Jiuxiang (Joshua) Gu

+1-240-615-6794 | gu.jiuxiang@gmail.com | gujiuxiang.com

gujiuxiang | gujiuxiang | Google Scholar (Cited by **12,369**)

Senior Research Scientist ◇ Adobe Research, Seattle, WA 98103

Research Interests & Summary

Research centers on **multimodal learning** across **1D (text)**, **2D (image/document)**, and **3D (video/scene)** modalities, covering both the **theoretical foundations** of deep learning (e.g., generalization, robustness, and alignment) and the **design of scalable, efficient foundation models**. Contributions include **architecture innovations**, **self-supervised objectives**, and **model compression and acceleration** for diffusion models and LLMs. These systems have been translated into real-world impact, powering applications such as Adobe Firefly, Acrobat AI Assistant, and the UCSF-JHU Opioid Industry Documents Archive. Publications span premier venues in machine learning (NeurIPS, ICML, ICLR, etc.), computer vision (CVPR, ICCV, ECCV, etc.), and natural language processing (ACL, etc.). Recognized multiple times among Stanford University's *Top 2% Most Cited Scientists*.

Education

- **Nanyang Technological University** *Jan 2016 - Oct 2019*
  Ph.D. in Artificial Intelligence, advised by Prof. Jianfei Cai, Prof. Tsuhan Chen, and Prof. Gang Wang Singapore
  - Authored **7** first-author papers in top-tier vision and AI venues (e.g., *ICCV*, *CVPR*, *ECCV*, *AAAI*, *Pattern Recognition*), with a thesis on *Bridging Images and Natural Language with Deep Learning*.

- **University of Chinese Academy of Sciences** *Sep 2010 - Jun 2013*
  Master of Electronic and Communications Engineering Beijing, China
  - Researched and developed a real-time SAR (HJ-1C) imaging system, combining FPGA-based PCIe data acquisition with GPU-accelerated parallel processing. *Focused on high-performance computing for accelerated SAR image reconstruction*

- **Jiangsu University** *Sep 2006 - Jun 2010*
  Bachelor of Electronic Information Science and Technology (Rank 1^{st}) Zhenjiang, China

Experience

- **Adobe Research** *Feb 2020 – Now*
  Senior Research Scientist Seattle, United States
  - Led research on multimodal intelligence and theoretical ML, advancing text (1D), image and document (2D), and video and 3D (temporal/spatial) generation, as well as visual reasoning and trustworthy AI
  - Led development of state-of-the-art LLMs and multimodal LLMs powering core capabilities in Adobe Firefly research and the Acrobat AI Assistant, enabling generative intelligence across creative and document workflows

- **Adobe Research** *Aug 2018 – Nov 2018*
  Research Intern San Jose, United States
  - Research on multimodal representation learning and structured visual semantics

- **Nanyang Technological University** *Feb 2015 – Jan 2016*
  Research Associate & SoC-AI System Lead Singapore
  - Led end-to-end development of real-time embedded AI systemsfrom algorithm design and dataset curation to SoC implementation, software stack (server, Android), and functional product prototyping

- **Chinese Academy of Sciences** *Jun 2013 – Feb 2015*
  Technical Lead, SoC Architecture and Verification Beijing, China
  - Led the design and verification of a 20M-gate SoC platform (Loongson), integrating Verilog logic, hardware/software co-simulation, and FPGA-based prototyping for first-silicon success

Professional Service

- **Area Chair:** ACL 25, ICLR 25, WACV 24/25
- **Program Committee Member/Reviewer:** AAAI 20–25, IJCAI 21–24, NAACL 21, ICLR 20–25, NeurIPS 20–25, CVPR 18–25, ECCV 18-20, ICCV 19-25
- **Journal Reviewer:** T-PAMI, IJCV, JMLR, T-IP, T-MM, T-NNLS, T-CSVT, T-SMC-B, PR, PRL, etc.

PATENTS AND PUBLICATIONS

† CORRESPONDING AUTHOR, ‡ PRIMARY MENTOR; C=CONF., J=JRNL., S=SUBMISSION, T=THESIS.

[C.1] Zhang, Jianyi and Zhou, Yufan and **Gu, Jiuxiang** and Wigington, Curtis and Yu, Tong and Chen, Yiran and Sun, Tong and Zhang, Ruiyi. **Artist: Improving the generation of text-rich images by disentanglement**. In **WACV**. (2025).

[C.2] **Gu, Jiuxiang**† and Liang, Yingyu and Sha, Zhizhou and Shi, Zhenmei and Song, Zhao. **Differential privacy mechanisms in neural tangent kernel regression**. In **WACV**. (2025).

[C.3] Owens, Deonna M and Rossi, Ryan A and Kim, Sungchul and Yu, Tong and Dernoncourt, Franck and Chen, Xiang and Zhang, Ruiyi and **Gu, Jiuxiang** and Deilamsalehy, Hanieh and Lipka, Nedim. **A multi-llm debiasing framework**. In Submission. (2025).

