

学界 | 为提高质量、稳定性和变化性不断推陈出新

2017-11-15 机器海岸线

选自 arXiv

作者: Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen 等

机器海岸线编译

参与: 方建勇

PROGRESSIVE GROWING OF GANs FOR IMPROVED QUALITY, STABILITY, AND VARIATION

Tero Karras
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

Jaakko Lehtinen
NVIDIA and Aalto University

{tkarras, taila, slaine, jlehtinen}@nvidia.com

论文链接:

<http://research.nvidia.com/sites/default/files/publications/karras2017gan-paper-v2.pdf>

摘要: 我们描述了一个新的生成对抗网络的训练方法。关键的思想是逐步发展生成器和鉴别器: 从低分辨率开始, 随着训练的不断进展, 我们增加了新的层, 可以模拟日益精细的细节。这既加快了训练速度, 又大大稳定了训练速度, 降低了我们产生前所未有的质量的图像, 例如 1024*1024 的 CELEBA 图像。我们还提出了一种简单的方法来增加生成的图像的变化, 并实现了最好的纪录, 在无人监督的 CIFAR10 中得分为 8.80。此外, 我们描述了几个实施细节, 这对于阻止发生器和鉴别器之间的不健康竞争是重要的。最后, 我们提出了一个评估 GAN 结果的新指标, 无论是在图像质量还是变化方面。作为额外的贡献, 我们构建了更高质量的 CELEBA 数据集版本。

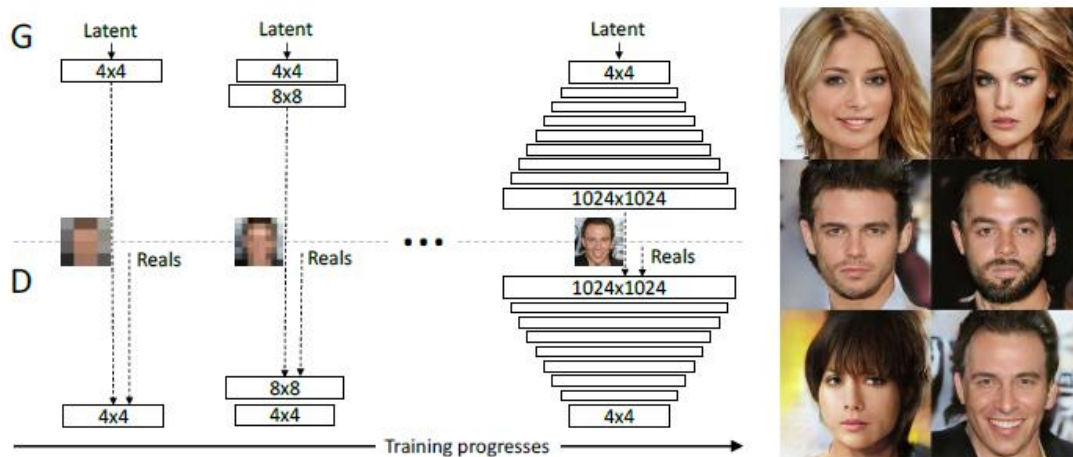


图 1: 我们的训练从具有 4×4 像素的低空间分辨率的生成器 (G) 和鉴别器 (D) 开始。随着训练的进展, 我们逐渐增加层到 G 和 D, 从而增加了生成的图像的空间分辨率。在整个过程中, 所有现有的层次都可以训练。这里 $N \times N$ 是指在 $N \times N$ 空间分辨率下工作的卷积层。这可以在高分辨率下实现稳定的合成, 并且可以大大加快训练。右侧为我们显示了在 1024×1024 分辨率下六个图像生成的例子。

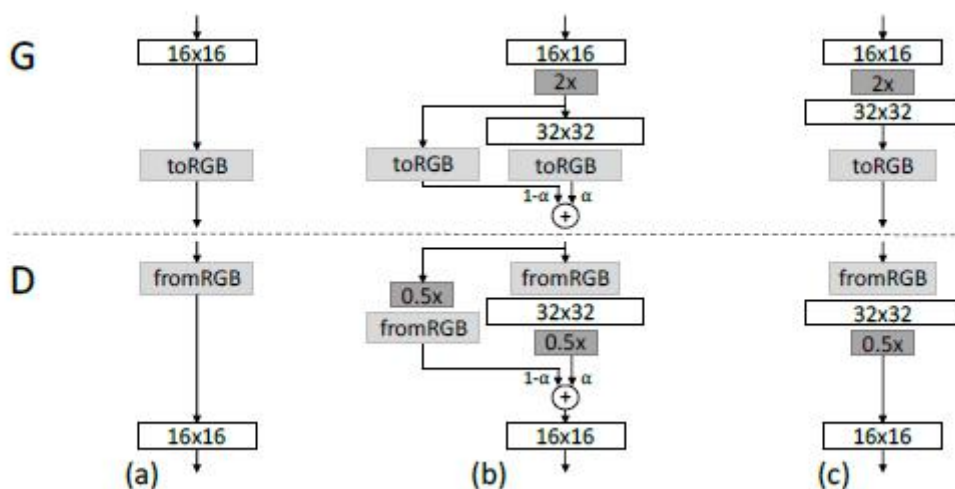


图 2: 当把生成器 (G) 和鉴别器 (D) 的分辨率加倍时, 我们平滑地淡入新层。这个例子说明了从 16×16 图像 (a) 到 32×32 图像 (c) 的转换。在过渡过程中 (b), 我们把像高分辨率的图像处理成一个残差块, α 的权重从 0 线性增加到 1。这里 $2 \times$ 和 $0.5 \times$ 指的是使用最近邻域滤波和平均值将图像分辨率加倍和减半。汇集, 分别。toRGB 表示一个将特征向量投影到 RGB 颜色的图层, 而 fromRGB 则反过来; 两者都使用 1×1 卷积。在训练鉴别器时, 我们会输入缩小的实际图像以匹配网络当前的分辨率。在分辨率转换过程中, 我们在实际图像的两个分辨率之间插值, 类似于生成器输出如何组合两个分辨率。

