第 5 章　生成对抗网络

生成对抗网络(Generative Adversarial Network，GAN)在诸多领域都取得了较好的应用效果，本章将以生成模型概述为切入点，介绍生成模型的基本概念和生成模型的意义及应用，在此基础上详细叙述 GAN，并分析 GAN 的延伸模型——SGAN 模型、CGAN 模型、StackGAN 模型、InfoGAN 模型和 Auxiliary Classifier GAN 模型的结构。

5.1　生成模型概述

深度神经网络的热门话题是分类问题，即给定一幅图像，神经网络可以告知你它是什么内容，或者属于什么类别。近年来，生成模型成为深度神经网络新的热门话题，它想做的事情恰恰相反，即给定一个类别，神经网络可以无穷无尽地自动生成真实而多变的此类别图像，如图 5.1 所示，它可以包括各种角度，而且会在此过程中不断进步。

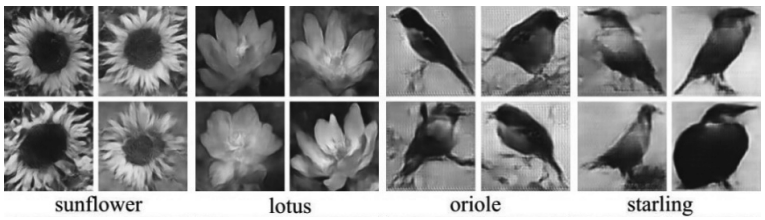

图 5.1　生成模型示例

5.1.1　生成模型的基本概念

在深度学习中，可以将其模型分为生成模型和判别模型两大类[1]。生成模型可以通过观察数据，学习样本与标签的联合概率密度分布 $P(x,y)$，然后生成对应的条件概率分布 $P(y|x)$，从而得到所预测的模型 $Y=f(x)$。判别模型强调直接从数据中学习决策函数[2]。生成模型的目标是给定训练数据，希望能获得与训练数据相同的新数据样本。判别模型的目标是找到训练数据的分布函数。在深度学习中，监督学习和非监督学习都包含其对应的生成模型，根据寻找分布函数的过程，可以把生成模型大致分为概率估计和样本生成。

概率估计是在不了解事件概率分布的情况下，通过假设随机分布，观察数据确定真正的概率密度分布函数，此类模型也可定义为浅层生成模型，典型的模型有朴素贝叶斯、混合高

斯模型和隐马尔可夫模型等。

样本生成是在拥有训练样本数据的情况下，通过神经网络训练后的模型生成与训练集类似的样本，此类模型也可以定义为深度生成模型，典型的模型有受限玻尔兹曼机、深度信念网络、深度玻尔兹曼机和广义除噪自编码器等。

5.1.2 生成模型的意义及应用

著名物理学家费曼说过一句话："只要是我不能创造的，我就还没有理解。"生成模型恰如其所描述的，其应用包括：

(1) 生成模型的训练和采样是对高维概率分布问题的表达和操作，高维概率分布问题在数学和工程领域有很广泛的应用[3]。

(2) 生成模型可以以多种方式应用到强化学习中。基于时间序列的生成模型可用来对未来可能的行为进行模拟；基于假设环境的生成模型可用于指导探索者或实验者，即使发生错误行为，也不会造成实际损失[4]。

(3) 生成模型可以使用有缺失的数据进行训练，并且可以对缺失的数据进行预测。

(4) 生成模型可以应用于多模态的输出问题，一个输入可能对应多个正确的输出，每一个输出都是可接受的[5]。图 5.2 是预测视频的下一帧图像的多模态数据建模示例。

