

# Shichao Li

✉ nicholas.li@connect.ust.hk

## Education

**The Hong Kong University of Science and Technology (HKUST)**

*Ph.D. in Computer Science and Engineering*

Program Advisor: Professor Kwang-Ting Tim Cheng | GPA - 3.9/4.3

Research Focus - computer vision, machine learning, 3D perception

**Zhejiang University (ZJU)**

*Bachelor of Engineering, Chu Kochen (CKC) Honors College*

Major: Information Engineering | GPA - 3.9/4.0 (91.3/100)

**Kowloon, Hong Kong**

*Sep 2017–Sep 2022*

**Hangzhou, Zhejiang**

*Sep 2013–Jul 2017*

## Experience

### Work Experience

**Department of AI models, DJI**

*Senior Computer Vision Engineer*

**Shenzhen**

*Dec 2022–*

- As a senior engineer focusing on deep learning and perception, I am responsible for the research, design, and training of cutting-edge AI models and delivering them in products such as a vehicle-side autonomous perception system and a cloud-based automatic labeling system. Key duties include improving system performance and maintaining technological competence. My technical experience includes, but is not limited to,
  - 2D/3D object detection from monocular/stereo images, point cloud, and event cameras
  - object pose, keypoint, and depth estimation
  - video object detection and time-to-contact estimation
- As a project leader, I am responsible for communicating and coordinating bandwidth requirements with upstream and downstream teams, allocating project team tasks, organizing meetings, tracking progress, and managing risks.
- As a mentor, I guide new employees and junior engineers in designing and training deep learning models, and mentor interns in pioneering research on multimodal deep perception models.

### Project Experience

**Long-range video object perception, DJI**

*Project Lead*

**Shenzhen**

*Sep 2024–*

- Situation: The popular BEV architectures for perception adopt voxel-based 3D scene representation, which consumes enormous memory and computation, making them difficult to apply to long-range perception scenarios. This project aims to achieve effective long-range object perception with deep-learning-based 2D video object detection from monocular or stereo inputs.
- Actions: (1) Invented a new system unique in the industry for long-range visual object perception with stereo cameras, and transferred it into products. (2) Designed and trained a new end-to-end multi-task video object detection model that performs effective stereo matching, object detection, object depth estimation, and object time-to-contact (TTC) estimation. (3) Proposed a new semi-supervised multi-task training scheme that effectively utilizes both labeled and unlabeled data from different sensor modalities to improve model generalization. (4) Lead and guide junior project team members in model design and training, and organize weekly meetings for task assignment and progress review.
- Results: (1) Compared to the previous classical architecture of cascading multiple individual modules, this end-to-end model performs multi-task perception directly from video inputs, reducing intermediate information loss. (2) The perception range and depth estimation performance of this model outperforms the previous system, and has been accepted for deployment in a new generation of products. (3) The novel semi-supervised training scheme contributes to the capability of utilizing massive unlabeled data for enhanced system performance, which was recommended for a patent application. (4) Successfully guide team members for model training and adding more product features.

## Offboard 4D autolabeling for dynamic objects, DJI

Shenzhen

Project Owner

Jul 2023–Sep 2024

- Situation: End-to-end models require a massive amount of 3D labels that are spatially accurate and temporally coherent for training. Manual labeling is time-consuming and demands an automatic approach.
- Actions: (1) Designed and implemented a 3D object detection and tracking model from camera and point cloud inputs based on the BEVFusion framework. (2) Designed and implemented an offboard auto-labeling pipeline consisting of object detection, tracking, state estimation, and instance-level refinement. (3) Responsible for communicating and coordinating the needs of data collection, mining, and auto-labeling.
- Results: (1) Compared to the previous point-cloud-based solution, the proposed multi-modal system achieves improved auto-labeling quality and leads to a publication and a patent application. (2) The system was widely used in the labeling process of massive training datasets of multiple models used in the products, producing, in total, over hundreds of millions of frames of labels.

## Learning-based semi-automatic object labeling from images, DJI

Shenzhen

Senior Engineer

Dec 2022–Jun 2023

- Situation: Clients' vehicles produce an enormous amount of realistic driving data that could be used as training data to improve model generalization; however, this data requires labeling, and manual labeling is time-consuming.
- Actions: Designed and implemented a high-resolution deep neural network that performs object keypoint and depth estimation given 2D region-of-interest inputs from annotators. 3D object trajectories are obtained from such model predictions and state estimation based on kinematic priors.
- Results: (1) Model predictions are used by annotators to improve the labeling efficiency by 20%. (2) 3D object trajectories produced by the semi-automatic tool are used for training deep models in products.

