Zhe Gan

Apple Dexter,
333 Dexter Ave N
Seattle, WA 98109

Phone: Provided upon request
Email: zhe.gan@apple.com
Homepage: http://zhegan27.github.io/

Research Interests

I am a Research Scientist and Manager at Apple AI/ML, primarily working on building large-scale vision and multimodal foundation models. Before joining Apple, I was a Principal Researcher at Microsoft Azure AI. My research interest mainly focuses on Vision-Language Multimodal Intelligence. I also have broad interests on other machine learning topics, such as sparse neural networks, adversarial training, and self-supervised visual representation learning.

Education

- Duke University, Durham, NC
  Ph.D., Electrical and Computer Engineering 09/2013 - 03/2018
- Peking University, Beijing, China
  M.S., Electrical Engineering 09/2010 - 07/2013
  B.S., Electrical Engineering 09/2006 - 07/2010

Experience

- **Apple AI/ML**
  Research Scientist and Manager. 12/2023 - now
  Research Scientist. 11/2022 - 12/2023
  Building large-scale vision and multimodal foundation models.
- **Microsoft Cloud and AI**
  Principal Researcher. 09/2021 - 11/2022
  Senior Researcher. 12/2018 - 09/2021
  Researcher. 04/2018 - 12/2018
  Building large-scale general-purpose multimodal foundation models that can be adopted to serve a wide range of Microsoft products, with downstream applications such as image classification, image retrieval, image captioning, object detection, segmentation, *etc*.
- **Information Initiative at Duke (iiD)** 09/2013 - 03/2018
  Research Assistant. Advisor: Prof. Lawrence Carin
  (*i*) Deep Bayesian Learning: developing deep generative models for computer vision and natural language processing applications, including VAE and GAN
  (*ii*) Bayesan Deep Learning: designing stochastic gradient variational inference algorithms and stochastic gradient MCMC methods for scalable Bayesian inference
- **Microsoft Research Redmond** 05/2017 - 08/2017
  Research Intern. Advisor: Xiaodong He, Lihong Li, Ph.D
  (*i*) Visual storytelling that aims to generate a story given an image input
  (*ii*) AttnGAN: text-to-image generation
- **Microsoft Research Redmond** 05/2016 - 08/2016
  Research Intern. Advisor: Xiaodong He, Jianfeng Gao, Li Deng, Ph.D
  (*i*) Image and video captioning
  (*ii*) Deep conflation for business data analytics
- **Adobe Research** 06/2015 - 09/2015
  Data Scientist Intern. Advisor: Hung Bui, Ph.D
  (*i*) Recurrent neural networks (RNN) for text classification and generation
  (*ii*) Variational autoencoder for text modeling

Publications

arXiv preprints

1. E. Amirloo*, J.-P. Fauconnier*, C. Roesmann*, C. Kerl, R. Boney, Y. Qian, Z. Wang, A. Dehghan, Y. Yang, **Z. Gan** and P. Grasch "Understanding Alignment in Multimodal LLMs: A Comprehensive Study", *arXiv preprint arXiv:2407.02477*
2. Y. Qian, H. Ye, J.-P. Fauconnier, P. Grasch, Y. Yang and **Z. Gan** "MIA-Bench: Towards Better Instruction Following Evaluation of Multimodal LLMs", *arXiv preprint arXiv:2407.01509*
3. X. Li, Y. Lu, **Z. Gan**, J. Gao, W. Wang and Y. Choi "Text as Images: Can Multimodal Large Language Models Follow Printed Instructions in Pixels?", *arXiv preprint arXiv:2311.17647*
4. Y. Qian, H. Zhang, Y. Yang and **Z. Gan** "How Easy is It to Fool Your Multimodal LLMs? An Empirical Analysis on Deceptive Prompts", *arXiv preprint arXiv:2402.13220*

2024

1. H. Zhang*, H. You*, P. Dufter, B. Zhang, C. Chen, H.-Y. Chen, T.-J. Fu, W. Y. Wang, S.-F. Chang, **Z. Gan** and Y. Yang "Ferret-v2: An Improved Baseline for Referring and Grounding with Large Language Models", *Conference on Language Modeling* (**COLM**), 2024
2. B. McKinzie*, **Z. Gan***, J.-P. Fauconnier, S. Dodge, B. Zhang, P. Dufter, D. Shah, X. Du, F. Peng, F. Weers, A. Belyi, H. Zhang, K. Singh, D. Kang, A. Jain, H. Hè, M. Schwarzer, T. Gunter, X. Kong, A. Zhang, J. Wang, C. Wang, N. Du, T. Lei, S. Wiseman, G. Yin, M. Lee, Z. Wang, R. Pang, P. Grasch, A. Toshev and Y. Yang "MM1: Methods, Analysis & Insights from Multimodal LLM Pre-training", *European Conf. on Computer Vision* (**ECCV**), 2024
3. K. You, H. Zhang, E. Schoop, F. Weers, A. Swearngin, J. Nichols, Y. Yang and **Z. Gan** "Ferret-UI: Grounded Mobile UI Understanding with Multimodal LLMs", *European Conf. on Computer Vision* (**ECCV**), 2024
4. Z. Lai*, H. Zhang*, B. Zhang, W. Wu, H. Bai, A. Timofeev, X. Du, **Z. Gan**, J. Shan, C.-N. Chuah, Y. Yang and M. Cao "From Scarcity to Efficiency: Improving CLIP Training via Visual-enriched Captions", *European Conf. on Computer Vision* (**ECCV**), 2024
5. J. Wu, J. Wang, Z. Yang, **Z. Gan**, Z. Liu, J. Yuan and L. Wang "GRiT: A Generative Region-to-text Transformer for Object Understanding", *European Conf. on Computer Vision* (**ECCV**), 2024
6. H. You*, H. Zhang*, **Z. Gan**, X. Du, B. Zhang, Z. Wang, L. Cao, S.-F. Chang and Y. Yang "Ferret: Refer and Ground Anything Anywhere at Any Granularity", *Int. Conf. Learning Representations* (**ICLR**), 2024 **Spotlight, Top 5% among all submissions**
7. T.-J. Fu, W. Hu, X. Du, W. Y. Wang, Y. Yang and **Z. Gan** "Guiding Instruction-based Image Editing via Multimodal Large Language Models", *Int. Conf. Learning Representations* (**ICLR**), 2024 **Spotlight, Top 5% among all submissions**
8. A. Jaiswal, **Z. Gan**, X. Du, B. Zhang, Z. Wang and Y. Yang "Compressing LLMs: The Truth is Rarely Pure and Never Simple", *Int. Conf. Learning Representations* (**ICLR**), 2024
9. W. Wu*, A. Timofeev*, C. Chen, B. Zhang, K. Duan, S. Liu, Y. Zheng, J. Shlens, X. Du, **Z. Gan** and Y. Yang "MOFI: Learning Image Representations from Noisy Entity Annotated Images", *Int. Conf. Learning Representations* (**ICLR**), 2024
10. J. Cho, L. Li, Z. Yang, **Z. Gan**, L. Wang and M. Bansal "Diagnostic Benchmark and Iterative Inpainting for Layout-Guided Image Generation", *Computer Vision and Pattern Recognition* (**CVPR**), Workshop on the Evaluation of Generative Foundation Models, 2024

