

Discovering Latent Graphs with GFlowNets for Diverse Conditional Image Generation



Bailey Trang¹, Parham Saremi^{4,5}, Alan Q. Wang²,
Fangrui Huang¹, Zahra TehraniNasab^{4,5}, Amar Kumar^{4,5},
Tal Arbel^{4,5}, Li Fei-Fei¹, and Ehsan Adeli^{1,2,3}

¹Dept. of Computer Science, Stanford University, Stanford, CA, USA

²Dept. of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

³Dept. of Biomedical Data Science, Stanford University, Stanford, CA, USA

⁴Center for Intelligent Machines, McGill University, Montreal, QC, Canada

⁵MILA - Quebec AI institute, Montreal, QC, Canada



Diversity in conditional image generation

Prompt: *“Sunset scene with mountain”*

Diversity in conditional image generation

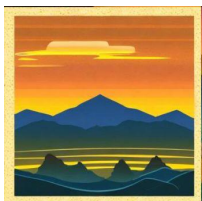
Prompt: *"Sunset scene with mountain"*

***SD: Stable Diffusion**

SD 3.5



SD 2-1



Diversity in conditional image generation

Prompt: *"Sunset scene with mountain"*

***SD: Stable Diffusion**

SD 3.5



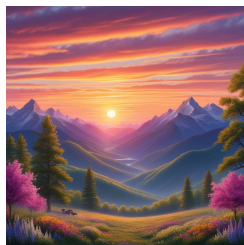
SD 2-1



Diversity in conditional image generation

A very detailed prompt still offers many variations in objects, layout, lighting, ...!

Prompt: *“A captivating sunset scene with mountains in spring unfolds in a breathtaking panorama. Majestic peaks, their rugged forms softened by a gentle blanket of fresh greenery, stand tall against the vibrant horizon. The sky transitions through a symphony of colors—from radiant oranges and soft pinks to deep purples and inky blues. The setting sun, a brilliant orb of fire, casts a golden glow that dances across the landscape, highlighting the budding foliage and the vibrant hues of blooming wildflowers. Silhouetted against this glowing canvas, the mountain ranges and valleys come alive with the fresh, new life of spring. In the foreground, a dense array of blossoming trees and wildflowers contribute vivid splashes of color, creating a striking contrast with the more muted tones of the peaks. The scene is imbued with a serene tranquility, heightened by the crisp, refreshing air and the gentle buzz of awakening nature, subtly enhancing the enchanting spring vista.”*



Diversity in conditional image generation

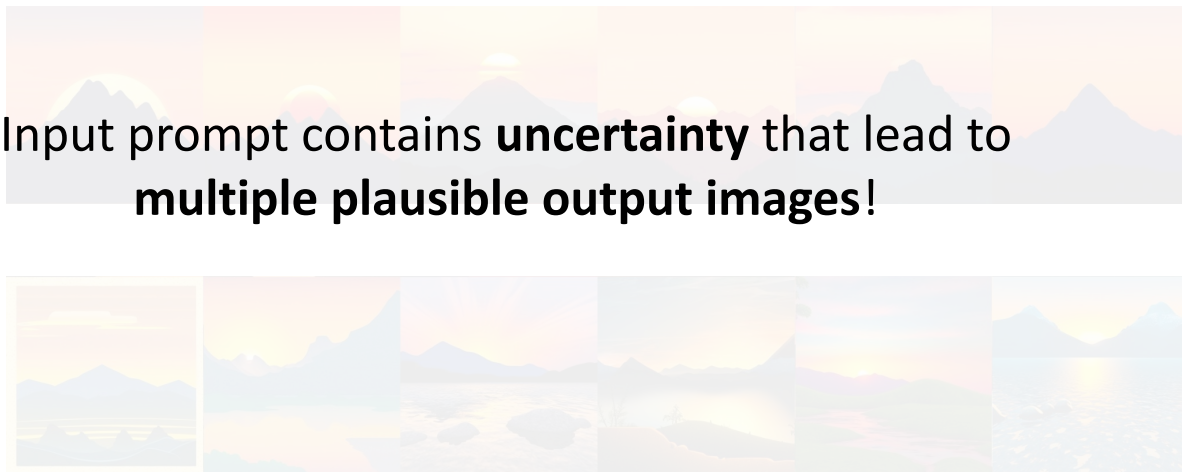
Prompt: “Sunset scene with mountain”

*SD: Stable Diffusion

SD 3.5

1. Input prompt contains **uncertainty** that lead to **multiple plausible output images!**

SD 2-1



Diversity in conditional image generation

Prompt: *"Sunset scene with mountain"*

***SD: Stable Diffusion**

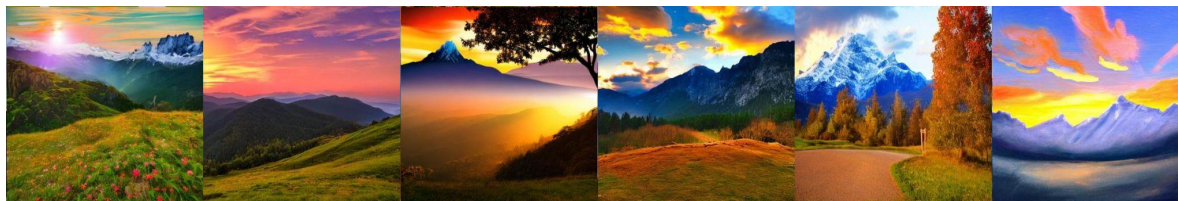
SD 3.5



SD 2-1



...how about this?



