

Delong Chen (陈德龙)

<https://chendelong.world> | delong.chen@connect.ust.hk | [Google Scholar](#) | [GitHub](#)

Education

Hong Kong University of Science and Technology (HKUST) Ph.D. Student in ECE Department, advised by Prof. Pascale Fung	Sept. 2023 – Present Hong Kong SAR
Hohai University B.Eng. in Computer Science, advised by Prof. Fan Liu	Sept. 2017 – Jun. 2021 Nanjing, China

Experience

Research Intern <i>Visual instruction tuning for multimodal LLMs</i>	Xiaobing.AI (Microsoft Xiaoice)	Mar. 2023 – Sept. 2023 Beijing, China
Research Intern <i>Vision-language learning in embodied AI domain</i>	Microsoft Research Asia (MSRA)	Oct. 2022 – Jan. 2023 Beijing, China
Research Intern <i>Large-scale contrastive vision language pretraining</i>	MEGVII Research (Face++)	Oct. 2021 – Oct. 2022 Beijing, China
Research Assistant <i>Vision-language foundation model for remote sensing</i>	Hohai University	Sept. 2021 – Sept. 2023 Nanjing, China
Summer Program <i>Linguistics and computation for natural language processing</i>	University of British Columbia	Jul. 2018 – Aug. 2018 Vancouver, Canada

Awards

Best Paper at AAAI 2023 Inaugural Summer Symposium Series - AI x Metaverse
Best Demo Award in IEEE International Conference on Multimedia and Expo (ICME) 2021
Best Dataset Paper Award in Long-Tailed Distribution Learning Workshop, IJCAI 2021
Best Presentation Award in International Conference on Big Data and Artificial Intelligence (BDAI) 2021
First-Class Outstanding Thesis in Jiangsu Province
Outstanding Thesis of Hohai University
Outstanding Graduate of Hohai University
Person of the Year of the in Jiangsu Province Nomination
Person of the Year of Hohai University in 2019
Third Prize of 2018 the 8th China Software Cup, East China Division Finals (team Leader)

Publications

* joint first author ✉ corresponding author

Vision-language Learning

Delong Chen, Samuel Cahyawijaya, Etsuko Ishii, Ho Shu Chan, Yejin Bang, Pascale Fung

The Pyramid of Captions

arXiv Preprint, 2024 [[arXiv](#)]

Delong Chen, Samuel Cahyawijaya, Jianfeng Liu, Baoyuan Wang, Pascale Fung

Subject-level Image Tokenization

arXiv Preprint, 2024. [[GitHub](#)] [[arXiv](#)] [[video](#)]

Delong Chen, Jianfeng Liu, Wenliang Dai, Baoyuan Wang

Visual Instruction Tuning with Polite Flamingo

Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI), 2024. [[GitHub](#)] [[arXiv](#)] (Oral)

[Delong Chen](#), Zhao Wu, Fan Liu, Zaiquan Yang, Huaxi Huang, Ying Tan, Erjin Zhou.

ProtoCLIP: Prototypical Contrastive Language Image Pretraining

IEEE Transactions on Neural Networks and Learning Systems (T-NNLS), 2023. [[GitHub](#)] [[arXiv](#)] [[doi](#)]

Fan Liu*, [Delong Chen](#)*, Zhangqingyun Guan, Xiacong Zhou, Jiale Zhu, Jun Zhou

RemoteCLIP: A Vision Language Foundation Model for Remote Sensing

IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2024. [[GitHub](#)] [[arXiv](#)] [[doi](#)]

[Delong Chen](#)*, Fan Liu, Xiaoyu Du, Ruizhuo Gao, Feng Xu

MEP-3M: A Large-scale Multi-modal E-commerce Product Dataset

IJCAI-21 Workshop on Long-Tailed Distribution Learning, 2021 (Best Dataset Paper)
Pattern Recognition, 2023. [[GitHub](#)] [[doi](#)]

Large Language Models

Samuel Cahyawijaya*, [Delong Chen](#)*, Yejin Bang*, Leila Khalatbari, Bryan Wilie, Ziwei Ji, Etsuko Ishii, Pascale Fung*

High-Dimension Human Value Representation in Large Language Models

arXiv Preprint, 2024. [[arXiv](#)]

Yejin Bang, [Delong Chen](#), Nayeon Lee, Pascale Fung

Measuring Political Bias in Large Language Models: What Is Said and How It Is Said

arXiv Preprint, 2024. [[arXiv](#)]

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, [Delong Chen](#), Ho Shu Chan, Wenliang Dai, Andrea Madotto, Pascale Fung

Survey of Hallucination in Natural Language Generation


arXiv Preprint, 2023 [[arXiv](#)]

Music, and Audio Processing

[Delong Chen](#), Fan Liu, Zewen Li, Feng Xu

VirtualConductor: Music-driven Conducting Video Generation System

IEEE International Conference on Multimedia & Expo (ICME) 2021. [[video](#)] (Best Demo)

Fan Liu, [Delong Chen](#) , Ruizhi Zhou, Sai Yang, Feng Xu

Self-Supervised Music Motion Synchronization Learning for Music-Driven Conducting Motion Generation

Journal of Computer Science and Technology (JCST), 2022 [[GitHub](#)] [[video](#)] [[doi](#)]

Zhuoran Zhao, Jinbin Bai, [Delong Chen](#), Debang Wang, Yubo Pan

Taming Diffusion Models for Music-driven Conducting Motion Generation

AAAI 2023 Inaugural Summer Symposium Series – AI x metaverse. [[GitHub](#)] [[arXiv](#)] [[doi](#)] (Best Paper)

Xinyu Zhou*, [Delong Chen](#)*, Yudong Chen

Towards Joint Modeling of Dialogue Response and Speech Synthesis based on Large Language Model

The 6th International Conference on Natural Language and Speech Processing (ICNLSP), 2023. [[GitHub](#)] [[arXiv](#)]

Few-shot Visual Recognition

Sai Yang, Fan Liu, [Delong Chen](#), Jun Zhou

Few-shot Classification via Ensemble Learning with Multi-Order Statistics

Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI), 2023. [[arXiv](#)] (Oral)

Fan Liu, Sai Yang, [Delong Chen](#), Huaxi Huang, Jun Zhou

Few-shot Classification Guided by Generalization Error Bound

Pattern Recognition, 2023. [[doi](#)]

Fan Liu*, [Delong Chen](#)*, Fei Wang, Zewen Li, Feng Xu

Deep Learning based Single Sample Face Recognition: A Survey

Artificial Intelligence Review, 2022. [[arXiv](#)] [[doi](#)]

Professional Services

Reviewer: ACL Rolling Review, ACM Multimedia (2023, 2024), ACM Transactions on Intelligent Systems and Technology (TIST), IEEE Transactions on Intelligent Transportation Systems (TITS), Artificial Intelligence Review, Journal of Digital Imaging

Teaching Assistant: A System View of Communications (2024 Spring, HKUST ELEC-1200)