2023

1. C. Li*, **Z. Gan***, Z. Yang*, J. Yang*, L. Li*, L. Wang and J. Gao "Multimodal Foundation Models: From Specialists to General-Purpose Assistants", *Foundations and Trends in Computer Graphics and Vision*, 2023. **A survey book on multimodal foundation models**
2. Y. Zhang, B. McKinzie, **Z. Gan**, V. Shankar and A. Toshev "Pre-trained Language Models Do Not Help Auto-regressive Text-to-Image Generation", *Neural Information Processing Systems* (**NeurIPS**), Workshop on I Can't Believe It's Not Better, 2023

3. Y.-L. Sung, L. Li, K. Lin, **Z. Gan**, M. Bansal and L. Wang "An Empirical Study of Multimodal Model Merging", *Conf. on Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2023
4. X. Zou*, Z.-Y. Dou*, J. Yang*, **Z. Gan**, L. Li, C. Li, X. Dai, H. Behl, J. Wang, L. Yuan, N. Peng, L. Wang, Y. J. Lee and J. Gao "Generalized Decoding for Pixel, Image, and Language", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
5. Z. Yang, J. Wang, **Z. Gan**, L. Li, K. Lin, C. Wu, N. Duan, Z. Liu, C. Liu, M. Zeng and L. Wang "ReCo: Region-Controlled Text-to-Image Generation", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
6. J. Zhou, L. Dong, **Z. Gan**, L. Wang and F. Wei "Non-Contrastive Learning Meets Language-Image Pre-Training", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
7. L. Li, **Z. Gan**, K. Lin, C.-C. Lin, Z. Liu, C. Liu and L. Wang "LAVENDER: Unifying Video-Language Understanding as Masked Language Modeling", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
8. T.-J. Fu*, L. Li*, **Z. Gan**, K. Lin, W. Wang, L. Wang and Z. Liu "An Empirical Study of End-to-End Video-Language Transformers with Masked Visual Modeling", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
9. C. Si, **Z. Gan**, Z. Yang, S. Wang, J. Wang, J. Boyd-Graber and L. Wang "Prompting GPT-3 To Be Reliable", *Int. Conf. Learning Representations* (**ICLR**), 2023
10. B. Wen, Z. Yang, J. Wang, **Z. Gan**, B. Howe and L. Wang "InfoVisDial: An Informative Visual Dialogue Dataset by Bridging Large Multimodal and Language Models", *arXiv preprint arXiv:2312.13503*
11. Z. Zhu*, Y. Wei*, J. Wang, **Z. Gan**, Z. Zhang, L. Wang, G. Hua, L. Wang, Z. Liu and H. Hu "Exploring Discrete Diffusion Models for Image Captioning", *arXiv preprint arXiv:2211.11694*

2022

1. **Z. Gan**, L. Li, C. Li, L. Wang, Z. Liu and J. Gao "Vision-Language Pre-training: Basics, Recent Advances, and Future Trends", *Foundations and Trends in Computer Graphics and Vision*, 2022. **A survey book on vision-language pre-training**
2. J. Wang, Z. Yang, X. Hu, L. Li, K. Lin, **Z. Gan**, Z. Liu, C. Liu and L. Wang "GIT: A Generative Image-to-text Transformer for Vision and Language", *Transactions on Machine Learning Research* (**TMLR**), 2022. **Our new multimodal foundation model that achieves 12 new SOTA on a diverse set of image/video captioning and QA tasks**
3. C. Wu*, J. Liang*, X. Hu, **Z. Gan**, J. Wang, L. Wang, Z. Liu, Y. Fang and N. Duan "NUWA-Infinity: Autoregressive over Autoregressive Generation for Infinite Visual Synthesis", *Neural Information Processing Systems* (**NeurIPS**), 2022
4. Z.-Y. Dou*, A. Kamath*, **Z. Gan***, P. Zhang, J. Wang, L. Li, Z. Liu, C. Liu, Y. LeCun, N. Peng, J. Gao, L. Wang "Coarse-to-Fine Vision-Language Pre-training with Fusion in the Backbone", *Neural Information Processing Systems* (**NeurIPS**), 2022
5. S. Shen*, C. Li*, X. Hu*, Y. Xie, J. Yang, P. Zhang, **Z. Gan**, L. Wang, L. Yuan, C. Liu, K. Keutzer, T. Darrell, A. Rohrbach and J. Gao "K-LITE: Learning Transferable Visual Models with External Knowledge", *Neural Information Processing Systems* (**NeurIPS**), 2022 **Oral**
6. Z. Yang, **Z. Gan**, J. Wang, X. Hu, F. Ahmed, Z. Liu, Y. Lu and L. Wang "UniTAB: Unifying Text and Box Outputs for Grounded Vision-Language Modeling", *European Conf. on Computer Vision* (**ECCV**), 2022 **Oral, Top 2.7% among all submissions**
7. T. Chen, Y. Cheng, **Z. Gan**, J. Wang, L. Wang, J. Liu and Z. Wang "Adversarial Feature Augmentation and Normalization for Visual Recognition", *Transactions on Machine Learning Research* (**TMLR**), 2022
8. Z. Dou, Y. Xu, **Z. Gan**, J. Wang, S. Wang, L. Wang, C. Zhu, P. Zhang, L. Yuan, N. Peng, Z. Liu and M. Zeng "An Empirical Study of Training End-to-End Vision-and-Language Transformers", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022

