



人工智能原理与技术



生成式对抗网络前沿

目录

CONTENTS

01

Progressive GAN

逐步生成，更稳更强大

02

SN-GAN

谱归一化层实现利普希茨连续

03

Self-Attention GAN

建模长距离依赖关系

04

BigGAN

BigGAN



Ian Goodfellow @goodfellow_ian · Jan 15

4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661

arxiv.org/abs/1511.06434 arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948



2014



2015



2016



2017



2018



31



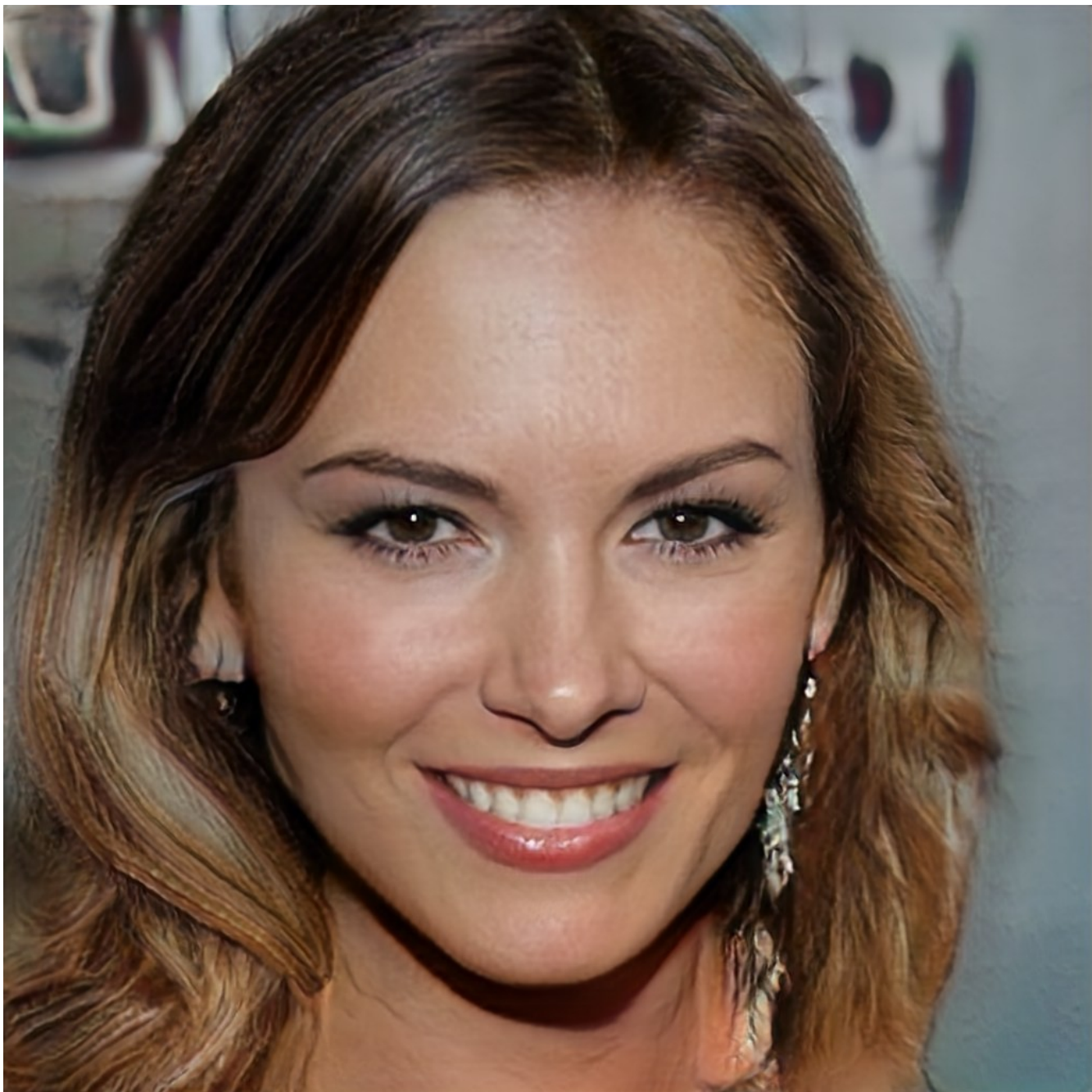
1.1K

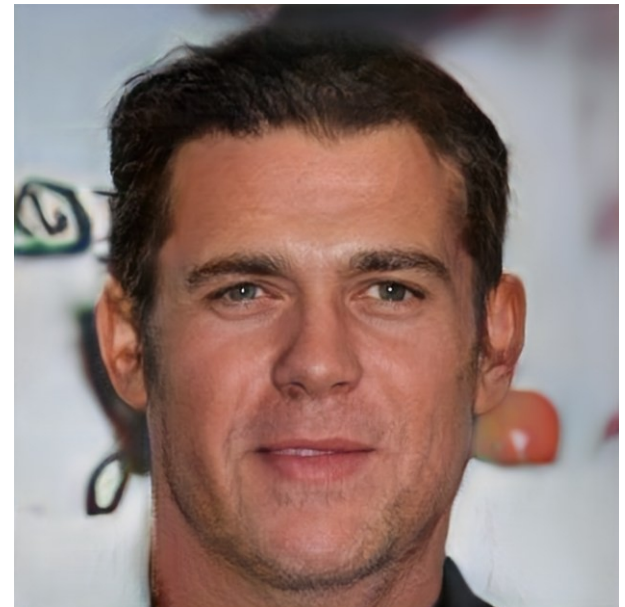


3.1K



GAN说，我有个梦想，那就是-做大做强！





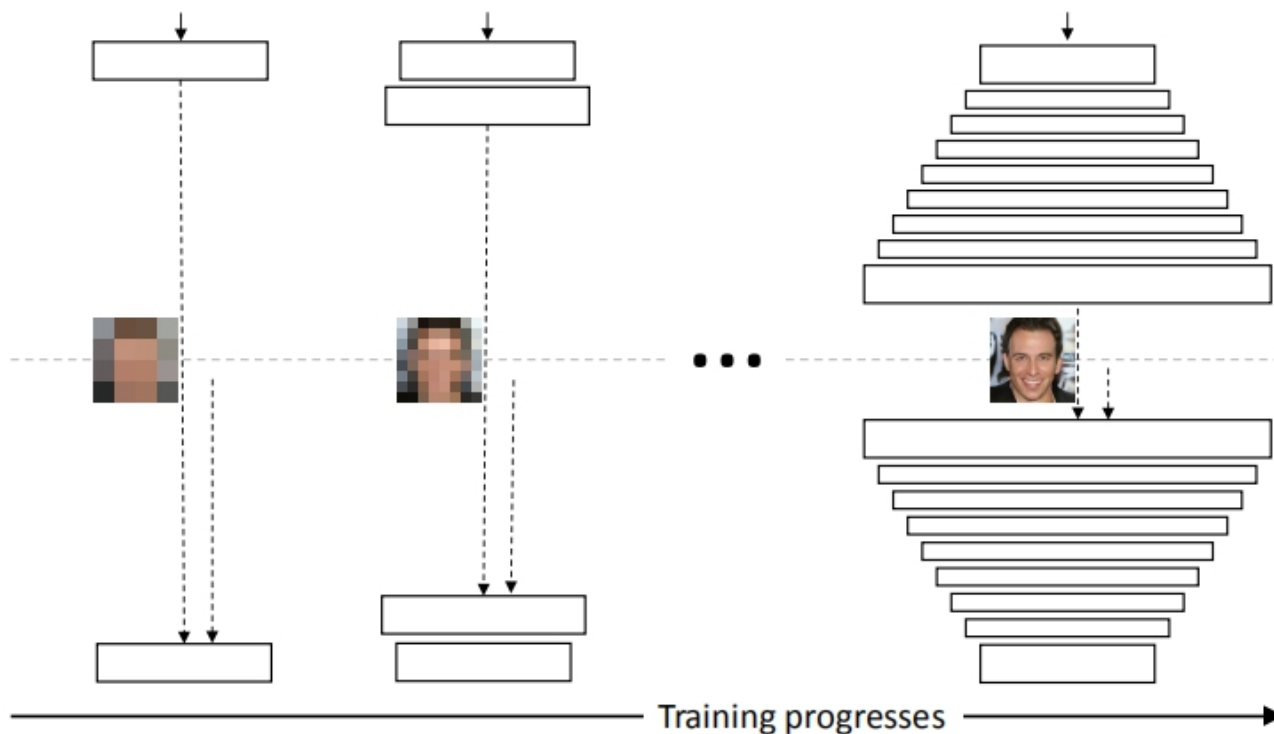
Progressive GAN

一点一点，稳扎稳打的来

解决思路：逐步训练

如何做大做强？

拒绝一口吃个胖子，要一步一个脚印~（逐步训练）



先训练一个小的GAN 生成一个小的图像
在上一个图像的基础上
再训练一个中等GAN 生成一个中等的图像

最后生成一个大的图像

Progressive GAN

