June 3–7, 2026 | Denver Convention Center

NVIDIA Experiences at CVPR 2026: **Tutorials, Workshop Talks, and Conference Talks**

WEDNESDAY 6/3 Workshops and Tutorials	THURSDAY 6/4 Workshops and Tutorials	FRIDAY 6/5 Main Conference	SATURDAY 6/6 Main Conference	SUNDAY 6/7 Main Conference
Full Day | Room 603 **Workshop on Autonomous Driving (WAD)** Sanja Fidler	Full Day | Mile High 2A **Agents in Interaction, From Humans to Robots** Umar Iqbal	9:15–10:30 a.m. | Mile High Ballroom 1A–2A **Oral 1C: Efficient Reasoning** Advancing Image Classification With Discrete Diffusion Classification Modeling #13 | O. Belhasin; M. Elad (NVIDIA/Technion)	10:15–11:30 a.m. | Bluebird Ballroom **Oral 3A: Generative Diffusion Modeling** PixelDiT: Pixel Diffusion Transformers for Image Generation #3 | W. Xiong; W. Nie; Y. Sheng; S. Liu	10:20 a.m. | NVIDIA Booth Theater #211 **Meet the Researcher: Deep Learning and Computer Vision** Shalini DeMello, NVIDIA Research
Full Day | 2E-2H **Deployment of Foundation Models for Embodied AI** Jan Kautz	Full Day | Exhibit Hall A–3B **Video Generative Models: Benchmarks and Evaluation** Yan Wang	10:20 a.m.–noon | NVIDIA Booth Theater #211 **Meet the Researcher: Vice President of Cosmos Lab** Ming-Yu Liu	10:20 a.m. | NVIDIA Booth Theater #211 **Accelerating Autonomous Driving Research With the Alpamayo Open Ecosystem** Boris Ivanovic	10:40 a.m. | NVIDIA Booth Theater #211 **Research Spotlight: 4D-RGPT: Toward Region-Level 4D Understanding via Perceptual Distillation** Min-Hung Chen
Full Day | Room 503 **Bridging Vision, Language, and Action: What's Missing in Actionable Visual Perception for Robotics** Marco Pavone, Ming-Yu Liu	Full Day | Room 603 **Exploring the Next Generation of Data (NeXD)** Tsung-Yi Lin	10:40 a.m. | NVIDIA Booth Theater #211 **Research Spotlight: Cosmos 3** Ming-Yu Liu	10:40 a.m. | NVIDIA Booth Theater #211 **The Next Generation of AV Foundation Models: Alpamayo 2 Super** Yan Wang	11 a.m. | NVIDIA Booth Theater #211 **Research Spotlight: Learning Physical Simulation of Neural Geometry** Nicholas Sharp
Full Day | Mile High 3A **AUTOPILOT Workshop** Jose Alvarez	Full Day | Room 604 **ERA: Embodied Reasoning in Action** Luke Song	11:45 a.m.–1:30 p.m. | Booth Theater **Professor Tech Talks**	11 a.m. | NVIDIA Booth Theater #211 **Closed-Loop Reinforcement Learning With AlpaGym** Yuxiao Chen	11:20 a.m. | NVIDIA Booth Theater #211 **Research Spotlight:**
Half Day: Morning | Room 709 **Embedded Vision Workshop 2026** Bowen Wen	Half Day: Morning | Mile High 2B **Tutorial: The Full Stack of Physical AI** R. Lo; J. Nunez; C. Yato; S. Huang; M. Patel	1–2:15 p.m. | Four Seasons Ballroom **Oral 2B: Materials and Lighting** PPISP: Physically-Plausible Compensation of Photometric Variations in Radiance Fields #11 | I. Deutsch; N. Moënne-Loccoz; G. State; Ž. Gojčič	11:20 a.m. | NVIDIA Booth Theater #211 **Benchmarking Progress in the Reasoning Era** Boris Ivanovic	2–3:15 p.m. | Bluebird Ballroom **Oral 6A: Geometric Learning** Learning Convex Decomposition via Feature Fields #3 | M. A. Uy; N. Sharp
Half Day: Morning | Room 106 **LatinX in Computer Vision (LatinX in CV)** Paula Ramos	Half Day: Afternoon | Mile High 2C **4D Digital Twins: Real-to-Sim-to-Real for Physical AI** Francis Williams	1:40 p.m. | NVIDIA Booth Theater #211 **Meet the Researcher: Deep Learning, Computer Vision, and Simulation** Sanja Fidler	1:40–1:55 p.m. | NVIDIA Booth Theater #211 **Meet the Researcher: Deep Learning and Computer Vision** Frank Wang, NVIDIA Research Director	2–3:15 p.m. | Mile High Ballroom 1A–2A **Oral 6C: Medical Vision** Medic-AD: Towards Medical Vision-Language Model's Clinical Intelligence #15 | G. Wong; K.C. Cheung