神经网络的发展大致可以分为神经网络的兴起、神经网络的萧条与反思、神经网络的复兴与再发展、神经网络的流行度降低和深度学习的崛起共 5 个阶段。

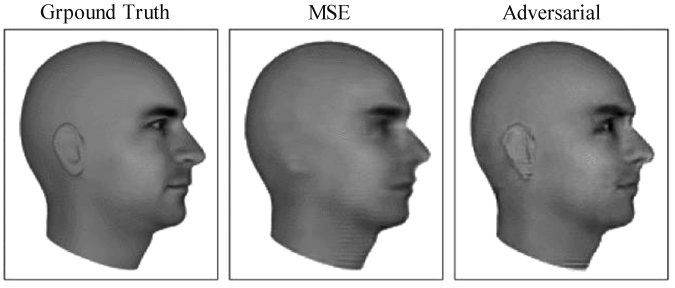

图 5.2 多模态数据建模示例

5.2 GAN 概述

2014 年，自蒙特利尔大学的博士生 Ian Goodfellow 提出 GAN 以来，对 GAN 的研究便成了深度学习领域的一大热门问题，GAN 甚至被称为过去十年深度学习领域最有趣的想法[6]。GAN 使用两个神经网络，将一个神经网络与另一个神经网络进行对抗，其潜力巨大，因为它们能学习模仿任何数据分布，因此，GAN 能在任何领域创造类似于图像、音乐、演讲、散文等形式的数据。图 5.3 是 GAN 的应用示例，在某种意义上，它们是机器人艺术家，它们的输出令人印象深刻，甚至人们能够被它们的输出深刻打动。

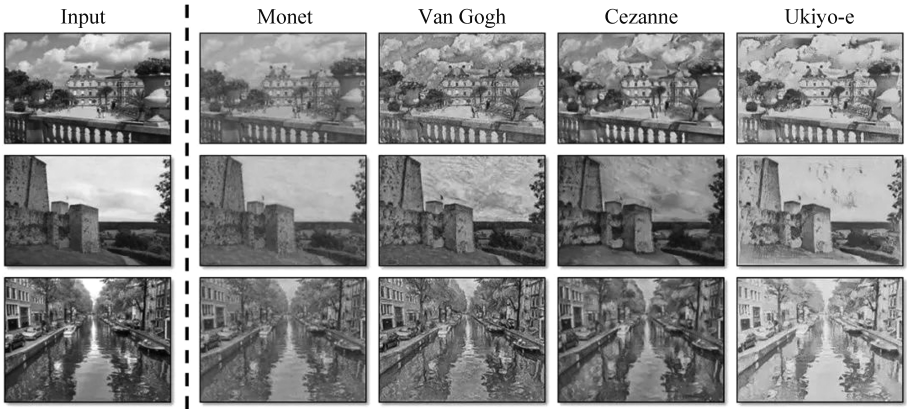

图 5.3　GAN 的应用示例

5.2.1　GAN 简介

GAN 对抗的两个网络分别为生成器 G 和鉴别器 D，生成器 G 是一个生成图像的网络，它接收一个随机的噪声 z，通过这个噪声生成图像，记为 $G(z)$，相当于编码器[4]。鉴别器 D 是一个判别网络，判别一幅图像是不是“真实的”。它的输入参数是 x，代表一幅图像，输出 $D(x)$代表 x 为真实图像的概率，$D(x)$为 1 时代表是真实的图像，$D(x)$为 0 时代表由生成器生成的图像，相当于解码器。

在训练过程中，生成器的目标就是尽量生成真实的图像去欺骗鉴别器。而鉴别器的目标是尽量把生成器生成的图像和真实的图像区分开。这样，生成器和鉴别器就构成了一个动态的“博弈过程”。在最理想的状态下，生成器可以生成足以“以假乱真”的图像，而对于鉴别器来说，它难以判定生成器生成的图像究竟是不是真实的，因此，理想情况下，当 $D(G(z))=0.5$ 时，得到一个生成式的模型 G，它可以用来生成所需的数据类型，鉴别器无法分辨。

5.2.2　GAN 的损失函数

GAN 的损失函数以及背后具有的数学原理是其另一个难点[7]。GAN 的目标是让生成器生成足以欺骗鉴别器的样本，从数学角度看，希望生成数据样本和真实数据样本拥有相同的概率分布，也可以说，生成数据样本和真实数据样本拥有相同的概率密度函数，即 $P_G(x)=P_{\text{data}}(x)$。

GAN 的损失函数源自二分类对数似然函数的交叉熵损失函数，有

$$L=-\frac{1}{N_i}[y_i\log p_i+(1-y_i)\log(1-p_i)] \tag{5-1}$$

式中，$y_i\log p_i$ 的作用是使来自真实数据的预测样本结果为 1，即当 $x\sim P_{\text{data}}(x)$时，$D(x)=1$，x 为真实样本数据，$D(x)$表示当数据为真实样本时鉴别器的输出结果。