9. X. Hu, **Z. Gan**, J. Wang, Z. Yang, Z. Liu, Y. Lu and L. Wang "Scaling Up Vision-Language Pre-training for Image Captioning", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022
10. K. Lin*, L. Li*, C.-C. Lin*, F. Ahmed, **Z. Gan**, Z. Liu, Y. Lu and L. Wang "SwinBERT: End-to-End Transformers with Sparse Attention for Video Captioning", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022
11. Z. Fang, J. Wang, X. Hu, L. Liang, **Z. Gan**, L. Wang, Y. Yang and Z. Liu "Injecting Semantic Concepts into End-to-End Image Captioning", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022
12. Z. Yang, **Z. Gan**, J. Wang, X. Hu, Y. Lu, Z. Liu and L. Wang "An Empirical Study of GPT-3 for Few-Shot Knowledge-Based VQA", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2022 **Oral, Leaderboard #1 on OK-VQA as of Nov. 4, 2021**
13. **Z. Gan**, Y.-C. Chen, L. Li, T. Chen, Y. Cheng, S. Wang, J. Liu, L. Wang and Z. Liu "Playing Lottery Tickets with Vision and Language", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2022 **Oral**
14. J. Chen, Y. Cheng, **Z. Gan**, Q. Gu and J. Liu "Efficient Robust Training via Backward Smoothing", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2022
15. T.-J. Fu, L. Li, **Z. Gan**, K. Lin, W. Wang, L. Wang and Z. Liu "VIOLET: End-to-End Video-Language Transformers with Masked Visual-token Modeling", *arXiv preprint arXiv:2111.12681*
16. Y. Nie*, L. Li*, **Z. Gan**, S. Wang, C. Zhu, M. Zeng, Z. Liu, M. Bansal and L. Wang "MLP Architectures for Vision-and-Language Modeling: An Empirical Study", *arXiv preprint arXiv:2112.04453*
17. J. Wang, X. Hu, **Z. Gan**, Z. Yang, X. Dai, Z. Liu, Y. Lu and L. Wang "UFO: A UniFied TransfOrmer for Vision-Language Representation Learning", *arXiv preprint arXiv:2111.10023*

2021

1. T. Chen, Y. Cheng, **Z. Gan**, L. Yuan, L. Zhang and Z. Wang "Chasing Sparsity in Vision Transformers: An End-to-End Exploration", *Neural Information Processing Systems* (**NeurIPS**), 2021
2. X. Chen, Y. Cheng, S. Wang, **Z. Gan**, J. Liu and Z. Wang "The Elastic Lottery Ticket Hypothesis", *Neural Information Processing Systems* (**NeurIPS**), 2021
3. T. Chen, Y. Cheng, **Z. Gan**, J. Liu and Z. Wang "Data-Efficient GAN Training Beyond (Just) Augmentations: A Lottery Ticket Perspective", *Neural Information Processing Systems* (**NeurIPS**), 2021
4. B. Wang*, C. Xu*, S. Wang, **Z. Gan**, Y. Cheng, J. Gao, A. H. Awadallah and B. Li "Adversarial GLUE: A Multi-Task Benchmark for Robustness Evaluation of Language Models", *Neural Information Processing Systems* (**NeurIPS**), Datasets and Benchmarks Track, 2021 **Oral**
5. L. Li*, J. Lei*, **Z. Gan**, L. Yu, Y.-C. Chen, R. Pillai, Y. Cheng, L. Zhou, X. Wang, W. Wang, T. Berg, M. Bansal, J. Liu, L. Wang and Z. Liu "VALUE: A Multi-Task Benchmark for Video-and-Language Understanding Evaluation", *Neural Information Processing Systems* (**NeurIPS**), Datasets and Benchmarks Track, 2021
6. J. Chen, **Z. Gan**, X. Li, Q. Guo, L. Chen, S. Gao, T. Chung, Y. Xu, B. Zeng, W. Lu, F. Li, L. Carin and C. Tao "Simpler, Faster, Stronger: Breaking The log-K Curse On Contrastive Learners With FlatNCE", *Neural Information Processing Systems* (**NeurIPS**), Workshop on Self-Supervised Learning, 2021
7. L. Li, J. Lei, **Z. Gan** and J. Liu "Adversarial VQA: A New Benchmark for Evaluating the Robustness of VQA Models", *Int. Conf. on Computer Vision* (**ICCV**), 2021 **Oral, Top 3% among all submissions**
8. C. Zhu, Y. Cheng, **Z. Gan**, F. Huang, J. Liu and T. Goldstein "MaxVA: Fast Adaptation of Stepsizes by Maximizing Observed Variance of Gradients", *European Conf. Machine Learning (***ECML***)*, 2021
9. X. Chen, Y. Cheng, S. Wang, **Z. Gan**, Z. Wang and J. Liu "EarlyBERT: Efficient BERT Training via Early-bird Lottery Tickets", *Association for Computational Linguistics (***ACL***)*, 2021 **Oral**
10. S. Wang, L. Zhou, **Z. Gan**, Y.-C. Chen, Y. Fang, S. Sun, Y. Cheng and J. Liu "Cluster-Former: Clustering-based Sparse Transformer for Question Answering", *Findings of Association for Computational Linguistics* (**Findings of ACL**), 2021 **Leaderboard #1 on NaturalQuestions as of Sep. 27, 2020**

