Large Scale GAN Training for High Fidelity Natural Image Synthesis (BigGAN)

Дроздова Анастасия, 181

Постановка задачи: генерация картинок по классу

State-of-the-art: SA-GAN

- Inception score (IS) = 52.52 ($\rightarrow$ max)
- Fréchet Inception Distance (FID) = 18.65 ($\rightarrow$ min)

Хотим улучшить:

- Метрики (получили: IS = 166.5, FID = 7.4)
- Увеличить разрешение изображений
- Генерация деталей




Scaling up GANs

Batch	Ch.	Param (M)	Shared	Skip-z	Ortho.	Itr $\times 10^3$	FID	IS
256	64	81.5	SA-GAN Baseline			1000	18.65	52.52
512	64	81.5	✗	✗	✗	1000	15.30	58.77($\pm$1.18)
1024	64	81.5	✗	✗	✗	1000	14.88	63.03($\pm$1.42)
2048	64	81.5	✗	✗	✗	732	12.39	76.85($\pm$3.83)
2048	96	173.5	✗	✗	✗	295($\pm$18)	9.54($\pm$0.62)	92.98($\pm$4.27)
2048	96	160.6	✓	✗	✗	185($\pm$11)	9.18($\pm$0.13)	94.94($\pm$1.32)
2048	96	158.3	✓	✓	✗	152($\pm$7)	8.73($\pm$0.45)	98.76($\pm$2.84)
2048	96	158.3	✓	✓	✓	165($\pm$13)	8.51($\pm$0.32)	99.31($\pm$2.10)
2048	64	71.3	✓	✓	✓	371($\pm$7)	10.48($\pm$0.10)	86.90($\pm$0.61)

Spectral Normalization

- Стабилизирует обучение дискриминатора
- Контролирует константу Липшица, ограничивает спектральную норму каждого слоя $g : h_{in} \to h_{out}$

$$\|g\|_{\mathrm{Lip}} = \sup_h \sigma(\nabla g(h))$$

$$\sigma(A) := \max_{\boldsymbol{h}:\boldsymbol{h}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{h}\|_2}{\|\boldsymbol{h}\|_2} = \max_{\|\boldsymbol{h}\|_2\leq 1} \|A\boldsymbol{h}\|_2$$ - макс. сингулярное значение

Пусть сеть имеет вид

$$f(\boldsymbol{x}, \theta) = W^{L+1} a_L(W^L(a_{L-1}(W^{L-1}(\ldots a_1(W^1\boldsymbol{x})\ldots))))$$

Spectral Normalization

$$f(\boldsymbol{x}, \theta) = W^{L+1} a_L(W^L(a_{L-1}(W^{L-1}(\dots a_1(W^1 \boldsymbol{x}) \dots))))$$

