CVPR2020之MSG-GAN：简单有效的SOTA 已付费

原创 bryant8 机器学习与生成对抗网络 3月11日

点击上方"**机器学习与生成对抗网络**"，关注"星标"

获取有趣、好玩的前沿干货！

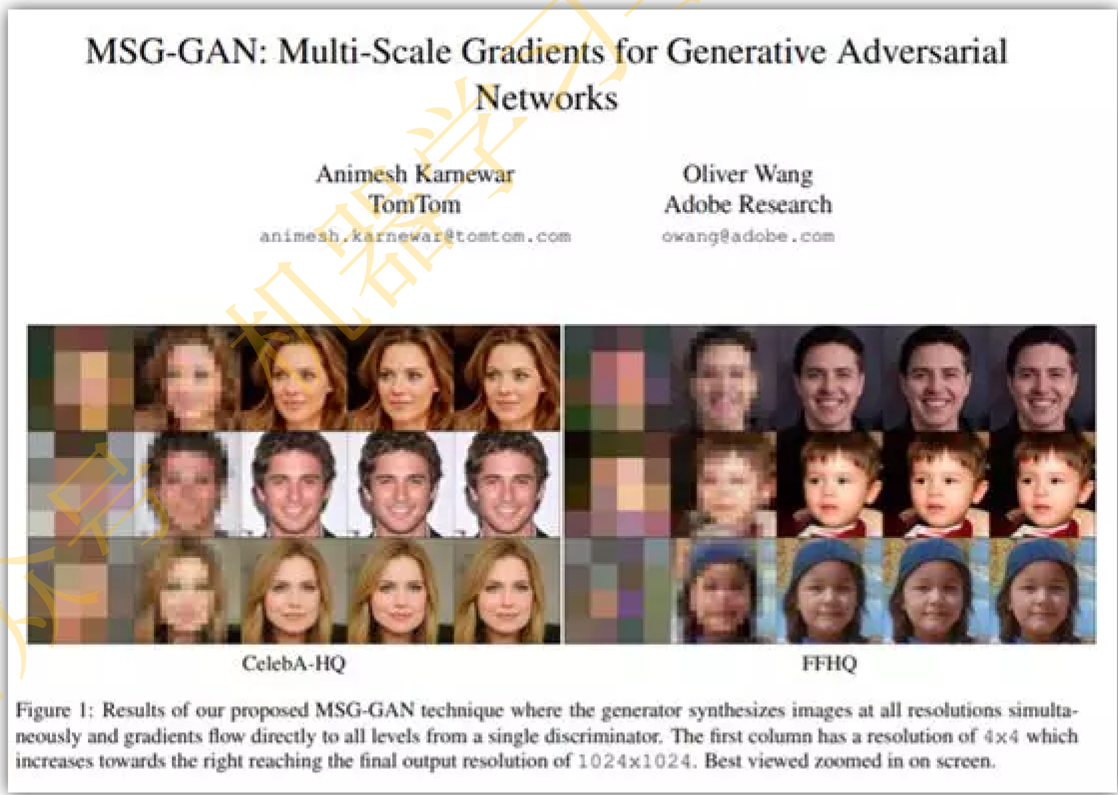
今天不知在哪儿、看到了说这篇**MSG-GAN**已被**CVPR2020**接收，其实去年2019年5月还是6月我就看到它了，当时试跑了一下开源的代码，印象中效果不错。今天就**极简**分享一下。

😅 😂

MSG-GAN: Multi-Scale Gradient GAN for Stable Image Synthesis

https://arxiv.xilesou.top/pdf/1903.06048.pdf

https://github.com/ akanimax/msg-stylegan-tf



MSG-GAN: Multi-Scale Gradients for Generative Adversarial Networks

Animesh Karnewar
TomTom
animesh.karnewar@tomtom.com

Oliver Wang
Adobe Research
owang@adobe.com

CelebA-HQ FFHQ

Figure 1: Results of our proposed MSG-GAN technique where the generator synthesizes images at all resolutions simultaneously and gradients flow directly to all levels from a single discriminator. The first column has a resolution of 4x4 which increases towards the right reaching the final output resolution of 1024x1024. Best viewed zoomed in on screen.

