

Cheng-Yen (Wesley) Hsieh

chengyeh@alumni.cmu.edu | (412) 996-5387 | [in chengyen-hsieh](https://github.com/chengyen-hsieh) | [WesleyHsieh0806](https://www.wesleyhsieh0806.com) | [Personal Website](#)

EDUCATION

Carnegie Mellon University (CMU), School of Computer Science
Master of Science in Computer Vision (MSCV)

Pittsburgh, PA
Dec. 2023

WORK EXPERIENCE

ByteDance

Research Scientist

San Jose CA

Mar. 2024 — Present

- Developed a transformer-based video generation model ensuring enhanced 3D consistency by utilizing camera trajectories estimated through structure-from-motion algorithms; designed 2D camera Plücker embeddings and epipolar attention modules to refine target camera pose conditioning.

Waymo

Software Engineer, Machine Learning Infrastructure, Intern.

Mountain View, CA

May. 2023 — Aug. 2023

- Proposed a differential learning rate (DLR) API that automatically divides model groups based on variable names, specified in the configuration files like proto, textproto, and dataclass.
- Created a modularized DLR training pipeline compatible with mainstream deep learning frameworks such as Tensorflow; Achieved a greater than 1.5% accuracy improvement across multiple tasks within the perception team.

ASUS Intelligent Cloud Services (AICS)

Software/Machine Learning Engineer Intern, Software Research Development Dept.

Taipei, Taiwan

Jul. 2021 — Sept. 2021

- Operated Spark database to process and analyze more than 1,000,000 advertisements (Ads) and customer data logs, building an advertisement recommendation system to predict user browsing preferences with decision trees.
- Added over 300 features to solve model underfitting by organizing narrative categorization into multi-levels, improving the system's CTR (Click Through Rate) to 110% compared to previous performance.

RESEARCH PROJECTS

Tracking Any Object Amodally ([Project Page](#))

CMU Graduate Researcher | MSCV Capstone Project

Pittsburgh, PA

Jan. 2023 — Present

- Developed TAO-Amodal dataset to enhance amodal tracking, comprehending complete object structures from partial visibility, by including 17k objects spanning 880 categories and adapted evaluation metrics.
- Presented a light-weight plug-in module, amodal expander, to amodalize any existing tracker with limited training data; Introduced PasteNOclude data augmentation technique to craft occlusion scenarios.
- Achieved significant 3.3% mAP and 1.6% mAP improvements on the detection and tracking of occluded objects, along with a >30% detection mAP boost for occluded people on the TAO-Amodal validation set.

Visual Question Answering with Vision Language Multi-Agent Debate

MIT Research Assistant (Remote Collaboration)

Pittsburgh, PA

Jul. 2023 — Present

- Integrated vision language model agents into a collaborative debate setting to improve language responses through the synthesis of reasoning processes from multiple agents;
- Enhanced > 2% accuracy across diverse tasks such as instance counting, object attributes, and relation analysis.

Self-Supervised Representation Learning for Multi-Label Visual Tasks ([Paper](#), [Github](#))

NTU Undergraduate Researcher

Taipei, Taiwan

Aug. 2020 — Feb. 2022

- Designed an unsupervised visual pretext task for downstream multi-label classification tasks.
- Utilized multi-scale pyramid and patch-level correlation learning to leverage information across distinct scales.
- Outperformed the strongest and SOTA methods by 3.3% mAP and 2.5% mAP on COCO and VOC, respectively.

Federated Learning for Efficient Image Recognition ([Paper](#), [Github](#))

NTU Undergraduate Researcher

Taipei, Taiwan

Jan. 2020 — Jun. 2021

- Designed computation-efficient FL-hyperdimensional computing (FL-HDC) to address high computation and communication burdens from previously training deep neural networks (DNN)s.
- Reduced communication costs by 23 times using bipolarization on model weights to cut communication costs, compensating for the precision loss of numerical values by proposing an adaptive retraining mechanism.

SELECTED PUBLICATIONS

- [1] **Cheng-Yen Hsieh**, Tarasha Khurana, Achal Dave, Deva Ramanan, "Tracking Any Object Amodally", submitted to the *Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024)*.
- [2] **Cheng-Yen Hsieh**, Chih-Jung Chang, Fu-En Yang, Yu-Chiang Frank Wang, "Self-Supervised Pyramid Representation Learning for Multi-Label Visual Analysis and Beyond", in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.
- [3] **Cheng-Yen Hsieh**, Yu-Chuan Chuang and An-Yeu Wu, "FL-HDC: Hyperdimensional Computing Design for the Application of Federated Learning," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2021, pp. 1-5, doi: 10.1109/AICAS51828.2021.9458526.