展开

举报

原创 213 粉丝 1780 获赞 882 评论 1369 访问 139万+

论文: LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS

最好的52.52提升到166.3, Frechet Inception Distance (FID, 越小越好) 由之前最好的18.65提升到9.6。

Github: https://github.com/AaronLeong/BigGAN-pytorch

原创 watersink 最后发布于2019-01-02 15:48:46 阅读数 7363 ☆ 收藏

Figure 1: Class-conditional samples generated by our model. 14845119

BigGAN作为GAN发展史上的重要里程碑,将精度作出了跨越式提升。在ImageNet (128*128分辨率)训练下,将Inception Score (IS,越大越好) 由之前

1. 通过2-4倍的增加参数量(增加channel),8倍的扩大batchsize,可以使GAN获得最大的性能提升。

Itr $\times 10^3$

1000

1000

1000

 $295(\pm 18)$

 $185(\pm 11)$

 $152(\pm 7)$

 $165(\pm 13)$

 $371(\pm 7)$

FID

18.65

15.30

14.88

 $9.54(\pm 0.62)$

 $8.73(\pm0.45)$

 $9.18(\pm 0.13)$ $94.94(\pm 1.32)$

 $8.51(\pm 0.32)$ $99.31(\pm 2.10)$

 $10.48(\pm 0.10)$ $86.90(\pm 0.61)$

IS

52.52

 $58.77(\pm 1.18)$

 $63.03(\pm 1.42)$

 $76.85(\pm 3.83)$

 $92.98(\pm 4.27)$

 $98.76(\pm 2.84)$

2. 通过使用截断技巧(truncation trick),可以使得训练更加平稳,但是需要在多样性和逼真度之间做平衡。

论文贡献:

- 3. 通过现存的和其他新颖的各种技术的集合,可以保证训练的平稳性,但是精度也会随之下降,需要在性能和训练平稳性之间做平衡。
- 背景知识:

其中, Z为服从正态分布N (0; I) 或者均匀分布U[-1; 1] 的随机噪声。

 $\min_{G} \max_{D} \mathbb{E}_{x \sim q_{\text{data}}(\boldsymbol{x})} [\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})} [\log (1 - D(G(\boldsymbol{z})))]$

Shared Hier. Ortho.

SA-GAN Baseline

Param:参数量

改进的式子如下,

ltr: 迭代次数

Batchsize和channel number增加:

Param (M)

81.5

81.5

81.5

81.5

173.5

160.6

158.3

158.3

2048 64 71.3

Batch

256

1024

2048

2048

2048

2048

2048 96

weighs和bias做投影,这样将会减少计算和内存的开销,将训练速度提升37%。

Ch.

	Table 1: Fréchet Inception Distance (FID, lower is better) and Inception Score (IS, higher is better) for ablations of our proposed modifications. <i>Batch</i> is batch size, <i>Param</i> is total number of parameters, <i>Ch.</i> is the channel multiplier representing the number of units in each layer, <i>Shared</i> is using shared embeddings, <i>Hier.</i> is using a hierarchical latent space, <i>Ortho.</i> is Orthogonal Regularization, and <i>Itr</i> 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.
Batch: batch size	
Ch: channel	

Hier: 使用分层的潜在空间(hierarchical latent space),也就是在生成器网络的每一层都会被喂入噪声向量。分层空间增加了计算量和内存消耗,但是 可以使得精度提升4%,训练速度加快18%。 Ortho: 使用正交正则化(Orthogonal Regularization) 最原始的正交正则化如下面式子,

Shared:使用共享嵌入(shared embeddings),在生成器的BN层中加入类别信息嵌入,将会大大的增加参数量,因此选择共享嵌入,也就是对每层的

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

其中W为权重矩阵, β 为超参数。 论文通过将对角线的元素全部去掉,实现了对其的改进,即放松了截断的限制,也保证了模型的平滑性。

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

IS: Inception Score,表示生成图像的质量,越大越好

其中,1表示一个全部为1的矩阵。

从上图可以看出,随着batchsize和channel数的增加,精度越来越高,同时论文还验证将深度增加为原来的2倍,将会使得性能下降。

FID: Frechet Inception Distance,表示生成图像的多样性,越小越好

截断技巧(truncation trick):