[C.4] Li, Xiang and Qiu, Kai and Chen, Hao and Kuen, Jason and **Gu, Jiuxiang** and Raj, Bhiksha and Lin, Zhe. **Imagefolder: Autoregressive image generation with folded tokens**. In **ICLR**. (2025).

[C.5] Chen, Jian and Zhang, Ruiyi and Zhou, Yufan and Yu, Tong and Dernoncourt, Franck and **Gu, Jiuxiang** and Rossi, Ryan A and Chen, Changyou and Sun, Tong. **LoRA-Contextualizing Adaptation of Large Multimodal Models for Long Document Understanding**. In **ICLR**. (2025).

[C.6] Shen, Xuan and Song, Zhao and Zhou, Yufa and Chen, Bo and Liu, Jing and Zhang, Ruiyi and Rossi, Ryan A. and Tan, Hao and Yu, Tong and Chen, Xiang and Zhou, Yufan and Sun, Tong and Zhao, Pu and Wang, Yanzhi and **Gu, Jiuxiang**†. **Numerical pruning for efficient autoregressive models**. In **AAAI**. (2025).

[C.7] Shen, Xuan and Song, Zhao and Zhou, Yufa and Chen, Bo and Li, Yanyu and Gong, Yifan and Zhang, Kai and Tan, Hao and Kuen, Jason and Ding, Henghui and Shu, Zhihao and Niu, Wei and Zhao, Pu and Wang, Yanzhi and **Gu, Jiuxiang**†. **LazyDiT: Lazy Learning for the Acceleration of Diffusion Transformers**. In **AAAI**. (2025).

[C.8] Jiang, Hanwen and Xu, Zexiang and Xie, Desai and Chen, Ziwen and Jin, Haian and Luan, Fujun and Shu, Zhixin and Zhang, Kai and Bi, Sai and Sun, Xin and **Gu, Jiuxiang** and Huang, Qixing and Pavlakos, Georgios and Tan, Hao. **MegaSynth: Scaling Up 3D Scene Reconstruction with Synthesized Data**. In **CVPR**. (2025).

[C.9] Chen, Jian and Zhang, Ruiyi and Zhou, Yufan and Yu, Tong and Dernoncourt, Franck and **Gu, Jiuxiang** and Rossi, Ryan A and Chen, Changyou and Sun, Tong. **LoRA-Contextualizing Adaptation of Large Multimodal Models for Long Document Understanding**. In **ICLR**. (2025).

[C.10] Xia, Yu and Mukherjee, Subhojyoti and Xie, Zhouhang and Wu, Junda and Li, Xintong and Aponte, Ryan and Lyu, Hanjia and Barrow, Joe and Chen, Hongjie and Dernoncourt, Franck and others. **From Selection to Generation: A Survey of LLM-based Active Learning**. In **ACL**. (2025).

[C.11] Shen, Xuan and Wang, Yizhou and Shi, Xiangxi and Wang, Yanzhi and Zhao, Pu and **Gu, Jiuxiang**†. **Efficient Reasoning with Hidden Thinking**. In Submission. (2025).

[C.12] Shen, Xuan and Ma, Weize and Zhou, Yufa and Tang, Enhao and Xie, Yanyue and Li, Zhengang and Gong, Yifan and Wang, Quanyi and Ding, Henghui and Wang, Yiwei and Wang, Yanzhi and Zhao, Pu and Lin, Jun and **Gu, Jiuxiang**†. **FastCar: Cache Attentive Replay for Fast Auto-Regressive Video Generation on the Edge**. In Submission. (2025).

[C.13] **Gu, Jiuxiang**† and Shi, Jing and Zhu, Wanrong and Jain, Rajiv and Wigington, Curtis and Kuen, Jason and Zhang, Ruiyi and Shen, Xuan and Sun, Tong. **ADOPT: A Multimodal Framework for Document Understanding and Generation**. In Submission. (2025).

[C.14] Shen, Xuan and Han, Chenxia and Zhou, Yufa and Xie, Yanyue and Gong, Yifan and Wang, Quanyi and Wang, Yiwei and Wang, Yanzhi and Zhao, Pu and **Gu, Jiuxiang**†. **DraftAttention: Fast Video Diffusion via Low-Resolution Attention Guidance**. In Submission. (2025).

[C.15] Zhao, Haozhe and Cai, Zefan and Si, Shuzheng and Chen, Liang and **Gu, Jiuxiang** and Xiao, Wen and Hu, Junjie. **MENTOR: Efficient Multimodal-Conditioned Tuning for Autoregressive Vision Generation Models**. In Submission. (2025).

[C.16] Cai, Zefan and Qiu, Haoyi and Zhao, Haozhe and Wan, Ke and Li, Jiachen and **Gu, Jiuxiang** and Xiao, Wen and Peng, Nanyun and Hu, Junjie. **From Preferences to Prejudice: The Role of Alignment Tuning in Shaping Social Bias in Video Diffusion Models**. In Submission. (2025).