11. J. Lei*, L. Li*, L. Zhou, **Z. Gan**, T. L. Berg, M. Bansal and J. Liu "Less is More: ClipBERT for Video-and-Language Learning via Sparse Sampling", *Computer Vision and Pattern Recognition (***CVPR***)*, 2021 **Oral with 3 Strong Accepts, Best Student Paper Honorable Mention**
12. L. Chen*, D. Wang*, **Z. Gan**, J. Liu, R. Henao and L. Carin "Wasserstein Contrastive Representation Distillation", *Computer Vision and Pattern Recognition (***CVPR***)*, 2021
13. S. Dai, **Z. Gan**, Y. Cheng, C. Tao, L. Carin and J. Liu "APo-VAE: Text Generation in Hyperbolic Space", *North American Chapter of the Association for Computational Linguistics (***NAACL***)*, 2021
14. B. Wang, S. Wang, Y. Cheng, **Z. Gan**, R. Jia, B. Li and J. Liu "InfoBERT: Improving Robustness of Language Models from An Information Theoretic Perspective", *Int. Conf. Learning Representations* (**ICLR**), 2021 **Leaderboard #1 on Adversarial NLI as of Oct. 9, 2020**
15. S. Yuan*, P. Cheng*, R. Zhang, W. Hao, **Z. Gan** and L. Carin "Improving Zero-Shot Voice Style Transfer via Disentangled Representation Learning", *Int. Conf. Learning Representations* (**ICLR**), 2021
16. Y. Fang*, S. Wang*, **Z. Gan**, S. Sun and J. Liu "FILTER: An Enhanced Fusion Method for Cross-lingual Language Understanding", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2021 **Leaderboard #1 on XTREME and XGLUE as of Sep. 8, 2020**
17. W. Chen, **Z. Gan**, L. Li, Y. Cheng, W. Wang and J. Liu "Meta Module Network for Compositional Visual Reasoning", *Winter Conf. on Applications of Computer Vision* (**WACV**), 2021 **Best Student Paper Honorable Mention**
18. L. Zhou, J. Liu, Y. Cheng, **Z. Gan**, and L. Wang "CUPID: Adaptive Curation of Pre-training Data for Video-and-Language Representation Learning", *arXiv preprint arXiv:2104.00285*
19. L. Li, **Z. Gan** and J. Liu "A Closer Look at the Robustness of Vision-and-Language Pre-trained Models", *arXiv preprint arXiv:2012.08673* **SOTA on 7 VQA robustness benchmarks as of April 23, 2021**
20. Y. Fang, S. Wang, **Z. Gan**, S. Sun, J. Liu and C. Zhu "Accelerating Real-Time Question Answering via Question Generation", *arXiv preprint arXiv:2009.05167*
21. D. Wang, Y. Yang, C. Tao, **Z. Gan**, L. Chen, F. Kong, R. Henao and L. Carin "Proactive Pseudo-Intervention: Contrastive Learning For Interpretable Vision Models", *arXiv preprint arXiv:2012.03369*
22. M. Cheng, **Z. Gan**, Y. Cheng, S. Wang, C. Hsieh and J. Liu "Adversarial Masking: Towards Understanding Robustness Trade-off for Generalization", *OpenReview*

2020

1. **Z. Gan**, Y.-C. Chen, L. Li, C. Zhu, Y. Cheng and J. Liu "Large-Scale Adversarial Training for Vision-and-Language Representation Learning", *Neural Information Processing Systems* (**NeurIPS**), 2020 **Spotlight, Top 4% among all submissions, SOTA on 6 Vision+Language tasks**
2. S. Sun, **Z. Gan**, Y. Cheng, Y. Fang, S. Wang and J. Liu "Contrastive Distillation on Intermediate Representations for Language Model Compression", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020
3. S. Wang, Y. Fang, S. Sun, **Z. Gan**, Y. Cheng, J. Jiang and J. Liu "Cross-Thought for Sentence Encoder Pre-training", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020
4. Y. Dong, S. Wang, **Z. Gan**, Y. Cheng, J. Cheung and J. Liu "Multi-Fact Correction in Abstractive Text Summarization", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020
5. L. Li*, Y.-C. Chen*, Y. Cheng, **Z. Gan**, L. Yu and J. Liu "HERO: Hierarchical Encoder for Video+Language Omni-representation Pre-training", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020 **SOTA on 8 Video+Language datasets, Leaderboard #1 on TVR and TVC as of Sep. 15, 2020**
6. Y. Zhang*, G. Wang*, C. Li, **Z. Gan**, C. Brockett and B. Dolan "POINTER: Constrained Progressive Text Generation via Insertion-based Generative Pre-training", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020
7. Y. Fang, S. Sun, **Z. Gan**, R. Pillai, S. Wang and J. Liu "Hierarchical Graph Network for Multi-hop Question Answering", *arXiv preprint arXiv:1911.03631 Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020 **Leaderboard #1 on HotpotQA as of Dec. 1st, 2019**

