CVPR 2025

IEEE/CVF Conference on Computer Vision and Pattern Recognition

CVPR *Nashville* JUNE 11-15, 2025

Adobe · amazon · Apple · ByteDance

captions™ · GMI · Google · intel

Meta · Qualcomm · SONY · Tencent

GOLD SPONSORS

Baidu 百度 · JOHNS HOPKINS WHITING SCHOOL of ENGINEERING · Lambda

NVIDIA · Toloka · VOXEL51 · WAYVE

SILVER SPONSORS

ANT RESEARCH · atmanity · Aurora · fal · ENCORD

firstsource · helm.ai · kitware · lakeFS · Meituan

Motional · nexdata · OAK RIDGE National Laboratory · OPPO · Pinterest

sapien · scale · TOYOTA RESEARCH INSTITUTE · WAYMO · ZOOX

Welcome to the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Nashville, Tennessee! CVPR is the premier and flagship annual meeting of IEEE/CVF and PAMI- TC, where researchers in our community present their latest advances in computer vision, pattern recognition, machine learning, robotics, and artificial intelligence, both in theory and practice. Our program includes invited keynotes, oral and poster presentations, tutorials, workshops, demos, exhibitions, and social events, all aimed at providing attendees with an exciting and enriching experience. CVPR 2025 is primarily an in-person conference, but for those who are unable to join us physically, we are pleased to offer a virtual component that will provide access to conference papers, posters, videos, and talks.

CVPR 2025 received 13,008 valid paper submissions, a 13% increase from last CVPR. The review process was managed by the 6 Program Co-Chairs, 22 Senior Area Chairs, and 708 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 12,593 reviewers. As in prior years, after receiving these initial reviews, authors had the opportunity to submit a rebuttal to the reviews. The process concluded with discussions among reviewers and ACs, finalizing of reviews, and ACs working in triplets to make final accept/reject decisions for each paper. At the end of this process, 2,878 papers were accepted, for a 22.1% overall acceptance rate. In keeping with the CVPR tradition, the PCs did not pre-define any target acceptance rate or number of papers to be accepted; the resulting acceptance rate reflects the community consensus and is consistent with past CVPRs.

All of the 2,878 accepted papers were invited to present posters at CVPR. In addition, 96 (3.3%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 387 (13%) were selected by ACs together with Senior ACs to be "highlights" because of their high quality and potential impact. The oral presentations will follow a three-track configuration. The highlights are flagged with a special annotation in the program. New this year, papers authored by outstanding reviewers are also flagged with a special annotation. ACs nominated 15 papers to be best paper award candidates, from which a committee convened by the PCs selected the award winners to be announced during the conference.

Due to the dramatic increase in the number of submissions and concerns about the deteriorating quality of reviews, the PCs decided to make several major changes to the reviewing process compared to the previous CVPR editions. These changes include imposing a limit on the number of submissions per author, requiring authors to participate in the review process when deemed qualified by the PCs, and rejecting papers submitted by highly irresponsible reviewers.

We would like to thank everyone involved in making CVPR 2025 a success. This includes the organizing committee, the Senior Area Chairs, Area Chairs, and reviewers, the authors, the demo session participants, contributing artists, and the donors and exhibitors. It is exciting to be part of a team of people around the world united in a common goal.

We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, Yoshitomo Matsubara for his work as technical chair, Lee Campbell and the Event Hosts team for their work on the website and virtual platform, Mike Weil and Hall Erickson for handling sponsorships and the exhibition, and Luba Elliot as our AI Art Coordinator. Last but not least, we thank all of you for attending CVPR 2025 and making it one of the top venues for computer vision research in the world. We hope that you also have some time to explore Nashville during the conference.

Enjoy CVPR 2025. We look forward to meeting you in person!

CVPR Program Chairs
Phillip Isola *(Massachusetts Institute of Technology)*
Hedvig Kjellström *(KTH Royal Institute of Technology)*
Vincent Lepetit *(Ecole des Ponts ParisTech)*
Fuxin Li *(Oregon State University)*
Hao Su *(University of California, San Diego)*
Siyu Tang *(ETH Zürich)*

CVPR 2025 General Chairs
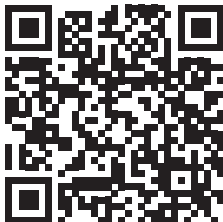Ce Liu *(Meta)*
Bryan Morse *(Brigham Young University)*
Cristian Sminchisescu *(Google DeepMind and Lund University)*

CVPR 2025 QR CODES

CVPR 2025 Virtual Platform	**CVPR 2025 Slido Site**	**Wayfinding App**	**Exhibit Hall Floorplan**
Access schedules, papers, workshops, tutorials, etc	Attendee engagement (Q&A, polls, etc.) for plenary sessions: keynotes, panels, orals	Find your way around the Music Center	Access location of exhibitors on the main floor

General Chairs
Ce Liu *(Meta)*
Bryan Morse *(Brigham Young University)*
Cristian Sminchisescu *(Google DeepMind and Lund University)*

Program Chairs
Phillip Isola *(Massachusetts Institute of Technology)*
Hedvig Kjellström *(KTH Royal Institute of Technology)*
Vincent Lepetit *(Ecole des Ponts ParisTech)*
Fuxin Li *(Oregon State University)*
Hao Su *(University of California, San Diego)*
Siyu Tang *(ETH Zurich)*

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

Workshop Chairs
Brian Clipp *(Kitware)*
Forrester Cole *(Google)*
Chen Sun *(Brown University)*
Lijuan Wang *(Microsoft)*

Tutorial Chairs
Katarina Doctor *(U.S. Naval Research Laboratory)*
Madhu Srinivasan *(AMD)*
Bolei Zhou (University of California, Los Angeles)

Demonstration Chairs
Rahul Garg *(Netflix)*
Neal Wadhwa *(Google)*

Finance Chair
Walter Scheirer *(University of Notre Dame)*

Publicity Chairs
Kosta Derpanis *(York University)*
Deblina Bhattacharjee *(University of Bath)*
Joakim Bruslund Haurum *(Aalborg University)*

Publications Chair
Eric Mortensen

Social Chairs
Vítor Albiero *(Meta)*
Yale Song *(Meta)*

Broadening Participation Chairs
Naresh Cuntoor *(BlueHalo Labs)*
Christopher Funk *(Kitware)*
Deepti Ghadiyaram *(Boston University)*
Roni Sengupta *(University of North Carolina at Chapel Hill)*
Saining Xie *(New York University)*

Senior PAMI-TC Ombud
David Forsyth *(University of Illinois at Urbana-Champaign)*
Linda Shapiro *(University of Washington)*

AI Art Curator
Luba Elliott *(Independent Curator)*

Technical Chair
Yoshitomo Matsubara *(Yahoo!)*

Doctoral Consortium Chairs
Nathan Jacobs *(Washington University in St. Louis)*
Abby Stylianou *(Saint Louis University)*

Web Developer
Lee Campbell *(Eventhosts)*

Conference Producer
Nicole Finn *(c to c events)*

To make it easier to navigate the growing number of CVPR workshops, we have grouped the workshops together into thematic "tracks." The workshops within each track cover closely related topics, and we've also tried our best to avoid having scheduling conflicts between them. We hope you enjoy the workshops!

Workshop Chairs
Brian Clipp, Forrester Cole, Chen Sun, Lijuan Wang

TABLE OF CONTENTS

Track on Computational Photography

Track on Computer Graphics

Track on Efficient Large / Edge Models

Track on Egocentric Vision

Track on Embodied AI

Track on Foundation Models

Track on Generative Models

Track on Human Computer Interfacing

Track on Human Modeling

Track on Image Classification / Recognition

Track on Medical

Track on Multimodal Learning

Track on Open World Learning

Track on Professional Topics

Track on Remote Sensing

Track on Responsible & Explainable Methods

Track on Retail / Consumer Products

Track on Robotic Manipulation

Track on Science

Track on Synthetic Data

Track on Video Understanding

Tuesday, June 10

11:00 –20:00   **Registration / Badge Pickup** (ExHall Concourse)

Wednesday, June 11

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 CVPR 2025 Tutorials page.

7:00 - 17:00   **Registration / Badge Pickup** (ExHall Concourse)

7:00 - 17:00   **Press Room** (203 B)

7:00 - 17:00   **Mother's Room** (Level 1 near Room 101 and on Level 3 near Exhibit Hall D)

7:00 - 17:00   **Prayer or Quiet Room** (203 A)

7:00 - 9:00   **Breakfast** (ExHall C)

8:00 - 18:00   **WORKSHOPS / TUTORIALS**

10:00 - 11:00   **Coffee Break** (ExHall D)

12:00 - 13:45   **Lunch** (ExHall C)

15:00 - 16:00   **Coffee Break** (ExHall D)

TUTORIALS

Foundations of Interpretable AI

Organizers: Rene Vidal, Jeremias Sulam, Aditya Chattopadhyay
Date: 6/11/2025
Time: 9:00 - 12:00
Location: 106 A

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.

Tackling 3D Deep Learning, Gaussian Splats and Physics Simulation with NVIDIA Kaolin Library, a Hands-On Lab

Organizers: Vismay Modi, Clement Fuji Tsang, Or Perel, Maria Shugrina
Date: 6/11/2025
Time: 9:00 - 12:00
Location: 107 B

Summary: 3D Deep Learning often demands extensive boilerplate work such as managing data, camera conventions, and visualizing novel 3D representations. NVIDIA's Kaolin Library, built on PyTorch, addresses these with tools like convenience APIs, reusable research modules, and GPU-optimized operations. The library's updates are designed to address the evolving needs of the research community. Recent examples include supporting emerging representations like 3D Gaussian Splats (3DGS), and physics-based simulations for dynamic 3D modeling. Initially developed for internal use, Kaolin is shared externally under an open-source license. The tutorial will provide hands-on coding experience to equip attendees with practical skills for using Kaolin. In this session, we explore interactive tools 3DGS viewing in Jupyter, how to create optimizable physical simulations, and finally convert between common 3D representations to export results. GPU back ends will be provided. By the end of the tutorial, attendees will be able to utilize Kaolin's features to streamline their research workflows and accelerate their projects.

Cognitive AI for the Future: Agentic Multimodal Models and RAG for Vision Language Applications, from Training to Deployment

Organizers: Zhuo Wu, Tiep Le, Gustavo A. Lujan, Yury Gorbachev, Adrian Boguszewski, Vasudev Lal and Raymond Lo
Date: 6/11/2025
Time: 9:00 - 12:00
Location: 401 AB

Summary: Cognitive AI represents a transformative leap in how machines understand and interact with the world. Despite its potential, practical challenges remain in making these systems accessible and applicable across diverse domains. This tutorial addresses how multimodal models [1], combined with Retrieval-Augmented Generation (RAG) [2] and agentic workflows [3], can enable cognitive AI systems to deliver personalized, context-aware solutions. With applications ranging from educational tools to assistive technologies for the elderly and disabled, this tutorial focuses on practical strategies for training, optimizing, and deploying these models and pipelines, making them both scalable and accessible to researchers and practitioners.

The 2nd Point Cloud Tutorial: All You Need To Know About 3D Point Cloud

Organizers: Xiaoyang Wu, Hengshuang Zhao, Li Fuxin, Julian Straub, Kaichun Mo

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 202 A

Summary: Point cloud is a data structure that is quite prevalent in 3D vision, which plays an important role in eras like 3D perception, 3D generation, autonomous driving, embodied AI, etc. However, there has not been a comprehensive resource that covers the state-of-the-art approaches and engineering details in point cloud processing. This tutorial aims to provide a comprehensive understanding of point cloud processing and analysis. Participants will delve into various aspects of point cloud data, exploring fundamental layers, network engineering considerations, pre-training technology, and acceleration library for point cloud processing. Through a combination of lectures, attendees will gain insights into the latest developments in the field and learn how to make informed choices when working with point cloud data. For the 2nd point cloud tutorial at CVPR 2025, we aim to move beyond traditional topics like backbone design and pre-training technologies covered in the 1st tutorial. This time, we will also explore challenges and opportunities in applications such as Autonomous Driving, Robotic Learning, and Egocentric Perception in AR/VR. With a diverse background spanning industry and academia, foundational research, and application-driven innovations, we offer a comprehensive perspective on the future of point cloud technology.

Scalable Generative Models in Computer Vision

Organizers: Nanye Ma, Oscar Michel, Saining Xie, Kaiming He, Jiatao Gu, Ruiqi Gao, Sherry Yang

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 202 B

Summary: Generative models have emerged as a transformative force in computer vision, enabling breakthroughs in image, video, and 3D content synthesis. Recent advancements in model architectures and generative frameworks have driven unprecedented scalability, allowing models to handle larger datasets, longer context lengths, and more complex distributions. This tutorial will provide a comprehensive discussion of these advancements, focusing on frontier techniques for scaling generative models and their applications to video synthesis, 3D reconstruction, and virtual world simulation. Attendees will gain insights into the design principles behind scalable models, learn about key technical innovations, and understand the broader implications for the future of computer vision. By addressing both theoretical and practical aspects, this tutorial aims to equip researchers with the knowledge to explore, build, and deploy next-generation scalable generative models.

From Video Generation to World Model

Organizers: Weichen Fan, Zhaoxi Chen, Ziwei Liu

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 204

Summary: In the past few years, the research community has witnessed remarkable advancements in generative models, especially in the realm of video generation. Generating compelling and temporal coherent videos is challenging but demanding. To overcome these challenges, early text-to-video (T2V) methods have explored the potential of text-to-image (T2I) pretraining, such as Make-A-Video, MagicVideo, and Lavie. With the success of Diffusion Transformers (DiT), the first T2V model, which can support generating up to 40 seconds and high-fidelity videos, named SORA, was proposed. The availability of large-scale high-quality video datasets are proved to be indispensable. Later methods, including CogVideoX and MovieGen, have further explored the potential of 3D VAE. However, the current largest T2V model still fails to maintain the physical standard in most of the generative videos. On the other hand, recent work such as Genie, Genie-2, and GameNGen has presented promising results towards action conditioned video generation, showing the great potential of controllable video generation toward world models. Thus, in this tutorial, we first would like to give a comprehensive background on text-to-video generation – by reviewing the previous and most recent advanced T2V methods. Then, we would like to discuss the connection, future directions, and potential solution from the current video generation model to the ultimate world model.

Volumetric Video in the Real World

Organizers: Amrita Mazumdar, Tianye Li, Sharath Girish, Aleksander Holynski, Shalini De Mello, David Luebke

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 205 B

Summary: Volumetric video, which encodes a time-varying 3D scene into a unified representation for novel-view synthesis of dynamic contents, has long been a grand challenge for achieving immersive experiences. High-quality volumetric video enables new and immersive applications, such as 3D video conferencing, 3D telepresence, and virtual tutoring in XR. Recent volumetric representation enables fast and high-quality reconstruction of dynamic 3D scenes.

As such, our tutorial summarizes practical challenges towards generating and distributing volumetric video in the real world. Specifically, invited talks in this tutorial will cover: (1) compression and performance optimization for 4D reconstruction, such as dynamic Gaussian splatting, quantization, and autoencoders; (2) volumetric video reconstruction from single or sparse-view captures; (3) reconstruction of indoor and urban scenes with dynamic content; (4) reconstruction and playback of dynamic 4D humans in the real-world; and (5) integration of volumetric video with vision-language models for other applications. Challenges across video domains, such as dynamic humans, automotive video, and synthetically generated video, will be thoroughly discussed.

Evaluating Large Multi-modal Models: Challenges and Methods

Organizers: Kaijie Zhu, Michael Saxon, Jindong Wang, Linyi Yang, Yue Feng, Yue Zhang

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 109

Summary: The proliferation of large multi-modal models (LMMs) has raised increasing concerns about their security and risks, which are mainly due to a lack of understanding of their capabilities and limitations. In this tutorial, our aim is to fill this gap by presenting a holistic overview of LMM evaluation. First, we discuss the recent advance of LMMs evaluation from the perspectives of what, where, and how to evaluate. Then, we present several key challenges in LMM evaluation such as data contamination and fixed complexity. Based on these challenges, we introduce how to overcome these challenges. Furthermore, our discussion covers key evaluation metrics including trustworthiness, robustness, and fairness, as well as performance across diverse downstream tasks in natural and social sciences. We conclude with an overview of widely-used code libraries and benchmarks that support these evaluation efforts. We hope that academic and industrial researchers continue to work to make LMMs more secure, responsible, and accurate. All materials will be available at https://LMM-understand.github.io/.

Evaluations and Benchmarks in Context of Multimodal LLM

Organizers: Hao Fei, Xiang Yue, Kaipeng Zhang, Long Chen, Jian Li, Xinya Du

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 208 A

Summary: Despite existing various emerging benchmarks for evaluating Multimodal Large Language Models (MLLMs), the evaluation of MLLMs validity and effectiveness might remain open to further discussion. This tutorial addresses the need for comprehensive and scientifically valid benchmarks in MLLM development. The tutorial will offer a systematic overview of current MLLM benchmarks and discuss necessary performance enhancements for achieving human-level AGI. We will introduce recent developments in MLLMs, survey benchmarks, and explore evaluation methods. Detailed discussions will cover vision-language capabilities, video modality evaluations, and expert-level skills across multiple disciplines. We'll further identify gaps in benchmarking the multimodal generalists, and introduce methods to comprehensively evaluate MLLMs. Finally, a special focus will be on addressing and mitigating the frequent hallucination phenomena in MLLMs to enhance model reliability. All the resources and materials will be made available online

Notes:

WORKSHOPS

LatinX in Computer Vision Research Workshop

Organizers: Lidia Talavera-Martínez, Willams de Lima Costa

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 105 A

Summary: This workshop aims to promote and increase the participation of the LatinX community in Computer Vision. The workshop will provide a platform for LatinX researchers at all levels to share academic, industrial, cultural, and social challenges; highlight prominent LatinX researchers and allies; offer resources and opportunities for career growth through sponsored registrations, mentoring, and resume sharing; and raise the visibility of women researchers within the LatinX community. While the event focuses primarily on researchers who identify as LatinX, everyone is invited to attend.

Global 3D Human Poses

Organizers: Tianjian Jiang, Manuel Kaufmann, Jie Song, Soyong Shin, Jiye Lee, Ye Yuan, Otmar Hilliges

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 105 B

Summary: The Global 3D Human Poses (G3P) workshop focuses on innovative techniques that incorporate trajectory data into pose estimation. By fostering collaboration among researchers and practitioners, the workshop will delve into new methodologies, address emerging challenges, and discuss the transformative potential of global pose estimation. Ultimately, the insights and innovations presented here are poised to push the boundaries of computer vision and pave the way for more robust, real-world applications in interactive systems and beyond.

8th Multimodal Learning and Applications Workshop

Organizers: Michael Ying Yang, Pietro Morerio, Paolo Rota, Bodo Rosenhahn, Vittorio Murino

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 106 B

Summary: The aim of this workshop is to generate momentum around multimodal learning and applications, and to encourage interdisciplinary interaction and collaboration between computer vision, multimedia, remote sensing, and robotics communities, that will serve as a forum for research groups from academia and industry. We expect contributions involving, but not limited to, image, video, audio, depth, IR, IMU, laser, text, drawings, synthetic, etc. Position papers with feasibility studies and cross-modality issues with highly applicative flair are also encouraged. Multimodal data analysis is a very important bridge among vision, multimedia, remote sensing, and robotics, therefore we expect a positive response from these communities.

