decades, SLAM (Simultaneous Localization and Mapping) has evolved significantly, transitioning from traditional methods to deep learning and, more recently, to Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting (3DGS). Since 2021, a surge of over 200 papers has reshaped the field, enabling new applications like realistic novel view synthesis. However, this rapid progress also raises challenges, such as lack of standardized benchmarks and understanding of key design choices. This workshop aims to unite researchers interested in dense neural SLAM, fostering discussion through keynotes, posters, and panels to explore emerging trends and future directions.

Representation Learning with Very Limited Resources: When Data; Modalities; Labels; and Computing Resources are Scarce

Organizers: Hirokatsu Kataoka, Yuki M. Asano, Iro Laina, Rio Yokota, Nakamasa Inoue, Rintaro Yanagi, Partha Das, Connor Anderson, Ryousuke Yamada, Daichi Otsuka, Yoshihiro Fukuhara

Date: 19-Oct

Time: 13:00 - 17:00

Location: 306 A

Summary: Modern vision and multimodal models depend on massive datasets and heavy compute, magnifying costs, energy use, bias, copyright, and privacy risks. The "DeepSeek shock" of January 2025 spotlighted the urgency of learning powerful representations under tight resource limits. Now in its third edition, our workshop continues to explore strategies for robust representation learning when data, labels, modalities, parameters, or compute are scarce. We focus on techniques such as synthetic and distilled data, self-supervision, transfer learning, sparsity, and low-rank adaptation that squeeze maximum performance from minimal resources.

SEA: 1st workshop on Sustainability with Earth observation and AI

Organizers: Zhuo Zheng, Junjue Wang, Xiaoyan Lu, Xinyu Dou, Gengchen Mai, Yanfei Zhong, Liangpei Zhang, Marshall Burke, David Lobell, Stefano Ermon

Date: 19-Oct

Time: 13:00 - 17:00

Location: 307 A

Summary: SEA brings together researchers, practitioners, and policy-makers to advance the state-of-the-art in applying artificial intelligence to Earth observation for sustainability challenges. Technically, this workshop explores how state-of-the-art EO data-tailored foundation models, efficient architectures, and novel learning paradigms can be leveraged or adapted to tackle pressing sustainability challenges. Topics include, but are not limited to, climate monitoring, disaster response, biodiversity, agriculture, urban development, clean energy, and social economics.

Short-Form Video Understanding: The Next Frontier in Video Intelligence

Organizers: Uttaran Bhattacharya, Ishita Dasgupta, Mehrab Tanjim, Chen-Yi Lu, Kunjal Panchal, Dinesh Manocha

Date: 19-Oct

Time: 13:00 - 17:00

Location: 313 B

Summary: Short-form videos (SVs) have proliferated as primary sources for entertainment, information, advertising, and social communication. Marketers are increasingly turning to SVs to reach their customers, and creative artists have begun to view SVs as a separate form of art and media for designing their content. Currently, SVs account for 90% of internet traffic and are estimated to be about 2.5 times more engaging than longer videos, driving their widespread popularity and diversity. Our workshop aims to consolidate efforts in SV understanding, highlight specific challenges, map the research landscape, and establish a foundation for future development in this rapidly expanding domain.

The 2nd Workshop on Efficient Computing under Limited Resources: Visual Computing

Organizers: Jinyang Guo, Zhenghao Chen, Yuqing Ma, Yifu Ding, Xianglong Liu, Jinman Kim, Wanli Ouyang, Dacheng Tao

Date: 19-Oct

Time: 13:00 - 17:00

Location: 327

Summary: This workshop explores efficient methodologies in visual computing, focusing on data-efficient techniques (e.g., image/video compression), label-efficient strategies (e.g., zero/few-shot learning), and model-efficient approaches (e.g., sparsification, quantization). By bringing together experts in these areas, we aim to foster the exchange of recent findings and discuss future directions. Given the growing importance of efficiency in practical deployments, this topic has attracted significant research interest. The workshop provides a platform for presenting novel perspectives and addressing core challenges in visual computing, ultimately driving advancements that bridge academic research with real-world applications.

The 6th Face Anti-Spoofing Workshop: Unified Physical-Digital Attacks Detection

Organizers: Jun Wan, Jiankang Deng, Jun Lan, Weiqiang Wang, Sergio Escalera, Hugo Jair Escalante, Xiaoming Liu, Ajian Liu, Hui Ma, Yanyan Liang, Zhen Lei, Isabelle Guyon

Date: 19-Oct

Time: 13:00 - 17:00

Location: 323 B

Summary: To tackle growing face spoofing threats, the 6th Face Anti-Spoofing Workshop@ICCV 2025 introduces a unified physical-digital attack detection task and releases UniAttackData+, a large-scale dataset with 2,875 subjects, 18,250 live videos, and 679,097 forgeries spanning 54 attack types, fostering research on generalizable FAS models.

Transparent & Reflective objects In the wild Challenges

Organizers: Alex Costanzino, Pierluigi Zama Ramirez, Fabio Tosi, Matteo Poggi, Luigi Di Stefano, Jean-Baptiste Weibel, Doris Antensteiner, Markus Vincze, Benjamin Busam, Guangyao Zhai, Weihang Li, Junwen Huang

Date: 19-Oct

Time: 13:00 - 17:00

Location: 302 B

Summary: Depth and pose estimation are critical for enabling machines to interact effectively with the real world. Depth estimation provides the spatial structure of a scene, pose estimation localises and orients objects within it, both fundamental for robotics, augmented reality, and 3D understanding. Traditional approaches achieved impressive results on standard benchmarks like KITTI and Middlebury. However, when these methods encounter reflective and transparent objects, their performance degrades significantly. This limitation is particularly problematic as these challenging materials are common in everyday environments. TRICKY 2025 features two complementary challenges encouraging the development of next-generation algorithms capable of advanced reasoning on non-Lambertian objects.

Visual Quality Assessment Competition

Organizers: Chris Wei Zhou, Jian Wang, Sizhuo Ma, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, Zhengzhong Tu, Hadi Amirpour, Shiqi Wang, Hanwei Zhu, Yixiao Li, Fan Huang, Shuo Xing, Fengjun Guo, Xin Li, Wei-Ting Chen, Xiaoshuai Hao, Ying Chen, Huasheng Wang, Pengxiang Xiao

Date: 19-Oct

Time: 13:00 - 17:00

Location: 303 B

Summary: The Visual Quality Assessment Competition (VQualA) Workshop at ICCV 2025 aims to advance perceptual quality evaluation in computer vision by addressing the limitations of traditional metrics such as PSNR and SSIM. Leveraging deep learning, generative models, and multimodal large language models (MLLMs), the workshop emphasizes human-aligned assessments. It features seven diverse challenges spanning low-level vision, document enhancement, face image quality, AIGC video evaluation, and visual comparison via MLLMs. Through both scalar metrics and comparative reasoning tasks, VQualA fosters more interpretable, robust, and perceptually meaningful evaluation. It unites academic and industrial communities to push the frontier of visual quality assessment forward.

Women in Computer Vision

Organizers: Aishwarya Jadhav, Supatta Viriyavisuthisakul, Elena Govi, Chaitra Desai, Ariana Bermudez Venegas, Uma Mudenagudi, Anuhya Thota

Date: 19-Oct

Time: 13:00 - 17:00

Location: 301 B

Summary: Women in Computer Vision Workshop (WiCV@ICCV 2025) aims to promote and increase the participation of female-identifying researchers in the computer vision community. The workshop features technical talks, poster sessions, a panel discussion, and a mentoring dinner to foster networking, visibility, and collaboration. WiCV provides a platform to present cutting-edge research, share career insights, and discuss challenges faced by women in CV. The event is open to all ICCV attendees and strongly encourages junior researchers and students to participate. Through community support and industry sponsorship, WiCV continues its mission to build a more inclusive and diverse research ecosystem.

Workshop on Graphic Design Understanding and Generation

Organizers: Kota Yamaguchi, Cherry Zhao, Rajiv Jain, Sanket Biswas, Akshay Gadi Patil, Yuhui Yuan

Date: 19-Oct

Time: 13:00 - 17:00

Location: 328

Summary: The workshop on Graphic Design Understanding and Generation (GDUG) aims to bring together researchers, creators, and practitioners to discuss the important concepts, technical perspectives, limitations, and ethical considerations surrounding recognition and generative approaches to graphic design and documents. While recent advances in generative AI are making impressive strides in creative domains, there is a disconnect between research attempts and the real-world workflow that involves graphics design, such as the creation of a website, posters, online advertisements, social media posts, infographics, or presentation slides, where creators do not paint pixels but instead work with structured documents, such as layered object representation, stylistic attributes, and typography.

Workshop on Knowledge-Intensive Multimodal Reasoning

Organizers: Arman Cohan, Xiangliang Zhang, Manling Li, Yapeng Tian, Minhao Cheng, Zeynep Akata, Yilun Zhao, Haowei Zhang, Tianyu Yang, Zhenting Qi, Yuyang Liu, Zhiyuan Hu, Simeng Han, Rui Xiao, Xiangru Tang

Date: 19-Oct

Time: 13:00 - 17:00

Location: 313 C

Summary: This workshop aims to advance the frontier of multimodal AI systems that can effectively reason across specialized domains requiring extensive domain knowledge. Recent advancements in multimodal AI—combining information from text, images, audio, and structured data—have unlocked impressive capabilities in general-purpose reasoning. However, significant challenges persist when these systems encounter scenarios demanding deep domain expertise in fields such as medicine, engineering, and scientific research. Such contexts require expert-level perception and reasoning grounded in extensive subject knowledge, highlighting the need for specialized strategies to handle domain-specific complexity. Through invited talks, panel discussions, and interactive poster sessions, researchers and practitioners from diverse backgrounds will share the latest developments, ongoing hurdles, and promising future directions for knowledge-intensive multimodal reasoning. The workshop aims to foster collaboration and stimulate innovation towards the development of next-generation multimodal AI systems capable of reliable, transparent, and contextually grounded reasoning in specialized, high-stakes environments.

