Research Summary

Mingyang Deng

Modern generative modeling has revolutionized machine learning, enabling powerful applications across vision, language, and science. Yet today's dominant methods—diffusion and autoregressive models—remain fundamentally indirect: they construct generation as a long chain of iterative refinements, which can be computationally expensive and difficult to adapt to new domains. My research aims to develop **direct, efficient, and adaptable generative paradigms** that can learn distributions end-to-end and rapidly specialize to real-world data.

1. Fast and Direct Generative Modeling. Traditional diffusion models simulate a long trajectory from noise to data, while autoregressive models generate token by token. Both approaches incur sequential bottlenecks. My recent work, *Mean Flows for One-step Generative Modeling* (NeurIPS 2025 Oral) [1], establishes a new formulation based on **mean-field dynamics**: instead of modeling stepwise velocity, we learn the *time-averaged* velocity over an interval and train a single-step generator to match it, thereby avoiding iterative denoising and its error accumulation.

2. Efficient Generation and Self-Refinement. A complementary direction in my research improves the efficiency of existing generative processes. In *Restart Sampling for Improving the Generative Process* (NeurIPS 2023) [3], we show that diffusion models can leverage their learned score function to **iteratively refine and self-correct** samples after initial generation, yielding substantial quality gains at lower inference cost. This demonstrates that the "laws" learned by a model—the underlying score function—can be reused to bootstrap better generations without external supervision, reducing reliance on dense human-curated data.

3. Modeling Continuous Distributions and Adaptation. Beyond discrete tokens, many real-world signals are continuous and structured. *Autoregressive Image Generation without Vector Quantization* (NeurIPS 2024 Spotlight) [2] introduces a continuous tokenizer, bridging discrete sequence modeling and continuous generation. This enables differentiable latent spaces that can be fine-tuned efficiently for domain-specific tasks. Building on this, I am exploring how such latent representations can be aligned with external structures—physical constraints, feature embeddings, or domain priors—to make generative models more data-efficient and adaptable to specialized settings like simulation, design, and multimodal reasoning.

4. Broader Vision. The unifying goal of my research is to build **generative models that learn and adapt like systems of interacting agents**: directly, efficiently, and with minimal external supervision. By grounding generation in physical and probabilistic principles, these models can generalize beyond abundant data regimes and support rapid adaptation to real-world domains—from scalable multimodal generation to structure-aware modeling of scientific or knowledge systems.

Selected Publications

- Zhengyang Geng, **Mingyang Deng**, Xingjian Bai, J. Z. Kolter, Kaiming He. *Mean Flows for One-step Generative Modeling.* NeurIPS 2025 (Oral). arXiv:2505.13447.
- Tianhong Li, Yonglong Tian, He Li, **Mingyang Deng**, Kaiming He. *Autoregressive Image Generation without Vector Quantization.* NeurIPS 2024 (Spotlight). arXiv:2406.11838.
- Yilun Xu*, **Mingyang Deng***, Xiang Cheng*, Yonglong Tian, Ziming Liu, Tommi Jaakkola. *Restart Sampling for Improving the Generative Process.* NeurIPS 2023. arXiv:2306.14878.

References

[1] Z. Geng, **M. Deng**, X. Bai, J. Z. Kolter, and K. He. Mean Flows for One-step Generative Modeling. *NeurIPS*, 2025. `https://arxiv.org/abs/2505.13447`.

[2] T. Li, Y. Tian, H. Li, **M. Deng**, and K. He. Autoregressive Image Generation without Vector Quantization. *NeurIPS*, 2024. `https://arxiv.org/abs/2406.11838`.

[3] Y. Xu*, **M. Deng***, X. Cheng*, Y. Tian, Z. Liu, and T. Jaakkola. Restart Sampling for Improving the Generative Process. *NeurIPS*, 2023. `https://arxiv.org/abs/2306.14878`.