5th Workshop on 3D Scene Understanding for Vision, Graphics, and Robotics

Organizers: Yixin Chen, Baoxiong Jia, Yao Feng, Songyou Peng, Chuhang Zou, Sai Kumar Dwivedi, Yixin Zhu, Siyuan Huang, Derek Hoiem, Marc Pollefeys, Song-Chun Zhu

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 106 C

Summary: The developments in AI technology have spurred calls for next-generation AI, e.g., Embodied AI and General AI, which enables systems to physically interact with their environments for comprehensive tasks in a human-like manner. However, new fundamental questions arise about how to sustain a more comprehensive understanding of the environment, unite these efforts, and facilitate the future development of next-generation AI. This year's focus will be exploring the fundamental aspects to enhance interaction between agents and 3D scenes in the new era of AI, promoting future directions and ideas to emerge within the next two to five years.

4th Edition of Computer Vision for Metaverse Workshop

Organizers: Giuseppe Serra, Ali Abdari, Alex Falcon, Beatrice Portelli, Vanessa Skliarova, Barbara Roessle, Daniel Sungho Jung, Shunlin Lu, Ji Hou, Bichen Wu, Djamila Aouada, Gyeongsik Moon

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 107 A

Summary: In the ever-growing areas of Augmented Reality (AR), Virtual Reality (VR), and the expansive Metaverse, computer vision brings together the digital and physical worlds seamlessly. Its ability to understand and interpret visual information pushes these immersive technologies to new levels, enhancing user experiences, driving creative innovations, and exploring new frontiers. On the other side, Natural Language Processing (NLP) is pivotal for deciphering human language and facilitating applications like translation and summarization. Large Language Models (LLMs) are now capable of human-level conversational skills, drastically enhancing human-machine interactions. As exemplified by CLIP and other multimodal foundational models, textual information plays a significant role in understanding visual data. Furthermore, as a consequence, these large models may contribute significantly to improving AR, VR, and the Metaverse, enabling hands-free navigation, voice-based commands, and immersive communication between avatars.

2nd MetaFood Workshop

Organizers: Yuhao Chen, Petia Radeva,
Jiangpeng He, Bhalaji Nagarajan,
Fengqing Zhu

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 108

Summary: Today, computer vision algorithms show near-perfect performance, better than human when there are clear, well curated and enough amount of data. However, there remains a substantial gap when it comes to applying state-of-the-art computer vision algorithms to food data, particularly when dealing with food in its natural, uncontrolled environment, often referred to as "data in the wild." This gap stems from the inherent challenges in noisy, watermarked, and low-quality food data readily available on the internet. The MetaFood Workshop (MTF) invites the CVPR community to engage with the food domain-related challenges. These challenges provide not only a demanding, real testing environment for the development of robust computer vision algorithms, but also an exciting opportunity to develop new algorithms in the fields of food data analysis and food digitization.

Computer Vision for Mixed Reality

Organizers: Rakesh Ranjan, Margarita Grinvald,
Vikas Chandra, Omer Shapira,
Andrea Colaco

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 109

Summary: With the advent of passthrough devices such as the Quest 3, Apple Vision Pro, and more recently, Orion AR glasses, users can now engage in deeply immersive experiences that blend the virtual and real worlds, often referred to as Mixed Reality (MR). Unlike traditional Virtual Reality (VR), MR presents unique challenges in computer vision, such as capturing and reconstructing real-world environments with high fidelity and augmenting them with virtual elements in a realistic manner, in real-time.

This workshop aims to provide the research community with a deeper understanding of these MR-specific challenges and explore novel methods in areas like view synthesis, scene understanding, and efficient on-device AI, among others. Attendees will benefit from the insights of a diverse committee with expertise in 3D computer vision, graphics, human visual perception, and efficient machine learning.

2nd GenAI Media Generation Challenge Workshop

Organizers: Sam Tsai, Ji Hou, Jialiang Wang,
Yaqiao Luo, Simran Motwani,
Xiaoliang Dai, Peizhao Zhang,
Kunpeng Li, Peter Vajda, Tao Xu,
Chih-Yao Ma

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 110 A
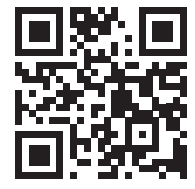
Summary: We are proud to announce the launch of the 2nd GenAI Media Generation Challenge (MAGIC), featuring a media generation track and auto-evaluation track: Media Generation Festival: For the first time, we are organizing a media generation festival with no restrictions on prompts. We would define a few different topics for which submitted media would compete in, and participants can submit their best generated videos or images for those specific topics. For each topic, we run a crowd sourced voting mechanism to determine the winners for each topic. Auto Evaluation Challenge: We are introducing an auto evaluation challenge for both text-to-image and text-to-video tasks. Participants can develop and submit their auto evaluation score for a preselect set of images and videos that we will provide and enter into the media generation festival track. Auto evaluation submissions would be to predict the outcomes from the crowd sourced voting mechanism in the media generation festival The auto evaluation method that achieves the best correlation with the final results would be the winners for this challenge.

Generalization in Robotics Manipulation Workshop and Challenges

Organizers: Shizhe Chen, Ricardo Garcia Pinel,
Jiafei Duan, Dieter Fox,
Cordelia Schmid, Ivan Laptev,
Sami Haddadin

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 110 B
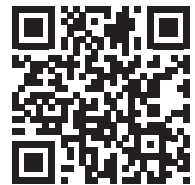
Summary: Robotic manipulation is one of the most fascinating and challenging problems in robotics, with broad applications in manufacturing, customer service, healthcare, household tasks and more. While learning-based visual policies have achieved impressive results such as manipulating Rubik's cubes, they are typically trained and tested in the same environments on specific tasks, lacking generalization capabilities to new scenes, objects and tasks. Recently, foundation models such as large language models (LLMs) and vision-language models (VLMs) have demonstrated strong abilities to encode vast amounts of world knowledge and generalize to new domains, offering a promising path forward for enhancing robots' generalization capabilities. In this workshop, we aim to unite researchers from different communities to push the boundaries of generalizable robotic manipulation, including foundation models, perception, planning, embodied AI, simulators, sim2real, among others.

Expanding Horizons in AI Benchmarking: Multimodal Approaches

Organizers: Laszlo Attila Jeni, Morteza Ziyadi, Hao Yang, Xu Zhang, Yang Zou, Zhaowei Cai, Maria Zontak, Davide Modolo, Ashwin Swaminathan, Liuyue Xie, Mosam Dabhi, Xiang Yue, Ce Zheng, Rohan Choudhury, Ananya Bal

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 202 C

Summary: The event will feature discussions highlighting the evolution and future of multimodal AI benchmarks, exploring methodologies and their real-world applications. It will conclude with a dynamic panel discussion, offering insights into the challenges and solutions for creating robust multimodal benchmarks and fostering interdisciplinary collaboration and innovation in AI benchmarking standards. This workshop aims to build a forum to discuss ongoing efforts in industry and academia, share best practices, and engage the community in working towards more comprehensive AI evaluation framework incorporating audio, visual, and textual inputs, addressing the limitations of current unimodal benchmarks.

8th Workshop on Efficient Deep Learning for Computer Vision

Organizers: Yung-Hsiang Lu, Shuai Zhang, George K Thiruvathukal
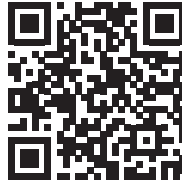
Date: 6/11/2025

Time: 9:00 - 12:00

Location: 207 A-D

Summary: Efficient computer vision on mobile, auto and edge devices significantly impacts daily life, technology, and industry. This workshop will explore the latest advancements in multimodal LLM, autonomous driving, Gaussian splatting avatars, and robotics. Additionally, discussions will delve into new optimization methods and applications, highlighting the 2025 IEEE Low Power Computer Vision Challenge (lpcv.ai), where winners of the three tracks will present their innovative solutions.

M&M: Multi-modal Models and Medicine

Organizers: Vishwesh Nath, Jeya Maria Jose Valanarasu, Zhihong Chen, Xueyan Mei, Weidi Xie, Vishal M. Patel, Bennett Allan Landman

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 208 A

Summary: Healthcare today stands at the intersection of technology and innovation, driven by diverse data sources—from clinical reports and electronic health records to medical imaging, vital signs, and numerous forms of unstructured data. While deep learning has significantly advanced medical imaging, the vast potential of integrating these abundant, multi-modal data streams remains largely untapped. This integration promises revolutionary improvements in patient outcomes, yet navigating this landscape poses unique and complex challenges due to the fragmented and isolated nature of healthcare data. This workshop addresses the critical questions facing researchers and practitioners: How can we effectively align and integrate multi-modal medical data? How do we tackle safety, privacy, interpretability, and the scarcity of clinically driven benchmarks?

The 2nd Workshop on Computer Vision for Videogames

Organizers: Iuri Frosio, Ekta Prashnani, David Durst, Raymond Rulon, Marguerite deCourcelle, Nicu Sebe, Georgios N. Yannakakis, Joohwan Kim

Date: 6/11/2025

Time: 9:00 - 12:00

Location: 210

Summary: Our aim is to bring together people working in Computer Vision (CV) and, more broadly speaking, Artificial Intelligence (AI), to talk about the adoption of CV/AI methods for videogames, that represents a large market share within the entertainment industry and a crucial domain for AI research. Our workshop will cover various aspects of videogames development and consumption, ranging from game creation, game servicing, player experience management, bot creation, cheat detection, and human computer interaction mediated by multimodal large language models. We believe that focusing on CV for videogames can bring the related research together cohesively for the foreseeable impact on the gaming market. Therfore, we plan to prioritize submissions that are specifically devoted to the application of state-of-the-art CV/AI methods FOR videogames, while assigning a lower priority to the submissions related to treating video games as test environments for CV/AI methods. We also intend to favor the presentation of novel datasets that can inspire further research in this field.

Sight and Sound

Organizers: Andrew Owens, Jiajun Wu, Kristen Grauman, Antonio Torralba, William T. Freeman, Andrew Zisserman, Hang Zhao, Ruohan Gao, Triantafyllos Afouras, Arsha Nagrani, Jean-Charles Bazin

Date: 6/11/2025
Time: 9:00 - 12:00
Location: 211

Summary: Since pretty much every video has an audio track, the prospect of learning from paired audio-visual data — either with new forms of unsupervised learning, or by simply incorporating sound data into existing vision algorithms — is intuitively appealing, and this workshop will cover recent advances in this direction. But it will also touch on higher-level questions, such as what information sound conveys that vision doesn't, the merits of sound versus other "supplemental" modalities such as text and depth, and the relationship between visual motion and sound. We'll also discuss how these techniques are being used to create new audio-visual applications, such as in the fields of speech processing and video editing.

The 2nd Workshop on Foundation Models for Medical Vision

Organizers: Jun Ma, Yuyin Zhou, Vishal M. Patel, Julia A Schnabel, Bo Wang

Date: 6/11/2025
Time: 9:00 - 12:00
Location: 212

Summary: The rapid growth of foundation models in various domains has been transformative, bringing unprecedented capabilities and advances in automated understanding. Medical vision, a pivotal segment of computer vision, is poised to greatly benefit from these advancements. This workshop delves into the integration and application of foundation models specific to the realm of medical imaging. We will cover state-of-the-art techniques for diverse medical data, such as echocardiogram, fundus, pathology, and radiology, as well as the practical challenges of implementing these models in clinical settings. Through expert-led sessions, interactive discussions, and international competitions, we aim to offer attendees a comprehensive understanding of the potential impact foundation models could have on the future of medical diagnostics and patient care.

The 6th Workshop on Fair, Data-Efficient, and Trusted Computer Vision

Organizers: Nalini K. Ratha, Srikrishna Karanam, Kuan-Chuan Peng, Mayank Vatsa, Richa Singh, Ziyan Wu, Michele Merler, Kush R. Varshney

Date: 6/11/2025
Time: 9:00 - 12:00
Location: Davidson C1

Summary: As humans and AI systems increasingly collaborate, it is critical that humans be able to trust AI. This workshop addresses this need by exploring research themes that enable trust, specifically fairness, explainability, and continuous learning from domain knowledge. It encompasses explainable AI, bias and scarcity detection, learning under bias and scarcity, and computational social science, aiming to foster cross-disciplinary research and discussion through a comprehensive treatment of these topics reflected in both paper submissions and invited keynote talks.

6th International Workshop on Large Scale Holistic Video Understanding

Organizers: Vivek Sharma, Shyamal Buch, Anurag Arnab, Ali Diba, Mohsen Fayyaz, Luc Van Gool, Joao Carreira, Manohar Paluri, Ehsan Adeli, Juergen Gall, David A Ross

Date: 6/11/2025
Time: 9:00 - 12:00
Location: Davidson C2

Summary: In recent years, the ability of computer systems to classify and analyze online videos has greatly improved. Significant advancements have been made in specific video recognition tasks, such as action and scene recognition. However, the comprehensive understanding of videos, known as holistic video understanding (HVU), has not received the attention it deserves. Current video understanding systems are specialized, focusing on narrow tasks.

For real-world applications like video search engines, media monitoring systems, and defining a humanoid robot's environment, integrating state-of-the-art methods is essential. To address this need, we are hosting a workshop focused on HVU. This workshop will cover recognizing scenes, objects, actions, attributes, and events in real-world videos.

We are introducing our HVU dataset, organized hierarchically with a semantic taxonomy for holistic video understanding. While many existing datasets focus on human action or sport recognition, our new dataset aims to broaden the scope and draw attention to the potential for more comprehensive video understanding solutions.

Our workshop will gather ideas related to multi-label and multi-task recognition in real-world videos, using our dataset to test and showcase research efforts.

10th New Trends in Image Restoration and Enhancement Workshop and Challenges

Organizers: Radu Timofte, Zongwei Wu, Florin-Alexandru Vasluianu, Yawei Li

Date: 6/11/2025

Time: 9:00 - 17:00

Location: Davidson C3

Summary: Image and video restoration, enhancement, and manipulation are key computer vision tasks with increasing importance across various fields.

The 10th edition of the NTIRE workshop seeks to provide a comprehensive overview of recent trends and advancements in these areas, facilitating interaction and potential collaboration between academic and industrial participants.

The NTIRE associated challenges gauge the state-of-the-art in topics such as super-resolution, efficiency, quality assessment, enhancement, normalization, removal of shadows, reflections and raindrops, HDR, light fields, raw restoration, reconstruction, event-based deblurring, cross-domain detection, depth estimation, night photography, and face restoration.

Building on the success of the previous editions, this event will feature presentations covering a wide selection of topics from 69 papers accepted for publication, organizers and winners of the 23 associated challenges, and invited talks provided by distinguished researchers.

3D Vision Language Model for Robotics Manipulation: Opportunities and Challenges

Organizers: Jiafei Duan, Muhammad Zubair Irshad, Ishika Singh, Vitor Campagnolo Guizilini, Rares Andrei Ambrus, Zsolt Kira

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 101 A

Summary: The intersection of 3D Vision-and-Language models (3D VLMs) in robotics presents a new frontier, blending spatial understanding with contextual reasoning. The Robo-3DVLM workshop seeks to explore the opportunities and challenges posed by integrating these technologies to enhance robot perception, decision-making, and interaction with the real world. As robots evolve to operate in increasingly complex environments, bridging the gap between 3D spatial reasoning and language understanding becomes critical. The workshop aims to drive conversations around the utility of 3D in robotic vision, the role of language in perception, and the limitations imposed by current data and hardware constraints. Through invited talks and interactive sessions, we aim to unite researchers from diverse disciplines to push the boundaries of multimodal learning in robotics, setting the stage for the next generation of intelligent systems.

4th Workshop on Computer Vision in the Wild

Organizers: Jianwei Yang, Chunyuan Li, Jiasen Lu, Reuben Tan, Qianhui Wu, Baolin Peng, Mu Cai, Xuehai He, Hao Zhang, Tianhe Ren, Feng Li, Shilong Liu, Xueyan Zou, Zhengyuan Yang, Xin Eric Wang, Yong Jae Lee, Lei Zhang, Jianfeng Gao

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 101 B

Summary: As artificial intelligence continues to evolve, the intersection of vision and language models is becoming increasingly crucial for real-world applications. The 4th Workshop on Computer Vision in the Wild (CVinW) at CVPR 2025 aims to foster discussions and innovations that push the boundaries of computer vision systems in unconstrained environments. Building on the success of our previous workshops: CVPR 2024 CVinW Workshop, CVPR 2023 CVinW Workshop and ECCV 2022 CVinW Workshop, this edition will focus on the next generation of large multimodal models (LMMs) and vision-language-action (VLA) systems, with an emphasis on temporal reasoning, video understanding, and physical interaction.

The 2nd Workshop on Equivariant Vision: From Theory to Practice

Organizers: Congyue Deng, Evangelos Chatzipantazis, Jiahui Lei, Yinshuang Xu, Stefanos Pertigkiozoglou, Minghan Zhu, Huazhe Xu, Thomas Mitchel, Leonidas Guibas, Kostas Daniilidis

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 101 C
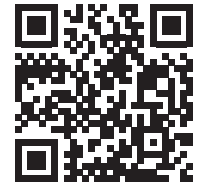
Summary: Exploiting symmetry in structured data is a powerful way to improve the generalization ability, data efficiency, and robustness of AI systems, which leads to the research direction of equivariant deep learning. Showing its effectiveness, it has been widely adopted in a large variety of subareas of computer vision, from 2D image analysis to 3D perception, as well as further applications such as medical imaging and robotics. The workshop will foster discussion and knowledge exchange among researchers actively working on equivariance, providing a platform to share methodologies and explore the latest advancements in this rapidly evolving field.

The 1st Workshop on Humanoid Agents

Organizers: Li Yi, He Wang, Wentao Zhu, Fangchen Liu, Bike Zhang, Koushil Sreenath, Yizhou Wang, Pieter Abbeel, Leonidas Guibas

Date: 6/11/2025
Time: 9:00 - 17:00
Location: 101 D

Summary: Building AI agents that behave just like human beings—i.e., Humanoid Agents—has long been a central goal in AI research. Recent advancements have brought us closer to this goal, whether through humanoid avatars in virtual worlds or humanoid robots in the real world. This workshop aims to facilitate the exchange of ideas, explore interdisciplinary intersections, and identify key challenges that can spark new insights and innovations. Additionally, it will serve as an opportunity to discuss ethical concerns and societal impacts, and strategies to address them through collective effort.

Navigating the Future: Ensuring Trustworthiness in Multi-Modal Open-World Intelligence

Organizers: Wei Ji, Hong Liu, Zhun Zhong, Zhe Zeng, Elisa Ricci, Andrew Gordon Wilson, Shin'ichi Satoh, Nicu Sebe

Date: 6/11/2025
Time: 9:00 - 17:00
Location: 101 E
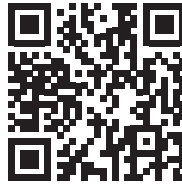
Summary: Today's interconnected world presents unique challenges for intelligent systems in processing and integrating diverse data modalities, including text, audio, and visual data. However, traditional closed-world paradigms can fall short when faced with unseen classes and novel scenarios, which frequently emerge in complicated real-world environments. We propose the consideration of open-world learning as a way to build intelligent systems that are highly adaptable while also being robust and trustworthy, capable of tackling highly dynamic and creative tasks. Here, the integration of privacy-preserving techniques is crucial as data sources expand, particularly in high-stakes applications such as autonomous navigation systems for public safety. These systems must discern and adapt to evolving traffic patterns, weather conditions, and user behaviors in real time, underscoring the necessity of continuous learning and resilience against adversities. By exploring these critical challenges, this workshop aims to foster discussions that advance the development of trustworthy, multi-modal systems capable of thriving in open-world contexts.

