# Cheng Perng Phoo

Research Scientist — Multimodal Perception

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


## Research Summary


*Research scientist 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.

## 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

## Peer-Reviewed Publications

(* Equal Contributions)

### MONITRS: Multimodal Observations of Natural Incidents Through Remote Sensing

- Shreelekha Revankar, Utkarsh Mall, **Cheng Perng Phoo**, Kavita Bala, and Bharath Hariharan.
- Conference on Neural Information Processing Systems (NEURIPS), 2025.
- Summary: We construct a large-scale multimodal dataset for natural incidents using remote sensing data and news articles.

### Towards LLM Agents for Earth Observation

- Chia Hsiang Kao, Wenting Zhao, Shreelekha Revankar, Samuel Speas, Snehal Bhagat, Rajeev Datta, **Cheng Perng Phoo**, Utkarsh Mall, Carl Vondrick, Kavita Bala, Bharath Hariharan
- TerraBytes: Towards global datasets and models for Earth Observation — a workshop at International Conference on Machine Learning (ICML), 2025.
- Summary: We introduce a new benchmark to evaluate the capabilities of LLM agents in performing Earth Observation tasks.

### DiSciPLE: Learning Interpretable Programs for Scientific Visual Discovery

- Utkarsh Mall, **Cheng Perng Phoo**, Mia Chiquier, Bharath Hariharan, Kavita Bala, Carl Vondrick
- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
- Summary: A neurosymbolic framework to learn programs explaining visual observations in visual spatial scientific domains.

### Scale-aware Recognition in Satellite Images under Resource Constraints

- Shreelekha Revankar, **Cheng Perng Phoo**, Utkarsh Mall, Bharath Hariharan, Kavita Bala
- International Conference on Learning Representations (ICLR), 2025
- Summary: We introduce a new approach to scale-aware recognition in satellite imagery under resource constraints.

### Learning 3D Perception from Others' Predictions

- Jinsu Yoo, Zhenyang Feng, Tai-Yu Pan, Yihong Sun, **Cheng Perng Phoo**, Xiangyu Chen, Mark Campbell, Kilian Q. Weinberger, Bharath Hariharan, and Wei-Lun Chao.
- International Conference on Learning Representations (ICLR), 2025
- Summary: We investigated how to create a 3D detector by leveraging predictions from other vehicles.

### AllClear: A Comprehensive Dataset and Benchmark for Cloud Removal in Satellite Imagery

- Hangyu Zhou*, Chia Hsiang Kao*, **Cheng Perng Phoo**, Utkarsh Mall, Bharath Hariharan, Kavita Bala
- Conference on Neural Information Processing Systems (NEURIPS), 2024.
- Summary: The largest dataset to investigate cloud removal leveraging temporal and multispectral information.

### Better Monocular 3D Detectors with LiDAR from the Past

- Yurong You*, **Cheng Perng Phoo***, Carlos Andres Diaz-Ruiz, Katie Luo, Wei-Lun Chao, Mark Campbell, Bharath Hariharan, Kilian Q. Weinberger.
- International Conference on Robotics and Automation (ICRA), 2024.
- Summary: Unlabeled LiDAR scans from repeated traversals could be used to improve camera-based 3D object detectors.

### Remote Sensing Vision-Language Foundation Models without Annotations via Ground Remote Alignment

- Utkarsh Mall*, **Cheng Perng Phoo***, Meilin Kelsey Liu, Carl Vondrick, Bharath Hariharan, Kavita Bala
- International Conference on Learning Representations (ICLR), 2024.
- Summary: We use ground images as intermediary to connect satellite imagery to natural language (encoded using CLIP), yielding VLMs without textual annoations.

### Pre-training LiDAR-based 3D Object Detectors through Colorization

- Tai-Yu Pan, Chenyang Ma, Tianle Chen, **Cheng Perng Phoo**, Katie Z Luo, Yurong You, Mark Campbell, Kilian Q Weinberger, Bharath Hariharan, Wei-Lun Chao
- International Conference on Learning Representations (ICLR), 2024.
- Summary: We pre-train a point cloud detector by tasking it to fill in the missing colors within the point cloud.

### Reward Finetuning for Faster and More Accurate Unsupervised Object Discovery

- Katie Z Luo*, Zhenzhen Liu*, Xiangyu Chen*, Yurong You, Sagie Benaim, **Cheng Perng Phoo**, Mark Campbell, Wen Sun, Bharath Hariharan, Kilian Q. Weinberger
- Conference on Neural Information Processing Systems (NEURIPS), 2023.
- Summary: We reframe object discovery as an RL problem and design a reward function to enable faster and more accurate discovery of objects in driving scenes without human supervision.

### Emergent Correspondence from Image Diffusion

- Luming Tang*, Menglin Jia*, Qianqian Wang*, **Cheng Perng Phoo**, Bharath Hariharan
- Conference on Neural Information Processing Systems (NEURIPS), 2023.
- Summary: Features from off-the-shelf image diffusion models could be used to identify semantic and geometric correspondence without further training.

### Distilling from Similar Tasks for Transfer Learning on a Budget

- Kenneth Borup, **Cheng Perng Phoo**, Bharath Hariharan.
- IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- Summary: We construct label- and compute-efficient models by identifying and distilling from suitable pre-trained models.

