

Meng-Jiun Chiou

Taipei, Taiwan

✉ mengjiun.chiou@u.nus.edu

🌐 [Personal Webpage](#)

🔗 [Scholar](#)

🐙 [Github](#)

in [Linkedin](#)

📝 [Medium](#)

Summary

I am a computer vision researcher at *Amazon*. I received my Ph.D. degree from the National University of Singapore in 2022. My research interest is *learning structured representations of visual scenes*, where related tasks include visual relationship detection, scene graph generation, action detection in video, etc. I have been tackling computer vision problems with multi-modal models using images, videos, texts, etc., and from an unbiased, real-world applications' view. I code mostly in Python and PyTorch. I speak Chinese, English and Japanese.

Experience

- Jul. 2022 – **Computer Vision Applied Scientist**, *Amazon*, Taipei, Taiwan.
Present I respond to researching and developing computer vision algorithms for intelligent devices at the Device AI team under Amazon Lab126.
- Oct. 2020 – **Computer Vision Research Intern**, *TikTok*, Singapore.
Jun. 2022 I worked on i) unbiased scene graph generation by proposing DLFE 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 – **Computer Vision Research Intern**, *ASUS Intelligent Cloud Services*, Singapore.
Oct. 2020 I 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 (MSP)**, *Microsoft*, Taipei, Taiwan.
Jun. 2014 I responded to i) developing Windows Apps, e.g., *NHK Reader* with 7K+ downloads, and ii) giving Microsoft Tech Talks on software development to university students in Taiwan.

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.
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 Scene*. 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 Aizawa Kiyoharu. 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 TikTok (ByteDance AI Lab) & National University of Singapore

- 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.**
Collaborator: **Dr. Houjing Huang**, TikTok, Singapore
- 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]
Collaborators: **Dr. Junhao Liew** and **Dr. Henghui Ding**, TikTok, Singapore
- 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**, specifically, **+5 averaged mean recall points (24%→29%)** or **+21 tail recall points (17%→38%)** than the previous SOTAs. [Paper] [Source Code] [Slides] [Poster] [Video]
Collaborators: **Dr. Henghui Ding** and **Dr. Changhu Wang**, TikTok, Singapore, **Prof. Jiashi Feng** and **Prof. Roger Zimmermann** and **Hanshu Yan**, National University of Singapore

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]
Collaborators: **Chun-Yu Liao** and **Li-Wei Wang**, ASUS Intelligent Cloud Services, Singapore, **Prof. Jiashi Feng** and **Prof. Roger Zimmermann**, National University of Singapore, Singapore

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]

Collaborators: **Prof. Jiashi Feng** and **Prof. Roger Zimmermann**, *National University of Singapore*, Singapore

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](#)]

Collaborators: **Dr. Zhenguang Liu** and **Dr. Yifang Yin** and **Prof. Roger Zimmermann**, *National University of Singapore*, Singapore

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](#)]

Collaborators: **Dr. Zhenguang Liu** and **Dr. Yifang Yin** and **Prof. Roger Zimmermann**, *National University of Singapore*, Singapore

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](#)]

Collaborators: **Xiao Ma** and **Yaqi Xie** and **Ziwei Xu**, *National University of Singapore*, Singapore
[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](#)]

Collaborators: **Li-An Yang** and **Kai-Sheng Yang**, *National Chiao Tung University*
[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](#)]

Collaborators: **Prof. Yamasaki Toshihiko** and **Prof. Aizawa Kiyoharu**, *University of Tokyo*, Japan

Academic Services

2018–Present **Program Committee**, *NeurIPS Workshop on Distribution Shifts in Real-World Applications* ('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* ('22/'21/'20), *IEEE TMM* ('21), *IEEE TIP* ('20), *ACM TOMM* ('20), *Springer MMSJ* ('19), *NUS MSCS Admission* ('21/'20)

2017–Present **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** at *iNTUition Hackathon 2017*, a 24-hour hackathon held at *Nanyang Technological University*, Singapore.

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 student to the *University of Tokyo*, awarded by *Japan Student Services Organization*.

- 2014 **Short Term Exchange Scholarship** for outbound exchange student, awarded by the *National Chiao Tung University*.
- 2014 **Xiao Yuan-Long Scholarship** for students with superb GPA, awarded by the *National Chiao Tung University*.

Skills

- Deep-learning Frameworks PyTorch, Lightning
- Programming Languages Python, Matlab, C, C++
- 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.