中图法分类号: TP301.6　　**文献标识码:** A　　**文章编号:** 1006－8961(2018)10－1433－17

论文引用格式: Cao Y J, Jia L L, Chen Y X, Lin N, Li X X. Review of computer vision based on generative adversarial networks [J]. Journal of Image and Graphics, 2018, 23(10): 1433-1449. [曹仰杰, 贾丽丽, 陈永霞, 林楠, 李学相. 生成式对抗网络及其计算机视觉应用研究综述[J]. 中国图象图形学报, 2018, 23(10): 1433-1449.] [DOI: 10.11834/jig.180103]

生成式对抗网络及其计算机视觉应用研究综述

曹仰杰, 贾丽丽, 陈永霞, 林楠, 李学相

郑州大学软件与应用科技学院, 郑州　450000

摘　要: **目的**　生成式对抗网络(GAN)的出现为计算机视觉应用提供了新的技术和手段,它以独特零和博弈与对抗训练的思想生成高质量的样本,具有比传统机器学习算法更强大的特征学习和特征表达能力。目前在机器视觉领域尤其是样本生成领域取得了显著的成功,是当前研究的热点方向之一。**方法**　以生成式对抗网络的不同模型及其在计算机视觉领域的应用为研究对象,在广泛调研文献特别是GAN的最新发展成果基础上,结合不同模型的对比试验,对每种方法的基本思想、方法特点及使用场景进行分析,并对GAN的优势与劣势进行总结,阐述了GAN研究的现状、在计算机视觉上的应用范围,归纳生成式对抗网络在高质量图像生成、风格迁移与图像翻译、文本与图像的相互生成和图像的还原与修复等多个计算机视觉领域的研究现状和发展趋势,并对每种应用的理论改进之处、优点、局限性及使用场景进行了总结,对未来可能的发展方向进行展望。**结果**　GAN的不同模型在生成样本质量与性能上各有优劣。当前的GAN模型在图像的处理上取得较大的成就,能生成以假乱真的样本,但是也存在网络不收敛、模型易崩溃、过于自由不可控的问题。**结论**　GAN作为一种新的生成模型具有很高的研究价值与应用价值,但目前存在一些理论上的桎梏亟待突破,在应用方面生成高质量的样本、逼真的场景是值得研究的方向。

关键词: 生成式对抗网络; 计算机视觉; 图像生成; 图像风格迁移; 图像修复

Review of computer vision based on generative adversarial networks

Cao Yangjie, Jia Lili, Chen Yongxia, Lin Nan, Li Xuexiang

Department of Software Engineering and Applied Science and Technology, Zhengzhou University, Zhengzhou　450000, China

Abstract: **Objective** The appearance of generative adversarial networks (GANs) provides a new approach and a framework for the application of computer vision. GAN generates high-quality samples with unique zero-sum game and adversarial training concepts, and therefore more powerful in both feature learning and representation than traditional machine learning algorithms. Remarkable achievements have been realized in the field of computer vision, especially in sample generation, which is one of the popular topics in current research. **Method** The research and application of different GAN models based on computer vision are reviewed based on the extensive research and the latest achievements of relevant literature. The typical GAN network methods are introduced, categorized, and compared in experiments by using generation samples to present their performance and summarized the research status and development trends in computer vision fields, such as high-quality image generation, style transfer and image translation, text-image mutual generation, image inpainting, and restoration. Finally, existing major research problems are summarized and discussed, and potential future research directions are presented. **Result** Since the emergence of GAN, many variations have been proposed for different fields, either structural improvement or development of theory or innovation in applications. Different GAN models have advantages and disad-

收稿日期: 2018-03-02; **修回日期:** 2018-04-08; **预印本日期:** 2018-04-15

基金项目: 郑州大学优秀青年教师发展基金项目(1521337044); 河南省高等学校重点科研项目(17A520016)

vantages in terms of generating examples, have significant achievements in many fields, especially the computer vision, and can generate examples such as the real ones. However, they also have unique problems, such as non-convergence, model collapse, and uncontrollability due to high degree-of-freedom. Priori hypotheses about the data in the original GAN, whose final goals are to realize infinite modeling power and fit for all distributions, hardly exits. In addition, the designs of GAN models are simple. A complex function model need not be pre-designed, and the generator and the discriminator can work normally with the back propagation algorithm. Moreover, GAN can use a machine to interact with other machines through continuous confrontation and learn the inherent laws in the real world with sufficient data training. Each aspect has two sides, and a series of problems are hidden behind the goal of infinite modeling. The generation process is extremely flexible that the stability and convergence of the training process cannot be guaranteed. Model collapse will likely occur and further training cannot be achieved. The original GAN has the following problems: disappearance of gradients, training difficulties, the losses of generator and discriminator cannot indicate the training process, the lack of diversities in the generated samples, and easy over-fitting. Discrete distributions are also difficult to generate due to the limitations of GAN. Many researchers have proposed new ways to address these problems, and several landmark models, such as DCGAN, CGAN, WGAN, WGAN-GP, EBGAN, BEGAN, InfoGAN, and LSGAN, have been introduced. DCGAN combines GAN with CNN and performs well in the field of computer vision. Furthermore, DCGAN sets a series of limitations for the CNN network so it can be trained stably and use the learned feature representation for sample generation and image classification. CGAN inputs the conditional variable (c) with the random variable (z) and the real data (x) to guide the data generation process. The conditional variable (c) can be category labels, texts, and generated targets. The straightforward improvement proves to be extremely effective and has been widely used in subsequent work. WGAN uses the Wasserstein distance to measure the distance between the real and generated samples instead of the JS divergence. The Wasserstein distance has the following advantages. It can measure distance even if the two distributions do not overlap, has excellent smoothing properties, and can solve the gradients disappearance problem to some degrees. In addition, WGAN solves the problems of instability in training, diversifies the generated examples, and does not require the careful balancing of the training of G and D. WGAN-GP replaces the weight pruning in WGAN to implement the Lipschitz constraint method. Experiments show that the quality of samples generated by WGAN-GP is higher than those of WGAN. It also provides stable training without hyperparameters and successfully trains various generating tasks. However, the convergence speed of WGAN-GP is slower, that is, it takes more time to converge under the same dataset. The EBGAN interprets GAN from the perspective of energy. It can learn the probability distributions of images with low convergence speed. The images BEGAN products are still disorganized, whereas other models have been able to express the outline of the objects roughly. However, the images generated by BEGAN have the sharpest edges and rich image diversities in the experiments. The discriminator of BEGAN draws lessons from EBGAN, and the loss of generator refers to the loss of WGAN. It also proposes a hyper parameter that can measure the diversity of generated samples to balance D and G and stabilize the training process. The internal texture of the generated images of InfoGAN is poor, and the shape of the generated objects is the same. As for the generator, in addition to the input noise (z), a controllable variable (c) is added, which contains interpretable information about the data to control the generative results, resulting in poor diversity. LSGAN can generate high quality examples because the object function of least squares loss replaces the cross-entropy loss, which partly solves the two shortcomings (i. e., low-quality and instability of training process). **Conclusion** GAN has significant theoretical and practical values as a new generative model. It provides a good solution to problems of insufficient sample, poor quality of generation, and difficulties in extracting features. GAN is an inclusive framework that can be combined with most deep learning algorithms to solve problems that traditional machine learning algorithms cannot solve. However, it has theoretical problems that must be solved urgently. How to generate high-quality examples and a realistic scene is worth studying. Further GAN developments are predicted in the following areas: breakthrough of theory, development of algorithm, system of evaluation, system of specialism, and combination of industry.

Key words: generative adversarial networks; computer vision; image generation; style transfer; image inpainting

0 引 言

生成式对抗网络(GAN)是 2014 年由 Goodfellow 等人[1]提出的一种生成式深度学习模型,该模型一经提出就成为了计算机视觉研究领域热点研究方向之一[2]。近年来,随着深度学习及移动设备的快速发展,图像处理[3]、图像风格迁移[4]、基于图像内容的检索与分类[5]、图像生成[6]等领域已经成为一个有巨大应用价值的课题。GAN 能够生成目标数据集,以弥补训练数据不足的缺陷,因而对深度学习意义重大;此外 GAN 在场景生成、图像翻译、文本与图像的相互生成、视频预测等领域都发挥了独特的作用。将从 GAN 及其在计算机视觉领域方面的研究进展与应用进行讨论。

计算机视觉是指使用计算机及相关设备对生物视觉的一种模拟,最终目标使计算机能够像人类一样通过视觉观察进而理解世界,具有自主适应环境的能力[7]。判断机器是否理解现实世界,可以看它能否创造出和真实世界一样的物体,当人类无法分辨看到的是真实影像还是计算机生成的虚假影像,即通过图灵测试[8]。在计算机出现以后,出现许多生成算法以描绘世界。传统的生成算法有梯度方向直方图(HOG)[9],尺度不变特征变换(SIFT)[10]等,这些算法采用手工提取特征与浅层模型相组合的方法实现目标的生成。其解决方案基本遵循 4 个步骤:图像预处理→手动特征提取→建立模型(分类器/回归器)→输出。而 GAN 等深度学习算法解决计算机视觉的思路是端到端(End to End)[11],即从输入直接到输出,中间采用神经网络自动学习特征,避免手动特征提取的繁琐操作,不需要人工干预。此外,GAN 与变分自动编码器(VAE)[12]、自回归模型(AR 模型)[13]等基于机器学习的生成模型相比有如下优势:GAN 理论上可以逐渐逼近任何概率分布,可以看作是一种非参数的产生式建模方法,若判别器训练良好,生成器可以生成与真实样本几乎相同的分布,因此,GAN 是渐进一致的。相比而言,VAE 会依赖预先假设的近似分布,而对近似分布的选择需要一定的经验信息;它还受变分方法本身的限制,最终学到的概率分布存在偏差。GAN 与自回归模型相比,它直接对整幅图像采样、评价和生成,生成目标的时间更短,一次产生一个样本,GAN 考虑全局信息且速度相对较快,因此在生成问题上使用 GAN 能够更快速、高效地解决问题,这也促使了对 GAN 的进一步研究。