### Unsupervised Domain Adaptation for Self-Driving from Past Traversal Features

- Zhang, Travis, Katie Luo, **Cheng Perng Phoo**, Yurong You, Wei-Lun Chao, Bharath Hariharan, Mark Campbell, and Kilian Q. Weinberger.
- BRAVO: roBustness and Reliability of Autonomous Vehicles in the Open-world — a workshop at IEEE/CVF International Conference on Computer Vision Workshops (ICCV), 2023.
- Summary: We used unlabeled LiDAR scans from repeated traversals to adapt a specialized 3D object detector for self-driving.

### Unsupervised Adaptation from Repeated Traversals for Autonomous Driving

- Yurong You*, **Cheng Perng Phoo***, Katie Luo*, Travis Zhang, Wei-Lun Chao, Bharath Hariharan, Mark Campbell, Kilian Q. Weinberger.
- Conference on Neural Information Processing Systems (NEURIPS), 2022.
- Summary: Unlabeled LiDAR scans from repeated traversals could be used to disambiguate foreground and background objects, yielding cleaner signals for self-training adaptation.

### Learning to Detect Mobile Objects from LiDAR Scans Without Labels

- Yurong You*, Katie Luo*, **Cheng Perng Phoo**, Wei-Lun Chao, Wen Sun, Bharath Hariharan, Mark Campbell, Kilian Q. Weinberger.
- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- Summary: Comparing unlabeled LiDAR scans from multiple traversals on the same location could uncover dynamic LiDAR points that could be used to train a mobile object detector in an unsupervised/self-supervised manner.

### Task2Sim: Towards Effective Pre-training and Transfer from Synthetic Data

- Samarth Mishra, Rameswar Panda, **Cheng Perng Phoo**, Chun-Fu Richard Chen, Leonid Karlinsky, Kate Saenko, Venkatesh Saligrama, Rogerio S. Feris.
- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- Summary: Different downstream tasks require different representations pre-trained on synthetic data generated using different configurations (lightings, object poses, etc). We use reinforcement learning to learn a policy that maps a compact task representation to the appropriate synthetic data configuration.

### Coarsely-labeled Data for Better Few-shot Transfer

- **Cheng Perng Phoo**, Bharath Hariharan.
- IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- Summary: Coarsely-labeled data can be cheap to acquire and can be used to learn a better representation for few-shot learning.

### Self-training for Few-shot Transfer across Extreme Task Differences

- **Cheng Perng Phoo**, Bharath Hariharan.
- International Conference on Learning Representations (ICLR), 2021. Oral Presentation. [53 / 2997 submissions]
- Summary: We can build strong neural representation for novel domains by (self-)training students to replicate pseudo-labels produced by a teacher from another, unrelated problem domain.

### Predicting risk of sport-related concussion in collegiate athletes and military cadets: a machine learning approach using baseline data from the CARE Consortium Study

- Joel Castellanos*, **Cheng Perng Phoo***, James T. Eckner, Lea Franco, Steven P. Broglio, Mike McCrea, Thomas McAllister, and Jenna Wiens.
- Sports medicine (2020): 1-13.
- Summary: Baseline tests conducted on college athletes and military cadets before each semester could contain information for identifying athletes/military cadets who are at a higher risk of experiencing a concussion.

### Heart Sound Classification based on Temporal Alignment Techniques

- José Javier González Ortiz, **Cheng Perng Phoo**, Jenna Wiens.
- Computing in Cardiology Conference (CinC), 2016.
- Summary: We use temporal alignment techniques such as dynamic time warping to extract features from heart sound recordings for identifying patients at risk of adverse cardiovascular outcomes.

## Academic Services

### Conference Reviewer

- Conference on Neural Information Processing Systems (NEURIPS) 2023, 2024
- Computer Vision and Pattern Recognition (CVPR) 2022, 2023, 2024, 2025
- European Conference on Computer Vision (ECCV) 2022, 2024
- International Conference on Computer Vision (ICCV) 2023
- International Conference on Learning Represenations (ICLR) 2024, 2025
- International Conference on Machine Learning (ICML) 2024

### Ph.D. Application Reviewer

- Computer Science, Cornell Univeristy 2023

## Teaching Experiences

**CS4780/5780: Machine Learning for Intelligent Systems** *Cornell Univerisity*
*Teaching Assistant for Kilian Weinberger, Chris De Sa* *Spring 2018*
Awarded **Outstanding TA.**

**CS4786/5786: Machine Learning for Data Science** *Cornell Univerisity*
*Teaching Assistant for Karthik Sridharan* *Fall 2017*

**EECS445: Introduction to Machine Learning** *University of Michigan, Ann Arbor*
*Instructional Aide for Jenna Wiens* *Winter 2017*

**EECS203: Discrete Mathematics** *University of Michigan, Ann Arbor*
*Instructional Aide* *Fall 2015, Fall 2016, Winter 2016*

## Skills

**Programming Languages**: Python, MATLAB, C/C++
**Machine Learning**: PyTorch, Tensorflow, scikit-learn, NumPy, SciPy, Pandas
**Others**: Bash, LaTeX, Linux

## Languages

**Mandarin**: Native Language
**English, Malay**: Fluent (speaking, reading, writing)