+ 关注她

ICLR'19谷歌大规模GAN训练论文深度解读

曼阿里

4 人赞同了该文章

曼殊 阿里-Al Chip Research Scientist

谷歌在ICLR'19上发表的论文Large Scale GAN Training for High Fidelity Natural Image Synthesis,成功的从复杂数据级ImageNet中生成了以假乱真的图片,本文将对论文的细节进行深度解读

尽管近来生成图像建模取得了许多进展,从复杂的数据集例如ImageNet中成功地生成高分辨率多样的图像样本依然是一个困难的目标。该文在大数据集上训练了GAN,并且研究了在这种规模下所特有的不稳定性(instability)。发现在生成器(generator)上应用正交正则化可以使其符合简单的阶段技巧"truncation trick",即对样本的保真度(fidelity)和多样性(variety)之间的权衡进行精确控制。本文的在分类别调解下的图像生成上得到了新的state-of-the-art结果。当使用128x128 ImageNet 训练时,模型取得了166.3的Inception Score (IS)和9.6的Frechet Inception Distance值,与之前的最好IS相比提高了52.52的IS和18.65的FID。

一、主要创新点

本文主要创新在于在大型数据集ImageNet实现了GAN模型性能的大幅度提升,技术上创新点如下:

(一)本文对网络结构进行了两个修改,改善了可扩展性(scalability);修改了正则化方法(regularization)提高了条件化性能(conditioning)。结果证明了扩大GAN模型的规模(多2-4倍参数,8倍batchsize)能够大幅度提高性能。

(二)本文对模型的修改使得模型可以使用一种截断技巧(truncation trick),即使用简单的采样技术实现对采样多样性(variety)和保真度(fidelity)之间的权衡的精确控制

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	X	X	X	1000	15.30	$58.77(\pm 1.18)$
1024	64	81.5	X	X	X	1000	14.88	$63.03(\pm 1.42)$
2048	64	81.5	X	X	X	732	12.39	$76.85(\pm 3.83)$
2048	96	173.5	X	X	X	$295(\pm 18)$	$9.54(\pm 0.62)$	$92.98(\pm 4.27)$
2048	96	160.6	1	X	X	$185(\pm 11)$	$9.18(\pm 0.13)$	$94.94(\pm 1.32)$
2048	96	158.3	/	/	X	$152(\pm 7)$	$8.73(\pm0.45)$	$98.76(\pm 2.84)$
2048	96	158.3	1	1	1	$165(\pm 13)$	8.51(±0.32)	99.31(+2.10)
2048	64	71.3	/	/	/	$371(\pm 7)$	$10.48(\pm 0.10)$	86.90(±0.61)

表1.数据从左到右依次为: batchsize, 通道(channel)数,参数数量,本文使用的三种技术(shared class embedding, hierarchical latent spaces, orthogonal regularization),训练所需的iteration数量,训练得到的GAN的FID值和IS值

二、技术改进

(一) Shared Class Embedding: 对于G(generator,生成器)的条件化,文章的GAN采用分享class embedding (shared class embedding),线性投射到每一层的gain和bias。与已有的G条件化方法(每层都有各自的class embedding),能够降低计算和存储开销,并加快训练速度37%(第六行第七行对比,295à185 iterations)。

(二)Hierarchical Latent Spaces: 噪声向量z输入到G的所有层,而不是传统的只输入给第一层。

这种方法使得G能够使用latent space来直接影响不同分辨率和分层(hierachy)中的特征。具体方法是将z给每个分辨率分一个块(chunk),然后与映射到每层的条件向量(conditional vector)c相连。这种方法能够提升4%的性能(第七行和第八行对比,IS 94.94à98.76),18%的训练速度(第七行和第八行对比,185à152 iterations)。

(三) Truncation Trick: 本文在采样(sampling)和训练(training)过程中使用了不同的latent

distribution。当训练使用的z~N(0,I),采样使用截断正态分布(落在大于阈值范围的值会被重采样

成小于阈值的值)时,这种做法可以显著GAN的IS和FID性能。这个 truncation 是通过对阈值外的样

本进行重采样来提高单个样本的质量,以牺牲样本多样性为代价。因此通过对采样阈值大小的控制(truncation trick),可以控制GAN生成图片的质量(保真度)与多样性之间的均衡。



文中Figure2(a)中四张图,从左到右的重采样阈值依次降低,生成图像的多样性(variety)降低,

保真度(fidelity)提高。推测原因是因为当阈值降低时,噪声z被向0方向截断,每个样本都接近G的输出分布。



Figure 2(b) 描述了大多类别的图片,当使用截断后的噪声时,生成图片会产生饱和效应(saturation artifacts),文章使用正交正则化(orthogonal regularization)解决了这个问题,使更多类别的模

Figure2(b) 使用GAN 生成图片的饱和效应

型符合上文提到的截断技巧(16%->60%) 表1中结果表明,增加8倍batchsize可以使IS提升46%,推测原因是大batchsize能够覆盖更多模式,

能,坏处是训练变得不稳定,最后训练会崩溃。 增加每层的宽度(通道数量)50%,即增加一倍的参数可以提高21%的IS,推测原因是提高了模型的容量。而提高模型的深度反倒降低了模型的性能。

