GAN整整6年了！是时候要来捋捋了！

原创 bryant 机器学习与生成对抗网络 2019-11-29

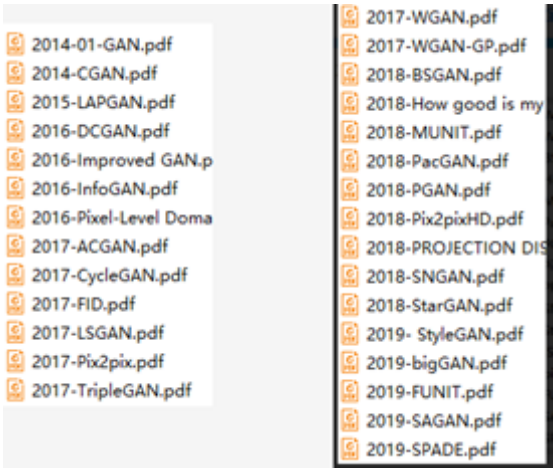
首先，这文章标题起成这样，如果觉得不舒服，哈哈，真的莫怪我，估计是受到现在的媒体文章“荼毒至深”的结果呢！

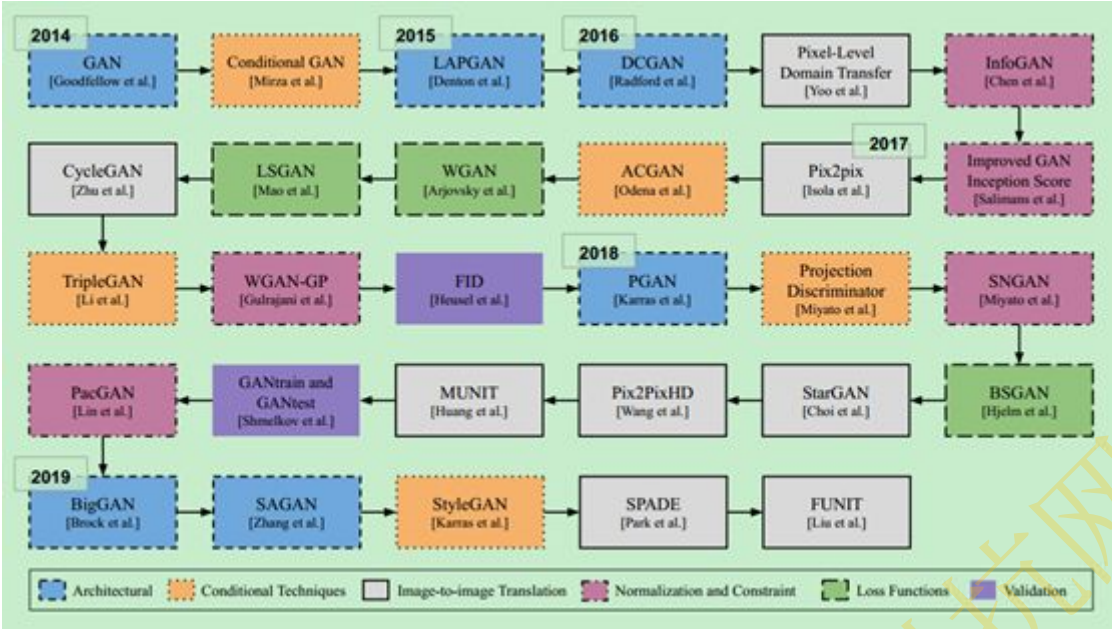
2014-2019，GAN已经诞生6年。GAN是一个生成模型，被广泛用于图像等信息生成。对于给定一批数据，GAN可以将高斯分布采样的噪声转化成和给定数据相似的新数据，从而达到源源不断地制作逼真的“合成数据”。具体的数学原理可参考：01-GAN公式简明原理之铁甲小宝篇。后来，为了更好地控制生成内容/效果，也加入了各种各样的条件信息，比如一幅图像的编码，那么可以达到图像风格转换的目的。