[C.17] Cai, Zefan and Xiao, Wen and Sun, Hanshi and Luo, Cheng and Zhang, Yikai and Wan, Ke and Li, Yucheng and Zhou, Yeyang and Chang, Li-Wen and **Gu, Jiuxiang** and Dong, Zhen and Anandkumar, Anima and Asi, Abedelkadir and Hu, Junjie. **R-KV: Redundancy-aware KV Cache Compression for Reasoning Models**. In Submission. (2025).

[C.18] Shen, Xuan and Wingenroth, Brian and Wang, Zichao and Kuen, Jason and Zhu, Wanrong and Zhang, Ruiyi and Wang, Yiwei and Ma, Lichun and Liu, Anqi and Liu, Hongfu and Sun, Tong and Hawkins, Kevin S. and Tasker, Kate and Alexander, G. Caleb and **Gu, Jiuxiang**†. **OIDA-QA: A Multimodal Benchmark for Analyzing the Opioid Industry Documents Archive**. In Submission. (2025).

[C.19] Zhu, Wanrong and Shi, Xiangxi and Zhou, Yufan and Zhang, Ruiyi and Sun, Tong and **Gu, Jiuxiang**†. **ADOPD-Instruct: A Large-Scale Multimodal Dataset for Document Editing**. In Submission. (2025).

[C.20] Zhang, Jianyi and Zhou, Yufan and **Gu, Jiuxiang** and Wigington, Curtis and Yu, Tong and Chen, Yiran and Sun, Tong and Zhang, Ruiyi. **ARTIST: Improving the Generation of Text-rich Images with Disentangled Diffusion Models**. In **WACV**. (2025).

[C.21] Li, Bingxuan and Wang, Yiwei and **Gu, Jiuxiang** and Chang, Kai-Wei and Peng, Nanyun. **Metal: A multi-agent framework for chart generation with test-time scaling**. In **ACL**. (2025).

[C.22] Qiu, Kai and Li, Xiang and Kuen, Jason and Chen, Hao and Xu, Xiaohao and **Gu, Jiuxiang** and Luo, Yinyi and Raj, Bhiksha and Lin, Zhe and Savvides, Marios. **Robust Latent Matters: Boosting Image Generation with Sampling Error Synthesis**. In Submission. (2025).

[C.23] Shen, Xuan and Ma, Weize and Liu, Jing and Yang, Changdi and Ding, Rui and Wang, Quanyi and Ding, Henghui and Niu, Wei and Wang, Yanzhi and Zhao, Pu and Lin, Jun and **Gu, Jiuxiang**†. **QuartDepth: Post-Training Quantization for Real-Time Depth Estimation on the Edge**. In **CVPR**. (2025).

[C.24] Li, Ming and Zhang, Ruiyi and Chen, Jian and **Gu, Jiuxiang** and Zhou, Yufan and Dernoncourt, Franck and Zhu, Wanrong and Zhou, Tianyi and Sun, Tong. **Towards Visual Text Grounding of Multimodal Large Language Model**. In Submission. (2025).

[C.25] Chu, Zhendong and Zhang, Ruiyi and Yu, Tong and Jain, Rajiv and Morariu, Vlad I and **Gu, Jiuxiang** and Nenkova, Ani. **Improving a Named Entity Recognizer Trained on Noisy Data with a Few Clean Instances**. In **NAACL**. (2024).

[C.26] Hong, Yicong and Zhang, Kai and **Gu, Jiuxiang** and Bi, Sai and Zhou, Yang and Liu, Difan and Liu, Feng and Sunkavalli, Kalyan and Bui, Trung and Tan, Hao. **Lrm: Large reconstruction model for single image to 3d**. In **ICLR**. Oral. (2024).

[C.27] Zhou, Yufan and Zhang, Ruiyi and **Gu, Jiuxiang** and Sun, Tong. **Customization assistant for text-to-image generation**. In **CVPR**. (2024).

[C.28] Li, Ming and Chen, Lichang and Chen, Jiuhai and He, Shwai and **Gu, Jiuxiang** and Zhou, Tianyi. **Selective reflection-tuning: Student-selected data recycling for llm instruction-tuning**. In **ACL**. (2024).

[C.29] **Gu, Jiuxiang**[†] and Shi, Xiangxi and Kuen, Jason and Qi, Lu and Zhang, Ruiyi and Liu, Anqi and Nenkova, Ani and Sun, Tong. **ADoPD: A large-scale document page decomposition dataset**. In **ICLR**. (2024).

[C.30] Cao, Shengcao and **Gu, Jiuxiang**[‡] and Kuen, Jason and Tan, Hao and Zhang, Ruiyi and Zhao, Handong and Nenkova, Ani and Gui, Liang-Yan and Sun, Tong and Wang, Yu-Xiong. **SOHES: Self-supervised open-world hierarchical entity segmentation**. In **ICLR**. (2024).

[C.31] Mathur, Puneet and Morariu, Vlad I and Garimella, Aparna and Dernoncourt, Franck and **Gu, Jiuxiang** and Sawhney, Ramit and Nakov, Preslav and Manocha, Dinesh and Jain, Rajiv. **DocScript: Document-Level Script Event Prediction**. In **COLING**. (2024).