8th International Workshop on Visual Odometry and Computer Vision Applications Based on Location Clues

Organizers: Guoyu Lu, Friedrich Fraundorfer, Yan Yan, Nicu Sebe, Chandra Kambhamettu

Date: 6/11/2025
Time: 9:00 - 17:00
Location: 102 A

Summary: Visual odometry has attracted substantial interest in computer vision, robotics and mechanical engineering communities, to name a few. This workshop aims to foster scalable algorithms and systems for accurate and real-time visual odometry, addressing the growing demands of location-aware applications. It will explore methods and applications leveraging location cues to enhance scene understanding, city navigation, and other context-rich problems, while emphasizing visual odometry and localization in mobile and robotics domains.

Uncertainty Quantification for Computer Vision

Organizers: Andrea Pilzer, Martin Trapp, Arno Solin, Gianni Franchi, Andrei Bursuc, Marcus Klasson, Angela Yao, Tuan-Hung Vu, Fatma Guney

Date: 6/11/2025
Time: 9:00 - 17:00
Location: 102 B

Summary: The UNcertainty quantification for Computer Vision (UNCV) Workshop aims to raise awareness and generate discussion regarding how predictive uncertainty can, and should, be effectively incorporated into models within the vision community. At the time of Generative AI (GenAI) we find this more crucial than ever. The workshop will bring together experts from machine learning and computer vision to create a new generation of well-calibrated and effective methods that know when they do not know.

The 4th Workshop on Federated Learning for Computer Vision

Organizers: Chen Chen, Guangyu Sun, Nathalie Baracaldo, Yang Liu, Peter Richtárik, Mi Zhang, Lingjuan Lyu, Nicholas Donald Lane, Ang Li, Bo Li, Mahdi Morafah

Date: 6/11/2025
Time: 9:00 - 17:00
Location: 103 B

Summary: This workshop aims at bringing together researchers and practitioners with common interest in federated learning for computer vision. This workshop is an attempt at studying the different synergistic relations in this interdisciplinary area. This day-long event will facilitate interaction among students, scholars, and industry professionals from around the world to discuss the future research challenges and opportunities.

Mobile AI workshop and associated challenges, 5th edition

Organizers: Andrey Ignatov, Radu Timofte
Date: 6/11/2025
Time: 9:00 - 17:00
Location: 103 C

Summary: Over the past years, mobile AI-based applications are becoming more and more ubiquitous. Various deep learning models can now be found on any mobile device, starting from smartphones running portrait segmentation, image enhancement, face recognition and natural language processing models, to smart-TV boards coming with sophisticated image super-resolution algorithms. The performance of mobile NPUs and DSPs is also increasing dramatically, making it possible to run complex deep learning models and to achieve fast runtime in the majority of tasks. While many research works targeted at efficient deep learning models have been proposed recently, the evaluation of the obtained solutions is usually happening on desktop CPUs and GPUs, making it nearly impossible to estimate the actual inference time and memory consumption on real mobile hardware. To address this problem, we introduce the first Mobile AI Workshop, where all deep learning solutions are developed for and evaluated on mobile devices.

Data Driven Autonomous Driving Simulation

Organizers: Azadeh Dinparastdjadid, Zan Gojcic, Maximilian Igl, Maximilian Naumann, Thomas Gilles, Ekaterina Tolstaya, Sanja Fidler, Shimon Whiteson
Date: 6/11/2025
Time: 9:00 - 17:00
Location: 104 A

Summary: On-road testing of autonomous vehicles presents significant challenges in terms of cost and safety, underscoring the importance of simulation as a crucial tool to accelerate the development of safe autonomous driving (AD), a technology with enormous real-world impact. To extract useful information from simulation, minimizing the sim-to-real gap by developing good agent behavior models, and ensuring faithful perception simulation is essential. While recent years have witnessed a surge of publications in this rapidly evolving field, several fundamental questions remain unanswered, with the research often scattered across different robotics and machine learning venues and research fields. This workshop aims to unite leading researchers from various specializations, including perception simulation, behavior modeling, planner development, and safety research, to foster interdisciplinary conversations and collaboration.

Workshop on 4D Vision: Modeling the Dynamic World

Organizers: Shangzhe Wu, Qianqian Wang, Gengshan Yang, Jiahui Lei, Ruoshi Liu, Yufei Ye, Congyue Deng, Tarasha Khurana, Aleksander Holynski, Carl Doersch
Date: 6/11/2025
Time: 9:00 - 17:00
Location: 104 B
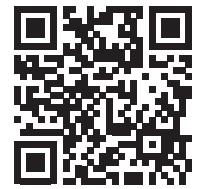
Summary: In recent years, we have seen remarkable progress in 3D computer vision, with increasingly robust and efficient models for reconstructing and generating 3D objects and scenes. 4D computer vision, as a natural extension of these efforts, is rapidly gaining traction. This workshop aims to establish a dedicated venue for discussions on this topic, bringing together researchers across various domains to exchange perspectives, identify challenges, and collectively accelerate progress in this space.

Visual Perception and Learning in an Open World

Organizers: Shu Kong, Neehar Peri, Yu-Xiong Wang, Andrew Owens, Abhinav Shrivastava
Date: 6/11/2025
Time: 9:00 - 17:00
Location: 104 C

Summary: Visual perception is crucial for a wide range of applications. Traditionally, visual perception models were developed under a closed-world paradigm, where data distributions and categorical labels were assumed to be fixed and known in advance. However, these closed-world models often prove brittle when deployed in the real open world, which is dynamic, vast, and unpredictable. Modern approaches to visual perception have shifted towards open-world models, such as pre-training foundation models on large datasets sourced from the open world (e.g., data collected from the Internet). These foundation models are then adapted to solve specific downstream tasks. While contemporary model training follows the principle of "open-world learning," our workshop seeks to address existing limitations, potential risks, new opportunities, and challenges.

2nd Workshop on Urban Scene Modeling: Where Vision meets Photogrammetry and Graphics (USM3D)

Organizers: Jack Langerman, Ruisheng Wang, Dmytro Mishkin, Ilke Demir, Renzhong Guo, Tolga Birdal, Sean (Xiang) Ma, Clement Mallet, Yang Wang, Shangfeng Huang

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 104 D

Summary: Classical 3D reconstruction has traditionally focused on low-level representations, and this workshop addresses the need for higher-level, structured and parametric representations like CAD models from images and point clouds, with implications for construction, manufacturing, urban planning, and related fields. The workshop aims to foster interdisciplinary collaboration between 3D vision researchers, photogrammetry, graphics, machine learning, and other domains where structured 3D representations are critical. To advance research in this area, the workshop introduces two large-scale datasets: S23DR, a collection of 3D models with corresponding multiview images, and Building3D, a city-scale dataset for building wireframe model generation from aerial LiDAR. By providing these resources and promoting collaboration, the workshop seeks to catalyze multi-view structured 3D reconstruction trends, bridge industry-academia gaps, and enable applications in urban planning, disaster management, and other critical areas.

12th Workshop on Fine-grained Visual Categorization

Organizers: Nico Lang, Elijah Cole, Suzanne Stathatos, Lukas Picek, Klara Janouskova, Christine Kaeser-Chen, Justin Kay, Joakim Bruslund Haurum, Xiangteng He, Mehmet Aygün, Serge Belongie, Oisin Mac Aodha, Subhransu Maji, Sara Beery, Grant Van Horn

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 104 E

Summary: FGVC12 will explore topics of broad interest to the computer vision community, specifically addressing self-supervision, limited data, and human-in-the-loop learning through the challenging lens of fine-grained learning. This focus extends beyond traditional computer vision, offering methodologies applicable to real-world scenarios in domains like ecology, biology, medicine, and art history, thus fostering participation from researchers outside the CVPR community. The workshop will feature innovative challenges, building upon successful past competitions like iNaturalist, which have previously introduced new datasets and fostered novel solutions. FGVC12 will feature not only leading researchers from the field of computer vision, but also experts from domains such as biomedical data science and ecology to promote discussion of open problems in these disciplines.

Computational Cameras and Displays

Organizers: Kristina Monakhova, Mark Sheinin, Fei Xia, Vishwanath Saragadam

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 205 A
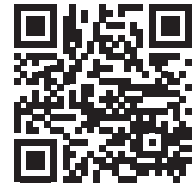
Summary: This workshop is designed to unite the computational camera and display communities in that it considers to what degree concepts from computational cameras can inform the design of emerging computational displays and vice versa, both focused on applications in computer vision. The Computational Cameras and Displays (CCD) workshop series serves as an annual gathering place for researchers and practitioners who design, build, and use computational cameras, displays, and imaging systems for a wide variety of uses. The workshop solicits posters and demo submissions on all topics relating to computational imaging systems.

CV4Science 2025: Using Computer Vision for the Sciences

Organizers: Utkarsh Mall, Ye Zhu, Jacob Berv, Siavash Golkar, Katherine Bouman, Subhransu Maji, David Fouhey

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 205 C

Summary: This workshop aims to: bring together researchers working on computer vision and diverse scientific domains to discuss the latest advancements, challenges, and opportunities at their intersections. The goal is to foster interdisciplinary collaboration, build community within the computer vision community, and highlight progress and researchers at the interface of computer vision and the sciences. AI advancements have become a transformative force, extending beyond their original domain to drive breakthroughs in scientific discovery—an impact highlighted by the 2024 Nobel Prizes in Physics and Chemistry. Computer vision, as one of the core areas in AI research, offers powerful tools for analyzing data, with applications spanning a wide range of scientific fields, from accelerating discoveries in astrophysics and biology to enhancing environmental monitoring and materials science.

EarthVision: Large Scale Computer Vision for Remote Sensing Imagery

Organizers: Ronny Hänsch, Devis Tuia, Jan Dirk Wegner, Loic Landrieu, Charlotte Pelletier, Hannah Kerner, Nathan Jacobs

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 208 B

Summary: Earth Observation (EO) and remote sensing are ever-growing fields of investigation where computer vision, machine learning, and signal/image processing meet. The general objective of the domain is to provide large-scale and consistent information about processes occurring at the surface of the Earth by exploiting data collected by airborne and spaceborne sensors. Earth Observation covers a broad range of tasks, from detection to registration, data mining, and multi-sensor, multi-resolution, multi-temporal, and multi-modality fusion and regression, to name just a few. It is motivated by numerous applications such as location-based services, online mapping services, large-scale surveillance, 3D urban modeling, navigation systems, natural hazard forecast and response, climate change monitoring, virtual habitat modeling, food security, etc. The sheer amount of data calls for highly automated scene interpretation workflows.

Demographic Diversity in Computer Vision

Organizers: Polina Kirichenko, Vikram V. Ramaswamy, Kyle Buettner, Sina Malakouti, Tarun Kalluri, Manmohan Chandraker, Adriana Kovashka, Olga Russakovsky

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 213
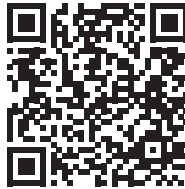
Summary: AI systems should serve all people with diverse values and perspectives around the world. However, as datasets scale, it's widely documented that they exhibit social biases of various forms, which translate to AI systems that cause real-world harm to under-represented demographic groups. A focused investigation of demographic biases in modern foundation models, their real-world impact and mitigation is thus critical to ensure equitable access to future models and their applications. This workshop will provide a shared discussion forum between interdisciplinary researchers in machine learning, computer vision, algorithmic fairness, computational ethics, policy of machine learning and human-computer interaction, to foster a common understanding of present demographic biases in foundation models, their impact and mitigation strategies. We encourage submissions and will facilitate discussions on diverse aspects of demographic fairness including geographic, cultural, linguistic, socioeconomic, ethnic, gender, race, and disability status, among others.

Foundation Models for Embodied Agents

Organizers: Manling Li, Ruohan Zhang, Jiayuan Mao, Wenlong Huang, Qineng Wang, Weiyu Liu, Xiaohan Zhang, Yonatan Bisk, Shenlong Wang, Yunzhu Li, Li Fei-Fei, Jiajun Wu

Date: 6/11/2025

Time: 9:00 - 17:00

Location: 214

Summary: We will cover the advances of the foundation models into Large Language Models Vision-Language Models, and Vision-Language-Action Models. In this workshop, 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 present a structured view to investigate the robot's decision-making process.

Workshop on Video Large Language Models

Organizers: Mubarak Shah, Larry S. Davis, Rene Vidal, Son Tran, Angela Yao, Salman Khan, Rita Cucchiara, Cees G. M. Snoek, Christoph Feichtenhofer, Chang Xu, Jayakrishnan Unnikrishnan, Afshin Dehghan, Mamshad Nayeem Rizve, Rohit Gupta, Swetha Sirnam, Ashmal Vayani, Omkar Chakradhar Thawakar, Muhammad Uzair Khattak, Dmitry Demidov

Date: 6/11/2025

Time: 9:00 - 17:00

Location: Grand A1

Summary: This workshop will explore the evolution, applications, and challenges of Video Large Language Models (VidLLMs), the latest advancement in multimodal LLMs. It will feature keynote talks from leading researchers, a panel discussion comparing VidLLMs with expert models, and a poster session. The workshop also includes three challenge tracks designed to evaluate VidLLMs' capabilities in compositional video retrieval, complex video reasoning and robustness, and multilingual video reasoning. These tracks aim to address key research areas such as training VidLLMs, their application in specialized computer vision tasks, and the challenges in evaluating their performance. Potential topics for invited papers include VidLLM methods/algorithms, data creation, evaluation and analysis, best practices, applications, and limitations, risks and safety.

3rd Workshop on Generative Models for Computer Vision

Organizers: Adam Kortylewski, Fangneng Zhan,
Tian Han, Alan Yuille,
Christian Theobalt
Date: 6/11/2025
Time: 9:00 - 17:00
Location: Grand A2

Summary: This workshop aims to foster collaboration between researchers in generative AI and computer vision to explore how visual recognition can benefit from recent advances in generative image modeling. The workshop will feature expert discussions on research results and future directions, specifically focusing on topics such as generative models as data source for training computer vision models, benchmarking with generative models, analysis-by-synthesis approaches, self-supervised learning, adversarial robustness, out-of-distribution generalization, and ethical considerations within generative modeling.

Embodied Intelligence for Autonomous Systems on the Horizon

Organizers: Hongyang Li, Kashyap Chitta,
Andrei Bursuc, Christos Sakaridis,
Jonah Philion, Florent Bartoccioni,
Ana-Maria Marcu, Huijie Wang
Date: 6/11/2025
Time: 9:00 - 17:00
Location: Grand B1

Summary: Autonomous systems, such as robots and self-driving cars, have rapidly evolved over the past decades. Despite this, several problems remain. Attempts have been made to develop more capable autonomous systems, such as integrating foundation models and utilizing large-scale data. However, the challenging problems have yet to be solved.

The motivation behind this workshop is to explore potential solutions, and discuss the challenges and opportunities associated with these approaches. We believe that this workshop serves as a brand-new perspective on the present and future of autonomous systems, and is necessary for both the robotics and computer vision communities.

Workshop on Autonomous Driving

Organizers: Vincent Casser, Alexander Liniger,
Jose M. Alvarez, Maying Shen,
Jannik Zürn, Chiyu Max Jiang,
Nadine Chang, Dragomir Anguelov,
John Leonard, Luc Van Gool
Date: 6/11/2025
Time: 9:00 - 17:00
Location: Grand C1

Summary: The CVPR 2025 Workshop on Autonomous Driving (WAD) brings together leading researchers and engineers from academia and industry to discuss the latest advances in autonomous driving. Now in its 8th year, the workshop has been continuously evolving with this rapidly changing field and now covers all areas of autonomy, including perception, behavior prediction and motion planning. In this full-day workshop, our keynote speakers will provide insights into the ongoing commercialization of autonomous vehicles, as well as progress in related fundamental research areas. Furthermore, we will host a series of technical benchmark challenges to help quantify recent advances in the field, and invite authors of accepted workshop papers to present their work.

Synthetic Data for Computer Vision Workshop

Organizers: Jieyu Zhang, Cheng-Yu Hsieh,
Zixian Ma, Rundong Luo,
Shobhita Sundaram, Wei-Chiu Ma,
Ranjay Krishna
Date: 6/11/2025
Time: 9:00 - 17:00
Location: Grand C2

Summary: The workshop aims to explore the use of synthetic data in training and evaluating computer vision models, as well as in other related domains. During the last decade, advancements in computer vision were catalyzed by the release of painstakingly curated human-labeled datasets. Recently, people have increasingly resorted to synthetic data as an alternative to laborintensive human-labeled datasets for its scalability, customizability, and costeffectiveness. Synthetic data offers the potential to generate large volumes of diverse and high-quality vision data, tailored to specific scenarios and edge cases that are hard to capture in real-world data. However, challenges such as the domain gap between synthetic and real-world data, potential biases in synthetic generation, and ensuring the generalizability of models trained on synthetic data remain. We hope the workshop can provide a forum to discuss and encourage further exploration in these areas.

Pixel-level Video Understanding in the Wild Challenge

Organizers: Henghui Ding, Nikhila Ravi,
Yunchao Wei, Jiaxu Miao,
Zongxin Yang, Yi Yang, Si Liu,
Yi Zhu, Elisa Ricci, Cees G. M. Snoek,
Song Bai, Philip Torr
Date: 6/11/2025
Time: 13:00 - 17:00
Location: 105 A

Summary: Pixel-level Scene Understanding is one of the fundamental problems in computer vision, which aims at recognizing object classes, masks and semantics of each pixel in the given image. Since the real-world is actually video-based rather than a static state, learning to perform video segmentation is more reasonable and practical for realistic applications. To advance the segmentation task from images to videos, we will present new datasets and competitions in this workshop, aiming at performing the challenging yet practical Pixel-level Video Understanding in the Wild (PVUW).

2nd Workshop on Neural Fields Beyond Conventional Cameras

Organizers: Ilya Chugunov, Tzofi Klinghoffer, Shengyu Huang, Wenzheng Chen, Daniel Gilo, Akshat Dave, Lingjie Liu, David B. Lindell, Or Litany, Ramesh Raskar

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 106 A

Summary: Neural fields have been widely adopted for learning novel view synthesis and 3D reconstruction from RGB images by modelling transport of light in the visible spectrum. This workshop focuses on neural fields beyond conventional cameras, including (1) learning neural fields from data from different sensors across the electromagnetic spectrum and beyond, such as lidar, cryo-electron microscopy (cryo-EM), thermal, event cameras, acoustic, and more, and (2) modelling associated physics-based differentiable forward models and/or the physics of more complex light transport (reflections, shadows, polarization, diffraction limits, optics, scattering in fog or water, etc.). Our goal is to bring together a diverse group of researchers using neural fields across sensor domains to foster learning and discussion in this growing area.

The 2nd Workshop on Computer Vision with Humans in the Loop

Organizers: Lei Zhang, Gang Hua, Nicu Sebe, Kristen Grauman, Yasuyuki Matsushita, Aniruddha Kembhavi, Ailing Zeng, Jianwei Yang, Xi Yin, Dongdong Chen, Heung-Yeung Shum

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 106 B

Summary: The ultimate goal of computer vision is to enable computers to perceive the world like humans. While the research community is still pursuing foundation models, we want to emphasize the importance of having humans in the loop to solve computer vision. For real-world applications and from a system perspective, computer vision systems need to be self-aware, meaning that the systems need to know when they do not know. Since the ultimate objective of visual perception is to facilitate downstream decisions, such self-awareness is very important as it endows the systems the capability to actively query humans for input or actively prompt for human control (such as in the Tesla Autopilot scenario). Considering both its importance and the recent research trend, we have organized this workshop to bring researchers, practitioners, and enthusiasts to explore and discuss the evolving role of human feedback in solving computer vision problems.

