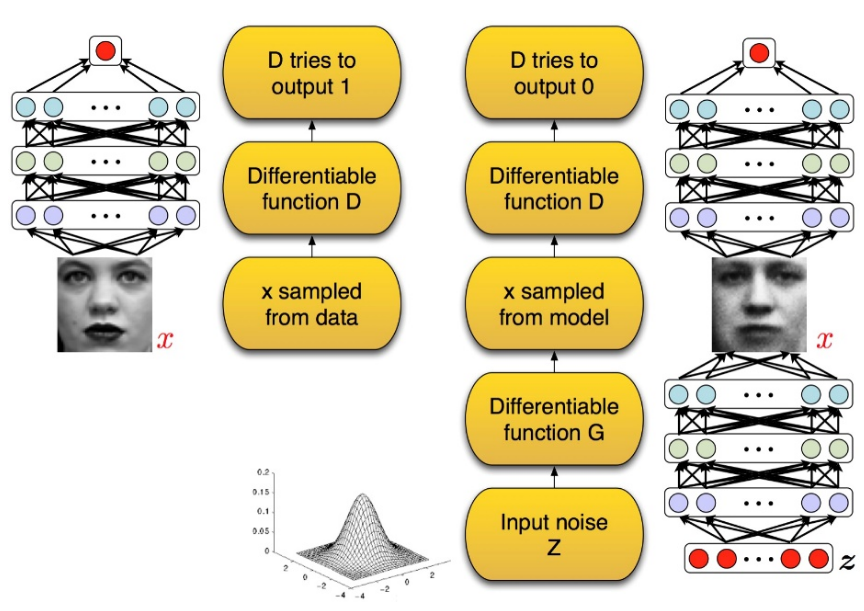
### 原始GAN

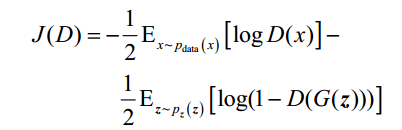
GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems. 2014: 2672-2680.



原始GAN提出了GAN网络的基本框架：

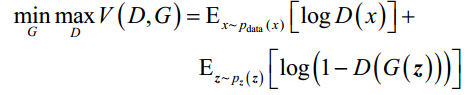
* 判别器的目的是正确区分真实数据和生成数据，从而最大化判别准确率
* 生成器则是尽可能逼近真实数据的潜在分布
* 为了在博弈中胜出，二者需不断提高各自的判别能力和生成能力，优化的目标就是寻找二者间的纳什均衡

由于判别器是一个二分类模型， 因此，可用交叉熵表示其目标函数



因此生成器G的目标函数J(G)=-J(D)

GAN的优化问题可描述为如下极大极小博弈问题:



存在的问题：

在原始GAN的（近似）最优判别器下

* 第一种生成器loss面临梯度消失问题
* 第二种生成器loss面临优化目标荒谬、梯度不稳定、对多样性与准确性惩罚不平衡导致mode collapse这几个问题

### WGAN

Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXivpreprint arXiv: 1701.07875, 2017

理论贡献：

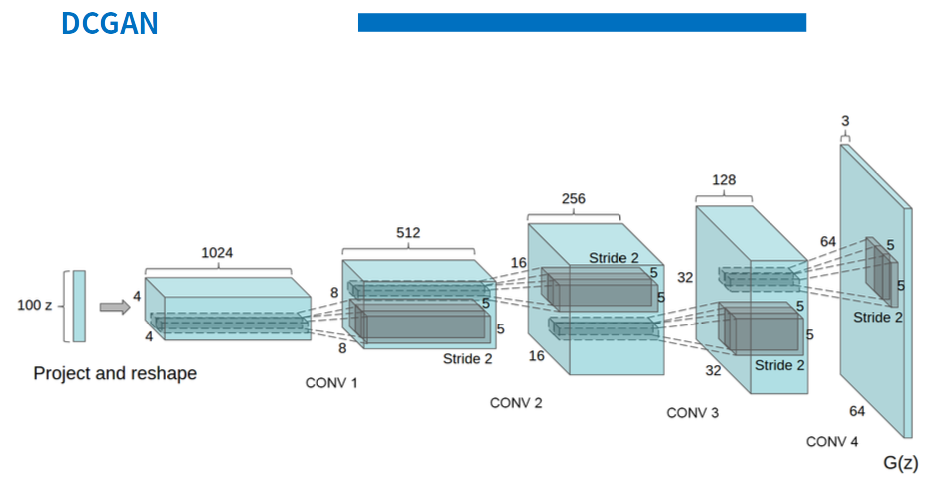
* 彻底解决GAN训练不稳定的问题，不再需要小心平衡生成器和判别器的训练程度
* 基本解决了collapse mode的问题，确保了生成样本的多样性
* 训练过程中终于有一个像交叉熵、准确率这样的数值来指示训练的进程，这个数值越小代表GAN训练得越好，代表生成器产生的图像质量越高
* 以上一切好处不需要精心设计的网络架构，最简单的多层全连接网络就可以做到