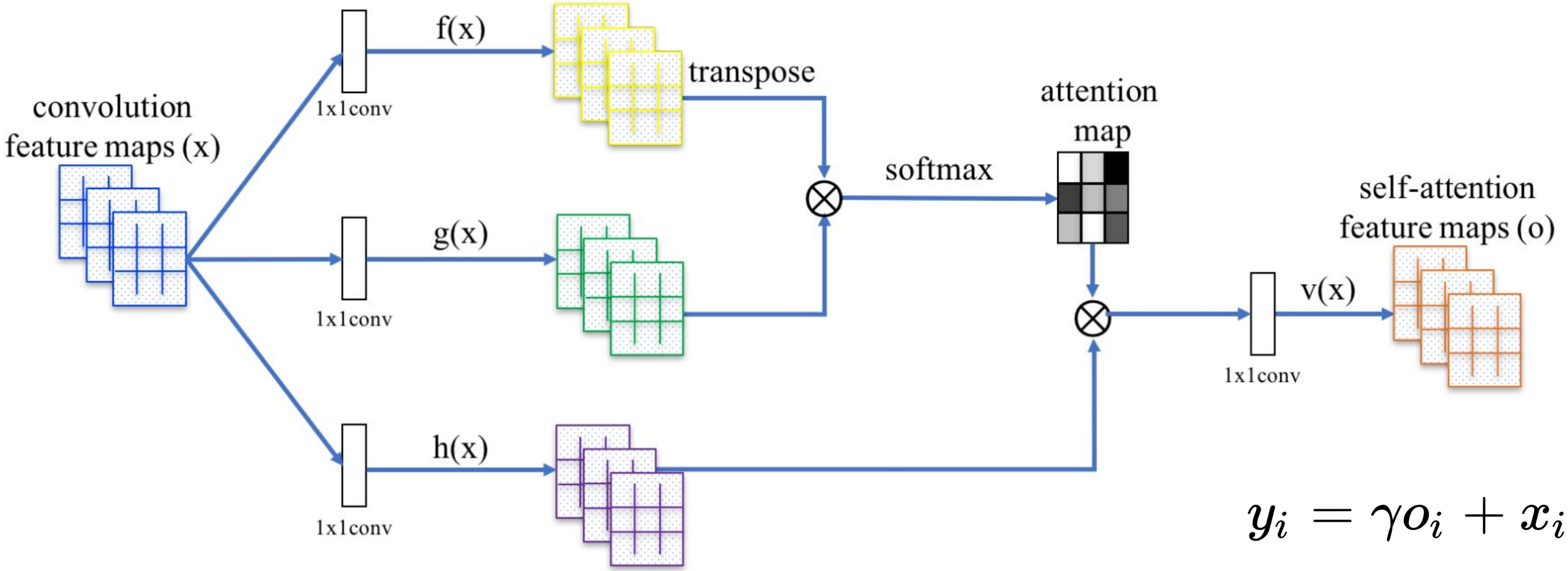
Для линейного слоя $g(h) = Wh$ $\quad \|g\|_{\text{Lip}} = \sup_h \sigma(W) = \sigma(W)$

$\|a_l\|_{\text{Lip}} = 1$ (ReLU, leaky ReLU)

Тогда из $\|g_1 \circ g_2\|_{\text{Lip}} \leq \|g_1\|_{\text{Lip}} \cdot \|g_2\|_{\text{Lip}}$ получаем: $\|f\|_{\text{Lip}} \leq \prod_{l=1}^{L+1} \sigma(W^l)$

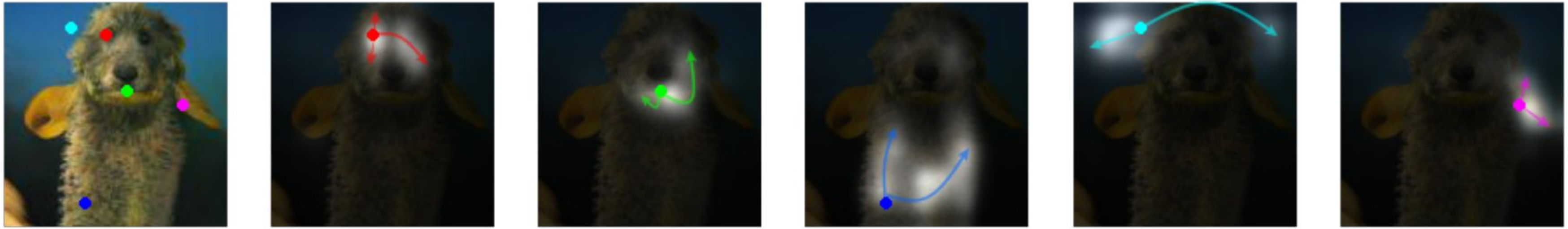
$$\overline{W}_{SN}(W) := W / \sigma(W)$$

Self-Attention GAN (SA-GAN)



$$y_i = \gamma o_i + x_i$$

SA-GAN



Hinge Loss:

$$
\begin{aligned}
L_D = & -\mathbb{E}_{(x,y)\sim p_{data}}[\min(0, -1 + D(x,y))] \\
& -\mathbb{E}_{z\sim p_z, y\sim p_{data}}[\min(0, -1 - D(G(z),y))] \\
L_G = & -\mathbb{E}_{z\sim p_z, y\sim p_{data}} D(G(z),y),
\end{aligned}
$$

