

# KAI CHEN

HKUST, Clear Water Bay, New Territories, Hong Kong SAR, P. R. China

Email: [kai.chen@connect.ust.hk](mailto:kai.chen@connect.ust.hk) ◊ Homepage: [www.cse.ust.hk/~kchenbf](http://www.cse.ust.hk/~kchenbf)

## RESEARCH OVERVIEW

---

My research aims at building reliable and generalizable AI systems from the **data-centric** perspective. Recent deep learning has witnessed superiority of the “**pre-training fine-tuning**” pipeline, empowered by training on massive amounts of datasets. Although remarkable, the intrinsic identity of fully supervised learning still poses AI systems with severe risks, especially when encountering unseen “**corner cases**” during deployment. Thus, a post-hoc “**corner case collection and fixing**” procedure is also essential to achieve ultimate reliability and trustworthiness of AI systems. Currently, I’m trying to answer the following questions,

- *Does more data always result in better performance?*
- *How to generate corner cases with controllable generative models (e.g., diffusion models and LLMs)?*
- *How to fix corner cases with minimum human intervention?*

**Research Areas:** Mixture-of-Experts, World Modeling for Autonomous Driving, (M)LLM Alignment

## EDUCATION

---

**Hong Kong University of Science and Technology**, Hong Kong SAR *Sep 2020 - Jun 2025*  
Ph.D. in **Computer Science and Engineering**  
GPA: 4.10/4.0  
Advisor: [Prof. Dit-Yan Yeung](#)

**Fudan University(FDU)**, Shanghai, China *Sep 2016 - Jun 2020*  
B.S. in **Computer Science**, Minor in **Economics** (Outstanding Graduates of Shanghai)  
Overall GPA: 3.70/4.0, Major GPA: 3.90/4.0, Ranking: 3/32  
Advisor: [Prof. Yanwei Fu](#)

**University of Manchester**, Manchester, UK *Sep 2018 - Jan 2019*  
Exchange student in the **Department of Computer Science**  
Advisor: [Dr. Tingting Mu](#)

## EXPERIENCE

---

**Mobile Intelligence Group (MIG), SenseTime** *Oct 2019 - April 2020*  
*Research Intern* *Advisor: [Dr. Wenxiu Sun](#), Sensesense*

- Research on real-time (portrait) instance segmentation deployable on mobile devices.

**Computer Vision Lab, Indiana University Bloomington (IUB)** *June 2019 - Sep 2019*  
*Global Talent Attraction Program (GTAP) Scholar* *Advisor: [Prof. David Crandall](#), IUB*

- Research on semi-supervised semantic segmentation and indoor 3D reconstruction.

## SELECTED HONORS

---

HKUST Research Travel Grant	Sep 2023
HKUST Postgraduate Scholarship	Sep 2020
Outstanding Graduates of Shanghai [ <a href="#">Wechat Push</a> ] (5%, by Shanghai Government)	April 2020
Scholarship for Outstanding Graduates (5%, by Fudan University)	April 2020
Oversea Visiting Student Stipend of (15,000 CNY, Fudan University)	Dec 2019
Joel & Ruth Spira Scholarship (1%, by <a href="#">Lutron Electronics</a> )	Mar 2019
National Scholarship (1%, by Ministry of Education of P.R.China)	Sep 2018
Scholarship for Outstanding Undergraduate Students (5%, by Fudan University)	Oct 2017

## PUBLICATIONS

---

Full publication list on my [Google Scholar](#). (\* denotes equal contribution)

### I. Mixture of Cluster-conditional Experts (MoCE)

*Q: Does more data always result in better performance during model pre-training and fine-tuning?*

