

Phone: (530)407-0090
Email: jian.chen@dolby.com
Page: puar-playground.github.io

Jian Chen

Senior AI Researcher

Google Scholar: [Jian Chen](#)
GitHub: [puar-playground](#)
LinkedIn: [jian-chen-1a0b9a11b](#)

I am a Senior Foundational AI Researcher at Dolby Laboratories in the Advanced Technology Group, where I focus on multimodal AI research aimed at enabling immersive digital experiences and ensuring data integrity and security. I also continue to collaborate actively with academic partners on the application and optimization of multimodal-LLMs and ML algorithms.

EXPERIENCE

Dolby Laboratories , Senior Foundational Audio AI Researcher,	06/2025 – present
Adobe , Intern Research Scientist/Engineer,	05/2024 – 11/2024
MBZUAI , Research Assistant,	10/2023 – 02/2024
University at Buffalo , Research Assistant,	09/2018 – 09/2021

EDUCATION

Ph.D. in Computer Science, <i>University at Buffalo, USA</i>	09/2018 – 06/2025
M.S. in Electrical Engineering, <i>Drexel University, USA</i>	09/2015 – 06/2017
B.S. in Applied Mathematics, <i>Hunan University, China</i>	09/2011 – 07/2015

RESEARCH INTERESTS

Audio AI, Multimodal Large Language Models, Generative Models, Representation Learning

SELECTED RESEARCH PROJECTS

Intern Research Scientist / Mentor: Ruiyi Zhang / Adobe	05/2024 – 11/2024
--	--------------------------

- Developed SV-RAG, a framework that enables multimodal large language models for multi-page document understanding. Our models are fine-tuned based on Phi-3-V, Paligemma, and InternVL2. SV-RAG rivals Google's Gemini-1.5-pro and Claude 3 Opus with the 4B InternVL2 backbone, achieving state-of-the-art accuracy in open-source models. ICLR 2025 [1]
- Created the [Multi-Modal Reading \(MMR\)](#) benchmark for evaluating Vision-Language Models' reading ability and tested multiple baseline models. submitted to IJCAI 2025 [2]

Research Assistant / Advisor: Changyou Chen / University at Buffalo	09/2022 – 06/2023
--	--------------------------

- Developed TextLap, a novel model for text-based layout planning, and created the InsLap dataset for document layout generation. Fine-tuned large language models to achieve state-of-the-art performance on graphical design benchmarks. EMNLP 2024 [3]
- Contributed to the development of [LLaVA-Read](#), responsible for part of the experimental work. LLaVA-Read is a multimodal large language model that uses dual visual encoders and a visual text encoder to enhance the understanding of text-rich images. [4]
- Contributed to the creation of the document dataset used for training [LaRA](#), a multimodal language model that enables reading capability through OCR input. CVPR 2024 [5]
- Developed LACE, a continuous diffusion model using constrained optimization for layout generation and editing. LACE is a unified model that integrates multiple conditional inputs, achieving state-of-the-art performance across several datasets. ICLR 2024 [6]
- Developed LRA-Diffusion, a method for learning from noisy labels by framing data labeling as a conditional generation problem. LRA-Diffusion employs a diffusion model to learn the label generation process, using pre-trained encoders as conditional inputs to mitigate overfitting. It has consistently ranked first on three leaderboards on [Papers with Code](#). NeurIPS 2023 [7]

- Designed TimedHN, a causal inference model capturing gene dependencies during cancer progression. The model uses binary mutation profiles as states and progression time as a latent variable, constructing a continuous-time Markov chain. An efficient optimization algorithm leveraging data sparsity was also developed to improve performance. [8]
- Proposed a neural network architecture replacing convolution with differentiable approximate string matching for microbial DNA and RNA sequence representation learning. This led to [AsMac](#), an efficient large-scale sequence comparison method. [9]

PROFESSIONAL AND VOLUNTEER SERVICE

Peer Reviewer

- 2025 ICCV, ICML, CVPR, ICLR, ACL
- 2024 ICML, TMLR, TETCI, PLOS ONE

AWARDS

Best Graduation Thesis: "Low-rank tensor optimization for video image recovery." Hunan University, 2015

SELECTED PUBLICATIONS

1. **Jian Chen**, Ruiyi Zhang, Yufan Zhou, Tong Yu, Jiuxiang Gu, Ryan A. Rossi, Changyou Chen, and Tong Sun. [SV-RAG: LoRA-Contextualizing Adaptation of MLLMs for Long Document Understanding](#). In *ICLR*, 2025
2. **Jian Chen**, Ruiyi Zhang, Yufan Zhou, Ryan Rossi, Jiuxiang Gu, and Changyou Chen. [MMR: Evaluating Reading Ability of Large Multimodal Models](#). *arXiv preprint arXiv:2408.14594*, 2024
3. **Jian Chen**, Ruiyi Zhang, Yufan Zhou, Jennifer Healey, Jiuxiang Gu, and Changyou Chen. [TextLap: Customizing Language Models for Text-to-Layout Planning](#). In *EMNLP Findings*, 2024
4. Ruiyi Zhang, Yufan Zhou, **Jian Chen**, Jiuxiang Gu, Changyou Chen, and Tong Sun. [LLaVA-Read: Enhancing Reading Ability of Multimodal Language Models](#). *arXiv preprint arXiv:2407.19185*, 2024
5. Ruiyi Zhang, Yanzhe Zhang, **Jian Chen**, Yufan Zhou, Jiuxiang Gu, Changyou Chen, and Tong Sun. [TRINS: Towards Multimodal Language Models That Can Read](#). In *CVPR*, 2024
6. **Jian Chen**, Ruiyi Zhang, Yufan Zhou, Rajiv Jain, Zhiqiang Xu, Ryan Rossi, and Changyou Chen. [Towards Aligned Layout Generation via Diffusion Model with Aesthetic Constraints](#). In *ICLR*, 2024
7. **Jian Chen**, Ruiyi Zhang, Tong Yu, Rohan Sharma, zhiqiang xu, Tong Sun, and Changyou Chen. [Label-Retrieval-Augmented Diffusion Models for Learning from Noisy Labels](#). In *NeurIPS*, 2023
8. **Jian Chen**. [Timed hazard networks: Incorporating temporal difference for oncogenetic analysis](#). *PLOS ONE*, 2023
9. **Jian Chen**, Le Yang, Lu Li, Steve Goodison, and Yijun Sun. [Alignment-free Comparison of Metagenomics Sequences via Approximate String Matching](#). *Bioinformatics Advances*, 2022
10. Jiayu Qin, **Jian Chen**, Rohan Sharma, Jingchen Sun, and Changyou Chen. A probability contrastive learning framework for 3d molecular representation learning. In *NeurIPS*, 2024