实验结果



Gulrajani et al. (2017) (128×128)



Our (256×256)

目录

CONTENTS

01

Progressive GAN

逐步生成，更稳更强大

02

SN-GAN

谱归一化层实现利普希茨连续

03

Self-Attention GAN

建模长距离依赖关系

04

BigGAN

BigGAN

Spectral Normalization GAN

利普希茨连续

Lipschitz约束简单而言就是：要求在整个 $f(\cdot)$ 的定义域内有

$$\frac{\|f(x) - f(x')\|_2}{\|x - x'\|_2} \leq M \quad (3)$$

其中，M是一个常数。满足公式(3)的函数 $f(\cdot)$ ，具体表现为：函数变化不会太快，其梯度总是有限的，即使最剧烈时，也被限制在小于等于M的范围。

WGAN、WGAN-GP从理论就要求判别器D满足**利普希茨连续条件**
实际应用发现，任意GAN的判别器如果能够满足利普希茨连续条件，都对生成数据的质量有益处

WGAN：权重截断 --> 权重二值化、梯度消失/爆炸 WGAN-GP：惩罚项 --> 收敛速度仍然慢于 DCGAN

有没有其他办法？

SN-GAN：每层权重除以该层矩阵谱范数即可满足利普希茨连续

Spectral Normalization GAN

SN-GAN

为简便分析，可将Discriminator看作是多层网络，因为CNNs可看作是特殊的多层网络。对于多层网络的第 n 层，其输入与输出关系可以表示为：

