Enhancing Accuracy and Efficiency in Diffusion Models

Chieh-Hsin LAI

Sony AI

Diffusion models are powerful generative tools for high-fidelity data generation across various applications. However, they encounter two main challenges: slow sampling speed and fixed dimensionality (resolution). In this talk, I will introduce approaches to improve efficiency and accuracy in training and sampling by leveraging a unified concept of "consistency" rooted in the mathematical structure of diffusion models.

To accelerate sampling, I will introduce the Consistency Trajectory Model (CTM), which condenses a pre-trained diffusion model into a single neural network. CTM produces scores (log-density gradients) in one forward pass, enabling seamless traversal between any initial and final time along the Probability Flow ODE. This capability allows for new deterministic and stochastic sampling methods, including long jumps along ODE trajectories.

To address dimensionality limitations, I will present Progressive Growing of Diffusion Autoencoder (PaGoDA), which enhances the generator's resolution beyond that of the pre-trained diffusion model. By using a pre-trained low-resolution diffusion model to encode high-resolution data into a structured latent space, PaGoDA progressively increases the decoder's resolution without needing to retrain models during upsampling, thereby improving efficiency.

I hope my talk will stimulate collaborations and discussions across disciplines and contribute to AI-based Natural Sciences research.