

# LICHENG YU

✉ [lichengyu24@gmail.com](mailto:lichengyu24@gmail.com)    📞 [+1-9198088511](tel:+19198088511)    🏠 [lichengunc.github.io](https://lichengunc.github.io)    [in licheng-yu](https://www.linkedin.com/in/licheng-yu)    [g Licheng Yu](https://www.google.com/search?q=Licheng+Yu)    [lichengunc](https://github.com/lichengunc)

I am a Research Scientist Manager at Facebook AI. My research interest lies in the intersection between Computer Vision and Natural Language Processing. I completed my PhD in Computer Science from University of North Carolina at Chapel Hill in 2019. My advisor is [Tamara L. Berg](#). My research goal is to build an artificial intelligent system that can communicate with people in a natural way, involving asking and answering questions, commonsense reasoning, and performing actions to better people's life.

## EDUCATION

<b>Ph.D in Computer Science</b> , <i>University of North Carolina at Chapel Hill</i>	2014.08 — 2019.05
<b>M.S in Information Engineering</b> , <i>Shanghai Jiao Tong University</i>	2011.09 — 2014.04
<b>B.S in Information Engineering</b> , <i>Shanghai Jiao Tong University</i>	2007.09 — 2011.07

## WORK EXPERIENCE

<b>Facebook AI</b> <i>Research Scientist Manager</i>	<b>Menlo Park, CA</b> 2023.06 — Present
<ul style="list-style-type: none"> <li>Supporting a talented team working on Video Generation and its various applications.</li> </ul>	
<i>Staff Research Scientist</i>	2022.08 — 2023.06
<ul style="list-style-type: none"> <li>Foundation Text-to-Image Generation model and its various applications.</li> <li>Multimodal (Vision+Language) Research and Applications for Ads.</li> </ul>	
<i>Senior Research Scientist</i>	2021.07 — 2022.08
<ul style="list-style-type: none"> <li>Vision+Language Research, e.g., large-scale multimodal pre-training, interactive visual search, etc.</li> <li>Multimodal Applications on E-Commerce, e.g., IG Shops, Marketplace, Visual Search, etc.</li> </ul>	
<i>Research Scientist</i>	2020.03 — 2021.07
<ul style="list-style-type: none"> <li>Vision+Language Research, e.g., visual question answering, trace-guided image captioning, etc.</li> <li>Multimodal Applications on Ads Ranking, building content-based triggers and features.</li> </ul>	
<b>Microsoft</b> <i>Researcher</i>	<b>Bellevue, WA</b> 2019.06 — 2020.03
<ul style="list-style-type: none"> <li>Vision+Language Research, e.g., image-text pre-training, video-text pre-training, image synthesis, etc.</li> </ul>	

## PUBLICATION

- [41] Hu Xu, Saining Xie, Po-Yao Huang, **Licheng Yu**, Russell Howes, Gargi Ghosh Luke Zettlemoyer, Christoph Feichtenhofe, "CiT: Curation in Training for Effective Vision-Language Data", in ICCV 2023.
- [40] Barry Menglong Yao, Yu Chen, Qifan Wang, Sijia Wang, Minqian Liu, Zhiyang Xu, **Licheng Yu**, Lifu Huang, "AMELI: Enhancing Multimodal Entity Linking with Fine-Grained Attributes", in arXiv:2305.14725.
- [39] Medhini Narasimhan, **Licheng Yu**, Sean Bell, Ning Zhang, Trevor Darrell, "Learning and Verification of Task Structure in Instructional Videos", in arXiv:2303.13519.
- [38] Tsu-Jui Fu, **Licheng Yu**, Ning Zhang, Cheng-Yang Fu, Jong-Chyi Su, William Yang Wang, Sean Bell, "Tell Me What Happened: Unifying Text-guided Video Completion via Multimodal Masked Video Generation", in CVPR 2023.
- [37] Yiwu Zhong, **Licheng Yu**, Yang Bai, Shangwen Li, Xueting Yan, Yin Li, "Learning Procedure-aware Video Representation from Instructional Videos and Their Narrations", in CVPR 2023.
- [36] Xiao Han, Xiatian Zhu, **Licheng Yu**, Li Zhang, Yi-Zhe Song, Tao Xiang, "FAME-ViL: Multi-Tasking Vision-Language Model for Heterogeneous Fashion Tasks", in CVPR 2023.
- [35] Sangwoo Mo, Jong-Chyi Su, Kevin Chih-Yao Ma, Mido Assran, Ishan Misra, **Licheng Yu**, Sean Bell, "RoPAWS: Robust Semi-supervised Representation Learning from Uncurated Data", in ICLR 2023.
- [34] Yunzhong He, Yuxin Tian, Mengjiao Wang, Feier Chen, **Licheng Yu**, Maolong Tang, Congcong Chen, Ning Zhang, Bin Kuang, Arul Prakash, "Que2Engage: Embedding-based Retrieval for Relevant and Engaging Products at Facebook Marketplace", in WWW 2023.
- [33] Suvir Mirchandani, **Licheng Yu**, Mengjiao Wang, Animesh Sinha, Wenwen Jiang, Tao Xiang, Ning Zhang, "FaD-VLP: Fashion Vision-and-Language Pre-training towards Unified Retrieval and Captioning", in EMNLP 2022.