改进后相比原始GAN的算法实现流程却只改了四点

* 判别器最后一层去掉sigmoid
* 生成器和判别器的loss不取log
* 每次更新判别器的参数之后把它们的绝对值截断到不超过一个固定常数c
* 不要用基于动量的优化算法（包括momentum和Adam），推荐RMSProp，SGD也行

### DCGAN

ZHU W, MIAO J, QING L, et al. Unsupervised representation learningwith deep convolutional generative adversarial networks. computerscience[J]. arXiv: arXiv1511.06434, 2015.



该模型的贡献：

（1）批量正则（batch\_norm，ＢＮ）：对 G的输出层和 D的输入层不用 BN，而其他层都用 BN，缓解了模型崩溃问题，有效避免了模型的振荡和不稳定问题；

（2）输出层（output layer）采用合适的激活函数，从之前的 sgd+momentum 到 adam；

（3）strided convolutional networks 替代传统 CNN 中的 pooling 层，是网络自动学习合适的采样核函数；

（4）提出了潜在空间（ latent space）概念，实现了语义的算术加减；

（5）语义遮盖（semantic mask），在隐空间上，设置变量控制显示或者消失物体；（6）辨别过滤器（discriminator filter）通过分析 filter，GANs 确实学习到了LSUN数据库中的床、窗口等特征。

模型结构上需要做如下几点变化：

* 将pooling层convolutions替代，其中，在discriminator上用strided convolutions替代，在generator上用fractional-strided convolutions替代。
* 在generator和discriminator上都使用batchnorm。
  + 解决初始化差的问题
  + 帮助梯度传播到每一层
  + 防止generator把所有的样本都收敛到同一个点。
  + 直接将BN应用到所有层会导致样本震荡和模型不稳定，通过generator输出层和discriminator输入层不采用BN可以防止这种现象。
* 移除全连接层
  + global pooling增加了模型的稳定性，但伤害了收敛速度。
* 在generator的除了输出层外的所有层使用ReLU，输出层采用tanh。
* 在discriminator的所有层上使用LeakyReLU。

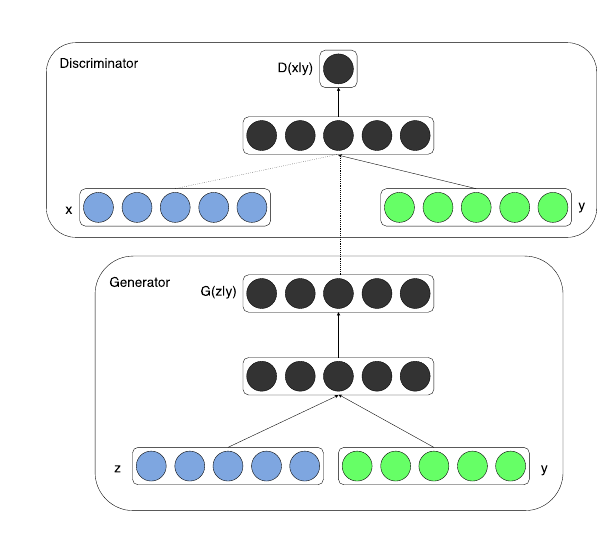
训练细节

* 预处理环节，将图像scale到tanh的[-1, 1]。
* mini-batch训练，batch size是128.
* 所有的参数初始化由(0, 0.02)的正态分布中随即得到
* LeakyReLU的斜率是0.2.
* 虽然之前的GAN使用momentum来加速训练，DCGAN使用调好超参的Adam optimizer。
* learning rate=0.0002
* 将momentum参数beta从0.9降为0.5来防止震荡和不稳定。

### CGAN

MIRZA M, OSINDERO S. Conditional generative adversarial nets[J].

