生成对抗网络在各领域应用研究进展

刘建伟 [1] 谢浩杰 [1] 罗雄麟 [1]

摘 要 随着深度学习的快速发展, 生成式模型领域也取得了显著进展. 生成对抗网络 (Generative adversarial network, GAN) 是一种无监督的学习方法, 它是根据博弈论中的二人零和博弈理论提出的. GAN 具有一个生成器网络和一个判别器网络, 并通过对抗学习进行训练. 近年来, GAN 成为一个炙手可热的研究方向. GAN 不仅在图像领域取得了不错的成绩, 还在自然语言处理 (Natural language processing, NLP) 以及其他领域崭露头角. 本文对 GAN 的基本原理、训练过程和传统 GAN 存在的问题进行了阐述, 进一步详细介绍了通过损失函数的修改、网络结构的变化以及两者结合的手段提出的 GAN 变种模型的原理结构, 其中包括: 条件生成对抗网络 (Conditional GAN, CGAN)、基于 Wasserstein 距离的生成对抗网络 (Wasserstein-GAN, WGAN) 及其基于梯度策略的 WGAN (WGAN-gradient penalty, WGAN-GP)、基于互信息理论的生成对抗网络 (Informational-GAN, InfoGAN)、序列生成对抗网络 (Sequence GAN, SeqGAN)、Pix2Pix、循环一致生成对抗网络 (Cycle-consistent GAN, Cycle GAN) 及其增强 Cycle-GAN (Augmented CycleGAN). 概述了在计算机视觉、语音与 NLP 领域中基于 GAN 和相应 GAN 变种模型的基本原理结构, 其中包括: 基于 CGAN 的脸部老化应用 (Face aging CGAN, Age-cGAN)、双路径生成对抗网络 (Two-pathway GAN, TP-GAN)、表示解析学习生成对抗网络 (Disentangled representation learning GAN, DR-GAN)、对偶学习生成对抗网络 (DualGAN)、GeneGAN、语音增强生成对抗网络 (Speech enhancement GAN, SEGAN) 等. 介绍了 GAN 在医学、数据增强等领域的应用情况, 其中包括: 数据增强生成对抗网络 (Data augmentation GAN, DAGAN)、医学生成对抗网络 (Medical GAN, MedGAN)、无监督像素级域自适应方法 (Unsupervised pixel-level domain adaptation method, PixelDA). 最后对 GAN 未来发展趋势及方向进行了展望.

关键词 生成对抗网络, 对抗学习, 自然语言处理, 计算机视觉, 零和博弈, 语音合成与分析

引用格式 刘建伟, 谢浩杰, 罗雄麟. 生成对抗网络在各领域应用研究进展. 自动化学报, 2020, **46**(12): 2500−2536

DOI 10.16383/j.aas.c180831

Research Progress on Application of Generative Adversarial Networks in Various Fields

LIU Jian-Wei[1] XIE Hao-Jie[1] LUO Xiong-Lin[1]

Abstract With the rapid development of deep learning, the field of generative models has also made significant progress. Generative adversarial network (GAN) is an unsupervised learning method based on the zero-sum game theory in game theory. GAN has a generator network and a discriminator network and trains through adversarial learning. In the past two years, GAN has become a hot research direction. GAN has not only achieved good results in the field of computer vision, but also emerged in natural language processing (NLP) and other fields. This paper expounds the basic principles of GAN, the training process and the problems existing in traditional GAN, and further introduces the principal structure of the GAN variant model proposed by the modification of the loss function, the change of the network structure and the combination of the two, e.g., conditional GAN (CGAN), Wasserstein-GAN (WGAN), WGAN-gradient penalty (WGAN-GP), informational GAN (InfoGAN), sequence GAN (SeqGAN), Pix2Pix, cycle-consistent GAN (CycleGAN) and augmented CycleGAN, and so on. Then in the areas of computer vision, speech synthetics and analysis and NLP, we review the structure of the principle networks and models, including Age-cGAN for face aging, two-pathway GAN (TP-GAN), disentangled representation learning GAN (DR-GAN), DualGAN, GeneGAN, speech enhancement GAN (SEGAN), gumbel-softmax GAN, and so forth. Then we also introduce the applications of GAN in the field of medicine, data enhancement,etc, including data augmentation GAN (DAGAN), medical GAN (MedGAN), unsupervised pixel-level domain adaptation method (PixelDA), and so on. Finally, the future trends and directions of GAN are prospected.

Key words Generative adversarial network (GAN), adverasrial learning, natural language processing (NLP), com-

收稿日期 2018-12-13 录用日期 2019-06-06
Manuscript received December 13, 2018; accepted June 6, 2019
国家自然科学基金 (21676295), 中国石油大学 (北京) 2018 年度前瞻导向及培育项目 “神经网络深度学习理论框架和分析方法及工具” (2462018QZDX02) 资助
Supported by National Natural Science Foundation of China (21676295) and Science Foundation of China University of Petroleum Beijing (2462018QZDX02)
本文责任编委 黎铭
Recommended by Associate Editor LI Ming
1. 中国石油大学 (北京) 自动化研究所 北京 102249
1. Research Institute of Automation, China University of Petroleum (Beijing), Beijing 102249

puter vision, zero-sum game, speech synthetics and analysis

Citation Liu Jian-Wei, Xie Hao-Jie, Luo Xiong-Lin. Research progress on application of generative adversarial networks in various fields. *Acta Automatica Sinica*, 2020, **46**(12): 2500−2536

自 2012 年以来, 深度学习的快速发展使得人工智能研究得到飞速进步. 当今, 人工智能发展正处于快速上升时期, 大量研究人员将精力以及资本投入到人工智能领域. 人工智能的发展是有目共睹的, 从无人机走进人们生活, 到 Goolge 人工智能围棋程序 AlphaGo 打败人类顶级选手, 无不证明了深度学习近年来的迅速发展. 从 AlphaGo 发展历程可以看出, 自 2016 年以来, 它的目标对手早已不是人类顶级选手, 而是与之前自己的版本进行较量, 开辟属于它的全新领域. AlphaGo 使用蒙特卡洛树搜索 (Monte Carlo tree search), 借助估值网络 (value network) 与策略网络 (policy network) 这两种深度神经网络来评估选点和选择落点[1].

此外, 深度学习的发展受神经网络的制约, 神经网络可以说是深度学习的灵魂, 其广泛的应用场景使得深度学习研究的深度和广度都得到了空前的提高. 本综述的生成对抗网络, 不论是生成器还是判别器均采用了神经网络, 并且在提及的多个应用领域中都将大量采用神经网络. 在过去的数年中, 神经网络的研究在图像、语音识别、自然语言处理领域等都取得了令人瞩目的成果. 但是神经网络也有参数多、训练难的特点, 其相应的改进也是层出不穷. 并且随着计算能力的飞速提升, 神经网络能够更快地训练更多的参数.

在生成式模型中, 生成对抗网络 (Generative adversarial network, GAN)[2] 是一类特殊的存在. 它的提出不仅使各个领域的发展达到新的高度, 更是促使人工智能领域走向了一个具有 “思想”的时代. 可以说, GAN 就是 “做梦”, 因为在自然界当中, 只有哺乳类才会做梦, 这就是 GAN 在人工智能 (Artificial intelligence, AI) 领域的份量. GAN 可以说就是一个具有对抗思想的网络结构. 尽管 GAN 的变种模型层出不穷, 并且用途广泛, 但是其核心一直没有发生变化, 即对抗思想一直没有发生变化. 关于对抗思想的介绍可参照王坤峰等[3] 提出的对抗思想, 即在博弈、竞争中包含着对抗的思想. GAN 的对抗思想就是在生成数据的过程中加入一个可以判断真实数据和生成数据的判别器, 使生成器 (Generator) 和判别器 (Discriminator) 相互对抗, 判别器的作用是努力地分辨真实数据和生成数据, 生成器的作用是努力改进自己从而生成可以迷惑判别器的数据. 当判别器无法再分别出真假数据, 则认为此时的生成器已经达到了一个不错的生成效果. 这种 GAN 对抗思想的提出可以说对生成式模型的发展具有重要的意义.

GAN 是一个无监督生成式模型. 模型主要分为两类, 一类是生成式模型, 另一类是判别式模型. 生成式模型会对 x 和 y 的联合分布 $p(x,y)$ 进行建模, 通过贝叶斯公式来求得 y 的条件后验概率 $p(y|x)$, 最后选择使 $p(y|x)$ 取得最大值的 y_i 作为模型的输出. 而判别式模型则会直接给出 $p(y|x)$ 的表达式. 二者之间存在的差异如下:

1) 生成式模型会对数据的分布做出一定的假设, 并且只有在满足这些假设时, 它才能在这些服从假设概率分布的数据上得到不错的效果. 若假设不成立, 则判别式模型将会有更好的学习效果.

2) 若需要对类别进行更新, 生成式模型只需要对新的 x 和 y 的联合概率分布 $p(x,y)$ 计算即可, 而判别式模型则需要对整个 $p(y|x)$ 进行重新训练.

3) 在对错误率进行分析方面, 生成式模型最终得到的错误率将比判别式模型的错误率更高, 但是生成式模型的抽样复杂性较低, 只需要很少的样本就可以使错误率收敛.

4) 对于无标签的数据, 生成式模型 (例如: 深度信念网络 (Deep belief network, DBN)) 能更好地利用数据本身所包含的信息.

5) 判别式模型通常需要解决凸优化问题.

以上是对生成式模型进行的简单分析, 下面对生成式模型进行讨论. 生成式模型主要分为变分自编码 (Variational auto-encoder, VAE) 和 GAN.

首先, VAE[4] 是基于变分思想的深度学习的生成式模型. 假设 x 为随机变量, z 为隐变量. VAE 提出了变分下界的概念, 通过变分函数 $q(z)$ 来对后验概率 $p(z|x)$ 进行替换, 并用 KL 散度度量两者的近似程度. 这样能简化在面对大规模复杂数据时的难求解问题. VAE 的好处在于它能很好地针对图像的特征进行建模.

与 VAE 相比, GAN 没有使用变分下界, 如果判别器训练良好, 那么生成器可以完美地学习到训练样本的概率分布. 换句话说, GAN 是渐进一致的, 而 VAE 是有偏估计. GAN 顾名思义包含了两个网络子模型, 生成器和判别器. 这里可以将两个网络分别比作造假币的罪犯 (生成器) 和警察 (判别器). 罪犯的任务是生成足够逼真的假钞来欺骗警察, 让

警察以为假钞就是真钞; 而警察的任务是判别钞票的真假. 最终警察将无法区别真钞和假钞. 生成器和判别器最终优化目标是达到纳什均衡[5].

既然两个生成模型都具有各自的优点, 若将 VAE 与 GAN 相结合, 那么 GAN 能够生成质量很好的图片, 特征明显且清晰. 而 VAE 则是将原始图片重构, 在编码器的作用下编码生成隐向量, 这个向量能够在服从高斯分布的情况下, 保留原图像的特征. VAE-GAN[6] 的提出实现了这个思想, 这样就可以使用 GAN 的判别器学习特征表示, VAE 为重构目标提供帮助. 其结构如图 1 所示. 这样做的好处在于 VAE + GAN 能够在生成高质量图像的同时保持模型的稳定.

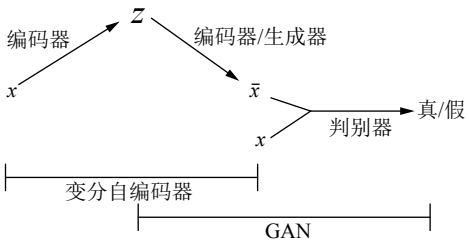


图 1 VAE + GAN 结构

Fig. 1 The structure of VAE + GAN

到目前为止, GAN 的主要应用场景集中在三大领域. 在图像处理领域, 例如: 在人脸识别和合成、图像超分辨率和图像转换等方面都取得不错的成绩; 在语音处理领域, GAN 也有了一定的发展, 例如: 语音增强和语音识别等; 此外, GAN 在自然语言处理领域也有一定的进展, 例如: 机器翻译、双语字典和语篇分析等.

除了以上三大领域, 本文还总结了一些比较新奇的其他领域的应用. 例如, 人体姿态估计、防止恶意软件攻击、物理应用、医学数据处理以及自动驾驶等.

可以说自从 Goodfellow 在 2014 年提出 GAN 之后, 尤其是近几年来, GAN 类的文章及应用呈井喷式爆发. 一方面, 各种应用场景给 GAN 的发展提出了挑战性的问题, 促使研究者根据应用场景研究新的 GAN 结构、模型和训练算法去解决计算机视觉、自然语言处理和语音处理中的问题; 另一方面, 新的 GAN 理论和模型的提出, 也拓展了人工智能在各领域中的应用广度和深度, 这也促使我们对近期 GAN 在各领域应用研究进展和重要文献进行总结及分析.