- [32] **Licheng Yu**, Jun Chen, Animesh Sinha, Mengjiao Wang, Yu Chen, Tamara L. Berg, Ning Zhang, "CommerceMM: Large-Scale Commerce MultiModal Representation Learning with Omni Retrieval", in KDD 2022.
- [31] Xiao Han, **Licheng Yu**, Xiatian Zhu, Li Zhang, Yi-Zhe Song, Tao Xiang, "FashionViL: Fashion-Focused Vision-and-Language Representation Learning", in ECCV 2022.
- [30] Yuxuan Wang, Difei Gao, **Licheng Yu**, Weixian Lei, Matt Feiszli, Mike Zheng Shou, "Generic Event Boundary Captioning: A Benchmark for Status Changes Understanding", in ECCV 2022.
- [29] Mingyang Zhou\*, **Licheng Yu**\*, Amanpreet Singh, Mengjiao Wang, Yu Zhou, Ning Zhang, "Unsupervised Vision-and-Language Pre-training via Retrieval-based Multi-Granular Alignment", in CVPR 2022 (Oral). (\* denotes equal contribution).
- [28] Jie Lei, Xinlei Chen, Ning Zhang, Mengjiao Wang, Mohit Bansal, Tamara L. Berg, **Licheng Yu**, "LOOPITR: Combining Dual and Cross Encoder Architectures for Image-Text Retrieval", in arxiv:2203.05465v1.
- [27] Linjie Li, Jie Lei, Zhe Gan, **Licheng Yu**, Yen-Chun Chen, Rohit Pillai, Yu Cheng, Luwei Zhou, Xin Eric Wang, William Yang Wang, Tamara L. Berg, Mohit Bansal, Jingjing Liu, Lijuan Wang, Zicheng Liu, "VALUE: A Multi-Task Benchmark for Video-and-Language Understanding Evaluation", in NeurIPS 2021.
- [26] Zihang Meng, **Licheng Yu**, Ning Zhang, Tamara L. Berg, Babak Damavandi, Vikas Singh, Amy Bearman, "Connecting What to Say With Where to Look by Modeling Human Attention Traces", in CVPR 2021.
- [25] Jie Lei, **Licheng Yu**, Tamara L. Berg, Mohit Bansal, "What Happens Next? Video-and-Language Future Event Prediction", in EMNLP 2020.
- [24] Linjie Li\*, Yen-Chun Chen\*, Yu Cheng, Zhe Gan, **Licheng Yu**, Jingjing Liu, "HERO: Hierarchical Encoder for Video+Language Omni-representation Pre-training", in EMNLP 2020 (\* denotes equal contribution).
- [23] Jize Cao, Zhe Gan, Yu Cheng, **Licheng Yu**, Yen-Chun Chen, Jingjing Liu, "Behind the Scene: Revealing the Secrets of Pre-trained Vision-and-Language Models", in ECCV 2020 (spotlight).
- [22] Yen-Chun Chen\*, Linjie Li\*, **Licheng Yu**\*, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, Yu Cheng, Jingjing Liu, "UNITER: Learning Universal Image-Text Representations", in ECCV 2020 (\* denotes equal contribution).
- [21] Jie Lei, **Licheng Yu**, Tamara L. Berg, Mohit Bansal, "TVR: A Large-Scale Dataset for Video-Subtitle Moment Retrieval", in ECCV 2020.
- [20] Jie Lei, **Licheng Yu**, Tamara L. Berg, Mohit Bansal, "TVQA+: Spatio-Temporal Grounding for Video Question Answering", in ACL 2020.
- [19] Yandong Li, Yu Cheng, Zhe Gan, **Licheng Yu**, Liqiang Wang, Jingjing Liu, "BachGAN: High-Resolution Image Synthesis from Salient Object Layout", in CVPR 2020.
- [18] Jingzhou Liu, Wenhui Chen, Yu Cheng, Zhe Gan, **Licheng Yu**, Yiming Yang, Jingjing Liu, "VIOLIN: A Large-Scale Dataset for Video-and-Language Inference", in CVPR 2020.
- [17] **Licheng Yu**, Xinlei Chen, Georgia Gkioxari, Mohit Bansal, Tamara L. Berg, Dhruv Batra, "Multi-target Embodied Question Answering", in CVPR 2019.
- [16] Hao Tan, **Licheng Yu**, Mohit Bansal, "Learning to Navigate Unseen Environments: Back Translation with Environmental Dropout", in NAACL 2019. (Rank 1 on VLN Leaderboard).
- [15] Jie Lei, **Licheng Yu**, Mohit Bansal, Tamara L. Berg, "TVQA: Localized Compositional Video Question Answering", in EMNLP 2018 (oral).
- [14] **Licheng Yu**, Zhe Lin, Xiaohui Shen, Jimei Yang, Xin Lu, Mohit Bansal, Tamara L. Berg, "MAttNet: Modular Attention Network for Referring Expression Comprehension", in CVPR 2018.
- [13] Anja Belz, TL Berg, **Licheng Yu**, "From image to language and back again", in JNLE 2018.
- [12] **Licheng Yu**, Mohit Bansal, Tamara L. Berg, "Hierarchically-Attentive RNN for Album Summarization and Storytelling", in EMNLP 2017.
- [11] **Licheng Yu**, Hao Tan, Mohit Bansal, Tamara L. Berg, "A Joint Speaker-Listener-Reinforcer Model for referring expressions.", in CVPR 2017 (spotlight).
- [10] Hongteng Xu, **Licheng Yu**, Mark Davenport, Hongyuan Zha, "A unified framework for manifold landmarking", in IEEE Transactions on Signal Processing, 2018.

