# 生成对抗网络 GAN 的研究综述

汪美琴[1]，袁伟伟[2]，张继业[1]

（1. 中国航天科工集团第二研究院 七〇六所，北京 100854；2. 中国人民解放军 32180 部队，北京 100012）

**摘 要**：为深入研究 GAN 这一热点模型，对基本 GAN 模型的原理、优点以及存在的问题进行分析；介绍 GAN 的发展以及不同的衍生模型，进一步说明 GAN 模型所做贡献，由此提出未来 GAN 衍生的改进方向的建议；描述 GAN 模型在图像、目标检测、文本合成、信息安全等各个领域的应用现状，总结其应用优势、局限性，针对其存在的问题，提出一些改善方法；对本文进行总结以及对该领域未来的研究方向提出一些展望。

**关键词**：生成对抗网络；机器学习；生成模型；图像生成；信息安全

**中图法分类号**：TP181 **文献标识号**：A **文章编号**：1000-7024（2021）12-3389-07

**doi**：10.16208/j.issn1000-7024.2021.12.012

# Overview of research on generative adversarial network GAN

WANG Mei-qin[1], YUAN Wei-wei[2], ZHANG Ji-ye[1]

(1. Institute 706, Second Academy of China Aerospace Science and Industry Corporation, Beijing 100854, China;
2. 32180 Unit of the People's Liberation Army of China, Beijing 100012, China)

**Abstract**: For studying the hot spot model GAN, the principle and advantages and disadvantages of the basic GAN model were analyzed. The development of GAN was introduced, as well as the different derivative models, to further illustrate the contribution made by the GAN model, and some suggestions for the improvement direction of GAN derivative were proposed. The application status of GAN model in image, object detection, text synthesis, cybersecurity and other fields was described, its application advantages and limitations were summarized, and some improvement methods were put forward according to the existing problems. The paper ended with the conclusion and future aspects in the field of research.

**Key words**: generative adversarial networks; machine learning; generative model; image generation; cybersecurity

## 0 引 言

随着科技的进步和智能化技术的发展，国内外都掀起了“智能化”的狂潮，人工智能和深度学习成为最近几年的热词。一般来说，根据数据集是否被标注，机器学习算法大致可以分为监督学习和无监督学习两类[1]。有监督学习需要依赖已知带标记的数据，虽然取得了不错的成绩，但是成本较高，因此无监督学习受到了越来越多的关注[1]。在无监督学习任务中，生成模型是最有前途的技术之一，早期的生成模型有受限玻尔兹曼机、深度信念网络、变分自动编码器、深度玻尔兹曼机等，然而由于泛化性没有很好，影响了它们的性能和结果。Goodfellow 等[2]引入了生成模型领域的一个新概念，即生成对抗网络(generative adversarial networks，GAN)，它是一种类似于对抗博弈游戏的训练网络，来源于博弈论中的二人零和博弈思想，通过让两个神经网络相互博弈的方式进行学习，在不断地优化迭代之后，使模型达到最优，即纳什平衡状态。GAN 作为一种新兴的无监督学习的生成模型一经提出便备受关注，在许多领域使用 GAN 的愿望也越来越大[3]。迄今为止，GAN 在图像生成、语音合成、目标检测、风格迁移、隐私保护等方面已经有一些研究和应用。

本文首先介绍了基本 GAN 模型的结构和优点，并分析了其存在的问题，接着对 GANs 的衍生模型进行了总结归纳，随之对其在不同领域尤其是计算机视觉和信息安全领域的潜在应用进行了综述，最后对 GAN 未来的发展趋势和研究提出了一些展望。

**收稿日期**：2021-03-22；**修订日期**：2021-08-16

**作者简介**：汪美琴（1997-），女，湖北黄冈人，硕士研究生，研究方向为计算机网络与安全；袁伟伟（1982-），男，河南平顶山人，博士，助理研究员，研究方向为网络安全；张继业（1978-），男，河南郑州人，硕士，研究员，硕士生导师，研究方向为计算机网络与安全。E-mail：1569677451@qq.com

