Cheng Perng Phoo

Researcher and Engineer — Multimodal Perception

chengperng@gmail.com | https://cpphoo.github.io/

Research Summary

Researcher and engineer specializing in multimodal perception and foundation models for real-world applications. Focuses on learning from limited supervision and unlabeled data across LiDAR, vision, language, and temporal modalities. Has led research efforts resulting in publications at NeurIPS, ICLR, CVPR, ICCV, and ICRA, with experience translating research ideas into large-scale industrial systems at Waymo.

Research Themes

- *Multimodal perception (LiDAR, vision, language, video)*
- *Learning from limited and weak supervision, including unlabeled and synthetic data*
- *Foundation models for autonomous driving and remote sensing*
- *Robustness to long-tailed and rare scenarios*

Research Experience

Software Engineer
Mountain View, CA

Waymo LLC
April 2025 - Current

- Leveraged state-of-the-art multimodal LLMs for perception in long-tailed, high-impact driving scenarios.
- Led research on LiDAR-based encoders for multimodal foundation models, improving generalization for road understanding.
- Designed camera–LiDAR sensor fusion architectures within multimodal LLM frameworks for autonomous driving.
- Collaborated cross-functionally to translate research prototypes into scalable training and evaluation pipelines.

Postdoctoral Research Scientist
Santa Clara, CA

Apple Inc. (via Harvey Nash)
June 2024 - March 2025

- Led research on multimodal video large language models operating on long-form video data.
- Built and maintained large-scale training infrastructure to support multimodal LLM research at billion-parameter scale.
- Developed a synthetic annotation pipeline to generate textual supervision for video-language learning.
- Conducted systematic evaluations of video LLM capabilities and failure modes, informing model design and benchmarking.

Graduate Research Assistant
Ithaca, NY

Cornell University Department of Computer Science
Aug 2017 - May 2024

- Advisor: Professor Bharath Hariharan
- Led research on learning perception models beyond internet applications, spanning autonomous driving and remote sensing.
- Developed methods for learning from unlabeled LiDAR and visual data, including adaptation from repeated traversals, mobile object discovery without supervision.
- Advanced data- and compute-efficient learning in novel domains by leveraging large-scale pre-trained foundation models such as CLIP and diffusion models.
- Proposed cross-domain and cross-sensor transfer techniques enabling robust perception in novel environments.

Research Internships

Research Intern @ FAIR Accel
Menlo Park, CA

Meta Fundamental AI Research (FAIR)
Jun 2022 - Aug 2022

- Conducted research on object state change modeling in egocentric video, contributing to Ego4D-related research.
- Proposed a novel state change embedding that could capture different degrees of state changes for an object.

Research Intern
Remote

MIT-IBM Watson AI Lab
Jun 2021 - Dec 2021

- Researched open-set semi-supervised learning and transfer learning for out-of-distribution data.
- Investigated low-level features and dynamic neural networks for open-set semi-supervised classification.

Selected Publications

[NEURIPS 2025] MONITRS: Multimodal Observations of Natural Incidents Through Remote Sensing
[ICLR 2024] Remote Sensing Vision-Language Foundation Models without Annotations via Ground Remote Alignment
[ICRA 2024] Better Monocular 3D Detectors with LiDAR from the Past
[ICCV 2023] Distilling from Similar Tasks for Transfer Learning on a Budget
[NeurIPS 2023] Emergent Correspondence from Image Diffusion
[NeurIPS 2022] Unsupervised Adaptation from Repeated Traversals for Autonomous Driving
[CVPR 2022] Learning to Detect Mobile Objects from LiDAR Scans Without Labels
[ICLR 2021] Self-training for Few-shot Transfer across Extreme Task Differences

Education

Cornell University, USA *Aug 2017 - May 2024*
Ph.D. in Computer Science
Advisor: Bharath Hariharan
Thesis: Toward Perception Models Beyond Internet Applications

University of Michigan, Ann Arbor, USA *Sep 2014 - May 2017*
B.S. in Computer Science and Pure Mathematics, GPA 3.78/4.00

Skills

Languages: Mandarin (Native), English (Fluent), Malay (Fluent)
Programming: Python (PyTorch, Hugging Face, PyTorch Lightning, vLLM); experience with large-scale model training and evaluation