Computer Science, 2014: 2672-2680



GAN 的最大优点体现在其对抗训练方式通过对 p(x)直接采样来逼近真实样本，利用反向传播即可获得梯度而不需复杂的马尔可夫链和推断过程，从而大大简化了计算。然而，原始GAN使用 z 作为先验，但生成式模型如何利用这个先验却是无法控制的。

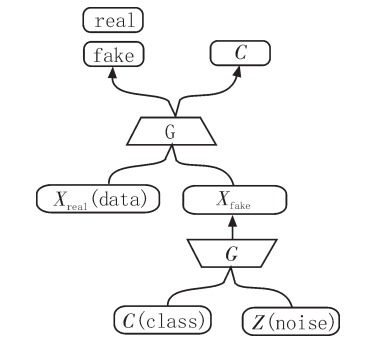
原始GAN的学习模式过于自由而导致GAN的训练过程和结果都不可控。为了提高GAN的稳定性，Mirza等提出条件生成式对抗网络，将条件变量y作为模型的附加信息以约束生成过程，这种条件变量可以是类别标签甚至还可以是不同模态的数据。

这项工作提出了一种带条件约束的GAN，在生成模型（D）和判别模型（G）的建模中均引入条件变量y（conditional variable y），使用额外信息y对模型增加条件，可以指导数据生成过程。

这些条件变量y可以基于多种信息，例如类别标签，用于图像修复的部分数据，来自不同模态（modality）的数据。如果条件变量y是类别标签，可以看做CGAN 是把纯无监督的 GAN 变成有监督的模型的一种改进。

### InfoGAN

CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]//Advances in Neural Information Processing Systems.2016: 2172-2180





* G：输入定义的潜在变量c和噪声
* 目标函数要使c和生成样本尽可能相关，即互信息最大

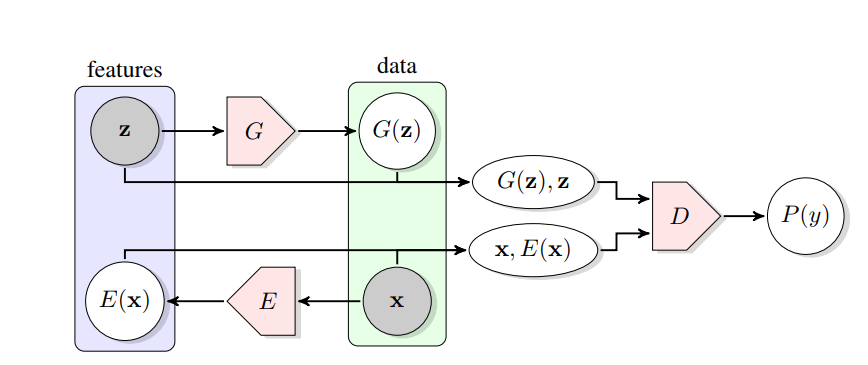
这篇paper所要达到的目标就是通过非监督学习得到可分解的特征表示。使用GAN加上最大化生成的图片和输入编码之间的互信息。最大的好处就是可以不需要监督学习，而且不需要大量额外的计算花销就能得到可解释的特征。

原始 GAN 模型输出为 G（z），没有利用生成网络输入自由度 noise z。InfoGAN做出了改进，将输出改为 G（z，c），将原始模型中的z拆解成两部分：可解释、有隐含意义的一组隐变量 c（c\_1、c\_2、…、c\_L）和不能再分解出有价值意义的 noise z。

InfoGAN 通过拆解先验的方式，让 GANs 能学习到更加可控的数据表示。首先，infoGAN中的 c对于 GANs的训练确实有帮助，使得生成模型生成更符合真实数据的结果，也能控制 c的维度，使得 info⁃GAN 能控制生成的图片在某一个特定语义维度的变化。利用这种更加细致的隐变量建模机制，Info-GAN将 GANs的发展又推进了一步

### BiGAN

DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning[J]. arXiv: arXiv1605.09782, 2016.