Monday, October 20

NOTE: Tutorial rooms are subject to change. Refer to the online site for up-to-date locations. Use the QR code for each tutorial to see its schedule. Here is the QR code for the ICCV2025 Tutorials page.

7:00 - 17:00	**Registration / Badge Pickup** (HCC Main Lobby)
7:00 - 17:00	**Nursing Pod**
7:00 - 17:00	**Prayer / Quiet Room / Mother's Room** (326 A)
7:00 - 9:00	**Breakfast** (Exhibit Hall II)
8:00 - 18:00	**Workshop & Tutorials**
10:00 - 11:00	**Coffee Break & Posters** (Exhibit Hall II)
12:00 - 13:45	**Lunch** (Exhibit Hall II / Overflow Seating Rooftop)
15:00 - 16:00	**Coffee Break & Posters** (Exhibit Hall II)
8:00 - 17:00	**POSTERS** (Exhibit Hall II)

TUTORIALS

Fourth Hands-on Egocentric Research Tutorial with Project Aria, from Meta

Organizer: James Fort
Date: 10/20
Time: 8:00 - 11:00
Location 326 B

Summary: Project Aria is a research device that is worn like a regular pair of glasses, for researchers to study the future of computer vision with always-on sensing. Sensors in Project Aria capture egocentric video and audio, in addition to eye-gaze, inertial, and location information. On-device compute power is used to encrypt and store information that, when uploaded to separate designated back-end storage, helps researchers build the capabilities necessary for AR to work in the real world. In this fourth tutorial, in addition to sharing research from academic partner program members, we will also provide an introduction to the second generation of Aria glasses 'Aria Gen 2', announced in February. As part of this introduction, we will provide live hands-on demo of Aria Research Kit (including Gen 2 glasses), describe how researchers can gain access to the Project Aria academic program, and demonstrate how open-source tools can be used to accelerate research for specific research challenges, including visual and non-visual localization and mapping, static and dynamic object detection and spatialization, human pose estimation, and building geometry estimation. We will review new open datasets from Meta and academic partners partners, including a dataset of 6000+ 3D objects with Aria captures for each object to facilitate novel research on

egocentric 3D object reconstruction, and a review of ego-perception challenges and benchmarks associated with all datasets, including a demonstration of methods for approaching each challenge.

Foundations of Interpretable AI

Organizer: Aditya Chattopadhyay
Date: 10/20
Time: 8:00 - 12:00
Location 306 B

Summary: In recent years, interpretability has emerged as a significant barrier to the widespread adoption of deep learning techniques, particularly in domains where AI decisions can have consequential impacts on human lives, such as healthcare and finance. Recent attempts at interpreting the decisions made by a deep network can be broadly classified in two categories, (i) methods that seek to explain existing models (post-hoc explainability), and (ii) methods that seek to build models that are explainable by design. This tutorial aims to provide a comprehensive overview of both approaches along with a discussion on their limitations. More specifically, this tutorial will consist of three lectures covering the following topics:

Post-hoc explainability methods; Explaining deep networks using Shapley values and statistical testing; Explainable-by-design deep networks.

From Segment Anything to Generalized Visual Grounding

Organizer: Andrew Westbury
Date: 10/20
Time: 9:00 - 12:30
Location 328

Summary: In this tutorial, Meta AI and its academic partners will overview frontier research on visual grounding. We will cover each building block necessary to move toward future general purpose visual grounding systems, including universal image and video encoding, multimodal language understanding, semantic instance segmentation and tracking, and the latest in 3D reconstructions methods. We will provide practical guidance on using SAM open source models, resources, and tooling to tackle the field's biggest open research problems. A new suite of SAM systems to be released this year will provide a foundation for our tutorial, offering practical entry points for each course component.

Foundation Models Meet Embodied Agents

Organizer: Manling Li
Date: 10/20
Time: 13:00 - 17:00
Location 306 B

Summary: An embodied agent is a generalist agent that can take natural language instructions from humans and perform a wide range of tasks in diverse environments. Recent years have witnessed the emergence of foundation models, which have shown remarkable success in supporting embodied agents for different abilities such as goal interpretation, subgoal decomposition, action sequencing, and transition modeling (causal transitions from preconditions to post-effects). We categorize the foundation models into Large Language Models (LLMs), Vision-Language Models (VLMs), and Vision-Language-Action Models (VLAs). In this tutorial, we will comprehensively review existing paradigms for foundations for embodied agents, and focus on their different formulations based on the fundamental mathematical framework of robot learning, Markov Decision Process (MDP), and design a structured view to investigate the robot's decision making process. This tutorial will present a systematic overview of recent advances in foundation models for embodied agents. We compare these models and explore their design space to guide future developments, focusing on Lower-Level Environment Encoding and Interaction and Longer-Horizon Decision Making.

RANSAC in 2025

Organizer: Daniel Barath
Date: 10/20
Time: 13:00 - 17:00
Location 328

Summary: RANSAC (Random Sample Consensus) has been a cornerstone of robust estimation in computer vision since its introduction in 1981. It remains highly relevant in 2025, as many vision applications still rely on detecting and handling outliers in data. This tutorial, "RANSAC in 2025", aims to provide a comprehensive update on the latest advancements of RANSAC and its family of algorithms. We will balance theoretical foundations (to understand how and why RANSAC works) with practical applications (to demonstrate its use in real-world vision problems). By covering both classic principles and cutting-edge improvements, the tutorial will equip researchers and practitioners with state-of-the-art techniques for robust model fitting in computer vision.

Towards Safe Multi-Modal Learning: Unique Challenges and Future Directions

Organizer: Xi Li
Date: 10/20
Time: 13:00 - 17:00
Location 326 B

Summary: Modern multi-modal learning leverages large models, such as large language models (LLMs), to integrate diverse data sources (e.g., text, images, audio, and video) and enhance understanding and decision-making. However, the inherent complexities of multi-modal learning introduce unique safety challenges that existing frameworks, primarily designed for uni-modal models, fail to address. This tutorial explores the emerging safety risks in multi-modal learning and provides insights into future research directions. We begin by examining the unique characteristics of multi-modal learning — modality integration, alignment, and fusion. We then review existing safety studies across adversarial attacks, data poisoning, jailbreak exploits, and hallucinations. Next, we analyze emerging safety threats exploiting multi-modal challenges, including risks from additional modalities, modality misalignment, and fused representations. Finally, we discuss potential directions for enhancing the safety of multi-modal learning. As multi-modal learning expands, addressing its safety risks is crucial. This tutorial lays the foundation for understanding these challenges and fostering discussions on trustworthy systems.

Notes:

WORKSHOPS

Anomaly Detection with Foundation Models

Organizers: Kuan-Chuan Peng, Ying Zhao, Abhishek Aich

Date: 20-Oct

Time: 8:25 - 12:15

Location: 314

Summary: The rapid advancement of foundation models in fields like healthcare, cybersecurity, and finance highlights the urgent need to improve their anomaly detection capabilities. Despite their growing application in high-stakes areas, the challenges of using these models for anomaly detection remain underexplored. The Anomaly Detection with Foundation Models (ADFM 2025) workshop aims to address this gap by focusing on the intersection of foundation models and anomaly detection. Our organizing and technical committee, composed of leading experts, provides a platform for advancing research and discussing the recent breakthroughs, and the technical and ethical implications of deploying these models. ADFM 2025 will foster interdisciplinary collaboration and contribute to the development of more reliable and effective anomaly detection systems in artificial intelligence.

Computer Vision for Materials Science

Organizers: Alexei Skurikhin, Alexander Hagen, Kai He, Kari Sentz, Joshua Stuckner, Katherine Sytwu

Date: 20-Oct

Time: 8:00 - 12:00

Location: 305 A

Summary: Computer vision and machine learning are critical tools to support large-scale materials characterization and development of new materials. Quantified structure features that are extracted from the data can be leveraged in statistical and machine learning models that establish processing-structure-property-performance (PSPP) relationships to identify non-linear and unintuitive trends in the high dimensional materials development space further accelerating materials development. The aim of workshop is to bring together cross-disciplinary researchers to demonstrate recent advancements in machine learning, computer vision, and materials microscopy, and discuss open problems such as representation learning, uncertainty quantification, and explainability in materials microscopy analysis.

Embodied Spatial Reasoning

Organizers: Wufei Ma, Yu-Cheng Chou, Xiwei Xuan, Artur Jesslen, Adam Kortylewski, Celso M. de Melo, Zhaoyang Wang, Rama Chellappa, Alan Yuille, Jieneng Chen

Date: 20-Oct

Time: 8:00 - 12:00

Location: Ballroom A

Summary: The 1st Embodied Spatial Reasoning Workshop at ICCV 2025 explores the integration of spatial understanding in intelligent agents. Focus areas include Embodied AI, which allows agents to perceive, reason, and act in real-world or simulated environments, and Spatial Reasoning, which involves interpreting spatial relations and sensory feedback. The workshop also delves into the development of Embodied World Models for building spatially coherent, semantically grounded internal representations, and Robot Spatial Reasoning, addressing challenges in planning and acting under uncertainty and task constraints. The goal is to advance robust, generalizable spatial reasoning for embodied agents.