本文首先介绍了广泛应用的 9 种 GAN 及其变种, 然后对 GAN 在计算机视觉、自然语言处理和语音处理中的应用进行了详细的梳理. 最后, 探索性地给出了未来 GAN 的发展趋势及研究方向.

1 GAN 及其变种模型

1.1 GAN 及其训练过程

本节主要介绍 GAN 的结构以及训练过程. 首先, 生成器和判别器通常由包含卷积或者全连接层的多层网络实现, 并且生成器和判别器网络激活函数必须是可微的.

若将生成器网络看作是一个将隐空间映射到数据空间的函数, 那么可将生成器网络表示为 $G: z \to \mathbf{R}^{|x|}$. 其中, $z \in \mathbf{R}^{|z|}$ 是来自隐空间的样本, $x \in \mathbf{R}^{|x|}$ 是图像, $|\cdot|$ 表示维数. 另外, 判别器网络的输出 $D: G(z) \to \{0,1\}$, 判别生成图像数据是否来自于实际数据概率分布中的样本.

其次, GAN 的训练包括两个方面: 1) 最大化判别器分类准确率的判别器的参数; 2) 找到最大程度混淆判别器的生成器的参数. GAN 的训练过程可以由图 2 来表示.

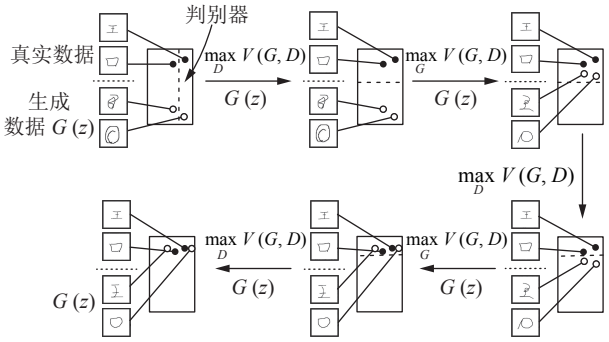


图 2 GAN 训练过程

Fig. 2 Training process of GAN

图 2 中的虚线相当于 GAN 中的判别器, 它的任务是准确判别出生成器生成的数据与真实数据. 首先, 优化判别器时, 需要画出虚线, 使它能够区分真实数据和生成数据; 优化生成器时, 生成数据会更加接近原始数据, 使得判别器难以区分数据的真假. 如此反复直到最后再也画不出区分真实数据与生成数据的虚线.

这两个过程在 GAN 中可以用一个值函数 $V(G,D)$ 来表示, 并把问题变为解决这个值函数的极小–极大问题:

$$\min_G \max_D V(G,D) = \mathrm{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathrm{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (1)$$

在训练过程中, 当更新一个模型生成器的参数

时, 另一个模型判别器的参数是固定的, 反之亦然. Goodfellow 等[2] 表明, 当固定生成器时, 存在唯一的最优判别器模型 $D^*(x)$:

$$D^*(x)=\frac{p_{\text{data}}(x)}{p_{\text{data}}(x)+p_g(x)} \tag{2}$$

另外, 当 $p_g(x)=p_{\text{data}}(x)$ 时, 生成器是最优的, 这相当于从 x 中抽取所有样本后, 判别器对由生成器产生的数据 $G(z)$ 与真实数据 x 的类标签取值为 0 或 1 的概率为 0.5. 换句话说, 当判别器不能区分真实样本和假样本时, 则生成器是最优的.

理想情况下, 根据上述训练方法先固定生成器训练判别器, 再更新生成器. 然而, 在实践过程中, 判别器由于不能被训练至类标签取值为 0 或 1 的概率为 0.5 的最佳状态, 而是通过少量的迭代进行训练, 并且生成器与判别器同时更新.

此外, 通常会将非饱和训练标准 $\max_G \log\ D(G(z))$ 用于生成器, 即训练生成器的时候使用 $\max_G \log D(G(z))$ 函数来替换 $\min_G \log(1-D(G(z)))$.

尽管 GAN 发展至今, 提出了很多应对不同场景 GAN 训练的解决方法, 但 GAN 训练具有挑战性, 因为 GAN 训练过程具有多种不稳定的原因. 其训练普遍存在以下三个问题:

1) 生成器模型和判别器模型训练过程难收敛[7].

2) 生成器会出现模式 “崩溃 (Mode collapse)” 现象. 即当输入不同时, 输出样本相同[8], 使生成数据缺乏多样性.

3) 判别器模型快速收敛到零[9], 没有为生成器模型提供可靠的参考信息.

1.2　GAN 变种模型及其评估标准

虽然 GAN 能从理论上完全逼近真实数据, 但是这种预先进行建模的方法太过自由. 若应对一些高分辨率的图像数据, 则基于简单的 GAN 无法对其生成结果进行控制, 生成结果无法预料. 因此在本小节中, 主要针对几类比较主流的 GAN 变种进行概述, GAN 的应用也主要集中在这几个方面.

1.2.1　条件生成对抗网络

条件生成对抗网络 (Conditional generative adversarial networks, CGAN)[10] 是一个对 GAN 进行条件约束的 GAN 变种网络. 这个网络在结构上分别在生成器和判别器中引入了条件变量 y. 通过使用条件变量 y 对模型增加更多信息, 可以说 GAN 从无监督网络变成了有监督网络. y 能够有效指导生成器的训练过程. CGAN 的结构如图 3 所示.

$G:(Z\times Y)\to \mathbf{R}^{|x|}$ 为生成器, 生成器的输入为

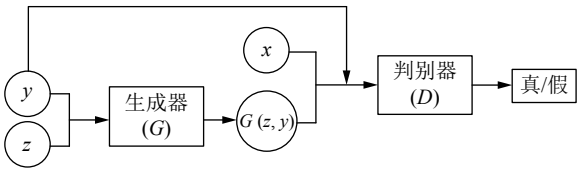


图 3　CGAN 结构

Fig. 3　The structure of CGAN

噪声数据 $z\in Z$ 和条件数据 $y\in Y$, 这个条件数据可以是一个类标签和低维数据, 也可以是图片数据和高维度数据, 且生成图像 $G(z|y)$.$D(G(z|y),y)\in\{0,1\}$ 为判别器, 它的输入为真实图像 x 或者生成图像 $G(z|y)$ 并结合条件 y, 判别在条件 y 下预测图像来自经验数据的概率分布或是来自生成器的概率分布. 这个 CGAN 类似于 GAN玩极小–极大游戏

$$\min_G\max_D\ V(D,G)=\mathrm{E}_{x\sim p_{\text{data}}(x)}[\log D(x|y)]+\mathrm{E}_{z\sim p_z(z)}[\log(1-D(G(z|y)))] \tag{3}$$

在后面的论述中, 有很多应用用到了 CGAN, 或是对 CGAN 进行进一步的改进, 如变换网络结构、增加一些附加惩罚损失函数等, 但将条件约束应用到 GAN 的思想是不变的. 可以说 CGAN 是 GAN 应用中最为重要的一部分.

1.2.2　深卷积生成对抗网络

卷积神经网络 (Convolutional neural network, CNN)[11] 的提出, 推动了计算机视觉领域的迅速发展. 深度卷积生成对抗网络 (Deep convolutional GAN, DCGAN) 模型成功地将卷积神经网络与 GAN 进行了融合, 利用卷积网络强大的特征提取能力来提高 GAN 的学习效果. DCGAN[7] 可以在分辨率更高的图像、更深的生成模型上稳定地训练. DCGAN 的改进主要是在网络结构上对 GAN 进行了改进. 生成器相当于反卷积网络, 而判别器相当于卷积网络. DCGAN 的生成器网络结构如图 4 所示.

DCGAN 的主要特点有:

1) 去掉生成器和判别器中的池化层. 在判别器中, 使用步幅卷积函数 (Strided convolutions) 替代池化函数. 在生成器中, 使用微步幅卷积函数 (Fractional-strided convolutions) 替代池化函数.

2) 在生成器和判别器中采用批量归一化技术. 这样做的好处是:

a) 解决了神经网络的训练结果依赖于初始连接矩阵权值和偏置向量权值设置问题;

b) 防止 DCGAN 反向传播过程出现梯度消失和梯度爆炸问题;

c) 防止生成器把所有输入样本都收敛到同一

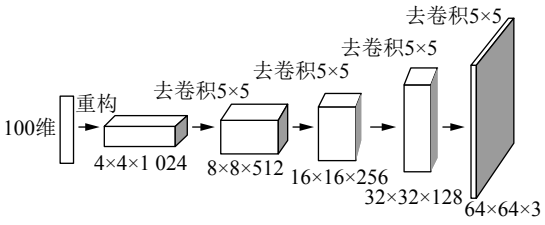


图 4 DCGAN 生成器网络结构

Fig. 4 The structure of DCGAN's generator

个输出上.

由于若将所有层都增加批量归一化, 会发生样本震荡和模型不稳定的情况. 因此在 DCGAN 中, 生成器的输出层和判别器的输入层不采用这种方法.

3) 在 CNN 中去掉全连接隐层, 虽然全连接层能够提高训练的稳定性, 但是会显著降低训练过程收敛的速度.

4) 生成器的最后一层的激活函数采用了 Tanh 函数, 其他层采用了 ReLU (Rectified linear units) 函数. 判别器网络每层用 Leaky ReLU 激活函数.

关于 DCGAN 的应用主要集中在图像处理方面, 可以说这个模型是最典型、应用最广泛的 GAN 变种模型.

1.2.3 Wasserstein-GAN 与 WGAN-GP

GAN 训练时会存在梯度消失和不稳定的问题. 首先, 梯度消失是由于生成对抗网络在训练过程中, 当判别器接近最优时, 由于生成器的损失函数 $\mathrm{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]$ 是最小化生成分布与真实分布之间的 JS 散度 (Jensen-Shannon divergence), 但是当这两个分布之间没有重叠或者重叠极小时, 都会使得 JS 散度为常数, 从而导致梯度消失. 其次, GAN 训练过程中存在不稳定的问题, 主要因为 GAN 存在生成器和判别器模型难以同时收敛从而导致模式 "崩溃".

WGAN (Wasserstein-GAN)[12] 对 GAN 进行了改进: 在 WGAN 中用 Wasserstein 距离, 也即 EM 距离 (Earth-Mover distance) 代替 JS 散度. 与 KL 散度 (Kullback-Leibler divergence)、JS 散度相比, Wasserstein 距离的优越性在于, 若两个分布之间没有重叠, Wasserstein 距离仍能反映分布的远近程度. 除此之外, KL 散度和 JS 散度是不连续变化的, 但是 Wasserstein 距离却是平滑变化的. 用梯度下降法优化模型参数时, KL 散度和 JS 散度都无法求导, 而 Wasserstein 距离却可以. 类似地, 在高维空间中, 如果两个分布不重叠或者重叠部分可忽略, 则 KL 散度和 JS 散度既无法反映概率分布的远近程度, 也无法求导, 得不到梯度. 但是 Wasserstein 却可以提供有意义的梯度, 这就是 Wasserstein 距离度量两个分布之间距离的优势所在.

Wasserstein 距离度量两个分布之间的距离为

$$W(P_r,P_g)=\inf_{\gamma\sim\prod(P_r,P_g)}\mathrm{E}_{(x,y)\sim\gamma}[\|x-y\|] \tag{4}$$

其中, $\prod(P_r,P_g)$ 是 P_r 和 P_g 组合起来的所有可能的联合分布的集合. 对于每一个可能的联合分布 $\gamma\in\prod(P_r,P_g)$ 而言, 可以从 $(x,y)\sim\gamma$ 中采样得到一个真实样本 x 和一个生成样本 y, 并计算出这对样本的距离 $\|x-y\|$. 此时计算出 γ 下样本对距离的期望 $\mathrm{E}_{(x,y)\sim\gamma}[\|x-y\|]$, 并在所有可能的联合分布中对这个期望取下界 $\inf_{\gamma\sim\prod(P_r,P_g)}\mathrm{E}_{(x,y)\sim\gamma}[\|x-y\|]$.