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

## 1 基本 GAN 模型简介

### 1.1 基本 GAN 模型的结构

生成对抗网络由两个重要部分组成，即判别器 D 和生成器 G，生成器 G 通过机器生成数据，目的是学习真实样本的分布，生成相似度逼近真实样本的伪样本[4]，而判别器 D 的作用是区分从数据获取的真实样本和由生成器 G 生成的伪样本。两个模型通过不断地对抗训练来迭代优化，使生成器 G 生成的数据分布最大可能接近于真实数据分布，当判别器 D 每次输出的概率基本为 1/2，即判别器相当于随机猜测样本是真是假，便说明模型达到了纳什均衡，即最优状态，这就是 GAN 的对抗性思想，如图 1 所示为 GAN 的模型结构，GAN 模型的训练过程可以分为 3 个阶段，第一阶段：固定判别器 D，训练生成器 G，第二阶段：固定生成器 G，训练判别器 D，然后循环阶段一和阶段二，不断进行训练。

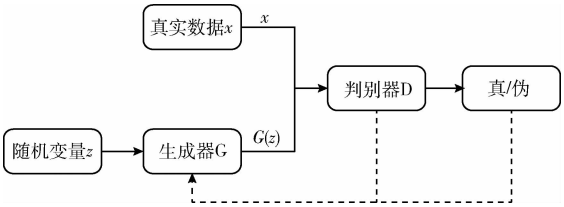


图 1 基本 GAN 模型结构

形式上，生成器 G 和判别器 D 的优化过程可以定义为二元极小极大博弈问题，其目标函数如下所示

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)]+ E_{z\sim P_{z(x)}}[\log(1-D(G(z)))]$$

式中：$P_{data(x)}$ 代表真实的数据分布，$P_{z(x)}$ 代表生成的数据分布，训练生成器以最小化目标函数，而训练判别器以使其最大化，最终使生成器生成类似于真实数据的数据[5]。

### 1.2 基本 GAN 模型的优点

由于生成器是一个简单的、确定的前馈网络[6]，因此 GAN 可以用简单的方式对生成的数据进行采样，而无需利用马尔科夫链和变分界线，也无需在学习过程中进行推断，因此在采样上计算较快且准确率较高。此外，GAN 模型可以并行生成样本，而无需利用与输入维数成比例的运行时间。再者，GAN 能够更好建立数据分布的模型，且已广泛应用于图像处理领域，因为事实证明能够很好的与图像配合使用[7]，能生成更清晰逼真锐利的图像。理论上，GAN 能够训练任何一种生成器网络，具有更加灵活的框架，与传统神经网络需要构建一个损失函数相比，GAN 可以学习损失函数。

### 1.3 基本 GAN 模型存在的问题

尽管 GAN 有许多优点，但是出于训练的目的，如何寻找纳什平衡解是一个挑战和难点，在实际训练中容易出现判别器 D 收敛、生成器 G 发散的现象，两个网络之间很难出现很好的同步，因此 GAN 面临着难训练、不稳定的问题。

GAN 存在模式崩溃现象，当生成器 G 学习到一个参数设置，可以生成对判别器 D 而言特别逼真的样本，由于此样本很容易骗过判别器 D，所以生成器 G 可能会一次又一次的生成相同的伪样本，最终始终生成同样的样本点，出现模式缺失，无法继续学习。

梯度消失也是 GAN 存在的问题之一，如果判别器 D 始终能够正确判断真实样本为真和生成样本为假，那么无论生成器 G 生成的样本多么好，判别器 D 都可以把它们分类为假样本，此时损失降为零，导致生成器 G 没有学习，就产生了梯度消失现象。

## 2 GAN 的发展及其衍生模型

针对基本 GAN 模型存在的难训练、不稳定、梯度消失、模式崩溃等问题，研究者们采用了许多方法去加以改进以期望解决这些问题，也随之出现了越来越多的衍生模型，有基于目标函数进行优化的，有基于架构进行优化的，这些方法虽然取得了一定的效果，但也是有着一定的代价的，接下来是具体的几种改进方式。

### 2.1 基于目标函数优化的 GANs

上文提到，生成器的目标是生成与真实数据分布接近的数据，GAN 的实质是生成器和判别器的对抗博弈问题，那么目标函数就是影响到基本 GAN 模型的直接因素。最初的 GAN 是通过最小化 Jensen-Shannon(JS) 散度来实现的，以最小化生成器的损失函数[1]，然而 JS 散度由于其自身函数域的问题，容易在训练中发生梯度消失的现象，导致生成数据真实数据分布差异较大。为了解决这类问题，文献[8] 提出了 f-GAN，认为可以根据任意凸函数下的 f-散度来构建 GAN 的目标函数，达到增强 GAN 稳定性的效果，但是实验结果表明，f-GAN 的结果具有较大的不确定性，更多的是一种推论，大部分问题仍然没有得到解决。

