

# Cheng Perng Phoo

POSTDOCTORAL RESEARCH SCIENTIST, APPLE INC.

+1(929)317-9729 | [chengperng@gmail.com](mailto:chengperng@gmail.com) | <https://cpphoo.github.io/>

## About Me

---

I am a postdoctoral research scientist at Apple Inc. My research lies at the intersection of computer vision and machine learning. Specifically, I focus on building perception systems capable of recognizing a wide range of concepts across various problem domains, such as remote sensing, medical imagery, and self-driving vehicles. Toward this goal, I have identified three key challenges: label efficiency, robust deployment, and multitasking. Currently, I am working on developing multimodal large language models capable of processing long videos and tackling multiple tasks simultaneously.

## Research Experiences

---

### Postdoctoral Research Scientist

APPLE INC. / HARVEY NASH

Sunnyvale, CA

June 2024 - Present

- Maintained infrastructure for training multimodal video large-language models (LLMs) with billions of parameters.
- Implemented a synthetic annotation pipeline for generating textual annotations for any videos for training multimodal LLMs.
- Explored strategies for assessing multimodal video LLMs, providing better understanding of their capabilities and limitations.

### Graduate Research Assistant

CORNELL GRAPHICS AND VISION GROUP

Cornell University

Aug 2018 - May 2024

- Advisor: Professor Bharath Hariharan
- Adapted visual-language models trained on internet imagery (e.g. CLIP) to remote sensing imagery without textual annotation [ICLR 2024].
- Leveraged LiDAR scans from different past traversals to improve camera-based 3D object detectors [ICRA 2024].
- Leveraged multiple large-scale pre-trained foundation models to create efficient models that could be trained with fewer annotated data and resources [ICLR 2021 (oral), ICCV 2023].
- Utilized features from off-the-shelf image diffusion models to establish correspondence between image pairs without training [NeurIPS 2023].
- Utilized unlabeled LiDAR scans from repeated traversals to adapt 3D object detectors to novel domains [NeurIPS 2022].
- Devised an approach to discover mobile objects from unlabeled LiDAR scans without any human annotations [CVPR 2022].

## Education

---

### Cornell University, USA

Ph.D. in Computer Science

Advisor: Bharath Hariharan

Thesis: Toward Perception Models Beyond Internet Applications

Aug 2017 - May 2024

## Selected Publications

---

[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

## Internships

---

- Meta Fundamental AI Research (FAIR) [Jun 2022 - Aug 2022]
- MIT-IBM Watson AI Lab [Jun 2021 - Dec 2021]

## Skills

---

**Languages:** Mandarin (Native), English (Fluent), Malay (Fluent)

**Coding:** Python (PyTorch, vLLM, Hugging Face, PyTorch Lightning, NumPy, pandas, scikit-learn)