Kaustav Kundu

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

Reseach Interest

Machine Learning. Building multi-modal models with limited supervision, which can be used across diverse *in-domain* and *out-of-domain* scenarios and can *reason* with in an environment of agents and humans.

Industry Experience

Senior Applied Scientist, *AWS AI* October 2019 — Present

- Leading a cross-team research effort to train a collection of **3B to 34B multi-modal models** on visual, text and audio input modalities for **hallucination free fine-grained reasoning**.
- *Tech lead (TL)* of a research team of 7 scientists which deployed 3 Moderation API launches in 2.5 years which currently has **>2K TPS** traffic and **high six digit annual revenue**.
- Led the research agenda by working backwards from customers to get **112x reduction in FPR** and **41% reduction in FNR**.
- Led the design, research and development of a *scalable active learning framework* which gave **67%** of the above improvements.
- Led the design and implementation of a *joint image classifier and detection model* in PyTorch which increased the label space from **150 to 257** labels and **improved AP by 58%** of existing labels.
- Led research projects that has led to **8 publications (3 orals)** in CVPR, ECCV, NeurIPS with an **acceptance record of 80%**.
- Led the development of the first end-to-end evaluation pipeline which can measure accuracy and backward compatibility metrics of 10M images on in-domain and out-of-domain test sets across 60+ customers under 4 hours on a single A100 instance.
- Led the research team and partnered with product, engineering and data teams to improve the science-engineering hand-off process that led to the fastest launch cycle of 4 months (from the usual yearly cycle) in the AWS Rekognition org.
- Mentored 9 full-time research scientists and 7 PhD interns.
- Research directions
  Multi-modal FMs: Hallucination free reasoning for fine-grained concepts in LLMs, finetuning w/ DPO.
  Image generation: LoRA, DreamBooth based fine-tuning to generate parts of object categories.
  Open set image recognition: 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

Skills

Frameworks/Libraries	Python, PyTorch, TensorFlow, NumPy, scikit-learn, WebDataset, pandas, Streamlit, FiftyOne
AI Modeling	Generative AI, Multi-modalily, Image generation, Orchestration, 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*. Ritwick Chaudhry, Kaustav Kundu, Jon Wu **NeurIPS 2024**

Peer Reviewed

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

Activities

Tutorial/Workshop Organizer

Cross-Model Compatibility in Computer Vision at **ICCV 2021**.
Workshop on Multimodal Content Moderation at **C**VPR 2024.

Conference/Journal Reviewer

Conferences: CVPR 2018 - 2024, ECCV 2018 - 2024, ICCV 2019 - 2023, ICLR 2021 - 2024, NeurIPS 2020 - 2024, ICML 2022 - 2024
Journals: T-PAMI 2018 - present

Teaching Assistant

Inference Algorithms and Machine Learning, Intro to ML, Probabilistic Graphical Models, Neural Networks, Intro to Image Understanding, Mathematical Expression and Reasoning for Computer Science, Intro to Visual Computing