Haoxuan You

Columbia University, 526 W 122nd St., NYC, NY, 10027 | Website | Last Updated in Apr/2024
hy2612@columbia.edu | haoxuanyou@gmail.com | Google Scholar: BhysChMAAAAJ | +1 6462263052

Education

Columbia University New York, NY
- Doctor of Philosophy (Ph.D.) in Computer Science *Sep. 2019 - May. 2024 (Expected)*
- Master of Philosophy (MPhil) in Computer Science *Sep. 2019 - May. 2023*
- Master of Science (MSC) in Computer Science *Sep. 2019 - Feb. 2021*
- Adviser: Prof. Shih-Fu Chang (Dean of Engineering School; Member of the National Academy of Engineering; Fellow of NAI, AAAS, ACM and IEEE)

Xidian University Xi'an, China
- Bachelor of Engineering in Electronic Information Engineering *Sep. 2014 - Jun. 2018*
- GPA: 3.77/4.0

Research Interests

Now: Vision-Language Learning; Multimodal Foundation Models;
Previous: 3D Vision Learning; Graph Neural Networks.

Experience

AI/ML, Apple Seattle
Research Intern *May. 2023 - Now*
- Advisor: Liangliang Cao, Yinfei Yang, Zhe Gan
- Research about Multimodal Foundation Model. Published a paper at ICLR 2024.

Google Research, Google NYC
Student Researcher *May. 2022 - Mar. 2023*
- Advisor: Mandy Guo, Jiahui Yu, Jason Baldridge
- Research in Large-Scale Bi-directional Image-Text Generation. Paper accepted at ICLR 2024.

Azure Cognitive Services Research, Microsoft Remote
Research Intern *Jun. 2021 - Aug. 2021*
- Advisor: Luowei Zhou
- Research about Modality-Shared Contrastive Image-Text Pre-training. Paper accepted at ECCV2022.

Paper - Preprint

** means equal contribution*

[1] Ferret-v2: An Improved Baseline for Referring and Grounding with Large Language Models.
Haotian Zhang, **Haoxuan You***, Philipp Dufter, Bowen Zhang, Chen Chen, Hong-You Chen, Tsu-Jui Fu et al.*
arXiv preprint arXiv:2404.07973 (2024).

[2] LLM-based Conversational AI Therapist for Daily Functioning Screening and Psychotherapeutic Intervention via Everyday Smart Devices
*Jingping Nie, Hanya Shao, Yuang Fan, Qijia Shao, **Haoxuan You**, Matthias Preindl, and Xiaofan Jiang*
arXiv preprint arXiv:2403.10779 (2024).

[3] Multimodal adaptive distillation for leveraging unimodal encoders for vision-language tasks
*Zhecan Wang, Noel Codella, Yen-Chun Chen, ... , **Haoxuan You**, Kai-Wei Chang, Shih-fu Chang, Lu Yuan*
arXiv preprint arXiv:2204.10496 (2022).

[4] Graph-MLP: Node Classification without Message Passing in Graph
*Yang Hu, **Haoxuan You**, Zhecan Wang, Zhicheng Wang, Erjin Zhou, and Yue Gao.*
arXiv preprint arXiv:2106.04051 (2021).

Paper - Publication

** means equal contribution*

[1] Ferret: Refer and Ground Anything Anywhere at Any Granularity
Haoxuan You*, Haotian Zhang*, Zhe Gan, Xianzhi Du, Bowen Zhang, Zirui Wang, Liangliang Cao, Shih-Fu Chang, Yinfei Yang*
*International Conference on Learning Representations(**ICLR**) 2024*

[2] CoBIT: A Contrastive Bi-directional Image-Text Generation Model
***Haoxuan You**, Mandy Guo, Zhecan Wang, Kai-Wei Chang, Jason Baldridge, Jiahui Yu*
*International Conference on Learning Representations(**ICLR**) 2024*

[3] Dataset Bias Mitigation in Multiple-Choice Visual Question Answering and Beyond
*Zhecan Wang, Long Chen, **Haoxuan You**, Keyang Xu, Yicheng He, Wenhao Li, Noal Codella, Kai-Wei Chang, Shih-Fu Chang*
*Findings of Empirical Methods in Natural Language Processing (**EMNLP-findings**) 2023*

[4] IdealGPT: Iteratively Decomposing Vision and Language Reasoning via Large Language Models
Haoxuan You*, Rui Sun*, Zhecan Wang*, Long Chen, Gengyu Wang, Hammad A Ayyubi, Kai-Wei Chang, Shih-Fu Chang*
*Findings of Empirical Methods in Natural Language Processing (**EMNLP-findings**) 2023*

[5] UniFine: A Unified and Fine-grained Approach for Zero-shot Vision-Language Understanding
Rui Sun, Zhecan Wang*, **Haoxuan You***, Noel Codella, Kai-Wei Chang, Shih-Fu Chang*
*Annual Meeting of the Association for Computational Linguistics **(ACL-findings)** 2023*