- Yunhao Gou\*, Zhili Liu\*, **Kai Chen\***, Lanqing Hong, Hang Xu, Aoxue Li, Dit-Yan Yeung, James Kwok, Yu Zhang. “Mixture of Cluster-conditional LoRA Experts for Vision-language Instruction Tuning”. *Arxiv preprint, 2023*. [\[link\]](#)
- Zhili Liu\*, **Kai Chen\***, Jianhua Han, Lanqing Hong, Hang Xu, Zhenguo Li, James Kwok. “Task-customized Masked Autoencoder via Mixture of Cluster-conditional Experts”. *In International Conference on Learning Representations (ICLR spotlight), 2023*. [\[link\]](#)
- Zhili Liu, Jianhua Han, **Kai Chen**, Lanqing Hong, Hang Xu, Chunjing Xu, Zhenguo Li. “Task-Customized Self-Supervised Pre-training with Scalable Dynamic Routing”. *In AAAI Conference on Artificial Intelligence (AAAI), 2022*. [\[link\]](#)

### II. Data Flywheel for (M)LLM Alignment

*Q: Can alignment via Reinforcement Learning be replaced with SFT by training on LLM-generated data?*

- Yunhao Gou\*, **Kai Chen\***, Zhili Liu\*, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung, James Kwok, Yu Zhang. “Eyes Closed, Safety On: Protecting Multimodal LLMs via Image-to-Text Transformation”. *Arxiv preprint, 2024*. [\[link\]](#)
- **Kai Chen\***, Chunwei Wang\*, Kuo Yang, Jianhua Han, Lanqing Hong, Fei Mi, Hang Xu, Zhengying Liu, Wenyong Huang, Zhenguo Li, Dit-Yan Yeung, Lifeng Shang, Xin Jiang, Qun Liu. “Gaining Wisdom from Setbacks: Aligning Large Language Models via Mistake Analysis”. *In International Conference on Learning Representations (ICLR), 2024*. [\[link\]](#)

### III. Visual World Modeling and Perception Corner Case (CODA) Generation with Geometric-aware Diffusion Models (GeoDiffusion)

*Q: How to controllably generate corner cases for visual perception models (e.g., object detectors)?*

- Yibo Wang\*, Ruiyuan Gao\*, **Kai Chen\***, Kaiqiang Zhou, Yingjie Cai, Lanqing Hong, Zhenguo Li, Lihui Jiang, Dit-Yan Yeung, Qiang Xu, Kai Zhang. “DetDiffusion: Synergizing Generative and Perceptive Models for Enhanced Data Generation and Perception”. *In IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR), 2024*. [\[link\]](#)
- Pengxiang Li\*, **Kai Chen\***, Zhili Liu\*, Ruiyuan Gao, Lanqing Hong, Dit-Yan Yeung, Huchuan Lu, Xu Jia. “TrackDiffusion: Tracklet-Conditioned Video Generation via Diffusion Models”. *Arxiv preprint, 2023*. [\[link\]](#)
- Ruiyuan Gao\*, **Kai Chen\***, Enze Xie, Lanqing Hong, Zhenguo Li, Dit-Yan Yeung, Qiang Xu. “MagicDrive: Street View Generation with Diverse 3D Geometry Control”. *In International Conference on Learning Representations (ICLR), 2024*. [\[link\]](#)
- Zhili Liu\*, **Kai Chen\***, Yifan Zhang, Jianhua Han, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung, James Kwok. “Implicit Concept Removal of Diffusion Models”. *Arxiv preprint, 2023*. [\[link\]](#)
- **Kai Chen\***, Enze Xie\*, Zhe Chen, Yibo Wang, Lanqing Hong, Zhenguo Li, Dit-Yan Yeung. “GeoDiffusion: Text-Prompted Geometric Control for Object Detection Data Generation”. *In International Conference on Learning Representations (ICLR), 2024*. [\[link\]](#)
- Kaican Li\*, **Kai Chen\***, Haoyu Wang\*, Lanqing Hong, Chaoqiang Ye, Jianhua Han, Yukuai Chen, Wei Zhang, Chunjing Xu, Dit-Yan Yeung, Xiaodan Liang, Zhenguo, Hang Xu. “CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving”. *In European Conference on Computer Vision (ECCV), 2022*. [\[link\]](#)

## IV. Object-level Self-supervised Visual Representation Learning (SSL)

*Q: How to perform object-level SSL w/o object GT for better transfer on downstream dense perception tasks?*