由于 Wasserstein 距离中的下界无法直接求解, 所以 WGAN 文献中通过一系列数学演算后得到近似 Wasserstein 距离公式

$$K\times W(P_r,P_g)\approx \max_{\|f_\omega\|_L\le K}\mathrm{E}_{x\sim P_r}[f_\omega(x)]-\mathrm{E}_{x\sim P_g}[f_\omega(x)] \tag{5}$$

其中, 函数 f 的一阶 Lipschitz 常数为 K, 把 f 用一个参数为 $\omega\in[-c,c]$ 的神经网络来表示, f 作为判别器. 另外, K 的变化会引起梯度发生 K 倍的变化, 并不会影响梯度的方向. 而 WGAN 实际上对 GAN 网络的修改只有以下 4 个方面:

1) 判别器的神经网络的最后一层去掉 sigmoid 函数.

2) 生成器和判别器的损失函数不取对数 (log) 函数.

3) 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, 即要将权重限制到一个范围内.

4) 不使用基于动量的优化算法 (例如 Momentum 算法和 Adam 算法) 进行参数更新. 使用的是均方根方法[13] 或者随机梯度下降方法.

WGAN 的提出基本解决了模式崩溃问题, 保证了生成样本的丰富性. 除此之外, 彻底解决了 GAN 训练不稳定的问题, 并且在训练过程中用 Wasserstein 距离来指示训练的进展过程, 这个数值越小代表 GAN 训练得越好, 即生成器生成的图像质量越好.

随后, 在 WGAN 的基础上又提出了梯度惩罚的 WGAN (WGAN-gradient penalty, WGAN-GP) 模型[14]. WGAN-GP 的提出解决了一些 WGAN 在具体的实验过程中存在训练困难和收敛缓慢的问题. WGAN 对 GAN 的主要改进是对 Lipschitz 连续性附加限制条件. 将 f 的权重限制在一定范围内, 即 $\omega\in[-c,c]$, 这会导致无论生成样本

与真实样本多么复杂, 判别器都将大部分数据的判别结果集中在阈值边界上. 此时, 判别器出现过度拟合现象, 使判别器不能充分发挥其判别能力, 这样的限制也会导致梯度爆炸或梯度消失问题.

因此文献 [14] 提出梯度惩罚 (Gradient penalty) 方法, 其本质是在原来的损失函数中增加了一个使梯度与 K 之间关联起来的罚项, 其中 WGAN-GP 中判别器的损失函数为

$$LD = \mathop{\mathrm{E}}_{\hat{x}\sim P_g}[D(\tilde{x})] - \mathop{\mathrm{E}}_{\hat{x}\sim P_r}[D(x)] + \lambda \mathop{\mathrm{E}}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \tag{6}$$

其中, $x_r \sim P_r, x_g \sim P_g, \varepsilon \sim \text{Uniform}[0,1]$, 并在 x_r 和 x_g 的连线上随机进行插值采样

$$\hat{x} = \varepsilon x_r + (1-\varepsilon)x_g \tag{7}$$

WGAN-GP 的贡献在于提出了一种全新的梯度惩罚策略, 解决了 WGAN 中的梯度消失及梯度爆炸的问题, 并且比 WGAN 具有更快的收敛速度. WGAN 的提出为 GAN 的稳定训练提供了极大的帮助, 在不同领域发挥着重要的作用.

1.2.4　InfoGAN

InfoGAN (Information GAN)[15] 在 CGAN 的基础上又进行了创新. 文献 [15] 在 2016 年被 OpenAI 组织评为年度五大突破之一. InfoGAN 将 GAN 与信息理论进行有效结合, InfoGAN 通过引入隐变量编码与生成数据之间的互信息约束, 使得模型能够学习到有价值的可解释性特征. InfoGAN 的提出利用了互信息的概念, 在信息论中, X, Y 之间的互信息 $I(X;Y)$ 可以用来衡量随机变量 X 中包含随机变量 Y 的信息量.

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{8}$$

如图 5 所示, 其中 H 表示熵, 那么互信息可以重新写为

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{9}$$

通过生成器生成的数据具有高度耦合的特点, 导致数据的每一个维度不代表具体的具有语义含义的特征. 由于 GAN 中对于连续噪声变量 z' 没有进行任何处理, 在 InfoGAN 中将 z' 表示成不可压缩的噪声 z 和隐变量编码 c 两部分, c 代表数据分布的结构化语义特征. 若 c 与生成数据 $G(z,c)$ 具有更多共同信息, 那么 c 和 $G(z,c)$ 应该具有高度相关性, 即互信息强. 而如果是无共同信息的话, 那么二者之间没有特定的关系, 即互信息接近于 0. 用目标函

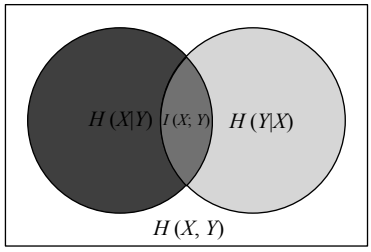


图 5　互信息图

Fig. 5　Mutual information map

数表示为

$$\min_G \max_D V_1(D,G) = V(D,G) - \lambda I(c; G(z,c)) \tag{10}$$

其中, λ 为超参数.

但是为了求互信息值 $I(c;G(z,c))$, 需要用到 G 生成的 $x_{\text{fake}} = G(z,c)z,c)z,c)$ 的条件后验概率 $p(c|x_{\text{fake}})$, 但是 $p(c|x_{\text{fake}})$ 很难求取, 需要求取 $p(c|x_{\text{fake}})$ 的下界. 于是 InfoGAN 中提出利用辅助分布 $Q(C|X_{\text{fake}})$ 逼近条件后验概率分布 $P(C|X_{\text{fake}})$, 所以利用神经网络对 $Q(C, X_{\text{fake}})$ 进行估计, 并得出 $P(C|X_{\text{fake}})$ 的下界. 除此之外, $Q(C, X_{\text{fake}})$ 将与判别器判别器共享全部卷积层, 并在卷积层后面增加一个全连接层, 输出 $Q(C|X_{\text{fake}})$, 通过这种方法简化计算. InfoGAN 的结构如图 6 所示.

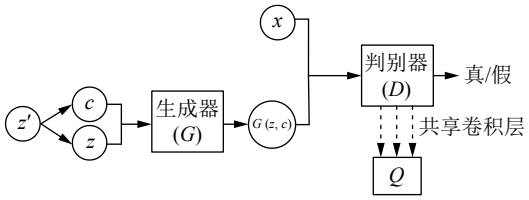


图 6　InfoGAN 结构

Fig. 6　The structure of InfoGAN

InfoGAN 利用互信息的手段对输入噪声 z 进行控制, 这样就能通过这个互信息 c 来对生成图像进行控制.

1.2.5　SeqGAN

SeqGAN (Sequence GAN)[16] 是将强化学习 (Reinforcement learning, RL) 与 GAN 成功进行融合的模型, 是强化学习思想与对抗思想的碰撞成果. RL 研究智能体和环境相互交互过程, 是通过 "试错" 的方式来学习最优策略的马尔科夫决策过程 (Markov decision process, MDP)[17]. RL 由 4 个基本部分组成, 分别是: 状态集合 S, 动作集合 A, 状态转移概率矩阵 P 和奖励函数 R. 定义策略 π 为状态空间到动作空间的映射. 智能体在当前状态 s_t 下根据策略 π 选择动作 a_t 作用于环境, 然后接收到环

境反馈回来的奖励 r_t，并以转移概率 $p^a_{s_t,s_{t+1}}$ 转移到下一个状态 s_t. RL 的目标是通过不断调整优化策略来最大化累积奖励值.

假定 $Y_{1:T}=(y_1,\cdots,y_t,\cdots,y_T)$ 为生成器的生成序列，$y_t\in\mathcal{Y}$，$\mathcal{Y}$ 是一个候选令牌词汇库，s 是取值于 $(y_1,\cdots,y_{t-1})$ 的状态集合. 在 SeqGAN 中将策略 π 看作一个由参数 θ 控制的生成器 $G_\theta(y_t|Y_{1:t-1})$，D_ϕ 是一个由参数 ϕ 控制的判别器.

SeqGAN 结构如图 7 所示. 在已知当前状态 s_{t-1} 时，通过蒙特卡洛树搜索 (Monte Carlo tree search, MCTS) 方法将当前状态之后要发生的动作 a_t 进行补全，并用 D_ϕ 对每一个完整序列进行评分. 将这个评分结果作为奖励 r_t 回传给 G_θ，并通过梯度算法对 G_θ 进行更新. 通过使用 RL 的方法，训练出一个可以产生下一个最优动作的生成网络.

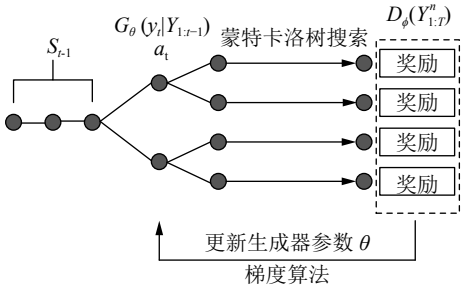


图 7 SeqGAN 结构

Fig. 7 The structure of SeqGAN

SeqGAN 的目标函数是为了生成更好的序列使奖励最大化

$$J(\theta)=\mathrm{E}[R_T|s_0,\theta]=\sum_{y_1\in Y}G_\theta(y_1|s_0)\times Q^{G_\theta}_{D_\phi}(s_0,y_1) \tag{11}$$

其中，R_T 是一个完整序列的奖励值，$J(\theta)$ 是在初始状态 s_0 与模型参数 θ 已知条件下，产生的完整序列的奖励的期望. 而 $Q^{G_\theta}_{D_\phi}(s_0,y_1)$ 是序列的动作值函数. $D_\phi(Y^n_{1:T})$ 为动作值函数的值

$$Q^{G_\theta}_{D_\phi}(a=y_T,s=Y_{1:T-1})=D_\phi(Y_{1:T}) \tag{12}$$

然而，$D_\phi(Y^n_{1:T})$ 只能针对完整的序列进行评分. 因此，若要估计中间状态的动作值函数，需要用到蒙特卡洛搜索方法中的 roll-out 策略 G_β 对未知的 $T-t$ 时间内的 $(y_{t+1},\cdots,y_T)$ 的值进行抽样. 定义一个 N 次的蒙特卡洛搜索算法为

$$\{Y^1_{1:T},\cdots,Y^N_{1:T}\}=MC^{G_\beta}(Y_{1:t};N) \tag{13}$$

综上所述，SeqGAN 的动作值函数为

$$Q^{G_\theta}_{D_\phi}(a=y_T,s=Y_{1:T-1})=\begin{cases}D_\phi(Y_{1:t}), & t=T\\ \dfrac{1}{N}\displaystyle\sum_{n=1}^{N}D_\phi(Y^n_{1:T}), & Y^n_{1:T}\in MC^{G_\beta}(Y_{1:t};N),\\ & t<T\end{cases} \tag{14}$$

其中，在已知 y_1 到 y_{t-1} 的情况下，需要通过蒙特卡洛搜索法对每一个路径的其余 $(y_{t+1},\cdots,y_T)$ 进行补全，并对 n 条完整的序列进行评分，作为 $D_\phi(Y^n_{1:T})$ 返回的奖励值.

虽然 GAN 在自然语言处理 (Natural language processing, NLP) 领域的应用一直没有突破，SeqGAN 的提出证明了 GAN 在 NLP 领域是一样具有影响力的，并在机器翻译等重要领域发挥着显著作用.

1.2.6 Pix2Pix

Pix2Pix[18] 用 CGAN 实现了图像翻译任务，即将图像内容从一个源域迁移到另一个目标域上，也可表示为图像移除一个域的属性，然后赋予另一个域的属性. 由于 Pix2Pix 的输入为配对图片，所以需要对 CGAN 进行相应的改进. Pix2Pix 的结构如图 8 所示，将随机噪声 z 与观测图像 x 作为生成器的输入，生成一个类似于真实图像 y 的生成图像 $G(x,z)$. 判别器则把 x 与 y 或者 x 与 $G(x,z)$ 的成对图像作为输入进行判别. 由此可见，Pix2Pix 将观测图像 x 作为 CGAN 中的条件信息.

如前文所示，GAN 能够学习到高维数据的特

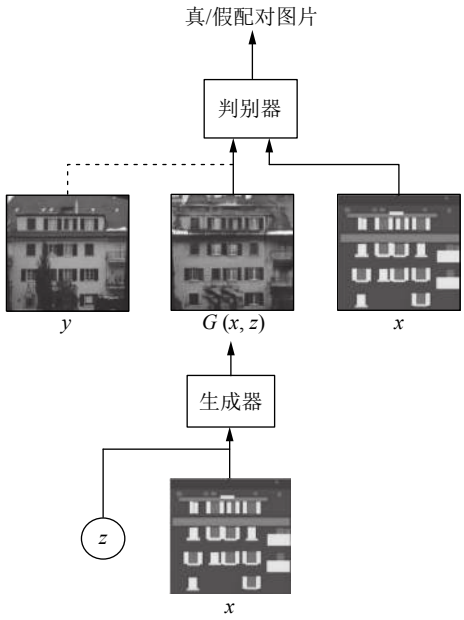


图 8 Pix2Pix 结构

Fig. 8 The structure of Pix2Pix

征, Pix2Pix 使用了 CGAN 的损失函数, 并将观测图像 x 作为条件信息

$$L_{\mathrm{CGAN}}(G,D)=\mathrm{E}_{x,y}[\log D(x,y)]+\mathrm{E}_{x,z}[\log(1-D(x,G(x,z)))] \tag{15}$$

其中, z 为随机输入噪声向量, x 为观测图像.