今天推荐一个关于生成对抗网络GAN（主要是在图像的生成和转换上）的一个简短综述：**The Six Fronts of the Generative Adversarial Networks**（https://arxiv.org/pdf/1910.13076.pdf）。主要从 **网络结构、条件信息、归一化和约束、损失函数、图像转换、评估准则** 六个方面做了简洁的梳理和回顾。

如果懒得下载读原文，那么可看看我速览该文之后、依其内容结构做的一个概要翻译，穿插有一些自己的“闲言碎语”，若有不当处，望纠正。

论文里有个很棒的 *Timeline of the GANs* 的图，提及的29个GAN基本算是蛮经典的了，已经它们打包下载好，可关注微信公众号“**学点诗歌和AI知识**”回复“**29GAN**”，来获得所有论文网盘下载地址哦。





好，开始转入正文……

GAN的问题和挑战

GAN令人头疼之处主要在于训练不稳定、收敛困难、难以精准控制生成内容，因此各种各样的**网络结构、损失函数、加入条件监督信息技巧和各种约束手段**被提出。另外，在训练时，各种超参数的选择对于最终的结果也影响巨大，比如bigGAN对batchsize、网络深度/宽度等超参数的实验就说明了这一点。论文《Are gans created equal? a large-scale study》甚至认为只要给够充分的时间去进行超参数调节和随机初始化，很多GAN变体提出的损失函数和技巧最终能达到的效果是难分上下的。

尽管GAN在人脸等相关数据集上表现惊人，但是在场景复杂、纹理形状多变的数据集的表现上往往大打折扣。如果图像的类别数量太多，或者不平衡，又极有可能导致别的问题，例如模式坍塌（生成图像单一）。事实上，我自己觉得GAN目前生成图像里，“大分辨率”和“精准控制图像内容”往往是鱼和熊掌难以兼得。

一. 网络结构

GAN最开始只能生成有较小的分辨率的图像，例如28X28大小的MNIST手写数字图像。因为对于判别器和生成器两者的训练火候本来就难以拿捏，往往要求两者旗鼓相当之下、丝毫不能有碾压对方之势头才能达到对抗的过程，有了较好的对抗才有最后理想的平衡。而越大的图像意味着更复杂的特征空间、更难的任务，让两者同时恰当地对抗真是太南了。

于是乎，经典的DCGAN率先出击，为尝试缓解这种“太南的局面”，提出去掉池化层、全连接层，使用BN等。但是改进效果实在有限，也只能生成在大一点点的图像。而LAPGAN引入一种金字塔式的pipeline，似乎也只能生成96X96的分辨率。随后的一些工作，也大都限于128X128（ACGAN），256X256的像素。2018年，Progressive GAN（PGAN）的出现，生成令人震撼地1024X1024人脸图像。事实上，到现在很多论文也是在128X128和256X256的图像大小上进行搞事情，因为再大，就南了，或者就崩了。

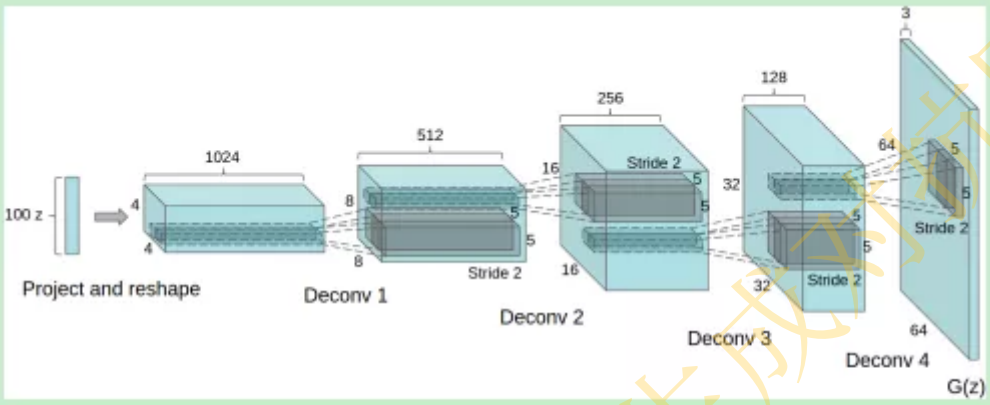


Fig. 3: DCGAN's generator architecture, reproduced from Radford *et al.* [24]. The replacement of fully-connected and pooling layers by (strided and fractional-strided) convolutional layers enabled GANs architecture to be deeper and more complex.