$$\mathbf{x}_n = a_n(W_n \mathbf{x}_{n-1} + \mathbf{b}_n) \quad (4)$$

其中， $a_n(\cdot)$ 是该层网络的非线性激活函数，可采用ReLU； W_l 是网络参数矩阵， $\mathbf{b}_l$ 是网络的偏置，为推导方便，对 $\mathbf{b}_l$ 进行省略处理，则(4)式可写为：

$$\mathbf{x}_n = D_n W_n \mathbf{x}_{n-1} \quad (5)$$

其中 D_n 是对角矩阵，用于表示ReLU的作用，当其对应输入为负数时，对角元素为0；当其对应输入为正数时，对角元素为1。于是，多层神经网络（假设是 N 层）输入输出关系可以写成：

$$f(\mathbf{x}) = D_N W_N \cdots D_1 W_1 \mathbf{x} \quad (6)$$

Lipschitz约束是对 $f(\mathbf{x})$ 的梯度提出的要求：

$$\|\nabla_x(f(\mathbf{x}))\|_2 = \|D_N W_N \cdots D_1 W_1\|_2 \leq \|D_N\|_2 \|W_N\|_2 \cdots \|D_1\|_2 \|W_1\|_2 \quad (7)$$

Spectral Normalization GAN

SN-GAN

此处 $\|W\|$ 表示矩阵W的谱范数，它的定义如下：

$$\sigma(A) := \max_{\|h\| \neq 0} \frac{\|Ah\|_2}{\|h\|_2} = \max_{\|h\|=1} \|Ah\|_2 \quad (8)$$