Multimodal Foundation Models for Biomedicine: Challenges and Opportunities

Organizers: Xiaohan Wang, Yuhui Zhang, Josiah Aklilu, Chenwei Wu, Zhi Huang, Liyue Shen, Hoifung Poon, Serena Yeung-Levy

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 107 A

Summary: Biomedical data spans diverse modalities across biological scales—from molecular genomics and cellular microscopy to tissue pathology, organ-level radiology, and patient-level electronic health records. While each modality provides unique insights, integrating these heterogeneous data sources remains a significant challenge in creating comprehensive biomedical understanding. The Multimodal Foundation Models for Biomedicine (MMFM-BIOMED) workshop brings together experts across disciplines to tackle this challenge.

The 4th Explainable AI for Computer Vision (XAI4CV) Workshop

Organizers: Sukrut Rao, Indu Panigrahi, Sunnie S. Y. Kim, Vikram V. Ramaswamy, Rajat Sahay, Avinab Saha, Dahye Kim, Miguel Ángel Fernández Torres, Lenka Tětková, Teresa Dorszewski, Bartlomiej Sobieski, Marina Gavrilova, Yuhui Zhang, Pushkar Shukla

Date: 6/11/2025

Time: 13:00 - 17:00

Location: 107 B

Summary: Explainability of computer vision systems is critical for people to effectively use and interact with them. This workshop provides a forum for researchers and practitioners to discuss the challenges and opportunities in explainable AI (XAI) for CV, addressing a critical need given the increasing deployment of these systems by: (1) initiating discussions across researchers and practitioners in academia and industry to identify successes, failures, and priorities in current XAI work; (2) examining the strengths, weaknesses, and underlying assumptions of proposed XAI methods and establish best practices in evaluation of these methods; and (3) discussing the various nuances of explainability and brainstorm ways to build explainable CV systems that benefit all involved stakeholders.

Image Matching: Local Features and Beyond

Organizers: Fabio Bellavia, Jiri Matas, Dmytro Mishkin, Luca Morelli, Fabio Remondino, Amy Tabb, Eduard Trulls, Kwang Moo Yi

Date: 6/11/2025
Time: 13:00 - 17:00
Location: 108
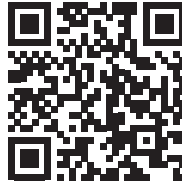
Summary: Matching two or more images across wide baselines is a core computer vision problem, with applications to stereo, 3D reconstruction, re-localization, SLAM, and retrieval, among many others. Until recently one of the last bastions of traditional handcrafted methods, they too have begun to be replaced with learned alternatives. Interestingly, these new solutions still rely heavily on design intuitions behind handcrafted methods. In short, we are clearly in a transition stage, and our workshop, held every year at CVPR since 2019, aims to address this, bringing together researchers across academia and industry to assess the true state of the field. We aim to establish what works, what doesn't, what's missing, and which research directions are most promising, while focusing on experimental validation.

Workshop on Human Motion Generation

Organizers: Rishabh Dabral, Guy Tevet, Amit Haim Bermano, Sigal Raab, Christian Theobalt, Chuan Guo, Peizhuo Li, Ikhsanul Habibie

Date: 6/11/2025
Time: 13:00 - 17:00
Location: 110 B

Summary: The prime goal of the workshop is to bring the human motion synthesis community together and foster discussions about the existing challenges and future direction. To enable this, we feature invited talks presented by a diverse group of leading experts spanning a variety of sub-domains. With this workshop, we hope to encourage cross-pollination of ideas coming from different vantage points as well as discuss the gap between the academic and the industrial perspectives of the topic.

Photo-realistic 3D Head Avatars

Organizers: Tobias Kirschstein, Simon Giebenhain, Tianye Li, Koki Nagano, Justus Thies, Matthias Nießner

Date: 6/11/2025
Time: 13:00 - 17:00
Location: 110 A

Summary: Photorealistic 3D head avatars will play a crucial role in future computer games, visual effects, movie production, and virtual telepresence. In this workshop, we bring together leading academic researchers and industry experts to discuss the technology behind 3D head avatars, current applications, and future trends. In particular, we focus on two key desiderata of 3D head avatars: achieving the highest possible rendering quality and controlling the avatar with a driving signal. To this end, the workshop hosts a challenge on the NeRSemble 3D Head Avatar Benchmark. Challenge participants are invited to submit their methods for two tasks: dynamic novel view synthesis on heads, and monocular FLAME-driven 3D head avatar reconstruction. The authors of the best-performing submission will receive a GPU prize and present their method alongside invited speakers in the workshop.

Multimodal Algorithmic Reasoning Workshop

Organizers: Anoop Cherian, Suhas Lohit, Honglu Zhou, Kevin A. Smith, Joshua B. Tenenbaum

Date: 6/11/2025
Time: 13:00 - 17:00
Location: 207 A-D

Summary: In this workshop, we plan to gather researchers working in neural algorithmic learning, multimodal reasoning, and cognitive models of intelligence to showcase their cutting-edge research, discuss the latest challenges, as well as bring to the forefront problems in perception and language modeling that are often overlooked but are pivotal in achieving true artificial general intelligence. An emphasis of this workshop is on the emerging topic of multimodal algorithmic reasoning, where a reasoning agent is required to automatically deduce new algorithms/procedures for solving real-world tasks, e.g., algorithms that use multimodal foundational models for analysis, synthesis, and planning, new approaches towards solving challenging vision-and-language mathematical (Olympiad type) reasoning problems, deriving winning strategies in multimodal games, procedures for using tools in robotic manipulation, etc. We hope to deep dive into this exciting topic at the intersection of multimodal learning and cognitive science to understand what we have achieved thus far in machine intelligence and what we are lacking in relation to the human way of thinking — through talks from outstanding researchers and faculty that could inspire the audience to search for the missing rungs on the ladder to true intelligence.

How to Stand Out in the Crowd?

Organizers: Anand Bhattad, Aditya Prakash, Unnat Jain, Angjoo Kanazawa, Georgia Gkioxari, Svetlana Lazebnik

Date: 6/11/2025
Time: 13:00 - 17:00
Location: 209 A-C

Summary: In today's AI landscape, visibility is harder than ever. The pace is breakneck, arXiv is overflowing, and the pressure to perform is real. So how do early-career researchers cut through the noise?

How do you define a research identity without chasing trends?

How do you publish with purpose, not just pace?

How do you explore emerging areas without getting lost in the noise?

How do you balance mentorship with momentum?

In its third year, this CVPR community-building workshop we bring voices across CV, NLP, ML, and Robot Learning — Andrea, Carl, Dima, Gül, Jia-Bin, Laura, Ludwig, Saining, Sara, and Shuran — to answer these questions and more. This is an open forum to share insights, frustrations, and hacks — because no one builds a research career alone.

Emergent Visual Abilities and Limits of Foundation Models

Organizers: Ashkan Khakzar, A. Sophia Koepke, Ameya Prabhu, Jindong Gu, Francesco Pinto, Arsha Nagrani, Boyi Li, Philip Torr, Trevor Darrell
Date: 6/11/2025
Time: 13:00 - 17:00
Location: 210

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

Visual information processing is being transformed by foundation models. Trained on massive datasets using self-supervised and generative methods, these models exhibit the emergence of sophisticated visual abilities—such as depth perception, object recognition, and part discovery — without explicit programming or supervision. This shift marks a new paradigm where neural models derive visual understanding from the intrinsic structures and patterns present in the data rather than supervisory signals associated with a visual task. Moreover, questions remain about how to systematically analyze and evaluate these emergent capabilities. Recent studies have also highlighted the models' visual limitations, emphasizing the need for innovative evaluation methods to identify these shortcomings. By evaluating and understanding both the capabilities and limits of these models, we can better compare different learning algorithms and architectures in terms of how they represent the visual world.

Workshop on Vision-based Assistants in the Real-world

Organizers: Apratim Bhattacharyya, Fadime Sener, Roland Memisevic, Bugra Tekin, Edoardo Remelli, Shugao Ma, Guodong Ding, Shweta Mahajan, Angela Yao
Date: 6/11/2025
Time: 13:00 - 17:00
Location: 211
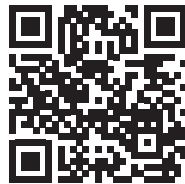
Summary: The ability of an AI system to truly comprehend and reason over a scene that is unfolding in real-time and in the real world is only starting to become reality. Progress in this area can have a huge impact by supporting daily tasks, such as cooking, health, fitness, driving, home improvement, and many others. This workshop seeks to serve as a forum for discussing recent progress and open challenges towards vision-based assistance for daily tasks and the technology to enable it. The VAR workshop features an exciting lineup of speakers and live demonstrations of existing systems. Additionally, as there are limited datasets and benchmarks for real-world vision-based assistants, the VAR workshop will host two new challenges designed to foster the development of such systems.

CVPR 2025 Photorealistic Avatar Challenge

Organizers: Ross Cutler, Julien Valentin, Justus Thies, Babak Naderi, Vishak Gopal
Date: 6/11/2025
Time: 13:00 - 17:00
Location: 212

Summary: The CVPR 2025 Photorealistic Avatar Challenge is intended to stimulate research in the field of photorealistic avatars. The challenge provides a test set and methodology to subjectively evaluate photorealistic avatars for news anchor and telecommunication scenarios. The evaluation task is audio/video driven self-reenactment. Test subjects will be sitting or standing but only the upper half of the body is rendered for tracks 1 and 2; track 3 will have head-only video. Speech, facial emotions, head turning, and hand gestures sequences are included. The inputs to the challenge are enrollment and test video clips. The output are self-reenactment video clips which we will evaluate with our subjective test framework and are used as the challenge metric.

Accessibility, Vision, and Autonomy Meet

Organizers: Eshed Ohn-Bar, Danna Gurari, Hernisa Kacorri, Kris M. Kitani, Chieko Asakawa, Jennifer Mankoff
Date: 6/11/2025
Time: 13:00 - 17:00
Location: 202 C

Summary: The overarching goal of this workshop is to gather researchers, students, and advocates at the intersection of accessibility, computer vision, and autonomous systems. Building upon the success of the previous CVPR workshop (with cross-disciplinary talks, posters, and challenges), this iteration will focus on addressing the lack of shared development tools and vision-based benchmarks for accessibility systems. The workshop will feature a multimodal challenge with synthetic and real-world benchmarks. By fostering discussion and actively engaging people with disabilities, the workshop aims to build a stronger community for accessibility research within computer vision, uncover research opportunities, and encourage the development of more effective and usable real-world visual reasoning models.

Real-to-Sim: Bridging the Gap between Neural Rendering and Robot Learning

Organizers: Wayne Wu, Bolei Zhou, Katerina Fragkiadaki
Date: 6/11/2025
Time: 13:00 - 17:00
Location: Davidson C1

Summary: The gap between simulation and real-world deployment remains a significant challenge in robotics and autonomous systems. This workshop addresses the potential of neural rendering to bridge this "sim-to-real" gap by enabling photorealistic simulations based on real-world data, facilitating robot learning. It aims to connect computer vision, robotics, and simulation communities, focusing discussions on reconstructing realistic environments and extracting real-world data distributions for robot training. Attendees will explore the current state, limitations, and future research directions in integrating neural rendering with robot learning to improve simulation-based training, fostering innovation in autonomous systems.

Thursday, June 12

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 CVPR 2025 Tutorials page.

Time	Event
7:00 - 17:00	**Registration / Badge Pickup** (ExHall Concourse)
7:00 - 17:00	**Press Room** (203 B)
7:00 - 17:00	**Mother's Room** (Level 1 near Room 101 and on Level 3 near Exhibit Hall D)
7:00 - 17:00	**Prayer or Quiet Room** (203 A)
7:00 - 9:00	**Breakfast** (ExHall C)
8:00 - 18:00	**WORKSHOPS / TUTORIALS**
10:00 - 11:00	**Coffee Break** (ExHall D)
12:00 - 13:45	**Lunch** (ExHall C)
15:00 - 16:00	**Coffee Break** (ExHall D)

TUTORIALS

Multi-Modal Computer Vision and Foundation Models In Agriculture in conjunction with IEEE CVPR 2025

Organizers: Paula Ramos-Giraldo, Chris Padwick, Jim Ostrowski, Ripudaman Singh Arora

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 107 B

Summary: With the recent success of computer vision and deep learning in various applications, there has been significantly increasing attention paid to its use in agriculture. Agriculture-related vision problems are of great economic and social value. For example, robotics has recently been reinvigorated with work on Vision-Language-Action models (c.f. https://www.physicalintelligence.company/blog/pi0). Building on these successes, researchers are using multi-modal computer vision foundation models to make progress on agricultural tasks and topics. Some relevant examples include 1) Agricultural models that leverage data from different remote sensing platforms, encompassing satellite, UAV, and even handheld data, encompassing a variety of different sensing modalities; 2) Multi-temporal yield prediction models that focus on semi-supervised/unsupervised domain adaptation with temporal attention based on multi-modality remote sensing data, weather, and management practices; 3) Multi-modal models that identify pests and weeds from hand-held imagery with a high degree of accuracy.

With the help of world-renowned/well-positioned researchers who will speak in detail and in-depth about their work, this tutorial will encourage research in machine learning, computer vision, and agriculture. Speakers will show the evolution and latest AI trends in Multi-modal Computer Vision and Foundation Models in Agriculture. And their groups' work.

Efficient Text-to-Image/Video modeling

Organizers: Srikumar Ramalingam, Sadeep Jayasumana, Ameesh Makadia, Richard Hartley

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 202 A

Summary: We are witnessing groundbreaking results in image-to-text and image-to-video models. However, the generation process with these models is iterative and computationally expensive, requiring multiple sampling steps through large models. There is a growing need to make these algorithms faster for serving millions of users without the use of too many GPUs/TPUs. In this course, we will focus on techniques such as progressive parallel decoding, distillation methods, and Markov Random Fields to achieve speedup text-to-image and text-to-video models. The course will also focus on highlighting the limitations of popular image evaluation techniques such as FID and providing efficient alternative approaches such as CMMD.

Animal Re-identification

Organizers: Lukas Picek, Sara Beery, Lukáš Adam, Blair Roberts Costelloe, Jason Allan Holmberg, Timm Haucke, Lasha Otarashvili

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 202 C

Summary: This tutorial will introduce the field of individual animal re-identification (ReID), a critical area in ML and CV that has profound ecological, conservation, and societal impacts. Accurate and scalable animal ReID allows for the long-term monitoring of endangered species, aids in combating wildlife poaching, and supports ecosystem management efforts. By enabling the identification of individuals without invasive tagging, ReID technology promotes ethical wildlife research while enhancing our ability to track animal movements, population dynamics, and behavioral patterns. This half-day hybrid tutorial will include four segments of several talks, delivered by four speakers (with some speakers presenting twice), aligning with CVPR's requirement of three or more speakers. The hybrid format will maximize accessibility and community engagement. Besides, we aim to host a panel discussion with various experts to encourage interactivity, answer questions, and inspire new directions for research in this area.

Edge AI in Action: Technologies and Applications

Organizers: Fabricio Batista Narcizo, Elizabete Munzlinger, Anuj Dutt , Shan Ahmed Shaffi, Sai Narsi Reddy Donthi Reddy

Date: 6/12/2025

Time: 9:00 - 12:00

Location: Davidson C3

Summary: Edge AI in Action: Technologies and Applications is a hands-on tutorial designed to equip participants with the knowledge and skills to develop, optimize, and deploy artificial intelligence models on resource-constrained edge devices. The tutorial explores the latest advancements in Edge AI and focuses on practical tools and techniques to address the challenges of low-latency, efficient, and reliable on-device inference.

Participants will gain experience through live coding sessions, covering topics such as model optimization using quantization and pruning, and deployment of lightweight large language models (LLMs) and computer vision models. Real-time demonstrations will showcase deployments on devices like Raspberry Pi, iPhones, and Android

smartphones, emphasizing performance and usability. Integrating hybrid cloud-edge architectures, leveraging tools such as Qualcomm's cloud services and Apple's MLX, will also be covered to highlight the potential for scalable and efficient AI applications.

This tutorial is intended for researchers and practitioners interested in the practical aspects of Edge AI, which has applications in autonomous systems, mobile computing, and real-time analytics. By the end of the session, attendees will have actionable insights and resources to implement cutting-edge AI solutions in real-world scenarios.

Computer Vision Over Homomorphically Encrypted Data

Organizers: Vishnu Boddeti, Wei Ao, Amina Bassit

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 205 B

Summary: Over the past decade, computer vision (CV) systems have become integral to applications spanning healthcare, surveillance, and personal devices. However, the sensitive nature of the data these systems handle and the proprietary nature of many CV models have raised critical concerns about data privacy and security. Fully homomorphic encryption (FHE) offers a transformative solution by enabling computations directly on encrypted data, ensuring privacy without exposing raw data during processing. This tutorial explores the integration of FHE into CV systems, a combination that addresses stringent privacy requirements and ensures secure computation outsourcing.

Despite its promise, combining FHE and CV poses unique challenges due to the computational complexity and nonlinearity of CV models, coupled with the operational constraints of FHE (e.g., restricted support for non-arithmetic operations). Recent advances have addressed these challenges, enabling practical applications like private image retrieval, encrypted image classification, and fully encrypted face recognition. These developments underscore the feasibility and transformative potential of encrypted CV systems.

This tutorial offers a comprehensive introduction to the mathematical foundations of FHE (e.g., ring theory, lattice-based cryptography, and learning with errors), an overview of key FHE schemes (e.g., BFV, BGV, TFHE, CKKS), and insights into leveraging FHE's SIMD capabilities for computational efficiency. We will also explore state-of-the-art research and applications at the intersection of CV and FHE and provide a hands-on demonstration of implementing private and encrypted CV tasks. Finally, we will discuss open research directions in this growing field, motivating participants to design secure, privacy-preserving CV systems leveraging FHE.

Continuous Data Cycle via Foundation Models

Organizers: Nadine Chang, Maying Shen, Zhiding Yu, Yan Chang, Jim Fan, Jose M. Alvarez

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 401 AB

Summary: Data has become more pivotal than ever, driving advancements from the first generation of deep learning models to the emergence of foundation models. The significant momentum behind foundation models has spurred their continuous integration into various applications, such as autonomous driving, medical diagnostics, and AI-powered chatbots. To ensure reliable and safe model development, the quality of the massive datasets these models depend on has gained increasing interest and attention. Given the sheer scale of raw data, it is essential to develop scalable methods to evaluate and select data based on the data quality and relevance to both general and specific tasks. Furthermore, foundation models themselves are now utilized to understand and uncover more data for further training, creating a feedback loop between models and datasets.

In this tutorial, we aim to discuss the challenges of the cyclic relationship between data and foundation models. Our talks will cover how foundation models affect future data efforts and how these data efforts can further improve foundation models. Specifically, we will discuss how to leverage foundation models for an efficient data engine that 1) strengthens models in various applications such as autonomous driving and embodied AI; 2) improves foundation models' performances.

Robotics 101: An Odyssey from A Vision Perspective

Organizers: Huijie Wang, Christos Sakaridis, Jean Oh, Zhixuan Liu

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 202 B

Summary: In recent years, there has been a growing interest in robotics in the vision community. Getting started in robotics research is challenging since it requires a wide range of knowledge. This comprehensive full-day tutorial covers all aspects of robotics. We will provide the necessary background to understand the different aspects of robotics. We will also explore technical advancements, key challenges, and potential avenues for future research. We have invited presenters from diverse domains to contribute to the tutorial, with non-overlapped content to existing events. This tutorial serves as a brand-new perspective from the vision community to discuss broad areas of robotics. This tutorial aims to provide an introduction to robotics, especially for researchers in the vision domain. The tutorial is divided into two parts: Perceive the World (AM session) focuses on understanding the real world, while Interact with the World (PM session) emphasizes the interaction with the physical world.

