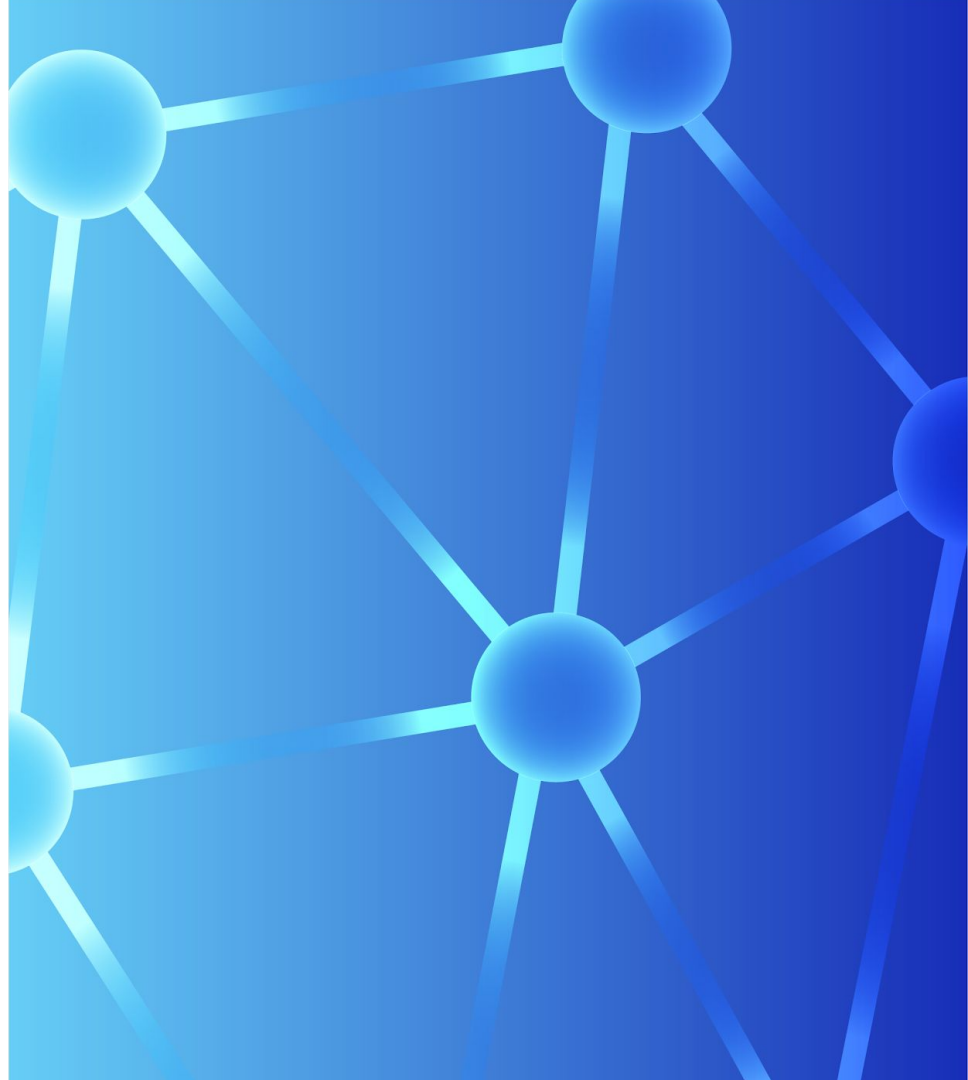


CVPR 2025 Tutorial: Efficient Text-to-Image/Video Modeling

Ameesh Makadia

12 June 2025

Google Research



Different perspectives on *efficiency*

Different perspectives on *efficiency*

Compression

More compact latent spaces → more efficient generation

Different perspectives on *efficiency*

Compression

More compact latent spaces → more efficient generation

Structured representations

Latent representation design that enables efficient modeling

Different perspectives on *efficiency*

Compression

More compact latent spaces → more efficient generation

Structured representations

Latent representation design that enables efficient modeling

Data sparsity

Generative models designed for data-sparse settings

Agenda

Part I - Compression (15 min)

Factorized latent representations for video

Part II - Structured representations (15 min)

Multiscale image generation with autoregressive models

Part III - Data sparsity (< 10 min)

Diffusion models from a single 3D shape