Half Day: Morning | Room 109 **CVPR Workshop on Subtle Visual Computing Editing Physiological Signals in Videos Using Latent Representations** T. Zhou; A. Paruchuri; J. Spjut; K. Aksit		2 p.m. | NVIDIA Booth Theater #211 **Partner Spotlight: Path Robotics—Physical AI for Manufacturing** Alexander Lonsberry	2–3:15 p.m. | Four Seasons Ballroom **Oral 4B: Embodied and Agentic Intelligence** NitroGen: An Open Foundation Model for Generalist Gaming Agents #8 | L. Magne; F. Hu; J. Kim; J. Kautz; L. Fan	2–3:15 p.m. | Mile High Ballroom 3A–4A **Oral 6D: Large-Scale Neural Modeling** Texvent: Asynchronous Event Data Simulation via Text Prompt #20 | K. C. Cheung; S. See
Half Day: Morning | Room 708 **Women in Computer Vision (WiCV)** Sanja Fidler		2:20 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: GRAIL** Umar Iqbal	2 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: LocateAnything** Zhiding Yu	
Half Day: Afternoon | Room 501 **End-to-End 3D Learning** Marco Pavone		2:40 p.m. | NVIDIA Booth Theater #211 **Roboflow: Cosmos 3 Reason for Video Analytics** Erik Kokalj	2:20 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: PPISP** Zan Gojic	
Half Day: Afternoon | Room 102/104 **Computer Vision With Small Data (CV4Smalls)** Jose Alvarez		3 p.m. | NVIDIA Booth Theater #211 **Kitware: Operationalizing Open Models for Clinical AI** Beatriz Paniagua	2:40 p.m. | NVIDIA Booth Theater #211 **Partner Spotlight: Radiance Fields—Create Better Radiance Field Datasets: Mobile Capture for Gaussian Splatting With SplatKing** Michael Rubloff	
Half Day: Afternoon | Four Seasons 4 **From Perception to Simulation: The Emergence of World Models in Multi-modal Reasoning** Ming-Yu Liu		3:20 p.m. | NVIDIA Booth Theater #211 **Voxel 51: Teaching Data Agents to Understand Physical AI Data With FiftyOne, MCP, and NVIDIA** Adonai Vera	3:20 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: FastGen** Julius Berner	
		3:40 p.m. | NVIDIA Booth Theater #211 **Foretellix: Foretellix+ Alpamayo Reference Kit** Rohan Bhasin	3:40 p.m. | NVIDIA Booth Theater #211 **Partner Spotlight: FieldAI** FieldAI	
		4 p.m. | NVIDIA Booth Theater #211 **Partner Spotlight: Parallel Domain—The Failure Flywheel: How to Turn Real-World Disengagements Into Solved Problems at Scale** Michael DiBenigno	4 p.m. | NVIDIA Booth Theater #211 **NVIDIA Tech Brief: AgentSkills** Itai Zadok	
		4:20 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: LocateAnything** Zhiding Yu	4:20 p.m. | NVIDIA Booth Theater #211	
		4:40 p.m. | NVIDIA Booth Theater #211 **Partner Spotlight: Lightwheel—Build Photorealistic Isaac Sim Environments With Gaussian Splats** Jonathan Stephens	4:40 p.m. | NVIDIA Booth Theater #211 **Research Spotlight: VIRAL** Zhengyi Luo	

June 3–7, 2026 | Denver Convention Center

NVIDIA Experiences at CVPR 2026: **Posters**

FRIDAY 6/5		SATURDAY 6/6		SUNDAY 6/7	
10:45 a.m.–12:45 p.m. | Exhibit Hall A–F Poster Session 1 (7 papers)	**4–6 p.m. | Exhibit Hall A and F Poster Session 2**	**11:45 a.m.–1:45 p.m. | Exhibit Hall F Poster Session 3 (18 papers)**	**4:45–6:45 p.m. | Exhibit Hall A Poster Session 4 (12 papers)**	**11:45 a.m.–1:45 p.m. | Exhibit Hall F Poster Session 5 (11 papers)**	**3:30–5:30 p.m. | Exhibit Hall A Poster Session 6 (13 papers)**