生成对抗网络（**GAN**）及其变体在图像合成任务取得了巨大成功，但众所周知，它们很难适应不同的数据集，部分原因是训练期间不稳定和对超参数敏感。对于这种不稳定，一种普遍的观点是：当真实分布和生成分布的支撑集不够重叠时，鉴别器反馈给生成器的梯度是无法提供有益的信息的。

在这项工作中，提出了多尺度梯度生成对抗网络（**MSG-GAN**），一种简单但有效的手段，通过在多个尺度上从鉴别器到生成器提供梯度。这可为高分辨率图像合成提供了一种稳定训练的方法，也可以替代常用的渐进式生长的**ProGAN**。作者表明**MSG-GAN**在不同的尺寸、分辨率和图像域的数据集，以及不同类型的损失函数和网络结构下，都可以使用相同的一组超参数稳定收敛。与先进的**GAN**相比，该方法在大多数情况下具有优势。

方法还是非常简明易懂的，网络结构如下：

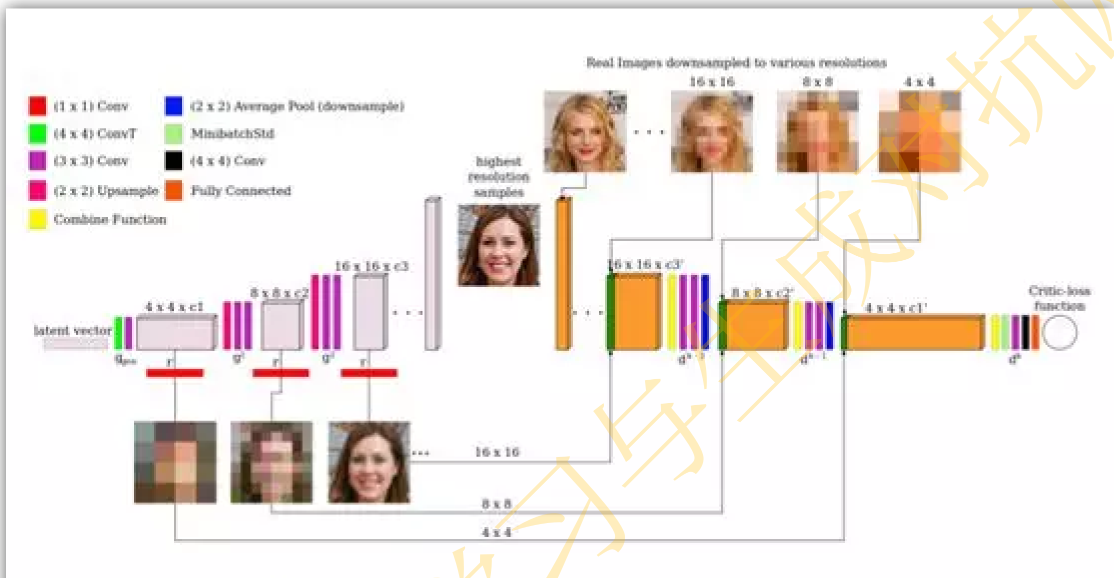


Figure 2: Architecture of MSG-GAN, shown here on the base model proposed in ProGANs [13]. Our architecture includes connections from the intermediate layers of the generator to the intermediate layers of the discriminator. Multi-scale images sent to the discriminator are concatenated with the corresponding activation volumes obtained from the main path of convolutional layers followed by a combine function (shown in yellow).

实现细节：

We evaluate our method on a variety of datasets of different resolutions and sizes (number of images); **CIFAR10** (60K images at 32x32 resolution); **Oxford flowers** (8K images at 256x256), **LSUN** churches (126K images at 256x256), **Indian Celebs** (3K images at 256x256 resolution), **CelebA-HQ** (30K images at 1024x1024) and **FFHQ** (70K images at 1024x1024 resolution).

For each dataset, we use the same initial latent dimensionality of **512**, drawn from a standard normal distribution $\mathcal{N}(0, \mathbb{I})$ followed by hypersphere normalization [13]. For all experiments, we use the same hyperparameter settings for MSG-ProGAN and MSG-StyleGAN (lr=0.003), with the only differences being the number of upsampling layers (fewer for lower resolution datasets).