$y_i\log p_i$ 可以表示为

$$E_{x\sim P_{\text{data}}(x)}\log(D(x)) \tag{5-2}$$

式中，$(1-y_i)\log(1-p_i)$的作用是使来自生成器样本的预测结果为0，即当$x\sim P(G(z))$时，$D(G(z))=0$，$G(z)$为来自生成器的生成样本数据。

$(1-y_i)\log(1-p_i)$可以表示为

$$E_{x\sim P_G(z)}\log(1-D(G(z))) \tag{5-3}$$

结合式(5-2)和式(5-3)，鉴别器的优化目标是最大化式(5-2)和式(5-3)之和，可以表示为

$$V(G,D)=E_{x\sim P_{\text{data}}(x)}\log(D(x))+E_{x\sim P_G(z)}\log(1-D(G(z))) \tag{5-4}$$

因此，在给定生成器的情况下，得到最优的鉴别器为

$$D_G^*=\arg\max_D V(G,D) \tag{5-5}$$

生成器的优化目标与鉴别器的优化目标相反，它需要鉴别器的结果尽可能地差，即当鉴别器达到最优时，需要对生成器进行优化，目的是最小化式(5-5)，于是可以得到最优的生成器表示为

$$G^*=\arg\min_G V(G,D_G^*) \tag{5-6}$$

综上所述，可以得到GAN中生成器和鉴别器的极大、极小博弈的损失函数，有

$$\min_G\max_D V(G,D)=E_{x\sim P_{\text{data}}(x)}\log(D(x))+E_{x\sim P_G(z)}\log(1-D(G(z))) \tag{5-7}$$

5.2.3 GAN的算法流程

以训练GAN生成人脸图片为例，GAN体系的核心结构如图5.4所示，其中包含以下4种类型。

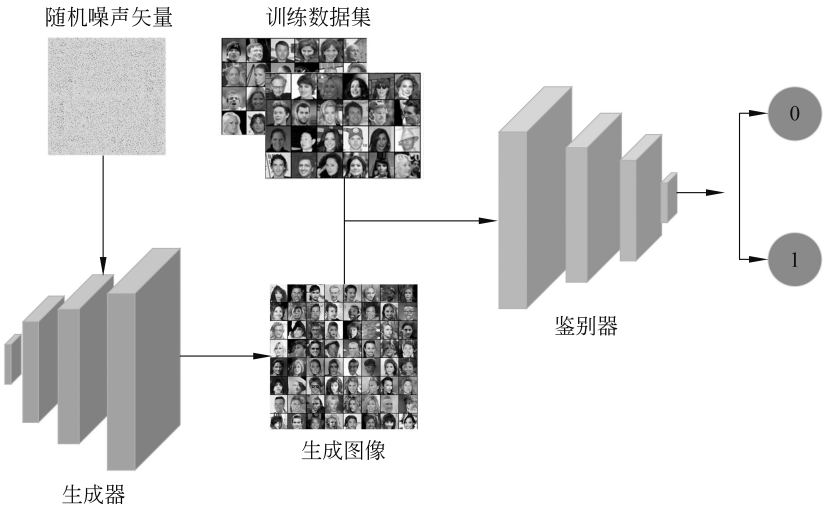

图5.4 GAN体系的核心结构

(1) 训练数据集：希望生成器以近乎完美的质量学习并模仿训练数据集，因此训练数据集由人脸图像组成，并将该数据集用作鉴别器的输入。

(2) 随机噪声矢量：作为生成器的原始输入z，这个输入是一个随机数向量，生成器使用它作为合成假样本的起点。

(3) 生成器：生成器将随机数向量 z 作为输入并输出假样本 x^*，它的目标是使生成的假样本与训练数据集中的真实数据样本无法区分。

(4) 鉴别器：鉴别器将来自训练数据集的真实数据样本 x 或生成器生成的假样本 x^* 作为输入，对于每个样本，鉴别器将确定并输出样本为真实数据样本的概率。

GAN 的生成器和鉴别器通过博弈的手段对两个网络进行迭代优化，在训练过程中采用交叉训练的方式进行，基本流程如下。

(1) 初始化生成器的参数 θ_G 和鉴别器的参数 θ_D。

(2) 从分布为 $P_{\text{data}}(x)$ 的数据集中采样 m 个真实数据样本 $\{x^{(1)},x^{(2)},\cdots,x^{(m)}\}$，同时从噪声先验分布 $P_G(z)$ 中采样 m 个噪声样本 $\{z^{(1)},z^{(2)},\cdots,z^{(m)}\}$，并且使用生成器获得 m 个生成数据样本 $\{\tilde{x}^{(1)},\tilde{x}^{(2)},\cdots,\tilde{x}^{(m)}\}$。

