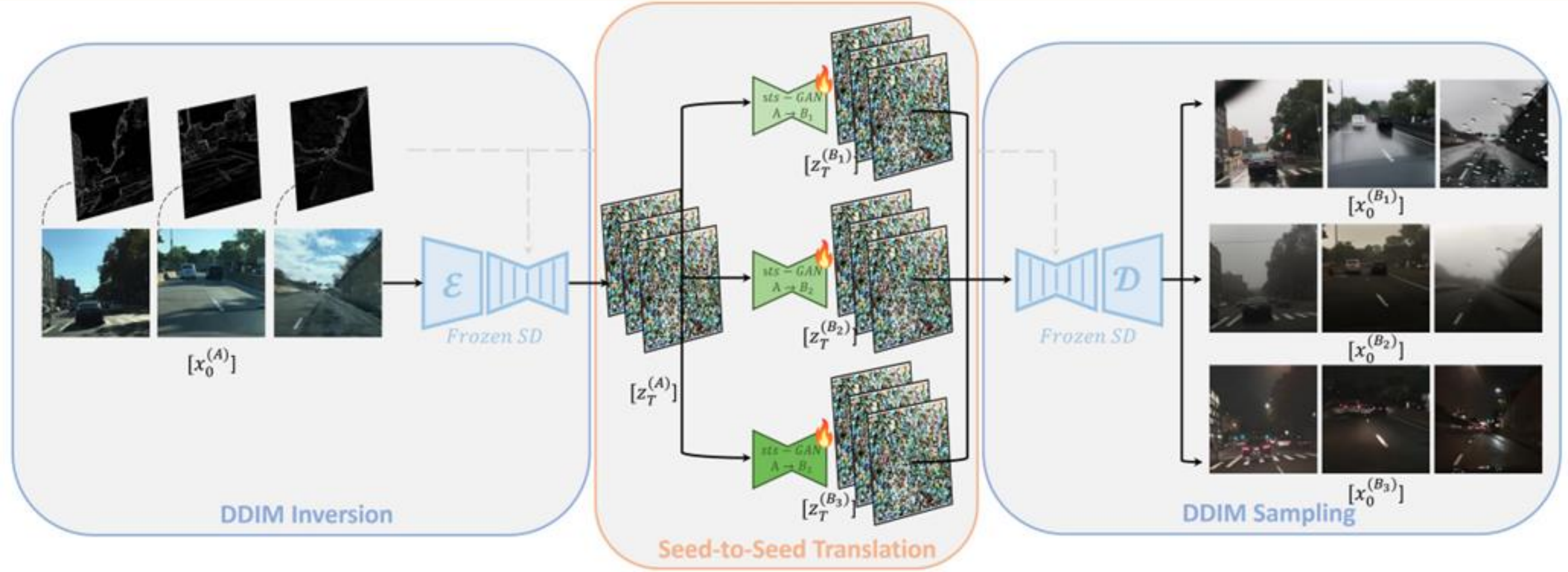
Seed-to-Seed:

