

# JIE LEI

Email: jayleicn@gmail.com  
Phone: 919-564-8651

Updated July 2024  
Website: <https://jayleicn.github.io/>  
Github: <https://github.com/jayleicn>

## WORK EXPERIENCE

---

**Meta AI** June 2022 - present  
*Research Scientist* Bellevue, WA  
· Work on vision and language research and products.

## EDUCATION

---

**University of North Carolina at Chapel Hill** Aug 2017 - May 2022  
*Ph.D.* in Computer Science  
Advisors: Tamara L. Berg and Mohit Bansal

**University of Electronic Science and Technology of China (UESTC)** Sep 2013 - Jun 2017  
*B.Eng* in Computer Science & Technology, Yingcai Honors College  
Advisor: Sinno Jialin Pan (NTU, Singapore)

## INTERN EXPERIENCE

---

**Facebook AI** May 2021 - Dec 2021  
*Research Scientist Intern with Licheng Yu, Xinlei Chen and Ning Zhang* Menlo Park, CA  
· Vision and language, especially image retrieval.

**Microsoft Cloud & AI** May 2020 - Aug 2020  
*Researcher Intern with Linjie Li, Luowei Zhou, Zhe Gan and Jingjing Liu* Bellevue, WA  
· End-to-end vision and language pretraining.

**Tencent AI Lab** May 2019 - Aug 2019  
*Research Intern with Liwei Wang, Yelong Shen and Dong Yu* Bellevue, WA  
· Recurrent transformer for coherent video paragraph captioning.

**Nanyang Technological University** Oct 2016 - Apr 2017  
*Research Intern with Sinno Jialin Pan* Singapore  
· Transfer learning in deep hash models for fast image retrieval.

**University of Manitoba** Jul 2016 - Oct 2016  
*Research Intern with Yang Wang* Winnipeg, Canada  
· Weakly supervised image classification using hierarchical image labels.

## PUBLICATIONS

---

- Check out my google scholar profile for the latest list of publications.
- [25] Zichun Yu, Fei Peng, **Jie Lei**, Arnold Overwijk, Wen-tau Yih, Chenyan Xiong, “Data-Efficient Pretraining with Group-Level Data Influence Modeling”, arXiv 2025
- [24] Yuanhao Xiong, Yixin Nie, Haotian Liu, Boxin Wang, Jun Chen, Rong Jin, Cho-Jui Hsieh, Lorenzo Torresani, **Jie Lei**, “UNICORN: A Unified Causal Video-Oriented Language-Modeling Framework for Temporal Video-Language Tasks”, EMNLP 2024
- [23] Raghuveer Thirukovalluru, Xiaolan Wang, Jun Chen, Shuyang Li, **Jie Lei**, Rong Jin, Bhuwan Dhingra, “SumCSE: Summary as a transformation for Contrastive Learning”, NAACL 2024 Findings
- [22] **Jie Lei**, Tamara L. Berg, Mohit Bansal, “Revealing Single Frame Bias for Video-and-Language Learning”, ACL 2023