8. Y. Cheng, **Z. Gan**, Y. Zhang, O. Elachqar, D. Li and J. Liu "Contextual Text Style Transfer", *Findings of Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2020
9. Y. Wei, **Z. Gan**, W. Li, S. Lyu, M.-C. Chang, L. Zhang, J. Gao and P. Zhang "MagGAN: High-Resolution Face Attribute Editing with Mask-Guided Generative Adversarial Network", *Asian Conf. on Computer Vision* (**ACCV**), 2020
10. S. Dai, Y. Cheng, Y. Zhang, **Z. Gan**, J. Liu and L. Carin "Contrastively Smoothed Class Alignment for Unsupervised Domain Adaptation", *Asian Conf. on Computer Vision* (**ACCV**), 2020
11. J. Cao, **Z. Gan**, Y. Cheng, L. Yu, Y.-C. Chen and J. Liu "Behind the Scene: Revealing the Secrets of Pre-trained Vision-and-Language Models", *European Conf. on Computer Vision* (**ECCV**), 2020 **Spotlight** (**Top 5% among all submissions**)
12. Y.-C. Chen*, L. Li*, L. Yu*, A. Kholy, F. Ahmed, **Z. Gan**, Y. Cheng and J. Liu "UNITER: UNiversal Image-TExt Representation Learning", *European Conf. on Computer Vision* (**ECCV**), 2020 **SOTA on 13 Vision+Language Datasets/Tasks, No. 1 on VCR and NLVR2 leaderboards as of Sep. 2019**
13. Y. Cheng, **Z. Gan**, Y. Li, J. Liu and J. Gao "Sequential Attention GAN for Interactive Image Editing", *ACM International Conference on Multimedia* (**ACMMM**), 2020
14. P. Cheng, W. Hao, S. Dai, J. Liu, **Z. Gan** and L. Carin "CLUB: A Contrastive Log-ratio Upper Bound of Mutual Information", *Int. Conf. Machine Learning* (**ICML**), 2020
15. L. Chen, **Z. Gan**, Y. Cheng, L. Li, L. Carin and J. Liu "Graph Optimal Transport for Cross-Domain Alignment", *Int. Conf. Machine Learning* (**ICML**), 2020
16. J. Xu, **Z. Gan**, Y. Cheng and J. Liu "Discourse-Aware Neural Extractive Text Summarization", *Association for Computational Linguistics (***ACL***)*, 2020
17. Y. Chen, **Z. Gan**, Y. Cheng, J. Liu and J. Liu "Distilling Knowledge Learned in BERT for Text Generation", *Association for Computational Linguistics (***ACL***)*, 2020
18. R. Zhang, C. Chen, **Z. Gan**, W. Wang, D. Shen, G. Wang, Z. Wen and L. Carin "Improving Adversarial Text Generation by Modeling the Distant Future", *Association for Computational Linguistics (***ACL***)*, 2020
19. Y. Li, Y. Cheng, **Z. Gan**, L. Yu, L. Wang and J. Liu "BachGAN: High-Resolution Image Synthesis from Salient Object Layout", *Computer Vision and Pattern Recognition (***CVPR***)*, 2020
20. J. Liu, W. Chen, Y. Cheng, **Z. Gan**, L. Yu, Y. Yang and J. Liu "VIOLIN: A Large-Scale Dataset for Video-and-Language Inference", *Computer Vision and Pattern Recognition (***CVPR***)*, 2020
21. R. Zhang, C. Chen, **Z. Gan**, Z. Wen, W. Wang and L. Carin "Nested-Wasserstein Self-Imitation Learning for Sequence Generation", *Artificial Intelligence and Statistics* (**AISTATS**), 2020
22. C. Zhu, Y. Cheng, **Z. Gan**, S. Sun, T. Goldstein and J. Liu "FreeLB: Enhanced Adversarial Training for Natural Language Understanding", *Int. Conf. Learning Representations* (**ICLR**), 2020 **Spotlight** (**Leaderboard #1 on GLUE, ARC Easy/Challenge and Commonsense QA as of Sep. 2019**)
23. W. Wang, H. Xu, **Z. Gan**, B. Li, G. Wang, L. Chen, Q. Yang, W. Wang and L. Carin "Graph-Driven Generative Models for Heterogeneous Multi-Task Learning", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2020 **Spotlight**
24. J. Hu, Y. Cheng, **Z. Gan**, J. Liu, J. Gao and G. Neubig "What Makes A Good Story? Designing Composite Rewards for Visual Storytelling", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2020 **Spotlight**

2019

1. W. Wang, C. Tao, **Z. Gan**, G. Wang, L. Chen, X. Zhang, R. Zhang, Q. Yang, R. Henao and L. Carin "Improving Textual Network Learning with Variational Homophilic Embeddings", *Neural Information Processing Systems* (**NeurIPS**), 2019
2. S. Sun, Y. Cheng, **Z. Gan**, and J. Liu "Patient Knowledge Distillation for BERT Model Compression", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2019
3. H. Wang, **Z. Gan**, X. Liu, J. Liu, J. Gao and H. Wang "Adversarial Domain Adaptation for Machine Reading Comprehension", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2019

4. D. Li, Y. Zhang, **Z. Gan**, Y. Cheng, C. Brockett, M. Sun and B. Dolan "Domain Adaptive Text Style Transfer", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2019
5. M. Jiang, Q. Huang, L. Zhang, X. Wang, P. Zhang, **Z. Gan**, J. Diesner and J. Gao "TIGEr: Text-to-Image Grounding for Image Caption Evaluation", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2019
6. L. Li, **Z. Gan**, Y. Cheng and J. Liu "Relation-Aware Graph Attention Network for Visual Question Answering", *Int. Conf. on Computer Vision* (**ICCV**), 2019
7. **Z. Gan**, Y. Cheng, A. Kholy, L. Li, J. Liu and J. Gao "Multi-step Reasoning via Recurrent Dual Attention for Visual Dialog", *Association for Computational Linguistics (***ACL***)*, 2019
8. L. Ke, X. Li, Y. Bisk, A. Holtzman, **Z. Gan**, J. Liu, J. Gao, Y. Choi, and S. Srinivasa "Tactical Rewind: Self-Correction via Backtracking in Vision-and-Language Navigation", *Computer Vision and Pattern Recognition (***CVPR***)*, 2019 **Oral**
9. Y. Li, **Z. Gan**, Y. Shen, J. Liu, Y. Cheng, Y. Wu, L. Carin, D. Carlson and J. Gao "StoryGAN: A Sequential Conditional GAN for Story Visualization", *Computer Vision and Pattern Recognition (***CVPR***)*, 2019
10. W. Wang, **Z. Gan**, H. Xu, R. Zhang, G. Wang, D. Shen, C. Chen and L. Carin "Topic-Guided Variational Autoencoders for Text Generation", *North American Chapter of the Association for Computational Linguistics (***NAACL***)*, 2019 **Oral**
11. L. Chen, Y. Zhang, R. Zhang, C. Tao, **Z. Gan**, H. Zhang, B. Li, D. Shen, C. Chen and L. Carin "Improving Sequence-to-Sequence Learning via Optimal Transport", *Int. Conf. Learning Representations* (**ICLR**), 2019
12. Q. Huang*, **Z. Gan***, A. Celikyilmaz, D. Wu, J. Wang and X. He "Hierarchically Structured Reinforcement Learning for Topically Coherent Visual Story Generation", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2019 **Spotlight**