[C.32] Zhang, Ruiyi and Zhang, Yanzhe and Chen, Jian and Zhou, Yufan and **Gu, Jiuxiang** and Chen, Changyou and Sun, Tong. **Trins: Towards multimodal language models that can read**. In **CVPR**. (2024).

[C.33] Biswas, Sanket and Jain, Rajiv and Morariu, Vlad I and **Gu, Jiuxiang** and Mathur, Puneet and Wigington, Curtis and Sun, Tong and Lladós, Josep. **DocSynthv2: A Practical Autoregressive Modeling for Document Generation**. In **CVPR**. (2024).

[C.34] Chu, Zhendong and Zhang, Ruiyi and Yu, Tong and Jain, Rajiv and Morariu, Vlad and **Gu, Jiuxiang** and Nenkova, Ani. **Self-Cleaning: Improving a Named Entity Recognizer Trained on Noisy Data with a Few Clean Instances**. In **NAACL**. (2024).

[C.35] Xiao, Wenxiao and **Gu, Jiuxiang**[‡] and Liu, Hongfu. **Category-aware active domain adaptation**. In **ICML**. (2024).

[C.36] Chen, Jian and Zhang, Ruiyi and Zhou, Yufan and Healey, Jennifer and **Gu, Jiuxiang** and Xu, Zhiqiang and Chen, Changyou. **TextLap: Customizing Language Models for Text-to-Layout Planning**. In **EMNLP**. (2024).

[C.37] Bai, Yue and Zhao, Handong and Lin, Zhe and Kale, Ajinkya and **Gu, Jiuxiang** and Yu, Tong and Kim, Sungchul and Fu, Yun. **Advancing Vision-Language Models with Adapter Ensemble Strategies**. In **EMNLP**. (2024).

[C.38] Qi, Lu and Kuen, Jason and Guo, Weidong and Shen, Tiancheng and **Gu, Jiuxiang** and Jia, Jiaya and Lin, Zhe and Yang, Ming-Hsuan. **High-Quality Entity Segmentation**. In **ICCV**. (2023).

[C.39] Mathur, Puneet and Jain, Rajiv and Mehra, Ashutosh and **Gu, Jiuxiang** and Dernoncourt, Franck and Tran, Quan and Kaynig-Fittkau, Verena and Nenkova, Ani and Manocha, Dinesh and Morariu, Vlad I and others. **LayerDoc: layer-wise extraction of spatial hierarchical structure in visually-rich documents**. In **WACV**. (2023).

[C.40] Verma, Gaurav and Rossi, Ryan A and Tensmeyer, Christopher and **Gu, Jiuxiang** and Nenkova, Ani. **Learning the visualness of text using large vision-language models**. In **ACL**. (2023).

[C.41] Qi, Lu and Kuen, Jason and Guo, Weidong and **Gu, Jiuxiang** and Lin, Zhe and Du, Bo and Xu, Yu and Yang, Ming-Hsuan. **AIMS: all-inclusive multi-level segmentation for anything**. In **NeurIPS**. (2023).

[C.42] Mathur, Puneet and Jain, Rajiv and **Gu, Jiuxiang** and Dernoncourt, Franck and Manocha, Dinesh and Morariu, Vlad I. **Docedit: language-guided document editing**. In **AAAI**. (2023).

[C.43] Zhang, Yanzhe and Zhang, Ruiyi and **Gu, Jiuxiang** and Zhou, Yufan and Lipka, Nedim and Yang, Diyi and Sun, Tong. **Llavar: Enhanced visual instruction tuning for text-rich image understanding**. In **NeurIPS** Workshop. (2023).

[C.44] **Gu, Jiuxiang**[†] and Ming, Yifei and Zhou, Yi and Kuen, Jason and Morariu, Vlad and Zhao, Handong and Zhang, Ruiyi and Barmpalios, Nikolaos and Liu, Anqi and Li, Yixuan and others. **A critical analysis of document out-of-distribution detection**. In **EMNLP**. (2023).

[C.45] Li, Ming and Chen, Lichang and Chen, Jiuhai and He, Shwai and Huang, Heng and **Gu, Jiuxiang** and Zhou, Tianyi. **Reflection-tuning: Data recycling improves llm instruction-tuning**. In **NeurIPS** Workshop. (2023).

[C.46] Gao, Jiahui and Zhou, Yi and Philip, LH and Joty, Shafiq and **Gu, Jiuxiang**[†]. **UNISON: Unpaired cross-lingual image captioning**. In **AAAI**. (2022).

[C.47] Huynh, Dat and Kuen, Jason and Lin, Zhe and **Gu, Jiuxiang** and Elhamifar, Ehsan. **Open-vocabulary instance segmentation via robust cross-modal pseudo-labeling**. In **CVPR**. (2022).

[C.48] Zhou, Yufan and Zhang, Ruiyi and Chen, Changyou and Li, Chunyuan and Tensmeyer, Chris and Yu, Tong and **Gu, Jiuxiang** and Xu, Jinhui and Sun, Tong. **Towards language-free training for text-to-image generation**. In **CVPR**. (2022).