Foundation & Generative Models in Biometrics

Organizers: Hatef Otroshi Shahreza, Vitomir Štruc, Luisa Verdoliva, Zhen Lei, Arun Ross, Sébastien Marcel

Date: 20-Oct

Time: 8:00 - 12:00

Location: 305 B

Summary: The ICCV 2025 Workshop on Foundation and Generative Models in Biometrics aims to bring together researchers to discuss state-of-the-art advancements, applications, and challenges in using foundation and generative models for biometric recognition, analysis, and security. While foundation models have gained significant attention in recent years, their applications in biometrics remain relatively underexplored. This workshop seeks to encourage discussions that inspire innovation and address the challenges of applying these advanced models in real-world biometric systems. The program will feature invited talks and paper presentations.

From street to space: 3D Vision AcrosS altiTudes

Organizers: Yujiao Shi, Yuanbo Xiangli, Zuzana Kukelova, Bo Dai, Richard Hartley, Hongdong Li

Date: 20-Oct

Time: 8:00 - 12:00

Location: 313 B

Summary: As large-scale 3D scene modeling becomes increasingly important for applications such as urban planning, robotics, autonomous navigation, and virtual simulations, the need for diverse, high-quality visual data is greater than ever. However, acquiring dense and high-resolution ground-level imagery at scale is often impractical due to access limitations, cost, and environmental variability. In contrast, aerial and satellite imagery provide broader spatial coverage but lack the fine-grained details needed for many downstream applications. Combining images from multiple altitudes — from ground cameras to aerial drones and satellites—offers a promising solution to overcome these limitations, enabling richer, more complete 3D reconstructions. How can we achieve coherent and accurate 3D scene modeling when our visual world is captured from vastly different altitudes—ground, aerial, and satellite—under varying conditions? Each altitude offers distinct advantages, but cross-altitude data fusion introduces significant challenges: sparse and incomplete views, visual ambiguities, spatio-temporal inconsistencies, image quality variations, dynamic scene changes, and environmental factors that alter topology over time. Traditional 3D reconstruction methods, optimized for dense and structured inputs, struggle with such heterogeneous multi-altitude data. Advances in multi-scale feature alignment, neural scene representations, and robust cross-view fusion offer promising solutions, but key challenges remain.

Generative AI for Biomedical Image Analysis: Opportunities; Challenges and Futures

Organizers: Yuanfeng Ji, Zhongying Deng, Xiangde Luo, Jin Ye, Xiyue Wang, Dan Lin, Junjun He, Jianfei Cai, Angelica I Aviles-Rivero, Carola-Bibiane Schönlieb, Shaoting Zhang, Ping Luo

Date: 20-Oct

Time: 8:00 - 12:00

Location: 316 B

Summary: The GAIA (Generative AI for Biomedical Image Analysis) workshop at ICCV 2025 explores how generative AI is transforming medical imaging and healthcare. The workshop focuses on three key areas: (1) Data synthesis and clinical modeling using generative models for anatomically accurate image creation and disease simulation, (2) Multimodal learning that integrates visual data with medical reports through large language models, and (3) Workflow automation streamlining medical imaging from acquisition to diagnosis. Bringing together experts from computer vision, healthcare, and AI research, the workshop addresses challenges in interpretability, regulatory compliance, and clinical reliability while showcasing opportunities for interdisciplinary collaboration in advancing biomedical image analysis.

Generative AI for Storytelling

Organizers: Andrew Shin, Yusuke Mori, Hiroaki Yamane, Hana Kopecka, Hajime Murai, Xianchao Wu, Lin Gu, Haitao Yu

Date: 20-Oct

Time: 8:00 - 12:00

Location: 302 A

Summary: Generative AI excels at producing stunning visuals but often fails at creating coherent, engaging stories. Storytelling demands consistency in character, plot, and setting—areas where current models fall short. This workshop explores combining advanced visual models, large language models, and multi-modal AI to generate narratives that are visually consistent and compelling. Our goal is to push generative AI beyond impressive graphics, enabling it to deliver dynamic, cohesive stories and broaden its role in content creation.

Generative Scene Completion for Immersive Worlds

Organizers: Ethan Weber, Hong-Xing "Koven" Yu, Lily Goli, Alex Trevithick, Angjoo Kanazawa, Jiajun Wu, Norman Müller, Christian Richardt

Date: 20-Oct

Time: 8:00 - 12:00

Location: 301 B

Summary: This workshop focuses on generative scene completion, which is indispensable for world models, VR/AR, telepresence, autonomous driving, and robotics. It explores how generative models can help reconstruct photorealistic 3D environments from sparse or partial input data by filling in occluded or unseen spaces. Topics include world models, generative models, inpainting, artifact removal, uncertainty, controllability, and handling of casual data. We will discuss how related directions like text-to-3D and single-image-to-3D compare with scene completion, where more input constraints must be satisfied. The workshop highlights key challenges and recent progress in transforming incomplete real-world captures into immersive environments.

Large-scale Video Object Segmentation

Organizers: Henghui Ding, Lingyi Hong, Chang Liu, Ning Xu, Linjie Yang, Yuchen Fan

Date: 20-Oct

Time: 8:00 - 12:00

Location: 318 A

Summary: The 7th LSVOS Workshop focuses on advancing research in Video Object Segmentation (VOS) and Referring Video Object Segmentation (RVOS). The VOS track is based on the MOSE and LVOS datasets, targeting long-term videos and complex real-world scenes with challenges like object disappearance and reappearance, inconspicuous small objects, heavy occlusions, and crowded environments. The RVOS track continues with the MeViS dataset, which emphasizes motion-based language expressions and demands fine-grained temporal reasoning. In addition to the challenges, the workshop hosts invited talks from leading researchers, covering topics such as vision-and-language, motion understanding, cognitive modeling, and embodied intelligence in video understanding.

Mobile Intelligent Photography and Imaging

Organizers: Shangchen Zhou, Xiaoming Li, Zongsheng Yue, Kang Liao, Peiqing Yang, Jianyi Wang, Yuekun Dai, Yikai Wang, Xinyu Hou, Zhouxia Wang, Haoying Li, Ruicheng Feng, Yihang Luo, Chongyi Li, Chen Change Loy

Date: 20-Oct

Time: 8:00 - 12:00

Location: 327

Summary: Developing and integrating advanced image sensors with novel algorithms in camera systems is prevalent with the increasing demand for computational photography and imaging on mobile platforms. However, the lack of high-quality data for research and the rare opportunity for in-depth exchange of views from industry and academia constrain the development of Mobile Intelligent Photography and Imaging (MIPI). The workshop's main focus is on MIPI, emphasizing the integration of novel image sensors and imaging algorithms.

The 3rd workshop on Binary and Extreme Quantization for Computer Vision

Organizers: Adrian Bulat, Zechun Liu, Haotong Qin, Ioanna Ntinou, Nic Lane, Georgios Tzimiropoulos

Date: 20-Oct

Time: 8:00 - 12:00

Location: 308 A

Summary: The 3rd edition of the Workshop seeks to explore novel directions for making deep learning models more efficient. We'll delve into low-bit quantization, a technique that significantly reduces model size and computational demand by representing model weights and activations with fewer bits. This is crucial for deploying models on-device, especially given the ever-growing model size. A core focus of the workshop will be to study ways of maintaining accuracy under extreme quantization, with recent breakthroughs demonstrating exciting potential for achieving this. Hear about the latest trends from our invited speakers and presented papers.

The Second Workshop on Multimodal Representation and Retrieval

Organizers: Xinliang Zhu, Arnab Dhua, Shengsheng Qian, Xin (Eric) Wang, Rene Vidal, Douglas Gray

Date: 20-Oct

Time: 8:30 - 12:30

Location: 308 B

Summary: Multimodal data is available in many applications like e-commerce production listings, social media posts and short videos. However, existing algorithms dealing with those types of data still focus on uni-modal representation learning by vision-language alignment and cross-modal retrieval. In this workshop, we target to bring a new retrieval problem where both queries and documents are multimodal. With the popularity of vision language modeling, large language models (LLMs), retrieval augmented generation (RAG), and multimodal LLM, we see a lot of new opportunities for multimodal representation and retrieval tasks. This event will be a comprehensive workshop focusing on the subject of multimodal representation and retrieval.

The Third Workshop on AI for 3D Content Creation

Organizers: Hezhen Hu, Georgios Pavlakos, Despoina Paschalidou, Nikos Kolotouros, Davis Rempe, Angel X. Chang, Kai Wang, Amlan Kar, Kaichun Mo, Daniel Ritchie, Leonidas Guibas

Date: 20-Oct

Time: 8:00 - 12:00

Location: 315

Summary: Generating realistic 3D content has been a long-standing problem and graphics, which has recently attracted increasing attention. This workshop aims to bring together researchers to explore recent advances and future directions toward building fully controllable 3D content generation pipelines. We focus on four key aspects: (1) Representations suitable for generating high-quality and controllable 3D assets; (2) Modeling techniques that enable scalable, diverse, and photorealistic generation of humans, objects, and scenes; (3) Interaction modeling for capturing dynamic human-object relations with physical realism; (4) Applications of 3D content creation in areas such as embodied AI, construction and digital design.