- [21] Feng Cheng, Xizi Wang, **Jie Lei**, David Crandall, Mohit Bansal, Gedas Bertasius, “VindLU: A Recipe for Effective Video-and-Language Pretraining”, CVPR 2023
- [20] Yan-Bo Lin, Yi-Lin Sung, **Jie Lei**, Mohit Bansal, Gedas Bertasius, “Vision Transformers are Parameter-Efficient Audio-Visual Learners”, CVPR 2023
- [19] Zineng Tang\*, Jaemin Cho\*, **Jie Lei**, Mohit Bansal, “PERCEIVER-VL: Efficient Vision-and-Language Modeling with Iterative Latent Attention”, WACV 2023
- [18] Zhenhailong Wang, Manling Li, Ruochen Xu, Luowei Zhou, **Jie Lei**, Xudong Lin, Shuohang Wang, Ziyi Yang, Chenguang Zhu, Derek Hoiem, Shih-Fu Chang, Mohit Bansal, Heng Ji, “Language Models with Image Descriptors are Strong Few-Shot Video-Language Learners”, NeurIPS 2022
- [17] **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”, arXiv 2022
- [16] Yan-Bo Lin, **Jie Lei**, Mohit Bansal, Gedas Bertasius, “ECLIPSE: Efficient Long-range Video Retrieval using Sight and Sound”, ECCV 2022 **Oral**
- [15] Xinya Du, Zixuan Zhang, ..., **Jie Lei**, ..., Shih-Fu Chang, Martha Palmer, Heng Ji, “Resin-11: Schema-guided event prediction for 11 newsworthy scenarios”, NAACL 2022 System Demo
- [14] Hao Tan\*, **Jie Lei\***, Thomas Wolf, Mohit Bansal, “VIMPAC: Video Pre-Training via Masked Token Prediction and Contrastive Learning”, CVPRW 2022
- [13] **Jie Lei**, Tamara L. Berg, Mohit Bansal, “QVHighlights: Detecting Moments and Highlights in Videos via Natural Language Queries”, NeurIPS 2021
- [12] 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”, NeurIPS 2021 Datasets and Benchmarks Track
- [11] Linjie Li, **Jie Lei**, Zhe Gan, Jingjing Liu, “Adversarial VQA: A New Benchmark for Evaluating the Robustness of VQA Models”, ICCV 2021 **Oral**
- [10] **Jie Lei**, Tamara L. Berg, Mohit Bansal, “mTVR: Multilingual Moment Retrieval in Videos”, ACL 2021
- [9] Jaemin Cho, **Jie Lei**, Hao Tan, Mohit Bansal, “Unifying Vision-and-Language Tasks via Text Generation”, ICML 2021
- [8] Zineng Tang\*, **Jie Lei\***, Mohit Bansal, “Improved Pre-Training from Noisy Instructional Videos via Dense Captions and Entropy Minimization”, NAACL 2021
- [7] **Jie Lei\***, Linjie Li\*, Luowei Zhou, Zhe Gan, Tamara L. Berg, Mohit Bansal, Jingjing Liu, “Less is More: ClipBERT for Video-and-Language Learning via Sparse Sampling”, CVPR 2021 **Best Student Paper Honorable Mention, Oral**
- [6] **Jie Lei**, Licheng Yu, Tamara L. Berg, Mohit Bansal, “What is More Likely to Happen Next? Video and Language Future Event Prediction”, EMNLP 2020
- [5] **Jie Lei**, Licheng Yu, Tamara L. Berg, Mohit Bansal, “TVR: A Large-Scale Dataset for Video-Subtitle Moment Retrieval”, ECCV 2020
- [4] **Jie Lei**, Liwei Wang, Yelong Shen, Dong Yu, Tamara L. Berg, Mohit Bansal, “MART: Memory-Augmented Recurrent Transformer for Coherent Video Paragraph Captioning”, ACL 2020
- [3] **Jie Lei**, Licheng Yu, Tamara L. Berg, Mohit Bansal, “TVQA+: Spatio-Temporal Grounding for Video Question Answering”, ACL 2020.
- [2] **Jie Lei**, Licheng Yu, Mohit Bansal, Tamara L. Berg, “TVQA: Localized, Compositional Video Question Answering”, EMNLP 2018 **Oral**
- [1] **Jie Lei**, Zhenyu Guo, Yang Wang. “Weakly Supervised Image Classification with Coarse and Fine Labels.” CRV 2017

## PROJECTS

---

- [1] **Jie Lei**, “AnimeGAN: Create Anime Face using Generative Adversarial Networks”, GitHub 2017.
- Implementation of GANs for anime face generation. 📄 Data & Code (**1K+ stars**)
  - Automatic pipeline for large-scale anime face dataset construction from the web.

## AWARDS

---

<b>Outstanding Reviewer</b> , CVPR 2022, New Orleans	2022
<b>Best Student Paper Honorable Mention</b> , CVPR 2021, Virtual	2021
<b>Adobe Research Fellowship</b> , Adobe, USA	2021
<b>Outstanding Undergraduate Thesis</b> , UESTC, China	2017
<b>Globalink Research Intern Award</b> (200 undergraduates across China), Mitacs, Canada.	2016

## ACADEMIC SERVICE

---

**Area Chair:** COLING 2022, EMNLP 2023

**Senior Program Committee:** IJCAI 2023

**Organizer:**

- Challenge: VALUE @ ICCV 2021
- Workshop: Transformer for Vision @ CVPR 2022
- Tutorial: Knowledge-Driven Vision-Language Pretraining @ AAAI 2023
- Tutorial: Knowledge-Driven Vision-Language Encoding @ CVPR 2023

**Journal and Conference Reviewer:** TPAMI, IJCV, CVPR 2019-2022, ICCV 2019, ECCV 2020, ACM MM 2020 Interactive Arts, AAAI 2021, ICCV 2021, ACL 2021-2022, NeurIPS 2021-2022, ICLR 2021, NeurIPS HAMILTONS 2020, EMNLP 2022, EMNLP 2024, COLM 2024