$\sigma(W)$ 是矩阵W的最大奇异值，对于对角矩阵D，有 $\sigma(D) = \max(d_1, \dots, d_n)$ ，即对角元素上最大的元素。由此，(7)可表示为：

$$\|\nabla_x(f(\mathbf{x}))\|_2 \leq \prod_{i=1}^N \sigma(W_i) \quad (9)$$

因为，ReLU所对应的对角矩阵的谱范数最大为1。为使 $f(\mathbf{x})$ 满足Lipschitz约束，可对(7)进行归一化：

$$\|\nabla_x(f(\mathbf{x}))\|_2 = \|D_N \frac{W_N}{\sigma(W_N)} \cdots D_1 \frac{W_1}{\sigma(W_1)}\|_2 \leq \prod_{i=1}^N \frac{\sigma(W_i)}{\sigma(W_i)} = 1 \quad (10)$$

由此可见，只需让每层网络的网络参数除以该层参数矩阵的谱范数即可满足Lipschitz=1的约束，由此诞生了谱归一化（Spectral Normalization）。

目录

CONTENTS

01

Progressive GAN

逐步生成，更稳更强大

02

SN-GAN

谱归一化层实现利普希茨连续

03

Self-Attention GAN

建模长距离依赖关系

04

BigGAN

BigGAN

Self-Attention GAN

原始GAN存在问题

局部效果很逼真，全局效果不逼真



Self-Attention GAN

原始GAN存在问题

局部效果很逼真，全局效果不逼真



放眼全局却有问题（5条腿？）

Self-Attention GAN

原始GAN存在问题



12

名

Self-Attention GAN

原始GAN存在问题

有腿没头?



12

名

Self-Attention GAN

引入注意力机制

原因何在？

原始GAN “眼界” 太窄，只关注局部，看不大整体

解决思路：

引入注意力机制，“拓宽视野”

- Self-Attention Generative Adversarial Network(SAGAN)是一个注意力驱动，长范围关联模型(attention-driven, long-range dependency modeling)。
- 传统的GAN在生成高分辨率的细节时，是基于低分辨率的feature map中的某一个小部分的。而SAGAN是基于所有的特征点(all feature locations)。
- 在训练时使用了光谱归一化(spectral normalization)来提升训练强度(training dynamics)。
- 可以很好的处理长范围、多层次的依赖(可以很好的发现图像中的依赖关系)
- 生成图像时每一个位置的细节和远端的细节协调好
- 判别器还可以更准确地对全局图像结构实施复杂的几何约束