为了避免使用 JS 散度造成的梯度为 0 的问题，文献[9] 提出了 LSGAN(least square GAN)，使用最小二乘损失函数代替传统 GAN 模型中的交叉熵损失函数，在判别器达到最优的时候，将生成器优化的目标变成了皮尔森卡方散度，解决了 GAN 模型中梯度消失的问题，并且在整个学习过程中比基本 GAN 模型更加稳定。但尽管如此，LSGAN 使用的还是属于 f-散度，在衡量两个分布的相似程度时，仍然避不开零测集的问题，训练依然有可能不稳定。

针对散度距离问题，以及训练不稳定的情况，文献[10] 提出了 WGAN(Wasserstein GAN)，它没有使用 Jensen-Shannon(JS) 散度，而是使用 Earth-Mover 距离来计算真实数据的概率分布与生成数据的相似度，能够解决训练

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

不稳定、梯度消失的问题，但是由于判别器使用了权值裁剪的方法，WGAN 有可能出现产生的数据样本质量较低，甚至在收敛过程中有时会失败的问题。

基于 WGAN 的不足，文献［11］提出了 WGAN-GP，通过用惩罚机制代替了权重裁剪以强制执行 Lipschitz 约束，以此来解决 WGAN 中权重剪裁导致的梯度消失、参数不集中等问题，相比 WGAN 有更好的稳定性和更多的多样性，但是该方法收敛速度慢，可能需要训练多次才能达到收敛状态。此外，文献［12］提出了 WGAN-LP，它使用了一个新的惩罚项来实施 Lipschitz 约束，能够进一步提高网络训练的稳定性[1]，但是梯度惩罚增加了计算复杂度。

为了使模型训练更加稳定，与 WGAN 类似，均值和协方差特征匹配 GAN（McGAN[13]）引入了新的积分概率度量标准（integral probability metrics，IPM）进行训练[14]，对判别器的二阶矩阵进行了约束，但是由于裁剪的使用最终限制了模型的容量，且需要张量分解，因此难以扩展到高阶矩匹配。

### 2.2 基于架构优化的 GANs

虽然采用目标函数优化的方法，能够有效改善基本 GAN 模型梯度消失、训练不稳定等问题，但是这些方法或多或少都存在一些缺陷，于是研究者们从架构的角度进行考虑和优化，与其它的模型和算法进行结合，并取得了一些成果，下面将列举一些常见的优化变体。

#### 2.2.1 基于条件优化的 GANs

为了控制基本 GAN 模型生成过程过于自由的问题，文献［15］提出了条件生成对抗网络（CGAN），在基本 GAN 模型的基础上增加了约束条件，提供了一个附加参数（标签）和潜在空间来改变生成器，判别器则以真实图像和标签作为输入，这种改进使得网络能够朝着既定的方向生成样本。但是 CGAN 仅仅是添加了约束条件，模型训练的稳定性仍然是不足的。

辅助分类器 GAN（auxiliary classifier GANs，AC-GAN[16]），在基本 GAN 模型的基础上增加了一个分类器能够对类别标签进行分类，从而达到辅助训练的目的，能够以半监督的方式生成多样化和高分辨率的可分辨图像，但是当训练数据较少时，训练的图片多样性不足，仍然可能出现梯度消失等问题。

在基本 GAN 模型中，生成器的输入是一段噪声 z，因此很难控制 z 与生成数据的语义特征的对应关系，为了增加 GAN 模型的可解释性，文献［17］提出了信息生成对抗网络（InfoGAN），在随机噪声中添加了可解释性的潜在编码，从而能够帮助控制数据的生成。但是由于添加了解释性信息，增加了计算的负担，使得生成图像的多样性也不足。

DCGAN[18]（deep convolutional GAN）的提出为 GAN 的发展做出了突出贡献，它将卷积神经网络 CNN 和 GAN 结合起来，填补了 CNN 在有监督学习和无监督学习成功之间的差距[19]。实验结果表明，DCGAN 在大部分场景中都能学习到特征，并能利用学习到的特征来完成新的任务，在大多数训练下是稳定的。但是 DCGAN 只对结构进行了改进，没有对优化方法进行改进，还有一定的改进空间。

为了使模型训练更稳定，BigGAN[20] 通过对生成器应用正交正则化使其可以使用简单的“截断技巧”，从而通过减小生成器输入的方差来精确控制样本的保真度和多样性之间的权衡[20]。该方法虽然能使得生成图像的品质更好，但是成本较高。

