

StyleGAN-XL

Обзор-рецензия на статью

Николаев Максим, БПМИ192

Авторы

- Axel Sauer, PhD student, University of Tübingen
- Katja Schwarz, PhD Student, University of Tübingen
- Andreas Geiger, Professor of Computer Science, University of Tübingen

StyleGAN3

- Бейзлайн StyleGAN-XL
- Сильно переработанная архитектура по-сравнению со StyleGAN2, основная задача была убрать артефакты при генерации видео

Projected GANs Converge Faster

NeurIPS 2021

- Авторы: Axel Sauer и Andreas Geiger.
- Статья посвящённая исследованию стабилизации и ускорения GANs.
- Из этой статьи авторы брали парадигмы для StyleGAN-XL.
- Так, авторы взяли латентное пространство равным 64, вместо 512.

Projected GANs Converge Faster

Task	Dataset	Model	Metric Name	Metric Value	Global Rank
Image Generation	ADE-Indoor	Projected GAN	FID	6.7	# 1
Image Generation	AFHQ Cat	Projected GAN	FID	2.16	# 1
Image Generation	AFHQ Dog	Projected GAN	FID	4.52	# 1
Image Generation	AFHQ Wild	Projected GAN	FID	2.17	# 2
Conditional Image Generation	ArtBench-10 (32x32)	Projected GAN	FID	11.837	# 6
Image Generation	Cityscapes	Projected GAN	FID-10k-training-steps	3.41	# 1
Image Generation	CLEVR	Projected GAN	FID-5k-training-steps	0.89	# 1
Image Generation	CUB 128 x 128	Projected GAN	FID	2.79	# 1

SwinIR: Image Restoration Using Swin Transformer

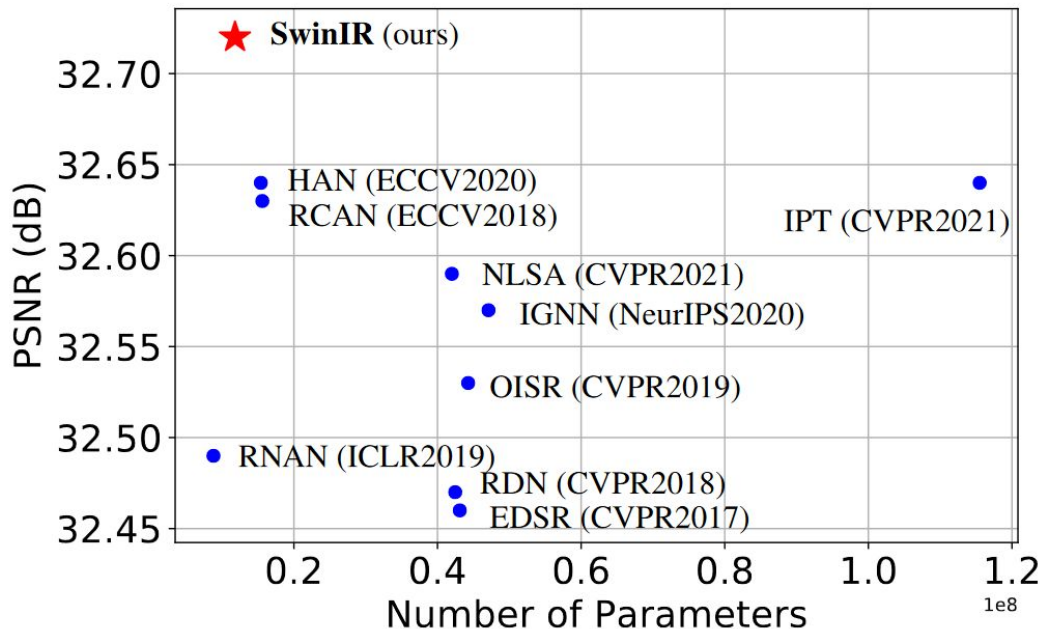


Figure 1: PSNR results v.s the total number of parameters of different methods for image SR ($\times 4$) on Set5 [3].

Конкуренты

Table 2. **Image Synthesis on ImageNet**. Empty cells indicate that the model was not available and the respective metric not evaluated in the original work.

Model	FID ↓	sFID ↓	rFID ↓	IS ↑	Pr ↑	Rec ↑	Model	FID ↓	sFID ↓	rFID ↓	IS ↑	Pr ↑	Rec ↑
Resolution 128²							Resolution 256²						
BigGAN	6.02	7.18	6.09	145.83	0.86	0.35	StyleGAN2	49.20					
CDM	3.52	128.80		128.80			BigGAN	6.95	7.36	75.24	202.65	0.87	0.28
ADM	5.91	5.09	13.29	93.31	0.70	0.65	CDM	4.88	158.70		158.70		
ADM-G	2.97	5.09	3.80	141.37	0.78	0.59	ADM	10.94	6.02	125.78	100.98	0.69	0.63
StyleGAN-XL	1.81	3.82	1.82	200.55	0.77	0.55	ADM-G-U	3.94	6.14	11.86	215.84	0.83	0.53
							StyleGAN-XL	2.30	4.02	7.06	265.12	0.78	0.53
Resolution 512²							Resolution 1024²						
BigGAN	8.43	8.13	312.00	177.90	0.88	0.29	StyleGAN-XL	2.52	4.12	413.12	260.14	0.76	0.51
ADM	23.24	10.19	561.32	58.06	0.73	0.60							
ADM-G-U	3.85	5.86	210.83	221.72	0.84	0.53							
StyleGAN-XL	2.41	4.06	51.54	267.75	0.77	0.52							

