

Summary

I am an applied scientist at Amazon Smart Home AI. I received my Ph.D. degree (Computer Science) from the National University of Singapore in 2022. My research interest during my Ph.D. is *learning structured representations of visual scenes* which covers visual relationship detection, scene graph generation and video understanding. I have 7 years+ experience in computer vision and machine learning research, while I am also experienced in Large Language Models and Supervised Fine-Tuning. I code in Python and PyTorch, and I am experienced in AWS cloud computing and MLOps.

Experience

- Jul. 2022 – **Applied Scientist**, *Amazon*, Sunnyvale, CA, U.S.A.
 - Present Researching and developing computer vision algorithms for intelligent devices at the Smart Home team under Amazon Devices and Services. Worked on i) home/room layout estimation via 3D reconstruction and 2D polygon fitting, ii) feasibility study of pet behavior understanding via video action detection, and iii) enhancing Alexa for a more ambient smart home experience with Large Language Models (LLM) and Supervised FineTuning (SFT).
- Oct. 2020 – **Research Intern**, *TikTok*, Singapore.
 - Jun. 2022 Worked on i) unbiased scene graph generation via positive-unlabeled learning that achieves SOTA debiasing performance, ii) revealing the weakness of single-positive multi-label learning methods by adding real-world biases, and iii) improving smoking video detection by 10% at the Trust & Safety team.
- Jun. 2020 – **Research Intern**, *ASUS Intelligent Cloud Services*, Singapore.
 - Oct. 2020 Worked on video human-object interaction (HOI) detection. Specifically, I introduced a new video HOI benchmark, *VidHOI* and proposed a spatial-temporal model *ST-HOI* which surpasses 2D/3D baselines.
- Jul. 2013 – **Software Development Intern**, *Microsoft*, Taipei, Taiwan.
 - Jun. 2014 As a Microsoft Student Partner, I developed multiple Windows Apps, e.g., *NHK Reader* with 7K+ downloads, and gave Microsoft Tech Talks on software development to Taiwan's college students.

Education

- 2017–2022 **Ph.D., Computer Science**, *National University of Singapore*, Singapore.
 - Supervised by Prof. Jiashi Feng and Prof. Roger Zimmermann. Worked on: Visual relationship detection, Scene graph generation, Human-object interaction recognition, Video (spatial-temporal) understanding
- 2012–2016 **B.Sc., Electrical and Computer Engineering**, *National Chiao Tung University (Currently, National Yang Ming Chiao Tung University)*, Hsinchu, Taiwan.
 - Overall GPA: 3.89/4.30 (or 3.90/4.00). Took various computer science courses.
- 2014–2015 **Exchange Program, Information & Communication Engineering**, *University of Tokyo*, Japan.
 - During the 1-year program, I worked on efficient look-up table based SVM classifiers for image classification at the *Multimedia Processing Lab*, supervised by Prof. Toshihiko Yamasaki and Prof. Kiyoharu Aizawa.

Publications

- 2022 **Meng-Jiun Chiou**. *Learning Structured Representations of Visual Scenes*. PhD thesis, National University of Singapore, 2022.
- 2021 **Meng-Jiun Chiou**, Roger Zimmermann, and Jiashi Feng. Visual relationship detection with visual-linguistic knowledge from multimodal representations. *IEEE Access*, volume 9, pages 50441–50451. IEEE, 2021.
- 2021 **Meng-Jiun Chiou**, Chun-Yu Liao, Li-Wei Wang, Roger Zimmermann, and Jiashi Feng. St-hoi: A spatial-temporal baseline for human-object interaction detection in videos. In *Proceedings of the ACM International Conference on Multimedia Retrieval Workshops (ACM ICMR-W'21)*, pages 9–17, 2021.
- 2021 **Meng-Jiun Chiou**, Henghui Ding, Hanshu Yan, Changhu Wang, Roger Zimmermann, and Jiashi Feng. Recovering the unbiased scene graphs from the biased ones. In *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM'21)*, pages 1581–1590, 2021.