给两个神经网络提供更好的梯度信息。这样做的好处是能够在更少的iteration内达到更好的最终性

三、所使用的GAN技术

1. 训练方法上使用spectral normalization来提升稳定性,与标准SN比,学习率降低一半;每个G step后进行2个D step

2.架构选择上使用SA-GAN,目标函数使用hinge-loss

3. 条件(conditioning)数据通过改变BatchNorm的增益(gain)和偏置(bias)来输入给G;对于

D,则使用特征与学习到的类别嵌入信号(class embedding)之间cosine相似度作为一个额外的evidence来分离真实样本和生成样本(projection)

4. 在G的权重上使用decay为0.9999的moving average
5. 使用128-512个TPU核,G的BatchNorm各项数据在所有device的batch中计算

1. easy vs hard class(Appendix-A), 作者发现有一些类别的图片比较容易生成,而其他一些类别难度较高

四、其他技术细节与结果(附录)

图片的质量可达以假乱真的程度。

2. interpolation and nearest neighbors(Appendix-A), 作者观察到Discrimniator 对数据有过拟合, 用此方法来证明Generator不是简单的"记忆" 训练数据

结构细节(Appendix-B), 给出了所使用的GAN的神经网络的结构细节
 实验细节(Appendix-C),给出了训练GAN所使用的一些技术细节

6. training collapse (Appendix-F), 训练大规模GAN时,最后训练会崩溃。可以通过使用技术如 R1

5.噪声向量z的分布(Appendix-E), 讨论了不同噪声向量分布latent space的选择对大规模GAN训练性能的影响

zero-centered gradient penalty 避免这种崩溃,但是最后达到的最好性能也会降低

8. 超参数值(Appendix-H),列出了训练网络所使用的超参数(hyper-parameter)的具体数值

7. negative results(Appendix-G),列出了一些失败了的试验结果

五、技术与结果分析 文章得到了GAN在ImageNet上的新的state-of-the-art结果,除了FID和IS性能达到目前最好,生成

显。因此作者在接受采访时自述,本文的结果是"算力的胜利" 使用文中提出的三个方法,对大规模GAN训练的性能体现在提升训练速度,例如,表1中第6行与第 9行,对于相同结构的网络,文中方法加速训练40%(295->165),提升性能FID 11%,IS 6.8%; 表1中第5行与第10行,加速49%(732->371),性能改善FID 15.4%, IS 13.1%

结果的性能提升,主要来自于网络规模的扩大。具体来讲,当batchsize扩大8倍,参数量扩大2

倍,所带来的性能提升高达80%(表1中第二行与第六行,IS 52.52->92.98),在规模扩大的基础

上,使用本文提出的三个方法,可以进一步提升性能,但是带来的性能收益不如扩大网络规模明

明显(即所需的iteration数量降低),两个case的提升都有~40%。

而对于性能的提升在网络规模较大时已不太明显(row6 vs row 9),因此文章工作对于性能的提升主要还是由于网络规模的扩大(8倍batchsize提升46%),但在规模较小时文章的方法也可以对性

从文中给出的数值结果看,该文章的技术从两方面提升了GAN的可扩展性(scalability),提升训

练速度,和提高GAN的性能。通过对相同结构的网络进行对比可以看出,其对训练速度的提升更为

后续可能会公布的结果上进行验证。 六、**现有的一些对文章的讨论**

以上讨论基于文章中表1的结果,但只给出了2个可以比较的case,上述的结论还需要在deepmind

https://blog.csdn.net/yH0VLDe8VG8ep9 VGe/article/details/83020981?...

48个小时。

才能见到成品。

能有所提升(row 5 vs row9, ~15%)

Ø blog.csdn.net
 "在这场实验中,Brock等人生成512像素图像,需要的是512块谷歌TPU,并且训练时间会持续24到

1. 训练史上最佳GAN用了512块TPU,一作自述:这不是算法进步,是算力进步

光有TPU加持可能还不够,你还得找一个电量充足的地方。大致估算一下,如果每个TPU每个小时需要200瓦的电量,那么在Brock的512像素实验中,每个TPU将耗费2450到4915度电,相当于一个普通美国家庭大约半年内的用电量"

2. google开放了训练大GAN的平台

史上最强GAN: 训练费10万起,现在免费体验,画风鬼畜又逼真 史上最强GAN: 训练费10万起,现在免费

体验,画风鬼畜又逼真 - 云+社区 - 腾讯云

BigGAN模型,要用128个Google TPU 核心。256×256、512×512模型需要的TPU核心数也相应上涨到了256个和512个。

更惊悚的是,用了这么多TPU的情况下,大部分模型还要训练24到48小时,也就是要等上一两天

根据论文附录中提供的细节,BigGAN是在TPU Pod上训练出来的。训练一个生成128×128图像的

最最最低配的那个,需要1.38万美元到2.76万美元,折合人民币9.6万元到19.3万元。
至于512×512的高清大GAN,训练费用最高可以达到11万美元,合人民币76万元。

按照Cloud TPU v2每TPU每小时4.5美元的价格来算,**训练一个基础版128×128的BigGAN,也就是**

编辑于 2019-03-18

人工智能 生成对抗网络 (GAN) 谷歌 (Google)

人工智能热门论文深度解读

文章被以下专栏收录

推荐阅读

关注专栏









新智元 发表于AI前沿研...

写下你的评论...