What is Next in Multimodal Foundation Models?

Organizers: Roei Herzig, Rogerio Feris, David M. Chan, Leonid Karlinsky, Tsung-Han Patrick Wu, Jiaxin Ge, Dantong Niu, Eli Schwartz, Assaf Arbelle, Nimrod Shabtay, Bo Wu, Jehanzeb Mirza, Wei Lin

Date: 20-Oct

Time: 8:00 - 12:00

Location: 326 B

Summary: The intersection of foundation models and multimodal learning is a significant and widely discussed topic that complements the main ICCV conference. This workshop aims to encourage an interdisciplinary discussion on recent advancements, ongoing challenges, and future directions in multimodal foundation models, which have achieved breakthroughs by applying techniques across computer vision, natural language, and robotics.

Workshop on Advanced Perception for Autonomous Healthcare

Organizers: Ziyan Wu, Meng Zheng, Benjamin Planche, Zhongpai Gao, Anwesa Choudhuri, Terrence Chen

Date: 20-Oct

Time: 8:00 - 12:30

Location: 306 A

Summary: This workshop explores cutting-edge technologies, focusing on AI and computer vision, to advance autonomous, efficient, and patient-centered healthcare. It addresses challenges like medical errors and the need for precise diagnostics amid rapid technological advancements. The event facilitates collaboration among AI researchers, clinicians, and industry professionals through invited talks and interactive sessions. These cover theoretical and practical applications of visual perception technologies to enhance workflow efficiency, diagnostic accuracy, reduce errors, and improve patient care. By fostering partnerships, the workshop tackles issues like staff shortages and rising healthcare costs, promoting innovative solutions for a more effective healthcare system.

Workshop on Cultural Continuity of Artists: Leveraging Artistic Legacies for AI-Driven Cultural Heritage

Organizers: Taehoon Kim, Daiwon Hyun, Nojun Kwak, Youngjoon Yoo, Sangdoo Yun, Everine Jo

Date: 20-Oct

Time: 8:00 - 12:00

Location: 307 A

Summary: The Workshop on Cultural Continuity of Artists (WCCA) brings together researchers, creators, and cultural institutions to explore how computer vision, multimodal AI, and XR technologies can safeguard and reinterpret artistic legacies. Our inaugural edition, co-located with ICCV 2025, highlights the visionary South Korean fashion designer André Kim and introduces a rich, newly curated dataset from his archives.

Workshop on Distillation of Foundation Models for Autonomous Driving

Organizers: Burhan Yaman, Yunsheng Ma, Xin Ye, Can Cui, Mahmut Yurt, Selim Engin, Sungyeon Park, Deyuan Qu, Xu Cao, Wenqian Ye, Chun-Hao Liu, Qi Chen, Yezhou Yang, Ziran Wang

Date: 20-Oct

Time: 8:00 - 12:00

Location: 316 C

Summary: The 2nd Workshop on Distillation of Foundation Models for Autonomous Driving (WDFM-AD) focuses on advancing the state of the art in deploying large foundation models—such as vision-language models (VLMs) and generative AI (GenAI) models—into autonomous vehicles through efficient distillation techniques. Building on the success of our previous workshops on large language and vision models for autonomous driving, WDFM-AD aims to bring together researchers and industry professionals to explore innovative approaches that accelerate the safe, efficient, and scalable adoption of cutting-edge AI technologies in autonomous vehicles.

Workshop on Neuromorphic Vision (NeVi): Advantages and Applications of Event Cameras

Organizers: Federico Becattini, Gaetano Di Caterina, Yulia Sandamirskaya, Gregory Cohen, Luca Cultrera, Lorenzo Berlincioni, Suzanne Little, Joseph Lemley, Gaurvi Goyal, Arren Glover, Axel von Arnim

Date: 20-Oct

Time: 8:00 - 12:00

Location: 303 B

Summary: Neuromorphic vision sensors, or event cameras, mimic biological vision by asynchronously detecting changes in illumination, enabling high temporal resolution, low power consumption, and no motion blur. These unique features support advanced applications in robotics, autonomous vehicles, and human behavior analysis, especially for motion-centric tasks. Event cameras excel in low-light conditions and fast dynamics, enabling real-time obstacle avoidance, emotion recognition, defect detection, and more. Their microsecond latency and high dynamic range offer significant advantages over conventional cameras. Moreover, their inherent data sparsity contributes to privacy preservation.

Workshop & Competition on Computationally Optimal Gaussian Splatting

Organizers: Forrest Iandola, Zechun Liu, Cheng Chang, Karthik Ganesan, Kareem Ibrahim, Enrique Torres Sanchez, Hugo Tessier, Miloš Nikolić, Andreas Moshovos

Date: 20-Oct

Time: 8:30 - 12:00

Location: 325 B

Summary: The first Workshop & Competition on Computationally Optimal Gaussian Splatting (COGS) welcomes researchers working on techniques for efficient 3D Gaussian Splatting (3DGS). While 3DGS has advanced rapidly, real-time rendering of Gaussian splats on resource-limited devices (e.g., smartphones, AR/VR headsets) remains a challenge. COGS aims to lower the barrier to entry and encourage new research in this area. The event will feature keynote talks from leading experts and a panel session exploring the current state and future directions of this promising technique.

Human-inspired Computer Vision

Organizers: Lucia Schiatti, Mengmi Zhang, Yen-Ling Kuo, Vittorio Cuculo, Andrei Barbu

Date: 20-Oct

Time: 9:00 - 12:00

Location: 309

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 the high performance reached by recent computer vision approaches, the relationship between machine and human vision remains unclear. Investigating such a relationship is timely and important both to improve machine vision, by identifying and tackling gaps between humans and machines, and to understand/enhance human vision, by developing interpretable models useful to explain neuroscientific and cognitive observations.

1st Workshop and Challenge on Category-Level Object Pose Estimation for Robotic Manipulation

Organizers: Yang You, Jiyao Zhang, Jiankai Sun, Leonidas Guibas, Chen Wang, Luca Carlone, Linfang Zheng, Mac Schwager, Cewu Lu, Hao Dong, Bowen Wen, Ruida Zhang, Weiyao Huang, Mingdong Wu, Yijia Weng, Yitong Peng, Ruihai Wu, Lixin Yang, Junxiao Kong, Qiaojun Yu

Date: 20-Oct

Time: 8:00 - 17:00

Location: 324

Summary: This workshop addresses the critical problem of category-level object pose estimation and its applications within complex robotic manipulation scenarios. Pose estimation, a fundamental challenge in both 3D computer vision and robotics perception, involves accurately determining an object's complete 6-degree-of-freedom (6DoF) pose, comprising its 3D rotation and translation. Our workshop specifically focuses on advancing category-level pose estimation methods under realistic and demanding robotic manipulation settings, particularly emphasizing articulated objects, dynamic environments with potential human-object interactions, and objects subject to severe occlusions and partial visibility.

1st Workshop on Multimodal Sign Language Recognition

Organizers: Hamzah Luqman, Raffaele Mineo, Maad Alowaifeer, Simone Palazzo, Motaz Alfarraj, Mufti Mahmud, Amelia Sorrenti, Federica Proietto Salanitri, Giovanni Bellitto, Concetto Spampinato, Silvio Giancola, Muhammad Haris Khan, Moi Hoon Yap, Ahmed Abul Hasanaath, Murtadha Aljubran, Sarah Alyami, Egidio Ragonese, Gaia Caligiore, Sabina Fontana, Senya Polikovsky, Sevgi Gurbuz, Kamrul Islam

Date: 20-Oct

Time: 8:00 - 17:00

Location: 319 B

Summary: The 1st Multimodal Sign Language Recognition Workshop (MSLR 2025) convenes researchers to explore vision, sensor, and generative based approaches. Emphasizing multimodal fusion of RGB video, depth maps, skeletal and facial keypoints, and radar data, MSLR showcases systems optimized for real-world variability and privacy. Topics include statistical and neural sign-to-text/text-to-sign translation, cross-lingual methods, multilingual support, advanced generative synthesis, and inclusive dataset creation. Through keynotes, presentations, and challenges on continuous and isolated sign language recognition tasks, attendees will engage with novel benchmarks, metrics, and ethical data curation. Focus areas also encompass privacy-preserving sensing and healthcare accessibility. All backgrounds are welcome to contribute.

2nd AI for Content Generation; Quality Enhancement and Streaming

Organizers: Marcos V. Conde, Radu Timofte, Eduard Zamfir, Julian Tanke, Takashi Shibuya, Yuki Mitsufuji, Varun Jain, Fan Zhang, Heather Yu

Date: 20-Oct

Time: 8:00 - 17:00

Location: 317 B

Summary: Welcome to the 2nd Workshop on AI for Content Generation, Quality Enhancement and Streaming. This workshop focuses on unifying new streaming technologies, computer graphics, and computer vision, from the modern deep learning point of view. Streaming is a huge industry where hundreds of millions of users demand everyday high-quality content on different platforms. Computer vision and deep learning have emerged as revolutionary forces for rendering content, image and video compression, enhancement, and quality assessment. From neural codecs for efficient compression to deep learning-based video enhancement and quality assessment, these advanced techniques are setting new standards for streaming quality and efficiency. Moreover, novel neural representations also pose new challenges and opportunities in rendering streamable content, and allowing to redefine computer graphics pipelines and visual content.

