Seed-to-Seed: Unpaired Image Translation in Diffusion Seed-Space

Or Greenberg, Eran Kishon, Dani Lischinski

THE HEBREW UNIVERSITY OF JERUSALEM

RESEARCH & DEVELOPMENT

SEED-TO-SEED? → TL;DR

Task & Motivation

Task: Unpaired Image translation, with explicit focus on cases that require a close adherence to the source image's semantics.

Common Alternatives:

- GAN-based methods- Content preservation is enforced via cycle consistency, but image appearance is usually of low quality.
- DM-based methods- High quality image generation, but content preservation is not explicitly enforced.

Challenge: combine the fidelity enforced by cycle consisteny in GANs with the high quality, yet iterative, Diffusion Models.

Demonstrated Applications:

- Time-of-day & weather adjustments of complex Automotive scense (Datasets: BDD100K, DENSE)
- Age & gender adjustments of human faces (Dataset: FFHQ)
- Traditional cat ↔ dog swap (Dataset: AFHQ)

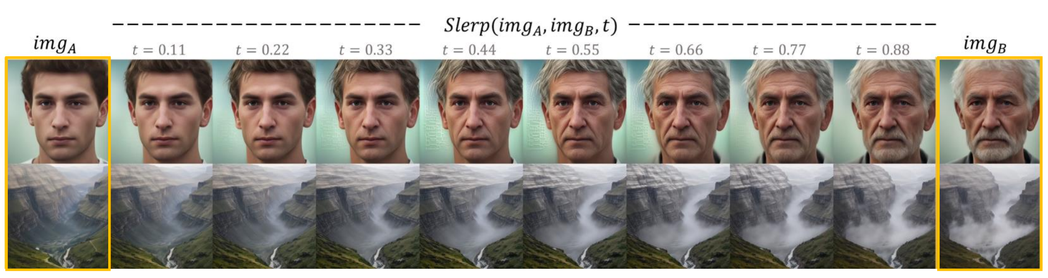
Main Innovation

- The "seed-space" of DMs has a **meaningful structure** which enables Image manipulation **within the seed space**, rather than along the sampling trajectory.
- We present **StS-GAN**: A CycleGAN-based Unpaired image-to-Image Translation model, via **Seed-to-Seed Translation.**
- A hybrid framework combining GAN-based translation with diffusion-based generation, **leveraging their complementary strengths**.

What's Next?

- Enhanced seed-editing techniques.
- Flexible content-preservation mechanism along the sampling trajectory.
- Control the Intensity of the target domain (e.g., fog density).

MEANINGFUL SEED SPACE



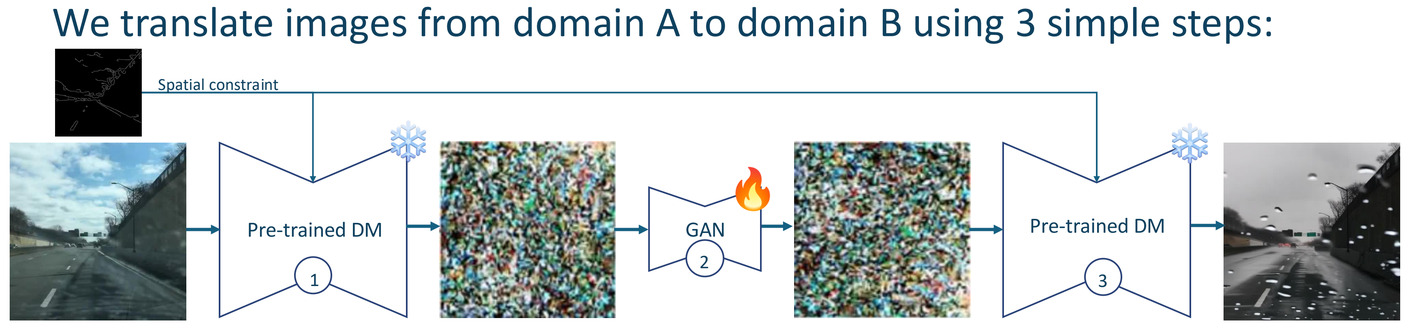
Spherical Interpolation (Slerp) between DDIM-inverted seeds from img_A and img_B yields a semantically coherent transformation between the images.

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

Spherical Interpolation (Slerp) between DDIM-inverted seeds from img_A and img_B yields a semantically coherent transformation between the images.

StS-GAN: METHOD

We translate images from domain A to domain B using 3 simple steps:



1. **inverting** a source image to a meaningful "*seed*"
2. **Translating** the "*seed*" using *task-optimized GAN to achieve target domain attributes*
3. **Sampling** *the manipulated "seed" with a spatially-constrained DM to generate the target image*