GAN 通过将简单的隐变量分布映射至任意复杂的数据分布来生成令人信服的自然图像，这表明 GAN 的生成器能对隐空间中的数据分布进行语义线性化，学习到隐空间中数据的良好特征表达。

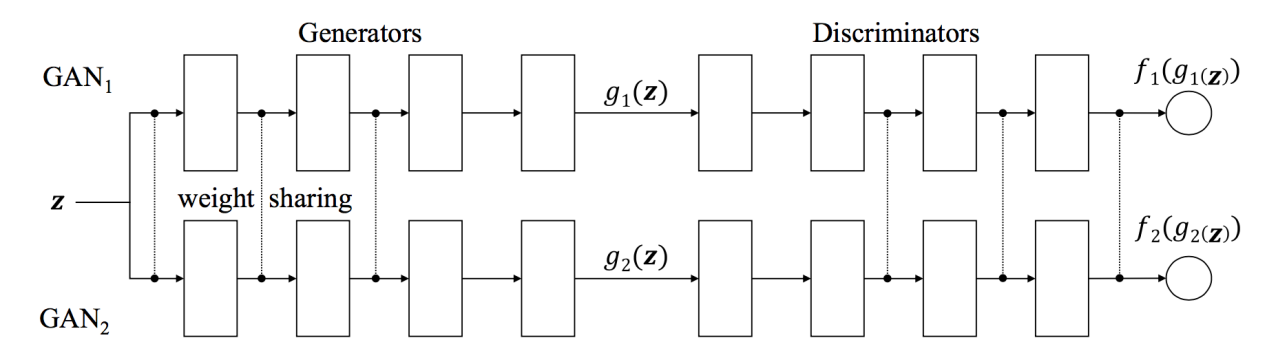
然而， GAN 缺少一种有效的推断机制，未能学习从数据空间映射至隐空间的逆映射。为解决这个问题， Donahue 等和 Dumoulin 等将单向的GAN 变为双向的 GAN，从而既能进行有效推断又保证了生成样本质量。

双向生成式对抗网络（BiGAN, Bidirectional GAN），除了学习标准 GAN 中的生成器 G 和判别器 D 外，还学习了一个将数据映射至隐式表达的编码器 E，如图所示。

通过编码器 E 学习到的特征表达有助于训练判别器 D，通过同时训练编码器和解码器以迷惑生成器，使判别器难以区分究竟是真实样本及其编码还是生成样本及其隐变量，换言之，双向生成对抗网络的生成器判别的是联合样本（x, z）而不是样本 x

### CoGAN

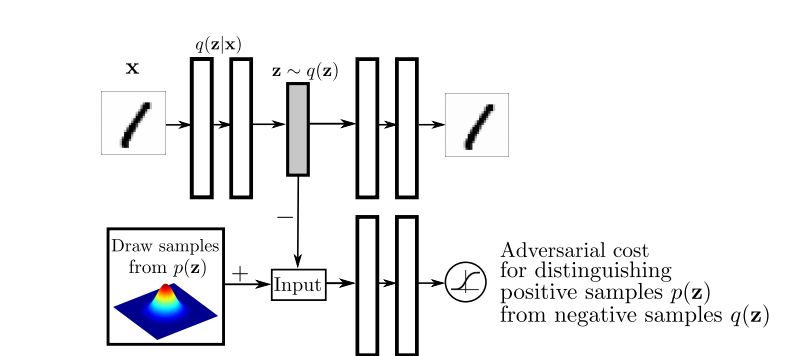
LIU M Y, TUZEL O. Coupled generative adversarial networks[C]//Advances in Neural Information Processing Systems, 2016:469-477

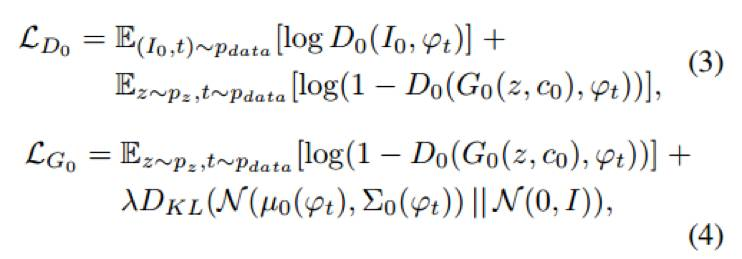


* 文章的思想是,利用网络层的权重共享约束,训练GAN网络.模型包括两个生成网络,两个判别网络
* 训练数据为不成对的两个域Domain1,Domain2的图片,我们希望的是训练的两个生成网络g1,g2能够在输入向量z相同的情况下,生成的图片高频信息相同,低频信息不同.因此在觉得高频特征的生成网络的前几层,将两个生成网络的权重共享,并且,将两个判别网络f1,f2的最后几层网络权重共享,如上图所示.
* 作者利用两个通过权值共享耦合的GAN网络，生成跨域样本。举个例子，输入一堆正常人脸和一堆戴眼镜的人脸，会生成长得一样的 with or without glasses 的人脸

### AAE

MAKHZANI A, SHLENS J, JAITLY N, et al. Adversarialautoencoders[J]. arXiv: arXiv1511.05644, 2015.