Unpaired Image Translation in Diffusion Seed Space

BMVC 2025

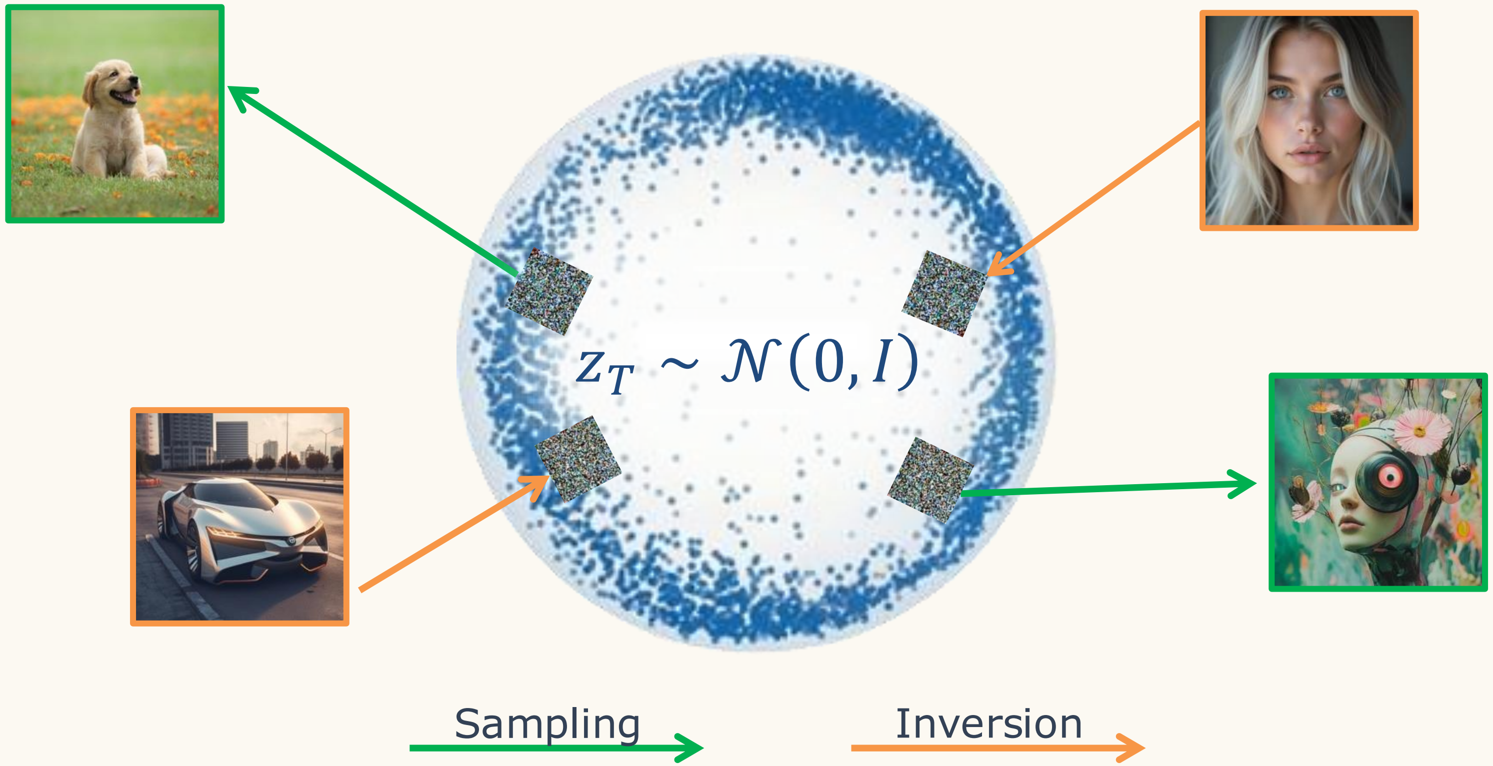


Or Greenberg, Eran Kishon, Dani Lischinski

The Seed Space

Seed-Space = The initial latent noise space

the high-dimensional vector space from which the initial random noise vector (the "seed") is sampled to initiate the diffusion process



Is the Seed-Space **Structured** and **accessible**?

Structured

Is the information encoded in the seed-space well structured?

Accessible

Is the information encoded in the seed accessible?

Can the seed-space be utilized for image manipulation, before the sampling process even starts?

Motivation: In-Seed-space vs. sampling-trajectory Manipulation

Along-Trajectory Manipulation

- **Inherent Conflict:** Generation/Edit instructions may conflict with information inherently encoded in the initial latent
- **Multi-Space Complexity:** Edits must span across multiple latent spaces at every sequential denoising step, making it highly complicated to disentangle features.

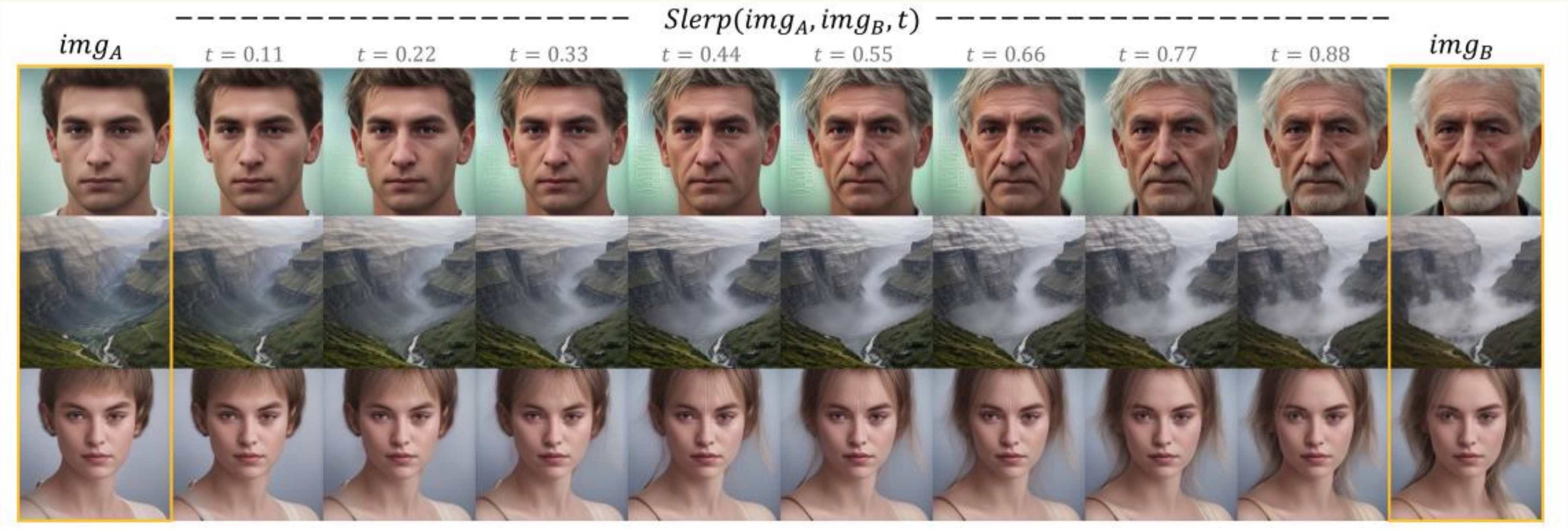
In-Seed-Space Manipulation

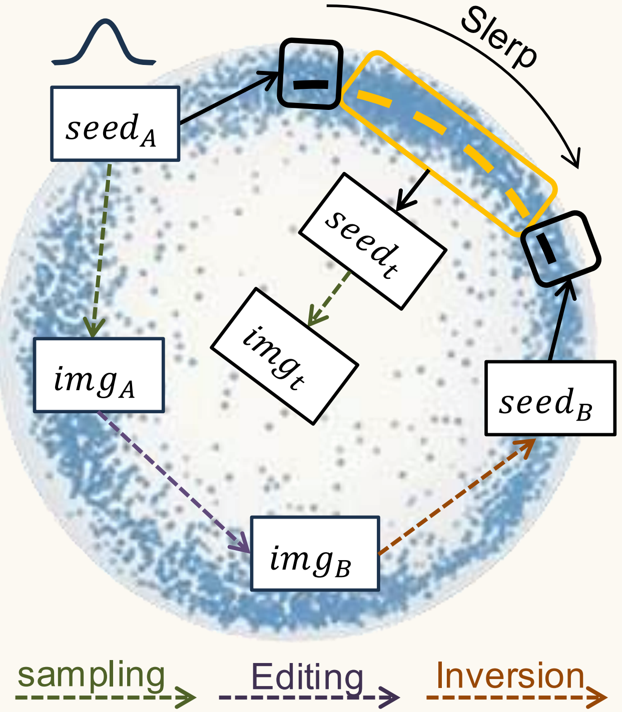
- **Seed/target alignment**: The information encoded in the initial seed fully aligns with the along-trajectory instructions
- **Single-Space Efficiency**: Bypasses step-by-step editing by operating entirely in a single, initial noise space prior to sampling.