2nd Workshop on Computer Vision for Ecology

Organizers: Sara Beery, Julia Chae, Mohamed Elhoseiny, Faizan Khan, Rupa Kurinchi-Vendhan, Andrew Temple, Edward Vendrow

Date: 20-Oct

Time: 8:00 - 17:00

Location: 323 B

Summary: The Computer Vision for Ecology workshop aims to bring together experts to foster discussion on the automation of ecological data collection, collation, and analysis. The goal is to establish a hub for the broader computer vision and ecology community at ICCV. The workshop encompasses applications of computer vision across a wide variety of ecological systems, spanning both terrestrial and aquatic systems, diverse geographic regions, and urban to wildland settings. 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 model deployment.

6th Workshop on Continual Learning in Computer Vision

Organizers: Jonghyun Choi, Marc Masana, Gido van de Ven, Liyuan Wang, Andrew D. Bagdanov, Evan Shelhamer, Dhireesha Kudithipudi

Date: 20-Oct

Time: 8:00 - 17:00

Location: 304 B

Summary: The Workshop on Continual Learning in Computer Vision (CLVision) aims to gather researchers and engineers from academia and industry to discuss the latest advances in Continual Learning. In this workshop, there will be regular paper presentations, invited speakers, and technical benchmark challenges to present the current state of the art, as well as the limitations and future directions for Continual Learning, arguably one of the most crucial milestones of AI.

9th AI City Challenge

Organizers: Zheng Tang, Shuo Wang, David C. Anastasiu, Ming-Ching Chang, Anuj Sharma, Norimasa Kobori, Jun-Wei Hsieh, Tomasz Kornuta, Rama Chellappa

Date: 20-Oct

Time: 8:00 - 17:00

Location: 313 A

Summary: The ninth AI City Challenge advanced real-world AI applications in transportation, automation, and safety, attracting 245 teams from 15 countries—a 17% increase. Featuring four tracks, the 2025 edition introduced challenges in 3D multi-camera tracking, traffic video question answering, warehouse spatial reasoning, and efficient fisheye camera detection. Tracks 1 and 3 utilized synthetic data from NVIDIA Omniverse. The evaluation platform ensured fair benchmarking with submission limits and held-out test sets. Public dataset releases reached over 30,000 downloads. Final rankings, announced post-competition, highlighted strong global participation and new benchmarks across multiple tasks, driving progress in intelligent visual perception and reasoning.

Advances in Image Manipulation Workshop and Challenges

Organizers: Radu Timofte, Andrey Ignatov, Marcos V. Conde, Dmitriy Vatolin, George Ciubotariu, Georgii Perevozchikov, Andrei Dumitriu, Florin Vasluianu, Chao Wang, Nikolai Karetin, Nikolay Safonov, Alexander Yakovenko

Date: 20-Oct

Time: 8:00 - 17:00

Location: 311

Summary: Image manipulation, restoration, and enhancement are key computer vision tasks serving as an important frontend of further tasks. Each step forward eases the use of images by people or computers. Not surprisingly then, there is an ever growing range of applications in fields such as surveillance, automotive, etc. or for mobile and wearable devices. 6th AIM workshop provides an overview of the advances in those areas, an opportunity for academic and industrial attendees to interact and explore collaborations. 32 papers and 18 associated competitions are gauging state-of-the-art on topics such as super-resolution, denoising, deblurring, ISP, segmentation, efficient models or quality assessment.

BioImage Computing

Organizers: Alexander Krull, Peter Bajcsy, Jan Funke, Dagmar Kainmueller, Khaled Khairy, Qingjie Meng, Virginie Uhlmann, Martin Weigert

Date: 20-Oct

Time: 9:00 - 17:00

Location: 325 A

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. 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.

Closing the Loop Between Vision and Language (Decade Mark)

Organizers: Mohamed Elhoseiny, Angel Chang, Anna Rohrbach, Marcus Rohrbach, Xin Eric Wang, Krishna Kumar, Kilichbek Haydarov, Eslam Abdelrahman, Austin Wang, Yiming Zhang, Tobias Wieczorek, Qianqi (Jackie) Yan

Date: 20-Oct

Time: 8:00 - 17:00

Location: 313 C

Summary: This workshop explores the intersection of Computer Vision and NLP, focusing on joint vision-language understanding. Recent advances, particularly in large-scale multimodal pretraining with transformers, have driven progress in various tasks. Topics include visual-linguistic representation learning, VQA, captioning, visual dialog, referring expressions, vision-and-language navigation, embodied QA, and text-to-image generation. We emphasize joint video-language understanding due to its unique challenges. Additionally, we welcome critical work on dataset and algorithmic bias, generalization issues, and efforts toward transparency and explainability.

Computer Vision for Biometrics; Identity & Behavior Science

Organizers: Abhijit Das, Mayank Vatsa, Richa Singh, Arun Ross, Vitomir Štruc, Antitza Dantcheva, Raghavendra Ramachandra

Date: 20-Oct

Time: 8:00 - 17:00

Location: 307 B

Summary: With growing global security concerns, biometric-based authentication and identification have become indispensable due to their reliability and robustness. Beyond physical biometrics, behavior understanding is emerging as a critical domain, aiming to interpret complex behavioral patterns that arise during interactions. Integrating both biometric and behavioral insights can lead to more secure, adaptive, and context-aware identity verification systems. Computer vision plays a pivotal role in analyzing and synthesizing biometric, identity, and behavior data. Recent advancements in behavior, biometric-based authentication and identification research, driven by deep learning and multimodal analysis, have significantly expanded the field. However, numerous challenges remain, including effective joint modeling of multimodal cues occurring at different time scales, handling the inherent uncertainty of machine-detectable behavioral evidence, and addressing long-term dependencies in human behavior and identity recognition. This workshop aims to bring together leading researchers, industry experts, and government agencies to discuss the latest breakthroughs in behavior, biometric-based authentication and identification research. It will serve as a platform to explore cutting-edge solutions, share innovative methodologies, and address the open challenges in this evolving field.

Computer Vision in Advertising and Marketing

Organizers: Alicja Kwasniewska, Subarna Tripathi, Maciej Szankin, Tz-Ying (Gina) Wu, Mateusz Ruminski, Sayantan Mahinder

Date: 20-Oct

Time: 8:00 - 17:00

Location: 303 A

Summary: The workshop will explore cutting-edge computer vision applications in digital advertising and marketing, covering fundamental visual understanding tasks, marketing optimization systems, brand intelligence, responsible AI practices, creative generation techniques, and emerging technologies that are transforming how brands connect with audiences through visual content. Key focus areas include multimodal data processing, visual similarity analysis, real-time bidding optimization, dynamic creative optimization, brand safety monitoring, and privacy-preserving analytics. The program will also address generative AI applications in advertising, automated visual optimization, and personalized content creation, while emphasizing ethical considerations and bias mitigation in marketing technology.

Ego-Exo Sensing for Smart Mobility

Organizers: Yi-Ting Chen, Katie Z Luo, Wei-Chiu Ma, Stephany Berrio Perez, Boris Ivanovic, Min-Hung Chen, Zhenzhen Liu, Zi-Hui Li

Date: 20-Oct

Time: 8:00 - 17:00

Location: 323 C

Summary: This workshop explores a holistic approach to ego-exo sensing, integrating vehicle sensors, roadside cameras, aerial imagery, and V2V communications to advance transportation intelligence and drive progress toward smart mobility. We examine how ego-exo sensing networks enhance safety-critical scenario detection and generation, comprehensive environmental perception, cooperative driving, and multi-agent decision making, among other crucial tasks shaping the future of mobility. This workshop bridges siloed research efforts to create unified approaches for heterogeneous sensor fusion that will define next-generation mobility systems.

Findings of the ICCV

Organizers: Margrit Betke, Yonatan Bisk, Juan C. Caicedo, Grigorios Chrysos, Trevor Darrell, Deepti Ghadiyaram, Boqing Gong, Derek Hoiem, Ziwei Liu, Bryan Plummer, Anna Rohrbach, Bryan Russell, Kate Saenko, Humphrey Shi, Kevin Shih

Date: 20-Oct

Time: 8:45 - 17:00

Location: 310

Summary: This workshop introduces the concept of a findings-style track to the computer vision community. NLP conferences have included Findings since 2020 to publish technically sound work, but which may not meet the main conference's threshold for novelty, impact, or excitement. There are many important results the community should be made aware of, and this venue provides an audience without delays for further submission iterations, or that might otherwise be lost if never published. This workshop provides a vehicle to discuss creating a computer vision Findings track and present Findings-quality papers to demonstrate their impact and benefits to inform future conferences.

Generating Digital Twins from Images and Videos

Organizers: Andrew Melnik, Chen Geng, Yujin Chen, Lei Ke, Qirui Wu, Jiayi Liu

Date: 20-Oct

Time: 8:00 - 17:00

Location: 319 A

Summary: In this workshop, we define Digital Twins as 3D models enriched with processes and semantic connections, similar to those in computer game environments. These models can range in fidelity from simplified 3D representations (Digital Cousins) to highly accurate reconstructions of real-world counterparts (Digital Twins). Our focus will be on methodologies and understanding required to improve the accuracy of these reconstructions and their associated processes using visual data, advancing from digital approximations to true Digital Twins. 3D Gaussian Splatting (3DGS) and diffusion models, have demonstrated impressive success in generating 3D representations from images and video. The next frontier in 3D representation is enriching models by integrating both physical and semantic object properties through generative AI and retrieval-based approaches.