3D Shape Analysis: From Classical Optimization to Learning-based Matching

Organizers: Nafie El Amrani, Lennart Bastian, Viktoria Ehm, Zorah Lähner, Florian Bernard, Daniel Cremers

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 204

Summary: The field of 3D shape analysis is concerned with the extraction of "useful" information from geometric data. Shape analysis has a high relevance for a wide range of applications, such as autonomous driving, biomedicine, or augmented/virtual reality. A core task of 3D shape analysis is shape matching, i.e. identifying correspondences between given shapes. While traditional shape matching methods rely on optimising a task-specific objective function, modern shape matching oftentimes involves data-driven components. We will first introduce traditional methods for shape matching, starting with the linear assignment problem and the quadratic assignment problem. We then present product graph formalisms in different settings, including 2D to 2D, 2D to 3D or shape to image, and 3D to 3D shape matching. We then discuss recent developments in learning-based shape correspondence methods, from learning shape correspondence with topological data structures to spectral approaches that provide efficient structure and circumvent annotations altogether. Furthermore, we discuss the practical relevance of these methods to application domains in image-to-image and shape-to-image correspondence, medical imaging and surgical navigation, and discuss how recent developments in foundation models play a role in shape analysis. Finally, the tutorial will conclude by addressing the challenges of shape matching, including handling partial shapes, and will explore potential future directions in the field.

Geospatial Computer Vision and Artificial Intelligence for Large-Scale Earth Observation Data

Organizers: Orhun Aydin, Philipe Ambrozio Dias, Dalton Lunga, Qiusheng Wu
Date: 6/12/2025
Time: 9:00 - 17:00
Location: 103 B

Summary: Earth observation (EO) data, primarily obtained through remote sensing, in-situ technologies , aerial photography, and UAV-based platforms, is central to a variety of multidisciplinary problems. The integration of computer vision (CV) advancements with EO data has extensive real-world applications across fields such as human dynamics, precision agriculture, disaster management, humanitarian assistance, and national security. Unlike traditional natural images used in CV benchmarks, EO imagery presents unique characteristics and challenges that include spatial & temporal awareness, high data volumes & diversity, and the need for multimodal reasoning across varied EO sensing types, representations, and acquisition conditions. This tutorial introduces current EO data sources, associated challenges, and computer vision-based analysis techniques. We will highlight the latest advancements in EO data sources, models, and applications, with a particular focus on open-source analysis ecosystem related to computer vision and deep learning for EO data, in the form of public notebooks and hands-on sessions. In the era of Foundation Models, this will expose the audience to discussions on development of cutting-edge geospatial foundation models and their application to both archived and live satellite data for environmental and climate monitoring.

Sense, Perceive, Interact & Render on Android XR

Organizers: Sean Fanello, Federico Tombari, Thabo Beeler, Andrea Colaco
Date: 6/12/2025
Time: 9:00 - 17:00
Location: 201 B

Summary: Google Android XR is a new operating system built for this next generation of computing. At the heart of this platform, Computer Vision and Machine Learning are pivotal in ensuring immersive user experiences. In this tutorial, in particular, we will describe how we built from the ground up the full Perception stack: from head tracking algorithms, all the way to photorealistic avatars and scene renderings. Additionally, researchers and engineers will have access to comprehensive references and documentation of the APIs used in this project.

The tutorial begins by emphasizing the significance of data capture, rendering, and groundtruth generation for Perception tasks such as hand, face, and eye tracking.

Next, we explore the construction of an efficient Perception stack, encompassing egocentric head tracking, hand tracking, face tracking, and eye tracking.

Furthermore, we demonstrate how these perception capabilities enable the creation of scalable and efficient photorealistic representations of humans and scenes.

Finally, we showcase use cases and experiences that leverage the full stack, highlighting its potential applications.

Identifying Sturcion in Data: All You Need to Know About Dimensionality, Reduction, Clustering, and More

Organizers: M. Saquib Sarfraz, Laurens van der Maaten, Brandon Duderstadt, Marios Koulakis, Constantin Marc Seibold
Date: 6/12/2025
Time: 13:00 - 17:00
Location: 106 B

Summary: As large-scale models and datasets grow, techniques for data curation, training objective definition, and quality monitoring are essential. Mastering them ensures efficient dataset exploration, robust model development, and impactful decision-making in AI and computer vision workflows.

In today's AI landscape, these techniques are crucial. Dimensionality reduction tools like t-SNE, UMAP and h-NNE pose a possibility to highlight meaningful structures in visualizations and draw insights. Clustering, on the other hand, organizes unstructured data into meaningful groups, aiding knowledge discovery, feature analysis, and retrieval-augmented generation (RAG).

These methods also address learned feature biases, errors, and redundancies affecting model performance. Dimensionality reduction, when done right, can help identify outliers and irregular patterns within the data. Robust clustering supports scalable embedding pipelines, enabling efficient data curation and querying. From k-means to DBSCAN and hierarchical approaches like FINCH, selecting the right method is key: including balancing scalability, managing noise sensitivity, and fitting computational demands.

This tutorial provides an in-depth exploration of the current state-of-the-art of data exploration techniques such as dimensionality reduction for data visualization and clustering methods, with a strong focus on their applications within computer vision. Attendees will gain a comprehensive understanding of both foundational and advanced techniques beyond classic methods like t-SNE and k-means. Through a blend of theoretical insights and hands-on applications, participants will learn how to effectively apply these methods to tasks such as big data analysis, representation learning, model development, pseudo-labeling, and data annotation.

Multimodal Mathematical Reasoning: Frontiers in Integrating Vision, Language, and Symbolic Representations

Organizers: Tianyu Yang, Yilun Zhao, Xiangliang Zhang
Date: 6/12/2025
Time: 13:00 - 17:00
Location: 107 B

Summary: This tutorial provides a focused overview of multimodal mathematical reasoning—a rapidly emerging field bridging computer vision, natural language processing, and symbolic logic. While vision-language models excel at recognizing objects and answering natural-language queries about images, extending these capabilities to mathematical tasks is non-trivial. Challenges include interpreting diagrams and charts, integrating symbolic notation, and performing logical multi-step reasoning.

We will survey representative datasets and benchmarks, highlight state-of-the-art models that fuse text, vision, and symbolic inputs, and discuss evaluation methods that emphasize explainability and logical rigor beyond simple accuracy. Attendees will gain insights into how advancements in computer vision, such as diagram parsing and fine-grained feature extraction, can enhance math-focused reasoning tasks. Finally, a panel discussion will examine applications in education, scientific discovery, and future research directions at the intersection of vision and mathematical reasoning.

Full-Stack, GPU-based Acceleration of Deep Learning and Foundation Models

Organizers: Maying Shen, Jason Clemons, Hongxu Yin, Pavlo Molchanov, Jose M. Alvarez, Jan Kautz

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 205 A

Summary: The best practice to develop and deploy large deep neural networks has been a very important topic across both the industry and the academia. In this edition of the tutorial, we wish to offer the most recent insights across the entire hardware-software stack to accelerate deep neural networks from convolutional architectures to the latest multimodal large language models. The major benefits are two-fold. On one hand, this tutorial will equip the deep learning practitioners with the recent advanced off-the-shelf tools to accelerate their models, and hence significantly reduce the design efforts across a wide range of sub-fields. In addition, through analyzing the approaches, challenges, and the current landscape with invited talks, we wish this tutorial to offer insights to the community and inspire the next generation of acceleration techniques that can oversee the design trade-offs of upstream software and downstream hardware, and hence fully explore the potential of the entire GPU stack. We believe this tutorial can greatly help practitioners, researchers, and entities across both industry and academia to greatly scale up the impact of their deep neural networks.

Power-efficient Neural Networks Using Low-precision Data Types and Quantization

Organizers: Thomas Pfeil, Markus Nagel, Tijmen Blankevoort

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 205 B

Summary: The growing size of neural networks, particularly in generative AI, poses significant challenges in terms of sustainability, time, and cost, hindering their study and practical application. Low-precision data types and computations, especially when natively supported by hardware, offer an effective solution, enabling broader research access and deployment on edge devices. However, to this end networks that are usually trained with high-precision data types have to be prepared for low-precision execution. In this tutorial, we review different low-precision data types and showcase typical challenges of their application, like outlier handling, on simple hands-on examples. In order to maintain the original task performance of neural networks, sophisticated quantization methods are required to compensate for quantization errors induced by low-precision data types. We introduce and compare the most common and effective methods to quantize neural networks and provide guidance for practitioners.

Intelligent Healthcare based on Cameras and Wireless Sensors

Organizers: Wenjin Wang, Daniel McDuff, Xuyu Wang

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 212

Summary: Understanding human health and extracting physiological-related metrics is an emerging research topic in computer vision that has grown rapidly recently. Without the need of any physical contact of the human body, cameras have been used to measure vital signs remotely (e.g. heart rate, heart rate variability, respiration rate, blood oxygenation saturation, pulse transit time, body temperature, etc.) from an image sequence of the skin or body, which leads to contactless, continuous and comfortable heath monitoring. The use of cameras also enables the measurement of human behaviors/activities and high-level visual semantic/contextual information leveraging computer vision and machine learning techniques, such as facial expression analysis for pain/discomfort/delirium detection, emotion recognition for depression measurement, body motion for sleep staging or bed exit/fall detection, activity recognition for patient actigraphy, etc. Understanding of the environment around the people is also a unique advantage of cameras compared to the contact bio-sensors (e.g., wearables), which facilitates better understanding of human and scene for health monitoring. In addition to camera based approaches, Radio Frequency (RF) based methods for health monitoring have also been proposed, using Radar, WiFi, RFID, and acoustic signals. Radar based methods mainly use Doppler/UWB/FMCW radar for health monitoring. They can obtain high monitoring accuracy for different applications such as sleeping staging and posture estimation. By using off-the-shelf WiFi device, for example WiFi RSS and CSI data from commodity WiFi NIC, we can monitor breathing and heart rates for single and multiple persons. In addition, WiFi signal can be used for other health related activity recognition based on machine learning algorithms. For acoustic based vital sign monitoring, the speaker and microphone of smartphones are used to build a sonar based sensing system to monitor breathing and heart rates. The rapid developments of computer vision and RF sensing have also given rise to new multi-modal learning techniques that expand the sensing capability by combining two modalities, while minimizing the need of human labels. The hybrid approach will also improve the performance of monitoring, such as using the camera images as beacon to gear human activity learning for the RF signals. The contactless monitoring of camera and RF will bring a rich set of compelling healthcare applications that directly improve upon contact-based monitoring solutions and improve people's care experience and quality of life, called "AI health monitoring", such as in care units of the hospital (patient monitoring), sleep/senior centers (sleep disorders like apnea), assisted-living homes (baby and elderly care), telemedicine and e-health, fitness and sports, driver monitoring in automotive, pilot/astronaut monitoring, etc.

Recent Advances in Vision Foundation Models

Organizers: Zhengyuan Yang, Linjie Li, Zhe Gan, Jianwei Yang

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 401 AB

Summary: We present our CVPR tutorial proposal on Recent Advances in Vision Foundation Models, a topic that has garnered significant attention from the computer vision community. Our tutorial will cover the most advanced directions in designing and developing vision foundation models, including the state-of-the-art approaches and principles in (i) learning vision foundation models for multimodal understanding and generation, (ii) scaling test-time compute and enabling the self-training of foundation models to improve themselves on reasoning and perception, and (iii) physical and virtual agents based on vision foundation models that can take actions for robotics and in virtual environments.

WORKSHOPS

5th International Workshop on Long-form Video Understanding

Organizers: Mike Zheng Shou, Yiqi Lin, Joya Chen, Ziyun Zeng, Linchao Zhu, Gedas Bertasius, Md Mohaiminul Islam, Gaoang Wang, Wei Li, Matt Feiszli, Lorenzo Torresani, Kristen Grauman, Jitendra Malik

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 105 A

Summary: Long-form videos are a rich source of knowledge and are increasingly considered as fundamental data for building a general world model. Building on the success of last year's workshop which attracted over 200 participants and 50+ competition teams, this year's event features talks pointing toward future directions, identifies suitable benchmarks, and introduces new data annotations designed to advance the development of world knowledge models capable of understanding long-form videos. Furthermore, the workshop's focus on event segmentation, sports streaming commentary, and the potential for world models in Augmented Reality (AR) extends its appeal to researchers in fields such as Cognitive Science, Machine Learning, Robotics, and video production.

Workshop on Foundation and Large Vision Models in Remote Sensing

Organizers: Saurabh Prasad, Jocelyn Chanussot, Begüm Demir, Biplab Banerjee, Danfeng Hong

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 209 A-C

Summary: This workshop will feature keynotes and presentations at the cutting-edge of foundation models and large vision models for remote sensing - it will bring together researchers working on both foundation and large vision models and geospatial image analysis to address the nuances presented by using such emergent models for remotely sensed imagery (e.g. a multitude of sensors with different sensing characteristics/specifications, diverse imaging modalities, ranging from passive-optical multi/hyperspectral to active-imaging such as SAR and LiDAR; limited ground-reference data etc.). Our emphasis will range from large vision and foundation models that are showing promise in the computer vision community to foundation models that are pre-trained on large-quantities of earth-observation imagery - this workshop will provide a venue for the community to present works that push the envelope on adapting these models for effective inference of multi-sensor, multi-temporal, multi-scale earth observation imagery.

Workshop on 3D-LLM/VLA: Bridging Language, Vision and Action in 3D Environments

Organizers: Jianing Yang, Shengyi Qian, Yining Hong, Valts Blukis, Xiaojian Ma, Yash Sanjay Bhalgat, Iro Laina, Joyce Chai, David Fouhey

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 106 A

Summary: This workshop addresses a critical gap in current AI research by focusing on the integration of language and 3D perception, which is essential for developing embodied agents and robots, especially considering the recent rise of multimodal LLMs and vision-language-action (VLA) models. The workshop will explore challenges and opportunities in this area, providing a platform for researchers to share their work, discuss future directions, and foster collaboration across disciplines including robotics, computer vision, natural language processing, and human-computer interaction.

SyntaGen: 2nd Workshop on Harnessing Generative Models for Synthetic Visual Datasets

Organizers: Khoi Nguyen, Anh Tuan Tran, Binh-Son Hua, Supasorn Suwajanakorn, Yi Zhou

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 106 B

Summary: Computer vision has been rapidly transformed by advancements in generative models, particularly in text-to-image generation with models like Imagen 3, Stable Diffusion 3, Flux, and DALLE-3, as well as text-to-video models such as Sora, Stable Video Diffusion, and Meta MovieGen. In the realm of 3D generation, models like Zero-123, Instant 3D, and the Large Reconstruction Model (LRM) have pushed the boundaries of 3D content creation. These innovations have enabled the development of highly realistic and diverse synthetic visual datasets, complete with annotations and rich variations, which are invaluable for training and evaluating algorithms in object detection, segmentation, representation learning, and scene understanding. The second SyntaGen Workshop aims to foster collaboration and knowledge exchange across the field, bringing together experts and practitioners to propel the development of generative models and synthetic visual datasets to new heights. Through talks, paper presentations, poster sessions, and panel discussions, the workshop will catalyze breakthroughs at the intersection of generative models and computer vision applications.

2nd Workshop on Efficient Large Vision Models

Organizers: Amir Habibian, Fatih Porikli, Auke Wiggers, Yung-Hsiang Lu, Vincent Tao Hu, Lanqing Guo, Qinghao Hu

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 106 C

Summary: This workshop focuses on the core principles of efficiency in large-scale vision models. How do we minimize redundant operations in generative models without compromising quality? Can autoregressive decoding and diffusion sampling be accelerated through parallelization? What are the trade-offs between compression, quantization, and expressivity? We seek to advance new directions in compact model representations, adaptive computation, parallel decoding, and structured sparsity—approaches that go beyond incremental optimizations and redefine how LVMs operate.

We invite researchers working on fast and scalable vision architectures, low-cost inference, and efficient generative models to share their insights. Whether through sampling acceleration, efficient transformers, new architectural paradigms, or theoretical limits of model compression, this workshop provides a platform to discuss how LVMs can be optimized for both performance and practicality.

Join us in shaping the next generation of vision models—where efficiency is not just a constraint, but a driving force for innovation.

Another Brick in the AI Wall: Building Practical Solutions from Theoretical Foundations

Organizers: Grigorios Chrysos,
Aggelina Chatziagapi,
Blerina Gkotse, Vikas Singh,
Sanmi Koyejo, Philip Torr,
Matthew B. Blaschko

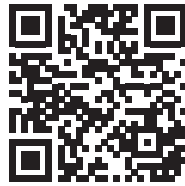**Date:** 6/12/2025

Time: 9:00 - 12:00

Location: 107 A

Summary: The shift towards foundation models has overshadowed the unique insights of deep learning theory, resulting in a loss of valuable knowledge and resources for the community. As machine learning and computer vision extend into new domains, such as biology, a deeper understanding of vision tasks becomes increasingly important. This workshop will provide a crucial platform for discussing the systematic challenges of integrating theory and practice. Concretely, to bridge the gap between theoretical research in machine learning and its practical applications, the workshop aims to explore how theoretical tools can be leveraged to perform rigorous worst-case analysis, crucial for deploying machine learning models in safety-critical societal domains like healthcare, education, and sustainability.

WorldModelBench: The 1st Workshop on Benchmarking World Models

Organizers: Heng Wang, Prithvijit Chattopadhyay,
Ming-Yu Liu, Mike Zheng Shou,
Jay Zhangjie Wu, Xihui Liu,
Deepti Ghadiyaram,
Gowthami Somepalli, Huaxiu Yao,
Wenhu Chen, Jiaming Song,
Humphrey Shi

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 108

Summary: World models are predictive systems that enable Physical AI agents to understand, decide, plan, and analyze counterfactuals through integrated perception, instruction processing, controllability, physical plausibility, and future prediction capabilities. The past year has witnessed significant advancements from both academic and industrial research teams, with various models utilizing different conditioning approaches (text, image, video, control) being released openly and commercially. While these developments enable applications in content creation, autonomous driving, and robotics, the models' diversity in training methods, data sources, architecture, and input processing necessitates critical evaluation. The WorldModelBench workshop aims to address this need by fostering discussions on evaluation criteria (physical correctness, prompt alignment, generalizability), metrics development, standardized methodologies, and crucial topics including accessible benchmarking, quantitative evaluation protocols, downstream task assessment, and safety/bias considerations in world models.

Test-time Scaling for Computer Vision

Organizers: Hang Su, Yinpeng Dong, Jindong Gu,
Yichi Zhang, Cihang Xie,
Lingjuan Lyu, Jun Zhu, Philip Torr,
Shiguang Shan, Wanli Ouyang

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 109

Summary: The Workshop on Test-time Scaling for Computer Vision (ViSCALE) at CVPR2025 aims to explore the frontiers of scaling test-time computation in vision models, addressing both theoretical advancements and practical implementations. We will discuss the suitability of test-time scaling for traditional vision tasks like perception and the extensions to multimodal and generative models, towards enhancing performance in critical domains. It will also cover solutions for efficient algorithms, considerations of robustness and safety, and novel problems in computer vision posed by test-time scaling. By bringing together experts, the workshop seeks to foster collaboration and innovation in applying this paradigm to push the limits of computer vision.