(b)

所谓的"截断技巧"就是通过对从先验分布 z 采样,通过设置阈值的方式来截断 z 的采样,其中超出范围的值被重新采样以落入该范围内。

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 administry of 14845119

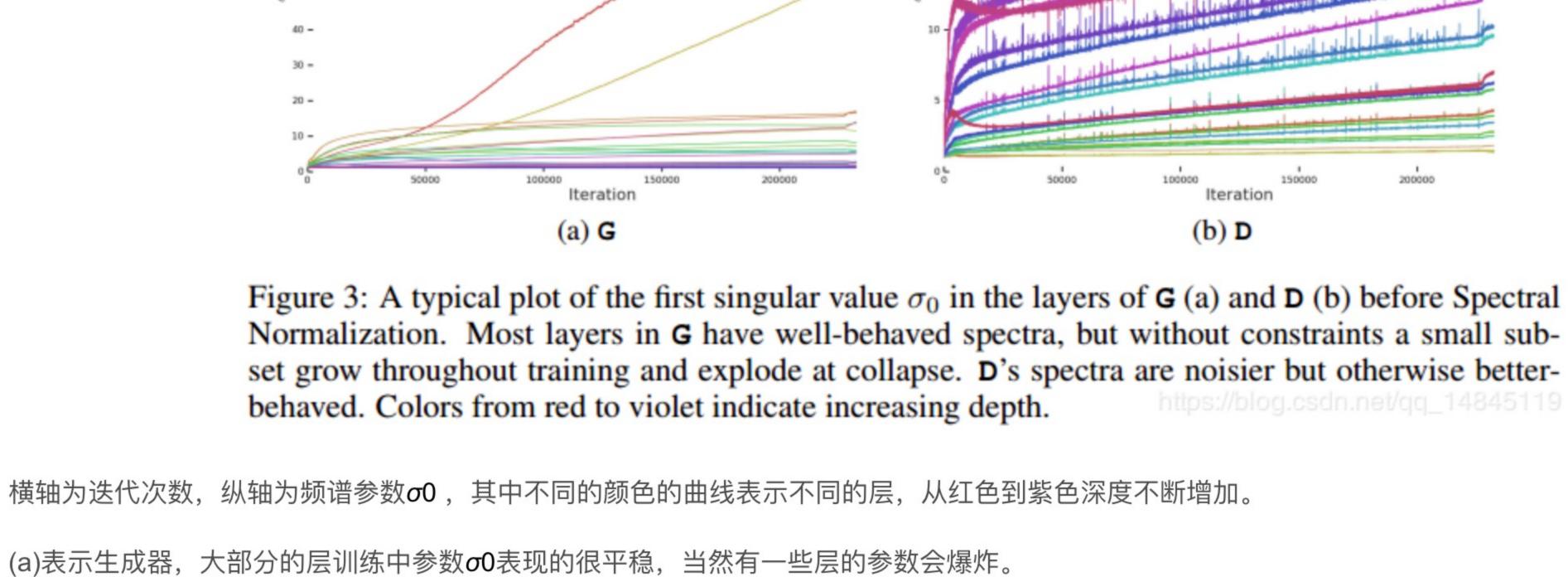
(b)对较弱的模型使用截断会带来饱和想象(伪影)。为了解决该问题,问题引入了正交正则化的改进版。

特征的平稳性:

(a)从左到右依次增加截断的区间,阈值分别为2,1.5,1,0.5,0.04,可以看出随着截断区间的增加,生成图像质量(FIDELITY)越来越高,但是多样性

Collapse .