2019年的StyleGAN号称GAN 2.0，因为它不再是简单地只在第一层接收噪声或隐变量Z，而是在生成器各层都注入，并且在Z送入“真正”的生成器之前，先经过多层全连接尝试将其解耦，如下图。

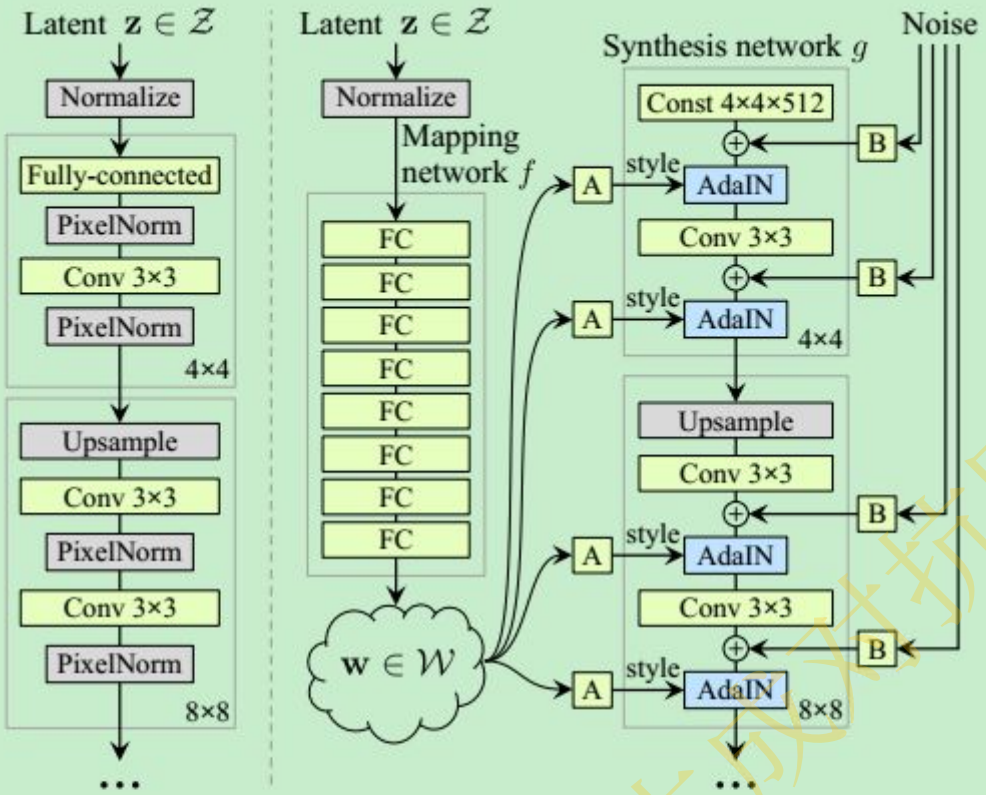


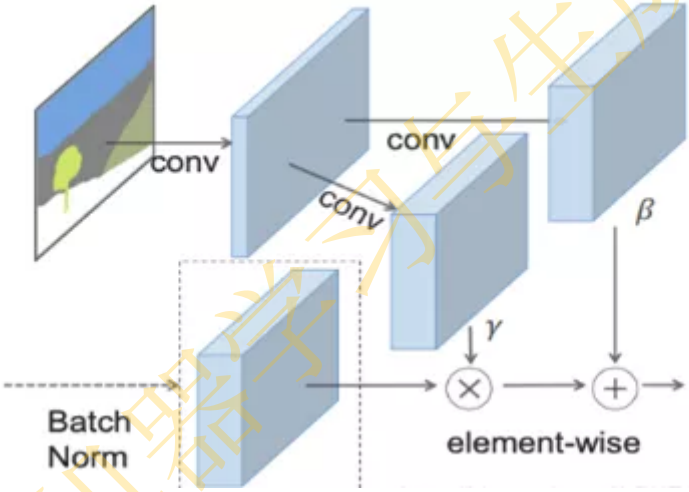
Fig. 5: Comparison between generators on PGAN [26] (on the left) and StyleGAN (on the right). First, a sequence of fully-connected layers is used to extract style information and disentangle the noise and class information. Next, style information feeds AdaIN layers, providing its parameters for shift and scale. Also, different sources of noise feed each layer to simplify the task of controlling stochastic aspects of generation (such as placement of hair and skin pores). Reproduced from Karras *et al.* [1].