[C.49] Qi, Lu and Kuen, Jason and Lin, Zhe and **Gu, Jiuxiang** and Rao, Fengyun and Li, Dian and Guo, Weidong and Wen, Zhen and Yang, Ming-Hsuan and Jia, Jiaya. **Ca-ssl: Class-agnostic semi-supervised learning for detection and segmentation**. In **ECCV**. (2022).

[C.50] Lai, Phung and Phan, NhatHai and Sun, Tong and Jain, Rajiv and Dernoncourt, Franck and **Gu, Jiuxiang** and Barmpalios, Nikolaos. **User-entity differential privacy in learning natural language models**. In **Big Data**. (2022).

[C.51] Zhou, Yufan and Zhang, Ruiyi and **Gu, Jiuxiang** and Tensmeyer, Chris and Yu, Tong and Chen, Changyou and Xu, Jinhui and Sun, Tong. **Tigan: Text-based interactive image generation and manipulation**. In **AAAI**. (2022).

[C.52] Wang, Haoyu and Zhao, Handong and Wang, Yaqing and Yu, Tong and **Gu, Jiuxiang** and Gao, Jing. **Fedkc: Federated knowledge composition for multilingual natural language understanding**. In **ACM Web**. (2022).

[C.53] Wang, Zihan and **Gu, Jiuxiang**‡ and Kuen, Jason and Zhao, Handong and Morariu, Vlad and Zhang, Ruiyi and Nenkova, Ani and Sun, Tong and Shang, Jingbo. **Learning adaptive axis attentions in fine-tuning: Beyond fixed sparse attention patterns**. In **ACL**. (2022).

[C.54] Ma, Haoyu and Zhao, Handong and Lin, Zhe and Kale, Ajinkya and Wang, Zhangyang and Yu, Tong and **Gu, Jiuxiang** and Choudhary, Sunav and Xie, Xiaohui. **Ei-clip: Entity-aware interventional contrastive learning for e-commerce cross-modal retrieval**. In **CVPR**. (2022).

[C.55] Mathur, Puneet and Morariu, Vlad and Kaynig-Fittkau, Verena and **Gu, Jiuxiang** and Dernoncourt, Franck and Tran, Quan Hung and Nenkova, Ani and Manocha, Dinesh and Jain, Rajiv. **Doctime: A document-level temporal dependency graph parser**. In **NNACL**. (2022).

[C.56] Xu, Li and Qu, Haoxuan and Kuen, Jason and **Gu, Jiuxiang** and Liu, Jun. **Meta spatio-temporal debiasing for video scene graph generation**. In **ECCV**. (2022).

[C.57] Mathur, Puneet and Dernoncourt, Franck and Tran, Quan Hung and **Gu, Jiuxiang** and Nenkova, Ani and Morariu, Vlad I and Jain, Rajiv and Manocha, Dinesh. **DocLayoutTTS: Dataset and Baselines for Layout-informed Document-level Neural Speech Synthesis.**. In **INTERSPEECH**. (2022).

[C.58] Qu, Haoxuan and Li, Yanchao and Foo, Lin Geng and Kuen, Jason and **Gu, Jiuxiang** and Liu, Jun. **Improving the reliability for confidence estimation**. In **ECCV**. (2022).

[C.59] Ming, Yifei and Cai, Ziyang and **Gu, Jiuxiang** and Sun, Yiyou and Li, Wei and Li, Yixuan. **Delving into out-of-distribution detection with vision-language representations**. In **NeurIPS** (**Code**). (2022).

[C.60] Wang, Zilong and **Gu, Jiuxiang**‡ and Tensmeyer, Chris and Barmpalios, Nikolaos and Nenkova, Ani and Sun, Tong and Shang, Jingbo and Morariu, Vlad I. **MGDoc: Pre-training with multi-granular hierarchy for document image understanding**. In **EMNLP**. (2022).

[C.61] Du, Mengnan and Manjunatha, Varun and Jain, Rajiv and Deshpande, Ruchi and Dernoncourt, Franck and **Gu, Jiuxiang** and Sun, Tong and Hu, Xia. **Towards interpreting and mitigating shortcut learning behavior of NLU models**. In **NAACL**. (2021).

[C.62] Qi, Lu and Kuen, Jason and **Gu, Jiuxiang** and Lin, Zhe and Wang, Yi and Chen, Yukang and Li, Yanwei and Jia, Jiaya. **Multi-scale aligned distillation for low-resolution detection**. In **CVPR**. (2021).

[C.63] Li, Peizhao and **Gu, Jiuxiang**‡ and Kuen, Jason and Morariu, Vlad I and Zhao, Handong and Jain, Rajiv and Manjunatha, Varun and Liu, Hongfu. **Selfdoc: Self-supervised document representation learning**. In **CVPR**. (2021).

[C.64] Yang, Huiyuan and Yin, Lijun and Zhou, Yi and **Gu, Jiuxiang**†. **Exploiting semantic embedding and visual feature for facial action unit detection**. In **CVPR**. (2021).

[C.65] **Gu, Jiuxiang**† and Kuen, Jason and Morariu, Vlad I and Zhao, Handong and Jain, Rajiv and Barmpalios, Nikolaos and Nenkova, Ani and Sun, Tong. **Unidoc: Unified pretraining framework for document understanding**. In **NeurIPS**. (2021).

