

Cheng-Yen (Wesley) Hsieh

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

EDUCATION

Carnegie Mellon University (CMU), School of Computer Science

Pittsburgh, PA

Master of Science in Computer Vision (MSCV)

Dec. 2023

National Taiwan University (NTU)

Taipei, Taiwan

Bachelor of Science in Electrical Engineering

Feb. 2022

WORK EXPERIENCE

ByteDance

San Jose CA

Research Scientist, AI-Lab-Research

Mar. 2024 — Present

Multimodal Diffusion Protein Language Model – AI for Science ([DPLM Home Page](#), [DPLM-2.1](#))

- Developed a multimodal protein foundation model that jointly generates protein sequences and 3D structures using a transformer-based discrete diffusion language model.
- Achieve a finer-grained supervision through bitwise discrete modeling and a hybrid approach for data-space modeling, significantly improving structural generation over conventional index-based tokens; Introduce representation alignment and geometry-aware architectural designs that effectively refines generation diversity.
- Explored the correlation between multimer and monomer data and leveraged multimer data to advance the structural modeling for both single and multi-chain proteins.
- Improve the folding RMSD from 5.52 to 2.36 on the PDB data dataset, outperforming 3B specialized folding baselines using a multimodal PLM with only 650M parameters.

3D-Consistent Diffusion Transformer for Video Generation

- 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

Mountain View, CA

Software Engineer, Machine Learning Infrastructure, Intern.

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)

Taipei, Taiwan

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

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](#))

Pittsburgh, PA

CMU Graduate Researcher | MSCV Capstone Project

Jan. 2023 — Mar. 2024

- 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 PasteNOcclude 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

Pittsburgh, PA

MIT Research Assistant (Remote Collaboration)

Jul. 2023 — Mar. 2024

- 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](#))

Taipei,

Taiwan

NTU Undergraduate Researcher

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.

Split Learning for Efficient Image Recognition ([Paper](#), [GitHub](#))

NTU Undergraduate Researcher

Taipei, Taiwan
Jun. 2021 — Feb. 2022

- Proposed circular convolution-based batch-wise compression for split learning to compress multiple features into one by exploiting the quasi-orthogonality of high-dimensional features with circular convolution and superposition.
- Reduced 16x communication costs with negligible accuracy drops on CIFAR10 and CIFAR100, while improving 1152x memory and 2.25x computation overhead compared to the SOTA dimension-wise compression method.

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.

PUBLICATIONS

- [1] **Cheng-Yen Hsieh**, Xinyou Wang, Daiheng Zhang, Dongyu Xue, Fei YE, Shujian Huang, Zaixiang Zheng, Quanquan Gu, "Elucidating the Design Space of Multimodal Protein Language Models", *ICML 2025 (Spotlight, Top 2.6% of all submissions)*.
- [2] **Cheng-Yen Hsieh**, Kaihua Chen, Achal Dave, Tarasha Khurana, Deva Ramanan, "TAO-Amodal: A Benchmark for Tracking Any Object Amodally", *arXiv preprint arXiv:2312.12433*, 2023.
- [3] **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.
- [4] **Cheng-Yen Hsieh**, Yu-Chuan Chuang, An-Yeu Wu, "C3-SL: Circular Convolution-Based Batch-Wise Compression for Communication-Efficient Split Learning", in *2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, 2022.
- [5] **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.