Pix2Pix 在 CGAN 的基础上, 为了保证输入图像和输出图像之间的相似度, 在 $L_{\mathrm{CGAN}}(G,D)$ 上加了一个 $L1$ 或 $L2$ 范数正则化项, 其中 $L1$ 范数正则化项如下:

$$L_{L1}(G)=\mathrm{E}_{x,y,z}[\|y-G(x,z)\|_1] \tag{16}$$

所以 Pix2Pix 的总损失函数变为

$$G^*=\arg\min_G\max_D L_{\mathrm{CGAN}}(G,D)+\lambda L_{L1}(G) \tag{17}$$

其中, λ 为权衡参数.

Pix2Pix 模型的特点主要在生成器的选择上, Pix2Pix 选择了 U-net[19] 作为生成器. U-net 由编码器和解码器组成, 由于第 i 层与第 $n-i$ 层的神经元个数相同, 所以将第 i 层的信息传递给第 $n-i$ 层的信息, 保存更多的特征信息, 换句话说 U-Net 是将第 i 层拼接到第 $n-i$ 层, 这样做是因为第 i 层和第 $n-i$ 层的图像大小是一致的, 这两层之间存在着相似的信息. Pix2Pix 模型为图像转换、图像边缘检测等领域的快速发展做出重要贡献. Pix2Pix 作为一个典型利用 CGAN 结构的神经网络, 为图像领域应用提供了很好的基础.

1.2.7　CycleGAN 与 Augmented CycleGAN

1.2.7.1　CycleGAN

CycleGAN (Cycle-consistent GAN)[20] 是将对偶学习与 GAN 进行结合的结果, CycleGAN 能将一类图片自动转化为另一类图片. 其中对偶学习[21] 是微软亚洲研究院 (Microsoft Research Asia, MSRA) 于 2016 年提出的一种用于机器翻译的强化学习方法, 用于解决海量数据配对标注的问题, 是一个无监督方法. 对偶学习是一种类似于双语翻译任务的双重学习机制, 利用原始任务和双重任务之间的反馈信号对模型进行训练.

CycleGAN 由一对镜像对称的 GAN 网络构成. 每个单向 GAN 都由两个生成器 $\{G,F\}$ 和两个判别器 $\{D_X,D_Y\}$ 组成. CycleGAN 中包含了两个映射关系 $G:X\to Y$ 和 $F:Y\to X$, 所以 CycleGAN 有 4 个损失函数. 其原理如图 9 所示.

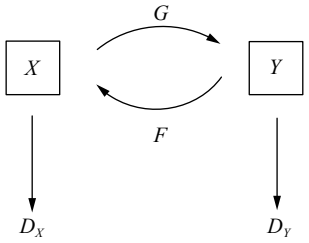


图 9　CycleGAN 原理图

Fig. 9　Principle of CycleGAN

图 9 中, 假定 $x\sim p_{\mathrm{data}}(x)$, $y\sim p_{\mathrm{data}}(y)$, 代表 X, Y 两个域中的数据服从的概率分布, 其中, $G:X\to Y$ 的单向损失损失函数为

$$L_{\mathrm{GAN}}(G,D_Y,X,Y)=\mathrm{E}_{y\sim p_{\mathrm{data}}(y)}[\log D_Y(y)]+\mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[1-D_Y(G(x))] \tag{18}$$

其中, G 会生成类似于 Y 中的样本 $G(x)$, D_Y 的任务是区分 $G(x)$ 和真实样本 y. 同样 $F:Y\to X$ 的单向损失函数为

$$L_{\mathrm{GAN}}(F,D_X,Y,X)=\mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[\log D_X(x)]+\mathrm{E}_{y\sim p_{\mathrm{data}}(y)}[1-D_X(F(x))] \tag{19}$$

其中, F 会生成类似于 X 中的样本 $F(y)$, D_X 的任务是区分 $F(y)$ 和真实样本 x.

CycleGAN 还增加了一项循环一致损失函数 $L_{\mathrm{cyc}}(G,F)$, $L_{\mathrm{cyc}}(G,F)$ 在两个单向损失函数上, 施加 $F(G(x))-x$ 和 $G(F(x))-y$ 的 L_1 范数的数学期望约束, 能够学习到更好的 $F(G(x))$ 和 $G(F(x))$ 映射关系

$$L_{\mathrm{cyc}}(G,F)=\mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[\|F(G(x))-x\|_1]+\mathrm{E}_{y\sim p_{\mathrm{data}}(y)}[\|G(F(x))-y\|_1] \tag{20}$$

合并上面 3 个损失函数可以得出 CycleGAN 总损失函数为

$$L(G,F,D_X,D_Y)=L_{\mathrm{GAN}}(G,D_Y,X,Y)+L_{\mathrm{GAN}}(F,D_X,Y,X)+\lambda L_{\mathrm{cyc}}(G,F) \tag{21}$$

其中, λ 为权衡参数, 用于权衡双向损失函数与循环一致损失函数的相对重要性.

虽然 Pix2Pix 也可用于图像翻译及风格转换等任务, 但它与 CycleGAN 的区别在于, Pix2Pix 模型要求的输入必须是成对数据, 而 CycleGAN 则对非成对数据也能进行训练.

1.2.7.2　SFSADFASDF

尽管 CycleGAN 能够从不成对数据中学习域间映射关系, 这样做的好处在于可以通过减少对配对数据的需求来提高结构化预测任务的性能. 但是在 CycleGAN 中假设底层域间映射关系几乎是确定性的, 并且是一对一的. 因而, CycleGAN 无法应用于多对多映射的学习任务. 但是不同域之间的映

射关系本应该是更加复杂的, 不都是一对一映射关系, 多对多映射能够更好地反映不同域之间关系. 文献 [22] 提出了一个名为 Augmented CycleGAN 的模型, 它可以学习域间的多对多映射. 假设源域为 A, 服从概率分布 $p_d(a)$, 目标域为 B, 服从概率分布 $p_d(b)$. 假设随机变量 Z_a 和 Z_b 服从标准高斯先验分布, 分别记为 $p(z_a)$ 和 $p(z_b)$, 并与 $p_d(a)$ 和 $p_d(b)$ 相互独立. 假设一对样本 z_a 和 z_b 分别是从 $p(z_a)$ 和 $p(z_b)$ 两个标准高斯分布中采样得到的, 样本 a 和 b 分别是从数据集概率分布 $p_d(a)$ 与 $p_d(b)$ 中采样得到的.

Augmented CycleGAN 由 4 个映射函数组成: 首先构造两个映射函数 $G_{AB}: A\times Z_b\mapsto B, G_{BA}: B\times Z_a\mapsto A$, G_{AB} 和 G_{AB} 可以看作是两个具有条件信息的生成器. 其次, 构造两个编码器映射函数 $E_A: A\times B\mapsto Z_a$ 和 $E_B: A\times B\mapsto Z_b$, 编码器能够通过随机结构化的映射来对循环一致性进行优化. G_{AB}, G_{BA}, E_A 和 E_B 分别由神经网络构成. Augmented CycleGAN 的原理如图 10 所示.

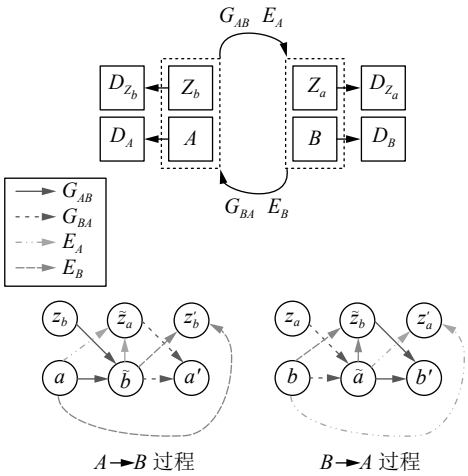


图 10 Augmented CycleGAN 原理图

Fig. 10 Principle of augmented CycleGAN

Augmented CycleGAN 的具体构造过程如下:

1) $A\to B$

如图 10 所示, 当给出 $(a, z_b)\sim p_d(a)p(z_b)$ 一对数据时, 通过 G_{AB} 和 E_A 生成数据对 $(\tilde{b}, \tilde{z}_a)$

$$\tilde{b}=G_{AB}(a, z_b),\ \tilde{z}_a=E_A(a, \tilde{b}) \tag{22}$$

$A\to B$ 单向损失函数为

$$L^B_{\text{GAN}}(G_{AB}, D_B)=\mathop{\text{E}}_{b\sim p_d(b)}[\log D_B(b)]+\mathop{\text{E}}_{\substack{a\sim p_d(a)\\ z_b\sim p(z_b)}}[\log(1-D_B(G_{AB}(a, z_b)))] \tag{23}$$

$$L^{Z_a}_{\text{GAN}}(E_A, G_{AB}, D_{Z_a})=\mathop{\text{E}}_{a\sim p(z_a)}[\log D_{Z_a}]+\mathop{\text{E}}_{\substack{a\sim p_d(a)\\ z_b\sim p(z_b)}}[\log(1-D_{Z_a}(\tilde{z}_a))] \tag{24}$$

2) $B\to A$

同样, 如图 10 所示, 当给出 $(b, z_a)\sim p_d(b)p(z_a)$ 一对数据时, 通过 G_{AB} 和 E_B 生成数据对 $(\tilde{a}, \tilde{z}_b)$

$$\tilde{a}=G_{BA}(b, z_a), \tilde{z}_b=E_B(b, \tilde{a}) \tag{25}$$

$B\to A$ 单向损失函数为

$$L^A_{\text{GAN}}(G_{BA}, D_A)=\mathop{\text{E}}_{a\sim p_d(a)}[\log D_A(a)]+\mathop{\text{E}}_{\substack{b\sim p_d(b)\\ z_a\sim p(z_a)}}[\log(1-D_A(G_{BA}(b, z_a)))] \tag{26}$$

$$L^{Z_b}_{\text{GAN}}(E_B, G_{BA}, D_{Z_b})=\mathop{\text{E}}_{b\sim p(z_b)}[\log D_{Z_b}]+\mathop{\text{E}}_{\substack{b\sim p_d(b)\\ z_a\sim p(z_a)}}[\log(1-D_{Z_b}(\tilde{z}_b))] \tag{27}$$

3) Augmented CycleGAN 的循环一致损失函数

在 Augmented CycleGAN 中从每一个单向转换过程 ($A\to B$ 或 $B\to A$) 经过一次循环后, 定义了两个循环一致损失函数.

a) 若从 $A\to B$ 开始, 则循环一致损失函数中关于 a 的循环一致损失函数 L^A_{cyc} 为

$$L^A_{\text{cyc}}(G_{AB}, G_{BA}, E_A)=\mathop{\text{E}}_{\substack{a\sim p_d(a)\\ z_b\sim p(z_b)}}\|a'-a\|_1 \tag{28}$$

其中, $a'=G_{BA}(\tilde{b}, \tilde{z}_a)$.

循环一致损失函数中关于 z_b 的循环一致损失函数 $L^{Z_b}_{\text{cyc}}$ 为

初始空段落

$$L^{Z_b}_{\text{cyc}}(G_{AB}, E_B)=\mathop{\text{E}}_{\substack{a\sim p_d(a)\\ z_b\sim p(z_b)}}\|z'_b-z_b\|_1 \tag{29}$$

其中, $z'_b=E_B(a, \tilde{b})$.

b) 若从 $B\to A$ 开始, 则循环一致损失函数中的关于 b 的循环一致损失函数 L^B_{cyc} 为

$$L^B_{\text{cyc}}(G_{BA}, G_{AB}, E_B)=\mathop{\text{E}}_{\substack{b\sim p_d(b)\\ z_a\sim p(z_a)}}\|b'-b\|_1 \tag{30}$$

其中, $b'=G_{AB}(\tilde{a}, \tilde{z}_b)$.