[C.66] Yao, Shuhan and **Gu, Jiuxiang** and Zhang, Huajun and Wang, Peng and Liu, Xiaochuan and Zhao, Tianyang. **Resilient load restoration in microgrids considering mobile energy storage fleets: A deep reinforcement learning approach**. In **PESGM**. Best Paper. (2020).

[C.67] Shi, Xiangxi and Yang, Xu and **Gu, Jiuxiang** and Joty, Shafiq and Cai, Jianfei. **Finding it at another side: A viewpoint-adapted matching encoder for change captioning**. In **ECCV**. (2020).

[C.68] **Gu, Jiuxiang**† and Kuen, Jason and Joty, Shafiq and Cai, Jianfei and Morariu, Vlad and Zhao, Handong and Sun, Tong. **Self-supervised relationship probing**. In . (2020).

[C.69] **Gu, Jiuxiang**† and Joty, Shafiq and Cai, Jianfei and Zhao, Handong and Yang, Xu and Wang, Gang. **Unpaired Image Captioning via Scene Graph Alignments**. In **ICCV**. (2019).

[C.70] **Gu, Jiuxiang**† and Zhao, Handong and Lin, Zhe and Li, Sheng and Cai, Jianfei and Ling, Mingyang. **Scene graph generation with external knowledge and image reconstruction**. In **CVPR**. (2019).

[C.71] Shi, Xiangxi and Cai, Jianfei and Joty, Shafiq and **Gu, Jiuxiang**†. **Watch It Twice: Video Captioning with a Refocused Video Encoder**. In **ACM MM**. (2019).

[C.72] **Gu, Jiuxiang**† and Cai, Jianfei and Wang, Gang and Chen, Tsuhan. **Stack-Captioning: Coarse-to-Fine Learning for Image Captioning**. In **AAAI**. Oral. (2018).

[C.73] **Gu, Jiuxiang**† and Cai, Jianfei and Joty, Shafiq and Niu, Li and Wang, Gang. **Look, Imagine and Match: Improving Textual-Visual Cross-Modal Retrieval with Generative Models**. In **CVPR**. Spotlight. (2018).

[C.74] **Gu, Jiuxiang**† and Joty, Shafiq and Cai, Jianfei and Wang, Gang. **Unpaired image captioning by language pivoting**. In **ECCV**. (2018).

[C.75] Bastan, Muhammet and Shi, Xiangxi and **Gu, Jiuxiang** and Heng, Zhao and Zhuo, Chen and Sng, Dennis and Kot, Alex C. **NTU ROSE Lab at TRECVID 2018: Ad-hoc Video Search and Video to Text.**. In *TRECVID*. (2018).

[C.76] **Gu, Jiuxiang**† and Wang, Gang and Cai, Jianfei and Chen, Tsuhan. **An empirical study of language cnn for image captioning**. In **ICCV**. (2017).

[C.77] Li, Zhuoling and Qu, Haoxuan and Kuen, Jason and **Gu, Jiuxiang** and Ke, Qiuhong and Liu, Jun and Rahmani, Hossein. **DiffIP: Representation Fingerprints for Robust IP Protection of Diffusion Models**. In Submission. (2024).

[J.1] Zhang, Zhehao and Rossi, Ryan A and Kveton, Branislav and Shao, Yijia and Yang, Diyi and Zamani, Hamed and Dernoncourt, Franck and Barrow, Joe and Yu, Tong and Kim, Sungchul and Zhang, Ruiyi and **Gu, Jiuxiang** and Derr, Tyler and Chen, Hongjie and Wu, Junda and Chen, Xiang and Wang, Zichao and Mitra, Subrata and Lipka, Nedim and Ahmed, Nesreen and Wang, Yu. **Personalization of large language models: A survey**. **TMLR**. (2025).

[J.2] Shen, Xuan and Wang, Yizhou and Shi, Xiangxi and Wang, Yanzhi and Zhao, Pu and **Gu, Jiuxiang**†. **Efficient Reasoning with Hidden Thinking**. In Submission. (2025).

[J.3] **Gu, Jiuxiang**† and Li, Chenyang and Liang, Yingyu and Shi, Zhenmei and Song, Zhao. **Exploring the frontiers of softmax: Provable optimization, applications in diffusion model, and beyond**. *arXiv preprint arXiv:2405.03251*. (2024).

[J.4] Zhou, Yufan and Zhang, Ruiyi and Zheng, Kaizhi and Zhao, Nanxuan and **Gu, Jiuxiang** and Wang, Zichao and Wang, Xin Eric and Sun, Tong. **Toffee: Efficient million-scale dataset construction for subject-driven text-to-image generation**. *arXiv preprint arXiv:2406.09305*. (2024).

[J.5] Zhang, Ruiyi and Zhou, Yufan and Chen, Jian and **Gu, Jiuxiang** and Chen, Changyou and Sun, Tong. **Llava-read: Enhancing reading ability of multimodal language models**. *arXiv preprint arXiv:2407.19185*. (2024).