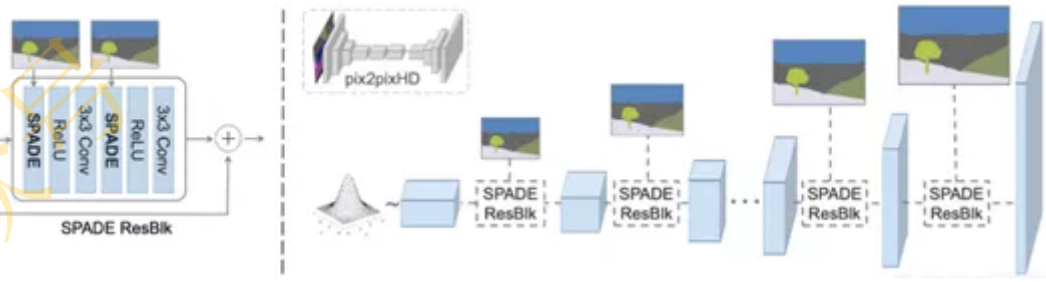
解耦的目标是使Z代表的隐空间由线性子空间组成，即某个子空间（每个维度）控制某种特征。但是隐空间Z往往可能是有纠缠现象的，通过可学习的全连接网络对Z解耦，使变化的因素变得更加线性。可能说的有点拗口，打一个不那么恰当的比方说，在生成猪八戒图像中，以往的GAN未很好解耦前，Z中某一维度可能同时控制生成二师兄的大眼睛和鼻子和肚子，而解耦后，通过分别修改每一层的输入，在不影响其他层的情况下，来控制该层级所表示的最后视觉特征。

二．条件信息

众所周知，原始GAN仅仅基于噪声作为生成器的输入，在期望生成指定的图像是无法实现的（不可控），为此，研究者尝试将条件信息或者标签信息引入，以期生成更理想的图像。

2014年，CGAN通过将类别标签和输入噪声的拼接作为生成器的输入，来试图达到控制生成图像的类别。在ACGAN中，生成器的输入也附加标签/条件信息，判别器的输出则有两部分，一部分判断真假，一部分输出类别。2018年的ICLR论文《cGANs with Projection Discriminator》加入条件信息的方式不再是拼接，而是选择将图像进入判别器后得到的feature与条件信息进行内积点乘。2019年的SPADE受到AdaIN的启发，在将语义分割式的标签图像合成真实图像时，生成器生成过程中归一化的参数是由语义标签图卷积而来的张量形式，保持着标签图在长宽维度上的空间信息再去scale和shift，从而达到生成更精细的图像的目的。





三．归一化和参数约束

在较早的DCGAN里，生成器和判别器都使用了batch normalization去来解决著名的internal covariate shift问题。而在渐进式训练的PGAN中，作者认为他们的问题不再是internal covariate shift, 而采用了一种没有参数的归一化方式 Pixelwise

Normalization；另外还提到了一种叫Equalized Learning Rate 的参数约束方法，这样可以更好地稳定训练过程，具体可以参考原论文。

著名的Spectral Normalization 谱归一化也是用来约束参数的，例如对判别器D的参数进行谱归一化后，可以使得D满足Lipschitz利普希茨连续性，以此达到网络对输入扰动具有较好的非敏感性，使训练过程更稳定、易收敛。同样地，在著名的WGAN中，提出Wasserstein distance距离作为衡量，并将其转换为求解最优的利普希茨连续函数的问题，为此进行参数约束：将过大的参数直接裁剪到一个阈值以下。更具体地原理解释，可参考相关解读：令人拍案叫绝的Wasserstein GAN