Three things everyone should ask about photorealistic virtual try-on.

Organizers: Aayush Bansal, Minh Vo

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 110 A

Summary: Virtual Try-On (VTON) promises to transform the apparel e-commerce industry, offering benefits for shoppers, businesses, and the environment. This workshop will address key challenges that must be overcome to realize VTON's full potential: achieving high-fidelity, rapid video try-ons; accurately predicting 3D garment size and improving 3D human body reasoning; and defining robust metrics for synthesis quality that avoid offensive results across diverse demographics. Addressing these VTON-specific challenges will necessitate fundamental advancements in generative image and video synthesis, offering broader impact within the computer vision and machine learning communities.

C3DV: 3rd Workshop on Compositional 3D Vision

Organizers: Habib Slim, Xiang Li,
Mahmoud Ahmed, Peter Vajda,
Wolfgang Heidrich,
Mohamed Elhoseiny,
Natalia Neverova

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 110 B

Summary: An important goal of 2D and 3D computer vision is visual understanding of complex textured shapes with varying styles and spatial orientations. The workshop focuses on compositional learning in 3D vision tasks, highlighted by a dataset challenge specifically designed for 3D compositional recognition. The organizers aim to stimulate new research and advancements in the challenging and relevant field of compositional 3D vision.

Women in Computer Vision

Organizers: Estefania Talavera Martinez,
Deblina Bhattacharjee,
Mengwei Ren, Himangi Mittal,
Karen Sanchez

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 105 B

Summary: Despite the expansion of computer vision field, the percentage of female faculty members and researchers both in academia and in industry is still relatively low. As a result, many female researchers working in computer vision may feel isolated and do not have a lot of opportunities to meet with other women. The goals of this workshop

are to: Raise the visibility of female computer vision researchers by presenting invited research talks by women who are role models in this field. Give opportunities to junior female students or researchers to present their work via a poster session and travel awards. Share experience and career advice for female students and professionals.

The 5th Workshop of Adversarial Machine Learning on Computer Vision: Foundation Models + X

Organizers: Tianyuan Zhang, Siyang Wu, Aishan Liu, Jiakai Wang, Siyuan Liang, Felix Juefei-Xu, Qing Guo, Xinyun Chen, Yew-Soon Ong, Xianglong Liu, Dawn Song, Alan Yuille, Philip Torr, Dacheng Tao

Date: 6/12/2025
Time: 9:00 - 12:00
Location: 205 A

Summary: The workshop will bring together researchers and practitioners from the computer vision and machine learning communities to explore the latest advances and challenges in adversarial machine learning, with a focus on the robustness of X-domain-specific foundation models. The program will consist of invited talks by leading experts in the field, as well as contributed talks and poster sessions featuring the latest research. In addition, the workshop will also organize a challenge on adversarial attacking foundation models. We believe this workshop will provide a unique opportunity for researchers and practitioners to exchange ideas, share latest developments, and collaborate on addressing the challenges associated with the robustness and security of foundation models. We expect that the workshop will generate insights and discussions that will help advance the field of adversarial machine learning and contribute to the development of more secure and robust foundation models for computer vision applications.

What is Next in Multimodal Foundation Models?

Organizers: Hilde Kuehne, Rogerio Feris, Leonid Karlinsky, Anna Kukleva, Ameya Prabhu, Wei Lin, Muhammad Jehanzeb Mirza, Sivan Doveh, Roei Herzig

Date: 6/12/2025
Time: 9:00 - 12:00
Location: 207 A-D

Summary: The intersection of foundation models with multimodal learning is an important and broadly discussed topic that complements the main CVPR conference. This workshop aims to foster interdisciplinary discussion on the latest advancements, remaining challenges, and future directions of multimodal foundation models, which have seen breakthroughs by adapting techniques across computer vision and NLP.

1st International Workshop on Interactive Video Search and Exploration

Organizers: Luca Rossetto, George Awad, Werner Bailer, Cathal Gurrin, Björn Þór Jónsson, Jakub Lokoč Stevan Rudinac, Klaus Schoeffmann

Date: 6/12/2025
Time: 9:00 - 12:00
Location: 208 A

Summary: The aim of the 1st international Workshop on Interactive Video Search and Exploration is to explore means to overcome current limitations in fully automated methods by focusing on human-machine teaming approaches for long-form video understanding. It will provide a venue to compare fully automated end-to-end approaches for video understanding and ones where humans and machines collaborate. The workshop will be centered around a challenge on text-based video retrieval and visual question answering in a large collection of long videos (3min to 30min per video, 1000h combined runtime). The challenge will be split into two tracks; the fully-automated track where queries are made available to participants beforehand and they have to solve them automatically without direct human intervention, and the interactive track where queries will be made available to participants during the workshop and they have to solve them interactively in a human-machine team under a strict time limit of five minutes.

Workshop on Distillation of Foundation Models for Autonomous Driving

Organizers: Burhaneddin Yaman, Yunsheng Ma, Xin Ye, Xu Cao, Wenqian Ye, Ana Jojic, Abhirup Mallik, Ziran Wang

Date: 6/12/2025
Time: 9:00 - 12:00
Location: 208 B

Summary: The 1st Workshop on Distillation of Foundation Models for Autonomous Driving (WDFM-AD) aims to advance the deployment of large foundation models—such as vision-language models (VLMs) and generative AI (GenAI) models—within autonomous driving systems through efficient distillation techniques. Building on the momentum of prior workshops focused on large language and vision models for autonomous driving, WDFM-AD provides a dedicated platform for researchers and industry practitioners to explore methods that bridge cutting-edge foundation model research with real-world deployment, particularly under the stringent latency and resource constraints of autonomous vehicles. By addressing the challenges of compressing, aligning, and deploying foundation models for self-driving, WDFM-AD seeks to accelerate their safe, efficient, and scalable integration into next-generation autonomous driving systems.

Spatial Intelligence for Cultural Heritage

Organizers: Marina Paolanti, Roberto Pierdicca, Fabio Remondino, Livio De Luca

Date: 6/12/2025
Time: 9:00 - 12:00
Location: 210

Summary: The Spatial Intelligence for Cultural Heritage (SINT4CH) workshop explores the role of AI in preserving and interacting with historical artefacts and sites. By integrating perception, action and learning, the concept of spatial intelligence aims to advance the engagement with cultural heritage (CH). This workshop will therefore explore topics such as 3D reconstruction, AI-driven restoration, and human-AI collaboration in the cultural heritage sector, with an emphasis on sustainable and ethical approaches to conservation. With progress in computer vision, robotic interaction, and large-scale language models, participants will explore techniques and future directions that augment human capabilities to ensure the long-term preservation and understanding of our CH.

ScanNet++ Novel View Synthesis and 3D Semantic Understanding Challenge

Organizers: Angela Dai, Yueh-Cheng Liu, Chandan Yeshwanth, Ben Mildenhall, Peter Kontschieder, Matthias Nießner

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 211

Summary: Recent advances in generative modeling and semantic understanding have spurred significant interest in synthesis and understanding of 3D scenes. In 3D, there is significant potential in application areas, for instance augmented and virtual reality, computational photography, interior design, and autonomous mobile robots all require a deep understanding of 3D scene spaces. The ScanNet++ workshop offers the first benchmark challenge for novel view synthesis in large-scale 3D scenes, along with high-fidelity, large-vocabulary 3D semantic scene understanding — where very complete, high-fidelity ground truth scene data is available. This is enabled through the new ScanNet++ dataset, which offers 1mm resolution laser scan geometry, high-quality DSLR image capture, and dense semantic annotations over 1000 class categories. In particular, existing view synthesis leverages data captured from a single continuous trajectory, where evaluation of novel views outside of the original trajectory capture is impossible. In contrast, our novel view synthesis challenge leverages test images captured intentionally outside of the train image trajectory, allowing for comprehensive evaluation of methods to test new, challenging scenarios for state-of-the-art methods.

Rhobin2025: The 3rd Rhobin Challenge on Reconstruction of Human-Object Interaction

Organizers: Xi Wang, Xianghui Xie, Nikos Athanasiou, Dimitrios Tzionas, Shashank Tripathi, Bharat Lal Bhatnagar, Alexey Gavryushin, Thiemo Alldieck, Muhammed Kocabas, Luc Van Gool, Marc Pollefeys, Gerard Pons-Moll

Date: 6/12/2025

Time: 9:00 - 12:00

Location: 212

Summary: The Rhobin workshop will continue provide a venue to present and discuss state-of-the-art research in the reconstruction of human-object interactions from images. The focus of this third workshop will go beyond image-based interaction reconstruction, extend to interaction tracking over time, and seek connections to relevant topics such as egocentric vision and dynamic scene interactions. The third Rhobin challenge will feature five tracks in total with tasks on human-object interaction tracking and image-based contact estimation, using datasets InterCap, DAMON, and BEHAVE.

VAND: Visual Anomaly and Novelty Detection - 3rd Edition

Organizers: Latha Pemula, Samet Akcay, Toby P. Breckon, Philipp Seeböck, Paul Bergmann, Paula Ramos-Giraldo, Yedid Hoshen, Guansong Pang, Jawad Tayyub, Thomas Brox

Date: 6/12/2025

Time: 9:00 - 12:00

Location: Davidson C1

Summary: Anomaly detection—also known as novelty or out-of-distribution detection—is a key challenge in computer vision and pattern recognition. From medical imaging to industrial inspection, spotting what doesn't belong is critical, yet notoriously hard. Why? Because anomalies can take unlimited forms, and most models see only a narrow slice of the possible "normal" during training.

The VAND workshop brings together cutting-edge research tackling this open-set problem across supervised, semi-supervised, and unsupervised methods, as well as few-, one-, and zero-shot approaches.

This year, we're also hosting two exciting challenges: (1) 'Adapt & Detect – Robust anomaly detection in real-world applications', and (2) 'VLM Anomaly Challenge – Few-shot learning for logical and structural anomaly detection using vision-language models'.

Join us to explore the next generation of models that can detect the unexpected.

The 3rd Workshop on Sign Language Recognition, Translation and Production

Organizers: Mohamed Ilyes Lakhal, Ozge Mercanoglu Sincan, Gül Varol, Harry Walsh, Edward Fish, Liliane Momeni, Necati Cihan Camgoz, Neil Fox, Kearsy Cormier, Bencie Woll, Richard Bowden

Date: 6/12/2025

Time: 9:00 - 12:00

Location: Davidson C2

Summary: Sign languages are visual languages and a key form of communication for deaf communities. Thanks to recent advances in deep learning and computer vision and the availability of larger datasets, significant progress has been made in sign language technologies. Following the first and second editions, this workshop is motivated by the desire to broaden participation in sign language research from the computer vision community. It aims to bring together researchers working on different aspects of vision-based sign language research and sign language linguists to explore recent advances and future directions in sign language recognition, translation, and production.

Workshop on Visual Concepts

Organizers: Yunzhi Zhang, Joy Hsu, Jiayuan Mao, R. Kenny Jones, Himanshu Gaurav Singh, Daniel Cohen-Or, Shangzhe Wu, Jiajun Wu

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 101 A

Summary: Visual concept discovery aims to extract compact and structured representations of the visual world, and recompose them to tackle novel intricate problems. It has played a crucial role in many core problems in computer vision research, including both discriminative and generative tasks. An important research question is to understand and design concept representations that facilitate better learning from various datasets and compositional reasoning. As an endeavor to answering this question, in this workshop, we gather together researchers in computer vision, multi-modal learning, machine learning, and cognitive science to discuss the development and interpretation of visual concept learning systems and their applications.

The 2nd Workshop on Populating Empty Cities – Intelligent Virtual Humans for Autonomous Systems

Organizers: Kwan-Yee Lin, Wayne Wu, Bolei Zhou, Matthias Nießner, Stella X. Yu

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 101 B

Summary: This workshop aims to explore the pathway toward building "Embodied Humans"—intelligent humanoid agents capable of both physical action and cognitive reasoning like humans—where the boundary between digital avatars and physical humanoid robots could be dissolved through their co-evolution across virtual and real worlds. We will examine this synergy's possibility through three core dimensions: 1) how humanoid robots learning foundational "genes" from avatars? 2) how virtual humans gain physical plausibility from robots' embodiment to enrich realism and interactivity? and 3) how both systems develop self-autonomy to perceive, plan, and act in dynamic, open-ended environments? Featuring academic researchers and industry experts as invited speakers and panelists, the workshop brings together perspectives from virtual avatar modeling and humanoid robot learning to explore how systems on both ends are progressing toward human-like capacities for perception, reasoning, and movement. Through advanced techniques—such as reinforcement learning, cognition modeling, motion and structure perception, geometric representations, multimodal simulation, and large language/vision/action models—we aim to understand how virtual humans are evolving beyond surface-level realism, and how humanoid robots are advancing beyond pre-scripted skills—enabling both to engage the world with situational understanding, behavioral adaptability, and autonomous intent. At the heart of this workshop lie two essential questions: What makes a virtual human real—not just to see, but to know? And what does it take for a humanoid robot to not just move, but to become?.

11th Workshop on Medical Computer Vision

Organizers: Yuankai Huo, Le Lu, Bennett Allan Landman, Daniel Moyer, Jie Ying Wu, Xiaoxiao Li, Chenyu You, Zhiyu Wan, Yucheng Tang, Nourhan Bayasi, Roza G Bayrak

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 101 C

Summary: The transformative potential of medical computer vision lies in its ability to revolutionize healthcare through enhanced diagnostic accuracy, improved treatment planning, and democratized access to high-quality care. The Medical Computer Vision (MCV) Workshop at CVPR will explore the impact of rapidly expanding deep learning methodologies within medical imaging. This workshop will address the challenges and opportunities associated with automating the analysis of complex medical images and fostering the translation of research into clinical practice, while also creating an engaging environment for students, clinicians, and researchers to collaborate and advance the field.

6th Embodied AI Workshop (EAI)

Organizers: Angel X Chang, Anthony Gerald Francis, Cem Gokmen, Changan Chen, Chengshu Li, Claudia Pérez-D'Arpino, David Hall, German Ros, Joel Jang, Lamberto Ballan, Luca Weihs, Mike Roberts, Minyoung Hwang, Oleksandr Maksymets, Rachith Prakash, Ram Ramrakhya

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 101 D

Summary: The Sixth Annual Embodied AI Workshop brings together researchers from computer vision, language, graphics and robotics to share the latest advances in embodied intelligent agents that see, talk, listen, reason, and act in bodies within interactive environments. This year's workshop focuses on Real World Applications, with topics including Embodied AI Solutions, Advances in Simulation, Generative Methods, and Foundation Models. The workshop will feature invited talks, a poster session, and panel discussions. Also, the sixth iteration of the workshop continues its tradition of highlighting several embodied AI challenges that advance the state of the art in the field.

Workshop on Pixel-level understanding with Vision Foundation Models

Organizers: Mennatullah Siam, Stella X. Yu, Sangwoo Mo, Leonid Sigal, Raoul de Charette, Tanzila Rahman, He Zhao, Aoran Xiao
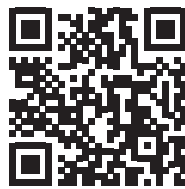
Date: 6/12/2025

Time: 9:00 - 17:00

Location: 101 E

Summary: In recent years, foundation models have gained significant traction and success. Such foundation models were shown to effectively adapt across various downstream tasks, with strong generalization capabilities, especially in zero-shot and few-shot scenarios. There is growing interest and progress specifically in vision foundation models (VFM). Some of the latest models include those trained using self supervision, such as the DINO series, and those utilizing image/text like CLIP, Flamingo, Llava and Cambrian. Various pixel-level vision foundation models have also emerged for image/video referring segmentation and depth estimation. Our workshop aims to bring together researchers dedicated to developing and adapting vision foundation models for pixel-level understanding tasks, including image/video segmentation, referring image/video segmentation and reasoning, tracking, depth estimation, and motion estimation. We will explore major directions in pixel-level understanding with vision foundation models and discuss the opportunities they present, particularly in low-resource settings that could have a positive societal impact. Additionally, we will discuss the risks associated with these models and explore methods to mitigate them. The workshop features seven invited talks, mixing emerging and established researchers, along with posters and selective spotlight presentations.

Multi-Agent Embodied Intelligent Systems Meet Generative-AI Era: Opportunities, Challenges and Futures

Organizers: Haibao Yu, Jianing Qiu, Yao Mu, Jiankai Sun, Li Chen, Walter Zimmer, Jiaru Zhong, Dandan Zhang, Fei Gao, Shanghang Zhang, Mac Schwager, Ping Luo, Zaiqing Nie, Tianxing Chen, Wenxian Yang, Ruiyang Hao, Chuanye Wang, Jiahao Wang, Siqi Fan

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 102 A

Summary: This workshop focuses on cooperative intelligence within multi-agent embodied intelligent systems. Artificial intelligence has propelled the development of embodied AI, particularly in autonomous vehicles, robotics, and drones. However, achieving full autonomy in complex and dynamic environments remains a formidable challenge for individual agents. Cooperative intelligence offers a transformative approach that allows agents to collaborate and interact with the infrastructure to handle a wide range of tasks more efficiently. Despite progress, challenges in coordinating multi-agent systems remain underexplored. Key hurdles include deciding what information to transmit, how to transmit, and how to fuse data across various levels like perception, prediction, and planning. Moreover, obtaining high-quality real-world datasets is difficult. Recent advances in foundational and generative models offer promising ways to overcome these obstacles. This workshop will explore opportunities, challenges, and future directions for multi-agent embodied intelligent systems in the generative-AI era.

3D Digital Twin: Progress, Challenges, and Future Directions

Organizers: Zhao Dong, Zhaoyang Lv, Zhengqin Li, Jiajun Wu, Hao Su, Manmohan Chandraker, Kalyan Sunkavalli, Jia Deng, Shuang Zhao, Lingjie Liu, Jerome Revaud, Hong-Xing Yu, Yunzhi Zhang, Leonidas Guibas

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 102 B

Summary: Despite the growing momentum around 3D reconstruction and generative AI in computer vision, a critical gap remains: how to create photorealistic, fully functional 3D digital twins that are indistinguishable from their real-world counterparts and enable practical applications. This workshop tackles that challenge by spotlighting 3D digital twin creation technologies and their broad impact across AR/VR, spatial and contextual AI, and robotics. Distinguished speakers from diverse disciplines will share cutting-edge digital twin creation techniques and real-world use cases. Additionally, we are excited to launch a benchmark and challenge for 3D digital twin creation, built on our Digital Twin Catalog (DTC) dataset and supported by open-source baselines. This initiative aims to spark meaningful discussion, foster collaboration, and accelerate progress in both academic research and practical deployment.

Visual Generative Modeling: What's After Diffusion?

Organizers: Tianhong Li, Yilun Xu, Tianyuan Zhang, Tim Dockhorn, Shuang Li, Arash Vahdat, Kaiming He

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 103 A

Summary: This year, our Visual Generative Modeling workshop at CVPR aims to explore what lies beyond diffusion models in visual generative modeling. We will discuss novel insights, alternative approaches, and new possibilities in modeling and generating visual data. Join us for a full-day event featuring keynote talks from top universities and industrial labs — all designed to ignite innovative ideas and novel research in visual generative modeling.

Computer Vision for Drug Discovery: Where Are We and What is Beyond?