[6] Learning Visual Representation from Modality-Shared Contrastive Language-Image Pre-training
Haoxuan You*, Luowei Zhou*, Bin Xiao*, Noel Codella*, Yu Cheng, Ruochen Xu, Shih-Fu Chang, Lu Yuan.*
*European Conference on Computer Vision (**ECCV**) 2022*

[7] Find Someone Who: Visual Commonsense Understanding in Human-Centric Grounding
***Haoxuan You**, Rui Sun, Zhecan Wang, Kai-Wei Chang, Shih-Fu Chang.*
*Findings of Empirical Methods in Natural Language Processing (**EMNLP-findings**) 2022*

[8] Understanding ME? Multimodal Evaluation for Fine-grained Visual Commonsense
*Zhecan Wang, **Haoxuan You**, Yicheng He, Wenhao Li, Kai-Wei Chang, Shih-Fu Chang*
*Empirical Methods in Natural Language Processing (**EMNLP**) 2022*

[9] SHREC'22 track: Open-set 3D object retrieval.
*Yifan Feng, Yue Gao, Xibin Zhao, Yandong Guo, Nihar Bagewadi, ..., **Haoxuan You**, Difei Zhu*
Computers & Graphics 107 (2022): 231-240

[10] Bridging the Gap between Recognition-level Pre-training and Commonsensical Vision-language Tasks.
*Yue Wan, Yueen Ma, **Haoxuan You**, Zhecan Wang, and Shih-Fu Chang.*
ACL 2022 Workshop on Commonsense Representation and Reasoning. 2022.

[11] Rethinking network design and local geometry in point cloud: A simple residual mlp framework
*Xu Ma, Can Qin, **Haoxuan You**, Haoxi Ran, Yun Fu.*
*International Conference on Learning Representations(**ICLR**) 2022*

[12] SGEITL: Scene Graph Enhanced Image-Text Learning for Visual Commonsense Reasoning
Zhecan Wang, **Haoxuan You***, Alireza Zareian, Liunian Li, Suji Park, Yiqing Liang, Kai-Wei Chang, Shih-Fu Chang.*
*36th AAAI Conference on Artificial Intelligence (**AAAI**) 2022*

[13] Unsupervised Vision-and-Language Pre-training Without Parallel Images and Captions
*Liunian Harold Li, **Haoxuan You**, Zhecan Wang, Alireza Zareian, Shih-Fu Chang, Kai-Wei Chang*
*Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**) 2021*

[14] Learning Visual Commonsense for Robust Scene Graph Generation
Alireza Zareian, Zhecan Wang*, **Haoxuan You***, Shih-Fu Chang.*
*European Conference on Computer Vision (**ECCV**) 2020*

[15] PointHop: An Explainable Machine Learning Method for Point Cloud Classification
*Min Zhang, **Haoxuan You***, Pranav Kadam, Shan Liu, C-C Jay Kuo. (*Corresponding Author)*
IEEE Transactions on Multimedia, VOL. 22, NO. 7, JULY 2020

[16] PointDAN: A Multi-Scale 3D Domain Adaption Network for Point Cloud Representation
Can Qin, **Haoxuan You***, Lichen Wang, C-C Jay Kuo, Yun Fu.*
*Conference on Neural Information Processing Systems (**NeurIPS**) 2019*

[17] Decoding EEG by Visual-guided Deep Neural Networks
*Zhicheng Jiao, **Haoxuan You**, Fan Yang, Xin Li, Han Zhang, Dinggang Shen.*
*28th International Joint Conference on Artificial Intelligence (**IJCAI**) 2019*

[18] Multi-modality latent interaction network for visual question answering
*Peng Gao, **Haoxuan You**, Zhanpeng Zhang, Xiaogang Wang, Hongsheng Li.*
*IEEE International Conference on Computer Vision (**ICCV**) 2019*

[19] Dynamic Fusion with Intra- and Inter-modality Attention Flow for Visual Question Answering
*Peng Gao, Hongsheng Li, **Haoxuan You**, Zhengkai Jiang, Pan Lu, Steven Hoi, Xiaogang Wang.*
*IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**) 2019*

[20] PVRNet: Point-View Relation Neural Network for 3D Shape Recognition
***Haoxuan You**, Yifan Feng, Xibin Zhao, Changqing Zou, Rongrong Ji, Yue Gao.*
*33rd AAAI Conference on Artificial Intelligence (**AAAI**) 2019*

[21] Hypergraph Neural Network
*Yifan Feng, **Haoxuan You**, Rongrong Ji, Yue Gao.*
*33rd AAAI Conference on Artificial Intelligence (**AAAI**) 2019*

[22] MeshNet: Mesh Neural Network for 3D Shape Representation
*YuTong Feng, Yifan Feng, **Haoxuan You**, Xibin Zhao, Yue Gao.*
*33rd AAAI Conference on Artificial Intelligence (**AAAI**) 2019*

[23] PVNet: A Joint Convolutional Network of Point Cloud and Multi-View for 3D Shape Recognition
***Haoxuan You**, Yifan Feng, Rongrong Ji, Yue Gao.*
*ACM International Conference on Multimedia (**ACM MM**) 2018*