除了对 GAN 模型本身的改进不断完善,针对不同机器视觉问题的模型也被提出,为很多领域带来新的解决问题的方法。GAN 的应用范围如此广泛是因为 GAN 是一个深度学习框架,采用二人零和博弈的思想,理论上可以生成任意分布,其与其他深度学习模型如卷积神经网络(CNN)[14-15]、循环神经网络(RNN)[16]、长短周期神经网络(LSTM)[17]等判别模型不同,GAN 是生成模型,通过从所给样本与标签中学习联合概率分布,生成与训练样本相似的分布或者生成标签描述的对象,如生成与训练集中同样的图像、视频、文字,或者生成文字描述的花、鸟图像;而 CNN、RNN、LSTM 等判别模型一般用于分类,将所给训练样本分成对应的类别,即学习所给样本的条件概率分布,其中 CNN 擅长处理图像,RNN 与其变体 LSTM 擅长处理文本和与时间相关的序列。从结构看,GAN 包含一对协同工作的网络:判别网络和生成网络,可以直接使用上述判别模型作为判别网络,而生成网络需要生成详细分布,这与特征提取相反,因此生成网络一般使用解卷积网络,它可以看成 CNN 的逆过程。由于 GAN 的特殊结构,可以使用任何网络作为其生成器与判别器,将从结构与应用详细介绍 GAN 模型的衍变与可能的发展方向。

首先介绍生成式对抗网络 GAN 的经典模型与工作原理,及其在计算机视觉方面的最新研究进展,结合应用场景介绍模型的创新与改进,并对其发展趋势进行展望。

1 生成式对抗网络

该节介绍 GAN 的网络结构及其两个重要组成部分:生成网络和判别网络,详细介绍了它们的工作过程与原理,也比较了 GAN 的优势与缺点。

1.1 GAN 网络结构

GAN 包括两个模型,生成模型(G)和判别模型(D)。生成器 G 和判别器 D 本质上都是函数,通常用深层神经网络来实现[18]。GAN 模型结构如图 1 所示,G 从真实样本中捕获数据分布映射到某个新的数据空间,输出生成的数据记作 $G(z)$,其分布记

作 $p_g(z)$，并尽量使其看上去和训练集中样本 $p_r(x)$ 一样。D 的输入包括真实数据 x 与生成数据 $G(z)$，输出是一个概率值或一个标量值，表示 D 认定输入是真实分布的概率，数值越大，是真实数据的概率越大，反之认为输入的是生成样本，D 根据输出反馈给 G，使 G 生成的数据与真实数据逐渐一致。理想状态下 D 无法分辨输入的是真实数据还是生成数据，即满足 $p_r(x)/(p_r(x)+p_g(z))$ 最小，此时认为 G 已经学到真实数据的分布，模型达到最优。

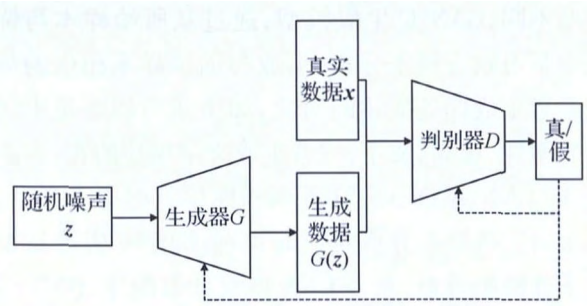

图 1　GAN 基本框架结构图

Fig. 1　The basic framework structure of GAN

GAN 对应于极小极大的双玩家游戏，又叫二人零和博弈[19]，在包含两个神经网络的零和游戏框架中相互竞争。系统中两名游戏者由两个函数表示，这两个函数对于它们各自的输入和参数都是可微的。判别器的函数用 D 表示，它的输入为 x，参数为 $\boldsymbol{\theta}^{(D)}$。生成器的函数用 G 表示，它的输入为 z，参数为 $\boldsymbol{\theta}^{(G)}$。两个玩家都有各自定义的损失函数。$D$ 需要通过更新 $\boldsymbol{\theta}^{(D)}$ 极小化 $J^{(D)}(\boldsymbol{\theta}^{(D)},\boldsymbol{\theta}^{(G)})$。$G$ 需要通过更新 $\boldsymbol{\theta}^{(G)}$ 极小化 $J^{(G)}(\boldsymbol{\theta}^{(D)},\boldsymbol{\theta}^{(G)})$。两个玩家的损失函数都依赖于对方的参数，但是却不能更新对方的参数，经过训练达到一个纳什均衡[20]。在该环境中，一个纳什均衡指的是一对参数 $(\boldsymbol{\theta}^{(D)},\boldsymbol{\theta}^{(G)})$，使得 $\boldsymbol{\theta}^{(D)}$ 是 $J^{(D)}$ 的一个极小值点，同时 $\boldsymbol{\theta}^{(G)}$ 是 $J^{(G)}$ 的一个极小值[21]。GAN 的优化实际上是一个极小极大化问题，其目标函数定义为

$$\min_G \max_D V(D,G) = E_{x\sim p_r(x)}[\ln D(x)] + E_{z\sim p_g(z)}[\ln(1-D(G(z)))] \tag{1}$$

式(1)的前一项代表当输入真实数据时判别器使得目标函数尽量大，判断其为真实数据；后一项代表当输入生成数据时，生成器使得 $G(z)$ 尽量小，即目标函数的值比较大，生成器欺骗判别器，使其误认为此时输入为真实数据，判别器努力鉴别其为假数据，两者相互博弈，最终达到一个纳什均衡。

1.2　生成网络

生成器用可微函数 G 表示，输入 z 是一个随机变量或者隐空间的随机变量，一般使用高斯变量或噪声，G 生成假样本分布 $G(z)$。G 网络只要求少量限制条件，对于输入变量，既可以把它输入到第一层，也可以输入到最后一层；还可以对隐层增加噪声，增加的方式可以是求和、乘积或做拼接。GAN 对输入变量 z 的维度没有限制，它通常是一个 100 维的随机编码向量。但需要注意 G 必须是可微的，因为经过判别器的“判断”会将它的梯度传回 G、D 来更新参数，否则误差无法传递。

1.3　判别网络

在 GAN 中，判别器 D 的主要目标是判断输入是否为真实样本并提供反馈机制，其与生成网络构成一个零和游戏。这个游戏由两个场景构成，在第一个场景中，从真实训练数据中采样 x，作为 D 的输入，D 输出的是一个 0 到 1 之间的数，表示 x 属于真实样本的概率。通常还会假设真实样本和伪造样本的先验比例是 1∶1。在第 1 个场景下，$D(x)$ 被训练地尽量输出接近 1 的概率值。在第 2 个场景中，从一个先验分布中采样出变量 z，将 $G(z)$ 作为 D 的输入，在这个场景中，两名玩家都要参与，D 的目标是使得输出 $D(G(z))$ 接近 0，而 G 的目标是使得它输出接近 1。两个玩家的模型经过足够的训练，游戏最终会达到一个纳什均衡，此时 $G(z)$ 与从真实样本中采样出的一样，而 $D(x)$ 对所有输入 x 的函数值都是 1/2，无法判断真假。

1.4　GAN 的优势与劣势

GAN 自出现以来，针对不同领域的许多变体被提出，它们或在结构上有所改进，或在理论有所发展，或在应用上有所创新。在 Goodfellow 等人[1]提出的原始 GAN 中，先验假设很少，对于数据没有做任何假设，它可以是任何分布，最终目标使 GAN 具有无限的建模能力，可以拟合一切分布。另外，GAN 模型设计简单，不必预先设计复杂函数模型，使用反向传播算法(BP)训练网络，生成器和判别器就能正常工作；GAN 为创建无监督学习模型提供了强有力的算法框架，它颠覆了传统人工智能算法，不是用人的思维去限定机器，而是用机器来“对话”机器，通过自身的不断对抗博弈，经过足够的数据训练，能够学到现实世界内在规律。

事情都有两面性，无限建模能力的目标背后隐藏一系列问题，由于生成过程过于自由，训练过程的稳定性和收敛性难以保证，容易发生模式崩塌，进而出现无法继续训练的情况；原始 GAN 存在如梯度消失，训练困难，生成器和判别器的损失无法指示训练进程，生成样本缺乏多样性，容易过拟合等问题；在由于 GAN 本身的局限性，它很难学习生成离散的分布，比如文本。到目前为止，许多新的 GAN 模型的提出或者训练技巧的改进都是为了增加模型的稳定性，提高生成结果的质量[22]。