四．损失函数

原始的GAN损失在衡量生成的数据分布和真实的数据分布上采用的是JS散度，而JS散度容易导致训练不稳定，梯度消失等问题。故涌出了诸如LSGAN、WGAN等一批试图使用更好的距离度量方式的GAN变体。此外，习惯了仅仅对输出图像的层面进行损失计算和反向传播之外，对中间的特征层feature map去计算损失也开始得到大家的重视，例如Perceptual Loss 越来越多地出现在GAN论文里。哈哈，不过呢，前面也提到过，**似乎也有**论文认为，尽管这一堆GAN的损失变体都声称它们取得了进展，但是在够计算力、训练程度下，大家实际上都彼此彼此啦。

个人认为，不同的损失也许在不同的场景下各有千秋、或者在相同场景下达到相同效果上有难易之分吧。

五．图像转换

这一块真的是GAN的主战场，应用也是极多。从一开始经典的pix2pix、cycleGAN到UNIT、MUNIT、starGAN到SPADE太多了额，这两三年image-to-image相关的论文盲猜一千篇以上了，网上随便一搜各种相关解读、应用也是以上百数千计，比如用来做风格迁移，人脸属性编辑/换脸，妆容迁移（参考我整理的论文：脸部妆容迁移！速览几篇用GAN来做的论文（本公众号查看）），医学图像合成/增强，虚拟穿衣/去衣/换衣（参考：虚拟换衣！这几篇最新论文不来GAN GAN（本公众号查看）），去雨、去噪去雾、去马赛克、去阴影，超分辨率等等，是在太多太多。

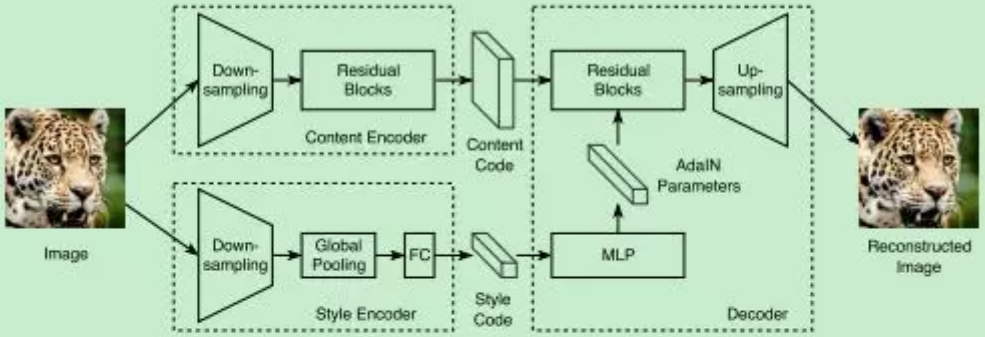


Fig. 10: The generator's architecture from MUNIT. The authors employ different encoders for content and style. The style information feeds a Multi-Layer Perceptron (MLP) that provide the parameters for the AdaIN normalization. This process incorporates the style information to the content, creating a new image. Reproduced from Huang *et al.* [37].

六. 评估方法

评估生成的图像也是极具挑战性的。定量指标例如IS和FID经常被用来评估生成图像，它们都是使用ImageNet预训练好的分类网络前向预测来打分。显而易见地，对于一些不出现在ImageNet数据中的图像类别，太不靠谱。定性指标例如人的主观评分，由于各有己见，甚至同一个人在不同时间、情绪下也会有不同，如何设计有代表或者更广泛地进行评估，也是不容易。目前，也有一些别的指标在不同的场景下或者任务中被提出，但没有一种指标是完美的，只有尽可能地多方面、多合理性的指标去进行评估。

更多相关知识分享请关注微信公众号：**学点诗歌和AI知识**