#### 2.2.2 基于自编码器优化的 GANs

除了添加约束条件之外，利用其它模型的优点来结合优化 GAN 模型也是非常好的方向，目前已有将 GAN 与 VAE、RBM 等融合在一起的诸多研究，其中基于 VAE 的优化则更为常见。

对抗自动编码器（adversarial autoencoders，AAE[21]）结合了 GAN 与自动编码器的思想，通过将自编码器的隐藏层编码向量的聚合后验与任意先验分布匹配，利用 GAN 来执行变分推论，这种方法能使得生成的图像质量更高，生成结果更加可控。但是由于很难把 AAE 扩展到高分辨率图片数据上，并且解码器是以重构误差为目的进行训练，而非 GAN 那样以骗过判别器为目的，因此可能更难生成非常新的图像。

BiGAN（bidirectional GAN[22]）结合了编码器和判别器的结构来进行优化，将实际数据的概率分布映射到隐空间，从而有助于学习如何提取相关特征。但是由于 BiGAN 新增了一个编码器，增加了优化函数的计算复杂性。

BEGAN（boundary equilibrium GAN[23]）使用自动编码器架构作为判别器，该方法是从 Wasserstein 距离得出的损失来匹配自动编码器的损失分布，而不是直接匹配数据分布，所以 BEGAN 收敛很快，但是在超参数的选取上有一定的难度。

#### 2.2.3 GANs 的其它优化

除了以上提到的两种架构优化方法外，还有将 GAN 与其它领域知识结合从而对 GAN 进行优化的研究。文献［24］提出了对抗网络的拉普拉斯金字塔（LAPGAN），该方法在金字塔框架内使用级联的卷积网络，以从粗到精的方式生成图像。在金字塔的每个级别，都使用生成对抗网络技术来训练单独的生成卷积模型。LAPGAN 模型收敛速度快，能够生成分辨率高的样本，但是必须在有监督的情况下进行训练。

### 2.3 未来 GAN 衍生的改进方向

GAN 是一个持续进行的研究领域，因此，要研究出单个能够改善所有不足的衍生模型是一项挑战，关于 GAN 的变体未来还有一定的研究空间。经过综合比较和分析上述提到的 GAN 不同衍生模型的优化方法、优势、不足，以及适合的应用场景（表 1），本文对未来 GAN 衍生的改进方