All models were trained with **RMSprop** and a learning rate of **0.003** for both generator and discriminator. We initialize the parameters of the generator and discriminator according to the standard normal $N(0, \mathbb{I})$ distribution. To match the previously published work, all StyleGAN and MSG-StyleGAN models were trained with Non-saturating GAN loss with 1-sided GP while ProGANs and MSG-ProGAN models were trained with the WGAN-GP loss function.

We also extend the MinBatchStdDev technique [13, 14], where the average standard deviation of a batch of activations is fed to the discriminator to improve sample diversity, to our multiscale setup. To do this, we add a separate MinBatchStdDev layer at the beginning of each block in the discriminator. This way, the discriminator obtains batch-statistics of the generated samples along with the straight-path activations at each scale, and can detect some degree of mode collapse by the generator.

其中，所使用的数据集如上所示，其中，**3千Indian Celebs**为作者自制。对每种数据集，都是从高斯分布采样**512**维噪声去生成。更详细的参数设置可参考原文。

实验：

在**256X256**图像上：

Dataset	Size	Method	# Real Images	GPUs used	Training Time	FID (↓)
Oxford Flowers (256x256)	8K	ProGANs*	12M	4 GTX1080-8GB	75 hrs	58.60
		MSG-ProGAN	1.7M	1 V100-32GB	44 hrs	28.27
		StyleGAN*	7.2M	2 V100-32GB	33 hrs	64.70
		MSG-StyleGAN	1.6M	2 V100-32GB	16 hrs	**19.60**
Indian Celebs (256x256)	3K	ProGANs*	9M	2 V100-32GB	37 hrs	67.49
		MSG-ProGAN	2M	2 V100-32GB	34 hrs	36.72
		StyleGAN*	6M	4 V100-32GB	18 hrs	61.22
		MSG-StyleGAN	1M	4 V100-32GB	7 hrs	**28.44**
LSUN Churches (256x256)	126K	StyleGAN*	25M	8 V100-16GB	47 hrs	6.58
		MSG-StyleGAN	24M	8 V100-16GB	50 hrs	**5.2**

Table 1: Experiments on mid-level resolution (*i.e.* 256x256) datasets. We use author provided scores where possible, and otherwise train models with the official code and recommended hyperparameters (denoted "*")

在**1024X1024**图像上：

Dataset	Size	Method	# Real Images	GPU Used	Training Time	FID (↓)
CelebA-HQ (1024x1024)	30K	ProGANs [14]	12M	-	-	**7.79**
		MSG-ProGAN	3.2M	8 V100-16GB	1.5 days	8.02
		StyleGAN [14]	25M	-	-	**5.17**
		MSG-StyleGAN	11M	8 V100-16GB	4 days	6.37
FFHQ (1024x1024)	70K	ProGANs*	12M	4 V100-32GB	5.5 days	9.49
		ProGANs [13]	12M	-	-	**8.04**
		MSG-ProGAN	6M	4 V100-32GB	6 days	8.36
		StyleGAN*	25M	4 V100-32GB	6 days	4.47
		StyleGAN [14]	25M	-	-	**4.40**
		MSG-StyleGAN	9.6M	4 V100-32GB	6 days	5.8

Table 2: Experiments on high resolution (1024x1024) datasets. We use author provided scores where possible, and otherwise train models with the official code and recommended hyperparameters (denoted "*").

Figure 5: During training, all the layers in the MSG-GAN synchronize across the generated resolutions fairly early in the training and subsequently improve the quality of the generated images at all scales simultaneously. Throughout the training the generator makes only minimal incremental improvements to the images generated from fixed latent points.