- [9] **Licheng Yu**, Patrick Poirson, Shan Yang, Alexander C. Berg, Tamara L. Berg, “Modeling Context in Referring Expressions”, in ECCV, 2016 (Spotlight).
- [8] Shan Yang, Zherong Pan, Tanya Ambert, Ke Wang, **Licheng Yu**, Tamara L. Berg, Ming C. Lin, “Detailed Garment Recovery from a Single-View Image”, in ACM Transactions on Graphics 2017.
- [7] **Licheng Yu**, Eunbyung Park, Alexander C. Berg, Tamara L. Berg, “Visual Madlibs: Fill-in-the-Blank Description Generation and Question Answering”, in ICCV, 2015.
- [6] Yi Xu, **Licheng Yu**, Hongteng Xu, Truong Nguyen, “Vector Sparse Representation of Color Image Using Quaternion Matrix Analysis.” in IEEE Transactions on Image Processing (TIP), 2015.
- [5] **Licheng Yu\***, Hongteng Xu\*, Hongyuan Zha, Yi Xu. “Dictionary Learning with Mutually Reinforcing Group-Graph Structures.” in AAAI, 2015 (\* denotes equal contribution).
- [4] **Licheng Yu**, Yi Xu, Hongteng Xu, Hao Zhang, “Quaternion-based Sparse Representation of Color Image.” in ICME, 2013 (Oral).
- [3] **Licheng Yu**, Yi Xu, Hongteng Xu, “Self-Example Based Super-resolution with Fractal-based Gradient Enhancement.” in ICME workshop, 2013.
- [2] **Licheng Yu**, Yi Xu, Bo Zhang, “Single Image Super-resolution via Phase Congruency Analysis.” in VCIP, 2013 (Oral).
- [1] **Licheng Yu**, Hongteng Xu, Yi Xu, Xiaokang Yang, “Robust Single Image Super-resolution based on Gradient Enhancement”, in APSIPA 2012.

## INTERN EXPERIENCE

### Facebook AI

Research Intern

Menlo Park, CA

2018.05 — 2018.08

- Embodied Question Answering - robot navigation in an unseen environment answering a given question.

### Adobe Research

Computer Vision Research Intern

San Jose, CA

2017.05 — 2017.08

- Referring Expression Comprehension - localize an object described by a sentence.
- Project page: <http://vision2.cs.unc.edu/refer>

### eBay Research

Computer Vision Research Intern

San Jose, CA

2016.05 — 2016.08

- Product attribute prediction and localization.
- US Patent 17,165,481: Visual aspect localization presentation
- US Patent 11,200,273: Parallel prediction of multiple image aspects

## ACTIVITIES

Reviewer of CVPR, ICCV, ECCV, ACL, EMNLP, NAACL, TPAMI	2014 – now
Organizer of VALUE Challenge, ICCV 2021	2021.10
Organizer of LVVU Workshop, CVPR 2020	2020.06
Organizer of Tutorial - Recent Advances in Vision-and-Language Research, CVPR 2020	2020.06
Winner of VQA 2020 Challenge	2020.06
Spotlight Presentation at CVPR 2017	2017.07
Spotlight Presentation at ECCV 2016	2016.09

## SKILLS

<b>Tools and Languages</b>	Python, PHP, Java, Javascript, C++/C, Lua, Git, $\LaTeX$ , Markdown
<b>Framework</b>	PyTorch, TensorFlow, Torch7, Caffe
<b>Communication</b>	Chinese, English