- 2020 **Meng-Jiun Chiou**, Zhenguang Liu, Yifang Yin, An-An Liu, and Roger Zimmermann. Zero-shot multi-view indoor localization via graph location networks. In *Proceedings of the 28th ACM International Conference on Multimedia (ACM MM'20)*, pages 3431–3440, 2020.
- 2019 Yifang Yin, **Meng-Jiun Chiou**, Zhenguang Liu, Harsh Shrivastava, Rajiv Ratn Shah, and Roger Zimmermann. Multi-level fusion based class-aware attention model for weakly labeled audio tagging. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM'19)*, pages 1304–1312, 2019.
- 2015 **Meng-Jiun Chiou**, Toshihiko Yamasaki, and Kiyoharu Aizawa. A fast table-based approach of bag-of-features for large-scale image classification. In *Proceedings of the ITE Annual Convention 2015 (ITE'15)*, pages 24A–1. The Institute of Image Information and Television Engineers, 2015.
- 2015 **Meng-Jiun Chiou**, Toshihiko Yamasaki, and Kiyoharu Aizawa. A fast method of visual words assignment of bag-of-features for object recognition. In *The 18th Meeting on Image Recognition and Understanding (MIRU'15)*, pages SS4–40, 2015.

Selected Projects

Affiliated with Amazon

- 2024 **Alexa Conversational Control for Smart Homes via Large Language Models.**
As an on-going project, we are working on an Alexa-related project aiming for a more ambient and conversational smart home experience with Large Language Models (LLM) and Supervised FineTuning (SFT). [[Video](#)] [[Media Coverage by Bloomberg](#)]
- 2023 **Alexa Map View: Room Layout Estimation via 3D Reconstruction and 2D Polygon Fitting.**
Worked on Alexa Map View (announced in Sep. 2023). Researched and prototyped a real-time home layout estimation system with RGB videos as input based on 3D mesh reconstruction methods, e.g., [NeuralRecon](#), and 2D polygon fitting approaches with rectilinear constraints for post-processing layouts, e.g. [Floorplan Fitting](#). Worked on it end-to-end including literature review, algorithm implementation, metric definition and evaluation, and failure analysis. [[Alexa Map View](#)] [[Video](#)]
- 2022 **Pet Behavior Understanding by Action Classification & Spatio-Temporal Action Detection.**
We took an early-stage initiative to study pet behavior understanding. I worked on the whole end-to-end experiment pipeline, i.e., problem definition (action classification/spatio-temporal action detection), data collection/cleaning/exploration, algorithm implementation, metric definition and error analysis. Training large 3D backbones with debiasing method like re-weighting, we obtained promising models for our goal.
- 2022 **Improving Smoking Video Detection with new Architectures and Augmentations.**
We implemented SOTA data augmentation techniques including Mixup, Cutout and CutMix, and various new visual backbones such as Swin Transformer and we ended up *improving the smoking video detection performance by around 10 percent measured by recall*.
- 2021 **Revealing the biases in Single-Positive Multi-Label Learning.**
We revealed that the current Single-Positive Multi-Label (SPML) methods do not consider labeling bias such as *bounded rationality* and *reporting bias*, and we showed that *adding theses real-world biases to the existing SPML models would undermine their performance*. [[Slides](#)]
- 2021 **Unbiased Scene Graph Generation with Positive-Unlabeled Learning.**
We introduced *Dynamic Label Frequency Estimation* (DLFE) for debiasing scene graph generation (SGG). Applying DLFE to SGG methods we got *new SOTA debiasing performance*, i.e., *+5 averaged mean recall (24%→29%) or +21 tail-part recall (17%→38%)* v.s. previous SOTAs. [[Paper](#)] [[Source Code](#)] [[Slides](#)] [[Poster](#)] [[Video](#)]

Affiliated with ASUS Intelligent Cloud Services & National University of Singapore

- 2020 **Human-Object Interaction Detection in Videos.**
We *introduced* a keyframe-centered, large-scale video human-object interaction detection benchmark named *VidHOI*. Proposed a strong baseline called *ST-HOI* *outperforming the 2D/3D baseline models by obtaining 74% relatively or 6.1% absolutely higher mAP (8.3%→14.4%)* on temporal-related HOIs. [[Paper](#)] [[Source Code & Dataset](#)] [[Slides](#)] [[Video](#)]