对抗自编码器（AAE,adversarial autoencoder)将利用自编码器得到的重构误差和对抗训练得到的隐变量与目标分布之间的误差进行组合，从而既能通过自编码器进行推断又能得到结构化的隐变量。

Larsen 等将 VAE 和 GAN 并入同一个无监督生成模型中，当将编码器和解码器看作一个生成模型整体时，这个生成模型和判别器构成了扩展的生成式对抗模型；若将解码器和判别器看作一个整体时，这个整体相当于解码器，并与编码器共同构成了扩展的自编码器，因此，该模型结合了 GAN和 VAE 的优点。

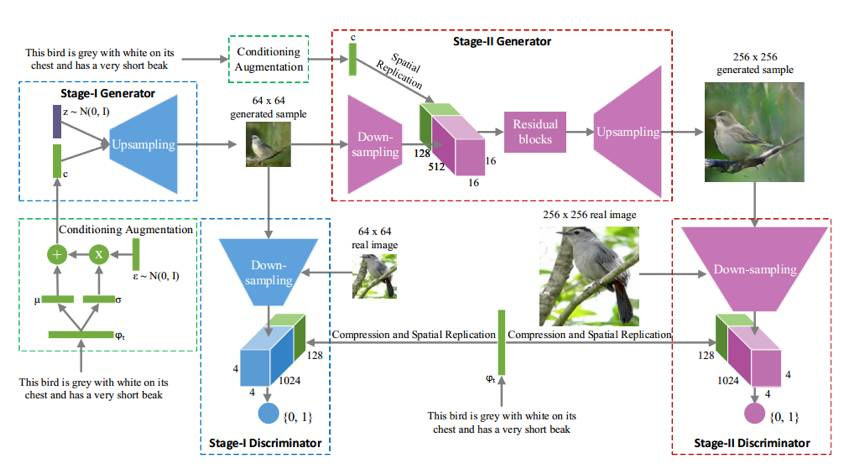
对抗自编码器分成两大部分网络结构：自编码部分（图上半部分）、GAN 判别网络（图下半部分），其中变量含义如下：x 是样本输入，z 是编码向量，p(z) 是隐含变量的先验分布， q(z|x) 是编码分布函数，p(x|z) 是解码分布函数，pd(x) 表示真实的数据分布，p(x) 表示模型数椐分布。

从网络可以看出分成三个部分：encoder, decoder 和 D。三者的 loss 如下：

* decoder 就是一般的重构误差：mse
* D 的 loss 就是 GAN 的 loss，就是是判别真假，只是它是在隐空间上判别；
* encoder 的 loss=重构 loss + 判别器的 loss

### StackGAN

HUANG X, LI Y, POURSAEED O, et al. Stacked generative adversarial networks[J]. arXiv: arXiv1612.04357, 2016



从文本描述生成高质量的图片是计算机视觉领域一个非常具有挑战性的工作。已有的文本-图像生成方法可以生成比较粗糙的模型，但是却不能生成必要的细节和生动的物体。

在本文中，作者提出了堆积生成对抗网络（StackGAN）来生成高质量的图片。

第一阶段的对抗生成网络利用文本描述粗略勾画物体主要的形状和颜色，生成低分辨率的图片。

第二阶段的对抗生成网络将第一阶段的结果和文本描述作为输入，生成细节丰富的高分辨率图片。

堆积对抗生成网络第一次在只给定文本的条件下生成真实的256\*256的图片。

第一步的对抗生成网络生成一个低分辨率的64\*64的图片和真实数据进行对抗训练得到粗粒度的生成模型。

第二阶段的StackGAN将第一阶段的生成结果和文本描述作为输入，用第二个对抗生成网络生成高分辨率的256\*256的图片。

两个阶段中的对抗生成网络参数可以参考下式：

### EBGAN

Zhao J, Mathieu M, LeCun Y. Energy- based Generative Adversarial Network[DB/OL].(2016- 03- 06）. http：//arxiv.org/abs/1609.03126.