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

向提出以下建议：

（1）从基本 GAN 模型内部结构的角度进行优化，尝试替换某个函数或者计算方法，从而实现对目标函数的改进，得到优化的 GAN 模型；

（2）尝试在基本 GAN 模型的基础上附加一些方法或者辅助工具，来控制 GAN 模型的结果朝目标方向改进；

（3）尝试结合其它模型的优点，来改进 GAN 训练不稳定等不足，比如目前有 VAE、RBM 与 GAN 的结合，未来可以根据应用需要，综合对算法进行优化；

（4）可以从多层次的角度进行改进，比如可以整体框架是 GAN 模型，而内部框架的实现可以采用其它框架或方法，每一级别也可以采用另外的框架或方法。

**表 1 GANs 的分析与比较**

| 名称 | 改进 | 优势 | 不足 | 应用场景 |
|---|---|---|---|---|
| f-GAN | 据任意凸函数下的 f-散度来构建 GAN 的目标函数 | 增强了 GAN 的稳定性 | 更多的是一种推论，具有较大的不确定性 | \ |
| LSGAN | 用最小二乘损失函数代替传统 GAN 模型中的交叉熵损失函数 | 解决了梯度消失的问题，更加稳定 | 可能会降低生成样本的多样性，训练时生成器可能会发生的梯度弥散问题 | 生成高质量的图像 |
| WGAN | 使用 Earth-Mover 距离来计算真实数据的概率分布与生成数据的相似度 | 解决了训练不稳定、梯度消失的问题 | 产生的数据样本质量较低，甚至在收敛过程中有时会失败 | 适合 GAN 模型不收敛、模式崩溃时使用 |
| WGAN-GP | 用惩罚机制代替了权重裁剪以强制执行 Lipschitz 约束 | 相比与 WGAN 有更好的稳定性和更多的多样性 | 收敛速度慢 | 模型参数不确定时使用 |
| WGAN-LP | 使用了一个新的惩罚项来实施 Lipschitz 约束 | 提高了网络训练的稳定性 | 梯度惩罚增加了计算复杂度 | 模型参数不确定时使用 |
| McGAN | 利用均值和协方差构建 IPM 进行训练，对判别器的二阶矩阵进行约束 | 使得模型训练更稳定 | 裁剪的使用最终限制了模型的容量；需要矩阵（张量）分解，难以扩展到高阶矩匹配 | 生成图像 |
| CGAN | 在基本 GAN 模型的基础上增加了约束条件 | 控制了 GAN 过于自由的问题，使网络朝着既定的方向生成样本 | 训练不稳定 | 无监督学习，半监督学习 |
| ACGAN | 附加类别标签输入生成器，判别器给出两个概率输出 | 生成多样丰富和高分辨率的图像 | 训练数据较少时，多样性不足 | 图像生成和分类 |
| InfoGAN | 在随机噪声中增加了可解释性的潜在编码 | 增加了对 GAN 模型的可解释性，能够控制图像的生成 | 增加计算负担，造成生成图像多样性不足 | 生成图像 |
| DCGAN | 将卷积神经网络 CNN 和 GAN 结合起来 | 提高了 GAN 的适用性和稳定性 | 只对结构进行改进，没有对优化方法进行改进 | 生成图像 |
| BigGAN | 对生成器应用正交正则化使其可以使用简单的“截断技巧”进行训练 | 使得模型训练稳定，且能使得生成图像的品质更好 | 模型大、参数多、成本较高 | 生成高品质图像 |
| AAE | 将自编码器的隐藏层编码向量的聚合后验与任意先验分布匹配，利用 GAN 来执行变分推论 | 生成图像质量更高，结果更可控 | 可能难以生成非常新的图像 | 生成图像 |
| BiGAN | 结合编码器和判别器的结构来进行优化，将实际数据的概率分布映射到隐空间 | 无监督 Bigan 比现有的弱监督模型具有更好的视觉特征学习性能 | 增加了优化函数的计算复杂性 | 生成图像 |
| BEGAN | 从 Wasserstein 距离得出的损失来匹配自动编码器的损失分布 | 收敛很快，而且判别器和生成器训练平衡 | 在超参数的选取上有一定的难度 | 生成高质量图像 |
| LAPGAN | 在金字塔框架内使用级联的卷积网络，在金字塔的每个级别使用 GAN 方法 | 收敛速度快，能生成分辨率高的样本 | 必须在有监督的情况下进行训练 | 生成高质量图像 |

## 3 GAN 的典型应用领域

由于 GAN 具有更灵活的框架，可以学习损失函数，即不需要为特定的应用去构建特定的损失函数，所以 GAN 在图像、视频、文本、安全、自然语言处理等诸多领域都有着广泛的应用。

### 3.1 图 像

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

GAN 具有很好的数据分布建模能力，而且其在图像领域的应用起步较早，所以目前已经取得了可观的成果，主要包括图像生成、图像翻译、图像修复等方面的应用。

3.1.1 图像生成

通过学习任何数据集的数据分布，GAN 可以对与原始数据集相似的新样本进行建模，加之端到端的工作方式，使得 GAN 相比传统的机器学习算法能更好地学习真实样本的特征分布和映射关系，因此在生成高分辨率图像方面具有更好的性能。Yang 等[25]提出了一种对抗性的图像生成模型(LR-GAN)，它通过考虑场景结构和上下文来生成清晰的图像，整个模型不受监督，并使用梯度下降法以端到端的方式进行训练。实验结果表明，LR-GAN 优于 DC-GAN，但是可能会出现输出图像与原始图像差异较大，或者图像失真的结果。为了生成更逼真的图像，Wang 等[26]提出了一种样式和结构模型(S2-GAN)，它有两个组成部分，Structure-GAN 用于生成图像结构，Style-GAN 将生成的图像结构作为输入以考虑图像样式。该模型首先对两个 GAN 进行独立训练，然后通过联合学习将它们合并在一起。虽然该方法生成效果较好，但是与最邻近的相比，生成的图像可能具有不同的风格和结构。