循环一致损失函数中的关于 z_b 的循环一致损失函数 $L^{Z_a}_{\text{cyc}}$ 为

$$L^{Z_a}_{\text{cyc}}(G_{BA}, E_A)=\mathop{\text{E}}_{\substack{b\sim p_d(b)\\ z_a\sim p(z_a)}}\|z'_a-z_a\|_1 \tag{31}$$

其中, $z'_a = E_A(b, \tilde{a})$.

综上所述, Augmented CycleGAN 的总损失函数为

$$\begin{aligned} L = {} & L^B_{\text{GAN}}(G_{AB}, D_B) + L^{Z_a}_{\text{GAN}}(E_A, G_{AB}, D_{Z_a}) + \\ & \gamma_1 L^A_{\text{cyc}}(G_{AB}, G_{BA}, E_A) + \gamma_2 L^{Z_b}_{\text{cyc}}(G_{AB}, E_B) + \\ & L^A_{\text{GAN}}(G_{BA}, D_A) + L^{Z_b}_{\text{GAN}}(E_B, G_{BA}, D_{Z_b}) + \\ & \gamma_3 L^B_{\text{cyc}}(G_{BA}, G_{AB}, E_B) + \gamma_4 L^{Z_a}_{\text{cyc}}(G_{BA}, E_A) \end{aligned} \tag{32}$$

其中, $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ 为权衡超参数.

Augmented CycleGAN 以无监督的方式学习多对多跨域映射. 定量和定性的实验结果验证了该模型在图像翻译任务中的有效性, Augmented CycleGAN也可以有效使用在半监督学习场景中. 因此, CycleGAN 与 Augmented CycleGAN 为 GAN 图像转换类应用提供了很好的技术.

1.2.8　GAN 评估方法

由于 GAN 的种类比较多, 对它们的评估需要依靠人工检验生成图像的视觉效果, 这样的评估十分耗费人力和时间, 并且评估结果会受到人的主观性的影响. 因此 GAN 的评估指标尤为重要. 目前对 GAN 进行定量评估的主要方法是通过评判生成分布 P_{g} 与真实分布 P_{r} 之间的相似性. GAN 的定量评估方法如下:

1) Inception 评分

借助 Google 外部模型 Inception 网络来评估生成图像的质量以及多样性[8]. Inception 评分可以通过下式求出:

$$IS\left(P_{\text{g}}\right) = \text{e}^{\text{E}_{x\sim P_{\text{g}}}[KL(p_M(y|x)||p_M(y))]} \tag{33}$$

其中, $p_M(y|x)$ 表示图像分类模型 M 在给定样本 x 下的标签条件后验概率分布. $p_M(y)$ 是通过 $p_M(y) = \int_x p_M(y|x)\ \text{d}P_{\text{g}}$ 求出的. Inception 打分不需要用到真实数据的概率分布, 只是针对人工生成概率分布进行评估.

2) Mode 评分

该评估方法为 Inception 评分方法的改进版本, 通过增加对真实分布的度量来对生成分布进行更好的评估[23]. 其计算式为

$$\begin{aligned} MS\left(P_{\text{g}}\right) = \exp\Big(\text{E}_{x\sim P_{\text{g}}}\big[& KL(p_M(y|x)||p_M(y)) - \\ & KL(p_M(y|x)||p_M(y^*))\big]\Big) \end{aligned} \tag{34}$$

其中, 真实概率分布 $p_M(y^*)$ 通过求积分 $p_M(y^*) = \int_x p_M(y|x)\text{d}P_{\text{r}}$ 计算得到. 通过减去真实分布, 式 (34) 即可度量真实概率分布和生成概率分布的差异.

3) 核最大均值差异 (Kernel maximum mean discrepancy, Kernel MMD)

利用不同的核函数 k 来度量生成概率分布与真实概率分布之间的相似性[24], 其计算式为

$$\begin{aligned} \text{MMD}^2(P_{\text{r}}, P_{\text{g}}) = {} & \text{E}_{\substack{x_{\text{r}}, x'_{\text{r}}\sim P_{\text{r}}, \\ x_{\text{g}}, x'_{\text{g}}\sim P_{\text{g}}}}[k(x_{\text{r}}, x'_{\text{r}}) - \\ & 2k(x_{\text{r}}, x_{\text{g}}) + k(x_{\text{g}}, x'_{\text{g}})] \end{aligned} \tag{35}$$

MMD 的值越低, 则两个概率分布之间的相似性越高.

4) Wasserstein 距离

如式 (4) 所示, Wasserstein 距离越小, P_{g} 与 P_{r} 越相似.

5) Fréchet inception 距离 (Fréchet inception distance, FID)[25]

FID 需先选取一个特征函数 ϕ, 并假设 $\phi(P_{\text{r}})$ 和 $\phi(P_{\text{g}})$ 为高斯随机变量, 并考虑两个高斯随机变量 $\phi(P_{\text{r}})$ 和 $\phi(P_{\text{g}})$ 的均值和方差计算 FID:

$$\begin{aligned} \text{FID}(P_{\text{r}}, P_{\text{g}}) = {} & \|\mu_{\text{r}} - \mu_{\text{g}}\| + \\ & \text{tr}(C_{\text{r}} + C_{\text{g}} - 2(C_{\text{r}}C_{\text{g}})^{\frac{1}{2}}) \end{aligned} \tag{36}$$

其中, μ_{r}, C_{r} 和 μ_{g}, C_{g} 分别为 $\phi(P_{\text{r}})$ 和 $\phi(P_{\text{g}})$ 的均值和协方差矩阵. FID 距离越小, P_{g} 生成的样本越自然, 并具有 P_{r} 相似的多样性.

6) 1-最近邻分类器 (1-nearest neighbor classifier, 1-NN)[26]

采用 1-NN 分类器进行评估时, 通过留一 (Leave-one-out, LOO) 准确率来评估 P_{g} 与 P_{r} 的差异程度. 假设 $S_{\text{r}} \sim P^n_{\text{r}}, S_{\text{g}} \sim P^m_{\text{g}}$ 分别是来自两个概率分布 P^n_{r} 和 P^m_{g} 的样本, 并且样本个数相等, 即 $|S_{\text{r}}| = |S_{\text{g}}|$, 当两个分布完全匹配时, LOO 准确率为 50%.

在文献 [27] 中对以上 6 种 GAN 评估方法进行了不同的测试, 实验结果表明 Kernel MMD 和 1-NN 分类器在判别能力、鲁棒性和效率方面都更具有优越性.

除此之外, 当 GAN 应用在不同领域时, 会采取不同的评估标准, 例如: 峰值信噪比 (peak signal to noise ratio, PSNR)、结构相似性 (structural similarity, SSIM)、平均主管意见评分 (mean opinion score, MOS) 等.

1.3　小结与分析

本节主要介绍了 GAN 及其主要的 GAN 变种以及 GAN 的评估方法. GAN 变种的变化主要分为 3 种:

1) 针对 GAN 的训练过程或者损失函数进行改进;

2) 将不同的神经网络应用在 GAN 上, 使其获

得显著的效果;

3) 将前两者进行融合的混合模型.

因此, 可将上述模型进行一个简单的归类, 其中, CGAN, WGAN, WGAN-GP 以及 InfoGAN 属于第 1 种变种; DCGAN 属于第 2 种变种; Augmented CycleGAN, CycleGAN, Pix2PixGAN 和 SeqGAN 属于第 3 种变种. 根据以上 3 种类型的变化, 使得 GAN 应用的领域大大拓宽. 不仅促进了在图像、声音和 NLP 领域的快速发展, 还增强了不同模型之间融合的可能性. 表 1 汇总了本节中不同模型的提出时间.

表 1 GAN 模型变种

Table 1 Variant of GAN model

年份	模型
2014	条件生成对抗网络 (CGAN)[10]
2015	深卷积生成对抗网络 (DCGAN)[7]
2017	Wasserstein-GAN (WGAN)[12]
2017	具有梯度惩罚项 (WGAN-GP)[14]
2016	信息生成对抗网络 (InfoGAN)[15]
2017	序列生成对抗网络 (SeqGAN)[16]
2017	基于CGAN的图像到图像翻译模型 (Pix2Pix)[18]
2017	循环生成对抗网络 (CycleGAN)[20]
2018	增强循环生成对抗网络 (Augmented CycleGAN)[22]

2 GAN 在图像领域的应用

GAN 在图像领域的应用是 GAN 的众多应用中最丰富多样的. 在这一节中, 对近年来 GAN 在图像领域的应用进行总结. 主要分为以下几类: 人脸图像识别、图像生成、图像超分辨率、图像复原、多视角图像生成、图像风格转化、文本到图像的生成、语义操作和图像自动填充.

2.1 人脸图像识别与图像生成

人脸识别技术是一个基于计算机技术对脸部图像进行特征提取并进行识别的过程. 在 GAN 的应用中这是一个比较常见且关键的应用领域.

2.1.1 基于 CGAN 的人脸图像生成

关于 CGAN 在人脸识别方面的应用研究主要集中在用 CGAN 来生成真实世界人脸的可能性, 并研究如何通过修改特定的条件信息来准确控制人脸属性[28], 其训练过程为: 首先, 生成器默认情况下输出随机 RGB 噪声. 其次, 判别器通过学习基本的卷积网络以区分人脸图像和随机噪声. 接着, 生成器学习正确的偏差 (肤色) 和人脸图像来混淆判别器. 最后, 判别器变得更加适应真正的面部特征, 以区分来自生成器的模型生成的图像和真实人脸图像. 此外, 判别器通过学习条件信息来寻找图像中的关键点做出判别. 实验证明, 在 Wild 数据集上条件信息可以用来准确控制生成器的输出, 即可通过条件信息来控制人脸属性, 实现人脸属性可控性.

另外, Age-cGAN 同样利用 CGAN 进行人脸图像生成[29], Age-cGAN 能够生成人脸的各个不同年龄的图像, 并取得了显著的进展. 在应用 Age-cGAN 时, 需要给出一个当前任务年龄和目标年龄作为条件信息, 其基本结构如图 11 所示.

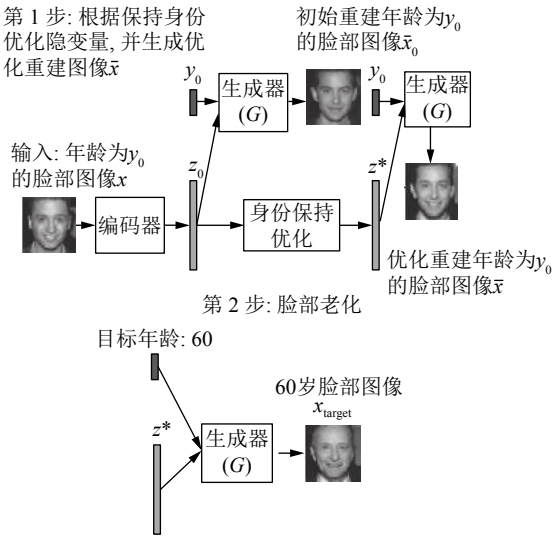


图 11 Age-cGAN 原理图

Fig. 11 Principle of Age-cGAN

按照 CGAN 的训练方法进行模型训练, 当 Age-cGAN 训练完成后, 执行以下两个步骤来进行脸部老化:

1) 给定年龄 y_0, 输入面部图像 x, 找到最佳隐向量 z^*, 如图 11 第 1 步中所示, 生成尽可能接近最初始面部的重构面部表示 $\bar{x} = G(z^*, y_0)$.

2) 给定目标年龄 y_{target}, 通过切换生成器输入的年龄, 如图 11 中的第 2 步中所示, 生成结果面部图像 $x_{\text{target}} = G(z^*, y_{\text{target}})$.

Age-cGAN 使用了 DCGAN 的生成器和判别器来构造模型. 其中, "身份保持"隐向量优化方法, 能够在重建过程中保留原始的身份信息. "身份保持"隐向量优化方法通过一个能够识别身份的人脸识别神经网络 FR 来对 x 和 $\bar{x}$ 的身份进行识别, 并通过计算 $FR(x)$ 和 $FR(\bar{x})$ 之间的欧氏距离, 最小化该距离来提升重建图像效果

$$z^* = \arg\min_z \|FR(x) - FR(\bar{x})\|_{L_2} \tag{37}$$

这种方法具有普遍意义, 不仅适用于面部老化学习,

还适用于其他面部改变的学习.

在上述两种模型中, 基于 CGAN 的人脸识别模型比以往的传统 GAN 具有更低的负对数似然, 而 Age-cGAN 中采用 OpenFace 软件 (用于检测两张照片是否属于同一人) 对 “身份保持”隐向量优化方法进行了评估, 其具有比逐像素优化方法更高的 FR (“OpenFace” face recognition) 分数, FR 分数提高到 82.9%.