[J.6] Wu, Junda and Li, Xintong and Yu, Tong and Wang, Yu and Chen, Xiang and **Gu, Jiuxiang** and Yao, Lina and Shang, Jingbo and McAuley, Julian. **Commit: Coordinated instruction tuning for multimodal large language models**. *arXiv preprint arXiv:2407.20454*. (2024).

[J.7] Chen, Jian and Zhang, Ruiyi and Zhou, Yufan and Rossi, Ryan and **Gu, Jiuxiang** and Chen, Changyou. **Mmr: Evaluating reading ability of large multimodal models**. *arXiv preprint arXiv:2408.14594*. (2024).

[J.8] Zhang, Zhehao and Rossi, Ryan and Yu, Tong and Dernoncourt, Franck and Zhang, Ruiyi and **Gu, Jiuxiang** and Kim, Sungchul and Chen, Xiang and Wang, Zichao and Lipka, Nedim. **VipAct: Visual-perception enhancement via specialized vlm agent collaboration and tool-use**. *arXiv preprint arXiv:2410.16400*. (2024).

[J.9] Nguyen, Chien Van and Shen, Xuan and Aponte, Ryan and Xia, Yu and Basu, Samyadeep and Hu, Zhengmian and Chen, Jian and Parmar, Mihir and Kunapuli, Sasidhar and Barrow, Joe and Wu, Junda and Singh, Ashish and Wang, Yu and **Gu, Jiuxiang** and Dernoncourt, Franck and Ahmed, Nesreen K. and Lipka, Nedim and Zhang, Ruiyi and Chen, Xiang and Yu, Tong and Kim, Sungchul and Deilamsalehy, Hanieh and Park, Namyong and Rimer, Mike and Zhang, Zhehao and Yang, Huanrui and Rossi, Ryan A. and Nguyen, Thien Huu. **A survey of small language models**. *arXiv preprint arXiv:2410.20011*. (2024).

[J.10] Li, Xiang and Qiu, Kai and Chen, Hao and Kuen, Jason and **Gu, Jiuxiang** and Wang, Jindong and Lin, Zhe and Raj, Bhiksha. **XQ-GAN: An Open-source Image Tokenization Framework for Autoregressive Generation**. In ArXiv. (2024).

[J.11] Wu, Junda and Lyu, Hanjia and Xia, Yu and Zhang, Zhehao and Barrow, Joe and Kumar, Ishita and Mirtaheri, Mehrnoosh and Chen, Hongjie and Rossi, Ryan A. and Dernoncourt, Franck and Yu, Tong and Zhang, Ruiyi and **Gu, Jiuxiang** and Ahmed, Nesreen K. and Wang, Yu and Chen, Xiang and Deilamsalehy, Hanieh and Park, Namyong and Kim, Sungchul and Yang, Huanrui and Mitra, Subrata and Hu, Zhengmian and Lipka, Nedim and Nguyen, Dang and Zhao, Yue and Luo, Jiebo and McAuley, Julian. **Personalized Multimodal Large Language Models: A Survey**. In Submission. (2024).

[J.12] Zhou, Yufan and Zhang, Ruiyi and **Gu, Jiuxiang** and Zhao, Nanxuan and Shi, Jing and Sun, Tong. **SUGAR: Subject-Driven Video Customization in a Zero-Shot Manner**. In Submission. (2024).

[J.13] Qi, Lu and Kuen, Jason and Wang, Yi and **Gu, Jiuxiang** and Zhao, Hengshuang and Torr, Philip and Lin, Zhe and Jia, Jiaya. **Open world entity segmentation**. **T-PAMI**. (2022).

[J.14] Lai, Phung and Phan, Hai and Xiong, Li and Tran, Khang and Thai, My and Sun, Tong and Dernoncourt, Franck and **Gu, Jiuxiang** and Barmpalios, Nikolaos and Jain, Rajiv. **Bit-aware randomized response for local differential privacy in federated learning**. In Openreview. (2022).

[J.15] **Gu, Jiuxiang**† and Wang, Zhenhua and Kuen, Jason and Ma, Lianyang and Shahroudy, Amir and Shuai, Bing and Liu, Ting and Wang, Xingxing and Wang, Gang and Cai, Jianfei and others. **Recent advances in convolutional neural networks**. **Pattern Recognition**. (2018).

[J.16] Shi, Xiangxi and Cai, Jianfei and **Gu, Jiuxiang** and Joty, Shafiq. **Video Captioning with Boundary-aware Hierarchical Language Decoding and Joint Video Prediction**. **Neurocomputing**. (2018).

[J.17] **Gu, Jiuxiang**† and Yang, Renzhong and Shi, Lu and Wei, Hongwei. **HJ-1C real-time image processing technology based on GPU**. *Journal of University of Chinese Academy of Sciences*. (2014).

[J.18] **Gu, Jiuxiang**† and Yang, Renzhong and Wei, Hongwei. **Research of RS Decoding Technology Based on GPU**. *Microelectronics & Computer*. (2013).