EBGAN从能量模型的角度对 GANs进行了扩展，给出了解读 GANs 的另外一种角度。EBGAN 将判别器看做是一个能量函数，这个能量函数在真实数据域附近的区域能量值比较小，而在其他非真实数据域区域都拥有较高能量值。

因此，EBGAN 中给予GANs 一种能量模型的解释，即生成器是以产生能量最小的样本为目的，而判别器则以对这些产生的样本赋予较高的能量为目的。从能量模型的角度来看待判别器的价值在于可以用更多更宽泛的结构和损失函数来训练 GANs结构，比如文中就用自编码器（AE）作为判别器实现整体的GANs框架。

### ImprovedGAN

Salimans T, Goodfellow I, Zaremba W, et al. Improved Techniques for Training GANs[DB/OL].（2016-04-06）[2016-12-26].http：//arxiv.org/abs /1504.01391.

对抗网络的目的是在高维非凸的参数空间中，找到一个纳什均衡，但是 GANs 网络使用梯度下降的方法只会找到低的损失，不能找到真正的纳什均衡。Improved GAN给出了5条有助于GANs稳定训练、提高网络的收敛的经验：

（1）特征匹配（feature matching）：让生成器产生的样本与真实样本在判别器中间层的响应一致，即使判别器从真实数据和生成数据中提取的特征一致，而不是在判别器网络的最后一层才做判断，有助于提高模型的稳定性；

（2）最小批量判断（minibatch discrimination）：在判别器中，不再每次对每一个生成数据与真实数据的差异性进行比较，而是一次比较一批生成数据与

真实数据的差异性。这种做法提高了模型的鲁棒性；

（3）历史平均（historical averaging）：在生成器和判别器的目标函数中各加一个约束项，使得网络训练过程中，梯度不容易进入稳定的轨道，继续向均衡点更新；

（4）单边标签平滑（one-sided label smoothing）：标签平滑将常规的 0、1 取值的二值标签替换为如0.1，0.9之类的平滑标签，增加网络的抗干扰能力。单边平滑对于取值0的标签保持不变；

（5）虚拟批量正则（virtual batch normalization，VBN）：BN 能够提高网络的收敛，但是导致 layer 的输出和本次 Batch 内的其他输入相关，通过引入一个参考集合，每次将当前数据 x加入参考集合构建一个新的虚拟的批量数据，然后在此基础上进行归一化，以此缓解原始BN操作所引起的波动问题。

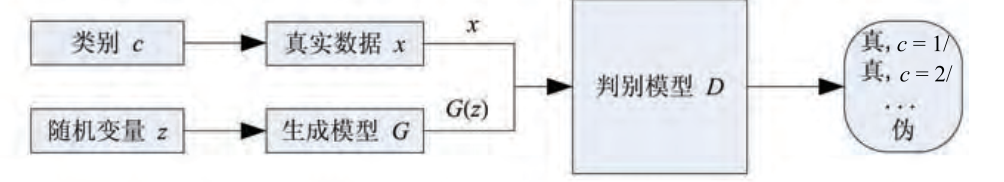
### Semi-GAN

Odena A. Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv: 1606.01583, 2016.

GAN 的训练只需要数据源的标注信息 (真或伪), 并根据判别器输出来优化. Odena提出了 Semi-GAN, 将真实数据的标注信息加入判别器 D

训练.

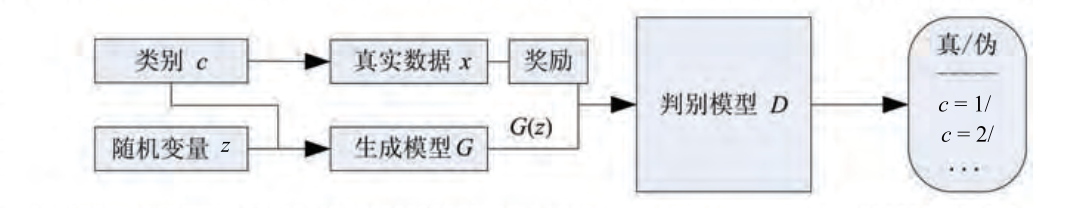
* G：不输入类别，只输入噪声来生成样本
* D：输入生成样本或者训练样本，最后的输出是输入图像所属类别或者输入图像为假，一共N+1维（N是总类别数）



### AC-GAN

Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifler GANs. arXiv preprint arXiv:1610.09585, 2016.

* G：与cGAN相同，同样是输入类标签和噪声
* D：不再输入类标签，另外在输入样本是否为真的同时，利用另一个分类器来判断输入样本的所属类别
* D训练的目标是最大LS+LC，G训练的目标是最大LC-LS



lt@202.17.10.69