Multi-Modal Reasoning for Agentic Intelligence

Organizers: Zhenfei Yin, Naji Khosravan, Tao Ji, Yin Wang, Roozbeh Mottagi, Iro Armeni, Zhuqiang Lu, Annie S. Chen, Yufang Liu, Zixian Ma, Mahtab Bigverdi, Amita Kamath, Chen Feng, Lei Bai, Gordon Wetzstein, Philip Torr

Date: 20-Oct

Time: 8:00 - 17:00

Location: 301 A

Summary: AI agents powered by Large Language Models (LLMs) have shown strong reasoning abilities across tasks like coding and research. With the rise of Multimodal Foundation Models (MFMs), agents can now integrate visual, textual, and auditory inputs for richer perception and decision-making. This workshop explores the development of Multimodal AI Agents across four categories: Digital, Virtual, Wearable, and Physical. We will discuss their applications in science, robotics, and human-computer interaction, as well as key challenges in cross-modal integration, real-time responsiveness, and interpretability. The goal is to advance robust, context-aware agents for complex, real-world environments.

Multimodal reasoning and slow thinking in large model era: towards system2 and beyond

Organizers: Chen Cheng, Chen Change Loy, David Clifton, Luc Van Gool, Shengwu Xiong, Peng Xu, Jiajun Zhang

Date: 20-Oct

Time: 8:00 - 17:00

Location: Ballroom B

Summary: This workshop aims to bridge the gap between computer vision and large language/reasoning models, focusing on complex tasks requiring advanced reasoning capabilities. We will explore how models can comprehend complex relationships through slow-thinking approaches like Neuro-Symbolic reasoning, Chain-of-Thought, and Multi-step Reasoning, pushing beyond traditional fixed tasks to understand object interactions within complex scenes. The goal is to bring together perspectives from computer vision, multimodal learning, and large language models to address outstanding challenges in multimodal reasoning and slow thinking in the context of large reasoning models, fostering more flexible and robust understanding in AI systems.

Robust and Interactable World Models in Computer Vision

Organizers: Shixiang Tang, Thu Nguyen-Phuoc, Zhenfei Yin, Amir Bar, Pengyu Zhang, Xu Jia, Yutong Bai, Lian Xu, Francesco Ferroni, Flora Salim, Tinne Tuytelaars, Jiajun Wu, Huchuan Lu, Yanyong Zhang, Philip H.S. Torr, Trevor Darrell

Date: 20-Oct

Time: 8:00 - 17:00

Location: 316 A

Summary: The workshop will focus on physical reliability and effective interactivity in world models for applications requiring precise physical reasoning and dense environmental interactions, such as robotics, autonomous systems, and multi-agent interactions. Beyond generating realistic predictions, world models must enforce physical consistency through differentiable physics, hybrid modeling, and adaptive simulation techniques. By bringing together researchers from machine learning, computer graphics, and physics-based modeling, the workshop will explore classical and cutting-edge approaches to aligning world models with real-world physics and extending them beyond simulation.

The 12th IEEE International Workshop on Analysis and Modeling of Faces and Gestures

Organizers: Joseph P. Robinson, Yun Fu, Sheng Li, Ming Shao, Yu Yin, Zhiqiang Tao

Date: 20-Oct

Time: 8:00 - 17:00

Location: 320

Summary: AMFG 2025 invites cutting-edge work in face, gesture, and multimodal recognition, where deep learning has unlocked unprecedented gains—but also raised concerns around generalization, transparency, and robustness. As models saturate benchmarks yet falter in real-world scenes with occlusion, motion, or lighting shifts, new challenges demand innovative solutions. Topics include detection and tracking, neural rendering, generative modeling, vision-language systems, kinship and soft biometrics, cross-modal fusion, benchmark creation, and ethical AI. With applications spanning HCI, surveillance, AR/VR, and behavioral science, AMFG aims to push beyond recognition into systems that interpret, adapt, and interact. Submit your work and shape the future of embodied vision.

Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities

Organizers: Yapeng Tian, Yuhang Zhao, Jon E. Froehlich, Chu Li, Yuheng Wu

Date: 20-Oct

Time: 8:00 - 17:00

Location: 318 B

Summary: The CV4A11y Workshop focuses on how new advances in vision foundation models and generative AI can help improve accessibility for people with disabilities. These technologies have great potential, but there are still important challenges, such as bias, limited data, lack of explainability, and real-world deployment issues. This workshop brings together experts in computer vision, AI, human-computer interaction, and accessibility to share new ideas, discuss open problems, and explore future directions. Our goal is to support the development of AI-powered tools that are more inclusive, useful, and effective in improving daily life for individuals with disabilities.

Vision-Language Modeling in 3D Medical Imaging

Organizers: Ibrahim Ethem Hamamci, Sezgin Er, Suprosanna Shit, Ezequiel De la Rosa, Anjany Sekuboyina, Murong Xu, Chinmay Prabhakar, Christian Bluethgen, Ayse Gulnihan Simsek, Omer Faruk Durugol, Sevval Nil Esirgun, Muhammed Furkan Dasdelen, Neslihan Simsek, Gulhan Ertan Akan, Mehmet Kemal Ozdemir, Melih Akan, Chenyu Wang, Weicheng Dai, Kayhan Batmanghelich, Xiaoman Zhang, Mohammed Baharoon, Luyang Luo, Pranav Rajpurkar, Pedro R. A. S. Bassi, Yixiong Chen, Wenxuan Li, Alan Yuille, Zongwei Zhou, Hadrien Reynaud, Bernhard Kainz, Chaoyi Wu, Weidi Xie, Benjamin Hou, Zhiyong Lu, Daguang Xu, Dong Yang, Pengfei Guo, Marc Edgar, Bjoern Menze

Date: 20-Oct

Time: 8:00 - 17:00

Location: 323 A

Summary: The VLM3D workshop brings together pioneers in vision-language modeling and 3D medical imaging to tackle the limitations of current models, which remain primitive and clinically unfit for real-world deployment. Through keynotes, discussions, and a dedicated benchmark challenge, we will explore why today's AI struggles with the complexity of 3D data and how to advance towards robust, deployable solutions. We aim to bridge the gap between research and clinical practice by defining critical next steps for generating reliable, interpretable, and clinically useful models. Join us to help shape the future of AI-driven 3D medical imaging.

Wild3D: 3D Modeling; Reconstruction; and Generation in the Wild

Organizers: Wei-Chiu Ma, Shenlong Wang, Yufei Ye, Linyi Jin, Lea Müller, Lingjie Liu, Despoina Paschalidou, Qixing Huang, Shubham Tulsiani, David Fouhey

Date: 20-Oct

Time: 8:00 - 17:00

Location: 312

Summary: Despite recent advances in 3D modeling, reconstruction, and generation, many methods remain limited to static scenes or dense viewpoints, making them less effective in real-world, dynamic, and often sparse or noisy settings. This workshop aims to gather researchers and practitioners focused on modeling, reconstructing, and generating dynamic 3D objects or scenes under challenging, in-the-wild conditions. Leveraging progress in 3D learning, the abundance of 2D/3D data, and powerful generative models, now is an opportune time to make 3D vision more robust and accessible. The workshop encourages contributions from standard 3D topics and broader 4D directions involving dynamics and video generation.

Workshop on Curated Data for Efficient Learning

Organizers: George Cazenavette, Kai Wang, Zekai Li, Xindi Wu, Tongzhou Wang, Peihao Wang, Ruihan Gao, Bo Zhao, Zhangyang Wang, Jun-Yan Zhu

Date: 20-Oct

Time: 8:00 - 17:00

Location: 304 A

Summary: The ICCV 2025 Workshop on Curated Data for Efficient Learning (CDEL) seeks to advance the understanding and development of data-centric techniques that improve the efficiency of training large-scale machine learning models. As model sizes continue to grow and data requirements scale accordingly, this workshop brings attention to the increasingly critical role of data quality, selection, and synthesis in achieving high model performance with reduced computational cost. Rather than focusing on ever-larger datasets and models, CDEL emphasizes the curation and distillation of high-value data—leveraging techniques such as dataset distillation, data pruning, synthetic data generation, and sampling optimization. These approaches aim to reduce redundancy, improve generalization, and enable learning in data-scarce regimes. The workshop will bring together researchers and practitioners from vision, language, and multimodal learning to share insights and foster collaborations around efficient, scalable, and sustainable data-driven machine learning.

Personalization in Generative AI Workshop

Organizers: Pinar Yanardag, Rinon Gal, Daniel Cohen-Or, Tuna Han Salih Meral, Enis Simsar, Nupur Kumari, Aysegul Dundar, Federico Tombari

Date: 20-Oct

Time: 8:00 - 17:00

Location: 317 A

Summary: Personalization in Generative AI Workshop (P13N) is a full-day workshop that brings together leading researchers and industry experts to explore cutting-edge personalization techniques in generative AI. The event will feature paper presentations, panel discussions, and a competition focusing on personalized generative models across images, and videos. Topics include advanced optimization methods for personalizing diffusion models, multi-subject composition, cross-modal personalization, AR/VR personalization, dataset curation and benchmarking, as well as ethical and privacy considerations.