Honors and Awards

- **Stanford Top 2% Scientists List** *2023 – 2025*
  Elsevier & Stanford University
  ◦ Recognized among the **worlds top 2% most-cited** scientists for multiple consecutive years.
- **Research Scholarship** *2016 – 2019*
  Nanyang Technological University (NTU)
  ◦ Awarded full PhD research scholarship.
  ◦ Received AAAI Student Travel Grant in 2018.
- **Academic Excellence and National Competitions** *2007 – 2010*
  Jiangsu University, China
  ◦ **Excellent Graduate** and **Outstanding** Undergraduate Thesis (2010).
  ◦ Merit Student and **1st Class** scholarship (2007); Merit Student and 2nd-class scholarship (2008, 2009).
  ◦ **1st Prize** in China Undergraduate Mathematical Contest in Modeling *(Provincial Level, 2008)*.
  ◦ **1st Prize** in China Undergraduate Electronic Design Contest *(Provincial Level, 2008)*.

Teaching Experience

- **Teaching Assistant of CZ3005: Artificial Intelligence** *Jan 2019 Dec 2019*
  Nanyang Technological University, Singapore
  - Conducted tutorials and lab sessions on AI topics such as search algorithms, knowledge representation, and machine learning.
  - Assisted in preparing course materials, including assignments and examination questions.
  - Held regular consultation hours to support students' understanding of complex AI concepts.
  - Evaluated and graded student submissions, providing constructive feedback to enhance learning outcomes.
- **Teaching Assistant of CZ1012: Introduction to Computing Systems** *Jan 2019 Dec 2019*
  Nanyang Technological University, Singapore
  - Facilitated lab sessions covering fundamental computing concepts, including computer architecture and operating systems.
  - Supported students in understanding system-level programming and debugging techniques.
  - Collaborated with faculty to develop lab exercises that reinforce theoretical knowledge.
  - Assisted in grading assignments and exams, ensuring fair and consistent assessment.

Invited Talks

- **Multimodal Learning for Captioning and Retrieval** *2018*
  Tokyo Institute of Technology
- **Self-Supervised Document Representation Learning** *2022*
  Brandeis University
- **Recent Advances in Intelligent Document Processing** *2024*
  Johns Hopkins University (JHU)
- **Recent Progress in Vision-Language Models for Documents** *2024*
  National Institutes of Health (NIH)
- **AI for Document** *2025*
  University of Queensland

Mentorship

- **Xiangxi Shi**, (ACM MM 2021, ECCV 2022, ICLR 2024; PhD from Oregon State University *2018 – 2025*
- **Jiahui Gao**, (AAAI 2022); PhD from University of Hong Kong *2019 – 2022*
- **Peizhao Li**, (CVPR 2021); now Research Scientist at Google *Summer 2020*
- **Mengnan Du**, (NAACL 2020); now Assistant Professor at NJIT *Summer 2020*
- **Zihan Wang**, (ACL 2022); PhD student at UCSD *Summer 2021*
- **Zilong Wang**, (EMNLP 2023); PhD student at UCSD *Summer 2021*
- **Dat Huynh**, (CVPR 2022); PhD from Northeastern University, now Senior Research Scientist at Meta *Summer 2021*
- **Yifei Ming**, (NeurIPS 2020, EMNLP 2021); PhD from UW-Madison, now Research Scientist at Salesforce *Summer 2022*
- **Shengcao Cao**, (ICLR 2024, CVPR 2025); PhD student at UIUC *Summer 2023*
- **Yicong Gong**, (ICLR 2024); PhD from ANU, now Research Scientist at Adobe Research *Summer 2023*
- **Wenxiao Xiao**, (ICML 2024); PhD student at Brandeis University *Summer 2024*
- **Xuan Shen**, (AAAI 2024 $\times$2, CVPR 2025, 4 on submission); PhD from Northeastern University *2024 – 2025*
- **Xiang Li**, (ICLR 2025, CVPR 2025); PhD from CMU, now Research Scientist at Google DeepMind *Summer 2024*
- **Zefan Cai**, (3 NeurIPS 2025 submissions); PhD from UW-Madison *2025*

References

Prof. Jianfei Cai
Dean, School of Computer Science and Engineering
Monash University
Email: jianfei.cai@monash.edu
Relationship: PhD Advisor

Prof. Tsuhan Chen
Deputy President and Provost
National University of Singapore (NUS)
Email: tsuhan@nus.edu.sg
Relationship: PhD Advisor

Prof. Gang Wang
CEO, CenoBots
Nanyang Technological University (NTU)
Email: gangwang6@gmail.com
Relationship: PhD Advisor

Prof. G. Caleb Alexander
Prof. of Epidemiology & Medicine
Johns Hopkins Bloomberg SPH (JHU)
Email: galexan9@jhmi.edu
Relationship: Academic Collaborator

Prof. Shafiq Joty
Director, Salesforce AI Research
Nanyang Technological University (NTU)
Email: srjoty@ntu.edu.sg
Relationship: Research Mentor

Prof. Hongfu Liu
Assistant Professor
Brandeis University
Email: hongfuliu@brandeis.edu
Relationship: Academic Collaborator