Organizers: Dawid Damian Rymarczyk, Ilknur Icke, Adriana Borowa, Ana Sanchez-Fernandez, Chao-hui Huang, Anne E Carpenter

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 103 C

Summary: The workshop aims to bridge the gap between computer vision, artificial intelligence, and the life sciences, with a focus on transformative advancements in drug discovery. By integrating innovative imaging modalities—such as Spatial Transcriptomics, Cell Painting, and Optical Pooled Screening—with state-of-the-art computer vision techniques, this workshop seeks to foster collaboration between experts in biomedical science, AI, and computer vision. The workshop highlights the potential for revolutionizing drug discovery processes, driving faster and more accu-

rate identification of therapeutic targets, and expediting the development of treatments for complex diseases. Addressing pressing challenges like cancer, neurodegenerative disorders, and pandemics, the focus lies on leveraging AI to analyze high-dimensional biological data, enhancing our understanding of disease mechanisms and responses to therapies.

Vision Meets Physics: Synergizing Physical Simulation and Computer Vision

Organizers: Fangyin Wei, Donglai Xiang, Qianli Ma, Yifei Li, Ming Lin, Chenfanfu Jiang, Shenlong Wang, David Levin I.W., Tsung-Yi Lin

Date: 6/12/2025
Time: 9:00 - 17:00
Location: 104 A

Summary: This workshop explores the evolving intersection of computer vision and physics, where two competing perspectives—physics-based simulations versus data-driven approaches like video foundation models—seek to model the world effectively. By bringing together researchers from both fields, the event aims to foster collaboration, identify synergies, and advance applications in scientific research, generative AI, robotics, gaming, and extended realities (XR). Through presentations and discussions, the workshop will promote interdisciplinary dialogue to develop next-generation technologies that combine physics-based and data-driven methods, ultimately enhancing realistic simulations for immersive environments, automated tasks, and seamless virtual-physical integration.

5th Workshop on CV4Animals: Computer Vision for Animal Behavior Tracking and Modeling

Organizers: Leandra Brickson, Urs Waldmann, Shangzhe Wu, Anna Zamansky, Gengshan Yang, Bastian Wandt, Garvita Allabadi, George Martvel, Andre Telfer, Xiaoxuan Ma

Date: 6/12/2025
Time: 9:00 - 17:00
Location: 104 C

Summary: Many biological organisms have evolved to exhibit diverse behaviors, and understanding these behaviors is a fundamental goal of multiple disciplines including neuroscience, biology, animal husbandry, ecology, and animal conservation. These analyses require objective, repeatable, and scalable measurements of animal behaviors that are not possible with existing methodologies that leverage manual encoding from animal experts and specialists. Recently, computer vision has been making a significant impact across multiple disciplines by providing new tools for the detection, tracking, and analysis of animal behavior. This workshop brings together experts across fields to stimulate this new field of computer-vision-based animal behavioral understanding.

7th Safe Artificial Intelligence for All Domains

Organizers: Timo Sämann, Oliver Wasenmüller, Markus Enzweiler, Peter Schlicht, Stefan Milz, Thomas Stauner, Joachim Sicking, Claus Bahlmann

Date: 6/12/2025
Time: 9:00 - 17:00
Location: 104 D

Summary: After the success of ML and AI-based approaches in outperforming traditional vision algorithms, recently a lot of research effort is dedicated to understanding of the limitations and the general behavior of AI methods in a broad range of computer vision applications. Specifically for a successful introduction of ML and AI in a wider range of products, safety is often a top priority. Being able to ensure safety of ML based computer vision is key to unlock its potential in a broad range of safety related applications and future products. In domains like automotive, aviation and the medical domain, it paves the way towards systems with a greater degree of autonomy and assistance for humans.

Vision Language Models For All: Building Geo-Diverse and Culturally Aware Vision-Language Models

Organizers: Shravan Nayak, Mehar Bhatia, Qian Yang, Kanishk Jain, Rabiul Awal, David Ifeoluwa Adelani, Spandana Gella, Siva Reddy, Vered Shwartz, Yash Goyal, Sjoerd van Steenkiste, Karolina Stanczak, Aishwarya Agrawal

Date: 6/12/2025
Time: 9:00 - 17:00
Location: 104 E

Summary: The CVPR community has long focused on evaluating AI systems for their general scene-understanding capabilities. However, as these models are deployed globally, it is essential that they also understand cultural concepts and values, ensuring they cater to the diverse needs of users. This workshop expands computer vision frontiers by bringing together researchers from computer vision, natural language processing, AI ethics, and cultural anthropology to discuss how we can build geo-diverse and culturally aware vision-language models (or AI models in general). Specifically, the workshop will focus on evaluating the types of tasks, benchmarks, and metrics we should develop to advance AI systems' capabilities in this area and explore promising approaches to overcome the challenges. Second, the workshop will benchmark progress in geo-diverse and cultural understanding of vision-language models through the CulturalVQA and GlobalRG challenges, which will test critical abilities such as visual question answering and grounding in culturally diverse scenarios. The insights from this workshop extend beyond computer vision, with significant implications for fields like healthcare, education, and e-commerce, where culturally aligned AI can enhance user experiences. Additionally, the workshop aims to inspire further research in AI ethics, fairness, and responsible AI deployment.

21th Workshop on Perception Beyond the Visible Spectrum (PBVS'2025)

Organizers: Riad I. Hammoud
Date: 6/12/2025
Time: 9:00 - 17:00
Location: 205 C

Summary: This established workshop series uniquely concentrates on computer vision utilizing modalities beyond the visible spectrum, including x-ray, infrared, SONAR, thermal, hyperspectral, and radar. Addressing a broad range of topics beyond specific applications, this CVPR PBVS'2025 workshop moves beyond merely adapting existing CV algorithms to explore innovative perspectives on exploitation and integration of non-visible spectrum data with traditional sensors. The workshop features multiple keynote speakers and hosts highly competitive challenges in CV, deep learning, VLM and multi-modal fusion.

Agent in Interaction, from Humans to Robots

Organizers: Yufei Ye, Homanga Bharadhwaj, Dandan Shan, Wei-Chiu Ma, Shubham Tulsiani, Abhinav Gupta, Michael J. Black

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 213

Summary: The goal of this workshop is to build communication among researchers who study human modeling and robotics, both of which can be considered as physical agents that interact with the world. In recent years, computer vision researchers have focused on creating digital twins of humans in virtual environments that behave like real humans, and at the same time roboticists have worked on building physical agents capable of interacting with the real world. We believe that the progress in one field can greatly benefit the other. On one hand, virtual humans can be considered as a special form of a robotic agent. On the other hand, robots can learn manipulation and locomotion from human demonstrations, including from simulated humans. Through the proposed workshop, we aim to bring these two fields together and explore common challenges - such as contact modeling, motion prediction, overcoming data paucity, the role of large-scale models, etc. We hope that our workshop will provide a suitable platforms for inspiring new research directions and solutions in this space.

Workshop on Computer Vision for Microscopy Image Analysis

Organizers: Mei Chen, Dimitris N. Metaxas, Steve Finkbeiner, Oren Kraus

Date: 6/12/2025

Time: 9:00 - 17:00

Location: 214

Summary: The emphasis of the 2025 CVMI workshop will be leveraging the advances in large foundational models (LFMs) to improve the multimodal analysis of microscopy and omics data with enhanced explainability. We aim to promote discussions on methods that offer explainability, interpretability, and usefulness for disease classification, prediction, and treatment, while integrating domain knowledge with LFMs. This workshop will bring together computer vision experts from academia, industry, and government who have made progress in developing computer vision tools for microscopy image analysis. It will provide a comprehensive forum on this topic and foster in-depth discussion of technical and application issues as well as cross-displinary collaboration. It will also serve as an introduction to researchers and students curious about this important and fertile field.

AI for Content Creation

Organizers: James Tompkin, Deqing Sun, Lu Jiang, Lingjie Liu, Fitsum Reda, Jun-Yan Zhu, Krishna Kumar Singh

Date: 6/12/2025

Time: 9:00 - 17:00

Location: Grand A1

Summary: AI for content creation plays a crucial role in domains such as photography, videography, virtual reality, gaming, art, design, fashion, and advertising, and lies at the intersection of computer vision, machine learning, computer graphics, and design. This workshop will provide attendees with a slice of cutting-edge techniques within this rapidly evolving field, considering both the fundamental technologies and practical challenges faced by designers and content creators, and will show successful applications of AI and deep learning in content creation. With invited speakers of world-class expertise in content creation, up-and-coming researchers, and posters from authors of submitted workshop papers, the workshop will help all to engage in a day filled with learning, discussion, and network building.

Foundation Models for V2X-Based Cooperative Autonomous Driving

Organizers: Walter Zimmer, Ross Greer, Max Peter Ronecker, Chuheng Wei, Haibao Yu, Rui Song, Xingcheng Zhou, Holger Caesar, Julie Stephany Berrio Perez, Alina Roitberg, Daniel Watzenig, Mohan Trivedi, Alois Knoll

Date: 6/12/2025

Time: 9:00 - 17:00

Location: Grand A2

Summary: The DriveX Workshop 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.

The 2nd Joint Egocentric Vision (EgoVis) Workshop

Organizers: Siddhant Bansal, Antonino Furnari, Tushar Nagarajan, Dima Damen, Giovanni Maria Farinella, Kristen Grauman, Jitendra Malik, Richard Newcombe, Marc Pollefeys, Yoichi Sato, David J. Crandall

Date: 6/12/2025

Time: 9:00 - 17:00

Location: Grand B1

Summary: Egocentric devices like wearable cameras, smart glasses, and AR/VR headsets are rapidly evolving to automatically recognize user actions, environments, gestures, and social interactions. This workshop serves as a central gathering point for the egocentric vision community to exchange ideas and explore this fast-growing field. It features challenges across five major datasets (EPIC-Kitchens, Ego4D, Ego-Exo4D, HoloAssist, HD-EPIC), keynote talks from leading experts, abstract presentations on emerging ideas, EgoVis award to seminal papers from 2023/2024, and poster sessions on pivotal papers—offering a comprehensive look at the future of egocentric perception and wearable AI.

Mechanistic Interpretability for Vision

Organizers: Tamar Rott Shaham, Yossi Gandelsman, Joanna Materzynska, Rohit Gandikota, Amil Dravid, Ashkan Khakzar, Eli Shechtman, Philip Torr

Date: 6/12/2025
Time: 9:00 - 17:00
Location: Grand C1
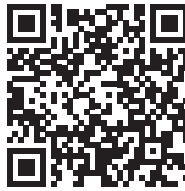
Summary: Mechanistic interpretability has proven transformative in understanding Large Language Models (LLMs), revealing intricate computational patterns and internal mechanisms. However, this powerful approach remains underutilized in computer vision, despite its potential to uncover fundamental properties of visual processing in neural networks. In this workshop, we will discuss methods to advance our understanding of vision models through mechanistic interpretability, potentially revealing novel insights about learned visual representations and emergent algorithms.

5th International Workshop on Event-based Vision

Organizers: Guillermo Gallego, Kostas Daniilidis, Cornelia Fermuller, Davide Migliore, Daniele Perrone

Date: 6/12/2025
Time: 9:00 - 17:00
Location: Grand C2

Summary: The Event-based Vision Workshop at CVPR is the premier venue for discussing exciting new ideas about neuromorphic cameras and their processing methods. It covers the sensing hardware, as well as the processing, data, and learning methods needed to take advantage of event-based cameras. The workshop aims to highlight an emerging field with the potential to overcome many of the limitations of frame-based systems (speed, power consumption, robustness to HDR illumination, etc.). This forum fosters community building around these novel cameras, capitalizing on a growing interest and increasing contributions at the main conference. Furthermore, the workshop seeks to connect with a broader audience by highlighting interdisciplinary links between computer vision, robotics, artificial intelligence, computational neuroscience, and psychology, as event cameras facilitate research into replicating the efficiency and robustness of the human visual system.

Exploring the Next Generation of Data

Organizers: Nadine Chang, Maying Shen, Jose M. Alvarez, Sifei Liu, Rafid Mahmood, Despoina Paschalidou

Date: 6/12/2025
Time: 9:00 - 17:00
Location: Davidson A2

Summary: Data is more crucial than ever, enabling the first generation of deep learning models to the new generation of foundation models. These foundation models are rapidly incorporating into several safety critical applications of human life. Thus, the large volume of data they rely on must be high-quality for safe model development. Due to the sheer volume of raw data, it is necessary to obtain a scalable ability to rank and select data by its inherent quality and value for both generic and specific tasks. Recently, foundation models themselves are used to discover even more data to feed into more foundation model training. This cyclic relationship between data and foundation models introduces another layer of complexity and biases to consider. Overall, this enormous challenge to discover the next generation of data requires several considerations: definition of quality data, bias-free data, scalability, generating data, ethical data gathering, continuous data gathering, and hallucination free foundation models for data mining. In this workshop, 7 leading experts across academia and industry will discuss how to tackle this large challenge together.

Open-World 3D Scene Understanding with Foundation Models

Organizers: Francis Engelmann, Ayça Takmaz, Jonas Schult, Alexandros Delitzas, Elisabetta Fedele, Zuria Bauer, Aikaterini Adam, Or Litany, Federico Tombari, Marc Pollefeys, Leonidas Guibas

Date: 6/12/2025
**Time:** 13:00 - 17:00
Location: 105 A

Summary: LThe developments in computer vision, graphics, and robotics have jointly spurred calls for next-generation AI systems that physically interact with their surroundings. Current research advances encompass 3D representations, large-scale foundation models, and end-to-end VLA approaches, but fundamental questions remain on how best to sustain environment comprehension, align efforts from diverse fields, and integrate scene understanding techniques to enhance physical interaction. The workshop seeks to unite current efforts, educate an interdisciplinary workforce with expertise across fields, and promote future developments in embodied and general AI.

Visual Modeling Challenges for 2D-3D Virtual Try-On

Organizers: Akshay Gadi Patil, Vidya Narayanan, Haoye Dong, Gerard Pons-Moll, Ming Lin

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 105 B
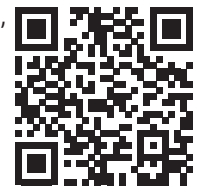
Summary: Virtual Try-On (VTO) is an emerging consumer application that enables users to perceive products on their individual bodies in a virtual or mixed reality space. This has been driven by advances in 3D modeling of humans and objects (such as furniture, shoes, garments etc.). These experiences enable users to visualize and "feel" products, especially in the beauty, fashion, and accessories space, virtually, before making purchases. This gives buyers opportunities to customize and personalize products. Successful integration of realistic try-on experiences on e-commerce websites can have significant environmental impact by reducing the need to return products, improving satisfaction of purchased items, and improving accessibility. Beyond fashion and beauty, adoption of try-on-like experiences, especially in an Augmented Reality (AR) setting, can be seen in applications such as remote collaboration, gaming, education, and health-care. Enabling such applications, however, is not without its unique challenges that necessitate solutions inspired by ideas in computer vision, 3D modeling and reconstruction (single and multi-view), geometry processing, and more recently, generative AI. Our workshop will focus on these key technical areas of strong interest for the CVPR community. Furthermore, the user-centric nature of this domain and inherent application to fashion requires researchers and technologists to understand the nature of bias in their data (catalog data is skewed towards a narrow BMI-range and demographic), trust and data-security issues (body image / data is often a necessary input) and the mental-health impact of try-on products for their users.

ReGenAI: 2nd Workshop on Responsible Generative AI

Organizers: Adriana Romero-Soriano, Reyhane Askari-Hemmat, Melissa Hall, Michal Drozdzal, Ye Zhu, Agata Lapedriza, Arantxa Casanova, Negar Rostamzadeh, Utsav Prabhu, Pinar Yanardag
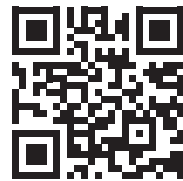
Date: 6/12/2025

Time: 13:00 - 17:00

Location: 106 A

Summary: This is the second edition of Responsible Generative AI (ReGenAI) workshop. This workshop aims to bring together researchers, practitioners, and industry leaders working at the intersection of generative AI, data, ethics, privacy and regulation, with the goal of discussing existing concerns, and brainstorming possible avenues forward to ensure the responsible progress of generative AI. We hope that the topics addressed in this workshop will constitute a crucial step towards ensuring a positive experience with generative AI for everyone.

Physics-inspired 3D Vision and Imaging

Organizers: Anagh Malik, Benjamin Attal, Dor Verbin, Xuaner Zhang, James Tompkin

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 106 C

Summary: 3D computer vision has become fundamental to technologies ranging from medical imaging to astronomy and from AR/VR to embodied intelligence. New sensors and imaging modalities like structured-light, time-of-flight, and light field microscopy are being developed to make 3D vision more tractable; but even with new types of sensor data, many problems in 3D vision tend to be ill-posed and hence to solve them we often rely on heuristics or data-driven priors. Unfortunately, these priors can fail in certain cases, especially for problems where ground truth data is not available, or for niche sensors where capturing large datasets is not feasible. A promising, but often overlooked, alternative is to incorporate knowledge of physics (e.g. physical light transport) into 3D computer vision algorithms, which can better constrain the solutions that they produce.

The goal of this workshop is to highlight work in 3D computer vision and imaging that makes use of physics-inspired modeling and physical-priors, showcasing their importance even with the prevalence of neural priors and big data. Examples include methods that apply physics-based approaches to inverse rendering, 3D microscopy, tomography, and light-in-flight imaging; or methods that combine such approaches with novel tools like neural radiance fields (NeRFs), 3D Gaussian Splatting (3DGS), and generative image/video models.

The 7th Workshop on Precognition: Seeing through the Future

Organizers: Khoa Luu, Nemanja Djuric

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 107 A

Summary: Vision-based detection and recognition studies have been recently achieving highly accurate performance and were able to bridge the gap between research and real-world applications. Beyond these well-explored detection and recognition capabilities of modern algorithms, vision-based forecasting will likely be one of the next big research topics in the field of computer vision. Vision-based prediction is one of the critical capabilities of humans, and the potential success of automatic vision-based forecasting will empower and unlock human-like capabilities in machines and robots.

One example application is in autonomous driving technologies, where vision-based understanding of a traffic scene and prediction of movement of traffic actors is a critical piece of the autonomous puzzle. Another area where vision-based prediction is used is the medical domain, allowing deep understanding and prediction of future medical conditions of patients. However, despite its potential and relevance for real-world applications, visual forecasting or precognition has not been the focus of new theoretical studies and practical applications as much as detection and recognition problems.

Through the organization of this workshop, we aim to facilitate further discussion and interest within the research community regarding this nascent topic. This workshop will discuss recent approaches and research trends not only in anticipating human behavior from videos but also precognition in multiple other visual applications, such as medical imaging, healthcare, human face aging prediction, early event prediction, autonomous driving forecasting, etc.

The 6th Annual International Workshop And Prize Challenge on Agriculture-Vision: Challenges and Opportunities for Computer Vision in Agriculture

Organizers: Chris Padwick, James P. Ostrowski, Ripudaman Singh Arora, Naira Hovakimyan, Humphrey Shi, Pan Zhao, Jing Wu, Leandro G. Almeida, Ignacio Ciampitti, Leonardo Lemes Bosche, Mojdeh Saadati, Smruti (Amar) Amayjyoti, Avinash Raju, Tejaishwarya Gagadam

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 107 B

Summary: With the recent success of computer vision and deep learning in various applications, there has been significantly increasing attention towards its use in agriculture, presenting both significant economic and social opportunities. This The 6th Annual International Workshop And Prize Challenge on Agriculture-Vision aims to foster research and applications at the intersection of computer vision and agriculture, addressing challenges in real-world agricultural scenarios, with a strong record from prior editions at CVPR 2020-2024. The workshop will feature a computer vision challenge, and invited speakers from diverse academic and industry backgrounds including computer vision, robotics, agriculture, and top industry practitioners. This event provides a platform to showcase current progress in interdisciplinary areas and encourage further research and development of Foundation Models In Agriculture.