2018

1. Y. Zhang, M. Galley, J. Gao, **Z. Gan**, X. Li, C. Brockett and B. Dolan "Generating Informative and Diverse Conversational Responses via Adversarial Information Maximization", *Neural Information Processing Systems* (**NeurIPS**), 2018
2. L. Chen, S. Dai, C. Tao, D. Shen, **Z. Gan**, H. Zhang, Y. Zhang and L. Carin "Adversarial Text Generation via Feature-Mover's Distance", *Neural Information Processing Systems* (**NeurIPS**), 2018
3. X. Zhang, R. Henao, **Z. Gan**, Y. Li and L. Carin "Multi-Label Learning from Medical Plain Text with Convolutional Residual Models", *Machine Learning for Healthcare* (**MLHC**), 2018 **Spotlight**
4. Y. Pu, S. Dai, **Z. Gan**, W. Wang, G. Wang, Y. Zhang, R. Henao and L. Carin "JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets", *Int. Conf. Machine Learning* (**ICML**), 2018
5. T. Xu, P. Zhang, Q. Huang, H. Zhang, **Z. Gan**, X. Huang and X. He "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks", *Computer Vision and Pattern Recognition (***CVPR***)*, 2018
6. W. Wang, **Z. Gan**, W. Wang, D. Shen, J. Huang, W. Ping, S. Satheesh and L. Carin "Topic Compositional Neural Language Model", *Artificial Intelligence and Statistics* (**AISTATS**), 2018
7. Y. Pu, M. R. Min, **Z. Gan** and L. Carin "Adaptive Feature Abstraction for Translating Video to Text", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2018

2017

1. **Z. Gan***, L. Chen*, W. Wang, Y. Pu, Y. Zhang, H. Liu, C. Li and L. Carin "Triangle Generative Adversarial Networks", *Neural Information Processing Systems* (**NeurIPS**), 2017
2. Y. Pu, W. Wang, R. Henao, L. Chen, **Z. Gan**, C. Li, and L. Carin "Adversarial Symmetric Variational Autoencoder", *Neural Information Processing Systems* (**NeurIPS**), 2017
3. Y. Pu, **Z. Gan**, R. Henao, C. Li, S. Han and L. Carin "VAE Learning via Stein Variational Gradient Descent", *Neural Information Processing Systems* (**NeurIPS**), 2017

4. Y. Zhang, D. Shen, G. Wang, **Z. Gan**, R. Henao and L. Carin "Deconvolutional Paragraph Representation Learning", *Neural Information Processing Systems* (**NeurIPS**), 2017
5. **Z. Gan**, Y. Pu, R. Henao, C. Li, X. He and L. Carin "Learning Generic Sentence Representations Using Convolutional Neural Networks", *Conf. on Empirical Methods in Natural Language Processing* (**EMNLP**), 2017 **Oral**
6. Y. Zhang, **Z. Gan**, K. Fan, Z. Chen, R. Henao, D. Shen and L. Carin "Adversarial Feature Matching for Text Generation", *Int. Conf. Machine Learning* (**ICML**), 2017
7. Y. Zhang, C. Chen, **Z. Gan**, R. Henao and L. Carin "Stochastic Gradient Monomial Gamma Sampler", *Int. Conf. Machine Learning* (**ICML**), 2017
8. **Z. Gan***, C. Li*, C. Chen, Y. Pu, Q. Su and L. Carin "Scalable Bayesian Learning of Recurrent Neural Networks for Language Modeling", *Association for Computational Linguistics (***ACL***)*, 2017 **Oral**
9. **Z. Gan**, C. Gan, X. He, Y. Pu, K. Tran, J. Gao, L. Carin and L. Deng "Semantic Compositional Networks for Visual Captioning", *Computer Vision and Pattern Recognition (***CVPR***)*, 2017 **Spotlight**
10. C. Gan, **Z. Gan**, X. He, J. Gao and L. Deng "StyleNet: Generating Attractive Visual Captions with Styles", *Computer Vision and Pattern Recognition (***CVPR***)*, 2017
11. **Z. Gan**, P. D. Singh, A. Joshi, X. He, J. Chen, J. Gao and L. Deng "Character-level Deep Conflation for Business Data Analytics", *Int. Conf. Acoustics, Speech and Signal Processing* (**ICASSP**), 2017
12. Y. Xian, Y. Pu, **Z. Gan**, L. Lu and A. Thompson "Adaptive DCTNet for Audio Signal Classification", *Int. Conf. Acoustics, Speech and Signal Processing* (**ICASSP**), 2017
13. Q. Su, X. Liao, C. Li, **Z. Gan** and L. Carin "Unsupervised Learning with Truncated Gaussian Graphical Models", *Proc. American Association of Artificial Intelligence* (**AAAI**), 2017 **Oral**