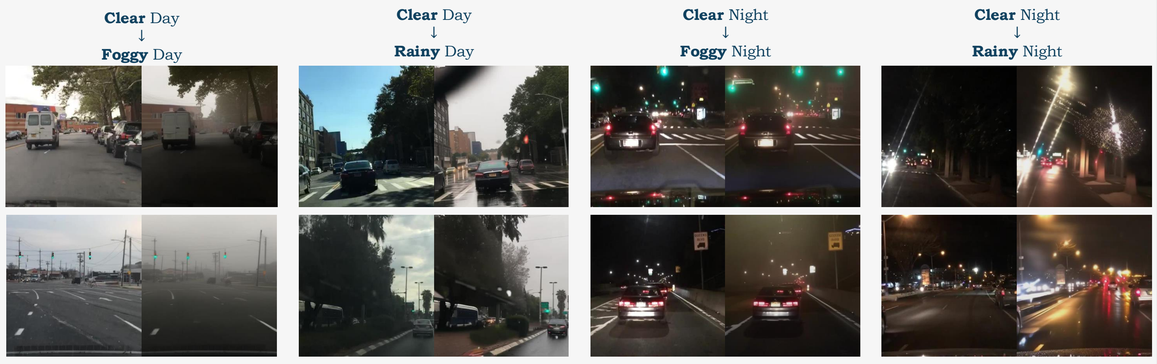
RESULTS

Quantitative comparison of {Source_Day}→{Clear_Night} translation (BDD100K)

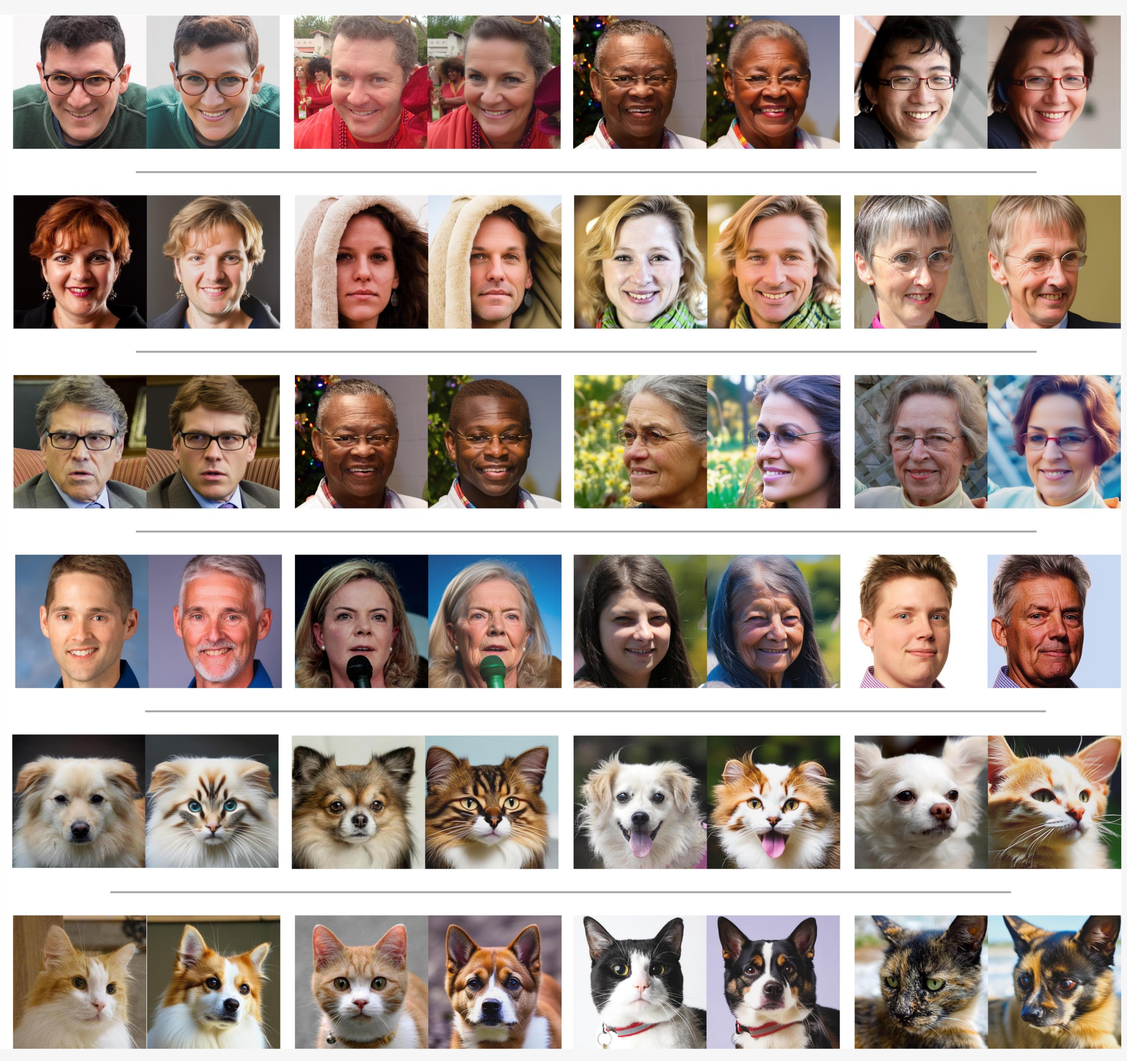
	GAN baselines					Diffusion baselines				StS (ours)
	CycleGAN	MUNIT	TSIT	AU-GAN	CycleGAN-turbo	SDEdit 0.5	SDEdit 0.7	PnP	ControlNet	
FID ↓	19.908	52.152	21.315	14.426	16.840	73.494	48.757	61.617	35.091	16.384
MMD ↓	58.395	260.081	56.484	45.970	49.845	242.001	161.666	172.808	95.171	41.344
KID ↓	4.539	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

Qualiative comparison of Day-to-Night translation (BDD100K)

Qualitative demonstration of different Automotive translations (BDD100K, DENSE)



Non-Automotive Applications (FFHQ, AFHQ)



Please visit our project page at