Affiliated with National University of Singapore

- 2020 **Visual Relationship Detection with External Knowledge.**
We introduced a novel Transformer-based multi-modal visual relation detection architecture, named Relational Visual-Linguistic BERT (*RVL-BERT*), enriched by the visual-linguistic knowledge from large-scale external datasets. *RVL-BERT achieved SOTA performance* on the SpatialSense dataset and competitive results on the VRD and VG datasets. [[Paper](#)] [[Source Code](#)]

2019 **Zero-Shot Indoor Localization with Floor Plans.**

We introduced a multi-view image-based indoor localization system named *GLN* achieving SOTA performance. Also proposed a zero-shot learning pipeline where we utilize the proposed *Map2Vec* location-aware embeddings. *Zero-shot GLN achieves promising results, e.g., 56.3% 5-meter localization error.* [[Paper](#)] [[Source Code](#)] [[Poster](#)] [[Video](#)]

2018 **Weakly-Labeled Audio Tagging with Attention-based Model.**

We introduced a multi-level attention-based audio tagging model making segment-level predictions with temporal modeling, followed by aggregations along both time and feature domains. *Our method achieves SOTA audio tagging results.* [[Paper](#)]

2017 **Real-Time On-Device Blind Navigation.**

The Light navigates blind people to move around smoothly in real time using *MobileNet* for object segmentation. It won 2nd prize at *iNTUition Hackathon 2017*. [[Project Page](#)] [[Source Code](#)]

[Affiliated with National Chiao Tung University](#)

2016 **Right Whale Identification with Fast R-CNN.**

We developed a right whale identification system face by training *Fast R-CNN* on a large-scale Kaggle dataset. [[Technical Report](#)] [[Source Code](#)]

[Affiliated with Univeristy of Tokyo](#)

2015 **Fast Image Recognition with Look-Up Tabled-based Bag-of-Features.**

Table-Based Bag-of-Features (Table-Based BoF) is a fast look-up table based method for finding bag-of-features-based indexes of query pictures without feature extraction. [[Paper](#)] [[Source Code](#)]

Academic Services

2018–Present **Program Committee**, *NeurIPS Workshop on Distribution Shifts (DistShift)* ('23/'22/'21), *ACMMM'22 Open-Source Program*, *BigMM'20 Graduate Student Consortium*, *CVPR'18 Workshop on Visual Understanding of Humans in Crowd Scene*

2018–Present **Reviewer**, *ACMMM* ('23/'22/'21/'20), *IEEE TMM* ('23/'21), *IET Computer Vision* ('23), *Journal of Imaging* ('23), *IEEE TIP* ('22/'20), *Springer MMSJ* ('19), *NUS MSCS Admission* ('21/'20)

2017–2021 **Teaching Assistant**, *Big-Data Analytics Technology* (NUS, '21), *Computer Vision and Pattern Recognition* (NUS, '19/'18), *Data Structures and Algorithms* (NUS, '17), *Special Friday Lecture for High School Students* (UTokyo, '15)

Scholarships & Awards

2017 **2nd Place**, *iNTUition Hackathon 2017*, a 24-hour hackathon at *Nanyang Technological University*.

2017 **NUS Research Scholarship** including full tuition waiver and monthly stipend, awarded by the *National University of Singapore*.

2015 **Helm Technology Scholarship** awarded by the *Helm Technology Inc.*, Taiwan.

2014 **Student Exchange Support Program** scholarship for exchange students to the *University of Tokyo*, awarded by *Japan Student Services Organization*.

2014 **Short Term Exchange Scholarship** for outbound exchange students, awarded by the *National Chiao Tung University*.

2014 **Xiao Yuan-Long Scholarship** for students with superb GPA, awarded by *National Chiao Tung University*.

Skills

Programming *PyTorch*, *Python*, *Matlab*, *C*, *C++*

Miscellaneous *AWS Cloud Computing* [[Course Cert](#)], *MLOps* [[Course Cert](#)], *Docker*

Language *Mandarin Chinese* (native speaker), *English* (fluent) and *Japanese* (fluent; *JLPT N1*)

Position of Responsibility

2013-2014 **Vice President**, *Chien-Kuo & Taipei First Girls' High School Alumni Association*, *National Chiao Tung University*.

I took on leadership roles to organize a variety of events for the two high schools' alumni.