Self-Attention GAN

SA-GAN

自注意力机制

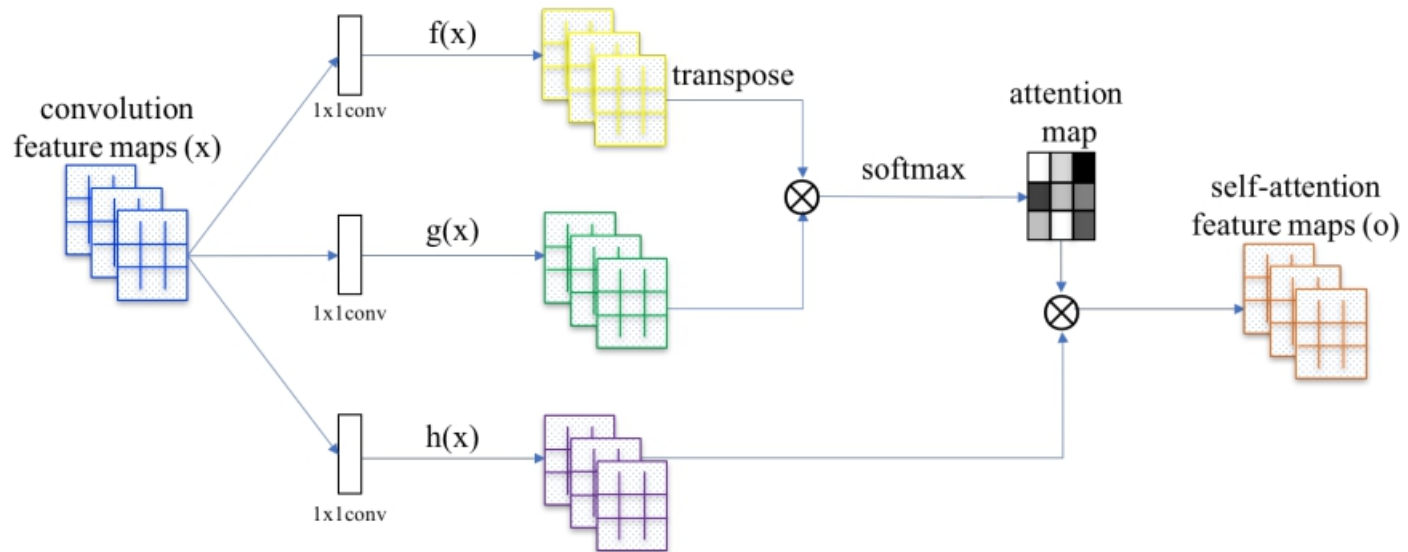
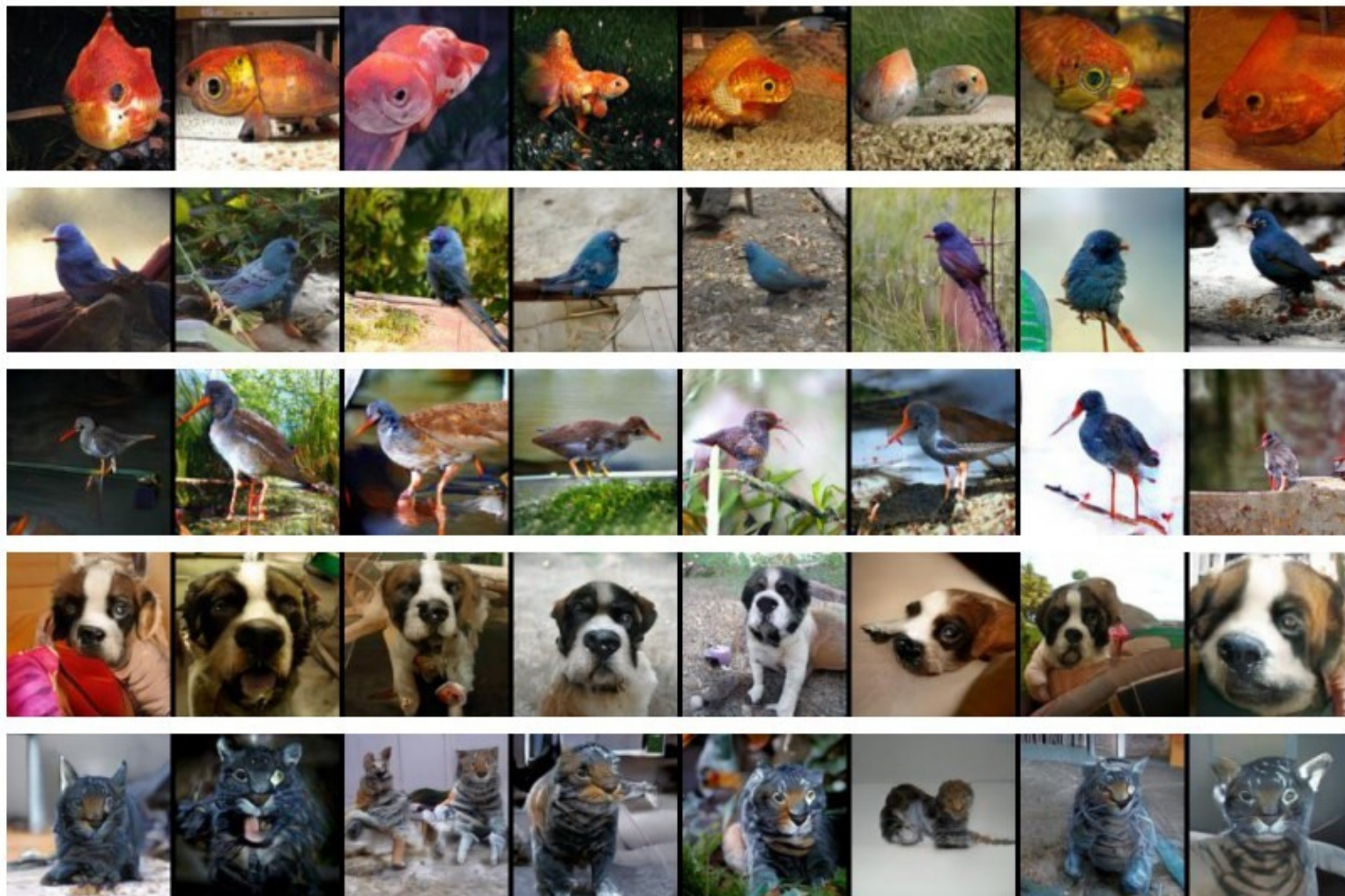