2 GAN 模型的衍化

针对原始 GAN 存在的问题，研究者们提出许多新的方法改进，本节将介绍几个有里程碑意义的改进模型，它们对比结果如表 1 所示。

表 1 典型 GAN 模型对比

Table 1 Comparisons of typical GAN models

GAN 模型	改进	优点	缺点	适用场景
CGAN[23]	增加一个条件变量 c 对模型增加约束条件，指导数据生成过程	对输入输出增加一个标签，能够生成指定目标，收敛也更快	对数据集要求高，需要有标签或标记好的数据集	适合有监督学习或者指定生成目标的场景
DCGAN[24]	GAN 与 CNN 结合；在结构上采用步幅卷积、微步幅卷积、批标准化、LRE-LU 等操作	稳定训练过程，易收敛，生成样本多样性丰富	训练不同数据需要调整参数，模型易崩溃，会出现梯度消失或爆炸	适合大部分场景，是使用率最高的模型
WGAN[25]	权重剪枝	训练过程更稳定，理论上解决梯度消失的问题	由于权重的不恰当剪枝，可能会出现梯度消失或爆炸	一般 GAN 不收敛，模型崩塌的情况
WGAN-GP[26]	取代权重剪枝采用梯度惩罚	不用平衡生成器与判别器，训练过程稳定，开箱即用，能直接处理文本	收敛慢，生成样本的多样性不如 DCGAN	模型参数不确定的情况，需要直接处理文本的场景

2.1 条件生成对抗网络

GAN 作为一种无监督学习方法，它从无标注的数据集中学习到概率分布规律，并表示出来，这个过程缓慢、自由。当数据集中图像内容复杂，规模较大，使用简单 GAN 模型很难控制生成的结果，机器理解的重点与人类理解存在偏差，最终导致生成结果与目标并不一致。一个自然的想法是增加约束条件，给生成器制定目标，文献[23]提出条件 GAN 即 CGAN。该 GAN 模型在输入随机变量 z 和真实数据 x 时，一同输入的还有条件变量 c，使用增加的信息 c 对模型增加约束条件，指导数据生成过程，条件 GAN 结构如图 2。条件变量 c 可以是类别标签，这样 CGAN 把无监督的 GAN 变成了一种有监督模型；c 也可以是一个文本输入，比如一段描述句子，镶嵌到与之对应的图片中，经过训练，模型可以“看图说话”；c 同样可以是对应的目标图片，这样 GAN 可以有目标地去学习；CGAN 不仅可以以类别为标签生成指定类别的图像，还可以以图像特征为条件变量，生成该图像的目标词向量。这个简单直接的改进被证明非常有效，广泛用于后续工作中。

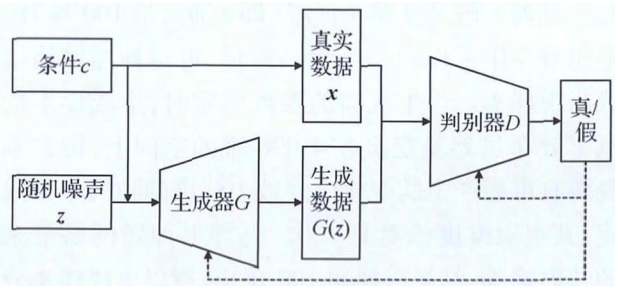

图 2 CGAN 基本框架

Fig. 2 The basic framework structure of CGAN

2.2 深度卷积生成对抗网络

GAN 发展的一个里程碑是 Radford 等人[24]提出的深度卷积生成对抗网络 DCGAN，其生成器结构如图 3 所示。它将计算机视觉领域表现很好的卷积神经网络 CNN 与 GAN 结合起来，其为 CNN 的网络拓扑结构设置了一系列的限制来使得它可以稳定地训练，使用学到的特征表示进行图像分类，得到好的效果验证模型的特征表达能力。DCGAN 的提出使

GAN 生成图像的质量有了保证，源于它对原始 GAN 做出的改进，首先判别器上使用步幅卷积和生成器上使用微步幅卷积代替池化[27]。不同于一般 CNN 用来提取特征，DCGAN 中的 CNN 结构需要生成图像，池化会忽略很多信息，而步幅卷积和微步幅卷积结构能够将大部分信息传给下一层，保证了生成图像的完整性和清晰度。其次引入批规范化(BN)操作[28]，这部分解决了梯度消失的问题，因为 BN 操作解决初始化差的问题，使梯度传播到每一层，防止生成器把所有样本收敛到同一点。再者移除全连接层和使用不同的激活函数，如 Adam 优化[29]，生成器使用 ReLU 激活函数[30]，判别器使用 leakyReLU[31] 激活函数。结果表明，DCGAN 在工程上取得了非常好的效果，此后的 GAN 结构在对比时一般以它为标准，也证实了 GAN 结构在生成样本领域的能力。

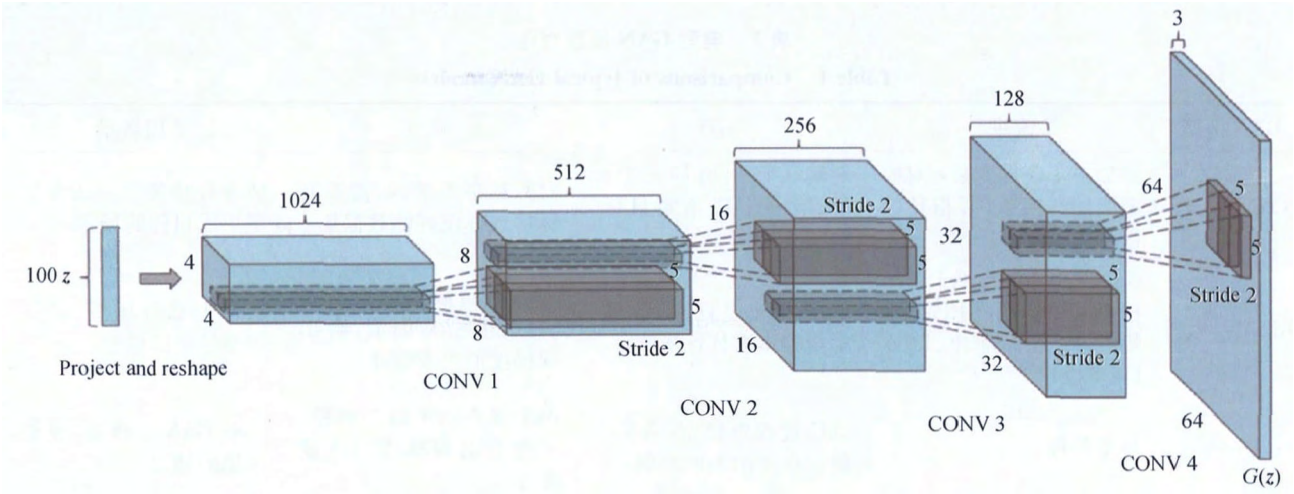

图 3　DCGAN 的生成器网络结构图

Fig. 3　The generator network structure of DCGAN

2.3　Wasserstein 生成对抗网络

GAN 训练中会出现梯度消失的问题，因为 GAN 的生成器一般是从某个低维(即 z 通常取 100 维)的随机分布中采样的一个编码向量，再经神经网络生成高维样本。当生成器的参数固定时，生成样本的概率分布虽然是定义在 4 096 维的空间上，但它本身所有可能产生的变化已经被 100 维的随机分布限定，其本质维度依然是 100。再加上神经网络带来的映射降维，结果可能比 100 还小，所以生成样本分布的支撑集就在 4 096 维空间中构成一个最多 100 维的低维流形，填不满整个空间，两个空间就很难有重合的部分，生成分布与真实分布的相似度衡量函数即 Jensen-Shannon(JS)散度变成一个常数，导致梯度消失，无法继续训练模型[32]。为了解决上述问题，文献[25]提出了 Wasserstein GAN(WGAN)，该模型使用 Wasserstein 距离(又称为 Earth-Mover(EM)距离)代替 JS 散度对真实样本和生成样本之间的距离进行度量。Wasserstein 距离比 JS 散度的优越性在于，即使两个分布没有重叠，Wasserstein 距离仍然能够较好地度量距离的远近，同时它又有优越的平滑特性，理论上可以解决梯度消失问题。除此之外，WGAN 还解决了训练不稳定的问题，不需要再小心平衡 G 和 D 的训练程度，而且生成的样本具有多样性。最重要的是在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过程，这个数值越小代表 WGAN 训练地越好，生成器产生的图像质量越高[33]。

2.4　改进 Wasserstein 生成对抗网络

文献[26]提出了一种改进的 WGAN 结构，命名为 WGAN-GP，它是具有梯度惩罚的 WGAN，取代 WGAN 中权重剪枝实施 Lipschitz 约束方法。实验证明该方法生成的样本质量比 WGAN 高，提供稳定的训练，几乎不需要超参数调参，能成功训练多种针对生成任务的 GAN 架构。但实验表明该方法的收敛速度较慢，同一数据集下需要更多的训练次数才能收敛。上述 DCGAN、WGAN、WGAN-GP 都是在优化方法或约束方法上改进了 GAN，并没有改变 GAN 的结构，本质上它们都是朝着能生成更好样本的方向去改进原始 GAN，增加约束条件。