Diversity in conditional image generation

Prompt: “Sunset scene with mountain”

*SD: Stable Diffusion

SD 3.5

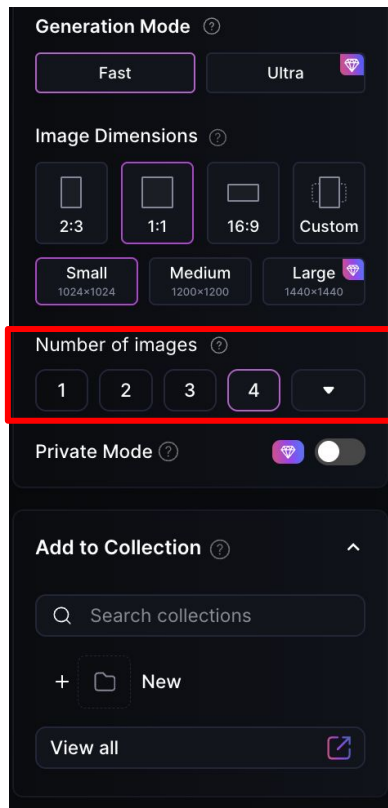
1. Input prompt contains **uncertainty** that lead to **multiple plausible output images!**

2. **Uncertainty** extends beyond object and layout choices to include **commonsense and domain-specific knowledge understanding.**

SD 2-1

...how about this?

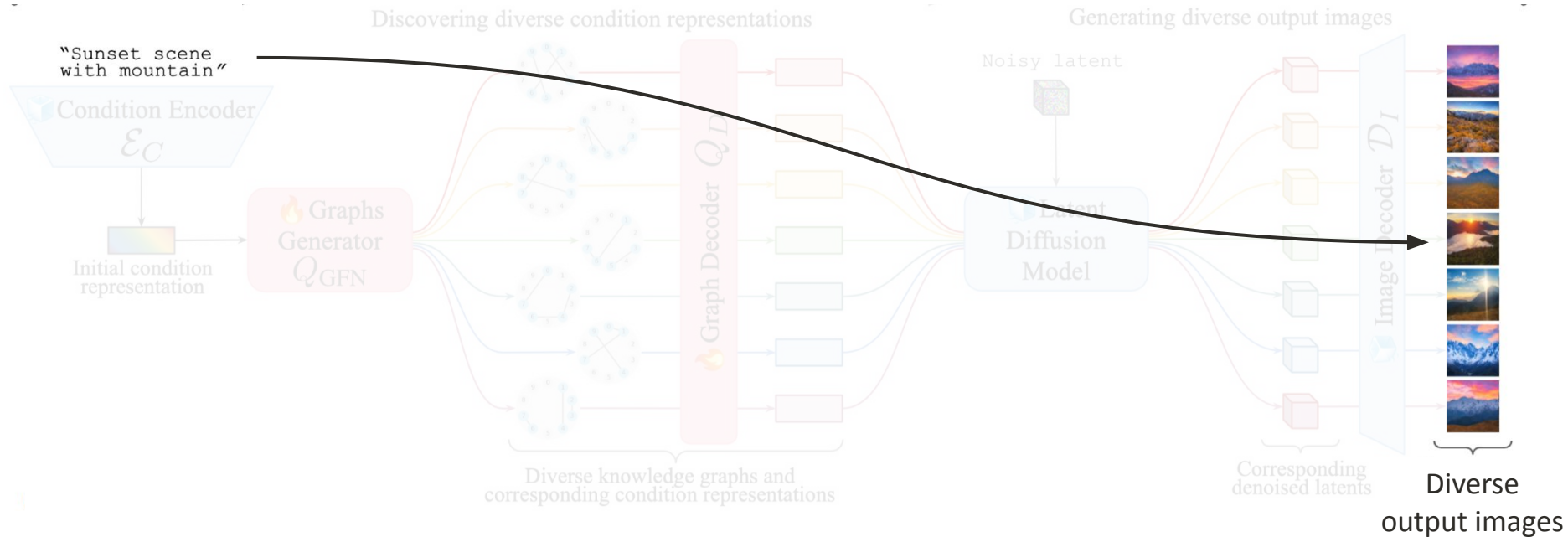
Diversity in conditional image generation