Counterfactual VLA: Self-Reflective VLA Model With Adaptive Reasoning #372 | W. Ding; Y. You; W. Luo; Y. Cao; A. Sharma; B. Ivanovic; B. Li; Y. Wang; M. Pavone	**SEASON: Mitigating Temporal Hallucination in Video LLMs** #364 | K.-P. Chang; Y.-C. Wang	**GraspGen-X: Cross-Embodiment 6-DOF Diffusion-Based Grasping** #619 | Y.-W. Chao; E. Coumans; C. Eppner; S. Birchfield; A. Murali	**SAGE: Scalable Agentic 3D Scene Generation for Embodied AI** #86 | X. Li; M. Li; Q. Ma; J. Xu; M.-Y. Liu; Y. Cui; T.-Y. Lin; F. Wei	**4D-RGPT: Region-Level 4D Understanding via Perceptual Distillation** #225 | S. Liu; S. Radhakrishnan; Y.-C. Wang; M.-H. Chen	**Latent Chain-of-Thought World Modeling for End-to-End Driving** #330 | K. Chitta; Y. You; Y. Wang; W. Luo; Y. Cao; M. Pavone; B. Ivanovic
Benchmarking Single-Factor Physical Video-to-Audio Generation #179 | S. Gururani; K. Shih; S. Lee; Z. Kong; A. Goel; M.-Y. Liu	**CARI4D: Category Agnostic 4D Reconstruction of Human-Object Interaction** #635 | B. Wen; Y. Chang; H. Rabeti; J. Li; Y. Yuan; S. Birchfield	**Attend Before Attention: Scalable Video Understanding via Autoregressive Gazing** #258 | H. Ye; A. Reite; J. Kautz; B. Li; P. Molchanov; D. Yin	**DuetGen: Towards General-Purpose Interleaved Multimodal Generation** #43 | Y. Cui; F. Ferroni; M. Li; Y. Balaji; H. Wang; T.-Y. Lin; M.-Y. Liu	**Flow4DGS-SLAM: Optical Flow-Guided 4D Gaussian Splatting SLAM** #442 | Y. Wang	**Plan, Imagine, Then Act: Steering Your VLA With Efficient Visually Grounded Planning** #95 | Y. Lu
EgoControl: Controllable Egocentric Video Generation via 3D Full-Body Poses #394 | U. Iqbal	**Fast-FoundationStereo: Real-Time Zero-Shot Stereo Matching** #32 | B. Wen; S. Birchfield	**PointWorld: Scaling 3D World Models for In-The-Wild Robotic Manipulation** #609 | Y.-W. Chao; A. Mousavian; M.-Y. Liu; K. Mo	**NitroGen (also in poster)** #8 | L. Magne; F. Hu; J. Kim; J. Kautz; L. Fan	**AV-Reasoner: Clue-Grounded Audio-Visual Counting for MLLMs** #453 | Z. Li	**DeltaQuant: 4-Bit Video Diffusion Models With Spatiotemporal Delta Smoothing** #688 | Y. Lin
Enhancing Vision Language Models for 4D Perception A. Badki; H. Su; J. Jiang; S. Liu; O. Gallo	**First Frame Is the Place to Go for Video Content Customization** #193 | F. Liu	**Scaling-Aware Data Selection for End-to-End Autonomous Driving Systems** #331 | N. Chang; M. Shen; R. Mahmood; J. M. Alvarez	**VideoITG: Multimodal Video Understanding With Instructed Temporal Grounding** #299 | D.-A. Huang; Z. Li; G. Liu; J. Kautz; J. M. Alvarez; Z. Yu	**Bi-Directional Autoregressive Diffusion for Large Complex Motion Interpolation** #668 | J. Gu	**DiffusionHarmonizer: Bridging Neural Reconstruction and Photorealistic Simulation** #680 | Y. Zhang; K. Tothova; Z. Wang; K. Yin; H. Turki; R. de Lutio; O. Litany; Ž. Gojčič
Fast-ThinkAct: Efficient VLA Reasoning via Verbalizable Latent Planning #469 | C.-P. Huang; Z. Yu; M.-H. Chen; J. Kautz; Y.-C. Wang; F.-E. Yang	**Masking Teacher and Reinforcing Student for Distilling VLMs** #275 | B.-K. Lee; Y.-C. Wang	**TokenGS: Decoupling 3D Gaussian Prediction From Pixels With Learnable Tokens** #105 | J. Ren; M. Tyszkiewicz; J. Huang; Ž. Gojčič	**4DSurf: High-Fidelity Dynamic Scene Surface Reconstruction** #103 | J. M. Alvarez	**CubeComposer: Spatio-Temporal 4K 360° Video Generation From Perspective Video** #373 | J. Gu	**LEAD: Minimizing Learner-Expert Asymmetry in End-to-End Driving** #335 | M. Igl; K. Chitta