1st Workshop on Long Multi-Scene Video Foundations: Generation; Understanding and Evaluation

Organizers: Vasco Ramos, Regev Cohen, Hila Chefer, Sivan Doveh, Jehanzeb Mirza, Hritik Bansal, Inbar Mosseri, Joao Magalhaes

Date: 20-Oct

Time: 13:00 - 17:00

Location: 303 B

Summary: This workshop aims to advance the state-of-the-art in long multi-scene video modelling, covering generation, understanding, evaluation, and ethical considerations. Long videos offer a powerful means of expression and communication, with applications in diverse fields such as entertainment, education, and health. However, current video generation and understanding techniques are typically confined to short, single-scene videos, limiting both our ability to create and comprehend complex video narratives. Thus, a growing need and research area is the development of methods for generating and understanding long-form videos of multiple dynamic scenes.

2nd Workshop on Audio-Visual Generation and Learning

Organizers: Shiqi Yang, Zhixiang Wang, Rodrigo Mira, Shoukang Hu, Vicky Kalogeiton, Stavros Petridis, Tae-Hyun Oh, Ming-Hsuan Yang

Date: 20-Oct

Time: 13:00 - 17:00

Location: 327

Summary: In this workshop, we aim to shine a spotlight on this exciting yet underinvestigated 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.

2nd Workshop on Scalable 3D Generation and 3D Geometric Scene Understanding

Organizers: Miaomiao Liu, Jose Alvarez, Mathieu Salzmann, Lingjie Liu, Hongdong Li, Richard Hartley

Date: 20-Oct

Time: 13:00 - 17:00

Location: 325 B

Summary: The objective of this workshop is to bring together engineers and researchers from academia and industry to discuss the current state-of-the-art methods in the field and challenges of computer vision for 3D scene generation, 3D scene reconstruction, 3D compositional scene geometric representation learning at a large scale. Moreover, this edition of workshop will also highlight 3D scene generation and understanding from multimodal data such as video, audio, and text, driven by its growing range of industrial applications.

Biometrics for Art

Organizers: Dzemila Sero, Estefanía Talavera, Tuğçe Arican, Katrien Keune, John Delanay, Karen Trentelman, Robert van Langh

Date: 20-Oct

Time: 13:00 - 17:00

Location: 307 A

Summary: The Biometrics for Arts (ArtMetrics) workshop aims to explore the intersection of biometrics, computer vision and the arts to provide a more nuanced understanding of an artwork's provenance and maker(s). In the same way as Biometrics serves as a tool for person identification from unique phenotypic or behavioural traits, Biometrics for Arts aims at artist recognition from unique attributes detected on works of art, thus fostering dialogue between engineers, computer scientists, heritage scientists, conservators, and art historians. This workshop will showcase innovative applications of computer vision in the visual art domain, emphasizing the role of technology in supporting conservation practices and enhancing the management of museum and private collections. Key topics include pattern recognition in works of art, AI-driven artistic generation, AI-driven analysis of multimodal imaging data of works of art, and digital restoration of different media. ArtMetrics seeks to inspire interdisciplinary collaboration, highlighting how computer vision can both interpret and enhance the diverse world of art.

Computer Vision in Plant Phenotyping and Agriculture

Organizers: Ian Stavness, Michael Pound, Feng Chen, Ronja Güldenring, Zane Hartley, Andrew French, Valerio Giuffrida

Date: 20-Oct

Time: 13:00 - 17:00

Location: 316 B

Summary: The CVPPA aims to advance computer vision techniques for applications in plant phenotyping and agriculture to support sustainable food, feed, fiber, and plant-based fuel production. The workshop seeks to highlight unsolved challenges, showcase current methods, and expand the research community at the intersection of plant and computer sciences. Topics include segmentation, tracking, detection, and reconstruction in agricultural contexts, open-source tools, and annotated datasets with benchmarks. 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.

Egocentric Body Motion Tracking; Synthesis and Action Recognition

Organizers: Lingni Ma, Yuting Ye, Robin Kips, Siyu Tang, Gen Li, Karen Liu, Boxiao Pan, Richard Newcombe

Date: 20-Oct

Time: 13:00 - 17:00

Location: Ballroom C

Summary: EgoMotion, in its second edition, is a continuation workshop focusing on human motion modeling using egocentric, multi-modal data from wearable devices. We focus motion tracking, synthesis, and understanding algorithms from egocentric/exocentric cameras, non-visual sensors, and high-level derived data. The workshop also covers research that applies egocentric motion for character animation, simulation, robotic learning etc. In addition to algorithms, the workshop promotes recent open-source projects, research platforms, datasets and associated challenges to encourage and accelerate research in the field. We will include live demo sessions to encourage discussions.

Human-Robot-Scene Interaction and Collaboration

Organizers: Jiayuan Gu, Xingyu Lin, Fangchen Liu, Yuexin Ma, Martin Magnusson, Sören Schwertfeger, Ye Shi, Hao Su, Jingya Wang, Lan Xu, Li Yi

Date: 20-Oct

Time: 13:00 - 17:00

Location: 315

Summary: Intelligent robots are advancing rapidly, with embodied agents increasingly expected to work and live alongside humans in households, factories, hospitals, schools, etc. For these agents to operate safely, socially, and intelligently, they must effectively interact with humans and adapt to changing environments. Moreover, such interactions can transform human behavior and even reshape the environment—for example, through adjustments in human motion during robot-assisted handovers or the redesign of objects for improved robotic grasping. Beyond established research in human-human and human-scene interactions, vast opportunities remain in exploring human-robot-scene collaboration. This workshop will explore the integration of embodied agents into dynamic human-robot-scene interactions.

International Workshop on Observing and Understanding Hands in Action

Organizers: Hyung Jin Chang, Rongyu Chen, Zicong Fan, Rao Fu, Kun He, Kailin Li, Take Ohkawa, Yoichi Sato, Linlin Yang, Lixin Yang, Angela Yao, Qi Ye, Linguang Zhang, Zhongqun Zhang

Date: 20-Oct

Time: 13:00 - 17:00

Location: 305 B

Summary: The ninth edition of this workshop will emphasize the use of multimodal LLMs for hand-related tasks. Multimodal LLMs have revolutionized the perceptions of AI, and demonstrated groundbreaking contributions to multimodal understanding, zero-shot learning, and transfer learning. Those models can process and integrate information from different types of hand data (or modalities), allowing the model to better understand complex hand-object/-hand interaction situations by capturing richer, more diverse representations.

Large Scale Cross Device Localization

Organizers: Zuria Bauer, Hermann Blum, Mihai Dusmanu, Linfei Pan, Qunjie Zhou, Marc Pollefeys
Date: 20-Oct
Time: 13:00 - 17:00
Location: 301 B

Summary: As computer vision moves into real-world use, robust localization across diverse devices is crucial. The CroCoDL workshop unites experts in vision, robotics, and AR to tackle cross-device, multi-agent localization. Focusing on 3D vision, visual localization, embodied AI, and AR/VR/MR, it bridges academic research and real-world deployment. The inaugural event features invited talks, papers, and a competition. It also introduces CroCoDL, a new large-scale benchmark with synchronized data from phones, headsets, and robots. By connecting efforts in structure-from-motion, neural rendering, and embodied AI, the workshop advances scalable localization across domains, sensors, and dynamic environments.

Multimodal Spatial Intelligence

Organizers: Songyou Peng, Jihan Yang, Kyle Genova, Thomas Funkhouser, Fei-Fei Li, Leonidas J. Guibas, Saining Xie
Date: 20-Oct
Time: 13:00 - 17:00
Location: 306 A

Summary: Our workshop will feature insightful keynote talks and a panel discussion around multi-modal spatial intelligence. Key topics include enhancing MLLMs' reasoning with images and 3D data, advancing 2D/3D perception, and enabling embodied AI. We will also delve into dynamic physical world modeling and critically examine the trust, ethics, and societal impact of these technologies. This workshop is a hub for advancing the future of spatially-aware AI, from core reasoning to real-world application and responsible deployment.

PHAROS - Adaptation; Fairness; Explainability in AI Medical Imaging

Organizers: Stefanos Kollias, Dimitrios Kollias, Xujiong Ye, Francesco Rundo
Date: 20-Oct
Time: 13:00 - 17:00
Location: Ballroom A

Summary: PHAROS-AFE-AIMI aims to present innovative approaches for predictive modeling using large medical image datasets, emphasizing deep learning models and transparent, human-centered integration of GenAI and LLMs in health services. It tackles key challenges at the intersection of computer vision and healthcare AI, including multidisease diagnosis, model explainability, fairness, domain adaptation, continual learning. With rising interest in trustworthy and interpretable AI, PHAROS-AFE-AIMI fosters discussion on responsible deployment in sensitive applications. PHAROS-AFE-AIMI is organised under PHAROS AI Factory, ensuring its topics having real-world relevance and strong foundation in cutting-edge research. Finally, the workshop includes two challenges (Multi-Source-Covid-19 Detection and Fair Disease Diagnosis).

Responsible Imaging

Organizers: Zhixiang Wang, Jian Wang, Yang Liu, Brandon Y. Feng, Zheng Wang, Yinqiang Zheng, Mingmin Zhao, Mohan Kankanhalli, Laura Waller
Date: 20-Oct
Time: 13:00 - 17:00
Location: 305 A

Summary: As imaging technologies advance, they surpass traditional capabilities, capturing and interpreting visual information beyond the limits of human perception. While these cutting-edge computational imaging systems push the boundaries of what can be seen and understood, they also quietly introduce critical ethical concerns related to privacy, safety, and robustness. Since these systems operate beyond human vision, many potential threats remain imperceptible, making them more difficult to detect and mitigate. This workshop aims to bring attention to these challenges and explore innovative solutions to deal with these challengings as imaging technologies push the boundaries of perceiving the invisible.