Diffusion models vs GANs

- EDM обогнал StyleGAN-XL на CIFAR-10 и ImageNet-64.
- StyleGAN-XL всё ещё топ 1 на больших датасетах.
- StyleGAN-XL значительно быстрее в инференсе.

Model	Inference Time ↓		
	Res. 128 ²	Res. 256 ²	Res. 512 ²
ADM	27.07	40.26	91.54
StyleGAN-XL	0.05	0.07	0.10

Diffusion models vs GANs

Task	Dataset	Model	Metric Name	Metric Value	Global
Image Generation	CIFAR-10	StyleGAN-XL	FID	1.85	# 1
Image Generation	FFHQ 1024 x 1024	StyleGAN-XL	FID	2.02	# 1
Image Generation	FFHQ 256 x 256	StyleGAN-XL	FID	2.19	# 2
Image Generation	FFHQ 512 x 512	StyleGAN-XL	FID	2.41	# 1
Image Generation	ImageNet 128x128	StyleGAN-XL	FID	1.81	# 1
Image Generation	ImageNet 256x256	StyleGAN-XL	FID	2.3	# 1

Diffusion models vs GANs

Image Generation	ImageNet 32x32	StyleGAN-XL	FID	1.10	# 1
Image Generation	ImageNet 512x512	StyleGAN-XL	FID	2.40	# 1
Image Generation	ImageNet 64x64	StyleGAN-XL	Inception Score	4.06	# 5
			FID	1.51	# 2
Image Generation	Pokemon 1024x1024	StyleGAN-XL	FID	25.47	# 1
Image Generation	Pokemon 256x256	StyleGAN-XL	FID	23.97	# 1

Сильные стороны

- SOTA для генерации больших и разнообразных наборов данных.
- Инференс быстрее, чем у диффузионок.
- Имеет распутанное латентное пространство и легко поддается редактированию.

Model	MSE ↓	PSNR ↑	SSIM ↑	FID ↓
BigGAN	0.10	10.85	0.26	47.48
StyleGAN-XL	0.06	13.45	0.33	21.73

Слабые стороны

- Модель не очень стабильна и её приходится обучать прогрессивно.
- Модель сильно больше обычных GAN, это плохо влияет на производительность.
- На датасетах с маленьким разрешением проигрывает SOTA диффузионкам
- Всё ещё проигрывает другим GAN в разнообразности и точности генерации

Дальнейшая работа

- Авторы предлагают провести дистилляцию, чтобы найти компромисс между качеством и производительностью.
- Также авторы предлагают использовать их метод на архитектуре StyleGAN 2, так как он лучше подходит для редактирования и инверсии, так как StyleGAN 3 семантически урезан.

Распутанность пространства



FFHQ 1024

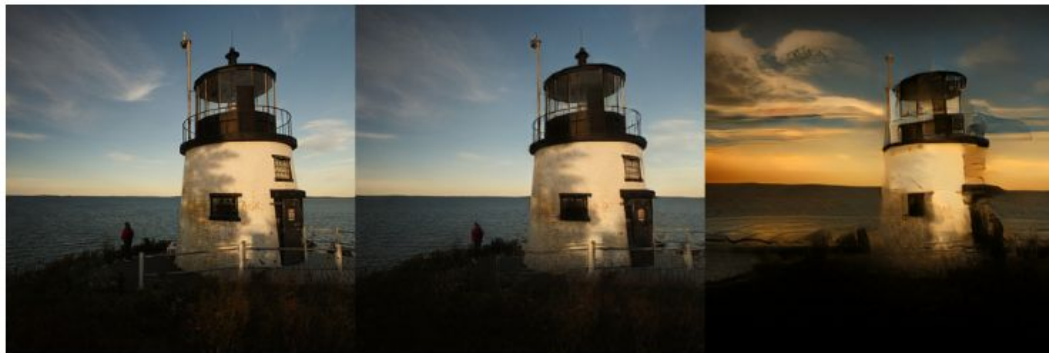


Инверсия



Fig. 9. **Inversion of a Given Source Image.** For BigGAN, we invert to its latent space z , for StyleGAN-XL we invert to style codes w .

Редактирование



Source

Inversion

Edit

BigGAN vs ADM vs StyleGAN-XL

