YIBO YANG

Research Scientist, AI/ML | Chan Zuckerberg Biohub

yiboyang1@gmail.com | yiboyang.com | Google Scholar

SUMMARY

Research scientist developing principled methods for generative modeling and probabilistic inference. My work spans diffusion modeling and guided generation for biological sequence design, generative and computationally efficient image/video compression, applications of optimal transport and deep learning to information theory, and statistical modeling in other scientific domains such as particle physics and astronomy.

INDUSTRY EXPERIENCE

Chan Zuckerberg Biohub (Redwood City, CA) Dec. 2024 – Present

Research Scientist, AI/ML

- Developed generative evaluations and inference-time guidance capability for a discrete diffusion model of biology sequences.
- Led a biological sequence design campaign from computational generation and candidate selection through successful wet-lab validation.

Amazon (New York, NY) Summer 2024

Research Intern | Host: Georgiana Dinu

- Developed a method for aligning and translating between language-model embedding spaces to enable efficient cross-model retrieval.

Vector Institute for AI (Toronto, Canada) Fall 2023

Research Intern | Hosts: Alireza Makhzani and Roger Grosse

- Studied monitoring and steering of transformer language models for AI safety.

Google Research (Mountain View, CA) Summer 2022

Research Intern, Neural Compression | Hosts: David Minnen and Lucas Theis

- Developed algorithms for instance-adaptive and computationally efficient neural image compression.

EDUCATION

University of California, Irvine 2024

M.S. and Ph.D. in Computer Science | Advisor: Stephan Mandt

University of Texas at Dallas 2018

B.S. in Computer Science, magna cum laude | Computing Scholars Honors Program

ACADEMIC SERVICE AND RECOGNITION

- Presented and co-organized neural compression tutorials at NeurIPS 2022, AAAI 2023, and UAI 2023; co-organized neural compression workshops at ICML 2023 and NeurIPS 2024.
- Best Student Presentation, ITA (Information Theory and Applications) Workshop 2024.
- Outstanding Reviewer: NeurIPS 2020/2021 and ICLR 2021.

SELECTED PUBLICATIONS (see Google Scholar for a full publication list)

Advances in Diffusion-Based Generative Compression
Yibo Yang, Stephan Mandt. *IEEE BITS – the Information Theory Magazine, 2026*.

> Presented a unifying review of recent diffusion-based generative compression methods through rate–distortion–perception theory, with connections to channel simulation and inverse problems.

Transformers for Mixed-type Event Sequences
Felix Draxler, Yang Meng, Kai Nelson, Lukas Laskowski, **Yibo Yang**, Theofanis Karaletsos, Stephan Mandt. *NeurIPS 2025* **(Spotlight)**.

> Developed a transformer-based point process with categorical and normalizing-flow heads to jointly model event times and mixed discrete/continuous attributes, with applications to electronic health records.

Progressive Compression with Universally Quantized Diffusion Models
Yibo Yang, Justus C. Will, Stephan Mandt. *ICLR 2025* **(Oral Presentation)**.

> Introduced a new diffusion formalism with uniform noise corruption, enabling tractable progressive compression from lossy to lossless reconstruction, learned end-to-end from data.

Estimating the Rate-Distortion Function by Wasserstein Gradient Descent
Yibo Yang, Stephan Eckstein, Marcel Nutz, Stephan Mandt. *NeurIPS 2023*.

> Applied optimal transport ideas to the rate–distortion estimation problem in information theory, yielding tighter bounds with finite-sample convergence bounds.

Computationally-Efficient Neural Image Compression with Shallow Decoders
Yibo Yang, Stephan Mandt. *ICCV 2023*.

> Replaced deep CNN decoders with shallow or linear transforms and compensated with stronger/iterative encoding, resulting in an **80%** reduction in decoding complexity with competitive rate–distortion performance.

An Introduction to Neural Data Compression
Yibo Yang, Stephan Mandt, Lucas Theis. *Foundations and Trends in Computer Graphics and Vision, 2023*.

> A research monograph surveying the theory and application of generative modeling approaches to data compression.

Learning to Simulate High Energy Particle Collisions from Unlabeled Data
Jessica N. Howard, Stephan Mandt, Daniel Whiteson, **Yibo Yang**. *Scientific Reports 2022*.

> Developed a physics-constrained Wasserstein autoencoder that learns to predict experimental particle collision data directly from theoretical models without labeled simulation pairs.

Towards Empirical Sandwich Bounds on the Rate-Distortion Function
Yibo Yang, Stephan Mandt. *ICLR 2022*.

> Proposed the first deep-learning-based algorithms for bounding the information-theoretic rate–distortion function from samples only.

Hierarchical Autoregressive Modeling for Neural Video Compression
Ruihan Yang, **Yibo Yang**, Joseph Marino, Stephan Mandt. *ICLR 2021*.

> Improved lossy video compression based on more expressive autoregressive generative modeling.

Improving Inference for Neural Image Compression
Yibo Yang, Robert Bamler, Stephan Mandt. *NeurIPS 2020*.

> Cast learning-based image compression as variational inference and closed fundamental approximation gaps with inference-time optimization based on a novel variational optimization algorithm in discrete space.