~~Generate diverse images~~

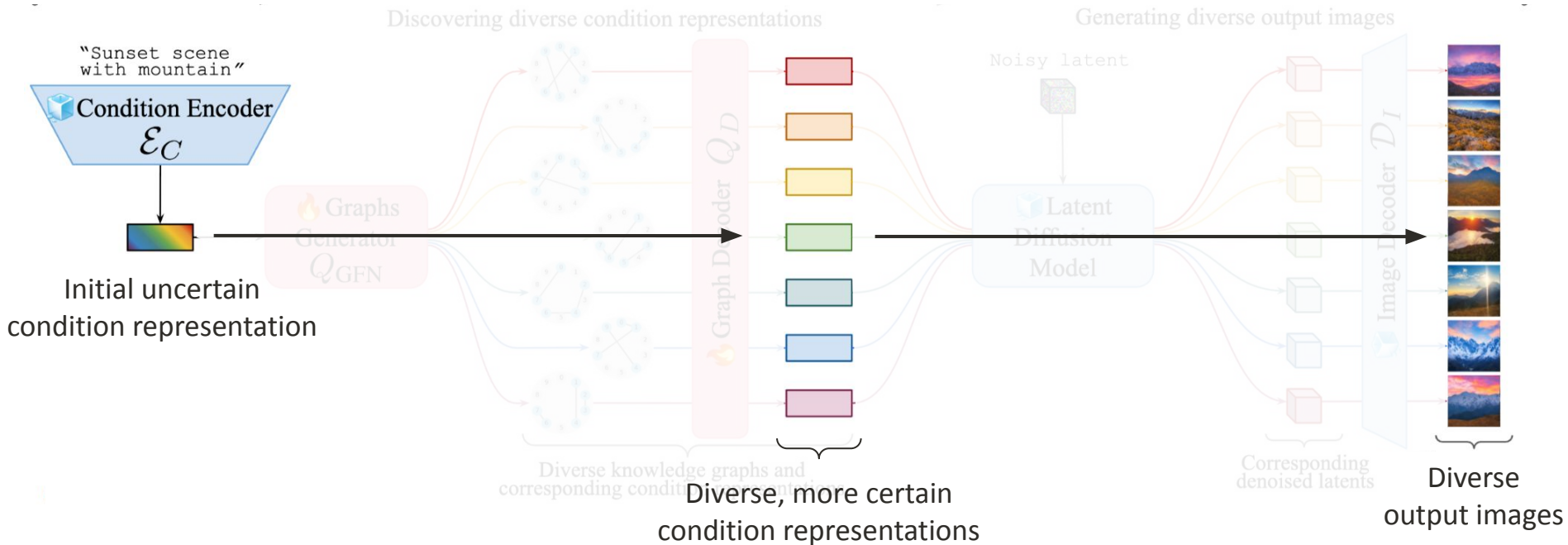
Repeat the generation process multiple times
(still have a high chance to fall into some
popular modes, not capturing diversity)

Rainbow



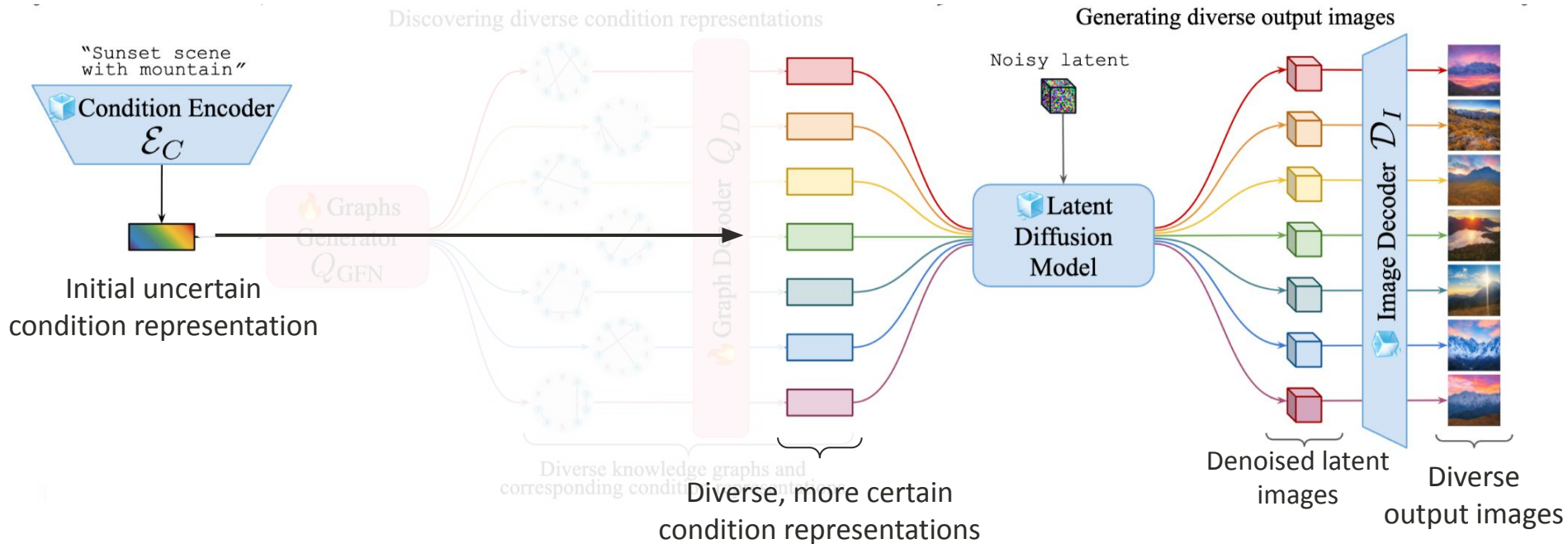
Rainbow generates diverse images at once, collectively reflect the input prompt.