## Internships

### SenseTime Autonomous Driving Group

Shanghai

Research Intern

Jan–Apr 2020

- Conduct 3D object detection research for autonomous visual perception

## Research Experience

### Vision and System Design Lab, HKUST

Hong Kong

Oct 2017–Jul 2022

- Stereo 3D object perception (2021–2022)
  - Proposed a new instance-aware multi-resolutional approach for stereo 3D object detection with enhanced performance, flexibility, and transferability.
  - Proposed the first architecture for joint stereo 3D object detection and implicit shape estimation.
- Monocular 3D Object Detection (2020–2021)
  - Proposed a novel approach for accurate RGB-based vehicle pose estimation through learning geometry-aware representation.
  - Proposed the first architecture for joint stereo 3D object detection and implicit shape estimation.
- Monocular 3D Human Pose Estimation (2019–2020)
  - Improved generalization ability of 2D-to-3D deep neural networks using synthetic data.
  - Proposed a novel evolution framework for data augmentation and a high-resolution model architecture, both contributed to achieving state-of-the-art performance.
- HKUST-Naver Collaborate Research Project on Mobile Computer Vision (2017-2019)
  - Developed a facial image analysis system on Windows and Android platforms consisting of face detection, 2D/3D facial landmark localization, head pose estimation, and graphics rendering modules.
  - Conduct facial image analysis using traditional local binary features and random forests, as well as deep learning.

## Selected Publications

**Shichao Li**, Peiliang Li, Qing Lian, Peng Yun and Xiaozhi Chen. Learning better representations for crowded pedestrians in offboard LiDAR-camera 3D tracking-by-detection, *IEEE International Conference on Robotics and Automation (ICRA)* [Github] [Link]

**Jinghang Li**, **Shichao Li\***, Qing Lian, Peiliang Li, Xiaozhi Chen, and Yi Zhou. Toward deep representation learning for event-enhanced visual autonomous perception: the eAP dataset, submitted to *IEEE*

*Transactions on Robotics* (Under review, \*equal contribution, project mentor) [Github] [Link]

**Shichao Li**, Zechun Liu, Zhiqiang Shen and Kwang-Ting Cheng. Stereo Neural Vernier Caliper, *Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)* [Github] [Link]

**Shichao Li**, Zengqiang Yan, Hongyang Li and Kwang-Ting Cheng. Exploring intermediate representation for monocular vehicle pose estimation, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Github] [Link]

**Shichao Li**, Lei Ke, Kevin Pratama, Yu-Wing Tai, Chi-Keung Tang and Kwang-Ting Cheng. Cascaded deep monocular 3D human pose estimation with evolutionary training data, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR, Oral Presentation)* [Github] [Link]

Lei Ke, **Shichao Li**, Yanan Sun, Yu-Wing Tai and Chi-Keung Tang. GSNet: Joint vehicle pose and shape reconstruction with geometrical and scene-aware supervision, *European Conference on Computer Vision (ECCV)* [Github] [Link]

**Shichao Li**, Xijie Huang, Zechun Liu, and Kwang-Ting Cheng. Joint stereo 3D object detection and implicit surface reconstruction, *Scientific Reports* [Github] [Link]

**Shichao Li**, Fei Sun, Di An and Sailing He. Increasing efficiency of a wireless energy transfer system by spatial translational transformation, *IEEE Transactions on Power Electronics* [Link]

**Shichao Li**, Wenchao Chen, Yandong Luo, Jun Hu, Pingqi Gao, Jichun Ye, Kai Kang, Hongsheng Chen, Erping Li and Wen-Yan Yin. Fully coupled multiphysics simulation of crosstalk effect in bipolar resistive random access memory, *IEEE Transactions on Electron Devices* [Link]

Zechun Liu, Zhiqiang Shen, **Shichao Li**, Koen Helwegen, Dong Huang and Kwang-Ting Cheng. How do Adam and training strategies help BNNs optimization, *International Conference on Machine Learning (ICML)* [Github] [Link]

Xijie Huang, Zhiqiang Shen, **Shichao Li**, Zechun Liu, Xianghong Hu, Jeffrey Wicaksana, Eric Xing and Kwang-Ting Cheng. SDQ: Stochastic Differentiable Quantization with Mixed Precision, *International Conference on Machine Learning (ICML)*

## Selected Awards

---

HKUST Research Travel Grant (2021)

SENG (School of Engineering) Academic Award for Continuing PhD Students (2020)

Hong Kong PhD Fellowship (2017–2021)

ZJU Outstanding Undergraduate Thesis Award (2017)

China National Scholarship (2015)

## Skills

---

- Programming: Python, C++
- Deep learning frameworks: PyTorch, MM Detection

## Languages

---

- Chinese
- English: TOFEL iBT 110/120, GRE Verbal 160/170