Opening the Sim-to-Real Door for Humanoid Pixel-to-Action Policy Transfer #617 | T. He; Z. Wang; F. Castañeda; L. Fan	**Scene-Centric Unsupervised Video Panoptic Segmentation** #333 | L. Leal-Taixe	**BOP-ASK: Object-Interaction Reasoning for Vision-Language Models** #233 | V. Blukis; G. Heinrich; S. Birchfield; J. Tremblay	**3D Instance Models Are Implicit Generalizable Spatial Learners** #517 | Y. Ge; Y. Sheng	**EasyV2V: High-Quality Instruction-Based Video Editing Framework** #169 | A. Mirzaei	**PhyCritic: Multimodal Critic Models for Physical AI** #66 | G. Liu; Y. Dong; J. Kautz; Z. Yu
Transition Matching Distillation for Fast Video Generation #429 | W. Nie; J. Berner; C. Liu; A. Vahdat	**Visual Sim-to-Real at Scale for Humanoid Loco-Manipulation** #582 | T. He; Z. Wang; Y. Yuan; F. Castañeda; L. Fan	**EgoEdit: Dataset, Real-Time Streaming Model, and Benchmark for Egocentric Video Editing** #168 | A. Mirzaei	**Captain Safari: A Real-Time World Engine** #365 | Y.-C. Chou (NVIDIA/JHU)	**FreeForm: Reduced-Order Deformable Simulation From Particle-Based Skinning** #359 | D. Xiang; V. Modi; G. Daviet; A. Chen; N. Sharp	**Scaling Parallel Sequence Models to Vision Foundation Models** #478 | C. McCarthy; H. Ye; J. Gu; J. Kautz; D. Yin; P. Molchanov; S. Liu
	Composite-Attribute Person Re-Identification via Pose-Guided Disentanglement #617 | C.-Y. Wang	**Event6D: Event-Based Novel Object 6D Pose Tracking** #80 | M. Kang; B. Wen	**EmbodiedSplat: Online Feed-Forward Semantic 3DGS for 3D Scene Understanding** #218 | Y. Wang	**InstructMix2Mix: Consistent Sparse-View Editing via Multi-View Personalization** #35 | O. Litany	**SpaceTools: Tool-Augmented Spatial Reasoning via Double Interactive RL** #87 | M.A. Uy; F. Ladhak; A. Murali; S. Birchfield; V. Blukis; J. Tremblay
		Frequency Switching Mechanism for Parameter-Efficient Multi-Task Learning #563 | F.-E. Yang; Y.-C. Wang	**Realiz3D: 3D Generation Made Photorealistic via Domain-Aware Learning** #526 | O. Litany	**LocateAnything3D: Vision-Language 3D Detection With Chain-of-Sight** #229 | L. Fan; J. Kautz; Z. Yu	**VGG-T³: Offline Feed-Forward 3D Reconstruction at Scale** #28 | S. Agostinho; Ž. Gojčič; L. Leal-Taixe; Q. Zhou
		Grounded 3D-Aware Spatial Vision-Language Modeling #228 | Y. Lu; P. Molchanov; V. Murali; J. Kautz; D. Yin; S. Liu	**SciEducator: Scientific Video Understanding via Deming-Cycle Multi-Agent System** #472 | X. Meng	**Mitigating Multimodal Hallucinations via Gradient-Based Self-Reflection** #365 | M. Shen; N. Chang; J. M. Alvarez	**∆ynamics: Language-Based Inference of Rigid-Body Dynamics From Videos** #575 | C.-Y. Wang
		Plenoptic Video Generation #177 | S. Tang; J. Gu; M.-Y. Liu; C.-H. Lin	**Text-Driven 3D Hand Motion Generation From Sign Language Data** #155 | M. Petrovich	**TR2M: Transferring Monocular Relative Depth to Metric Depth** #518 | Y. Huang	**Learning Convex Decomposition via Feature Fields (also in poster)** #3 | M.A. Uy; N. Sharp
		Spatial Retrieval Augmented Autonomous Driving #330 | Z. Zhang et al. (NVIDIA/Waymo/ Stanford/Tsinghua)	**Thinking in 360°: Humanoid Visual Search in the Wild** #94 | M. Pavone		**Medic-AD (also in poster)** #15 | G. Wong; K.C. Cheung
		OpenVoxel: Training-Free 3D Scene Understanding #232 | J. Choe; Y.-C. Wang; C. Sun			**Texvent (also in poster)** #20 | K.C. Cheung; S. See
		LAM: Language Articulated Object Modelers #165 | Y. Ge			
		M³KG-RAG: Multi-Hop Multimodal Knowledge Graph-Enhanced RAG #68 | W. Byeon			
		Decompose, Mix, Adapt: Parameter-Efficient NN Recombination and Compression #478 | S. Prabhumoye			
		Registration-Free Learnable Multi-View Capture of Faces #26 | G. Retsinas			
		PixelDiT: Pixel Diffusion Transformers for Image Generation #3 | W. Xiong; W. Nie; Y. Sheng; S. Liu			