Rainbow



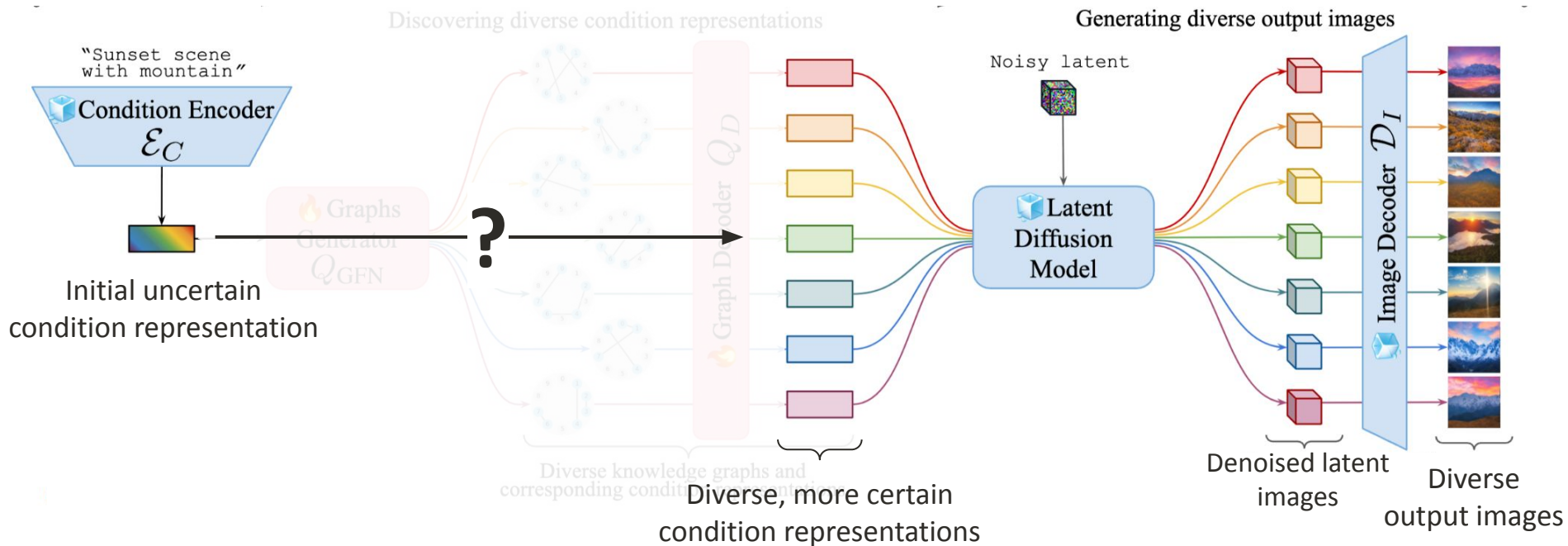
Rainbow decomposes the original condition into diverse representations.

Rainbow



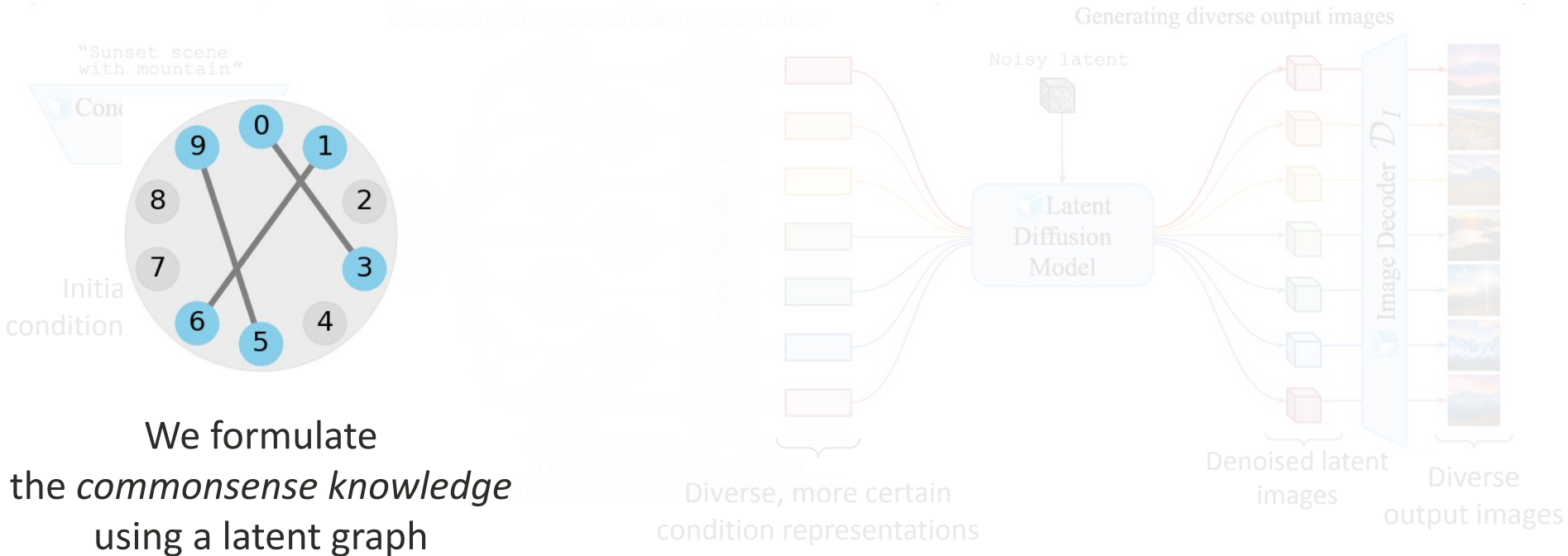
Diverse condition representations generate diverse output images.