In-Seed-Space manipulation is possible only if the seed-space is both <u>Structured</u> and <u>Accessible</u>

The Meaningful Seed-space

EXP #1: Structured Seed-Space

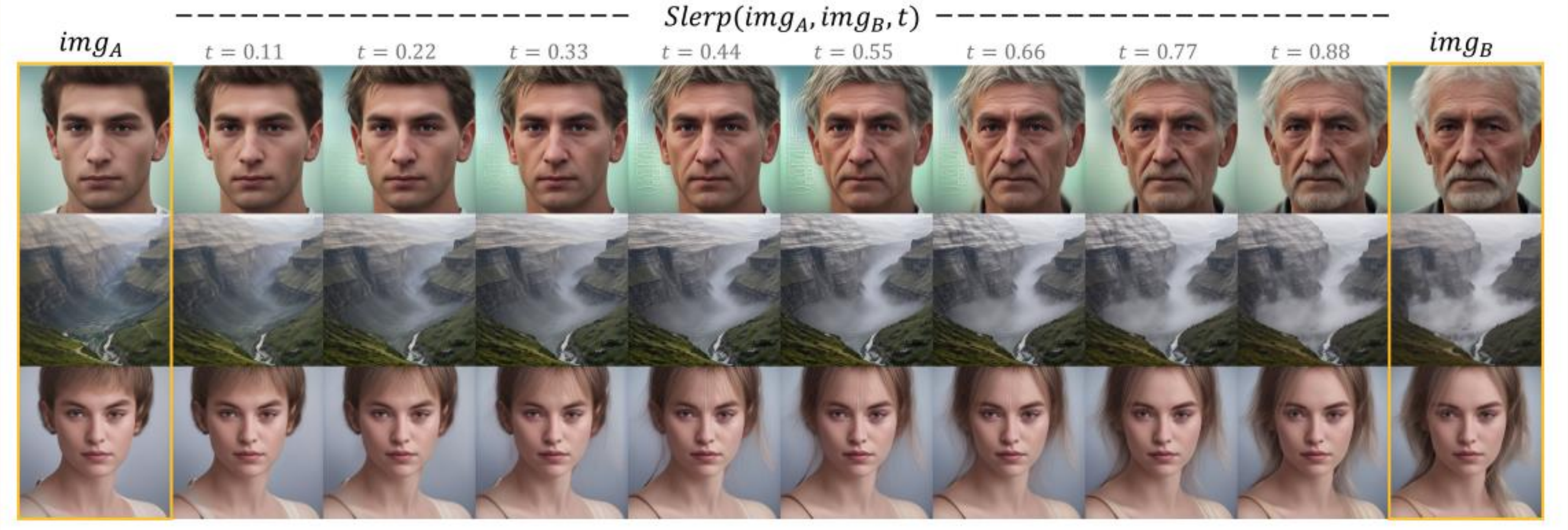


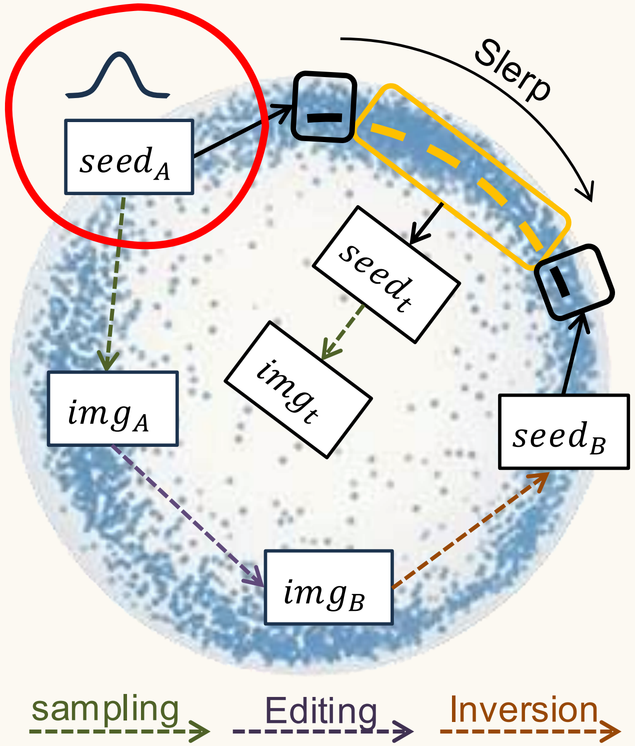


Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

EXP #1: Structured Seed-Space

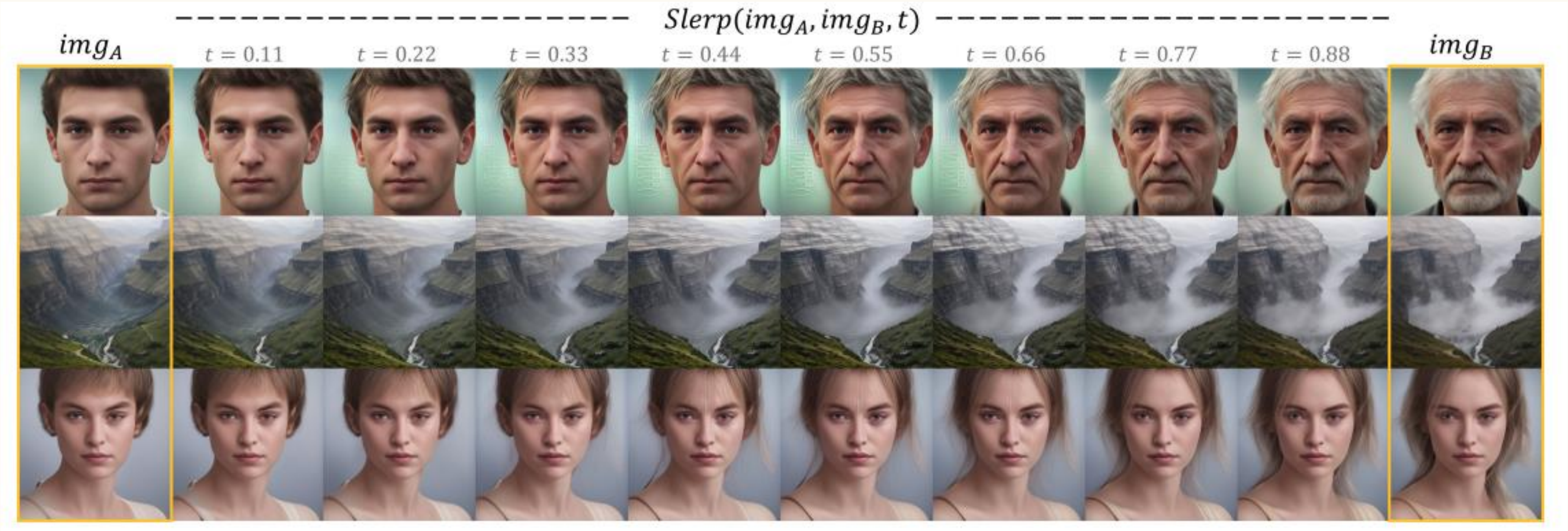


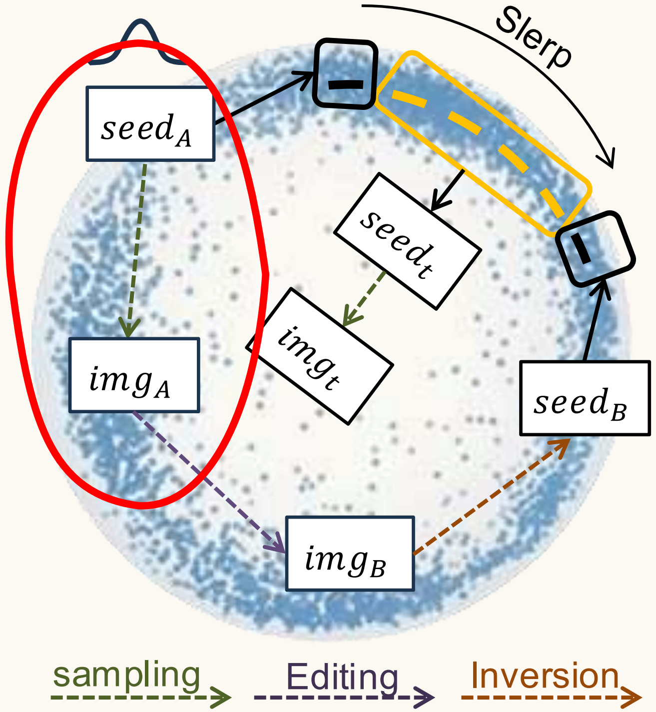


Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

EXP #1: Structured Seed-Space

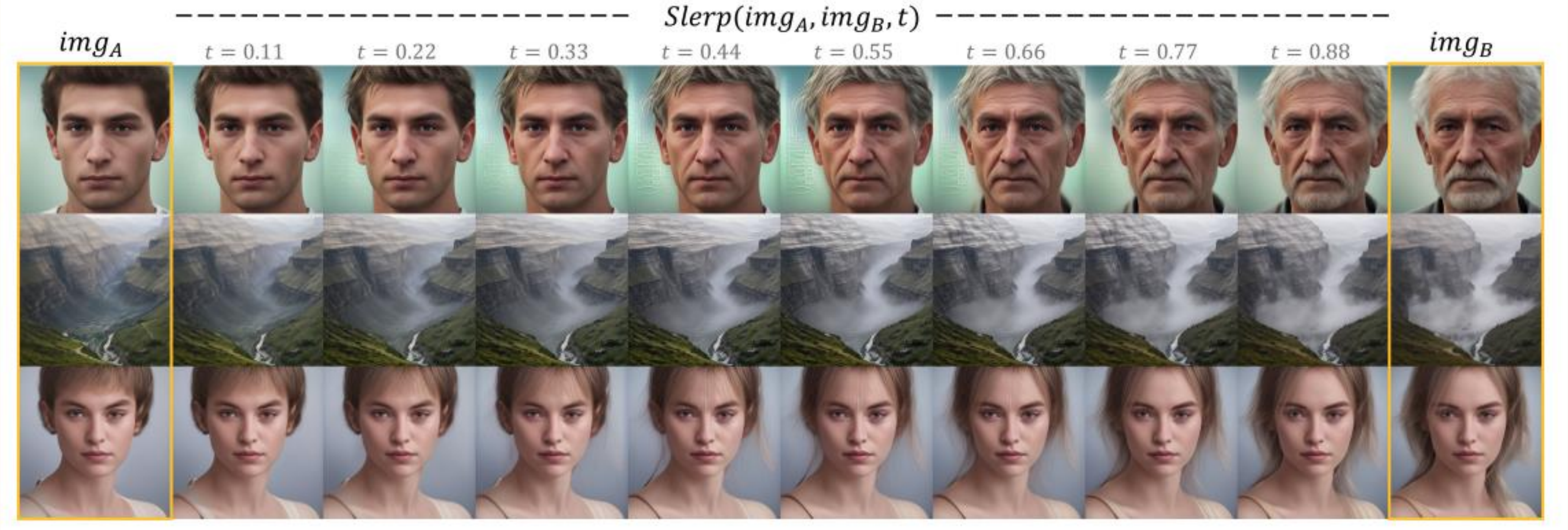


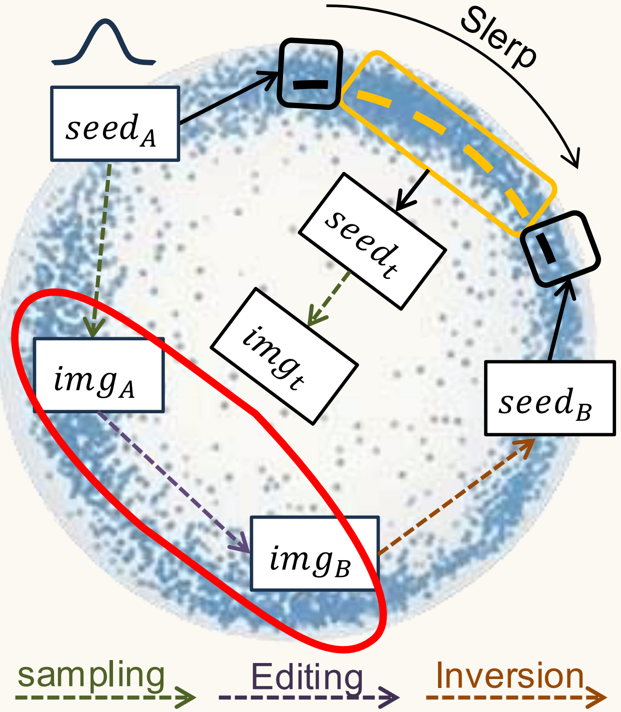


Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

EXP #1: Structured Seed-Space

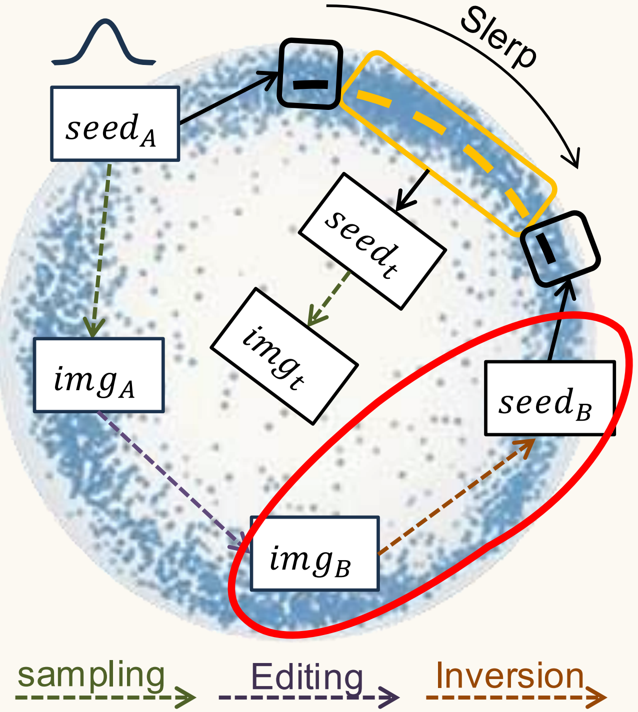




Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

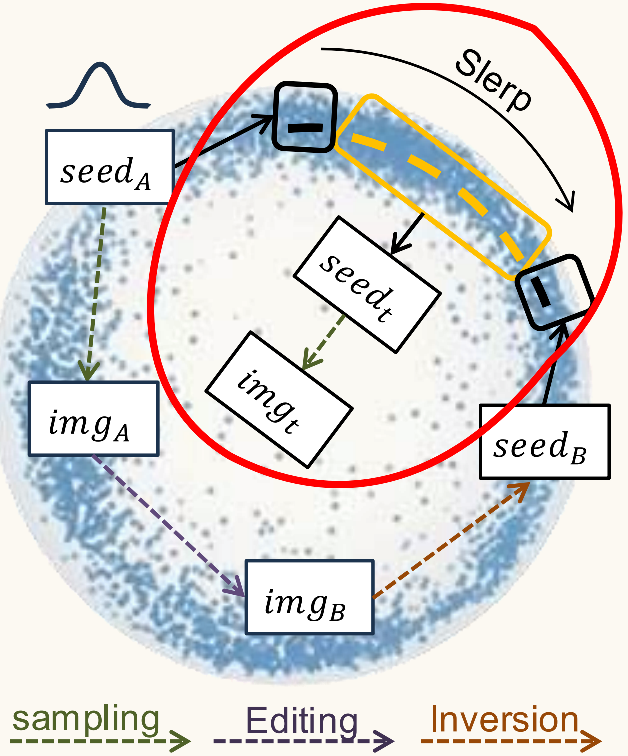
EXP #1: Structured Seed-Space



Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

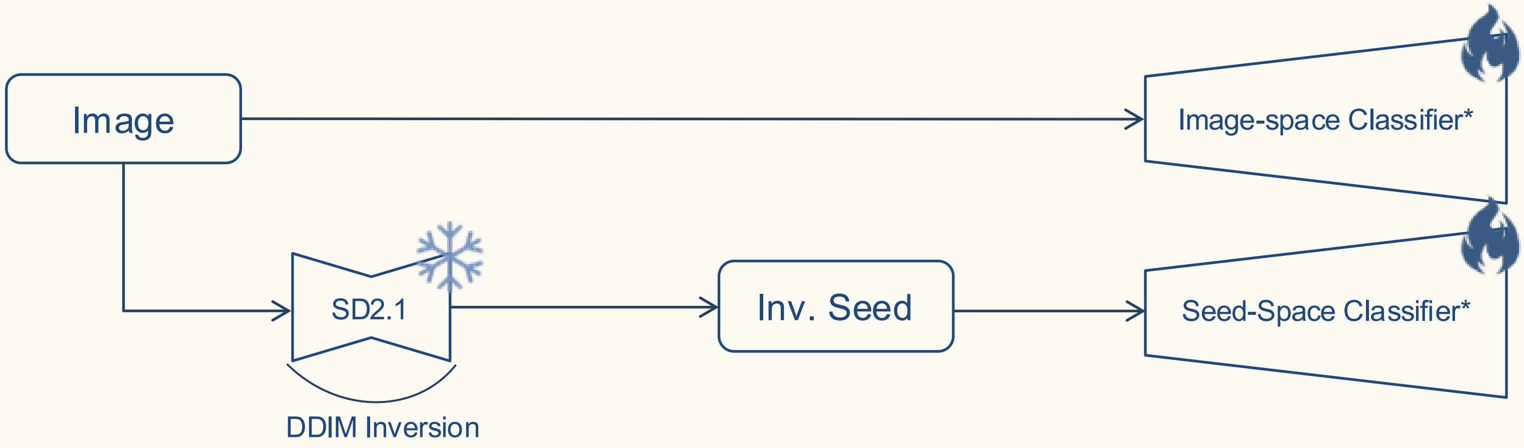
EXP #1: Structured Seed-Space



Spherical interpolation between $seed_A$ and $seed_B$ yialds a semantically coherent transformation between the corresponding images img_A and img_B

The Meaningful Seed-space

EXP #2: Accessible Seed-Space



Task	seeds	images