2.5 不同 GAN 模型实验对比

为了测试不同 GAN 的实际性能,将 DCGAN 作为基准模型,在 CGAN、WGAN、WGAN-GP、基于能量的 GAN(EBGAN)[34]、边界均衡 GAN(BEGAN)[35]、Information GAN(INFOGAN)[36]、最小二乘 GAN(LSGAN)[37]上进行对比实验,实验代码来自网络(https://github.com/hwalsuklee/tensorflow-generative-model-collections)[38],该实验使用同一数据集,在相关参数一致的情况下检验生成结果。实验使用 Fashion-MNIST[39]数据集与 MNIST 数据集[15],它是一个新建的替代 MNIST 手写体数据集的图像数据库。数据集包括 10 种类别的共 7 万个不同商品的正面图片,分别是 T 恤、裤子、套衫、裙子、外套、凉鞋、汗衫、运动鞋、包、踝靴,对应着数字标注编号 0 ~ 9,这与 MNIST 保持一致,此外,Fashion-MNIST 的名称、大小、格式和训练集、测试集划分与原始的 MNIST 完全一致,包括 60 000 张训练图片与 10 000 张测试图片,大小 28 × 28 像素的灰度图片。使用这个数据集和在 MNIST 上的实验结果对比,因为 MNIST 过于简单,其中的许多数字只需要一个像素就能区分开,实验也表明使用 MNIST 不利于机器学到更详细的特征,因为它不涉及内部的纹理表示,实际的问题远比识别分类 MNIST 中的数字更困难。该实验列出了在该数据集验证不同模型生成图像的能力,实验结果如图 4 所示。

图 4 是 8 个不同 GAN 模型及其生成图像的结果,所有模型以 DCGAN 结构为基础,实验设置判别器包括一个输入层、输出层,两个卷积层、全连接层,生成器包括一个输入层、输出层,两个全连接层、反卷积层,网络内部采用的激活函数、批标准化操作也相同。从图 4 的第一个迭代看生成图像的质量,WGAN 收敛最快,生成的图像较清晰,与其他模型相比,图像的边缘与背景易区分因为它使用权重剪枝的方法,模型更加的稳定因此生成器能更快的学到图像分布规律。而 WGAN-GP 在实际中收敛速度最慢,最终生成结果也表明,它在 40 个迭代下生成的结果并不清晰,其他模型在同样情况下模型已收敛。经过更进一步的实验,其在 60 个迭代下能够生成清晰地图像。原因在于其相比 WGAN 把权重剪枝改为根据判别器的输入计算出权重梯度,并针对梯度的范数进行惩罚,要训练的参数更多,所以收敛慢,但它是开箱即用,不需要任何调整,学习率的改变对它影响很小,模型非常稳定。图 4 中 CGAN 比较特别,在模型中了加入控制条件,能够生成指定类别的图像,且收敛速度更快。EBGAN 从能量的角度诠释 GAN,实验结果表明此方法也可以学习图像的概率分布,但其收敛速度很慢,其他模型已经能够大致表达图像轮廓,它生成的图像仍旧杂乱无章。实验中 BEGAN 生成的图像边缘最清晰,图像多样性丰富,该模型的判别器借鉴 EBGAN,生成器借鉴了 WGAN 损失的定义方法; 该论文还提出了一个衡量生成样本多样性的超参数来均衡 D 与 G,从而稳定训练过程。InfoGAN 生成图像的内部纹理并不好,同类物体外形形似。因其生成器中除了输入噪声 z,还加入控制变量 c,它包含对数据的可解释信息以期控制生成结果,造成多样性差。LSGAN 生成图像质量很高,源于将 GAN 的目标函数由交叉熵损失替换为最小二乘损失,部分解决了生成图片质量不高及训练过程不稳定两个缺陷。整体生成结果表明 DCGAN 生成的图像更具多样性,特别是图像内部的纹理与细节更丰富。

3 GAN 在计算机视觉领域的应用

GAN 在计算机视觉的许多方面都表现非凡,从最初的图像生成,到后面的一系列应用,越来越多新的 GAN 框架被提出并应用到新的领域,由于 GAN 自身的对抗特性它能不断地自我提升,在生成样本领域取得了比传统方法更显著的效果。本节将介绍 GAN 在视觉上的应用及为了实现目标任务在结构上做出的改变。

3.1 生成高质量图像

GAN 最初的应用是在图像生成与建模上,无论以监督学习或无监督学习的方式,GAN 都能学习真实数据的分布。研究者一直致力于使得生成的图像更接近真实的图像,较成功的有 DCGAN、WGAN、STACKGAN[40]等模型。DCGAN 将 GAN 与深度 CNN 结合,对模型施加约束、提升训练技巧,使得 DCGAN 稳定性增加,在不同数据集上都取得了良好的生成结果,已经成为 GAN 模型的基准。此外,它的生成器能进行有趣的矢量算术加减,证明生成图片不是对数据库中的图片元素的记忆,而是特定过滤器已经学会绘制特定图像。由于一般数据集中图片数量非常多,以 ImageNet[41]为例,包含千万级别

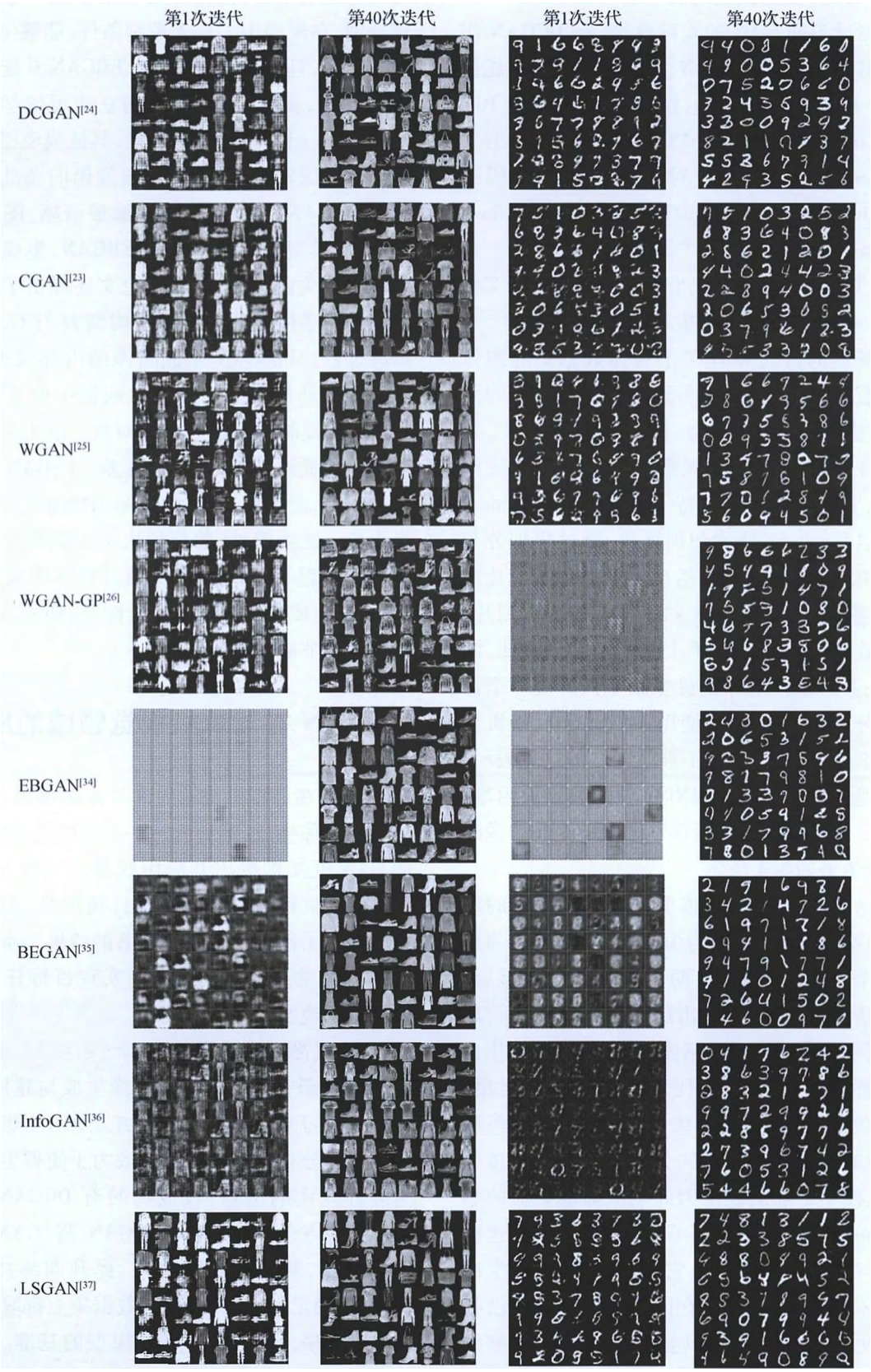

图 4　不同 GAN 模型生成图像结果

Fig. 4　Generated images obtained from different GAN models

的图像，图片的像素都很低，在这样的数据集上训练分类足够，但生成的图片不会清晰，如何提高生成图片的质量？文献 [42] 提出的拉普拉斯金字塔生成对抗网络(LAPGAN) 是一个串联网络，来源于同一张原始图的不同分辨率图像的集合按照金字塔从塔顶到塔底，图像分辨率越来越高。LAPGAN 采用这种原理先用低分辨率的样本生成低分辨率的图像，再将生成的低分辨率图像作为下一阶段输入的一部分和对应的高分辨率样本生成对应的高分辨率图像，每一个阶段的生成器都对应一个判别器，判断该阶段图像是生成还是真实的，其模型如图 5 所示。LAPGAN 的优点是每一个阶段的生成器都能学到不同的分布，传递到下一层作为补充信息，经过几次特征提取，最终生成图像的分辨率得到较大提升，生成结果更逼真。LAPGAN 除了采用以上方法，还结合了 CGAN，将无监督的方式转化为有监督的学习，效率提升明显。表 2 是 DCGAN 与 LAPGAN 模型的比较。