Training configuration	CELEB-A						LSUN BEDROOM					
	Sliced Wasserstein distance $\times 10^3$					MS-SSIM	Sliced Wasserstein distance $\times 10^3$					MS-SSIM
	128	64	32	16	Avg		128	64	32	16	Avg	
(a) Gulrajani et al. (2017)	12.99	7.79	7.62	8.73	9.28	0.2854	11.97	10.51	8.03	14.48	11.25	0.0587
(b) + Progressive growing	4.62	2.64	3.78	6.06	4.28	0.2838	7.09	6.27	7.40	9.64	7.60	0.0615
(c) + Small minibatch	75.42	41.33	41.62	26.57	46.23	0.4065	72.73	40.16	42.75	42.46	49.52	0.1061
(d) + Revised training parameters	9.20	6.53	4.71	11.84	8.07	0.3027	7.39	5.51	3.65	9.63	6.54	0.0662
(e*) + Minibatch discrimination	10.76	6.28	6.04	16.29	9.84	0.3057	10.29	6.22	5.32	11.88	8.43	0.0648
(e) Minibatch stddev	13.94	5.67	2.82	5.71	7.04	0.2950	7.77	5.23	3.27	9.64	6.48	0.0671
(f) + Equalized learning rate	4.42	3.28	2.32	7.52	4.39	0.2902	3.61	3.32	2.71	6.44	4.02	0.0668
(g) + Pixelwise normalization	4.06	3.04	2.02	5.13	3.56	0.2845	3.89	3.05	3.24	5.87	4.01	0.0640
(h) Converged	2.95	2.38	1.98	5.16	3.12	0.2880	3.26	3.06	2.82	4.14	3.32	0.0633

表 1: 生成的图像与训练图像(第 5 节)和多尺度结构相似性(MS-SSIM)之间的切片 Wasserstein 距离(SWD)在 128 128 的几个训练设置生成的图像。对于 SWD, 每列代表拉普拉斯 金字塔, 最后给出了四个距离的平均值。



图 3: (a) - (g) 与表 1 中的行相对应的 CELEBA 示例。这些是有意不聚合的。(h) 我们的收敛结果。请注意, 有些图像显示锯齿, 有些图像不锐利 - 这是数据集的一个缺陷, 模型学习如实地复制。

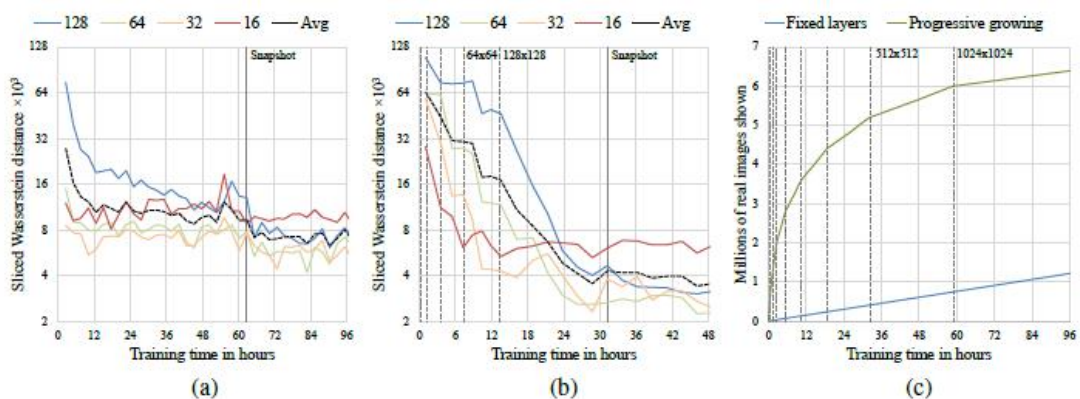


图 4: 渐进式增长对训练速度和收敛的影响。计时是在使用 NVIDIA Tesla P100 的单 GPU 设置上测量的。(a) Gulrajani 等人关于挂钟时间的统计相似性。(2017) 以 128 * 128 分辨率使用 CELEBA。每张图表示拉普拉斯金字塔一个层次上的切片 Wasserstein 距离, 垂直线表示停止表 1 中训练的点。(b) 启用渐进式增长的相同图形。竖直的虚线表示将 G 和 D 的分辨率加倍的点。(c) 逐步增长对 1024 * 1024 分辨率的原始训练速度的影响。



图 5: 使用 CELEBA-HQ 数据集生成 1024×1024 的图像。有关更多的结果, 请参阅附录 F, 以及潜在的空间插值附带的视频。在右侧, 来自 Marchesi (2017) 的较早的百万像素 GAN 的两幅图像显示了有限的细节和变化。



图 6: LSUN 卧室的视觉质量比较; 从引用的文章复制的图片。



图 7: 从不同的LSUN 类别生成的 $256 * 256$ 图像的选择。

本文为机器海岸线编译，转载请联系 fangjianyong@zuaa.zju.edu.cn 获得授权。