3.1.2 图像翻译

图像翻译是指将图像从一个域映射到另一域中的对应图像[27]。早期基于神经网络的图像翻译是采用卷积神经网络来实现的，但是其训练效率比较低，由于 GAN 算法不受形式约束，使用灵活，可以同时解决许多不同的任务，研究者们尝试将 GAN 应用到图像翻译领域。Isola 等[28]提出了基于 CGAN 的有监督学习模型(Pix2pixGAN)，用来完成成对的图像转换。Pix2pixGAN 不仅学习从输入图像到输出图像的映射，而且学习损失函数来训练该映射。结果表明，该方法生成的样本更真实，训练速度更快。但是该模型是监督模型，仍然需要带有标签和标记的数据，于是针对许多任务的配对的训练数据不可用的情形，Zhu 等[29]提出了一种无监督学习模型(CycleGAN)，它可以学习将图像从源域 X 转换为目标域 Y，并且不需要成对的图片作为训练数据。虽然该模型生成的图像质量不如 Pix2pixGAN，但是其应用场景更丰富灵活。

3.1.3 图像修复

传统的图像修复方法通常是根据周围的像素点估计待修复的像素点，但是这种算法大多比较复杂，而且对于大面积的图像损毁很难修复。由于 GAN 是通过对抗博弈的方式来进行训练的，所以在图像修复方面不用受限于可用的图像统计信息，且能使得修复效果更加自然。Yu 等[30]提出了一种生成式图像修复系统，可以使用自由形式的掩模和输入来完成图像。该系统基于从数百万张图像中学习的门控卷积，无需额外的标记工作，解决了将所有输入像素都视为有效像素的问题。此外，还提出了基于补丁的 GAN 丢失(SN-PatchGAN)，用于使训练快速稳定。但由于 GAN 模型本身存在的问题，导致在修复图像时，可能会出现过度平滑或模糊的情况。

### 3.2 其它应用

随着 GAN 的发展，除了在图像领域外，在目标检测、视频预测、文本合成图像、隐私保护、医学图像分割等方面也有着很好的表现。

3.2.1 目标检测

目标检测虽然已经取得了长远的发展，但是其中的小目标、大姿态等问题仍然是经典难题，传统的方法可能出现由于丢失高频造成的模糊等现象，而 GAN 良好的生成性能以及在图像领域优秀的表现，使得用 GAN 来解决此类问题成为可能。Li 等[31]提出了一种新的感知生成网络(perceptual GAN) 模型，通过缩小小对象与大对象之间的表示差异来改善小对象的检测，实验结果表明，该方法产生的对抗块扰动大、攻击效果好。但是缺陷在于牺牲了视觉保真度。

3.2.2 视频生成

传统的视频生成方法很难轻松地处理好帧不连续性的问题以及无文本的生成方案。为了提高生成结果的连续性，Tulyakov 等[32]提出了用于视频生成的运动和内容分解的生成对抗网络(MoCoGAN) 框架，通过将随机向量序列映射到视频帧序列来生成视频。该方法表明利用 GAN 来进行视频生成能够有效的进行建模，但可能会缺乏对语义的理解。

3.2.3 文本合成图像

在文本合成图像方面，主要目标是将视觉概念从字符转化为像素，生成具有逼真的细节的高分辨率图像。由于 GAN 模型能够更加充分的利用文本信息，对细节特征关注更多，所以能通过更细粒度的约束来提升生成效果。Zhang 等[33]提出了堆栈式生成对抗网络(StackGAN)，使用 GAN 架构从文字描述中合成图像，将两个 GAN 叠加进行分段式训练，尽管这种分段式模式可能会出现每个任务找不到重点，最终导致生成失败的情况，但是该方法提高了合成图像的多样性，并加强了 GAN 训练的稳定性。

3.2.4 信息安全

GAN 是深度学习领域的一项重大突破，信息安全领域的学者们也对其展开了相应的研究，相比于传统的方法而言，GAN 以半监督的方式训练检测器，有助于解决带标签样本少的问题，而且 GAN 通过学习实际的数据分布，在对实际的高维和复杂数据建模时具有更好的优势。Kim 等[34]提出了一种称为转移深度卷积生成对抗网络(tDCGAN) 的新方法，能够生成伪造的恶意软件，并将其与真实的恶意软件区分开，该方法使用深度自动编码器(DAE) 作为生成器，再将经过训练的判别器通过迁移学习用于恶意软件的检测。

### 3.3 GAN 应用问题的改善方法

由上文可以发现，针对不同的领域和问题，GAN 相对于传统的方法具有一定的优势，但是仍然存在一些问题没有解决，许多应用仍处于起步阶段，在未来还有较大的研

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

究空间。下面就预期的解决方案和发展方向提出一些建议：

（1）解决资源限制问题。由于 GAN 的训练是有资源限制的，所以它的安全性是小范围的，没有泛化，未来可以试图解决资源限制问题，提高 GAN 的普适性；

（2）解决对抗样本对 GAN 性能、精度、稳定性的影响。加了扰动的生成样本原本没有达到和真实样本一样的程度，但是由于添加了扰动，骗过了判别器，使得 GAN 网络训练出来的图像或者目标没有达到最优的效果便停止了；