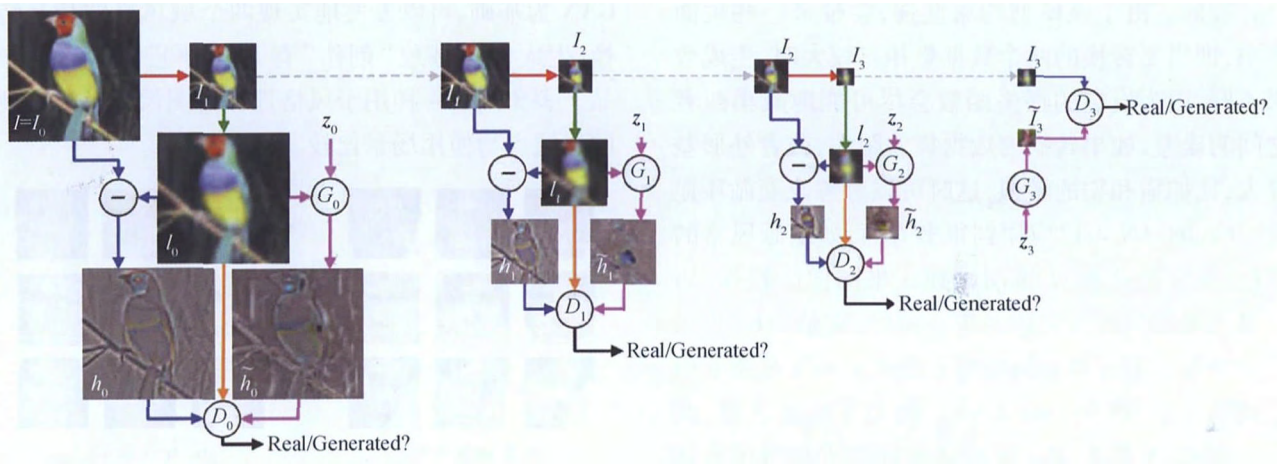

图 5　LAPGAN 的网络结构图

Fig. 5　Network structure of LAPGAN

表 2　DCGAN 与 LAPGAN 的对比

Table 2　Comparisons between DCGAN and LAPGAN

GAN 模型	改进	优点	缺点	适用场景
DCGAN [24]	将生成对抗网络与卷积神经网络结合；生成器能进行矢量算术加减	稳定训练过程；易收敛；生成样本多样性丰富；可以在无监督下学习	生成的图像分辨率低	适合大部分图形生成，是使用率最高的模型
LAPGAN [42]	上采样用 CGAN 生成拉普拉斯金字塔；下采样用高斯金字塔；用级联的卷积网络从低像素生成高像素图像	针对残差的逼近和学习较容易；逐级独立训练提高输入记忆样本的难度；增大 GAN 学习能力	必须在有监督下训练	需要生成高分辨率图像的场景

3.2　风格迁移与图像翻译

GAN 中一个有趣的应用是风格迁移，即把图像从一种风格转换到另一种风格。文献 [43] 中提出的 pix2pix 模型，它是一对一的图像风格迁移模型，使用两个数据集 *A* 和 *B*，如数据集 *A* 中是鞋子的轮廓图，数据集 *B* 中是与之对应的真实鞋子图像。该模型是以 CGAN 为基础，一个数据集中的图像作为输入，另一个作为条件输入，也叫生成目标，损失函数计算两者之间的误差，经过训练，给定一张图片就能生成另一种风格的图片。为了使生成的图片更加逼真，该模型做了以下优化：生成器使用 U-NET 架构 [44]，它是改进的 CNN 网络，在风格迁移中，有许多像素排列不变，因此 U-NET 结构除了正常的卷积池化操作，还有一部分直接传递到下一层，保证了图像内容不变；其次判别器使用卷积 PatchGAN 分类器，经过试验，使用局部分类器分类的结果比全局更好，参数的数量大规模缩减，提升了训练的速度和效率。在如此多的条件约束下，pix2pix 模型经过足够的训练，可以实现逼真的艺术风格转换。

同样是风格迁移的应用，文献[45]中提出的循环一致GAN（CycleGAN），打破了pix2pix模型数据集只能是成对图片的限制，论文提出的循环GAN，实现自我约束，通过对原域图像两步变换：先将其映射到目标域，再返回原域得到二次生成图像，从而消除了在目标域图像配对的要求，使用生成器网络将图像映射到目标域，通过匹配生成器与判别器，能提高生成图像的质量。将二次生成图像与原始图像对比，当二者分布一致，可以判断生成的目标域图像也是合理的。由于该模型约束性强，会带来一些负面影响，即当要转换的两个数据集相差较大时，生成效果不好，因为设定的损失函数会尽可能地减小两者之间的误差，如果转换的数据集差距大，或者外形差异大，比如猫和狗的转换，这时可以考虑改变循环损失。CycleGAN可以应用到很多方面，如绘画风格的转换，季节的迁移，2维图画到3维图像的转换，历史名人图像到真人的转换等。此外，文献[46]提出了一种基于生成对抗网络的方法来学习发现跨领域之间的关系，称为DiscoGAN。利用发现的关系，两个不同GAN耦合在一起形成的网络成功将风格从一个域迁移到另一个域，同时保留关键属性，如在保留面部主要特征的情况下，实现性别的转换（如图6所示）。文献[47]提出的域迁移网络（DTN）能够实现无监督的跨域图像生成。它采用复合损失函数，包括多种GAN损失和规范的组件能在保持实体原有身份的同时产生令人信服的以前没有的新形象。文献[48]提出耦合生成对抗网络，可以在没有任何对应图像元组情况下学习联合分布，能够在多领域实现图像变换。GAN在图像的风格迁移上有独特优势，源于GAN的两个网络能够相互制衡，相互“理解”。文献[49]提出以GAN为基础的无监督方法学习从一个域到另一个域的像素空间变换。以GAN为基础，可以方便地实现两个域风格的相互转换，生成目标域或“创作”有某一种艺术风格的作品。表3列出4种用于风格迁移与图像翻译的模型的优缺点与使用场景比较。

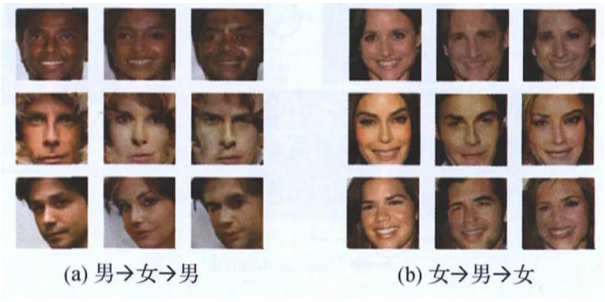

图6 DiscoGAN模型实现性别转换

Fig. 6 Gender transition by DiscoGAN model
((a) male→female→male; (b) female→male→female)

表3 用于风格迁移与图像翻译的不同GAN模型对比

Table 3 Comparison of different GAN models used in Style Transfer and ImageTranslation

GAN模型	改进	优点	缺点	适用场景
pix2pix[43]	采用CGAN模型；生成器采用U-NET架构；判别器使用PatchGAN分类器；实现像素级的图像迁移	参数大规模减少，提升训练速度和效率；生成的图像非常逼真	对数据集要求高，必须是一对一的成对数据集	成对图像之间的风格迁移生成，如不同风格的地图、实物与轮廓图等
CycleGAN[45]	提出循环GAN，实现自我约束，通过对原域图像两步变换返回原域，提高生成图像的质量。	对数据集要求低；只需提供包含两种风格的图像就能实现两种风格之间的转换	生成目标图像的质量比pix2pix低	大部分风格转换场景，如季节转换，艺术作品风格的迁移等
DiscoGAN[46]	DiscoGAN利用两个不同GAN耦合一起发现跨域关系；实现将风格从一个域迁移到另一个域，同时保留关键属性	实现一对一双向映射，避免模型崩溃，提升生成图像质量	对数据集要求高，必须是一对一的成对数据集	不同域之间图像的相互生成，如性别转换，包与鞋子相互转换、汽车与椅子相互转换等
DTN[47]	生成器网络包含一个输入函数和一个学习函数，并综合多项复合损失函数，解决了给定面部图像的表情符号生成问题	生成具有视觉吸引力的脸部表情符号，并且比人类创建的表情符号捕捉更多面部特征	由于输入函数的不对称性以及新源域中较低的信息内容，产生的结果缺乏吸引力	人物照片生成动漫图像