Rainbow



Diverse condition representations generate diverse output images.

Rainbow

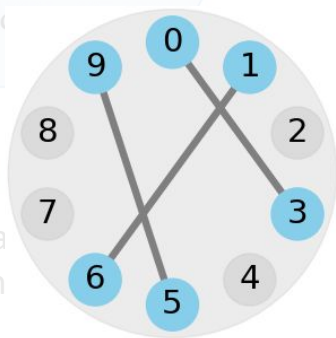


Rainbow

"Sunset scene
with mountain"

Context

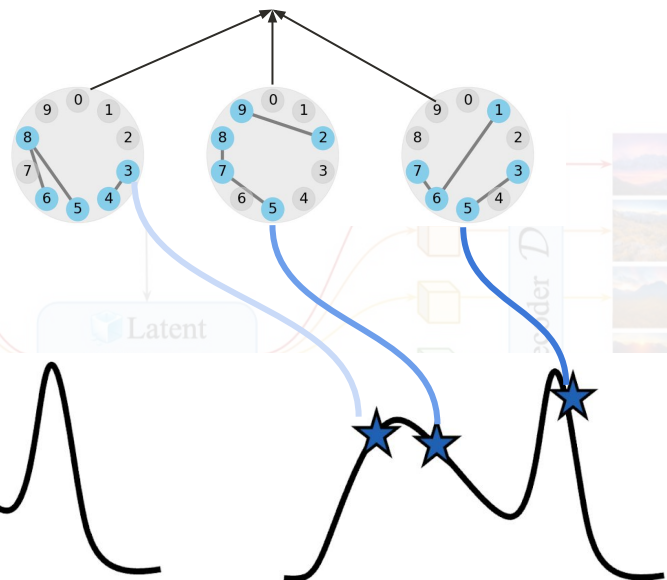
Initial
condition



We formulate
the *commonsense knowledge*
using a latent graph

Rainbow employs
GFlowNets to sample
multiple sub-graphs
that collectively
represent a common
context!

Some given context



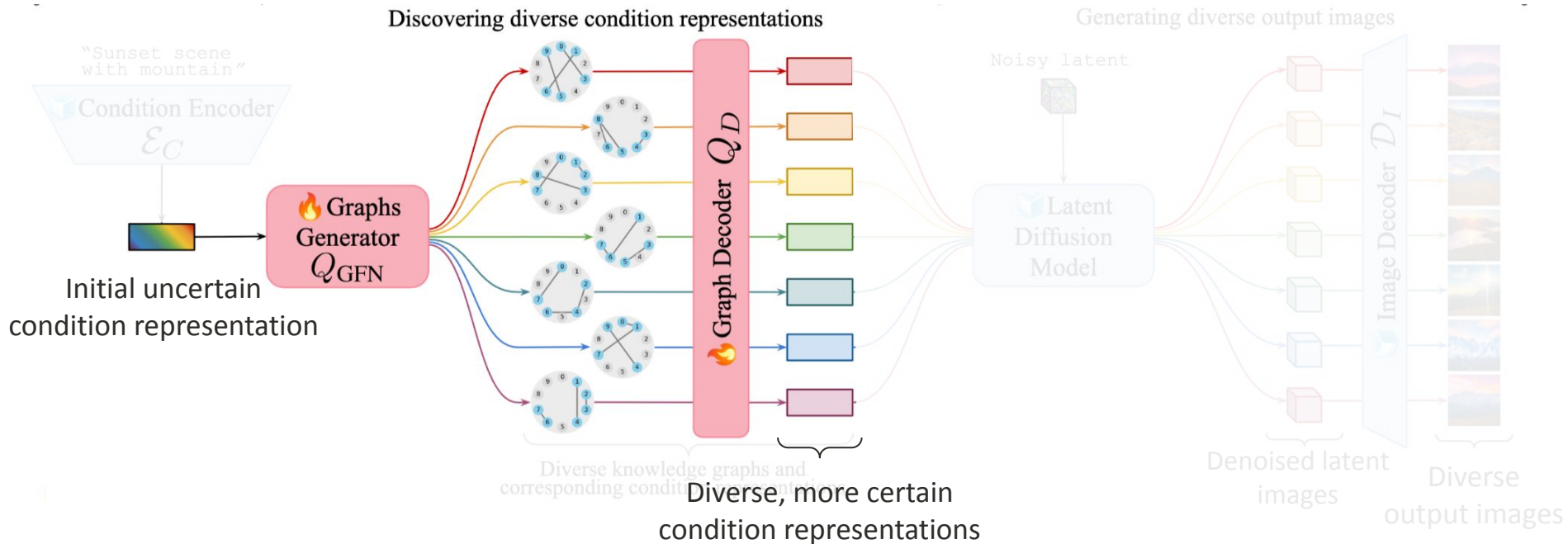
Reward-Maximization
(**Converge to a single mode**)

Reward-Matching
(**Cover Multiple Modes**)