(3) 固定生成器，使用梯度上升策略训练鉴别器使其能够更好地判断样本是真实数据样本还是生成数据样本，有

$$\nabla_{\theta_D}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^i)+\log(1-D(G(z^i)))] \tag{5-8}$$

(4) 循环多次对鉴别器的训练后，使用较小的学习率对生成器进行优化，固定鉴别器，生成器使用梯度下降策略进行优化，有

$$\nabla_{\theta_G}\frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^i))) \tag{5-9}$$

(5) 多次更新之后，理想状态是生成器生成一个鉴别器无法分辨的样本，即最终鉴别器的分类准确率是 0.5。

图 5.5 是 GAN 在训练过程中的鉴别器和生成器变化图，循环多次优化鉴别器，再优化生成器，因为想先拥有一个有一定效果的鉴别器，它能够比较正确地区分真实数据样本和生成数据样本，这样才能够根据鉴别器的反馈对生成器进行优化。

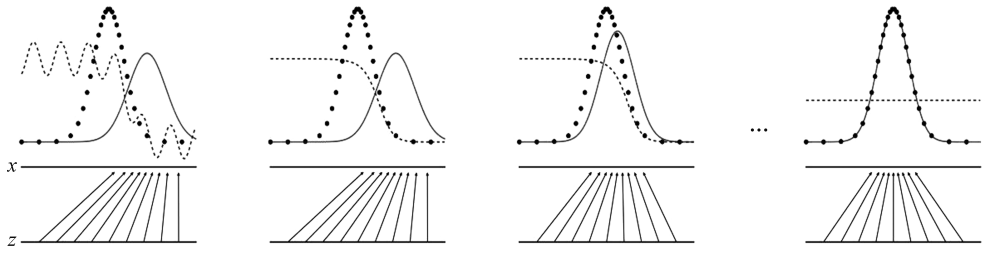

图 5.5　GAN 在训练过程中的鉴别器和生成器变化图

图 5.5 中，圆点线代表真实数据样本分布，实线代表生成的数据样本分布，虚线代表鉴别器判别概率的分布情况，z 表示噪声。从图 5.5 中可以看出，图 5.5(a)处于初始状态，此时生成数据样本和真实数据样本的差距比较大，而且鉴别器也不能对它们很好地进行区分，因此需要先对鉴别器进行优化。对鉴别器优化若干轮后，进入图 5.5(b)所示状态，此时鉴别器已能够很好地区分真实数据样本和生成数据样本，但生成数据样本和真实数据样本的分布差异还是非常明显的，因此需要对生成器进行优化。经过训练后真实数据样本和生成数据

样本的差异缩小很多，也就是图 5.5(c)所示状态。经过若干轮对鉴别器和生成器的训练后，生成数据样本和真实数据样本的分布已经完全一致(图 5.5(d))，此时鉴别器无法再区分，已达到预期的结果。

5.2.4 GAN 的算法分析

GAN 的训练过程都是无监督的，数据集是没有经过人工标注的，只知道这是真实的图像，如全是人脸图像，系统中的鉴别器并不知道该图像的类型，它只需要分辨真假。生成器也不知道自己生成的数据是什么类型，它只需要尽可能地学习真数据集欺骗鉴别器。

正由于 GAN 的无监督，在生成过程中，生成器就会按照自己的意思天马行空地产生一些诡异的图像，同时鉴别器可能给出一个很高的分数，如人脸极度扭曲的图像，这就是无监督学习目的性不强所导致的。

在 GAN 中，生成器和鉴别器两个网络有相互竞争的目标，即若一个网络变得更好，另一个网络必将变得更差。博弈论可能会认为这是一种零和博弈，一方的收益等于另一方的损失，所有的零和博弈都有一个纳什均衡点[8]。在纳什均衡点上，任何一方都不能通过改变自己的行为改善自己的处境或获得回报。

