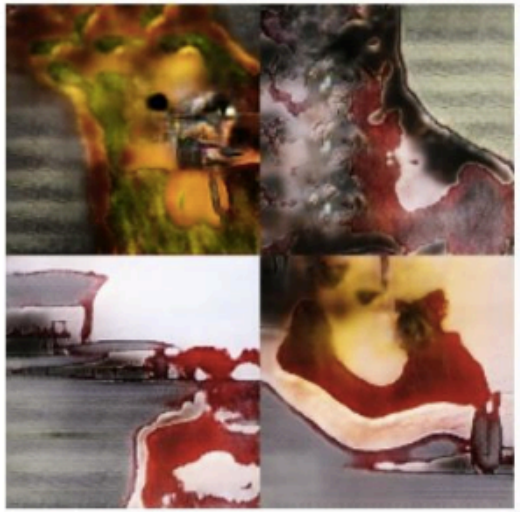
LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS

Чистяков Глеб

162 гр.



2015

2016

2018

2019

Metrics

Inseption Score

[Barratt et. al, 2018]

$$\mathbf{IS}(G) = \exp\left(\mathbb{E}_{\mathbf{x}\sim p_g} D_{KL}(\, p(y|\mathbf{x}) \,\|\, p(y) \,) \right)$$

Y – метка класса, X – сгенерированное изображение

$p(y) = \int_{\mathbf{x}} p(y|\mathbf{x}) p_g(\mathbf{x})$ - предельное распределение классов

Frechet Inseption Distance

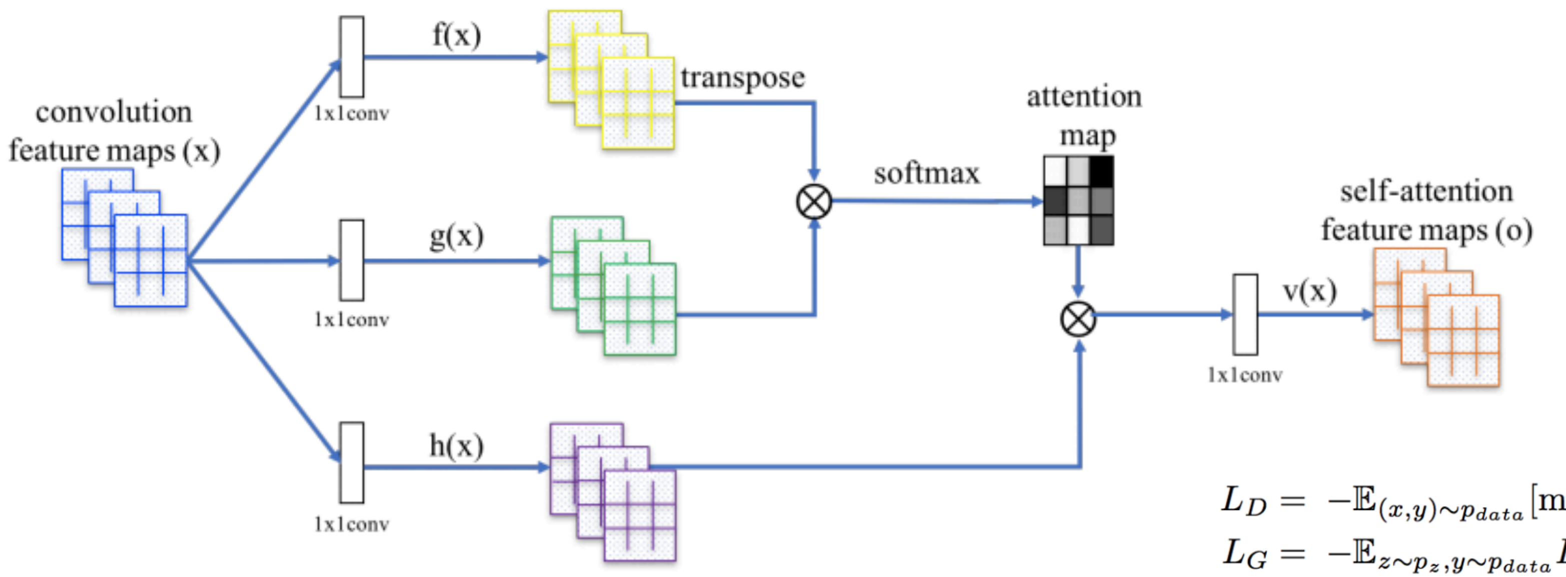
[DeVries et. al., 2019]