(VARIETY) 越来越差。



(b)表示判别器,大部分的层都有很大的噪声

因此,首先考虑对 σ 0进行正则化操作。

生成器特征的平稳性:

最终结果,不管有无频谱归一化(Spectral Normalization),这样的改进都可以提升训练的平稳性,但是不能解决爆炸的问题。因此论文对判别器进行改 进。

 $z \sim \mathcal{N}(0, I) \in \mathbb{R}^{128}$

V0,u0表示奇异值向量, $\sigma clamp$ 或者被设置为固定的值 σreg 或者被设置为 $r \cdot sg(\sigma 1)$,表示 $\sigma 1$ 的的r倍。

在对权重矩阵进行奇异值分解,前3个值(σ 0, σ 1, σ 2),对每个矩阵具有最大的影响。

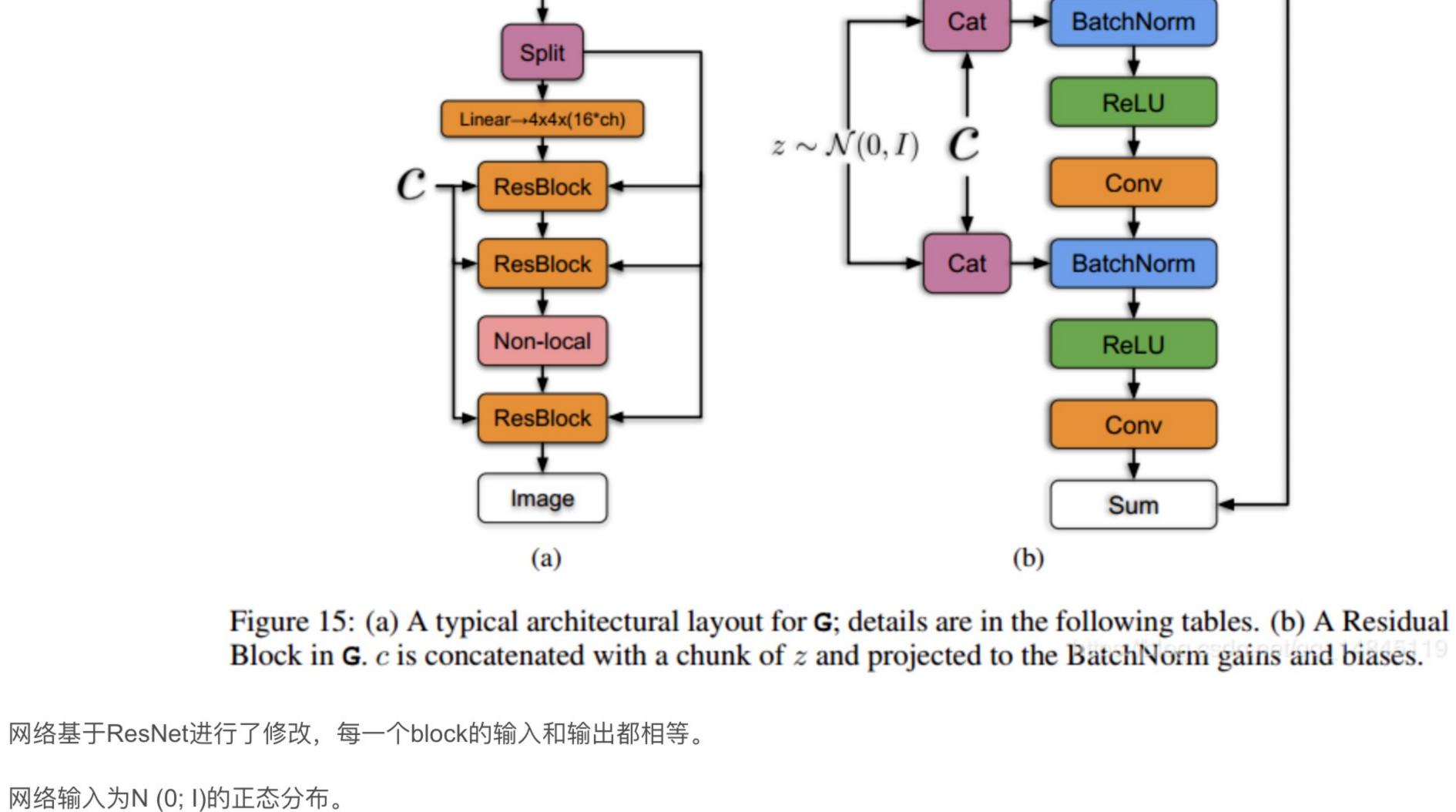
判别器特征的平稳性: 论文假设频谱噪声是导致训练不平稳的因素,因此很自然的想法是进行梯度惩罚,通过对R1进行0均值的正则化惩罚来进行改善。

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

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

其中,r表示常数,10。 论文实验表明,惩罚越大,训练越平稳,但是精度也越来越低。

网络结构:



左侧实心线c表示共享嵌入(shared embeddings),右侧实心线表示分层的潜在空间(hierarchical latent space)。(a)表示整体网络结构,(b)表

实验:

SA-GAN

BigGAN

BigGAN

BigGAN

256

512

Param (M)

317.1

99.4

207.9

355.7

128

Shared

Hier.