3.3 文本与图像的相互生成

GAN 能够在图像数据集上经过训练生成与真实分布相似的分布,如 GAN 在人脸数据集上经过训练,生成人分辨不出真假的人脸图像。这些都是直接学习特征再生成分布。一个有创造力的应用是根据输入的文本生成相应的图像,这比单纯的生成图像更有难度,因为涉及文本的语义理解。文献[50]提出的深层卷积结构和 GAN 框架,在文本和图像之间搭建了一个桥梁,将视觉概念从字符转换为像素。它展示了模型的功能,从详细的文字描述中产生鸟和花的合理图像。具体实现方法如下: 在生成网络中,高斯噪声同文本描述嵌到一起作为输入,经过学习,生成文本所描述的图像; 在判别网络中,进入其中的有 G 生成的假图像、真实的图像及对应的描述、真实的图像错误的描述,它们一同输入到判别网络中,随着判别器能力的增强,生成网络生成逼真图像的能力也随之增强。该论文生成的只是简单的描述性语句,一个物体加上一些修饰颜色或状态的词语。比如"有着重叠的粉红色尖花瓣并且围绕着黄色短花蕊的花"。比较好的改进有 StackGAN,不同于一般的 GAN 网络结构,它分为两个阶段生成图像,第一阶段生成图片比较粗糙,经过第二阶段生成分辨率更高的图像。经过对结构的改变或者增加更多的约束,GAN 可以生成更高质量的目标图片。若需要,可以添加更多阶段,生成有丰富细节和细腻纹理的图像,图 7 为生成"有着棕色喙,黑、白、棕色身体的鸟"。

从文本到图像的生成难度大且限制多,相反,根据给定图像生成文字描述比较容易,经过训练,机器能够简单描述图像内容,但如何让机器像人类一样看图说话,写出文章? 文献[51]提出一种半监督的段落生成框架,它通过对局部语义区域进行推理和利用语言知识合成语义连贯的段落描述。文中提出的循环主题转换生成对抗网络(RTT-GAN)构建了一个结构化段落生成器和多级段落识别器之间的对抗框架。段落发生器通过在每个步骤中引入基于区域的视觉和语言注意机制来循序地产生句子。表 4 是用于文本与图像的相互生成的模型比较。

图 7 StackGAN 分两个阶段生成一幅图像

Fig. 7 StackGAN generates an image in two phases((a) result of the first phase; (b) result of the second phase)

3.4 图像还原与修复

目前人脸识别的结果越来越精确,已经应用到地铁、火车站、机场等人群密集的场所来快速准确地识别行人,甚至抓到很多在逃的犯罪嫌疑人。然而,这种密集人群检测很困难,同一时刻,镜头里会有各种形态不同表情的人,特别是出现在镜头中的只有一个侧面或局部被遮挡,这样就不能根据已有面部信息辨别人的身份。如何利用科技手段,从局部得到整体信息,这是一个亟待解决的问题。文献[52]受人类视觉识别过程启发,结合 GAN 的强大性能,提出了双路径 GAN(TP-GAN),它能够考虑整体结构和局部信息,合成的图像逼真且保留了原有身份特征。使用不同角度的侧面照、或在不同的光照条件下、或保持不同的姿势,TP-GAN 都能根据已有信息合成人的正脸信息,合成的图像和真实图像非常接近。为了实现以上描述,TP-GAN 做了以下改变,它的生成网络有两条路径,一条专注于推理全局结构,另一条则推理局部的纹理,分别得到两个特征地图,将两个特征地图融合在一起,用于最终合成。合成的正面视图和真实的相片进入判别器进行判断。不仅如此,还将正面人脸的分布信息并入一个 GAN,由此对恢复过程进行了很好的约束。除此之

表 4　用于文本与图像相互转化的不同 GAN 模型对比

Table 4　Comparison of different GAN models used in Text and Image Transforming to each other

GAN 模型	改进	优点	缺点	适用场景
文献[50]	改进了判别器的输入，让其学习文本描述与图片内容的对应关系；通过插值方法生成大量新的文本描述；逆向的风格迁移分析	实现了从文本描述到图像的生成；简单场景下能够生成逼真的图像表示	目前只能生成简单的描述，当描述复杂时，生成的图像不清晰，需要经过大量训练	简单的文本到图像的生成，如花的生成、鸟的生成
StackGAN[40]	将两个 GAN 叠加在一起，第一个 GAN 根据文本描述生成相对粗糙的图像，第 2 个 GAN 修正之前生成的图像并添加细节	通过分阶段生成，最终生成的图像清晰度提高	将复杂的生成任务分成两个阶段，可能出现每个任务找不到重点，导致生成任务失败	用于从文本到清晰图像的生成
RTT-GAN[51]	RTT-GAN 构建了一个结构化段落生成器和多级段落识别器之间的对抗框架。通过在每个步骤中引入基于区域的视觉和语言注意机制来循序的产生句子。	在半监督的条件下生成段落，通过对局部语义区域进行推理和利用语言知识来合成各种语义连贯的段落描述	在无监督条件下无法工作	让机器像人类一样看图说话，写出文章

外，TP-GAN 组合多种 loss，合成缺失部分，以保留面部突出特征。

GAN 在图像修复领域有广泛的应用。文献[53]提出一种新颖的语义图像修复方法，通过调整可用数据生成缺失内容。文中使用上下文和先前的损失来搜索潜在图像中损坏图像的最接近编码，然后将该编码通过生成模型来推断丢失的内容。该方法成功地预测了大量缺失区域的信息，并实现了像素级的逼真度。文献[54]提出了一种基于上下文的像素预测驱动的无监督视觉特征学习算法。文中提出上下文编码器，它是经过训练的卷积神经网络，能生成以其周围环境为条件的任意图像区域的内容。训练上下文编码器时，使用标准像素重建损失加上对抗性损失，能够补全图像并产生更清晰的结果。它指出上下文编码器在学习时，不仅捕获了外观，而且捕获了视觉结构的语义，此外还可用于语义修复任务。图 8 为该模型在不同图像上的修复结果。文献[55]提出一种使用深度生成模型的面部补全算法。它基于神经网络直接生成缺失区域的内容，通过引入重建损失，两个对抗性损失和语义解析损失的组合进行训练，确保了像素忠实度和局部全局内容的一致性。它能处理任意形状的大面积缺失像素，并产生逼真的面部完成结果。

除了修复局部图像，GAN 在超分辨率重建上也取得了显著成果，超分辨率重建即从低分辨率图像得到高分辨率图像，是图像反模糊化的过程。文献[56]提出的 SRGAN 使用 GAN 完成图像的超分辨率重建，将下采样失真的图像恢复如照片一样逼真，

(a) 局部图像的修复结果　(b) 整体图像的修复结果　(c) 复杂场景图像的修复结果

图 8　不同图像修复结果

Fig. 8　Results of different image inpainting ((a) local images; (b) whole images; (c) complex scene images)

为了实现在大规模放大图像时纹理细节更逼真,它提出了一个包含对抗性损失和内容损失的感知损失函数。对抗性损失将问题的解决方案推向自然流形图像,使用判别网络进行训练,以区分超分辨图像和原始照片图像,另外文中使用感知相似性驱动的内容损失而不是像素空间的相似性来使纹理更加丰富。文献[57]提出一种基于条件 GAN 和内容损失的端到端学习模型 DeblurGAN,它可以处理由相机抖动和因物体运动而产生的模糊。该网络架构获得了动态去模糊的最新技术成果,它的生成器网络将模糊图像作为输入并产生对清晰图像的估计,在训练期间,判别网络将生成图像和清晰图像作为输入并估计它们之间的距离,根据 VGG-19 在真实图像和恢复图像的特征图之间激活差异,总损失包括来自判别器的损失和感知 WGAN 损失,生成的图像更加清晰。表 5 是用于图像修复与超分辨率的不同模型的对比。

GAN 在 2 维数据分布建模方面的卓越性能极大地改善了很多不合理的低级视觉问题,通过与编码器、CNN、上下文语义以及组合多种损失使生成的图像更加的逼真。相比于其他的生成模型更加的灵活,效果更好。

表 5　用于图像还原与修复的 GAN 模型对比

Table 5　Comparison of GAN models used inImage Inpainting and Restoration

GAN 模型	改进	优点	缺点	适用场景
TP-GAN[53]	将从数据分布得来的先验知识和人脸领域知识结合,提出双路径 GAN,一条专注于推理全局结构,另一条则推理局部的纹理	根据单一的图像合成正面人脸视图,合成的图像非常逼真且很好地保留了身份特征,能应对大量不同的姿势	当旋转的角度过大时,生成的面部细节与真实照片存在差异	应用在人脸分析的工作或者需要通过侧脸鉴定身份信息的场景,如寻找嫌疑人等
文献[53]	使用上下文和先前的损失来搜索潜在图像中损坏图像的最接近编码,然后将该编码通过生成模型来推断丢失的内容	不需要伪装训练,生成图像比较尖锐,实现了像素级的逼真度	大面积缺失的情况下,生成结果不真实	通过调整可用数据生成缺失内容,如恢复被遮挡的部分
文献[54]	提出上下文编码器;使用标准像素重建损失加上对抗性损失,能够补全图像并产生更清晰的结果	上下文编码器在学习时捕获外观与视觉结构的语义,此外还可用于语义修复任务	由于在无监督条件下训练,生成结果没有在监督训练情况下真实	用于上下文的像素预测驱动的无监督视觉特征学习
文献[55]	为了确保像素忠实度和局部全局内容的一致性,引入重建损失,两个对抗性损失和语义解析损失的组合进行训练	能处理任意形状的大面积缺失像素,并产生逼真的面部完成结果	损失函数太多,每一个损失的权值选择较困难	面部补全,它可以直接生成缺失区域的内容
SRGAN[56]	使用对抗性损失和内容损失的感知损失函数,感知相似性驱动的内容损失生成的纹理更加丰富	能够生成比较接近原图的清晰图像,而且内部的纹理放大后仍然丰富	生成的图像在视觉上比较清晰,但是放大足够倍数会出现许多不存在的纹路	用于需要提升分辨率,且对像素重建后的品质要求较高的场景
DeblurGAN[57]	提出一种基于条件 GAN 和内容损失的端对端学习模型,总损失包括来自批评者和感知损失的 WGAN 损失,使图像更加清晰	它可以处理由相机抖动和因物体运动而产生的模糊	目前只能一定程度的改善运动产生的模糊效果	用于改善由于相机抖动或物体快速移动产生的模糊图像