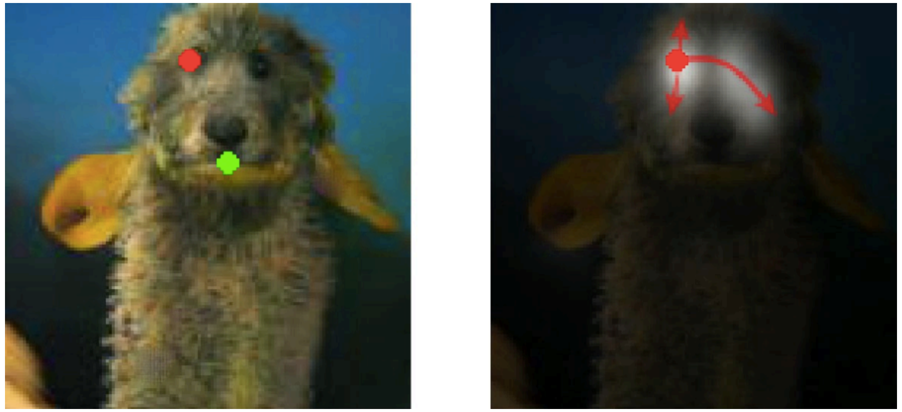
$$\mathtt{FID}(x, g) = ||\mu_x - \mu_g||_2^2 + \mathrm{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}})$$

X – реальное изображение, G – сгенерированное
μ – среднее значение, Σ – ковариация

Self-Attention GAN

[Zhang et al., 2018]





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

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{W_g x} \qquad \boldsymbol{x} \in \mathbb{R}^{C\times N},\ \boldsymbol{W_f} \in \mathbb{R}^{\bar{C}\times C},\ \bar{C} = C/8$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{W_f x} \qquad \boldsymbol{W_g} \in \mathbb{R}^{\bar{C}\times C}$$

$$s_{ij} = \boldsymbol{f}(\boldsymbol{x_i})^T \boldsymbol{g}(\boldsymbol{x_j})$$

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})} \qquad \beta \in \mathbb{R}^{N\times N}$$

$\beta_{j,i}$ indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region.

$$\boldsymbol{h}(\boldsymbol{x_i}) = \boldsymbol{W_h x_i} \qquad \boldsymbol{W_h} \in \mathbb{R}^{C\times C}$$

$$\boldsymbol{o_j} = \sum_{i=1}^{N} \beta_{j,i} \boldsymbol{h}(\boldsymbol{x_i}) \qquad \boldsymbol{o} \in \mathbb{R}^{C\times N}$$

$$\boldsymbol{y_i} = \gamma \boldsymbol{o_i} + \boldsymbol{x_i}$$

where γ is initialized as 0 so the model will explore the local spatial information first before refining it with self-attention.

Scaling up GANs and Modifications

Размер батча * 8 → Увеличение IS на 46%

Количество каналов в каждом слое * 2 → Увеличение IS на 21%

Общий BatchNorm embedding, который линейно проецируется на gain и biase каждого слоя → Снижение затрат на вычисление и память
Увеличение скорости на 37%

Direct skip connections (skip-z) → Улучшение производительности на 4%
Повышение скорости на 18%.

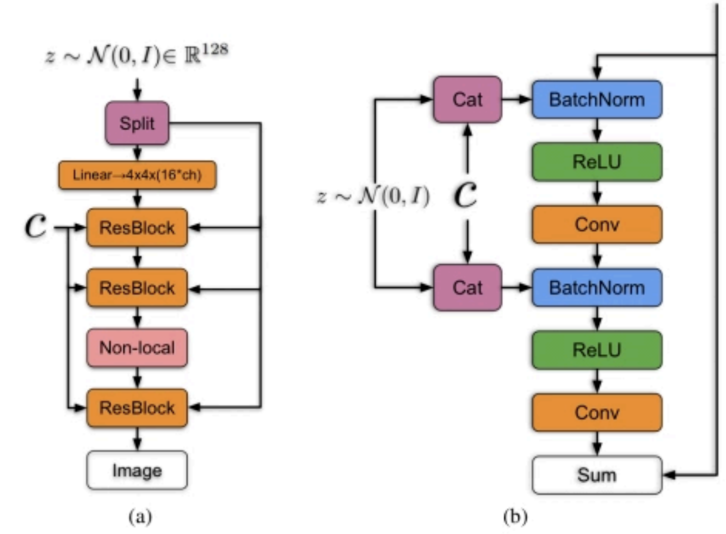


Figure 15: (a) A typical architectural layout for **G**; details are in the following tables. (b) A Residual Block in **G**. c is concatenated with a chunk of z and projected to the BatchNorm gains and biases.

Truncation Trick

Уловка усечения - усечение вектора *z* путем повторной выборки значений с величиной, превышающей выбранный порог

качество

разнообразие

(a) (b)

Figure 2: (a) The effects of increasing truncation. From left to right, the threshold is set to 2, 1, 0.5, 0.04. (b) Saturation artifacts from applying truncation to a poorly conditioned model.