IMAGENET实验: FID / (max IS) FID/IS (min FID) / IS FID / (valid IS) Model Res. SN-GAN 27.62 / 36.80 N/A N/A N/A

N/A

 $7.7 \pm .1/126.5 \pm .1$

 $7.7 \pm .1/178 \pm 5$

9.3 / 202.5

N/A

 $9.6 \pm .4/166.3 \pm 1$

 $9.3 \pm .3/233 \pm 1$

10.9 / 241.4

(min FID) / IS

48.6 / 23.1

22.4 / 21.0

17.1 / 23.3

13.0 / 28.0

N/A

 $25 \pm 2/206 \pm 2$

 $25 \pm 5/295 \pm 4$

24.4/275

FID / (max IS)

49.1 / 23.9

60.9 / 35.8

51.6 / 38.1

46.2 / 47.8

18.65 / 52.52

 $8.7 \pm .6/98.8 \pm 2.8$

 $8.2 \pm .2/154 \pm 2.5$

10.9 / 154.9

示每个block结构。

3), scores at the best FID	(Column 4), scores	olutions. We report scores without the IS of validation data (Core computed over at least three	olumn 5), and scores at

(a) 128×128 (b) 256×256 (c) 512×512 Figure 4: Samples from our model with truncation threshold 0.5 (a-c) and an example of class leakage in a partially trained model (d).

FID

48.38

23.48

18.84

13.75

23.27

24.78

27.86

30.61

Ortho.

JFT-300M实验:

Table 3: Results on JFT-300M at 256×256 resolution. The FID and IS columns report these				
scores given by the JFT-300M-trained Inception v2 classifier with noise distributed as $z \sim \mathcal{N}(0, I)$				
(non-truncated). The (min FID) / IS and FID / (max IS) columns report scores at the best FID and				
IS from a sweep across truncated noise distributions ranging from $\sigma = 0$ to $\sigma = 2$. Images from the				
JFT-300M validation set have an IS of 50.88 and FID of 1.94. https://blog.csdn.net/qq_14845119				

1. 增加网络深度会使得精度降低。 2. 在判别器上使用贡献嵌入参数的方法,对参数的选择非常敏感,刚开始有助于训练,后续则很难优化。 3. 使用WeightNorm 替换BatchNorm 会使得训练难以收敛,去掉BatchNorm 只有Spectral Normalization 也会使得难以收敛。

负面结果:

- 6. 相比采用3*3的滤波器,采用5*5的滤波器会使精度有略微提升,而7*7则不会。
- 4. 判别器中增加BatchNorm 会使得训练难以收敛。 5. 在128*128的输入情况下,改变attention block对精度没提升,在256*256输入的情况下,将attention block上移一级,会对精度有提升。
- 7. 使用膨胀卷积会降低精度 8. 将生成器中的最近邻插值换为双线性插值会使得精度降低。
- 9. 在共享嵌入中使用权值衰减(weight decay),当该衰减值较大(10-6) 会损失精度,较小(10-8) 会起不到作用,不能阻止梯度爆炸。 10. 在类别嵌入中,使用多层感知机(MLP)并不比线性投影(linear projections)好。
- 11. 梯度归一化截断会使得训练不平稳。
- 总结: BigGAN是GAN发展史上新的里程碑,论文贡献包括,大batchsize,大channel数,截断技巧,训练平稳性控制等。

凸 点赞 2 ☆ 收藏 🖸 分享 …

watersink

想对作者说点什么



发布了213 篇原创文章·获赞 882·访问量 139万+

显示推荐内容

CSDN Greener V2.1.4 By GitHub :: AdlerED 风车小站 9个月前 请问论文中改进后的正交正则化公式怎么理解呢?

关注