GAN 到达纳什均衡点，满足如下条件时，有

(1) 生成器生成的数据样本与训练数据集中的真实数据样本无法区分。

(2) 鉴别器最多只能随机猜测一个特定的数据样本是真或是假。

从鉴别器的角度看，当每个假样本 x^* 与来自训练数据集的真实数据样本 x 不可区分时，鉴别器就不能用任何方法分辨它们，因为鉴别器收到的数据样本有一半是真的，一半是假的，鉴别器能做的事情最好就是掷硬币，以 50%的概率将每个数据样本归类为真的或假的。从生成器的角度看，生成器同样处于一个点上，它从进一步的调整中没有任何收益，因为它产生的数据样本已经与真实的数据样本无法区分，所以，即使对它用来将随机噪声向量 z 转化为假数据样本 x^* 的过程进行微小的改变，也可能给鉴别器一个如何从真实数据样本中辨别假数据样本的线索，从而使生成器变得更糟。

实际应用中，由于在非零博弈中达到收敛复杂性较大，因此几乎不可能找到 GAN 的纳什均衡。事实上，GAN 收敛仍然是 GAN 研究中最重要的课题。

5.3 GAN 模型

GAN 常用于图像生成、人脸生成、图像转换、超分辨率和文本到图像的转换等领域，本节主要介绍 SGAN 模型、CGAN 模型、StackGAN 模型、InfoGAN 模型和 AC-GAN 模型。通过分析这些模型，读者可以充分体会它们在不同领域的应用。

5.3.1 SGAN 模型

半监督生成对抗网络(Semi-Supervised GAN，SGAN)模型作为 GAN 的一种，其鉴别器是多分类器，该鉴别器不只是区分两个类(真和假)，还要学会区分 $N+1$ 类，N 是训练数据集中的类数，生成器生成的伪样本增加一个类[9]。

要将鉴别器变成半监督分类器，除了计算其输入是否为实数的概率外，鉴别器还需要学习它所训练的每个原始数据集类的概率，通过该概率，鉴别器能够将一个信号发送回生成器，从而有可能提高生成器创造真实图像的能力。SGAN 模型的鉴别器须将其梯度发送回生成器，因为这是生成器在训练期间调整参数的唯一信息来源，然而，在真实图像的情况下，鉴别器还必须为每个单独的数据集类输出单独的概率，为此，可以将 Sigmoid 激活函数输出转换为具有 N 个类输出的 Softmax 函数，前 $N-1$ 个类为数据集的单个类概率(0～$N-2$)，第 N 个类是来自生成器的所有假图像，如果将第 N 类概率设置为 0，那么 $N-1$ 个概率的和表示先前使用 Sigmoid 激活函数计算的相同概率。图 5.6 是 SGAN 模型的结构示意图。

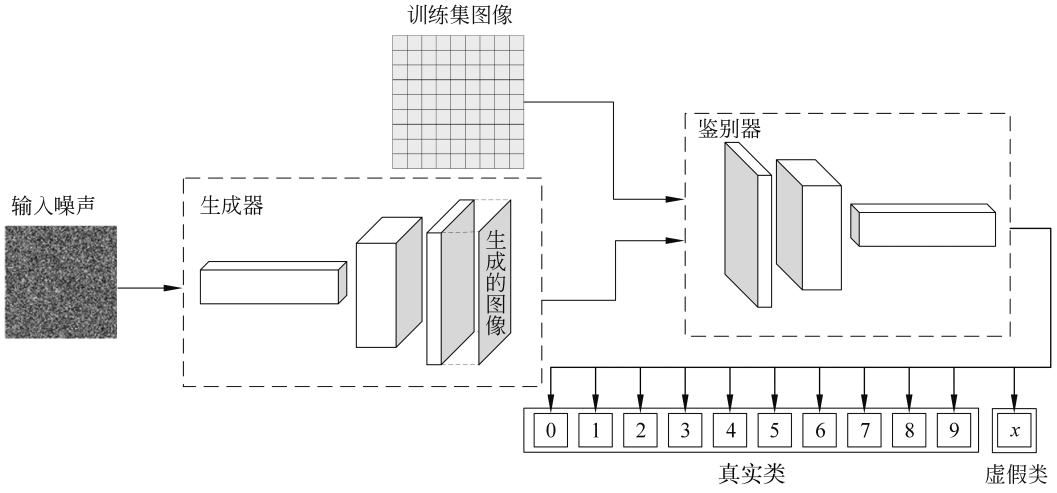

图 5.6　SGAN 模型的结构示意图

SGAN 模型生成器的目的与原始 GAN 相同，接收一个随机数向量并生成假样本，力求使假样本与训练数据集别无二致[10]。但是，SGAN 鉴别器与原始 GAN 实现有很大不同，它接收 3 种输入数据用于训练，具体情况如下。