Orthogonal Regularization

[Brock et al., 2017]

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

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

1 – матрица из единиц

W - весовая матрица, а β - гиперпараметр

Characterizing Instability: Generator

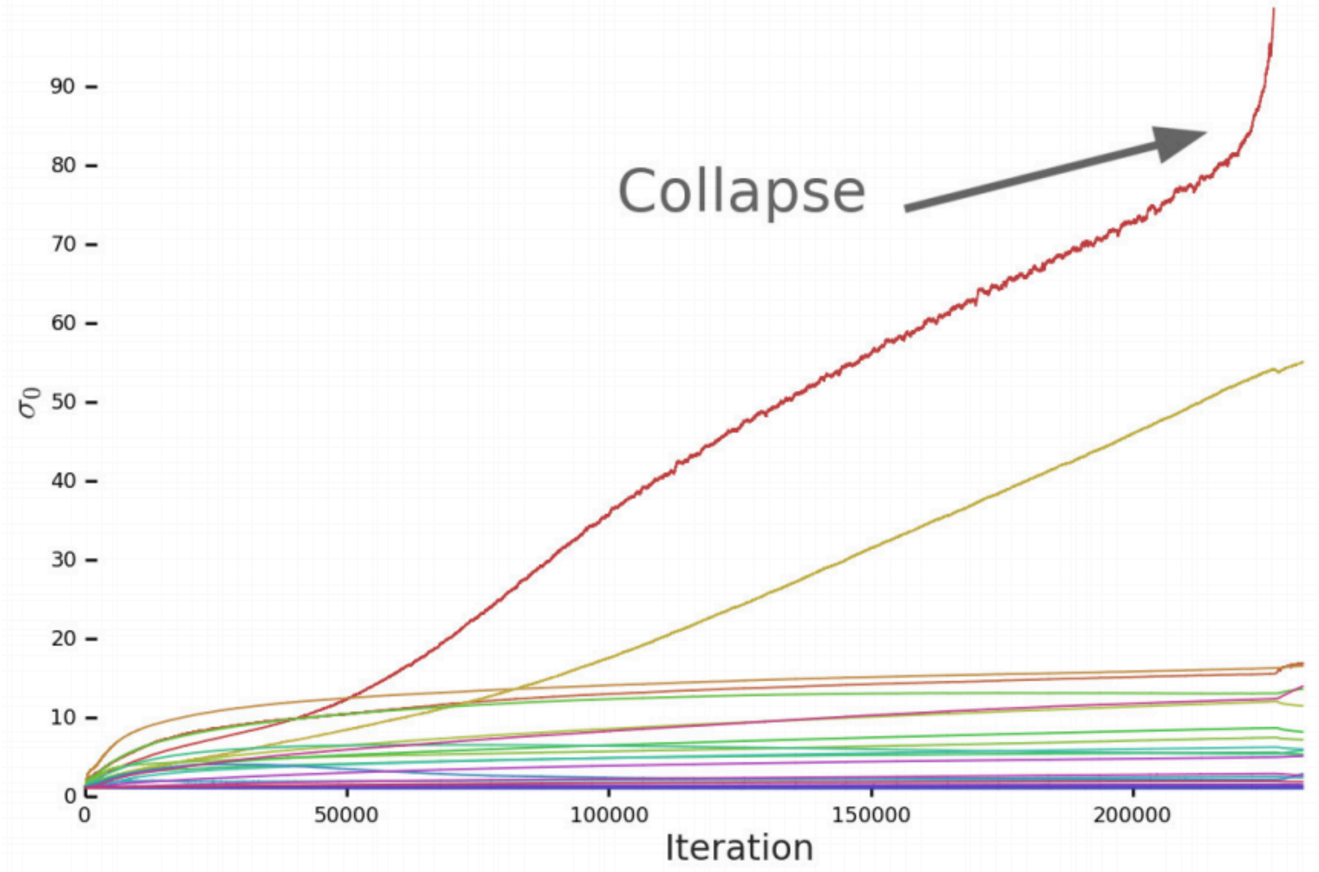
σ_0 σ_1 σ_2 – верхние сингулярные значения каждой матрицы

Regularize the top σ_0:

- σ_{reg}
- $r * sg(\sigma_1)$ - with sg the stop-gradient operation

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

where σ_{clamp} is set to either σ_{reg} or $r \cdot sg(\sigma_1)$



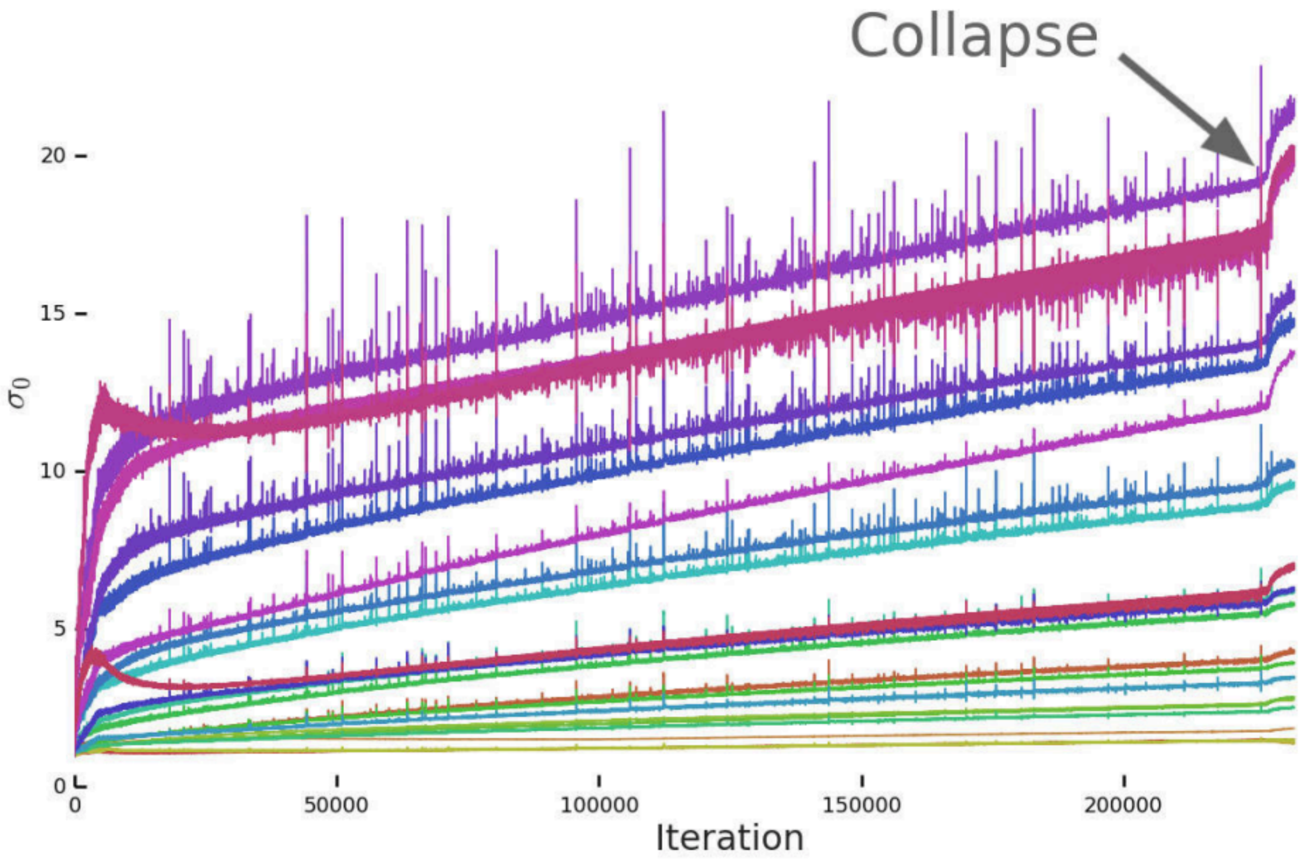
Characterizing Instability: Discriminator

[Mescheder et al., 2018]

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

При γ = 10:
- Тренировка становится стабильней
- Производительность уменьшается
  IS уменьшается на 20%

Различные γ, Dropout, L2 – аналогичное поведение



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)

Table 1: Fréchet Inception Distance (FID, lower is better) and Inception Score (IS, higher is better) for ablations of our proposed modifications. *Batch* is batch size, *Param* is total number of parameters, *Ch.* is the channel multiplier representing the number of units in each layer, *Shared* is using shared embeddings, *Skip-z* is using skip connections from the latent to multiple layers, *Ortho.* is Orthogonal Regularization, and *Itr* indicates if the setting is stable to 10^6 iterations, or it collapses at the given iteration. Other than rows 1-4, results are computed across 8 random initializations.

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	$7.5/152.8$	$7.7/181.4$	$11.5/241.5$	$39.7/298$

Table 2: Evaluation of models at different resolutions. We report scores without truncation (Column 3), scores at the best FID (Column 4), scores at the IS of validation data (Column 5), and scores at the max IS (Column 6). Standard deviations are computed over at least three random initializations.

Quiz

- Описать baseline модель SA-GAN (общую схему). Какие методы масштабирования и модификации могут улучшить результаты модели?
- Что такое Truncation Trick? Написать формулу ортогональной регуляризации.