2016

1. Y. Zhang, **Z. Gan** and L. Carin "Generating Text via Adversarial Training", *NeurIPS Workshop*, 2016
2. Y. Xian, Y. Pu, **Z. Gan**, L. Lu and A. Thompson "Modified DCTNet for Audio Signals Classification", *Journal of the Acoustical Society of America*, 2016
3. Y. Pu, **Z. Gan**, R. Henao, X. Yuan, C. Li, A. Stevens and L. Carin "Variational Autoencoder for Deep Learning of Images, Labels and Captions", *Neural Information Processing Systems* (**NeurIPS**), 2016
4. J. Song, **Z. Gan** and L. Carin "Factored Temporal Sigmoid Belief Networks for Sequence Learning", *Int. Conf. Machine Learning* (**ICML**), 2016
5. C. Li, A. Stevens, C. Chen, Y. Pu, **Z. Gan** and L. Carin "Learning Weight Uncertainty with Stochastic Gradient MCMC for Shape Classification", *Computer Vision and Pattern Recognition (***CVPR***)*, 2016 **Spotlight**
6. C. Chen, D. Carlson, **Z. Gan**, C. Li and L. Carin "Bridging the Gap Between Stochastic Gradient MCMC and Stochastic Optimization", *Artificial Intelligence and Statistics (***AISTATS***)*, 2016 **Oral**

2015

1. **Z. Gan**, C. Li, R. Henao, D. Carlson and L. Carin "Deep Temporal Sigmoid Belief Networks for Sequence Modeling", *Neural Information Processing Systems* (**NeurIPS**), 2015
2. R. Henao, **Z. Gan**, J. Lu and L. Carin "Deep Poisson Factor Modeling", *Neural Information Processing Systems* (**NeurIPS**), 2015
3. **Z. Gan**, C. Chen, R. Henao, D. Carlson and L. Carin "Scalable Deep Poisson Factor Analysis for Topic Modeling", *Int. Conf. Machine Learning* (**ICML**), 2015
4. **Z. Gan**, R. Henao, D. Carlson and L. Carin "Learning Deep Sigmoid Belief Networks with Data Augmentation", *Artificial Intelligence and Statistics* (**AISTATS**), 2015
5. **Z. Gan**, X. Yuan, R. Henao, E. Tsalik and L. Carin "Inference of Gene Networks Associated with the Host Response to Infectious Disease", Chapter 13 of Book *Big Data Over Networks*. Cambridge University Press. In Press.

PhD Dissertation

1. **Z. Gan** "Deep Generative Models for Vision and Language Intelligence", Duke University.

Tutorial and Workshop

1. C. Li, **Z. Gan**, H. Zhang, J. Yang, L. Li, Z. Yang, K. Lin, J. Gao and L. Wang "Recent Advances in Vision Foundation Models", *Computer Vision and Pattern Recognition (***CVPR***)*, 2024
2. L. Li, **Z. Gan**, C. Li, J. Yang and Z. Yang "Recent Advances in Vision Foundation Models", *Computer Vision and Pattern Recognition (***CVPR***)*, 2023
3. **Z. Gan**, L. Li, C. Li, J. Yang, P. Zhang, L. Wang, Z. Liu and J. Gao "Recent Advances in Vision-and-Language Pre-training", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022
4. M. Luo, T. Gokhale, Z. Fang, P. Banerjee, Y. Yang, C. Baral, D. Teney, **Z. Gan**, K. Marino, T. Wang and S. Aditya "O-DRUM: Workshop on Open-Domain Retrieval Under a Multi-Modal Setting", *Computer Vision and Pattern Recognition (***CVPR***)*, 2022
5. **Z. Gan**, C. Li, J. Yang and P. Zhang "Microsoft Vision+Language Summer Talk Series", 2021
6. P. Anderson, Y. Artzi, **Z. Gan**, X. He, L. Li, J. Liu, X. Wang, Q. Wu and L. Zhou "From VQA to VLN: Recent Advances in Vision-and-Language Research", *Computer Vision and Pattern Recognition (***CVPR***)*, 2021
7. **Z. Gan**, L. Yu, Y. Cheng, L. Zhou, L. Li, Y.-C. Chen, J. Liu and X. He "Recent Advances in Vision-and-Language Research", *Computer Vision and Pattern Recognition (***CVPR***)*, 2020
8. P. Knees and **Z. Gan** "The ACM Multimedia 2020 Interactive Arts Exhibition"

Professional Activities

Senior Area Chair: ACL 2025, EMNLP 2024

Area Chair for Top-tier AI Conferences:

- 2024: NeurIPS, ICML, ICLR, CVPR, WACV, ACL, NAACL, COLM
- 2023: NeurIPS, ICML, ICLR, CVPR, EMNLP, AAAI, IJCNLP-AACL
- 2022: NeurIPS, ICML, ECCV, ACL, NAACL, EMNLP, AAAI
- 2021: NeurIPS, ICML, ICLR, ACL
- 2020: NeurIPS
- 2019: NeurIPS

Senior Program Committee (SPC) Member: AAAI 2021/2020

Action Editor: Transactions on Machine Learning Research (TMLR), ACL Rolling Review

Interactive Arts Chair: ACMMM 2020

Awarded as Outstanding SPC Member: AAAI 2020

Awarded as Top/Outstanding Reviewer: EMNLP 2020, ICML 2020, NeurIPS 2018

Conference Reviewer/PC Member:

- 2022: ICLR, CVPR, WACV
- 2021: CVPR, ICCV, WACV; NAACL, EMNLP
- 2020: ICML, ICLR, IJCAI; CVPR, ECCV, ACMMM; ACL, EMNLP, COLING, AACL, CoNLL
- 2019: ICML, ICLR, AAAI; CVPR, ICCV, ACMMM; EMNLP, CoNLL
- 2018: NeurIPS, EMNLP, CVPR, ACCV
- 2016: NIPS

Journal Reviewer:

- Transactions on Pattern Analysis and Machine Intelligence (PAMI)
- Journal of Machine Learning Research (JMLR)