3.5　其他应用

GAN 除了应用在图像生成、图像修复与还原、风格迁移与图像翻译等领域外,其在计算机视觉的其他领域也表现出巨大的潜力,如使用模拟与无监督的学习方法,输入合成图像,经判别器鉴别,以提升合成图像质量,这也是一个有前景的方向[58];再者如视频预测[59],能够合理预测下一帧发生场景的具有时空卷积架构的视频生成对抗网络[60];能够分解运动和内容的 GAN(MoCoGAN)[61],它将随机噪声向量依次映射到视频帧来生成视频剪辑,实现未来帧预测;GAN 对象检测,文献[62]提出的感知生成对抗网络,通过缩小小对象与大对象的表示差异

改善小对象检测; GAN 能够生成时间序列,如音乐生成[63]、重症监护室的 ICU 记录生成[64]、电子健康记录生成[65]等; GAN 通过 3D 对抗建模学习物体形状的潜在概率空间[66],使用生成网络将真实图像合成新型 3D 视图[67]; GAN 检测多光谱影像变化[68]; 使用 GAN 在有限的训练数据中生成逼真的结果[69]; 同时 GAN 在医疗影像分割[70],自动驾驶等[71]都取得了良好的实践效果。

4 结 语

4.1 总结

GAN 作为一种生成模型,对于解决样本不足、生成质量差、提取特征难度大等问题提供了一种较好的解决方案。对基于深度学习的生成对抗网络在计算机视觉方面的应用进行了分析总结,不仅深入分析了 GAN 在理论模型方面的改进,而且重点介绍了 GAN 在视觉方面的几类突出的应用,并通过实验验证了不同 GAN 算法的优缺点及适用应用场景。GAN 本身是一个有包容性的框架,它可以和许多深度学习模型结合起来,解决传统机器学习模型所不能解决的问题。

4.2 发展趋势

目前,GAN 虽然仍存在一些理论上亟待解决的问题,但在实践上获得巨大成功。GAN 不仅是实现无监督学习的途径之一,它与监督学习、半监督学习的结合能加速训练,通过对生成器添加指定生成目标或加入语义控制条件等指导最终的生成结果,实现对模型的深度控制,弥补 GAN 本身的不足。未来,GAN 在以下方面将取得进一步发展:

1) 理论突破。GAN 提出的对抗博弈思想为生成模型提供了崭新的思路,其核心思想是通过对抗训练学习真实分布的特征,模仿并生成与真实分布相同的分布。已有许多 GAN 模型被提出以解决不同领域的问题,但它们都面临同样的困境,即由于 GAN 自身理论的不完善,生成样本的质量有待提高。为此,从理论层面取得突破,解决 GAN 自身不收敛、模型崩溃、训练困难等问题,找出导致以上问题的根本原因并改进是未来研究的重要方向之一。

2) 算法拓展。进一步拓展 GAN 算法的应用范围,吸收机器学习中最新的理论与研究成果并与之相结合,如 GAN 与强化学习结合,解决 GAN 处理离散变量时效果不佳的弱点,即利用强化学习中的策略梯度算法,使 GAN 可以用于离散的场景,进一步增强 GAN 适用范围; GAN 与对抗样本应用于解决深度学习系统的安全问题,GAN 可以生成加入不同噪音的样本,进一步研究有利于抵抗对抗样本的算法,增强深度学习系统的鲁棒性等。

3) 评估体系。GAN 模型的评估与比较缺乏科学、统一的标准。GAN 作为新的生成模型,目前还没有相关的指标能够从性能、准确率、过拟合程度、生成样本的视觉质量等方面对不同的模型综合评估,因此提出一个更精确的评价指标,采用统一的标准,构建标准化、通用化的科学评估体系是亟待解决的问题。

4) 专用系统。应用 GAN 解决更加具体的计算机视觉应用问题,即从目前解决一类问题向解决具体实际应用问题转变,针对该问题在已有基础上设计更具针对性的方案并开发专用的深度学习系统,如生成特定场景系统或者提升图片特定部分分辨率的专门系统,或与游戏结合,直接生成完整的游戏场景与人物,进一步生成高质量的视觉场景等。

5) 行业融合。GAN 与某些特殊行业的交叉融合,有利于生成不易获取的样本数据并作为真实数据的补充,如当有关医学的数据集不足时,在已有数据的基础上生成更多相似样本等。从人工智能长远发展来看,利用 GAN 提升机器理解世界的能力,让机器拥有“意识”是值得研究的问题。

参考文献(References)

[1] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. arXiv preprint arXiv: 1406.2661, 2014.

[2] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507. [DOI: 10.1126/science.1127647]

[3] Sonka M, Hlavac V, Boyle R. Image Processing, Analysis and Machine Vision[M]. Boston, MA: Springer, 1993. [DOI: 10.1007/978-1-4899-3216-7]

[4] Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 702-716. [DOI: 10.1007/978-3-319-46487-9_43]

[5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the

25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2012: 1097-1105.

[6] Cappelli R, Erol A, Maio D, et al. Synthetic fingerprint-image generation [C] //Proceedings of the 15th International Conference on Pattern Recognition. Barcelona, Spain: IEEE, 2000: 471-474. [DOI: 10.1109/icpr.2000.903586]

[7] Szeliski R. Computer Vision: Algorithms and Applications [M]. London: Springer, 2011. [DOI: 10.1007/978-1-84882-935-0]

[8] Baird H S, Coates A L, Fateman R J. PessimalPrint: a reverse Turing test [J]. International Journal on Document Analysis and Recognition, 2003, 5(2-3): 158-163. [DOI: 10.1007/s10032-002-0089-1]

[9] Owechko Y. Specific emitter identification using histogram of oriented gradient features: US, US 20100061630 A1 [P]. 2010-03-11.

[10] Choi J Y, Sung K S, Yang Y K. Multiple vehicles detection and tracking based on scale-invariant feature transform [C] //Proceedings of 2007 IEEE Intelligent Transportation Systems Conference. Seattle, WA, USA: IEEE, 2007: 528-533. [DOI: 10.1109/itsc.2007.4357684]

[11] Mo J, Walrand J. Fair end-to-end window-based congestion control [J]. IEEE/ACM Transactions on Networking, 2000, 8(5): 556-567. [DOI: 10.1109/90.879343]

[12] Pu Y C, Gan Z, Henao R, et al. Variational autoencoder for deep learning of images, labels and captions [EB/OL]. 2016-09-28 [2017-11-13]. https://arxiv.org/pdf/1609.08976.pdf.

[13] Kingma D P, Welling M. Auto-encoding variational Bayes [EB/OL]. 2014-05-01 [2017-11-19]. https://arxiv.org/pdf/1312.6114.pdf.

[14] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551. [DOI: 10.1162/neco.1989.1.4.541]

[15] LeCun Y, Bottou L, Bengio Y, et al. Gradient-Based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324. [DOI: 10.1109/5.726791]

[16] Graves A, Fernández S, Schmidhuber J. Multi-dimensional recurrent neural networks [C] //Proceedings of the 17th International Conference on Artificial Neural Networks. Porto, Portugal: Springer, 2007: 549-558. [DOI: 10.1007/978-3-540-74690-4_56]

[17] Graves A. Long short-term memory [M] //Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer, 2012: 1735-1780. [DOI: 10.1007/978-3-642-24797-2_4]

[18] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. Cambridge: The MIT Press, 2016: 26-29.

[19] He D, Chen W, Wang L W, et al. A game-heoretic machine learning approach for revenue maximization in sponsored search [C] //Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. Beijing, China: AAAI, 2013: 206-212.

[20] Ratliff L J, Burden S A, Sastry S S. Characterization and computation of local Nash equilibria in continuous games [C] //Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing. Monticello, IL, USA: IEEE, 2013: 917-924. [DOI: 10.1109/allerton.2013.6736623]

[21] Goodfellow I, NIPS 2016 tutorial: generative adversarial networks [EB/OL]. 2017-04-03 [2018-03-01]. https://arxiv.org/pdf/1701.00160.pdf.

[22] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs [EB/OL]. 2016-06-10 [2017-12-05]. https://arxiv.org/pdf/1606.03498.pdf.

