KAUSTAV KUNDU

✉ kkundu10@gmail.com   🌐 https://kaustav-kundu.github.io/   ⑧ Google Scholar

RESEACH INTEREST

Machine Learning. Advancing generative AI models that excel at fine-grained spatio-temporal reasoning and instruction following while minimizing hallucinations and preserving consistency.

INDUSTRY EXPERIENCE

Staff Research Engineer, *Meta MSL* - Video Generation   June 2025 — Present
- Built large-scale video captioning pipelines and data-ablation experiments powering next-generation video-generation research, **expanding pretraining datasets from O(10M) to O(100M) clips**.
- Curated pretraining and mid-training datasets of O(100M) videos from a 1B+ candidate pool using distribution-aware sampling to capture long-tail concepts and improve T2V "interestingness".
- Led cross-stage dataset analyses and qualitative "watch parties" to connect bottom-of-funnel impact with top-of-funnel curation decisions, improving data quality and filter design.
- Bridge research and data-curation teams to embed scalable curation and evaluation frameworks directly into model training workflows.

Senior Applied Scientist, *AWS AI*   October 2019 — June 2025
- Led a cross-team research effort to train a collection of **3B to 34B multimodal models** on visual, text, and audio input modalities for **fine-grained reasoning with reduced hallucination**.
- *Tech lead (TL)* of a research team of 7 scientists which deployed 3 Moderation API launches in 2.5 years which had **>2K TPS** traffic and **seven-digit annual revenue**.
- Led the research agenda by working backwards from customer requirements to get **112x FPR reduction** and **41% FNR reduction**.
- Led the design, research and development of recognition systems such as, (a) *scalable active learning framework* which gave **67%** of the above improvements, (b) *joint image classification and detection model* in PyTorch with a non-homogeneous label set **improving AP by 58%** on existing labels.
- Led research projects that has led to **8 publications (3 orals)** in CVPR, ECCV, NeurIPS with an **acceptance record of 80%**.
- Mentored 9 full-time research scientists and 9 PhD interns.
- Research directions
  Multi-modal FMs:
     Instruction tuning, preference learning for multi-turn conversations, fine-grained reasoning with reduced hallucination.
     Using Aderversarial LLM Agents for multi-turn conversations and reducing hallucinations.
  Open set image recognition:
     Generalizaing preserving adaptation of FMs.
     Active learning with limited supervision, Self-supervised learning, Backward compatibility

Applied Scientist, *Amazon Go*   March 2018 — October 2019
- Led the research and development of the **first end-to-end computer vision based real-time action detection system** in an Amazon Go store. This enabled the stores to sell non-discrete items (such as beverages, hot foods) in the Amazon Go stores. Was deployed from 2019 to (at least) 2022 in more than 10 1P stores across North America and Europe.
- Designed and implemented a multi-task based real-time deep learning model which enabled the Amazon Go stores to remove depth sensors and **reduce CapEx by 50% without loss in accuracy**.
- Research directions
  Image level: Multi-task learning, Person detection, Semantic Segmentation
  Video-level: End-to-end real-time action detection of varying action durations from multiple RGB streams

Research Intern, *Apple (SPG)*   June 2016 — September 2016
- Designed the research direction and implemented deep structured Models to predict lane boundaries in autonomous driving scenes which improved F1 score on actual datasets by 48%.

Research Intern, *Apple (SPG)*   May 2015 — August 2015
- Implemented a CNN based deep learning network to predict the future 0.5s trajectory of each vehicle in the autonomus driving scenes. Explored use of simulation data and homography transformation to reason in the BEV space.

EDUCATION

PhD (ABD) in Computer Science, *University of Toronto*   January 2014 — December 2017
 Thesis Title: Efficient Search Strategies in 3D for Visual Scene Understanding.
 Advisors: Raquel Urtasun and Sanja Fidler

Masters in Computer Science, *Toyota Technological Institute at Chicago*   September 2012 — December 2013
 Thesis Title: Joint Semantic Segmentation and Depth Prediction in 3D Point Cloud.
 Advisor: Raquel Urtasun

BTech (Hons.) in Computer Science and Engineering, *IIIT Hyderabad*   August 2008 — May 2012
 Thesis Title: Geometry directed browser for personal photographs.
 Advisor: P J Narayanan

| Frameworks/Libraries | Python, PyTorch, TensorFlow, NumPy, scikit-learn, WebDataset, pandas, Streamlit, FiftyOne |
| AI Modeling | Multi-modal LLM, Generative AI, Real-time deep learning models |

AWARDS

- Best paper honorable mention award at **CVPR 2017**
- Outstanding reviewer at **CVPR 2018**, **CVPR 2021**