- Transactions on Image Processing (TIP)
- Transactions on Knowledge and Data Engineering (KDE)
- Journal of Selected Topics in Signal Processing (STSP)
- Transactions on Multimedia Computing Communications and Applications (TOMM)
- Transactions on Audio, Speech and Language Processing (ASL)
- Science China
- Transactions on Cybernetics, IET Computer Vision, Entropy, Artificial Intelligence

Workshop Reviewer/PC Member:

- 2023: 3rd ACL Workshop on Advances in Language and Vision Research
- 2021: AAAI Workshop on Optimal Transport and Structured Data Modeling
- 2021: 4th ICCV Workshop on Closing the loop between Vision and Language
- 2021: NeurIPS Workshop on Disentanglement and Controllable Generation for Vision and Language
- 2021: 2nd NAACL Workshop on Advances in Language and Vision Research
- 2020: ACL Workshop on Advances in Language and Vision Research
- 2019: ICCV Workshop on Closing the loop between Vision and Language
- 2019: ICLR Workshop on Deep Generative Models for Highly Structured Data
- 2018: ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models

Talks

- "Methods, Analysis Insights from Multimodal LLM Pre-training", *CVPR Tutorial*, Seattle, US, June 2024
- "MM1: Methods, Analysis Insights from Multimodal LLM Pre-training", *CVPR Workshop*, Seattle, US, June 2024
- "Recent Advances in Vision Foundation Models", *CVPR Tutorial*, Vancouver, Canada, June 2023
- "Towards Building Multimodal Foundation Models", *WACV Tutorial*, Zoom, January 2023
- "Towards Building Multimodal Foundation Models", *Ohio State University (OSU)*, Zoom, November 2022
- "Towards Building Multimodal Foundation Models", *Apple AI/ML*, Zoom, October 2022
- "Finding Universal Lottery Tickets in Large Neural Networks Efficiently", *Fudan University*, Zoom, August 2022
- "Big Models, Few-shot Learning, and Model Evaluation for Vision-Language Pre-training", *CVPR Tutorial*, New Orleans, USA, June 2022
- "Neural Networks for NLP", *Duke Machine Learning Summer School*, Zoom, June 2022
- "Vision-Language Pre-training for Multimodal Intelligence", *Google Brain*, Zoom, May 2022
- "Playing Lottery Tickets with Vision and Language", *AAAI*, Zoom, February 2022
- "Compressing Transformers with Knowledge Distillation and Pruning", *Baidu Research*, Zoom, January 2022
- "How Much Can GPT-3 Benefit Few-Shot Visual Reasoning?", *Microsoft Research Summit*, October 2021
- "Vision-and-Language Pre-training: Basics, Recent Advances, and Future Directions", *University of California, Merced*, Zoom, October 2021
- "Large-scale Vision-and-Language Pre-training for Multimodal Learning", Keynote at the 3rd Workshop on Continual and Multimodal Learning for Internet of Things, *IJCAI*, Zoom, August 2021
- "Recent Advances in Vision-Language Pre-training", *University of Bristol*, Zoom, June 2021

- "Recent Advances in Vision-Language Pre-training", *Wuhan University*, Zoom, May 2021
- "Recent Advances in Vision-Language Pre-training", *University of California, Santa Cruz (UCSC)*, Zoom, May 2021
- "Vision-Language Pre-training", *Student Forum on Frontiers of AI (SFFAI)*, Zoom, April 2021
- "Large-Scale Adversarial Training for Vision-and-Language Representation Learning", *NeurIPS*, Zoom, December 2020
- "Visual QA and Reasoning", *CVPR Tutorial*, Zoom, June 2020
- "Deep Generative Models for Vision and Language Intelligence", *Ph.D. Final Defense*, Durham, NC, February 2018
- "Deep Generative Models for Vision and Language Intelligence", IBM Thomas J. Watson Research Center, Yorktown, NY, October 2017
- "Deep Generative Models for Vision and Language Intelligence", NVIDIA, Santa Clara, CA, September 2017
- "Deep Generative Models for Vision and Language Intelligence", Apple, Cupertino, CA, September 2017
- "Learning Generic Sentence Representations Using Convolutional Neural Networks", *EMNLP*, Copenhagen, Denmark, September 2017
- "Semantic Compositional Networks for Visual Captioning", *CVPR*, Hawaii, July 2017
- "Semantic Compositional Networks for Visual Captioning", *Ph.D. Preliminary Exam*, Durham, NC, April 2017
- "Deep Generative Models for Sequence Learning", *Ph.D. Qualifying Exam*, Durham, NC, December 2015

Competitions

- 2022/06: First human parity on TextCaps, and Rank 1st on NoCaps and VizWiz leaderboards
- 2021/09: Rank 1st on OK-VQA leaderboard
- 2021/06: Rank 1st on TextCaps Challenge 2021
- 2020/10: Rank 1st on Adversarial NLI leaderboard
- 2020/09: Rank 1st on NaturalQuestions leaderboard
- 2020/09: Rank 1st on TVR and TVC leaderboards
- 2020/09: Rank 1st on XTREME and XGLUE leaderboards
- 2020/05: Rank 4th on VQA Challenge 2020
- 2019/12: Rank 1st on HotpotQA leaderboard
- 2019/10: Rank 1st on VCR and NLVR2 leaderboards
- 2019/09: Rank 1st on GLUE, ARC Easy/Challenge and Commonsense QA leaderboards
- 2019/06: Rank 2nd in Visual Dialog Challenge 2019
- 2018/09: Rank 3rd in Visual Dialog Challenge 2018

Awards

- CVPR 2021 Best Student Paper Honorable Mention
- WACV 2021 Best Student Paper Honorable Mention
- AAAI 2020 Outstanding Senior Program Committee Member Award
- ECE Fellowship, Duke University, 2013
- National Scholarship, Department of Minister of Education of China, 2010-2013