Conditional BatchNorm

$$\hat{x}_i = \frac{x_i - \mu_\mathcal{B}}{\sqrt{\sigma_\mathcal{B}^2 + \epsilon}}$$

$$y_i = \gamma \hat{x}_i + \beta = \mathrm{BN}_{\gamma,\beta}(x_i)$$

$$\hat{\beta}_c = \beta_c + \Delta\beta_c$$

$$\hat{\gamma}_c = \gamma_c + \Delta\gamma_c$$

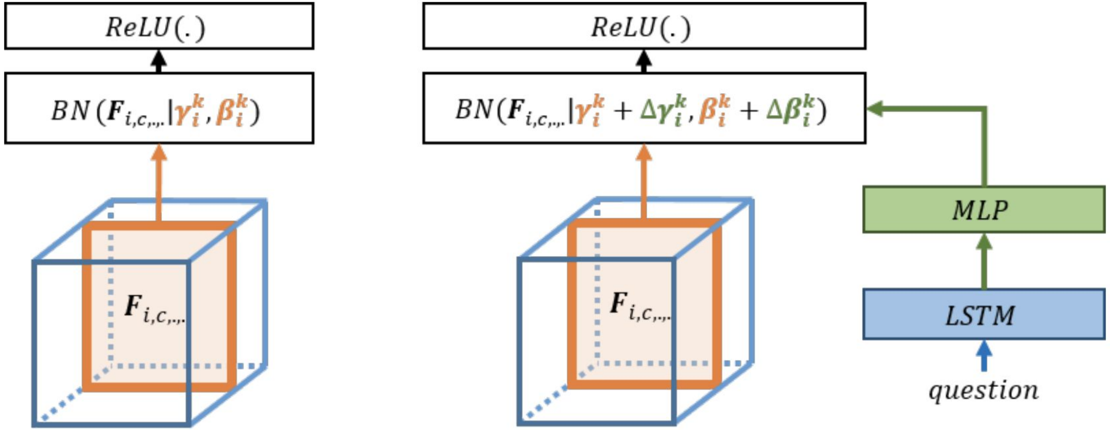


Figure 2: An overview of the computation graph of batch normalization (left) and conditional batch normalization (right). Best viewed in color.

Truncation Trick

Выбираем случайный вектор: $z \sim \mathcal{N}(0, I)$, перевыбираем, если величина больше какого-то порога.