类似于 Age-GAN, Li 等[30] 提出了一种基于 CGAN 的利用全局和局部特征生成具有年龄跨度图像的 GAN 模型 (Global and local consistent age GAN, GLCA-GAN). GLCA-GAN 由两部分组成:

1) 通过全局网络学习整个面部结构并模拟整个面部的老化趋势;

2) 3 个关键面部切片被 3 个局部网络预先进行特征压缩提取, 旨在模仿关键面部部位的细微变化.

除此之外, 通过学习残差面部图像 (定义为输入面部图像与其对应的生成面部图像之间的差异) 保留了与年龄属性无关区域中的大部分细节, 并采用身份保持损失函数和年龄保持损失函数以更好地保持身份信息和提高不同年龄合成的准确性, 另外还采用像素损失函数来保存面部输入的详细信息.

2.1.2　多姿态人脸图像生成

多姿态人脸图像生成是从单一侧的人脸图像生成完整的正面人脸图像, 也可称为人脸转正. 本文中主要介绍两种人脸转正 GAN: 通过两条路径进行重新图像合成的双路径生成对抗网络和具有独特姿态编码的表示解析生成对抗网络.

1) 双路径生成对抗网络

双路径生成对抗网络 (Two-pathway GAN, TP-GAN)[31] 将 GAN 应用于人脸转正, 人脸转正是从单一侧脸图像合成为高清的正面人脸图像的技术. TP-GAN 的提出受人类对物体进行视图合成过程的启发: 首先, 人通过以往的经验就物体进行大概的草图描绘, 这个过程可衍生为 TP-GAN 中的全局路径 (Global pathway). 其次, 人需要对该草图中的细节进行特征填充, 该过程可理解为对一个图像中局部不同区域的特征提取, 这个过程可衍生为 TP-GAN 中的局部路径 (Local pathway). 在人脸识别中, 人脸的五官会有不同特征但是又具有一致性, 因此 TP-GAN使用多条局部路径的 CNN 学习到图像的局部特征, 使用全局路径的 CNN 学习到图像的全局特征.

如图 12 所示, TP-GAN 的生成器通过全局路径和局部路径的双路 CNN 建模合成函数 G_{θ_G} (由参数 θ_G 决定), 每条路径均包含编码器 E 和解码器 D, 分别记为 $\{G_{\theta_E^g}, G_{\theta_D^g}\}$ 和 $\{G_{\theta_E^l}, G_{\theta_D^l}\}$, 其中 g 和 l 分别表示全局结构路径和局部纹理路径, 如图 12 上半部分所示. 在全局路径中, $G_{\theta_E^g}$ 的输出是瓶颈 (bottleneck) 层, 瓶颈层用交叉熵损失函数 $L_{\text{cross-entropy}}$ 来实现分类任务. I_F 代表真实正脸图像, 而 I_P 是在不同姿态下的人脸图像, 任务是从不同姿态下的人脸图像恢复成正脸图像. 其中 $\{I_F, I_P\} \in W \times H \times C$, C 为颜色通道.

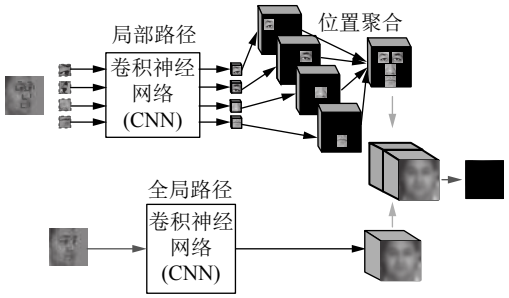


图 12　TP-GAN 生成器结构原理图

Fig. 12　Principle of TP-GAN's generator

在图 12 的局部路径网络中, 为了有效整合来自全局和局部路径的信息实现特征融合, 首先将 4 条局部路径的输出特征张量融合成一个与全局特征张量具有相同空间分辨率的单特征张量. 然后, 简单地把每条路径的特征张量拼接在一起, 产生一个融合的增广特征张量, 然后将其馈送到卷积层, 以产生最终的合成输出. 通过对 G_{θ_G} (由参数 θ_G 决定) 与 D_{θ_D} (由参数 θ_D 决定) 进行交替训练, 优化以下极小–极大问题:

$$\min_{\theta_D}\max_{\theta_G}\left(\mathrm{E}_{I^F\sim P(I^F)}\log D_{\theta_D}(I^F)+\mathrm{E}_{I^P\sim P(I^P)}\log(1-D_{\theta_D}(G_{\theta_G}(I^P)))\right) \tag{38}$$

TP-GAN 的总损失函数 L_{syn} 由以下几部分组成:

a) 像素损失函数

$$L_{\text{pixel}}=\frac{1}{W+H}\sum_{x=1}^{W}\sum_{y=1}^{H}\left|I_{x,y}^{\text{pred}}-I_{x,y}^{\text{gt}}\right| \tag{39}$$

其中, $I^{\text{pred}}=G_{\theta_G}(I^P)$ 为预测图像.

b) 对称损失函数

$$L_{\text{sym}}=\frac{1}{\frac{W}{2}\times H}\sum_{x=1}^{\frac{W}{2}}\sum_{y=1}^{H}\left|I_{x,y}^{\text{pred}}-I_{W-(x-1),y}^{\text{pred}}\right| \tag{40}$$

c) 对抗损失函数

$$L_{\text{adv}}=\frac{1}{N}\sum_{n=1}^{N}-\log D_{\theta_D}(G_{\theta_G}(I_n^P)) \tag{41}$$

d) 身份保持损失函数 (Identity preserving loss)

$$L_{\text{ip}} = \sum_{i=1}^{2} \frac{1}{W_i \times H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \left| F(I^P)^i_{x.y} - F(G(I^{\text{pred}}))^i_{x,y} \right| \tag{42}$$

其中, W_i 和 H_i 代表倒数第 i 层空间的维数.

e) 总损失函数

$$L_{\text{syn}} = L_{\text{pixel}} + \lambda_1 L_{\text{sym}} + \lambda_2 L_{\text{adv}} + \lambda_3 L_{\text{ip}} + \lambda_4 L_{\text{tv}} \tag{43}$$

其中, L_{tv} [32] 是对总损失函数进行变分正则化, 使输出图像比较平滑.

TP-GAN 通过最小化特殊设计的 $L_{\text{cross-entropy}}$ 和 L_{syn} 的加权组合损失函数来优化网络参数 θ_G. 对于具有 N 个训练对的训练集 $\{I_n^F, I_n^P\}_{n=1}^N$, TP-GAN 优化问题为

$$\hat{\theta}_G = \frac{1}{N} \arg\min_{\theta_G} \sum_{n=1}^{N} \left\{ L_{\text{syn}}(G_{\theta_G}(I_n^P), I_n^F) + \alpha L_{\text{cross-entropy}}(G_{\theta^g{}_E}(I_n^P), y_n) \right\} \tag{44}$$

其中, α 是权衡参数.

TP-GAN 在较大视角姿态范围内的人脸识别上取得不错的结果. 如: 大多数以往的人脸转正技术只能在±60°的姿势范围内生成正脸图像. 但是 TP-GAN 能从更大的角度还原正脸图像并取得不错的效果, 如图 13 所示.

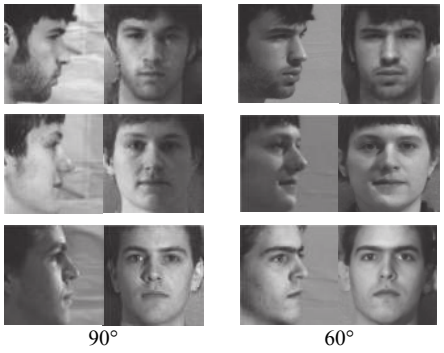


图 13 TP-GAN 实验效果图

Fig. 13 Experiment results of TP-GAN

2) 表示解析学习生成对抗网络

表示解析学习生成对抗网络 (Disentangled representation learning GAN, DR-GAN)[33] 类似于 TP-GAN,也是 GAN在人脸转正领域上的应用. DR-GAN 的核心思想是将 GAN 和身份表示进行结合, 可以从判别器看出, 判别器不仅用来判别生成图像还能够生成身份表示 (ID) 和姿态码 c. 该模型引入了一个姿态码 c 作为生成模型生成器的输入并参考了对抗自编码器的思想. DR-GAN 中生成器将脸部图像、姿势码 c 和随机噪声向量 z 作为输入, 目标是生成与可以愚弄判别器的具有相同身份的目标姿势人脸. DR-GAN 可以学习一个不变的姿态表示 (c)和其他变化身份表示 (ID), 这对于实现姿态不变人脸识别 (Pose-invariant face recognition, PIFR) 目标是理想的模型.

DR-GAN 模型有两种:

a) 单图像输入 DR-GAN

首先, 给定一个带有标签 $y = \{y_d, y_p\}$ 的输入图像 x, y_d 代表身份, y_p 代表姿态. 单图像输入的 DR-GAN 的目的主要有两个:

i) 学习 PIFR 的姿态不变身份表示;

ii) 合成具有相同身份 y_d, 但可由姿态码 c 生成指定的不同姿态的人脸图像 $\hat{x}$.

单图像输入 DR-GAN 的结构如图 14 所示, G_{enc} 是生成编码器, 用于提取隐向量特征. G_{dec} 是生成解码器, 用于重构图像. 判别器是一个由两部分组成的多任务 CNN: $D = [D^d, D^p]$. D^d 用于身份分类, 共 $N^d + 1$ 类. D^p 用于姿态分类, 共 N^p 类. $\hat{x} = G(x, c, z)$ 表示生成图像. 所以, 判别器和生成器的损失函数分别为

$$\max_D V_D(D,G) = \mathrm{E}_{x,y\sim p_d(x,y)}\left[\log D^d_{y^d}(x) + \log D^p_{y^p}(x)\right] + \mathrm{E}_{\substack{x,y\sim p_d(x,y),\\ z\sim p_z(z), c\sim p_c(c)}}\left[\log(D^d_{N^d+1}(G(x,c,z)))\right] \tag{45}$$

$$\max_G V_G(D,G) = \mathrm{E}_{\substack{x,y\sim p_d(x,y),\\ z\sim p_z(z), c\sim p_c(c)}}\left[\log\left(D^d_{y^d}\left(G\left(x_1,\cdots,x_n,c,z\right)\right)\right) + \log\left(D^p_{y^t}\left(G\left(x,c,z\right)\right)\right)\right] \tag{46}$$

其中, D_i^d 和 D_i^p 是 D^d 和 D^p 中的第 i 个元素.

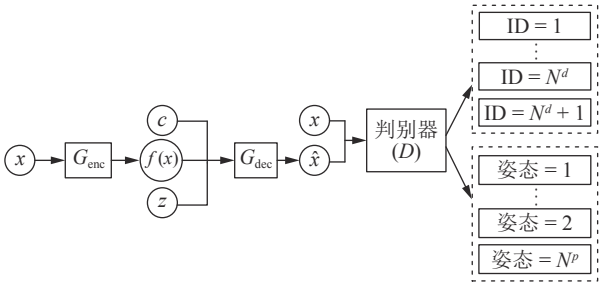


图 14 DR-GAN 结构图 (单图像)

Fig. 14 The structure of DR-GAN (single image)

b) 多图像输入 DR-GAN

单图像输入 DR-GAN 是通过处理单个输入图像 x 来提取身份并对脸部图像进行旋转, 而多图像

输入 DR-GAN 通过多个图片作为输入 $(x_1, x_2, \cdots, x_n)$ 提取出更多的身份信息. 在多图像输入 DR-GAN 的生成器结构中, 如图 15 所示, G_{enc} 作用不仅是提取隐向量特征, 同样需要估计置信系数 ω. 多图像输入 DR-GAN 只是在生成器上做了一些变化, 对多个 G_{enc} 的输出进行整合

$$f(x_1, x_2, \cdots, x_n) = \frac{\sum_{i=1}^{n} \omega_i f(x_i)}{\sum_{i=1}^{n} \omega_i} \tag{47}$$

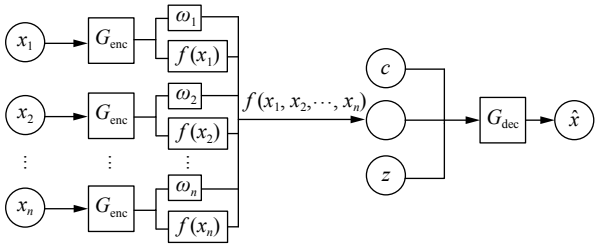


图 15　DR-GAN 生成器结构图 (多图像)

Fig. 15　The structure of DR-GAN's generator (mutiple image)

由于与单图像输入的区别只是有多个输入, 所以判别器的损失函数与单图像输入的完全相同. 但是生成器的损失函数有所不同, 如下所示:

$$\begin{aligned}\max_G\ & V_G(D,G) = \\ & \sum_{i=1}^{n}\Big[\mathrm{E}_{\substack{x_i,y_i\sim p_d(x,y),\\ z\sim p_z(z),c\sim p_c(c)}}[\log(D^d_{y^d}(G(x_i,c,z)))+ \\ & \log(D^p_{y^t}(G(x_i,c,z)))]\Big]+ \\ & \mathrm{E}_{\substack{x_i,y_i p_d(x,y),\\ z\sim p_z(z),c\sim p_c(c)}}\Big[\log(D^d_{y^d}(G(x_1,\cdots,x_n,c,z)))+ \\ & \log(D^p_{y^t}(G(x_1,\cdots,x_n,c,z)))\Big]\end{aligned} \tag{48}$$

其中, 多图像输入 DR-GAN 中 $\hat{x} = G(x_1, \cdots, x_n, c, z)$ 为生成器输出图像. DR-GAN 在 Multi-PIE, CFP 和 IJB-A 数据库上都获得了不错的人脸识别性能.

