

# 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)    🌐 [Licheng Yu](https://www.google.com/search?q=Licheng+Yu)    📄 [lichengunc](https://github.com/lichengunc)

I am now a Senior Research Scientist at Facebook AI. My research interest lies in the intersection of 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 the questions, conduct 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 Electrical and Computer Engineering</b> , <i>Georgia Institute of Technology</i>	2011.09 — 2014.04
<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>Senior Research Scientist</b> <i>Facebook AI</i>	<b>2020.03 — Present</b> <i>Menlo Park, CA</i>
<ul style="list-style-type: none"> <li>Vision+Language Research, including unsupervised multimodal pre-training, trace-guided image captioning, etc.</li> <li>Multimodal Applications on Commerce and Ads Ranking.</li> </ul>	
<b>Researcher</b> <i>Microsoft Dynamics 365 AI Research</i>	<b>2019.06 — 2020.03</b> <i>Bellevue, WA</i>
<ul style="list-style-type: none"> <li>Vision+Language Research, including Image-Text Pre-training, Video-Text Pre-training, Image Synthesis, etc.</li> </ul>	

## PUBLICATION

- [29] **Licheng Yu**, Jun Chen, Animesh Sinha, Mengjiao MJ Wang, Yu Chen, Tamara L. Berg, Ning Zhang, "CommerceMM: Large-Scale Commerce MultiModal Representation Learning with Omni Retrieval", in arXiv:2202.07247v1.
- [28] 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. (\* denotes equal contribution).
- [27] Linjie Li, Jie Lei, Zhe Gan, **Licheng Yu**, Yen-Chun Chen, Rohit Pillai, Yu Cheng, Luowei 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, Wenhu 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 Special Sessions of APSIPA 2012

## INTERN EXPERIENCE

<b>Research Intern</b> <i>Facebook AI</i>	<b>2018.05 — 2018.08</b> <i>Menlo Park, CA</i>
<ul style="list-style-type: none"> <li>Embodied Question Answering - robot navigation in an unseen environment answering a given question.</li> </ul>	
<b>Computer Vision Research Intern</b> <i>Adobe Research</i>	<b>2017.05 — 2017.08</b> <i>San Jose, CA</i>
<ul style="list-style-type: none"> <li>Referring Expression Comprehension - Localize an object described by a sentence.</li> <li>Project page: <a href="http://vision2.cs.unc.edu/refer">http://vision2.cs.unc.edu/refer</a></li> </ul>	
<b>Computer Vision Research Intern</b> <i>eBay Research</i>	<b>2016.05 — 2016.08</b> <i>San Jose, CA</i>
<ul style="list-style-type: none"> <li>Product attribute prediction and localization.</li> <li>US Patent 17/165,481: Visual aspect localization presentation</li> <li>US Patent 11,200,273: Parallel prediction of multiple image aspects</li> </ul>	

## 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, $\text{\LaTeX}$ , Markdown
<b>Framework</b>	PyTorch, TensorFlow, Torch7, Caffe
<b>Communication</b>	Chinese, English