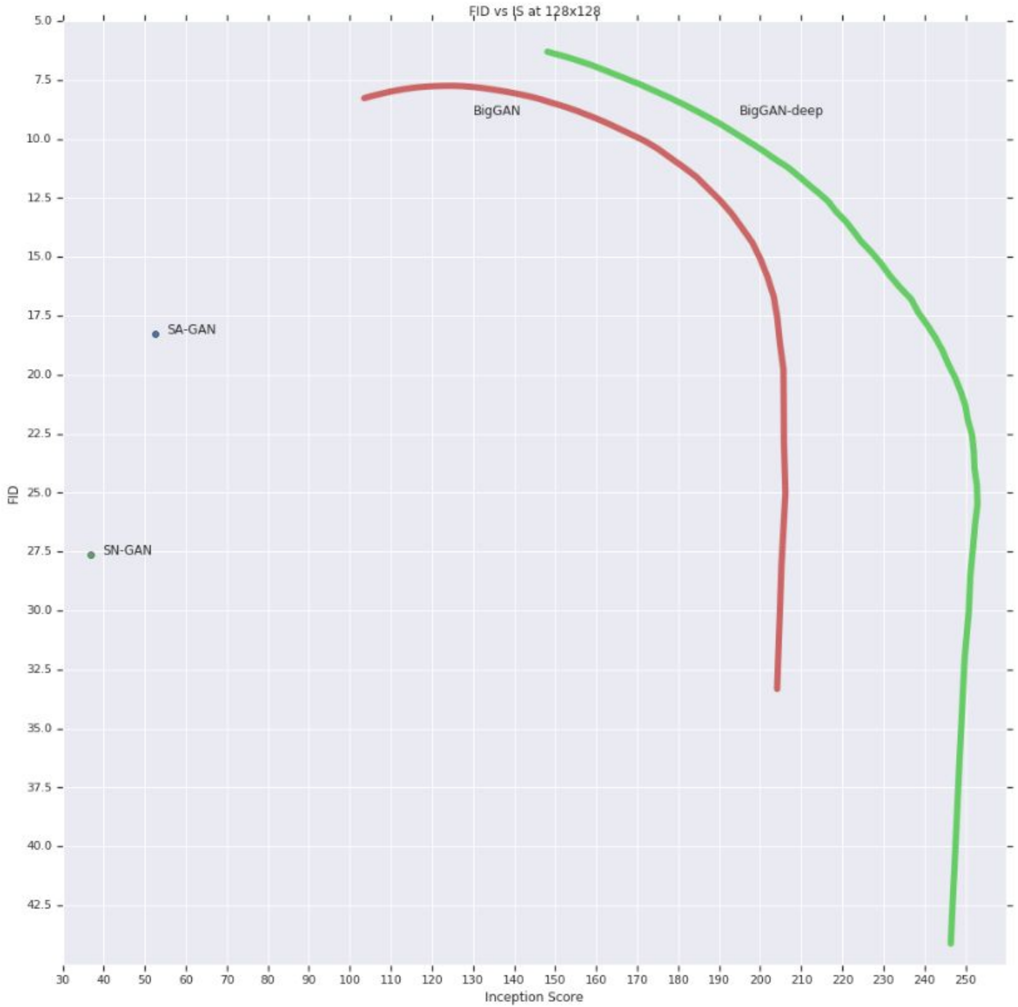
2 1 0.5 0.04

Truncation Trick

Выбор между

разнообразием (FID) и

правильностью (IS)



Orthogonal Regularization - Generator

Изначальная версия:

$$R_\beta(W) = \beta \|W^\top W - I\|_{\mathrm{F}}^2$$

W - матрица весов

β - гиперпараметр

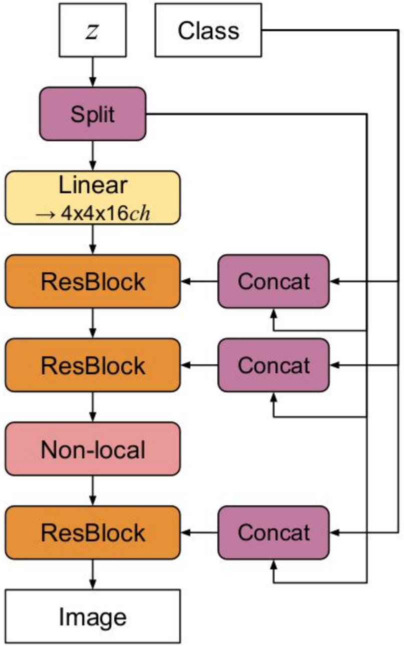
Ослабленные ограничения:

$$R_\beta(W) = \beta \|W^\top W \odot (\mathbf{1} - I)\|_{\mathrm{F}}^2$$

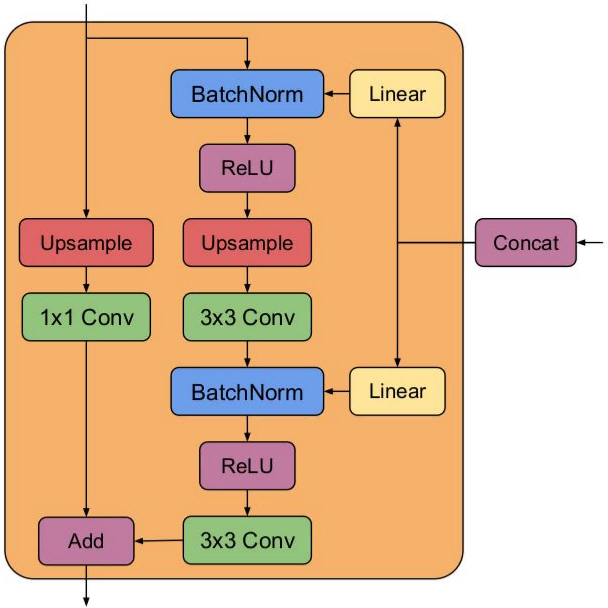
$\mathbf{1}$ - матрица из 1

минимизирует попарные косинусные расстояния

BigGAN - Generator



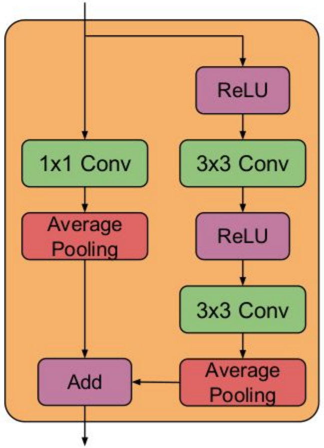
ResBlock



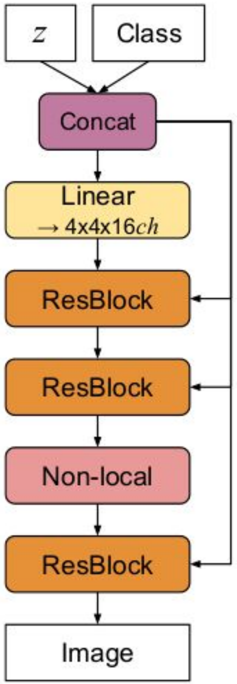
BigGAN - Discriminator

RGB image $x \in \mathbb{R}^{256\times256\times3}$
ResBlock down $ch \rightarrow 2ch$
ResBlock down $2ch \rightarrow 4ch$
Non-Local Block (64×64)
ResBlock down $4ch \rightarrow 8ch$
ResBlock down $8ch \rightarrow 8ch$
ResBlock down $8ch \rightarrow 16ch$
ResBlock down $16ch \rightarrow 16ch$
ResBlock $16ch \rightarrow 16ch$
ReLU, Global sum pooling
Embed$(y)\cdot\boldsymbol{h}$ + (linear $\rightarrow 1$)

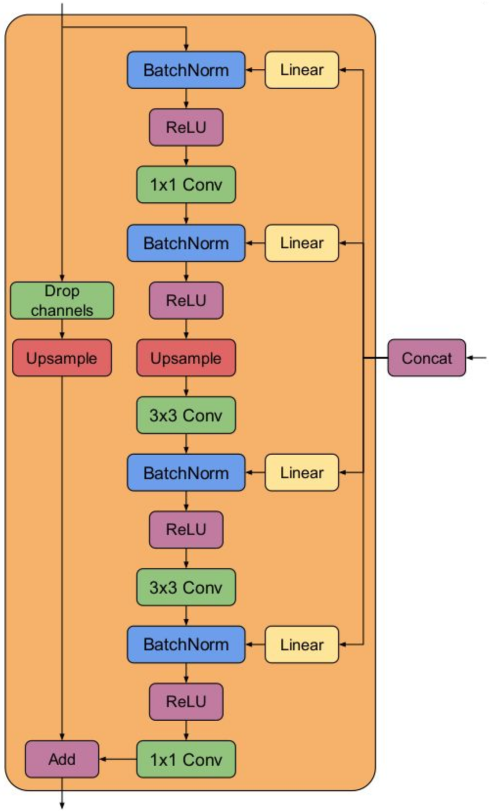
Resblock



BigGAN-deep - Generator



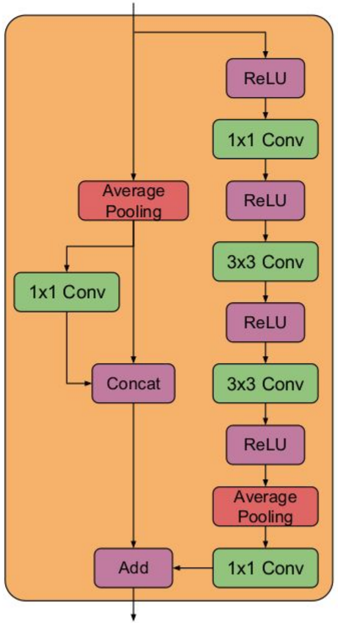
ResBlock



BigGAN-deep - Discriminator

RGB image $x \in \mathbb{R}^{256\times256\times3}$
3×3 Conv $3 \to ch$
ResBlock down $ch \to 2ch$
ResBlock $2ch \to 2ch$
ResBlock down $2ch \to 4ch$
ResBlock $4ch \to 4ch$
Non-Local Block (64×64)
ResBlock down $4ch \to 8ch$
ResBlock $8ch \to 8ch$
ResBlock down $8ch \to 8ch$
ResBlock $8ch \to 8ch$
ResBlock down $8ch \to 16ch$
ResBlock $16ch \to 16ch$
ResBlock down $16ch \to 16ch$
ResBlock $16ch \to 16ch$
ReLU, Global sum pooling
Embed$(y)\cdot\boldsymbol{h}$ + (linear $\to 1$)