在上述两个模型的实验比较中, 在大角度学习场景, TP-GAN 具有比 DR-GAN 更好的实验效果, 将 DR-GAN 的 Rank-1 识别率提高了 5% ~ 10%, 并且两个方法在大角度识别方面 ($\geq 60°$) 均具有比基于 CNN 以及基于 LDA (Latent dirichlet allocation) 方法更高的 Rank-1 识别率.

2.1.3　图像纹理合成

视觉纹理合成的目的是, 从一个视觉样本纹理中归纳出该视觉样本的纹理特征, 用来生成具有该种纹理的任意的新的图像.

空间生成对抗网络 (Spatial GAN, SGAN)[34] 是一种基于 GAN 的纹理合成模型. SGAN 通过将输入噪声分布空间从单个向量扩展到整个空间张量, 创建了一个非常适合纹理合成任务的结构.

SGAN 的结构如图 16 所示, 生成器通过 DCGAN 中的微步幅卷积层将一个空间噪声阵列 $Z \in \mathbf{R}^{l\times m\times d}$ 转换为 RGB 图像 $X \in \mathbf{R}^{h\times w\times 3}$, 其中, l 和 m 为空间维数, d 为信道数. 判别器的输入为生成图像 X 或通过对真实图像 I 进行矩形补丁提取的 X'. SGAN 的目标函数为

$$\begin{aligned}\min_G \max_D\ & V(D,G) = \\ & \frac{1}{lm}\sum_{\lambda}^{l}\sum_{\mu}^{m}\mathrm{E}_{Z\sim p_Z(Z)}[\log(1-D_{\lambda\mu}(G(Z)))]+ \\ & \frac{1}{lm}\sum_{\lambda}^{l}\sum_{\mu}^{m}\mathrm{E}_{X'\sim p_{\text{data}}(X)}[\log D_{\lambda\mu}(X')]\end{aligned} \tag{49}$$

其中, $\lambda \in [1, l]$ 和 $\mu \in [1, m]$ 是位置坐标.

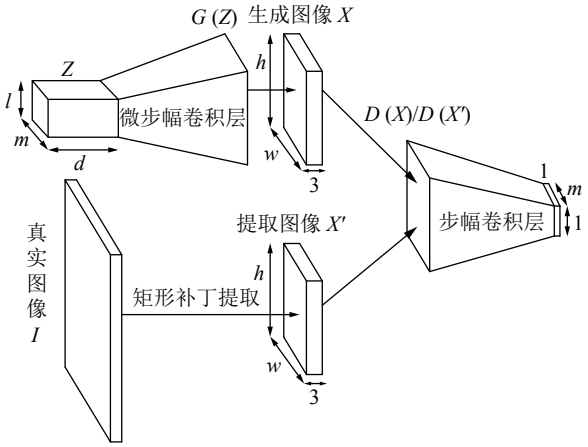


图 16　SGAN 结构原理图

Fig. 16　The structure of SGAN

这是第一个基于 GAN 的完全数据驱动的纹理合成方法, SGAN 的优点在于:

1) 能够生成高质量的纹理图像;

2) 输出纹理合成图像的大小具有非常高的可扩展性;

3) 生成纹理合成图像的速度较快;

4) 能够在复杂纹理中融合多个不同源图像的能力.

除此之外, 根据马尔科夫原理的马尔科夫生成对抗网络 (Markovian GAN, MGAN)[35], 解决了深度神经网络的计算复杂性问题, MGAN 预先通过前

馈的、跨步幅的卷积网络进行计算, 而不是先前研究工作中使用的数值反卷积运算. MGAN 能够捕获马尔科夫斑块的统计特征, 生成任意维度的输出, 并在生成器和判别器上巧妙运用了 VGG19 和 RELU 函数等技术, 能够有效实现图像纹理合成. MGAN 在对图像进行纹理合成方面, 具有与 Texture network[36] 相同的合成速度.

除此之外, BigGAN[37] 的提出再次刷新了人类的认知, 其生成的图片的纹理和背景都十分逼真, BigGAN 将 IS 分数提高到 166 分 (真实图片的 IS 分数为 233 分), 在 FID 指标上, 也实现了大幅超越. BigGAN 训练过程使用了超大的每批次样本个数 (2 048), 使其生成图片的性能大幅提升. 文中还采取了截断技巧保证了生成图像的平滑性, 以及对生成器和判别器的稳定性控制方法, 但是 BigGAN 的算法复杂性很高.

2.2 图像超分辨率

图像超分辨率技术 (Super-resolution) 是指从观测到的低分辨率图像重新生成出对应的高分辨率图像的技术. 这项技术在监控设备、卫星图像和医学影像等领域都有重要的应用价值.

超分辨率主要分为两类: 首先是从多幅低分辨率图像重建出高分辨率图像; 另一个是从单幅低分辨率图像重建出高分辨率图像 (Single image super-resolution, SISR). 在 GAN 应用中, SR 主要是基于单幅低分辨率的重建方法.

尽管深度卷积神经网络在图像超分辨率领域的生成精度与速度方面取得了巨大进步, 但是仍然不能解决当图像经过大幅放大时, 如何保持图像的纹理细节问题. 因此, 针对传统超分辨率方法中存在的过于平滑的问题, 提出了图像超分辨率生成对抗网络 (Super resolution GAN, SRGAN)[38], 利用参数化的残差网络作为生成器, 用 VGG 网络作为判别器, 并且得到了不错的纹理细节学习效果.

SRGAN 的目标是: 训练一个生成函数 G, 使其能够预测给定的输入低分辨率 (Low resolution, LR) 图像的高分辨率 (High resolution, HR) 部分. 为了达到这个目的, 通过优化一个与图像分辨率有关的损失函数 l^{SR}, 训练一个前向卷积神经网络 G_{θ_G} 作为生成器, G_{θ_G} 的参数 $\theta_G = \{W_{1:L}; b_{1:L}\}$ 由一个 L 层深度网络的权重和偏置组成, 对于一个给定的训练图像 $I_n^{\mathrm{HR}} \in \mathbf{R}^{rW \times rH \times C}$, $n = 1, \cdots, N$, 设与 I_n^{HR} 对应的低分辨率图像为 $I_n^{\mathrm{LR}} \in \mathbf{R}^{W \times H \times C}$, 其中, r 定义为在训练中将 I^{HR} 通过高斯滤波器进行下采样得到 I^{LR} 的下采样因子. 因此生成器需要优化的问题为

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{\mathrm{SR}}(G_{\theta_G}(I_n^{\mathrm{LR}}), I_n^{\mathrm{HR}}) \tag{50}$$

进而可以得到 SRGAN 的总损失函数为

$$\min_{\theta_G} \max_{\theta_D} \mathrm{E}_{I^{\mathrm{HR}} \sim p_{\mathrm{train}}(I^{\mathrm{HR}})}[\log D_{\theta_D}(I^{\mathrm{HR}})] + \mathrm{E}_{I^{\mathrm{LR}} \sim p_G(I^{\mathrm{LR}})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{\mathrm{LR}})))] \tag{51}$$

SRGAN 的创新点在于与图像分辨率有关的 l^{SR} 的设计, 将逐像素损失替换为内容损失. SRGAN 提出的与图像分辨率有关的 l^{SR} 由以下两部分加权组成:

1) 图像内容损失函数

$$l^{\mathrm{SR}}_{\mathrm{VGG}/i,j} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{\mathrm{HR}})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{\mathrm{LR}})_{x,y})^2 \tag{52}$$

其中, $\phi_{i,j}$ 为 VGG19 网络中在第 i 个最大池化层之前的第 j 次卷积得到的特征映射. $W_{i,j}$ 和 $H_{i,j}$ 表示 VGG 中各特征映射的维数. $l^{\mathrm{SR}}_{\mathrm{VGG}/i,j}$ 把某一层的特征映射的逐像素损失作为内容损失 (而不是最后输出结果的逐像素损失), 并可使得低维特征 $\phi_{i,j}(G_{\theta_G}(I^{\mathrm{LR}})_{x,y}$ 和高维特征 $\phi_{i,j}(I^{\mathrm{HR}})_{x,y}$ 保持一致.

2) 对抗损失函数

$$l^{\mathrm{SR}}_{\mathrm{Gen}} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{\mathrm{LR}})) \tag{53}$$

在这个对抗损失函数中, 用负对数求和替换原来的损失函数 $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{\mathrm{LR}}))]$, 类似于 GAN 中的非饱和训练方法.

综合以上对目标损失函数的改进, SRGAN 在图像超分辨率领域取得了不俗的效果. 如图 17 所示, 左侧为应用双三次插值法得到的图像超分辨率

双三次插值 SRGAN

图 17 SRGAN 实验效果

Fig. 17 Experiment result of SRGAN

结果; 右侧为应用 SRGAN 得到的图像超分辨率结果. 其中, 将 SRGAN 以及文献 [38] 中提出的 SRResNet, 与近邻采样 (Nearest neighbor, NN) 以及双三次插值法进行比较 (这是目前存在的 4 种最先进的超分辨率方法), 其中 SRGAN 与 SRResNet 相比较, 在 PSNR、SSIM 和 MOS 指标上都具有更高的分数.

CycleGAN 是一个能够学习将不成对的图像数据从图像域 X 映射到另一个图像域 Y 的应用对偶原理的变种 GAN 模型. 其中对 CycleGAN 的扩展 c-CycleGAN[39], 使得从 X 到 Y 的映射受到属性条件 Z 的影响. 若以人脸验证网络提取的人脸特征向量作为 Z, c-CycleGAN 方法在保持人脸图像超分辨率方面具有有效性.

另外, 对 WGAN-GP 与其他 GAN 模型在 DCGAN, ResNet 以及 MLP (Multi-layer perceptron) 上进行实验[40], 结果表明 WGAN 对 SISR 具有有效性.

2.3 图像复原与多视角图像生成

1) 图像复原

在实际应用中, 图像常常会被噪声损坏, 或是一些不可抗力因素导致图像受到破坏. 图像修复问题就是还原图像中缺失损坏的部分, 也就是基于图像中已有的信息, 来复原图像中的缺失部分.

一般来讲, 由于缺乏高级上下文信息, 仅从单幅图像提取信息实现图像复原的现有方法通常产生的图像复原结果并不令人满意. 因此, 文献 [41] 提出了利用 GAN 进行语义图像修复的方法. 文中利用 GAN 网络, 并将损失函数定义为上下文损失函数与先验损失函数之和. 实验结果表明 GAN 在图像复原上具有有效性.

Demir 等[42] 提出的 PGGAN (Patch-global GAN) 将全局生成对抗网络 (Global GAN, G-GAN)[43] 的判别器和在 Pix2Pix 模型中的基于 Markovian 方法的补丁生成对抗网络 (Patch GAN) 的判别器进行组合成为一个新的判别器. PGGAN的生成器则由一个 ResNet 网络组成. 实验结果表明, PGGAN与目前的图像复原方法相比, 生成图像在质量和视觉上会有更好的效果. PGGAN 具有比基于 CNN和 GAN 的 CE (Context encoders) 模型[44] 更低的 $L1$ 和 $L2$ 损失, 以及更高的 PSNR 和 SSIM 分数.