Day/Night	98.37%	98.47%
Cat/Dog	90.10%	98.53%
Older/Younger	92.60%	97.90%

* Shared architecture (ResNet18)

Seed-derived classifiers perform nearly as well as image-derived ones, confirming that a simple network can pull significant semantic information across scene (time of day), object (dog/cat), and sub-object (age) level attributes directly from the inverted seed.

Unpaired Image Translation

Structural Fidelity

- Complex scenes (e.g., automotive driving datasets)
- strict adherence to source semantics and geometry

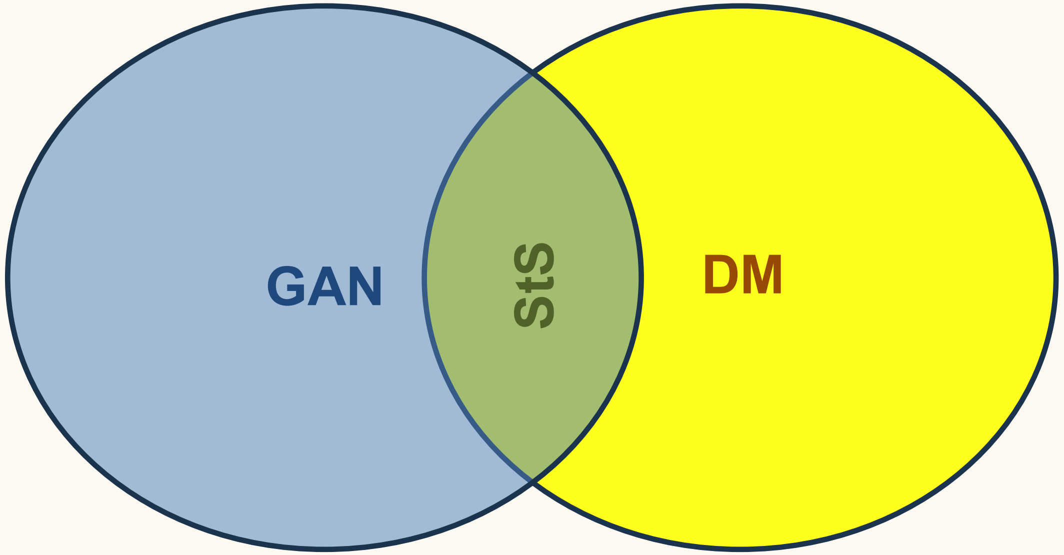
High Quality

- Natural image manifold
- Complex target domain effects

The Paradigm Gap

GANs: Strong structural fidelity via cycle-consistency **but** suffer from stochastic artifacts and lack realism.

Diffusion: High-quality texture generation capabilities and semantic understanding **but** prone to structural drift during the sampling trajectory.



Proposed Framework: Seed-to-Seed (StS)

We propose a hybrid framework combining GAN-based Seed-translation with diffusion-based generation, **leveraging their complementary strengths**.

1

Inversion

DDIM inversion of source image $x_0^{(A)}$ to establish seed $z_T^{(A)}$.