(1) 带有标签的真实图像 X，与任何常规的监督分类问题相同，提供图像标签对。

(2) 无标签真实图像 X，分类器只知道这些图像是真实的。

(3) 来自生成器的图像 X^*，鉴别器会把它们归类为假的。

所有这些数据的组合使分类器能够从更广泛的角度进行学习，从而能够更精确地确定正确的结果。

SGAN 模型的生成器和鉴别器通过两个目标设置损失函数，具体情况如下。

(1) 为使得鉴别器能够帮助生成器产生真实的图像，可以通过学习区分真实和虚假的数据样本。

(2) 使用生成器的图像与标注后和未标注的训练图像对数据集精确分类。

对于 SGAN 模型的损失函数，除计算鉴别器的损失值，还必须计算生成器中有监督训练样本的损失 $D(x,y)$。所以，SGAN 模型有两种损失值，即有监督损失和无监督损失，具体为

$$L = L_{\text{supervised}} + L_{\text{unsupervised}} \tag{5-10}$$

$$L_{\text{supervised}} = -E_{x,y\sim P_{\text{data}}(x,y)} \log p_{\text{model}}(y \mid x, y < K+1) \tag{5-11}$$

$$L_{\text{unsupervised}} = -\left\{ \begin{matrix} E_{x\sim P_{\text{data}}(x)} \log[1 - p_{\text{model}}(y = K+1 \mid x)] \\ + E_{x\sim G}[\log p_{\text{model}}(y = K+1 \mid x)] \end{matrix} \right\} \tag{5-12}$$

SGAN 模型引入半监督学习的思想将鉴别器改进为多分类器，这是其与普通 GAN 最大的区别。

5.3.2 CGAN 模型

如果 GAN 的生成器和鉴别器都基于一些额外的信息 y，则 GAN 可以扩展为一个条件模型，y 可以是任何类型的辅助信息，如类标签或来自其他形式的数据，通过将鉴别器和生成器作为额外的输入层执行条件作用，进一步提出条件生成对抗网络(Conditional GAN，CGAN)模型[11]，如图 5.7 所示。

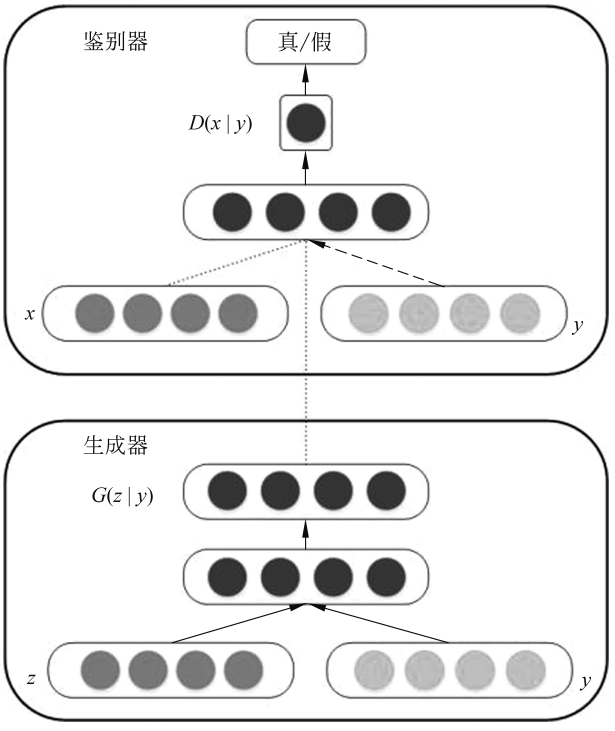

图 5.7 CGAN 模型的结构示意图

在生成器中，预先输入的噪声 z 和 y 组合在隐含层中，对抗训练框架在隐含层具有相当大的灵活性。

在鉴别器中，x 和 y 作为输入和鉴别函数，将条件输入和先验噪声作为多层感知机的单个隐含层的输入，可以使用更高阶的交互允许复杂的生成机制。

CGAN 模型的损失函数为

$$\min_G \max_D V(G,D) = E_{x\sim P_{\text{data}}(x)} \log(D(x \mid y)) + E_{x\sim P_G(z)} \log(1 - D(G(z \mid y))) \tag{5-13}$$