GFlowNets algorithm

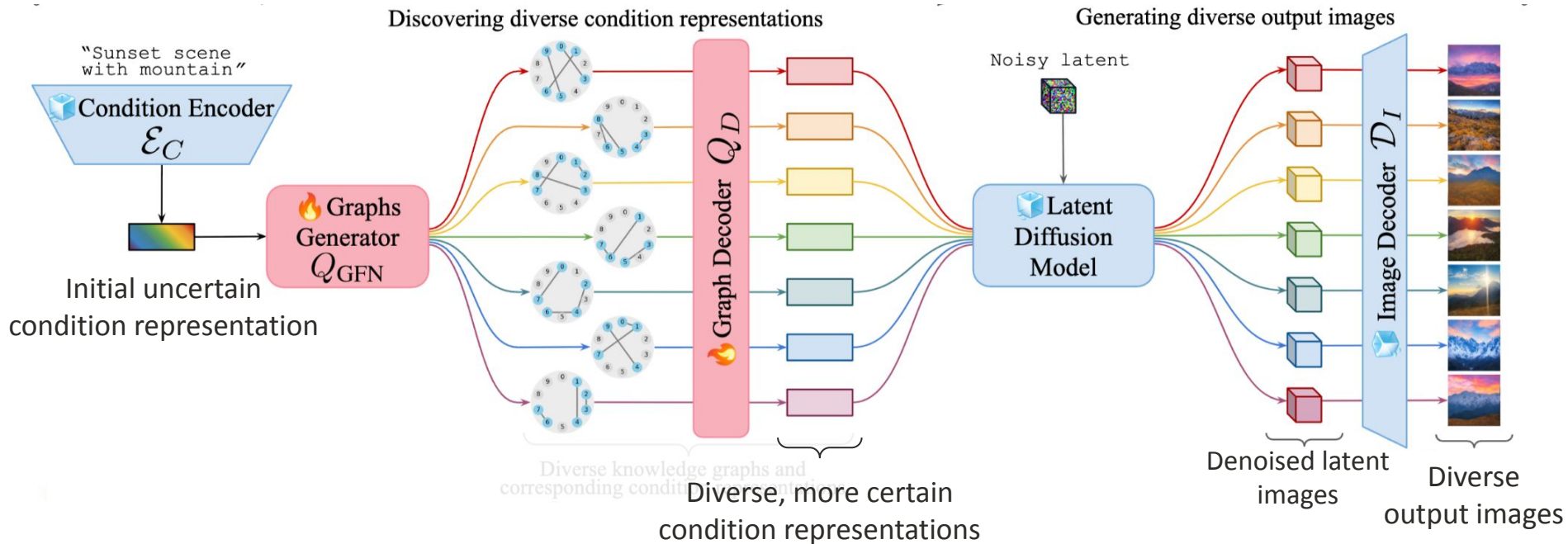
*GFlowNets: [Generative Flow Networks](#)

Rainbow



Input condition $\rightarrow$ Diverse subgraphs $\rightarrow$ Diverse condition representations

Rainbow



Input condition $\rightarrow$ Rainbow $\rightarrow$ Diffusion model $\rightarrow$ Diverse output images

Qualitative results - Diverse image generation

Rainbow generates diverse layouts, captures multiple seasons!

"Sunset scene with mountain"

"Sunset scene with mountain in a specific season"

SD3.5 [62]

SD2-1 [62]

CADS[67]

PAG [84]

Rainbow (Ours)



Qualitative results - Diverse image generation

...even in prompt that does not mention *the "season" cue!*

"A forest with animals"