2

Translation

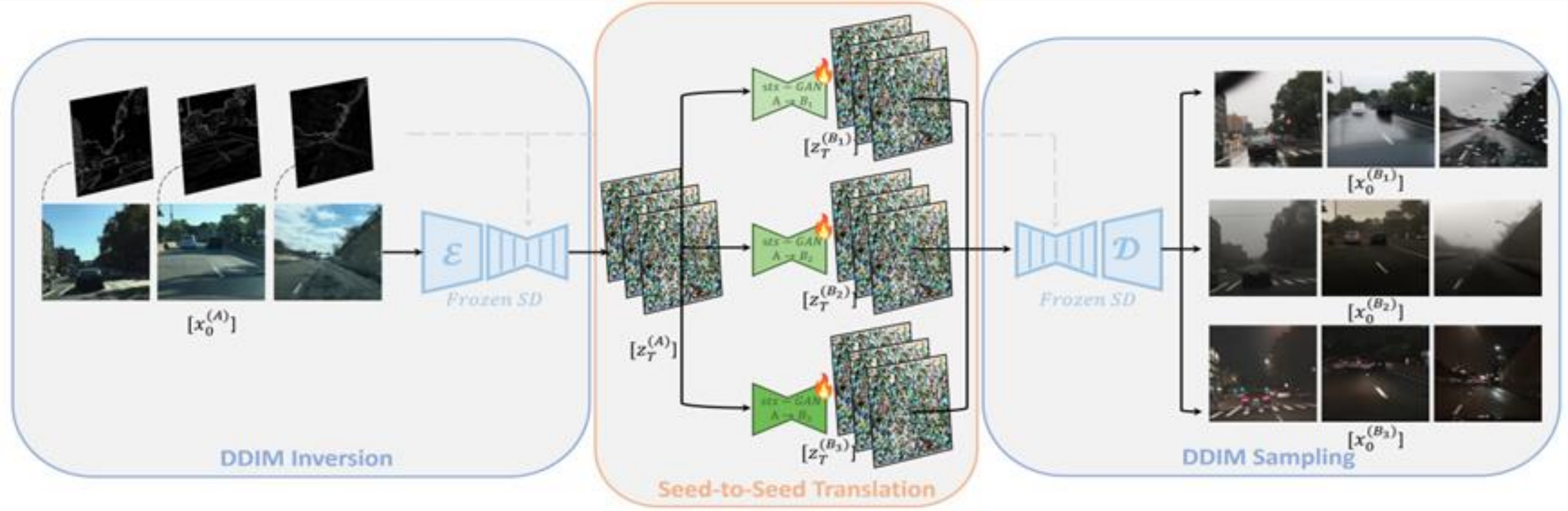
sts-GAN maps source seed $z_T^{(A)}$ to target seed $z_T^{(B)}$

3

Sampling

Sample $z_T^{(\mathrm{B})}$ via a pre-trained diffusion backbone to $x_0^{(\mathrm{B})}$

- A ControlNet-like spatial constraint is used at sampling time to prevent trajectory drift and to enable high guidance-scale usage



Configuration	FID ↓	SSIM ↑
ControlNet+RS	35.091 (+114%)	0.493 (+2%)
ControlNet+IS	49.572 (+202%)	**0.756 (-49%)**
ControlNet+IS+ST (StS)	**16.384**	0.505

Results

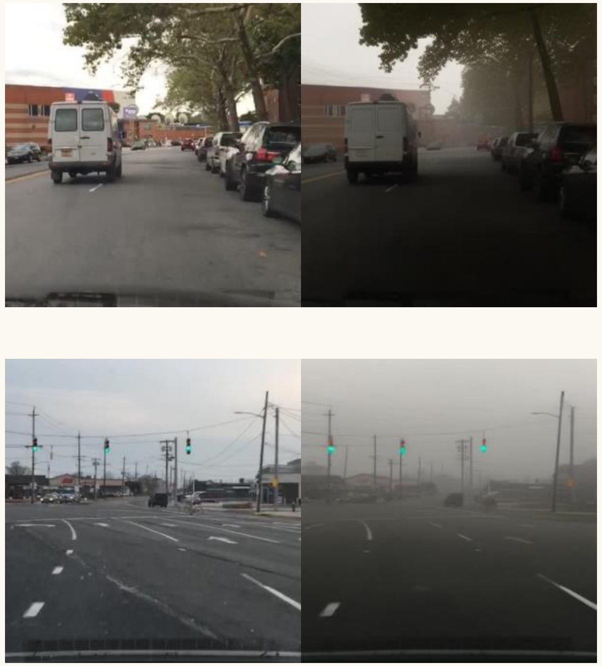
	GAN baselines					Diffusion baselines				StS (ours)
	CycleGAN	MUNIT	TSIT	AU-GAN	CycleGAN-turbo	SDEdit 0.5	SDEdit 0.7	PnP	ControlNet	
FID ↓	19.908	52.152	21.315	14.426	16.840	73.494	48.757	61.617	35.091	16.384
MMD ↓	58.395	260.081	56.484	45.970	49.845	242.001	161.666	172.808	95.171	41.344
KID ↓	4.539	12.968	4.446	3.985	4.215	12.097	9.185	9.575	6.340	3.718
SSIM ↑	0.469	0.308	0.3929	0.463	0.431	0.661	0.603	0.768	0.493	0.505