UniLight: Unifying Evaluation Metrics for Image-based Lighting; Relighting and Compositing

Organizers: Anand Battad, Jean-Francois Lalonde, Javier Vazquez-Corral, Roni Sengupta, Mathieu Garon, Yannick Hold-Geoffroy
Date: 20-Oct
Time: 13:00 - 17:00
Location: 318 A

Summary: Recent advancements in image editing applications such as relighting, compositing, harmonization, and virtual object insertion have opened up new horizons in visual media, augmented reality, and virtual production—especially with the rise of powerful image generative models. However, evaluating the quality of results for these applications is still a significant challenge. Traditional image quality metrics are not always effective in capturing the perceptual realism and subtle effects these technologies aim to achieve. Additionally, relying on user studies can be time-consuming and introduces variability, making comparing methods consistently challenging. To address these issues, this workshop explores and develops standardized evaluation metrics to bridge the gap between quantitative assessment and qualitative perception.

Visual object tracking and segmentation challenge workshop

Organizers: Matej Kristan, Jiři Matas, Alan Lukežič, Luka Čehovin Zajc, Michael Felsberg, Pavel Tokmakov, Hyung Jin Chang, Gustavo Fernández

Date: 20-Oct

Time: 13:00 - 17:00

Location: 302 A

Summary: The VOTS2025 workshop is the thirteenth annual benchmarking activity of the VOT initiative, which has successfully identified key trends in tracking research, most recently the rise of video segmentation models as a promising direction for general object tracking. Continuing to connect the tracking community, VOTS2025 pushes the boundaries of tracking research. The workshop will present results of 32 trackers from three sub-challenges, focusing on holistic targets, targets undergoing topological transformations, and real-time tracking. Additionally, the program features presentations of winning methods, a panel discussion, and keynotes outlining future directions in object tracking and video understanding.

Workshop on Computer Vision Systems for Document Analysis and Recognition

Organizers: Axel De Nardin, Silvia Zottin, Silvia Cascianelli, Alessio Fagioli, Marco Raoul Marini, Claudio Piciarelli, Romeo Lanzino, Luigi Cinque, Fabio Galasso, Rita Cucchiara, Gian Luca Foresti

Date: 20-Oct

Time: 13:00 - 17:00

Location: 308 A

Summary: In today's rapidly digitalizing world, the ability to analyze documents automatically is becoming increasingly important in our daily life. Document Analysis plays a growing role in both industrial and cultural contexts, highlighting the need for AI systems capable of handling highly diverse documents, presenting significant challenges. This workshop seeks to address these issues by fostering interdisciplinary collaboration. By bringing together researchers and professionals from different domains, it aims to facilitate knowledge exchange, promote innovation, and advance the development of intelligent, adaptable solutions for Document Analysis in a wide range of applications.

Workshop on Scene Graphs and Graph Representation Learning

Organizers: Azade Farshad, Maëlic Neau, Iro Armeni, Federico Tombari, Ehsan Adeli, Nassir Navab

Date: 20-Oct

Time: 13:30 - 17:30

Location: 314

Summary: The workshop focuses on the topic of scene graphs and graph representation learning for visual perception applications in different domains. Through a series of keynote talks, the audience will learn about defining, generating and predicting scene graphs, as well as about employing them for other tasks. Oral presentations of accepted submissions to the workshop will further enrich discussed topics with state-of-the-art advancements and engage the community. The objective is for attendees to learn about current developments and application domains of scene graphs and graph representation learning, as well as to draw inspiration and identify commonalities across these domains. Furthermore, this workshop will create an opportunity to discuss limitations, challenges, and next steps from research, practical, and ethical perspectives.

Critical Evaluation of Generative Models and their Impact on Society

Organizers: Noa Garcia, Amelia Katirai, Kento Masui, Mayu Otani, Yankun Wu

Date: 20-Oct

Time: 13:00 - 17:00

Location: 309

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.

TrustFM: Workshop on Trustworthy Foundation Models

Organizers: Chenliang Xu, Jure Leskovec, Dan Hendrycks, Jindong Wang, Lingjuan Lyu, Hangfeng He, Ting Wang, Zhiheng Li

Date: 20-Oct

Time: 13:00 - 17:00

Location: 308 B

Summary: Foundation models are revolutionizing the way we interact with AI—powering everything from search engines to scientific discovery. But as their reach expands, so do the risks. Can we truly trust these systems—before putting them to use? From AlexNet to LLaVA , the pace of innovation is staggering. Yet one thing remains constant: the urgent need for trustworthiness. In the foundation model era, we ask: What does trust mean at scale? Can classical insights still guide us? This workshop brings together researchers, engineers, and thought leaders to confront these challenges head-on. We'll explore how to create models that are not just powerful, but robust, fair, interpretable, and accountable.

Fairness and Ethics in AI: facing the ChalLEnge through Model Debiasing

Organizers: Vito Paolo Pastore, Enzo Tartaglione, Irina Voiculescu, Jaegul Choo, Vittorio Murino

Date: 20-Oct

Time: 8:00 - 12:00

Location: 302 B

Summary: Despite the unprecedented surge, the presence of biases within AI models is a critical concern, perpetuating disparities and ethical dilemmas. In the last years, the scientific community has increasingly focused on understanding and addressing model bias, as evidenced by a significant uptick in research across various disciplines. In this context, we present the second edition of the workshop FAILED. 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. Join us in this collaborative endeavor and be a part of this transformative journey!

Notes:

FLOOR PLAN LEVEL 1 Kamehameha Exhibition Hall and Lobby



LEGEND

- Parking
- Bicycle parking
- Escalator (2nd FL parking)
- Escalator (3rd & 4th FL)
- Elevator
- Accessible passenger loading zone
- Restroom (Men)
- Restroom (Women)
- Restroom (Universal)
- ATM
- Vending area
- Water fountain
- Bottle filler
- Nursing pod
- Defibrillator
- Smoking area
- LCD board
- Automatic entry door
- Entrance
- Wi-Fi area
- E-Bar
- 1801 CAFÉ + BAR

Complimentary Wi-Fi Area
Connect to : Hawaii Free Wi-Fi

HAWAI'I CONVENTION CENTER
Where Business and Aloha Meet

ASM GLOBAL

Hawai'i Convention Center
1801 Kalākaua Avenue, Honolulu, Hawai'i 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaiiconvention.com

HawaiiConventionCenter @HawaiiConventionCenter
@HIConventionCtr Hawaii-Convention-Center

v2025

FLOOR PLAN LEVEL 3 Meeting Rooms

Complimentary Wi-Fi Area
Connect to : Hawaii Free Wi-Fi



LEGEND

- Business center
- 808 Café
- Escalator (2nd FL parking)
- Escalator (3rd & 4th FL)
- Elevator
- Restroom (Men)
- Restroom (Women)
- Water fountain
- Bottle filler
- Defibrillator
- First aid
- Art display
- Smoking area
- LCD board
- Entrance
- Wi-Fi area
- Vending area

HAWAI'I CONVENTION CENTER
Where Business and Aloha Meet

ASM GLOBAL

Hawai'i Convention Center
1801 Kalākaua Avenue, Honolulu, Hawai'i 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaiiconvention.com

HawaiiConventionCenter @HawaiiConventionCenter
@HIConventionCtr Hawaii-Convention-Center

v2025

FLOOR PLAN LEVEL 4 Ballroom and Rooftop Garden



LEGEND

- Escalator (3rd & 4th FL)
- Elevator
- Restroom (Men)
- Restroom (Women)
- Water fountain
- Defibrillator
- Smoking area
- LCD board
- Wi-Fi area

ROOFTOP GARDEN FEATURES:

- 2.5 acre open air rooftop garden
- Floor Load Limit, 100 lbs per net sq. ft.
- Grassy stage area approx. 37 x 42 ft.

Complimentary Wi-Fi Area
Connect to : Hawaii Free Wi-Fi

HAWAI'I CONVENTION CENTER
Where Business and Aloha Meet

ASM GLOBAL

Hawai'i Convention Center
1801 Kalākaua Avenue, Honolulu, Hawai'i 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaiiconvention.com

HawaiiConventionCenter @HawaiiConventionCenter
@HIConventionCtr Hawaii-Convention-Center

v2025

Notes:

NURSING POD

The Hawaii Convention Center (HCC) now offers a **private, comfortable and clean space to pump or nurse**.

HCC's **ADA compliant Mamava lactation pod** is spacious, secure and offers a 60'' turnaround and grab bars for easy wheelchair access. The pod includes the following features:

- Secure access with Mamava app or access code
- Two benches
- Dual outlet and USB port
- Shelf, mirror, and coat hook
- Ceiling vents and exhaust fan
- Occupancy-activated lighting

Look for the nursing pod located at the **Lobby Level near Exhibit Hall 1**.

To unlock, users may download the Mamava app or use HCC's house phone (located beside the pod) and dial "0" to request for the access code.

For more information about the Mamava app, click HERE.









Information about the Mamava app



ICCV
OCT 19-23, 2025
HONOLULU
HAWAII