SD3.5



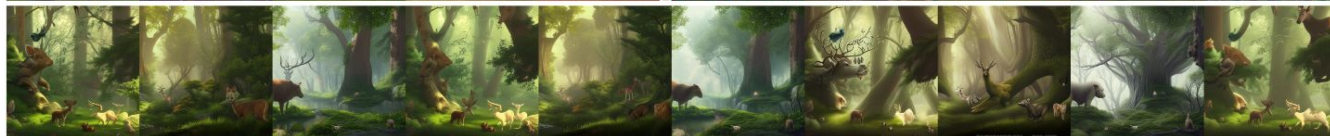
SD2-1



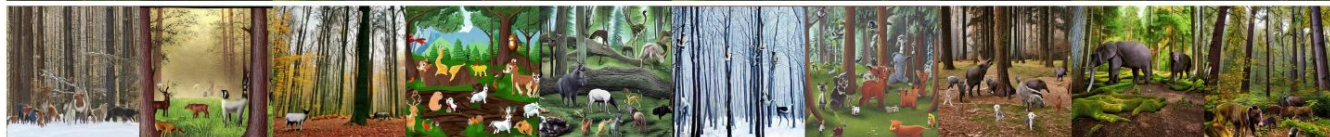
CADS



PAG

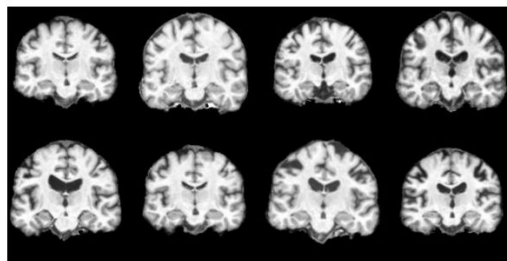


Rainbow

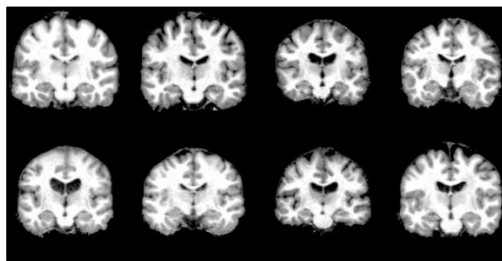


Qualitative results - Diverse image generation

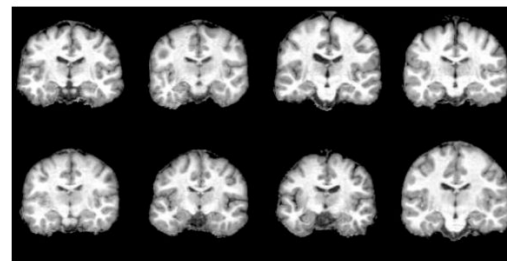
Rainbow can capture the vary ventricle size in MRIs of people in 60s!



(a) Actual samples from dataset



(b) Generations by **Rainbow**



(c) Generations by baseline LDM

Figure 5: Comparison of MRI image generations for 65-year-old male individual. Compared to actual samples from males aged 63-65, **Rainbow** captures greater diversity in details like ventricle sizes, while the baseline LDM generates images with less variation.

Qualitative results - Diverse image generation

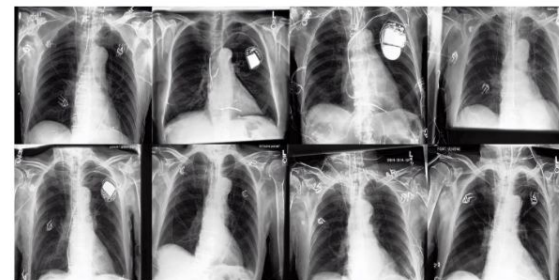
Rainbow can generate diverse support devices!



(a) Generations by RadEdit



(b) Generations by SD1.5

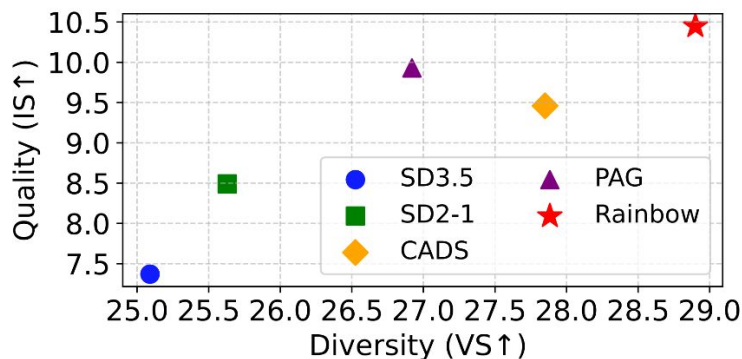


(c) Generations by Rainbow

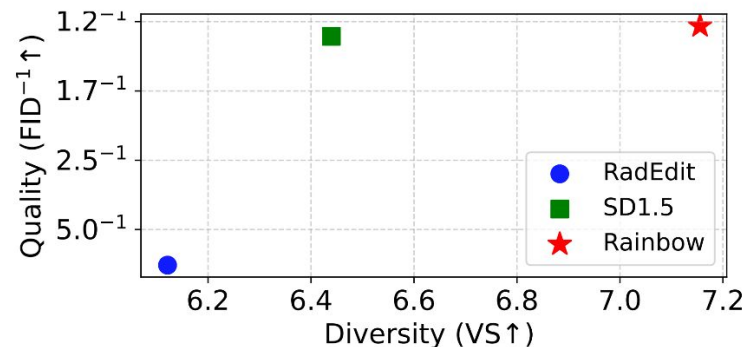
Figure 6: Comparison of chest X-ray image generations for the prompt *Chest X-ray with support devices*. Rainbow is able to provide high-quality images while generating a more diverse set of medical devices compared to the baselines - (a) and (b).

Quantitative Results - Diversity vs Quality

Rainbow generate diverse high-quality images!