GANs

DMs

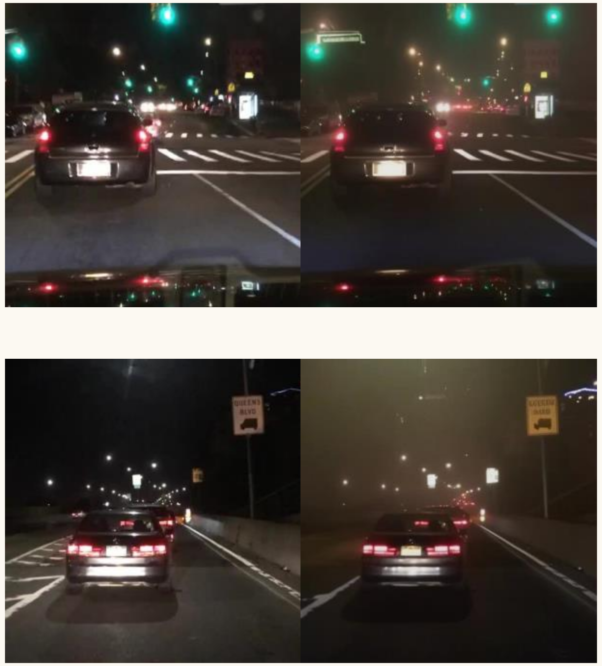
Results: Other Domains

Clear Day
↓
Foggy Day

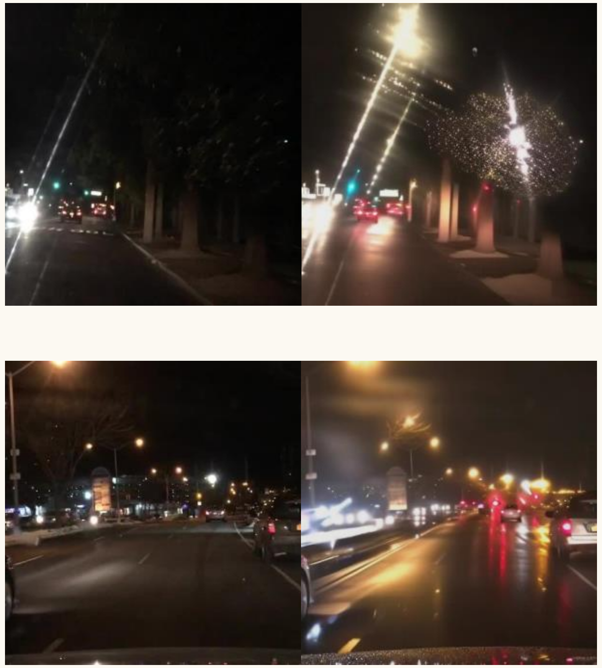


Clear Day
↓
Rainy Day

Clear Night
↓
Foggy Night



Clear Night
↓
Rainy Night



Results: Non-Automotive

Male → Female

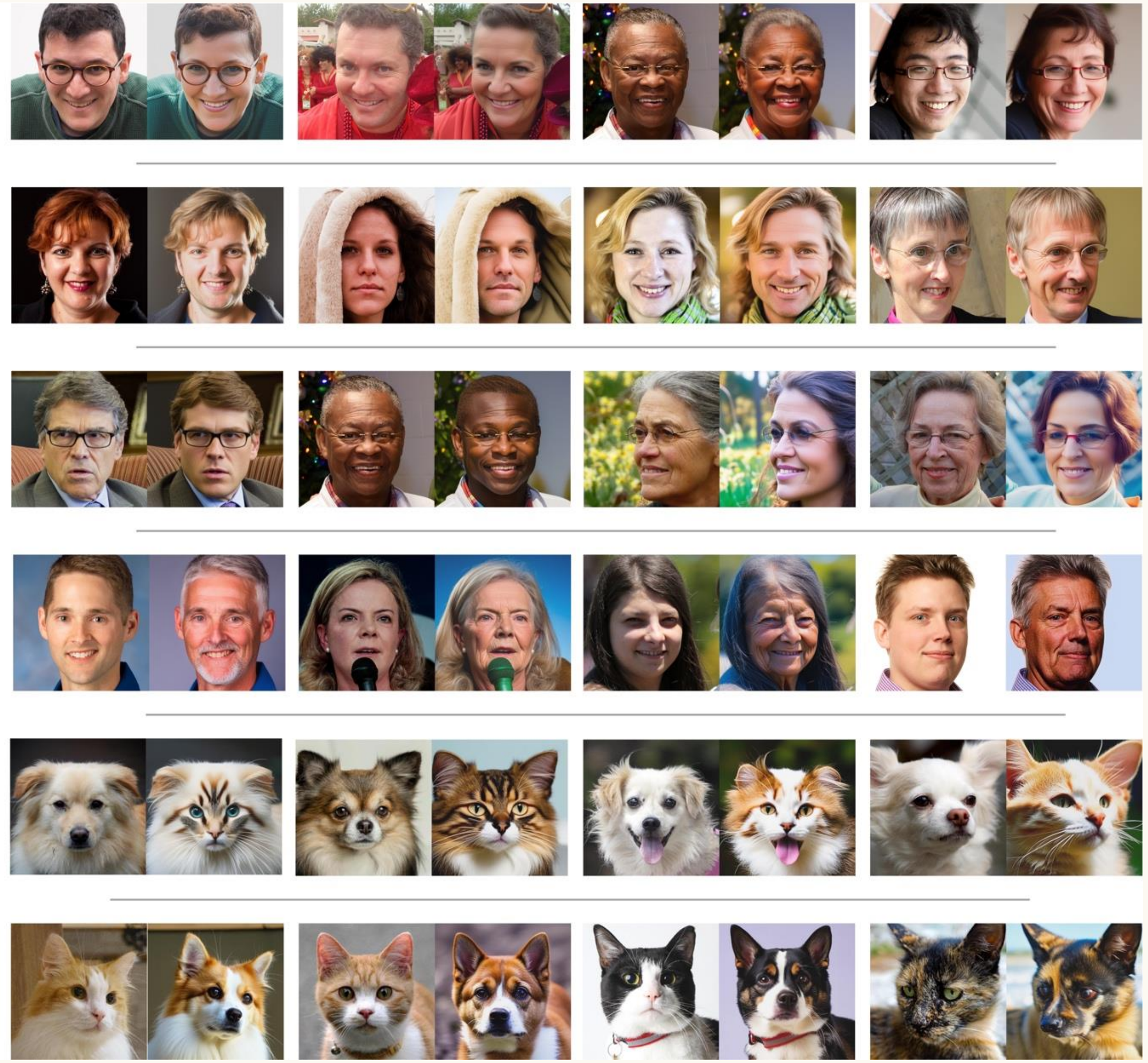
Female → Male

Older → Younger

Younger → Older

Dog → Cat

Cat → Dog



TL;DR

- **The Meaningful "Seed-Space".**
  The space of initial latents of pre-trained LDM has a meaningful structure and is accessible using a simple trainable network

- **The Manipulable "Seed-Space".**
  We present StS-GAN. A CycleGAN-based Unpaired image-to-Image Translation model, via **Seed-to-Seed Translation.**

- **The Best of The Two Paradigms .**
  Our Seed-to-seed approach proposes a hybrid framework combining GAN-based translation with diffusion-based generation, leveraging their complementary strengths.

For more details, results and comparisons please view our project page

contact: or.greenberg@gm.com

Thank You!

And BTW, We (General Motors) are hiring!!!

Questions?