- **Kai Chen\***, Zhili Liu\*, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung. “Mixed Autoencoder for Self-supervised Visual Representation Learning”. In *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023. [\[link\]](#)
- **Kai Chen**, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung. “MultiSiam: Self-supervised Multi-instance Siamese Representation Learning for Autonomous Driving”. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. [\[link\]](#)
- Jianhua Han, Xiwen Liang, Hang Xu, **Kai Chen**, Lanqing Hong, Jiageng Mao, Chaoqiang Ye, Wei Zhang, Zhenguo Li, Xiaodan Liang, Chunjing Xu. “SODA10M: A Large-Scale 2D Self/Semi-Supervised Object Detection Dataset for Autonomous Driving”. In *Datasets and Benchmarks Track, Neural Information Processing Systems (NeurIPS)*, 2021. [\[link\]](#)

### Early Works

- Md. Alimoor Reza, **Kai Chen**, Akshay Naik, David Crandall, Soon-Heung Jung. “Automatic Dense Annotation for Monocular 3D Scene Understanding”. In *IEEE Access Journal (IEEE Access)*, 2020 [\[link\]](#)
- Md Alimoor Reza, Akshay Naik, **Kai Chen**, David Crandall. “Automatic Annotation for Semantic Segmentation in Indoor Scenes”. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2019 [\[link\]](#)

## PATENTS

---

- [\[CN116665219A\]](#) **GeoDiffusion: Text-Prompted Geometric Control for Object Detection Data Generation**. Enze Xie, **Kai Chen**, Lanqing Hong, Zhenguo Li. *Published in May 26th, 2023*.
- [\[CN115731530A\]](#) **MultiSiam: Self-supervised Multi-instance Siamese Representation Learning for Autonomous Driving**. **Kai Chen**, Lanqing Hong, Hang Xu, Zhenguo Li. *Published in Aug. 24th, 2021*.

## ACADEMIC SERVICES

---

### Workshop Organizer/Program Committee

- The 2nd [SSLAD](#) workshop at ECCV 2022.
- The 1st [SSLAD](#) (Self-supervised Learning for Next-generation Industry-level Autonomous Driving) workshop at ICCV 2021.

### Conference Reviewer

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2022-2024
- IEEE International Conference on Computer Vision (ICCV) 2023
- European Conference on Computer Vision (ECCV) 2022-2024
- International Conference on Learning Representations (ICLR) 2023-2024
- Neural Information Processing Systems (NeurIPS) 2021-2023
- International Joint Conferences on Artificial Intelligence (IJCAI) 2023-2024
- AAAI Conference on Artificial Intelligence (AAAI) 2022
- International Conference on Robotics and Automation (ICRA) 2022
- Asian Conference on Computer Vision (ACCV) 2024

### Journal Reviewer

- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)
- IEEE Transactions on Image Processing (TIP)
- IEEE Access

## TEACHING

---

- **HKUST COMP 2012** - Object-Oriented Programming and Data Structures, Teaching Assistant, Fall 2021.
- **HKUST COMP 2012** - Object-Oriented Programming and Data Structures, Teaching Assistant, Spring 2021.

## INVITED TALKS

---

- [AI TIME Online] Gaining Wisdom from Setbacks: Aligning Large Language Models via Mistake Analysis. [\[Recording\]](#)
- [TechBeat Online] Gaining Wisdom from Setbacks: Aligning Large Language Models via Mistake Analysis. [\[Recording\]](#)
- [VALSE 2023@Wuxi] Mixed Autoencoder for Self-supervised Visual Representation Learning. [\[Recording\]](#)
- [VALSE 2023@Wuxi] CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving. [\[Recording\]](#)

## TECHNICAL SKILLS

---

<b>Program Languages</b>	Python, Matlab, C/C++/C#, SQL, $\text{\LaTeX}$
<b>Framework</b>	Pytorch, Tensorflow
<b>Language</b>	Native in Mandarin Chinese, Fluent in English and Japanese CET-4(649), CET-6(619), TOEFL-iBT(101)