（3）结合其它的机器学习算法。综合改进 GAN 的目标函数和结构框架，建立准确合理的生成模型，并且考虑不同领域结合时的安全性和鲁棒性，多尝试并且能够反推，突破往往是无意中发现的；

（4）数据集的选取和创新。不同的数据集对训练结果会有一定的影响，未来可以尝试在数据集的选取方面进行研究和创新。

## 4 GAN 未来的发展方向

随着 GAN 模型的不断探索，生成对抗模型极大地促进了图像处理领域的快速发展，同时它们在安全和医学等其它领域中也发挥着越来越重要的作用。虽然目前 GAN 的发展仍然面临着诸多挑战。但是相信经过更多的尝试和努力，GAN 会具有更广阔的应用前景，下面对该领域未来的发展和研究方向提出一些展望：

（1）通用的度量标准。目前尚未实现能够全面评估 GAN 模型的度量标准，不同的衍生模型在不同方面各有各的优势，因此无法绝对评判一个模型的好坏，新的模型层出不穷，制定一个通用、合理化、标准的度量标准有助于引导新的研究成果前进的方向，是急需解决的问题。

（2）完善的理论体系。虽然 GAN 一经提出，其“对抗训练”的思想便受到广大研究学者的欢迎，关于其研究也越来越多，但其存在的梯度消失、模式崩溃、训练不稳定等问题还是没有得到完美的解决，或许未来可以考虑从其最基本的模型结构出发，找到这些问题的根本原因，研究并完善其理论体系。

（3）结构拓展。虽然目前有许多衍生模型将其它的算法应用到 GAN 领域，但更多的是简单的叠加使用，如何将 GAN 与其它理论研究更完美的融合在一起，使 GAN 的应用范围更广泛，是未来需要考虑的问题。

## 5 结束语

本文对基本 GAN 模型的结构进行了介绍，并分析了 GAN 的优缺点。其次，分类总结了基于目标函数优化和基于架构优化的衍生模型，并通过观察 GAN 的不同变体的发展历程，对未来 GAN 衍生的发展方向提出了一些建议。然后介绍了 GAN 模型的一些典型的应用模型，包括图像生成、图像翻译、图像修复、视频生成、文本合成、信息安全等领域，并着重分析了这些模型的应用相比于传统方法的优势、存在的问题，就这些问题提出了一些改善方法。接着对 GAN 未来发展方向进行了展望，希望能为读者在研究主题或开发方法时提供指引。最后对本文进行了总结。

## 参考文献：


[1] Pan Z，Yu W，Yi X，et al. Recent progress on generative adversarial networks（GANs）：A survey［J］. IEEE Access，2019，7：36322-36333.

[2] Goodfellow I，Pouget-Abadie J，Mirza M，et al. Generative adversarial nets［C］//Advances in Neural Information Processing Systems，2014：2672-2680.

[3] Shu L，Yongchun Z. MIT technology review released “top 10 global breakthrough technologies” in 2018［J］. China Scitechnology Business，2018，3：34-37.

[4] Wu SQ，Li XM. Survey on research progress of generating adversarial networks［J］. Journal of Frontiers of Computer Science and Technology，2020，14（3）：377-388.

[5] Moti Z，Hashemi S，Namavar A. Discovering future malware variants by generating new malware samples using generative adversarial network［C］//9th International Conference on Computer and Knowledge Engineering，2019：319-324.

[6] Hong Y，Hwang U，Yoo J，et al. How generative adversarial networks and their variants work：An overview［J］. ACM Computing Surveys，2019，52（1）：1-43.

[7] Kumar S，Dhawan S. A detailed study on generative adversarial networks［C］//5th International Conference on Communication and Electronics Systems，2020：641-645.

[8] Nowozin S，Cseke B，Tomioka R. f-GAN：Training generative neural samplers using variational divergence minimization［J］. Advances in Neural Information Processing Systems，2016，6：271-279.

[9] Mao X，Li Q，Xie H，et al. Least squares generative adversarial networks［C］//IEEE International Conference on Computer Vision，2017：2813-2821.

