Self-Attention Generative Adversarial Networks

Han Zhang¹², Ian Goodfellow², Dimitris Metaxas¹, Augustus Odena² ¹Rutgers University, ²Google Brain





Which GAN paper are we talking about?



What did we do?

- Add Self-Attention blocks to Generator and Discriminator
- Spectral Normalization (Miyato et al., ICLR, 2018) in both Generator and Discriminator
- Different learning rate for Generator and Discriminator (TTUR: Heusel et al., NIPS, 2017)

What were the results?

Model	Inception Score	Intra-FID	FID
AC-GAN	28.5	260	١
SNGAN-projection	36.8	92.4	27.62
Our SAGAN	52.52	83.7	18.65

Comparison of SAGAN and AC-GAN (A. Odena et al., ICLR, 2017), SNGAN-projection (T. Miyato et al., ICLR, 2018) on ImageNet

What's wrong with convolutions?





DCGAN (Radford et al, ICLR, 2016)

Improved GAN (Salimans et al, NIPS, 2016)

- + Excel at synthesizing image classes with few structural constraints
- Fail to capture geometric or structural patterns

What is Self-Attention?







Image Transformer (Parmar et al, ICML, 2018) Non-local Neural Networks (Wang et al, CVPR, 2018)

+ Models long-range dependencies more efficiently

Self-Attention in GANs

 $\mathcal{Y}_{:}=\frac{1}{C(x)}\sum_{i}f(x_{i},x_{i})g(x_{i})$

Self-Attention in GANs



Self-Attention in GANs

$$\begin{aligned} g(x_{i}) &= \frac{1}{C(x_{i})} \sum_{j} f(x_{i}, x_{j}) g(x_{j}) \\ f(x_{i}, x_{j}) &= \exp(\langle \Theta(x_{i}) | \phi(x_{j}) \rangle \\ &= \frac{1}{2} \sum_{j} f(x_{i}, x_{j}) \quad \exists \ a \quad \text{"softmax}^{n} \\ g(x_{j}) \quad \exists \ a \quad \text{"softmax}^{n} \end{aligned}$$

What does self-attention do?



Generator allocates attention according to similarity of color and texture

What does self-attention do?



Adjacent query points may result in very different attention maps

What does self-attention do?



Attention lets you draw dogs with 4 legs!

There's a learned distinction between background and foreground

Spectral Normalization in the Generator

• SN-GAN (Miyato et al., ICLR, 2018) only adds SN to the

discriminator.

- We add it to the generator as well
- The fact that this works suggests that the motivation given in e.g.

the WGAN paper is incomplete

What does Spectral Norm do?

• Spectral Normalization (SN) in both G and D



Baseline: Very unstable

SN on G/D: More stable

Different learning rates for G and D

- Regularized Discriminators use more steps than generators
- This is wasteful! It throws out useful computations and it makes training go slower.
- If you use a lower learning rate for the generator, everything works better.

What do different learning rates do?

• Imbalanced learning rate, TTUR, for G and D updates



Best FID: 33.39

Best FID: 18.65

Results (all ImageNet classes)

Model	no attention	SAGAN				Residual			
		$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$	$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$
FID	22.96	22.98	22.14	18.28	18.65	42.13	22.40	27.33	28.82
IS	42.87	43.15	45.94	51.43	52.52	23.17	44.49	38.50	38.96

Comparison of Self-Attention and Residual block on GANs

Adding attention at "higher" layers is better.

Results (all ImageNet classes)

Model	no attention	SAGAN				Residual			
		$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$	$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$
FID	22.96	22.98	22.14	18.28	18.65	42.13	22.40	27.33	28.82
IS	42.87	43.15	45.94	51.43	52.52	23.17	44.49	38.50	38.96

Comparison of Self-Attention and Residual block on GANs

This comparison shows that attention really helps.

Results (all ImageNet classes)

Model	no attention	SAGAN				Residual			
		$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$	$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$
FID	22.96	22.98	22.14	18.28	18.65	42.13	22.40	27.33	28.82
IS	42.87	43.15	45.94	51.43	52.52	23.17	44.49	38.50	38.96

Comparison of Self-Attention and Residual block on GANs

This comparison shows that attention is better than just adding the same number of parameters to convolutions.

Intra-FID for example classes

goldfish

Model	SNGAN- projection	Our SAGAN	indigo
Goldfish	58.1	44.4	bunting
Indigo bunting	66.8	53.0	
Redshank	60.1	48.9	redshank
Saint bernard	55.3	35.7	agint
Tiger cat	90.2	88.1	bernard



tiger cat

More Info

Paper: arxiv.org/pdf/1805.08318

Code: github.com/brain-research/self-attention-gan

Poster: 06:30 -- 09:00 PM @ Pacific Ballroom #11