(a) Quantitative comparison on Natural Images

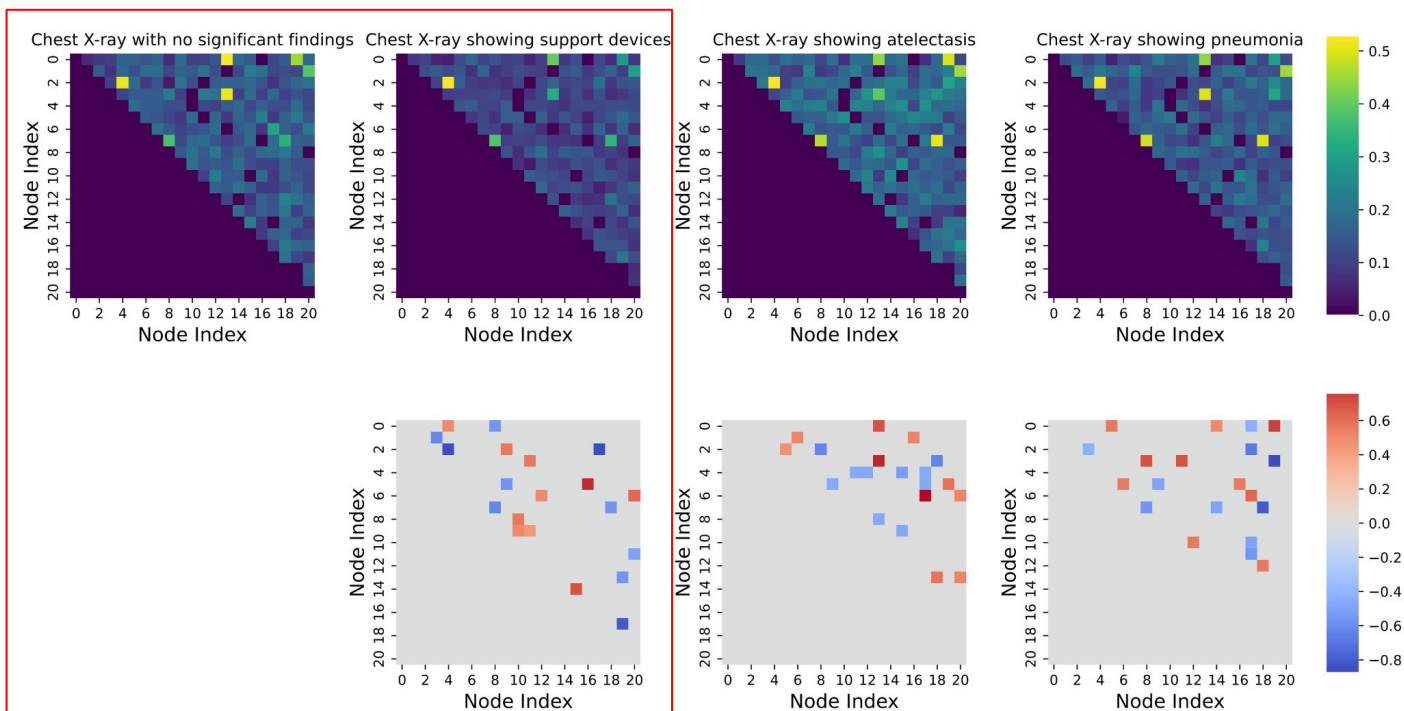


(b) Quantitative comparison on Chest X-rays

Figure 4: **Quantitative analysis on diversity and image quality of SD-based Rainbow.** Rainbow consistently outperforms SD baselines in diversity with higher Vendi Score (VS) across domains and image quality with higher Inception Score (IS) in natural images and FID^{-1} in chest X-rays.

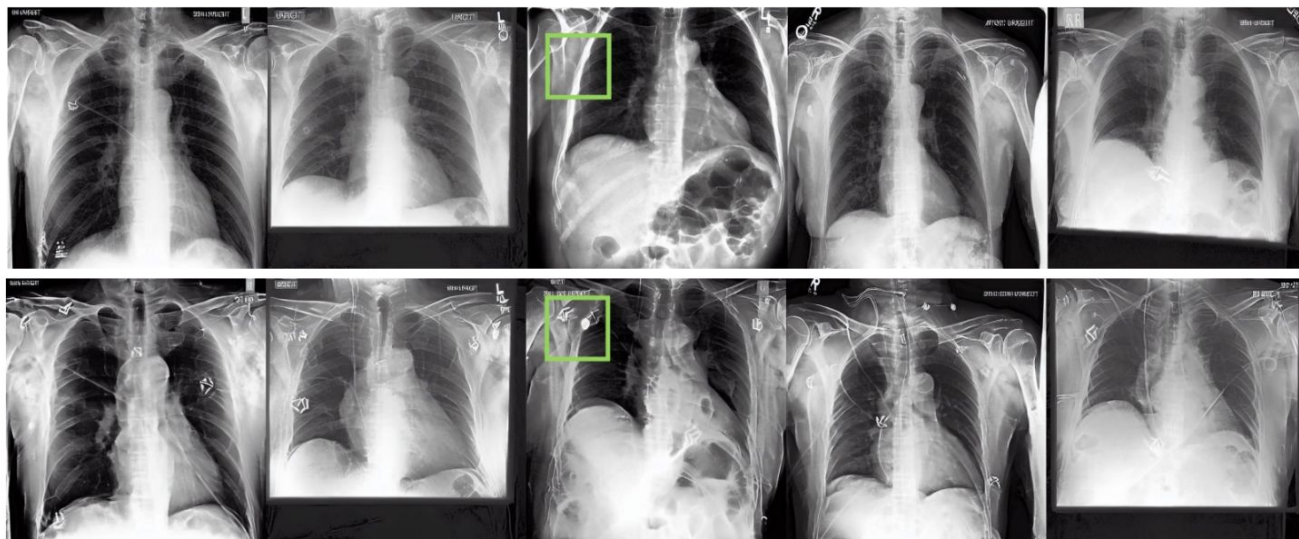
Latent graph interpretation

Set of edges can be interpret to meaningful features!



Latent graph interpretation

Generations with added “support devices” edges introduce devices!



Latent graph interpretation

Generations with added *season* edges introduce clear seasonal features!



Base prompt: *"Sunset scene with mountain"*



Base prompt + **Spring** edges



Base prompt + **Winter** edges



Thank You

Contact:

Ehsan Adeli (eadeli@stanford.edu)
Bailey Trang (trangn@stanford.edu)