[10] Arjovsky M，Chintala S，Bottou L. Wasserstein GAN［J］. arXiv e-prints，2017：arXiv：1701.07875.

[11] Gulrajani I，Ahmed F，Arjovsky M，et al. Improved training of Wasserstein GANs［C］//Advances in Neural Information Processing Systems，2017：5768-5778.

[12] Petzka H，Fischer A，Lukovnicov D. On the regularization of Wasserstein GANs［C］//International Conference on Learning Representations，2018：1-24.

[13] Mroueh Y，Sercu T，Goel V，et al. McGan：Mean and covariance feature matching GAN［C］//34th International Conference on Machine Learning，2017：2885-2899.

[14] Ghosh B，Dutta IK，Totaro M，et al. A survey on the progression and performance of generative adversarial networks［C］//11th International Conference on Computing，Commu-


(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

nication and Networking Technologies，2020：1-8.

[15] Mirza M，Osindero S. Conditional generative adversarial nets [J]. arXiv e-prints，2014：arXiv：1411.1784.

[16] Odena A，Olah C，Shlens J. Conditional image synthesis with auxiliary classifier GANs [C] //34th International Conference on Machine Learning，2017：4043-4055.

[17] Chen X，Duan Y，Houthooft R，et al. InfoGAN：Interpretable representation learning by information maximizing generative adversarial nets [C] //Neural Information Processing Systems，2016：2172-2180.

[18] Radford A，Metz L，Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. arXiv：1511.06434，2015.

[19] Ullah H，Khan SD，Ullah M，et al. Generative adversarial networks：A short review [C] //IS and T International Symposium on Electronic Imaging Science and Technology，2020：312-1-312-7.

[20] Brock A，Donahue J，Simonyan K. Large scale GAN training for high fidelity natural ImageSynthesis [J]. arXiv e-prints，2018：arXiv：1809.11096.

[21] Makhzani A，Shlens J，Jaitly N，et al. Adversarial autoencoders [J]. arXiv e-prints，2015：arXiv：1511.05644.

[22] Donahue J，Krhenbühl P，Darrell T. Adversarial feature learning [J]. arXiv e-prints，2016：arXiv：1605.09782.

[23] Berthelot D，Schumm T，Metz L. BEGAN：Boundary equilibrium generative adversarial net-works [J]. arXiv e-prints，2017：arXiv：1703.10717.

[24] Denton E，Chintala S，Szlam A，et al. Deep generative image models using a Laplacian pyramid of adversarial networks [J]. Advances in Neural Information Processing Systems，2015，1：1486-1494.

[25] Yang J，Kannan A，Batra D，et al. LR-GAN：Layered recursive generative adversarial net-works for image generation [J]. arXiv e-prints，2017：arXiv：1703.01560.

[26] Wang X，Gupta A. Generative image modeling using style and structure adversarial networks [C] //European Conference on Computer Vision. Springer International Publishing，2016：318-335.

[27] Saxena D，Cao J. Generative adversarial networks (GANs)：Challenges，solutions，and future directions [J]. arXive-prints，2020：arXiv：2005.00065.

[28] Isola P，Zhu JY，Zhou T，et al. Image-to-image translation with conditional adversarial networks [C] //IEEE Conference on Computer Vision and Pattern Recognition，2017：5967-5976.

[29] Zhu JY，Park T，Isola P，et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] //IEEE International Conference on Computer Vision，2017：2242-2251.

[30] Yu J，Lin Z，Yang J，et al. Free-form image inpainting with gated convolution [C] //IEEE/CVF International Conference on Computer Vision，2019：4470-4479.

[31] Li J，Liang X，Wei Y，et al. Perceptual generative adversarial networks for small object detection [C] //IEEE Conference on Computer Vision and Pattern Recognition，2017：1222-1230.

[32] Tulyakov S，Liu MY，Yang X，et al. MoCoGAN：Decomposing motion and content for video generation [C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition，2018：1526-1535.

[33] Zhang H，Xu T，Li H，et al. StackGAN：Text to photo-realistic image synthesis with stacked generative adversarial networks [C] //IEEE International Conference on Computer Vision，2017：5908-5916.

[34] Kim JY，Bu SJ，Cho SB. Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders [J/OL]. Information Sciences，2018，460-461：83-102. http：//hfcaf253cb3a601b84ef2sxwf0u55fx96q6pbu.fayx.eds.tju.edu.cn/science/article/pii/S00200255183034.

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net