ResBlock

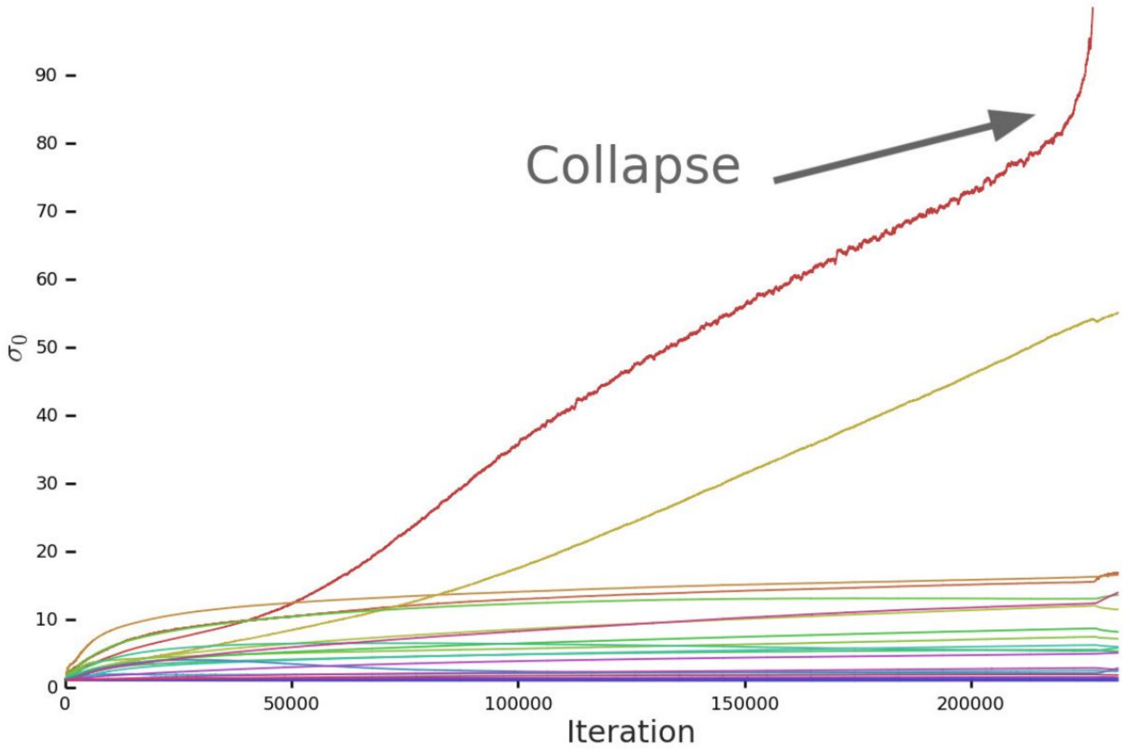


Instabilities - Generator

Улучшаем стабильность

регуляризацией:

$$W = W - \max(0, \sigma_0 - \sigma_{clamp}) v_0 u_0^\top$$

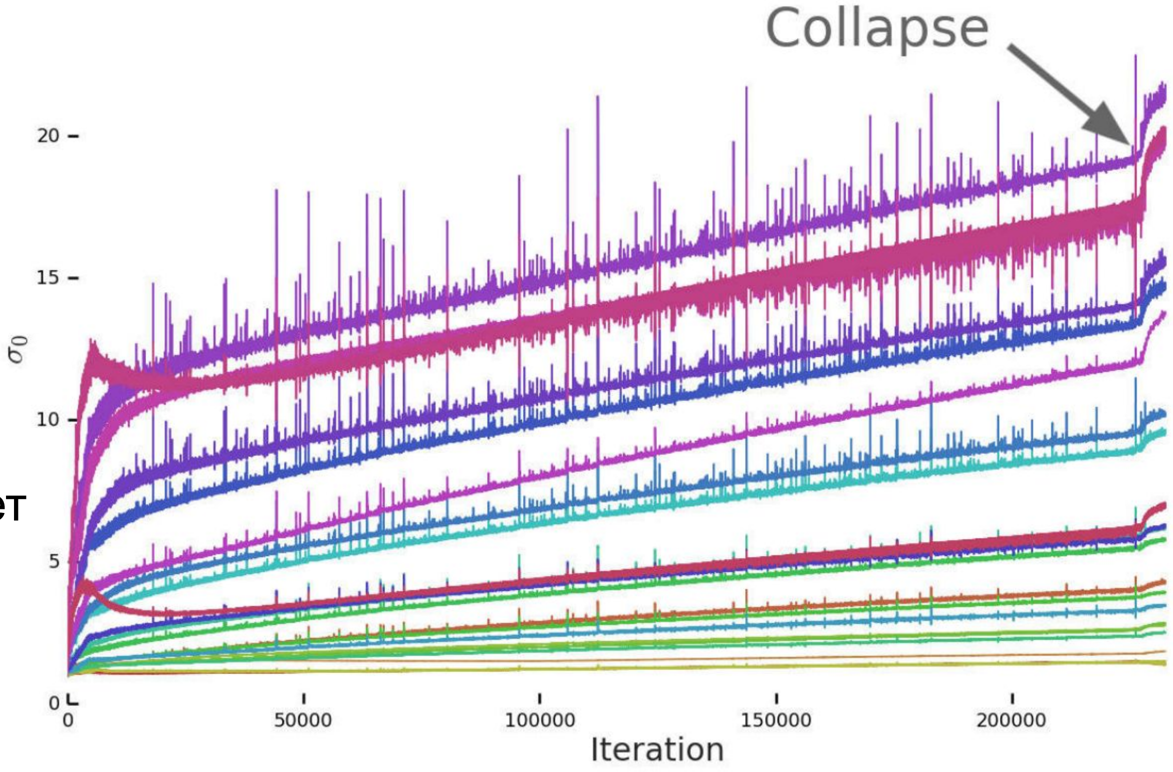


Instabilities - Discriminator

Регуляризуем Якобиан

дискриминатора:

$$R_1 := \frac{\gamma}{2}\mathbb{E}_{p_{\mathcal{D}}(x)}\left[\|\nabla D(x)\|_F^2\right]$$



Дискриминатор запоминает

обучающ. выборку

Experiments and results

ImageNet ILSVRC 2012

Model	Res.	FID/IS	(min FID) / IS	FID / (valid IS)	FID / (max IS)
SN-GAN	128	$27.62/36.80$	N/A	N/A	N/A
SA-GAN	128	$18.65/52.52$	N/A	N/A	N/A
BigGAN	128	$8.7 \pm .6/98.8 \pm 3$	$7.7 \pm .2/126.5 \pm 0$	$9.6 \pm .4/166.3 \pm 1$	$25 \pm 2/206 \pm 2$
BigGAN	256	$8.7 \pm .1/142.3 \pm 2$	$7.7 \pm .1/178.0 \pm 5$	$9.3 \pm .3/233.1 \pm 1$	$25 \pm 5/291 \pm 4$
BigGAN	512	$8.1/144.2$	$7.6/170.3$	$11.8/241.4$	$27.0/275$
BigGAN-deep	128	$5.7 \pm .3/124.5 \pm 2$	$6.3 \pm .3/148.1 \pm 4$	$7.4 \pm .6/166.5 \pm 1$	$25 \pm 2/253 \pm 11$
BigGAN-deep	256	$6.9 \pm .2/171.4 \pm 2$	$7.0 \pm .1/202.6 \pm 2$	$8.1 \pm .1/232.5 \pm 2$	$27 \pm 8/317 \pm 6$
BigGAN-deep	512	$7.5/152.8$	$7.7/181.4$	$11.5/241.5$	$39.7/298$

JFT-300M (BigGAN, Res 256)

Ch.	Param (M)	Shared	Skip-z	Ortho.	FID	IS	(min FID) / IS	FID / (max IS)
64	317.1	✗	✗	✗	48.38	23.27	$48.6/23.1$	$49.1/23.9$
64	99.4	✓	✓	✓	23.48	24.78	$22.4/21.0$	$60.9/35.8$
96	207.9	✓	✓	✓	18.84	27.86	$17.1/23.3$	$51.6/38.1$
128	355.7	✓	✓	✓	13.75	30.61	$13.0/28.0$	$46.2/47.8$

References

BigGAN

SA-GAN

Spectral Normalization for GANs

Modulating early visual processing by language

RunwayML - BigGAN

BigGAN - trained