PUBLICATIONS

Under Review

- *Generalization preserving fine-tuning with embedding projections*. <u>Kaustav Kundu</u>, Ritwick Chaudhry, Jon Wu **ICLR 2026**
- *GHOST: Generating Hard negatives to overcome Object-Specific Textual hallucinations in MMLMs.* Sheng Cheng, Ritwick Chaudhry, Zhizhong Li, Rahul Duggal, Jon Wu, <u>Kaustav Kundu</u> **ICLR 2026**

Peer Reviewed

- *Robustness Preserving Fine-tuning using Neuron Importance*. Guangrui Li, Rahul Duggal, Aaditya Singh, Bing Shuai, <u>Kaustav Kundu</u>, Jon Wu. **ECCV 2024**
- *Towards Omnisupervised Instance Segmentation with Foundation Models*. Arnav Das, Ritwick Chaudhry, <u>Kaustav Kundu</u>, Davide Modolo. **CVPRW 2024**
- *Hierarchical Self-supervised Representation Learning for Movie Understanding*. Fanyi Xiao, <u>Kaustav Kundu</u>, Joseph Tighe, Davide Modolo. **CVPR 2022**
- *Id-Free Person Similarity Learning*. Bing Shuai, Xinyu Li, <u>Kaustav Kundu</u>, Joseph Tighe. **CVPR 2022**
- *What to Look at and Where: Semantic and Spatial Refined Transformer for Detecting Human-Object Interactions*. ASM Iftekhar, Hao Chen, <u>Kaustav Kundu</u>, Xinyu Li, Joseph Tighe, Davide Modolo. **CVPR 2022 (Oral)**
- *TubeR: Tubelet Transformer for Video Action Detection*. Zhao et al. **CVPR 2022 (Oral)**
- *Positive-congruent training: Towards regression-free model updates*. Sijie Yan, Yuanjun Xiong, <u>Kaustav Kundu</u>, Shuo Yang, Siqi Deng, Meng Wang, Wei Xia, Stefano Soatto. **CVPR 2021 (Oral)**
- *Exploiting weakly supervised visual patterns to learn from partial annotations*. <u>Kaustav Kundu</u>, Erhan Bas, Michael Lam, Hao Chen, Davide Modolo, Joseph Tighe. **NeurIPS 2020**
- *Pose Estimation for Objects with Rotational Symmetry*. Enric Corona, <u>Kaustav Kundu</u>, Sanja Fidler. **IROS 2018**
- *SurfConv: Bridging 3D and 2D Convolution for RGBD Images*. Hang Chu, Wei-Chiu Ma, <u>Kaustav Kundu</u>, Raquel Urtasun, Sanja Fidler. **CVPR 2018**
- *3D Object Proposals using Stereo Imagery for Accurate Object Class Detection*. Xiaozhi Chen*, <u>Kaustav Kundu</u>*, Yukun Zhu, Humin Ma, Sanja Fidler, Raquel Urtasun. **TPAMI 2017**
- *Annotating Object Instances with a Polygon-RNN*. Lluís Castrejón, <u>Kaustav Kundu</u>, Raquel Urtasun, Sanja Fidler. **CVPR 2017** (Best Paper Honorable Mention Award)
- *Exploiting Semantic Information and Deep Matching for Optical Flow*. Min Bai*, Wenjie Luo*, <u>Kaustav Kundu</u>, Raquel Urtasun. **ECCV 2016**
- *Monocular 3D Object Detection for Autonomous Driving*. Xiaozhi Chen, <u>Kaustav Kundu</u>, Ziyu Zhang, Humin Ma, Sanja Fidler, Raquel Urtasun. **CVPR 2016**
- *3D Object Proposals for Accurate Object Class Detection*. Xiaozhi Chen*, <u>Kaustav Kundu</u>*, Yukun Zhu, Andrew Berneshawi, Humin Ma, Sanja Fidler, Raquel Urtasun. **NeurIPS 2015**
- *Rent3D: Floor-Plan Priors for Monocular Layout Estimation*. Chenxi Liu*, Alexander Schwing*, <u>Kaustav Kundu</u>, Raquel Urtasun, Sanja Fidler. **CVPR 2015 (Oral)**
- *Geometry Directed Browser For Personal Photographs*. Aditya Deshpande, Siddharth Choudhary, P J Narayanan, Krishna Kumar Singh, <u>Kaustav Kundu</u>, Aditya Singh, Apurva Kumar. **ICVGIP 2012 (Oral)**

ACTIVITIES

Area Chair
CVPR 2026
Tutorial/Workshop Organizer
Cross-Model Compatibility in Computer Vision at **ICCV 2021**.
Workshop on Multimodal Content Moderation at **CVPR 2024**.
Conference/Journal Reviewer
Conferences: CVPR 2018 - 2025, ECCV 2018 - 2024, ICCV 2019 - 2023, ICLR 2021 - 2025, NeurIPS 2020 - 2024, ICML 2022 - 2025
Journals: T-PAMI 2018 - present