Zhi-Yi Chin

+886 972752710 | joycenerd.cs09@nycu.edu.tw | joycenerd.github.io | joycenerd | zhi-yi-chin-b7927645 | Zhi-Yi Chin

Education

National Yang Ming Chiao Tung University (NYCU) *Feb. 2021 - Sep. 2023*

MSc in Computer Science and Engineering

- **GPA: 4.25/4.3**
- Enriched Vision Applications Lab (EVA), Advisor: Dr. Wei-Chen Chiu
- **Selected Courses:** Deep Learning and Practice (**A+**), Reinforcement Learning (**A+**), Selected Topics in Visual Recognition using Deep Learning (**A+**), Computer Vision (**A+**)

National Chung Cheng University (CCU) *Sep. 2017 - Jan. 2021*

B.S. in Computer Science and Information Engineering

- **GPA:** Overall: **4.18/4.3**, Major: **4.21/4.3**, Ranking: **1/43**
- Machine Vision and Learning Lab (MVL), Advisor: Dr. Chen-Kuo Chiang
- **Honors:** Presidential Honor Award * 6 (F'17, S'18, F'18, S'19, F'19, S'20), College Student Research Scholarship
- **Selected Courses**: Machine Learning (**A+**), Statistics (**A+**), Object-Oriented Programming (**A+**), Data Structure (**A+**), Automatic Car Based on Learning Algorithm (**A+**), Compiler Design (**A+**)

Research Experiences

Oxford AI Governance Initiative, University of Oxford, *Advisor: Fazl Barez* *July. 2025 - CURRENT*

Research Fellow

- Researching scalable interpretability methods for LLM capability analysis and safety benchmarking.

Reinforcement Learning and Bandits Lab, NYCU, *Advisor: Ping-Chun Hsieh & Pin-Yu Chen & Mario Fritz* *Feb. 2025 - CURRENT*

Research Assistant

- Researching post-hoc interpretation of black-box T2I model misbehavior.
- Researching on RL backdoor attack detection and mitigation. *[Preprint]*

Enriched Vision Application Lab, NYCU, *Advisor: Wei-Chen Chiu & Pin-Yu Chen & Mario Fritz* *Aug. 2021 - Jan. 2025*

Research Assistant

- Researching on Red-teaming tool for safe T2I model developer. *[ICML'24]* *[Preprint]*
- Researched on salieny-guided masking as a novel data augmentation for contrastive-based vision SSL. *[WACV'24]*
- Researched on point cloud augmentation for non-color datasets. *[code]*

Machine Vision and Learning Lab, CCU, *Advisor: Chen-Kuo Chiang* *Mar. 2020 - Jan. 2021*

Undergraduate Researcher

- Researched on multi-target multi-camera vehicle tracking system. *[CVPRW'21]*
- Researched on applications for 6DoF robotic arms in calligraphy.

Selected Publications († indicates equal contribution)

[1] **Zhi-Yi Chin**†, Chieh-Ming Jiang†, Pin-Yu Chen, Ching-Chun Huang, Wei-Chen Chiu "Prompting4Debugging: Red-Teaming Text-to-Image Diffusion Models by Finding Problematic Prompts" *In Proceedings of the International Conference on Machine Learning (ICML) 2024.* *[project]* *[code]*

[2] **Zhi-Yi Chin**, Mario Fritz, Pin-Yu Chen, Wei-Chen Chiu "In-Context Experience Replay Facilitates Safety Red-Teaming of Text-to-Image Diffusion Models" *In Submission*

[3] Sze-Ann Chen, **Zhi-Yi Chin**, Chi-Yu Li, Ping-Chun Hsieh "Plan2Cleanse: Monte-Carlo Planning for Efficient Backdoor Detection and Mitigation in Reinforcement Learning" *In Submission*

[4] **Zhi-Yi Chin**†, Chieh-Ming Jiang†, Pin-Yu Chen, Ching-Chun Huang, Wei-Chen Chiu "Masking Improves Contrastive Self-Supervised Learning for ConvNets, and Saliency Tells You Where" *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2024.* *[code]*

[5] Yun-Lun Li, **Zhi-Yi Chin**, Ming-Ching Chang, Chen-Kuo Chiang "Multi-Camera Tracking by Candidate Intersection Ratio Tracklet Matching" *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops 2021.*

Projects

3D Point Cloud Augmentation via SRN *Jan. 2022*

MediaTek Research Project *[code]* *[slides]*

- Design a 3D point cloud augmentation based on a novel view synthesis method, scene representation networks, and use PointNet to evaluate our augmented point clouds quality.
- Replace instance object id with image features from ResNet to apply our method on unseen objects and do interpolation later on.
- Proposed method is successful in ModelNet10 and generates the augmented data by intra-class interpolation with ShapeNet in the latent space of SRN encoder.
- Observe limitation of novel view synthesis method on non-textured data.

Reimplementation Challenge *Jul. 2021*

Reinforcement Learning *[code] [report] [slides]*

- Reimplement ICLR 2018 paper: MAXIMUM A POSTERIORI POLICY OPTIMISATION in Pytorch.
- Successfully replicate the results in Cartpole, Hopper and Acrobot in MuJoCo environment.

Generative Models as Data Augmentation *Sep. 2021*

Deep Learning and Practice *[code] [slides] [video]*

- Investigate image transformation by exploring walks in the latent space of GAN.
- Use GAN steerability as an data augmentation technique.
- Conclude that GAN steerability is a better data augmentation technique compare to transformation done in the data space.

RSNA Pneumonia Detection *Jan. 2022*

Visual Recognition Using Deep Learning *[code] [report] [slides]*

- Design a two stage method, which first use a classification model to classify pneumonia, then use a detection model to locate the disease.
- Get the best results when using EfficientNet as classification model with 0.2 classification probability threshold when testing, and YOLOR as detection model. This method can reduce false positive results.
- Boost the final accuracy 2% by resizing the predicted bounding box to 87.5% of the original size.

Calendar Helper, Google *Aug. 2019*

Software Product Sprint Developer *[code]*

- A multifunctional Webapp for to-do lists and calendars.
- Using Javascript and JQuery as front-end and Java as back-end and host the Webapp on Google cloud console.
- Highlights: tagging system, nice dashboard design, synchronize with Google Calendar.

Honors & Awards

Reviewers, ICLR (2025), CVPR (2025)

Presidential Honor Award - top 4% students (6 times), CSIE Dept. at CCU *Fall '17, Spring '18, Fall '18, Spring '19, Fall '19, Spring '20*

College Student Research Scholarship, Ministry of Science and Technology, Taiwan *2020*

Google Student Travel Scholarship, Grace Hopper Celebration *Oct. 2019*