8th Workshop and Competition on Affective & Behavior Analysis in-the-wild

Organizers: Dimitrios Kollias, Stefanos Zafeiriou, Irene Kotsia, Panagiotis Tzirakis, Eric Granger, Simon L Bacon, Marco Pedersoli, Alan Cowen

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 202 A

Summary: The ABAW Workshop is a premier platform highlighting the latest advancements in multimodal analysis, generation, modeling, and understanding of human affect and behavior in real-world, unconstrained environments. It emphasizes cutting-edge systems that integrate facial expressions, body movements, gestures, natural language, voice and speech to enable impactful research and practical applications. The workshop fosters interdisciplinary collaboration across fields such as computer vision, AI, human machine interaction, psychology, robotics, ethics & healthcare. The workshop further addresses complex challenges like algorithmic fairness, demographic bias & data privacy, making it a vital forum for building equitable, generalizable & human-centered AI systems. By uniting experts from academia, industry & government, the workshop promotes innovation, drives knowledge exchange, and inspires new directions in affective computing, behavior modelling and understanding & human-computer interaction. Finally, the Workshop includes a Competition with 6 challenges, including valence-arousal estimation, basic & compound expression recognition, action unit detection, emotional mimicry intensity estimation and ambivalence/hesitancy recognition.

11th IEEE International Workshop on Computer Vision in Sports

Organizers: Rikke Gade, Anthony Cioppa, Thomas B. Moeslund, Graham Thomas, Adrian Hilton, Jim Little, Michele Merler, Silvio Giancola

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 108

Summary: Sports is said to be the social glue of society. It allows people to interact irrespective of their social status, age etc. With the rise of the mass media, a significant quantity of resources has been channeled into sports in order to improve understanding, performance, and presentation. For example, areas like performance assessment, which were previously mainly of interest to coaches and sports scientists are now finding applications in broadcast and other media, driven by the increasing use of on-line sports viewing which provides a way of making all sorts of performance statistics available to viewers. Computer vision has recently started to play an important role in sports as seen in for example football where computer vision-based graphics in real-time enhances different aspects of the game. Computer vision algorithms have a huge potential in many aspects of sports ranging from automatic annotation of broadcast footage, through to better understanding of sport injuries, coaching, and enhanced viewing. So far, the use of computer vision in sports has been scattered between different disciplines. The ambition of this workshop is to bring together practitioners and researchers from different disciplines to share ideas and methods on current and future use of computer vision in sports.
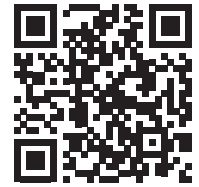
4th Monocular Depth Estimation Challenge

Organizers: Matteo Poggi, Fabio Tosi, Ripudaman Singh Arora, Anton Obukhov, Jaime Spencer, Chris Russell, Simon Hadfield, Richard Bowden

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 109

Summary: Monocular depth estimation (MDE) is an important low-level vision task with applications in fields such as augmented reality, robotics, and autonomous vehicles. In 2024, the field was dominated by generative approaches, with DepthAnything representing the transformer-based solution and Marigold being a denoising diffusion model based on the popular Text-to-Image LDM Stable Diffusion. Even before that, there has been an increased interest in self-supervised systems capable of predicting the 3D scene structure without requiring ground-truth LiDAR training data. The automotive industry accelerated the development of these systems thanks to the vast quantities of data and the ubiquity of stereo camera rigs. However, the evaluation process has remained focused on in-domain evaluation, relying on simple metrics and sparse LiDAR data.
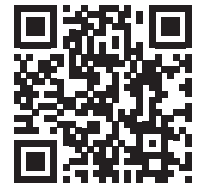
Multi-Modal Learning for Materials Science

Organizers: Weike Ye, Santosh K Suram, Helge Sören Stein, Mathew J. Cherukara, Jiarui Zhang

Date: 6/12/2025

Time: 13:00 - 17:00

Location: 110 A

Summary: The Multimodal Learning for Materials Science (MM4Mat) workshop aspires to shape and nurture interdisciplinary discussions and collaborations between machine learning experts and materials scientists. By bringing together world-leading experts in multimodal AI and data-driven materials discovery, this workshop will highlight innovative methodologies and their real-world applications in materials science.
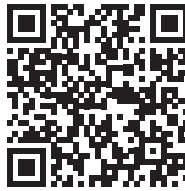
Workshop on 3D Human Understanding

Organizers: Qianli Ma, Siwei Zhang, Shashank Tripathi, Rawal Khirodkar, Yan Zhang, Yao Feng, Georgios Pavlakos, Siyu Tang

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 110 B

Summary: Understanding humans in 3D is essential in computer vision, graphics, robotics, and relevant fields. The complexities associated with analyzing human behaviors, postures, and interactions are multifaceted, making them a challenging yet crucial aspect of vision research. Decoding these intricacies is essential for a range of applications, such as sophisticated augmented reality interfaces, real-time safety monitoring systems, high-fidelity digital avatar modeling, and lifelike human motion generation. Recent developments in the field have brought to light new challenges, including multi-human interaction analysis, fine-grained motion capture, clothed body reconstruction, virtual try-ons, tracking under extreme occlusion crowding, LLM-based human understanding, and motion and behavior synthesis. Through this workshop, our aim is to shed light on these evolving challenges and foster a dialogue that could pave the way for innovative solutions. We seek to critically assess the existing methodologies, understand their limitations, and collaboratively discuss novel approaches to push the boundaries of 3D human understanding in computer vision.

1st Workshop on Experimental Model Auditing via Controllable Synthesis

Organizers: Viraj Uday Prabhu, Prithvijit Chattopadhyay, Sriram Yenamandra, Hao Liang, Krish Kabra, Guha Balakrishnan, Judy Hoffman, Pietro Perona

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 208 B

Summary: This workshop is dedicated to exploring techniques for auditing the behavior of machine learning models – including (but not limited) to performance, bias, and failure modes – by the controlled synthesis (via generation or simulation) of data. Of special interest are algorithms for generating data (images, text, audio, etc.) and benchmarking that provide reliable insights into model behavior by minimizing the impact of potential confounders. We also welcome work on the broader topic of using synthetic or quasi-synthetic data for model debugging, broadly construed, with the goal of providing a venue for interdisciplinary exchange of ideas on this emerging topic.

AI for Creative Visual Content Generation, Editing and Understanding

Organizers: Ozgur Kara, Fabian Caba Heilbron, Anyi Rao, Victor Escorcia, Ruihan Zhang, Mia Tang, Dong Liu, Maneesh Agrawala, James Matthew Rehg

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 207 A-D

Summary: Visual content creation is booming, yet producing engaging visual content remains a challenging task. This workshop aims to highlight machine learning technologies that accelerate and enhance creative processes in visual content creation and editing, including im-

age animation, text-to-visual content generation, and content translation. Moreover, we believe that advancing technology to better understand edited visual content can enable novel platforms for creating compelling media. We seek to bridge the gap between technical and creative communities by bringing together researchers from computer vision, graphics, and the arts, fostering interdisciplinary collaboration and exploring opportunities in this under-explored area.
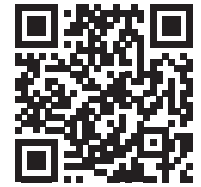
2nd Workshop on Efficient and On-Device Generation (EDGE)

Organizers: Felix Juefei-Xu, Tingbo Hou, Yang Zhao, Licheng Yu, Zhisheng Xiao, Xiaoliang Dai, Qifei Wang, Tao Xu, yanwu xu, Ali Thabet, qiang liu, Xuan Ju, Ruiqi Gao, Xi Yin, Haolin Jia, Xide Xia, Peizhao Zhang, Peter Vajda

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 208 A

Summary: The Second Workshop on Efficient and On-Device Generation (EDGE) at CVPR 2025 will focus on the latest advancements of generative AI in the computer vision domain, with an emphasis on efficiencies across multiple aspects. We encourage techniques that enable generative models to be trained more efficiently and run on resource-constrained devices, such as mobile phones and edge devices. Through these efforts, we envision a future where these permeating generative AI capabilities become significantly more accessible with virtuous scalability and plateauing carbon footprint.

The 4th Workshop on Transformers for Vision

Organizers: Gedas Bertasius, Rohit Girdhar, Zhiding Yu, Lucas Beyer, Gül Varol, Alaaeldin El-Nouby, Tyler Zhu, Feng Cheng, Yan-Bo Lin, Md Mohaiminul Islam, Yi-Lin Sung, Jaemin Cho, Ce Zhang, Yue Yang, Ziyang Wang, Mohit Bansal, Shilong Liu, Hao Zhang, Fuxiao Liu, Xiaolong Li, Subhashree Radhakrishnan, Shiyi Lan, Jose M. Alvarez

Date: 6/12/2025
Time: 13:00 - 17:00
Location: 209 A-C

Summary: Transformers have recently emerged as promising and versatile deep neural architecture in various domains. Since the introduction of Vision Transformers (ViT) in 2020, the vision community has witnessed an explosion of transformer-based computer vision models with applications ranging from image classification to dense prediction (e.g., object detection, segmentation), video, self-supervised learning, 3D and multi-modal learning. This workshop presents a timely opportunity to bring together researchers across computer vision and machine learning communities to discuss the opportunities and open challenges in designing transformer models for vision tasks.

Domain Generalization: Evolution, Breakthroughs, and Future Horizons

Organizers: Muhammad Haris Khan, Biplab Banerjee
Date: 6/12/2025
Time: 13:00 - 17:00
Location: 211

Summary: Domain generalization has established itself as a vital and viable research approach for mitigating domain shifts. This method is particularly appealing because it makes minimal assumptions about target domain data and eliminates the need for re-training or adaptation steps. As a result, DG is increasingly crucial for enhancing the robustness of AI models in real-world applications. Despite recent advancements in visual domain generalization, the scope of research has predominantly been limited to single tasks (such as classification or segmentation), single modalities (specifically images), straightforward domain shifts, and closed-world scenarios. However, real-world applications often involve more complex conditions, including multiple modalities, diverse tasks, advanced neural architectures, extreme environmental conditions, and resource-limited scenarios—all areas that have seen limited exploration in DG research. In this workshop, we aim to expand the boundaries of DG by fostering new ideas and stimulating discussion on these relatively unexplored and challenging aspects of domain generalization.

Workshop on Perception for Industrial Robotics Automation

Organizers: Vage Taamazyan, Aarrushi Shandilya, Agastya Kalra, Huaijin Chen, Krzysztof Marcin Choromanski, Martin Sundermeyer, Phil Nelson, Satya Mallick, Stan Birchfield, Tim Salzmann, Tomas Hodan, Yang Qian
Date: 6/12/2025
Time: 13:00 - 17:00
Location: 210

Summary: This workshop addresses the gap between cutting-edge computer vision research and its practical application in industrial robotics, specifically addressing challenges in tasks like reliable, scalable, and cost-effective bin picking. The workshop brings together researchers and practitioners to discuss topics including 3D scene understanding, embodied AI, and robot learning, focusing on developing robust solutions by considering factors like embodiment, camera choice, and data needs. Complementing the workshop, the Perception Challenge for Bin Picking offers a practical platform for participants to tackle real-world 6DoF pose estimation problems using a robot-in-the-loop evaluation, providing a more realistic performance assessment than traditional vision-only metrics. The workshop and challenge together aim to accelerate the adoption of vision-guided robotics and enhance industrial automation efficiency.

Catch UAVs that Want to Watch You: Detection and Tracking of Unmanned Aerial Vehicle (UAV) in the Wild and the 4th Anti-UAV Workshop & Challenge

Organizers: Jian Zhao, Jianan Li, Lei Jin, Miguel Bordallo Lopez, Liang Li, Xinyi Ying, Tianyang Xu, Yawen Cui, Sadaf Gulshad, Shin'ichi Satoh, Hongyuan Zhang, Jianshu Li, Jiaojiao Zhao, Zhiqi Cheng, Mengmi Zhang, Zaiping Lin, Miao Li, Zheng Wang, Zechao Li, Yunchao Wei, Junliang Xing, Shengmei Jane Shen, Qi Wang, Xuelong Li

Date: 6/12/2025
Time: 13:00 - 17:00
Location: Davidson C1

Summary: Video object detection and tracking in wild scenarios are fundamental yet challenging problems in computer vision and public security. This workshop serves as a forum for researchers across object detection, visual tracking, and related fields to share advancements and innovative concepts relevant to UAV detection and tracking.

VizWiz Grand Challenge

Organizers: Danna Gurari
Date: 6/12/2025
Time: 13:00 - 17:00
Location: Davidson C2

Summary: Our goal for this workshop is to educate researchers about the technological needs of people with vision impairments while empowering researchers to improve algorithms to meet these needs. A key component of this event will be to track progress on five dataset challenges, where the tasks are to recognize objects in few-shot learning scenarios, answer visual questions, ground answers, recognize visual questions with multiple answer groundings, locate objects in few-shot learning scenarios, classify images in a zero-shot setting. The second key component of this event will be a discussion about current research and application issues, including invited speakers from both academia and industry who will share their experiences in building today's state-of-the-art assistive technologies as well as designing next-generation tools.

The 1st Workshop on Enforcing Geometric, Physical, Topological, and Functional Inductive Bias in 3D Generation

Organizers: Qixing Huang, Congyue Deng, Lin Gao, Hanwen Jiang, Lingjie Liu, Biao Zhang, Ruqi Huang, Anand Bhattad, Roni Sengupta, Despoina Paschalidou
Date: 6/12/2025
Time: 13:00 - 17:00
Location: Davidson C3

Summary: The current era of generative AI thrives on scaling data and models, yet fundamental challenges remain, particularly in preserving geometric, physical, topological, and functional priors in 3D generation. This workshop aims to unite researchers to advance the integration of inductive biases in 3D generative models, providing tools that are simple, efficient, and impactful. By addressing the challenges of modeling inductive biases—which require interdisciplinary knowledge and often incur training costs—this workshop seeks to foster a community dedicated to these methods. As generative AI faces limits in training data and model scalability, incorporating inductive biases offers a path forward, enabling smaller models with reduced data requirements. While the focus is on 3D inductive biases due to their abundance of priors, the workshop's relevance extends to the broader computer vision community. Key topics include embedding geometrical, physical, topological, and functional operators into networks, enforcing constraints with losses, discovering relations from data, and 3D editing with guidance.

City Map
Entrance Points from Hotels

COMMERCE

RENAISSANCE HOTEL

BROADWAY

8TH AVENUE

CAMBRIA HOTEL

7TH AVENUE

EMBASSY SUITES & 1 HOTEL

6TH AVENUE

REP. JOHN LEWIS WAY SOUTH

HILTON HOTEL

TAXI STAND

DEMONBREUN

JW MARRIOTT HOTEL

4

P3

SHUTTLE DROP-OFF

3

2

PARKING GARAGE

P2

MUSIC CITY CENTER

P1

THE WESTIN HOTEL

1

OMNI HOTEL

KOREAN VETERANS BOULEVARD (KVB)

8TH AVENUE S.

LaFAYETTE

6TH AVENUE

MARGARITAVILLE HOTEL

REP. JOHN LEWIS WAY SOUTH

AC HOTEL

THE JOSEPH HOTEL

4TH AVENUE

1 Level 1

Rep. John Lewis Way South and Korean Veterans Blvd.

AC Hotel by Marriott
Drury Plaza Hotel
Hampton Inn & Suites
Holiday Inn & Suites
Hyatt Place
The Joseph Hotel
Margaritaville Hotel
Omni Nashville
Residence Inn by Marriott

2 Level 1

Rep. John Lewis Way South and Demonbreun

Four Seasons Hotel
Hilton Nashville Downtown
Renaissance Hotel

3 Level 2

6th Ave and Demonbreun

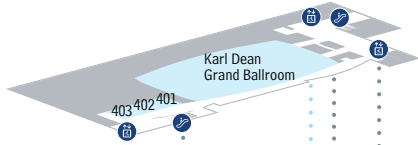Hotel shuttle drop-offs
Parking Garage

4 Level 3

8th Ave and Demonbreun

1 Hotel Nashville
Cambria Hotel
Embassy Suites by Hilton
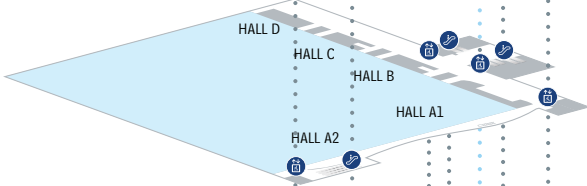JW Marriott
The Westin
Renaissance Hotel

Music City Center.

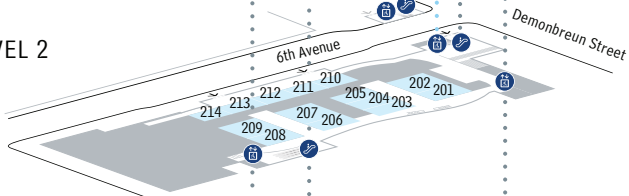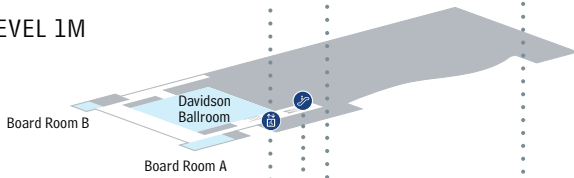201 Rep. John Lewis Way South
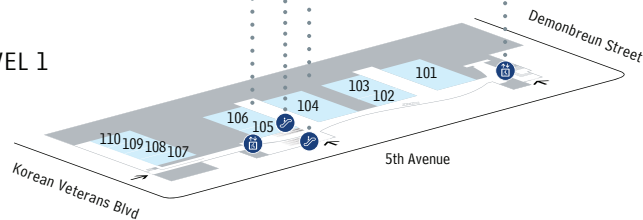
Nashville, TN 37203

615.401.1400

MAPS

LEVEL 4

Karl Dean Grand Ballroom

403 402 401

LEVEL 3

HALL D
HALL C
HALL B
HALL A1
HALL A2

6th Avenue

Demonbreun Street

LEVEL 2

214 213 212 211 210
205 204 202 201
209 208 207 206 203

LEVEL 1M

Board Room B

Davidson Ballroom

Board Room A

LEVEL 1

Demonbreun Street

103 101
102
106 104
105
110 109 108 107

Korean Veterans Blvd

5th Avenue

Icon	Label	Icon	Label
🚹	Men's Restroom	🚼	Family Restroom	🛗 Escalator
🚺	Women's Restroom			🛗 Elevator

LEVEL 1

Korean Veterans Blvd.

B A 110
109
108
B A 107

C B A 106
E D C 104
105

C B A 103
E D C 101
102

Ramp
To Davidson Ballroom

Demonbreun Street

LEVEL 4

Karl Dean Grand Ballroom

C2
A1
A2
C1
B1
403 401
402 B A

LEVEL 3

HALL D
HALL C
HALL B
HALL A2 HALL A1

To 6th Street & Parking Garage
To Grand Ballroom
To Lower Levels

LEVEL 2

Parking Garage

6th Avenue

Demonbreun Street

Ramp Down to Davidson Ballroom
214 213 212 211 210

C B A 209
D C B A 207
C B A 205
C B 202

208 B A
206 B A
204
203 B A 201

Stairs Down to Davidson Ballroom

LEVEL 1M

Board Room B

Ramp Up

C2 C3
C1
B2 Davidson
B1 Ballroom
A3
A2 A1

Board Room A

CVPR *Nashville* JUNE 11-15, 2025