Figure 2: The proposed self-attention mechanism. The $\otimes$ denotes matrix multiplication. The softmax operation is performed on each row.

Self-Attention GAN

实验结果

局部和整体都很好



Self-Attention GAN

量化的实验结果

Model	Inception Score	FID
AC-GAN [31]	28.5	/
SNGAN-projection [17]	36.8	27.62*
SAGAN	52.52	18.65

Table 2: Comparison of the proposed SAGAN with state-of-the-art GAN models [19, 17] for class conditional image generation on ImageNet. FID of SNGAN-projection is calculated from officially released weights.

目录

CONTENTS

01

Progressive GAN

逐步生成，更稳更强大

02

SN-GAN

谱归一化层实现利普希茨连续

03

Self-Attention GAN

建模长距离依赖关系

03

BigGAN

BigGAN

BigGAN

实验结果



哪一行是假的？

BigGAN

实验结果



哪一行是假的？

BigGAN

实验结果



哪一行是假的？

BigGAN

增大Batch Size

Batch	Ch.	Param (M)	Shared	Hier.	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(± 1.18)
1024	64	81.5	✗	✗	✗	1000	14.88	63.03(± 1.42)
2048	64	81.5	✗	✗	✗	732	12.39	76.85(± 3.83)
2048	96	173.5	✗	✗	✗	295(± 18)	9.54(± 0.62)	92.98(± 4.27)
2048	96	160.6	✓	✗	✗	185(± 11)	9.18(± 0.13)	94.94(± 1.32)
2048	96	158.3	✓	✓	✗	152(± 7)	8.73(± 0.45)	98.76(± 2.84)
2048	96	158.3	✓	✓	✓	165(± 13)	8.51(± 0.32)	99.31(± 2.10)
2048	64	71.3	✓	✓	✓	371(± 7)	10.48(± 0.10)	86.90(± 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, *Hier.* is using a hierarchical latent space, *Ortho.* is Orthogonal Regularization, and *Itr* either indicates that the setting is stable to 10^6 iterations, or that it collapses at the given iteration. Other than rows 1-4, results are computed across 8 different random initializations.