GAN 虽然能生成新的数据，但是无法确切地控制新样本的类型，如手写数字集，无法通过 GAN 指定要生成的具体数字，而 CGAN 模型可以轻松地解决这个问题。

5.3.3　StackGAN 模型

从文本中生成逼真的图像有广泛的应用，包括图片编辑、计算机辅助设计等。基于给定文本描述，虽然 CGAN 模型能够生成与文本含义高度相关的图像，但训练 GAN 从文本描述生成高分辨率的逼真图像是非常困难的，如在 GAN 模型中添加更多的上采样层生成高分辨率图像（如 256×256 像素），会导致训练不稳定，并产生无意义的输出。StackGAN 模型将文本到图像的合成问题分解为 Stage-Ⅰ GAN 和 Stage-Ⅱ GAN 两个更容易处理的子问题[12-13]，StackGAN 模型将两个 GAN 堆叠在一起，从而形成一个能生成高分辨率图像的网络。Stage-Ⅰ以文本描述为条件生成具有基本颜色和粗略草图的低分辨率图像。通过对 Stage-Ⅰ的结果和文本进行条件设置，Stage-Ⅱ GAN 学会捕捉 Stage-Ⅰ GAN 忽略的文本信息，并为对象绘制更多的细节，产生更高分辨率的图像（如 256×256 像素）[14]。

Stage-Ⅰ GAN 和 Stage-Ⅱ GAN 网络主要由文本编码器（text encoder）、条件增强网络（conditioning augmentation network）、生成器网络（generator network）、鉴别器网络（discriminator network）和嵌入压缩网络（embedding compressor network）组成。

对于 Stage-Ⅰ GAN 和 Stage-Ⅱ GAN 的文本编码器和条件增强网络，其中文本编码器的唯一目的是将文本描述（t）转换为文本嵌入（ϕt），文本编码器网络将句子编码为高维（如 1024 维）文本嵌入。从文本编码器获取文本嵌入（ϕt）之后，将它们传输到一个全连接层，以生成均值等于 μ_0 和标准差等于 σ_0 的值，将它们用于创建对角线协方差矩阵，即将 σ_0 放在矩阵 $\Sigma(\phi t)$ 的对角线。最后，使用 μ_0 和 Σ_0 创建高斯分布，具体表示为

$$N(\mu_0(\phi t),\Sigma_0(\phi t)) \tag{5-14}$$

从刚创建的高斯分布中获取高斯条件增强变量 $\hat{c}_0$，有

$$\hat{c}_0=\mu_0+\sigma_0 N(0,I) \tag{5-15}$$

添加 CA 网络（Conditioning Augmentation network）能增加网络的随机性，通过捕获具有各种姿势和外观的对象，可以使生成网络更强大，它能产生更多的图文对。使用大量的图文对，可以训练一个抗干扰的强大网络。

Stage-Ⅰ GAN 的生成器网络是具有多个上采样层的深度卷积神经网络，生成网络是 CGAN 模型，其条件是高斯条件变量 $\hat{c}_0$ 和随机噪声变量 z，z 是从高斯分布 p_z 采样的随机噪声变量（尺寸为 N_z），生成器网络生成的图像可以表示为 $s_0=G_0(z,\hat{c}_0)$。鉴别器网络是一个深度卷积神经网络，其中包含一系列下采样卷积层，下采样层从图像生成特征图，无论它们是来自真实数据还是由生成器网络生成的图像，将特征映射连接到文本嵌入，使用压缩和空间复制将文本嵌入转换为连接所需的格式，空间压缩和复制包括一个全连接层，该层用于将文本嵌入压缩为一个 N_d 维输出，然后通过在空间上复制文本将其转换为 $N_d\times N_d\times N_d$ 维张量，将特征图以及压缩和空间复制的文本嵌入沿通道维合并。最后，具有一个节点的全连接层用于二分类。