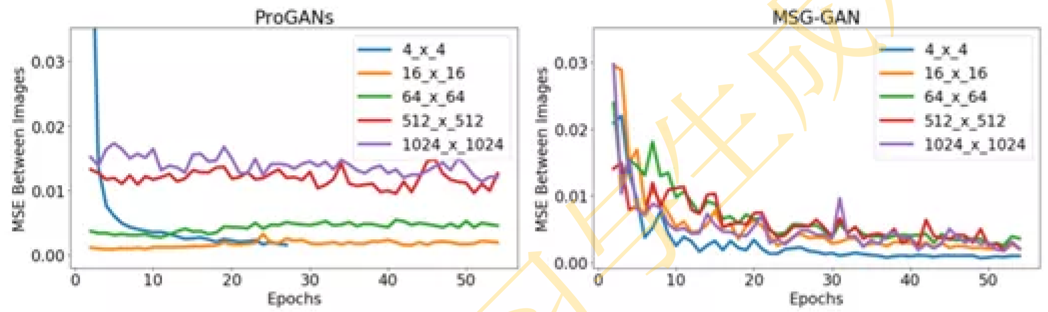


Figure 6: Image stability during training. These plots show the MSE between images generated from the same latent code at the beginning of sequential epochs (averaged over 36 latent samples) on the CelebA-HQ dataset. MSG-ProGAN converges stably over time while ProGANs [13] continues to vary significantly across epochs.

对学习率的鲁棒性实验：

Method	# Real Images	Learning rate	IS (↑)
Real Images	-	-	11.34
MSG-ProGAN	12M	0.003	**8.63**
MSG-ProGAN	12M	0.001	8.24
MSG-ProGAN	12M	0.005	8.33
MSG-ProGAN	12M	0.01	7.92

Table 3: Robustness to learning rate on CIFAR-10. We see that our approach converges to similar IS scores over a range of learning rates.

不同尺寸层面连接提供梯度的消融探究、损失函数探究：

Level of Multi-scale connections	FID (↓)
No connections (DC-GAN)	14.20
Coarse Only	10.84
Middle Only	9.17
Fine Only	9.74
All (MSG-ProGAN)	**8.36**
ProGAN*	9.49

Table 4: Ablation experiments for varying degrees of multiscale gradient connections on the high resolution (1024x1024) FFHQ dataset. Coarse contains connections at (4x4) and (8x8), middle at (16x16) and (32x32); and fine at (64x64) till (1024x1024).

Method	Combine function	FID (↓)
MSG-ProGAN	ϕ_{lin_cat}	11.88
	ϕ_{cat_lin}	9.63
	ϕ_{simple}	**8.36**
MSG-StyleGAN	ϕ_{simple}	6.46
	ϕ_{lin_cat}	6.12
	ϕ_{cat_lin}	**5.80**

Table 5: Experiments with different combine functions on the high resolution (1024x1024) FFHQ dataset.

一些局限：

使用渐进逐步训练（**ProGAN**那种），在较低的分辨率下训练迭代速度更快，而**MSG-GAN**的每次迭代都需要相同的时间。

在**FFHQ**和**CelebAHQ**的人脸数据集上，没有超过**StyleGAN**的生成质量。造成这种情况的原因很多，其中包括超参数选择不当，或者**StyleGANs**架构更适合这些数据集。

此外，由于在**MSGStyleGAN**中进行了多尺度修改，因此方法无法利用**mixing regularization trick [14]**。

公众号近期荐读：

GAN整整6年了！是时候要来捋捋了！

新手指南综述 | GAN模型太多，不知道选哪儿个?

数百篇GAN论文已下载好！搭配一份生成对抗网络最新综述！

有点夸张、有点扭曲！速览这些GAN如何夸张漫画化人脸！

天降斯雨，于我却无！GAN用于去雨如何？

脸部转正！GAN能否让侧颜杀手、小猪佩奇真容无处遁形？

容颜渐失！GAN来预测？

强数据所难！SSL（半监督学习）结合GAN如何？

弱水三千，只取你标！AL（主动学习）结合GAN如何？

异常检测，GAN如何gan ？

虚拟换衣！速览这几篇最新论文咋做的！

脸部妆容迁移！速览几篇用GAN来做的论文

【1】GAN在医学图像上的生成，今如何？

01-GAN公式简明原理之铁甲小宝篇

GAN&CV交流群，无论小白还是大佬，诚挚邀您加入！

一起讨论交流！长按备注【进群】加入：

更多分享、长按关注本公众号：