增大Batch Size 就可以提高图像质量

BigGAN

增大Batch Size

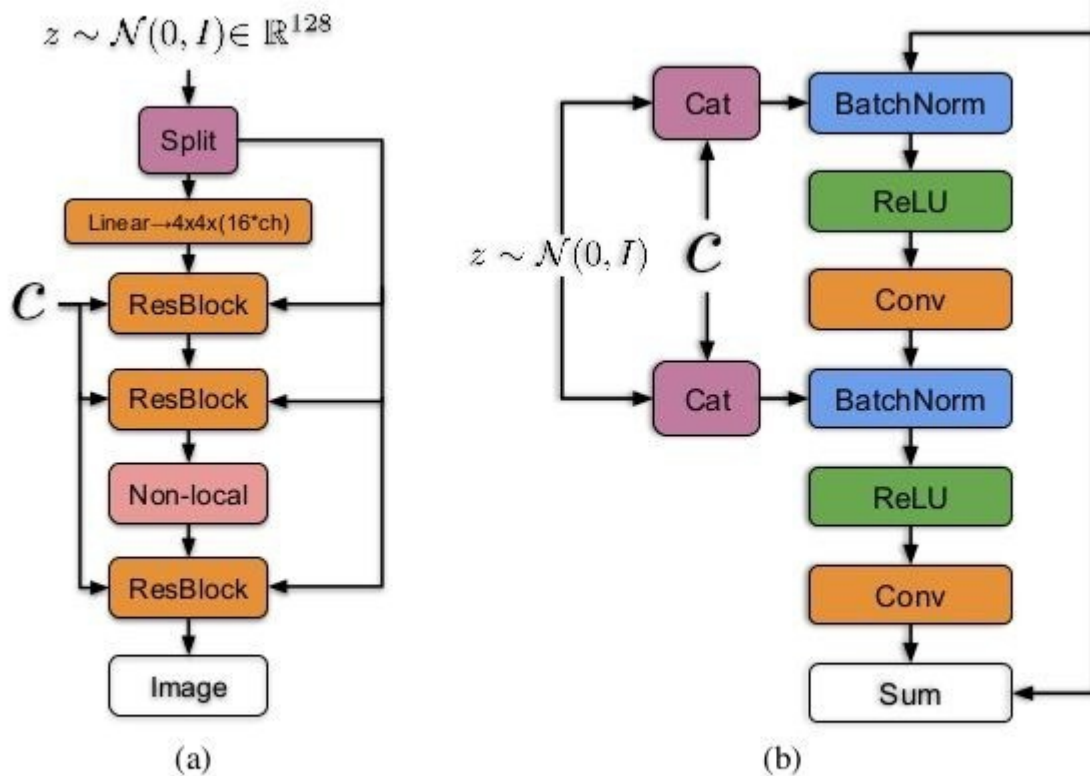


Figure 15: (a) A typical architectural layout for $\mathbf{G}$; details are in the following tables. (b) A Residual Block in $\mathbf{G}$. c is concatenated with a chunk of z and projected to the BatchNorm gains and biases.

噪声不单单加在输入层，还加在中间层

BigGAN

截断技巧

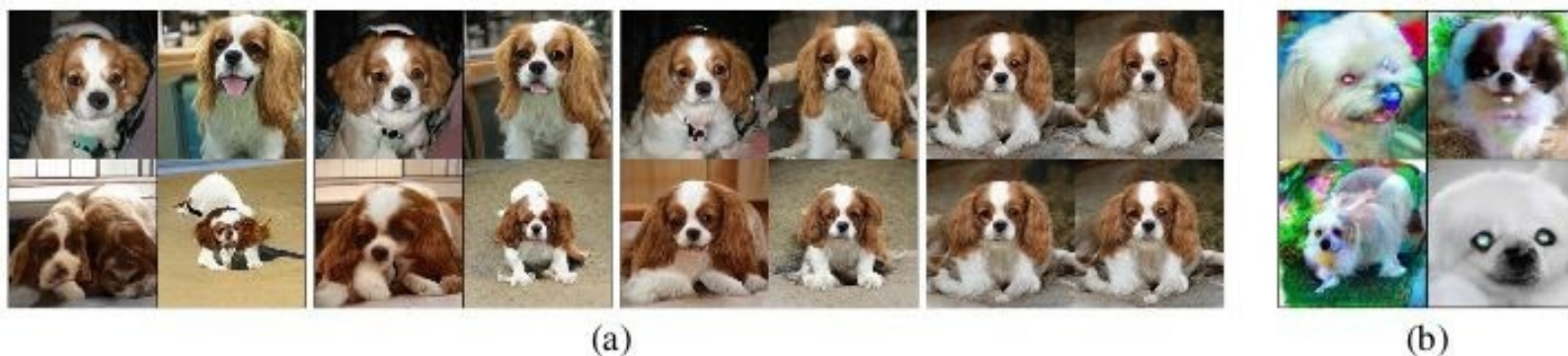


Figure 2: (a) The effects of increasing truncation. From left to right, threshold=2, 1.5, 1, 0.5, 0.04. (b) Saturation artifacts from applying truncation to a poorly conditioned model.

在一个分布中采样噪声，如果噪声超过阈值，则丢弃再次采样
阈值越大，图片多样性越好，但图片质量会略微下降
阈值越小，图片多样性越差，但图像质量更好





BigGAN

文章

Paper: Large Scale GAN Training for High Fidelity Natural Image Synthesis

<https://www.paperweekly.site/papers/2366>



THANKS