鉴别器网络损失表示为

$$L^{(D_0)}=E_{(I_0,t)\sim p_{\text{data}}}[\log D_0(I_0,\phi)]+$$

$$E_{z\sim p_z,t\sim p_{\text{data}}}[\log(1-D_0(G_0(z,c_0),\phi_t))] \tag{5-16}$$

生成器网络损失表示为

$$L^{(G_0)}=E_{z\sim p_z,t\sim p_{\text{data}}}[\log(1-D_0(G_0(z,c_0),\phi_t))]+\lambda D_{KL}(N(\mu_0(\phi t),\Sigma_0(\phi t))\parallel N(0,I)) \tag{5-17}$$

Stage-Ⅱ GAN 的主要组件是生成器网络和鉴别器网络[8]。生成器网络是深层卷积神经网络，Stage-Ⅰ GAN 的结果(即低分辨率图像)通过几个下采样层生成图像特征，将图像特征和文本条件变量沿通道尺寸连接在一起，将连接的张量送入一些残差块，这些残差块学习跨图像和文本特征的多峰表示，最后一个操作的输出被输入一组上采样层，它们会生成高分辨率图像。鉴别器网络是一个深度卷积神经网络，并且包含额外的下采样层，因为图像的大小比 Stage-Ⅰ GAN 中的鉴别器网络大，是一个可识别是否匹配的鉴别器，这能够更好地匹配图像和条件文本，在训练期间，鉴别器将真实图像及其对应的文本描述作为正样本对，而负样本对则由两组组成：第一组是具有不匹配文本嵌入的真实图像；第二组是具有相应文本嵌入的合成图像。

Stage-Ⅱ GAN 中的生成器网络和鉴别器网络也可以通过使鉴别器网络的损失最大并使生成器网络的损失最小进行训练。

生成器损失表示为

$$L^{(D_0)}=E_{(I,t)\sim p_{\text{data}}}[\log D_0(I,\phi)]+E_{s_0\sim p_{G_0},t\sim p_{\text{data}}}[\log(1-D(G(s_0,\hat{c}),\phi_t))] \tag{5-18}$$

Stage-Ⅰ GAN 和 Stage-Ⅱ GAN 的两个生成器网络都以文本嵌入为条件，主要区别在于 Stage-Ⅱ GAN 是在 Stage-Ⅰ GAN 结果的基础上生成高分辨率图像，它以低分辨率的图像为条件，并再次嵌入文本，以纠正 Stage-Ⅰ GAN 结果中的缺陷。Stage-Ⅱ GAN 完成了之前忽略的文本信息，以生成更逼真的细节。

Stage-Ⅱ GAN 中的鉴别器网络损失表示为

$$L^{(D_0)}=E_{s_0\sim p_{G_0},t\sim p_{\text{data}}}[\log(1-D(G(s_0,\hat{c}),\phi_t))]+\lambda D_{KL}(N(\mu_0(\phi t),\Sigma_0(\phi t))\parallel N(0,I)) \tag{5-19}$$

5.3.4 InfoGAN 模型

无监督生成对抗网络(Information Maximizing GAN，InfoGAN)模型是 GAN 的信息论扩展，能够以完全无监督的方式学习分离表示[15]。另外，InfoGAN 模型能最大化小部分潜在变量和观测值之间的互信息。InfoGAN 模型不需要任何形式的监督，可以分离离散的和连续的潜在因素，扩展到复杂的数据集，且通常不需要比 GAN 更多的训练时间[16]。

InfoGAN 模型将输入噪声向量分解为不可压缩噪声 z 和隐含编码 c，并将数据分布进行结构化语义特征，采用 $c_1,c_2,\cdots,c_L$ 表示结构化的隐含变量集，代表生成数据不同的特征维度，如 MNIST 数据集可以由一个取值范围为 0～9 的离散随机变量表示数字，使用隐含编码 c 表示所有的潜在变量 c_i，为生成器网络提供不可压缩噪声 z 和隐含编码 c，因此生成器的形式变为 $G(z,c)$。然而，在 GAN 中，生成器可以通过找到满足式(5-20)的解来忽略额外的隐含编码，使其完全不起作用，即