[23] Mirza M, Osindero S. Conditional generative adversarial nets [J]. arXiv preprint arXiv: 1411.1784, 2014.

[24] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. 2016-12-20 [2018-02-28]. https://arxiv.org/pdf/1511.06434.pdf.

[25] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [EB/OL]. 2017-12-06 [2018-02-23]. https://arxiv.org/pdf/1701.07875.pdf.

[26] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs [EB/OL]. 2017-12-25 [2018-01-19]. https://arxiv.org/pdf/1704.00028.pdf.

[27] Zeiler M D, Taylor G W, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning [C] //Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2018-2025. [DOI: 10.1109/iccv.2011.6126474]

[28] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C] //Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015: 448-456.

[29] Kingma D P, Ba J. Adam: a method for stochastic optimization [J]. arXiv preprint arXiv: 1412.6980, 2014.

[30] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines [C] //Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, Israel: Omnipress, 2010: 807-814.

[31] Xu B, Wang N Y, Chen T Q, et al. Empirical evaluation of rectified activations in convolutional network [EB/OL]. 2015-11-27 [2018-04-28]. https://arxiv.org/pdf/1505.00853.pdf.

[32] Wang K F, Gou C, Duan Y J, et al. Gener ative adversarial networks: the state of the art and beyond [J]. Acta Automatica Sinica, 2017, 43(3): 321-332. [王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望 [J]. 自动化学报, 2017, 43(3): 321-332.] [DOI: 10.16383/j.aas.2017.

y000003]

[33] 郑华滨. 令人拍案叫绝的 Wasserstein GAN [EB/OL]. 2017-04-02 [2018-01-20]. https://zhuanlan.zhihu.com/p/25071913.

[34] Zhao J B, Mathieu M, LeCun Y. Energy-based generative adversarial network [EB/OL]. 2017-03-06 [2017-12-23]. https://arxiv.org/pdf/1609.03126.pdf.

[35] Berthelot D, Schumm T, Metz L. BEGAN: boundary equilibrium generative adversarial networks [EB/OL]. 2017-05-31 [2018-01-08]. https://arxiv.org/pdf/1703.10717.pdf.

[36] Chen X, Duan Y, Houthooft R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [EB/OL]. 2016-06-12 [2017-12-27]. https://arxiv.org/pdf/1606.03657.pdf.

[37] Mao X D, Li Q, Xie H R, et al. Least squares generative adversarial networks [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2018: 2813-2821. [DOI: 10.1109/ICCV.2017.304]

[38] Hwalsuklee. Tensorflow-generative-model-collections [EB/OL]. 2017-09-06 [2017-10-19]. https://github.com/hwalsuklee/tensorflow-generative-model-collections.

[39] Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms [EB/OL]. 2017-09-15 [2018-02-23]. https://arxiv.org/pdf/1708.07747.pdf.

[40] Zhang H, Xu T, Li H S, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 5908-5916. [DOI: 10.1109/iccv.2017.629]

[41] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 248-255. [DOI: 10.1109/CVPR.2009.5206848]

[42] Denton E, Chintala S, Szlam A, et al. Deep generative image models using a Laplacian pyramid of adversarial networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 1486-1494.

[43] Isola P. Zhu J Y, Zhou T H, et al. Image-to-Image translation with conditional adversarial networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 5967-5976. [DOI: 10.1109/CVPR.2017.632]

[44] Ronneberger O. Invited Talk: U-Net convolutional networks for biomedical image segmentation [M]//Maier-Hein K H, Fritzsche G, Deserno T M, et al. Bildverarbeitung für die Medizin 2017. Berlin, Heidelberg: Springer, 2017. [DOI: 10.1007/978-3-662-54345-0_3]

[45] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image translation using cycle-consistent adversarial networks [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2242-2251. [DOI: 10.1109/iccv.2017.244]

[46] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017.

[47] Taigman Y, Polyak A, Wolf L. Unsupervised cross-domain image generation [EB/OL]. 2016-11-07 [2017-12-24]. https://arxiv.org/pdf/1611.02200.pdf.

[48] Liu M Y, Tuzel O. Coupled generative adversarial networks [EB/OL]. 2016-09-20 [2018-01-12]. https://arxiv.org/pdf/1606.07536.pdf.

[49] Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptationwith generative adversarial networks [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 95-104. [DOI: 10.1109/CVPR.2017.18]

[50] Reed S, Akata Z, Yan X C, et al. Generative adversarial text to image synthesis [C]//Proceedings of the 33rd International Conference on Machine Learning. New York, USA: JML, 2016: 1060-1069.

[51] Liang X D, Hu Z T, Zhang H, et al. Recurrent topic-transition GAN for visual paragraph generation [EB/OL]. 2017-03-23 [2018-01-08]. https://arxiv.org/pdf/1703.07022.pdf.

[52] Huang R, Zhang S, Li T Y, et al. Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2458-2467. [DOI: 10.1109/ICCV.2017.267]

[53] Yeh R A, Chen C, Lim T Y, et al. Semantic image inpainting with deep generative models [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6882-6890. [DOI: 10.1109/CVPR.2017.728]

[54] Pathak D, Krähenbühl P, Donahue J, et al. Context encoders: feature learning by inpainting [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 2536-2544. [DOI: 10.1109/CVPR.2016.278]

[55] Li Y J, Liu S F, Yang J M, et al. Generative face completion [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 5892-5900. [DOI: 10.1109/CVPR.2017.624]

[56] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//Pro-

ceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 105-114. [DOI: 10.1109/CVPR.2017.19]

[57] Kupyn O, Budzan V, Mykhailych M, et al. DeblurGAN: blind motion deblurring using conditional adversarial networks [EB/OL]. 2018-04-03 [2018-03-27]. https://arxiv.org/pdf/1711.07064.pdf.

[58] Shrivastava A, Pfister T, Tuzel O, et al. Learning from simulated and unsupervised images through adversarial training [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2242-2251. [DOI: 10.1109/CVPR.2017.241]

[59] Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error [EB/OL]. 2016-02-26 [2017-12-19]. https://arxiv.org/pdf/1511.05440.pdf.

[60] Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics [EB/OL]. 2016-10-26 [2017-12-23]. https://arxiv.org/pdf/1609.02612.pdf.

[61] Tulyakov S, Liu M Y, Yang X D, et al. MoCoGAN: decomposing motion and content for video generation [EB/OL]. 2017-12-14 [2018-02-17]. https://arxiv.org/abs/1707.04993.

[62] Li J A, Liang X D, Wei Y C, et al. Perceptual generative adversarial networks for small object detection [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 1951-1959. [DOI: 10.1109/CVPR.2017.211]

[63] Yang L C, Chou S Y, Yang Y H. MidiNet: a convolutional generative adversarial network for symbolic-domain music generation [EB/OL]. 2017-07-18 [2017-12-22]. https://arxiv.org/pdf/1703.10847.pdf.

[64] Esteban C, Hyland S L, Rätsch G. Real-valued (Medical) time series generation with recurrent conditional GANs [EB/OL]. 2017-12-04 [2018-02-04]. https://arxiv.org/pdf/1706.02633.pdf.

[65] Choi E, Biswal S, Malin B, et al. Generating Multi-label discrete electronic health records using generative adversarial network [EB/OL]. 2018-01-11 [2018-01-28]. https://arxiv.org/pdf/1703.06490v1.pdf.

[66] Wu J J, Zhang C K, Xue T F, et al. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling [EB/OL]. 2017-01-04 [2017-12-26]. https://arxiv.org/abs/1610.07584.pdf.

[67] Park E, Yang J M, Yumer E, et al. Transformation-grounded image generation network for novel 3D view synthesis [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 702-711. [DOI: 10.1109/cvpr.2017.82]

[68] Gong M G, Niu X D, Zhang P Z, et al. Generative adversarial networks for change detection in multispectral imagery [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(12): 2310-2314. [DOI: 10.1109/lgrs.2017.2762694]

[69] Gurumurthy S, Sarvadevabhatla R K, Babu R V. DeLiGAN: generative adversarial networks for diverse and limited data [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 4941-4949. [DOI: 10.1109/CVPR.2017.525]

[70] Xue Y, Xu T, Zhang H, et al. SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation [EB/OL]. 2017-07-16 [2018-01-26]. https://arxiv.org/pdf/1706.01805.pdf.

[71] Ghosh A, Bhattacharya B, Chowdhury S B R. SAD-GAN: synthetic autonomous driving using generative adversarial networks [EB/OL]. 2016-11-27 [2018-01-29]. https://arxiv.org/pdf/1611.08788.pdf.

作者简介

曹仰杰,1976 年生,男,副教授,硕士生导师,主要研究方向为机器学习与视觉计算,高性能计算。
E-mail: caoyj@zzu.edu.cn

林楠,通信作者,女,副教授,硕士生导师,主要研究方向为物联网与智能计算、计算机视觉与深度学习。
E-mail: linnan@zzu.edu.cn

贾丽丽,女,硕士研究生,主要研究方向为深度学习与计算机视觉。E-mail: jialilics@163.com

陈永霞,女,讲师,主要研究方向为物联网,智能计算和人工智能。E-mail: rjchenyx@zzu.edu.cn

李学相,男,教授,硕士生导师,主要研究方向为高性能计算,云计算和人工智能。E-mail: lxx@zzu.edu.cn