2) 多视角图像生成

多视角图像生成是从单个输入的视图生成多视角的图像. 当看到单张图片时, 就能知道这个物体的 3D 视角图像. 那么如何解决这个从复杂的 2D 图像到 3D 图像的推理任务呢? 为了实现这个任务, 文献 [45] 提出了一种新的图像生成模型, 它结合了变分推理和 GAN 的优势, 称为 VariGAN. VariGAN模型是以粗到细的方式生成目标图像. 第 1 步, 通过变分推理对输入图像的整体外观 (例如, 形状和颜色) 进行模拟并生成其多视角的粗略图像; 第 2 步, 对生成的多视角粗略图像进行超分辨率操作, 生成更加精细丰满的图像.

VariGAN 在 MVC 和 DeepFashion 数据集上具有比条件 VAE (Conditional VAE, cVAE)[46] 以及 CGAN更高的 IS 和 SSIM 分数.

2.4 图像转换

图像转换研究的问题是对图像的风格进行转换. 图像的风格是一个十分抽象的概念. 对于人眼来说, 能够有效地辨别出不同画家不同流派绘画的风格. 但是在计算机眼中, 图像的风格本质上就是一些像素, 通过多层网络找出更复杂、更内在的特征. 所以图像的风格理论上可以通过多层网络来提取图像里面可能含有的一些与风格有关的显著的特征来获取.

1) DualGAN 实现域自适应图像风格转换

尽管用于跨域图像到图像转换的 CGAN 最近取得了很大进展, 但是当面对大量未标记的图像时, 由于人类标注成本昂贵, 这时可以说 CGAN 是无法应用的. 受自然语言翻译中的对偶学习的启发, 利用对偶学习的 DualGAN[47] 模型实现了跨域图像生成. DualGAN 仅使用未标记的数据, 就能比 CGAN 生成更好的图像输出. 另一方面, 对于涉及基于语义的标签学习任务, DualGAN 比 CGAN 表现优异.

如图 18 所示, 假定 U 和 V 分别为源域和目标域, z, z' 是随机噪声, 源域上的图像 $u \in U$ 通过源域生成器 G_A 转换到目标域 V, 得到 $G_A(u, z)$. 使用目标域 V 中的判别器 D_A 对 $G_A(u, z)$ 进行判别, 然后使用生成器 G_B 将 $G_A(u, z)$ 转换回源域 U, 其输出 $G_B(G_A(u, z), z')$ 作为 u 的重构版本. 相同地, 目标域上的图像 $v \in V$ 被生成器 G_B 转换到源域 U, 从而得到 $G_B(v, z')$, 然后使用生成器 G_A 将 $G_B(v, z')$ 转换回目标域 V, 其输出 $G_A(G_B(v, z'), z)$ 作为 $v \in V$ 的重构版本. 判别器 D_A 用 v 作为正样本、$G_A(u, z)$ 作为负样本进行训练, 同样 D_B 取 u 为正样本、$G_B(v, z')$ 为负样本. 生成器 G_A 和 G_B 被优化用来生成负样本输出以迷惑相应的判别器 D_A 和 D_B, 并且最小化两个重构损失 $\|G_A(G_B(v, z'), z) - v\|$ 和

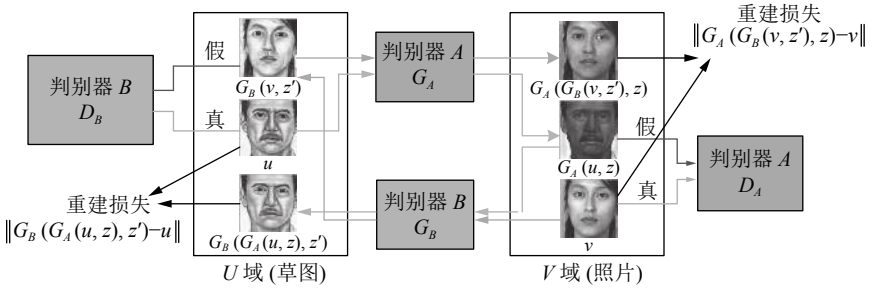


图 18 DualGAN 结构

Fig. 18 The structure of DualGAN

$\|G_B(G_A(u,z),z')-u\|$.

DualGAN 由于缺乏标签的条件信息, 使其图像风格转换效果不如 CGAN 好. 所以对这种基于对偶学习的模型来说, 条件信息的加入或许可使其具有更好的效果. 除此之外, DualGAN 对比于传统的 GAN 模型, 图像转换效果只有微弱的提升.

2) 图像中的对象替换

除了 DualGAN 外, GeneGAN[48] 也可用于将图像中的对象替换为另一幅图像中的另一个对象. 由于人的表情具有不同的风格, 因此若想在另一幅人脸图像上实现表情变化, 需要大量配对图片进行训练. GeneGAN 的创新点在于, 它通过对输入两个图片的表情进行特征提取, 并对特征编码进行重新组合, 解码生成新的图片, GeneGAN 可以使用不成对的图片完成生成式模型学习任务. 它的结构如图 19 所示.

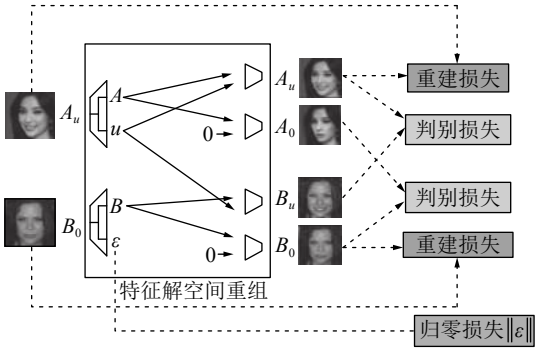


图 19 GeneGAN 训练过程

Fig. 19 Training process of GeneGAN

GeneGAN 的工作原理是: 首先, 从单幅图像生成一个对象特征向量. 接着, 将对象特征移植到其他图像以生成具有类似对象的新颖图像. 所以该模型由两部分组成: 将图像分解为背景特征部分和对象特征部分的编码器, 以及可以组合背景特征和对象特征以产生图像的解码器. 编码–解码器相当于 GAN 中的生成器, 对图像进行重新组合和重建, 再通过右侧的判别损失 (判别器) 进行判别.

如图 19 所示, 其中编码特征 A, u, B, ε 以及由特征重组的图像 $x_{A0}, x_{Bu}, x'_{Au}, x'_{B0}$ 由以下过程组成:

$$\begin{aligned} &(A,u)=Encoder(x_{Au}), \quad (B,\varepsilon)=Encoder(x_{B0})\\ &x_{A0}=Decoder(A,0), \quad x_{Bu}=Decoder(B,u)\\ &x'_{Au}=Decoder(A,u), \quad x'_{B0}=Decoder(B,0) \end{aligned} \tag{54}$$

其中, 强制将 x_{B0} 编码的 ε 为零, 以确保 x_{A0} 不应包含关于 x_{B0} 的任何信息的约束, 并且可以安全地丢弃 x_{B0} 的对象部分中包含的信息. 判别器为传统 GAN 的判别器结构, 而生成器的损失函数由以下 4 部分组成:

a) 标准 GAN 损失函数 L_{GAN}, 其度量生成的图像的真实性.

$$\begin{aligned} L^0_{\text{GAN}}&=-\text{E}_{z\sim P_{=0}}[\log D(x_{A0},z)]\\ L^{\neq 0}_{\text{GAN}}&=-\text{E}_{z\sim P_{\neq 0}}[\log D(x_{Bu},z)] \end{aligned} \tag{55}$$

其中, $P_{=0}, P_{\neq 0}$ 分别代表没有对象的图像分布和有对象的图像所服从的概率分布.

b) 重构损失函数 $L_{\text{reconstract}}$, 它的作用是度量在一系列编码和解码之后重建原始输入的程度.

$$\begin{aligned} L^{Au}_{\text{reconstruct}}&=\|x_{Au}-x'_{Au}\|_1\\ L^{B0}_{\text{reconstruct}}&=\|x_{B0}-x'_{B0}\|_1 \end{aligned} \tag{56}$$

c) 归零损失函数 L_0, 由于将编码特征 ε 强制置零, 因此 L_0 反映了物体特征与背景特征的分离程度, 即编码特征 ε 的一范数.

$$L_0=\|\varepsilon\|_1 \tag{57}$$

d) 平行四边形损失函数 $L_{\text{parallelogram}}$, 其在图像像素值中强制引入子代与父代之间的约束, 这是一个可选的损失函数.

$$L_{\text{paralelogram}}=\|x_{Au}+x_{B0}-x_{A0}-x_{Bu}\|_1 \tag{58}$$

GeneGAN 通过提取图像的编码特征, 生成子代与父代图像的生成模式, 是一个新颖的图像转换模型. 为以后生成逼真且更具多样性的图像奠定了

基础.

3) 由结构 GAN (Structure-GAN) 和风格 GAN (Style-GAN) 组成的网络 S²-GAN

传统端到端的图像生成方法会忽略图像中两个比较重要的因素: 结构和风格. 结构指的是图像中实体的 3D 结构; 风格则是在结构的基础上进行的纹理合成. S²-GAN 结合了这两个重要因素, 因此, S²-GAN[49] 由两个 GAN 网络组成, 一个是 Structure-GAN, 另一个是 Style-GAN.

RGB 图像 $X=(X_1,\cdots,X_M)$ 以及与 RGB 对应的曲面法线贴图 $C=(C_1,\cdots,C_M)$, 从均匀噪声分布中采样的 $\tilde{Z}=(\tilde{z}_1,\cdots,\tilde{z}_M)$ 和 $\hat{Z}=(\hat{z}_1,\cdots,\hat{z}_M)$. 其基本结构如图 20 所示. Structure-GAN 从 $\hat{z}$ 中采样生成一个曲面法线贴图 (Surface normal maps); 再由 Style-GAN 将曲面法线贴图与另一个隐向量 $\tilde{z}$ 作为输入并生成一个 2D 图像.

a) Structure-GAN

Structure-GAN 中的生成器由 10 层网络组成. 从给定的 100 维 $\hat{z}$ 作为输入完全连接到一个 $9\times9\times64$ 的 3D 块中, 并进行反卷积操作生成曲面法线贴图. 生成器按照 DCGAN 使用批量归一化, 并在每层之后的 ReLU 激活. 在最后一层, 应用 Tanh 激活. 判别器由 6 层网络组成. 将图像作为输入, 输出单个数字, 该数字预测输入曲面法线贴图是真实的还是生成的. 将 DCGAN 中的 LeakyReLU 作为激活函数. 但是不在此处使用批量归一化.

b) Style-GAN

Style-GAN 是给定来自 Kinect 的 RGB 图像和曲面法线贴图, 并以表面法线为条件生成图像的 GAN. 生成器为 CGAN. 条件信息为曲面法线贴图, 作为生成器和判别器的附加输入. 增加曲面法线作为判别器的附加输入不仅可以使生成图像看起来更真实, 而且还可使生成图像与曲面法线贴图相匹配. 在训练判别器时, 只需考虑真实的 RGB图像及其相应的曲面法线作为正例. 从曲面法线中提取更多信息, 使 Style-GAN 能够生成更高分辨率的 $128\times128\times3$ 图像.

Style-GAN 的生成器的输入为 $128\times128\times3$ 曲面法线贴图和 100 维 $\tilde{z}$, 首先它们分别经过卷积和反卷积层, 然后连接形成 $32\times32\times192$ 的特征映射. 并将这个特征映射通过 7 层卷积和反卷积层, 输出是 $128\times128\times3$ 的 RGB 图像. 在判别器的选择上采用了与 Structure-GAN 中的类似架构, 但是判别器的输入是曲面法线贴图和图像 $(128\times128\times6)$ 的串联. Style-GAN 的判别器和生成器损失函数分别为

$$L^D(X,C,\tilde{Z})=\sum_{i=1}^{\frac{M}{2}}L(D(C_i,X_i),1)+\sum_{i=\frac{M}{2}+1}^{M}L(D(C_i,G(C_i,\tilde{z}_i)),0)\tag{59}$$

$$L^G(C,\tilde{Z})=\sum_{i=\frac{M}{2}+1}^{M}L(D(C_i,G(C_i,\tilde{z}_i)),1)\tag{60}$$

其中, $L(y^*,y)=-[y\log(y^*)+(1-y)\log(1-y^*)]$.

通过独立训练 Structure-GAN 和 Style-GAN 之后, 合并所有网络并联合训练它们. 完整模型包括来自 Structure-GAN 的曲面法线生成, 并且基于它通过 Style-GAN 生成图像. S²-GAN 生成的图像具有可解释性, 并且与以往的 DCGAN、GAN 相比能够生成更逼真的图像.

当没有正确配对的数据对应关系时, 实现两组独立的无监督图像学习任务是具有一定困难的. 文献 [50] 提出了一种全新的基于深度注意力机制的 GAN (Deep attention GAN, DA-GAN). DA-GAN 能够在高度结构化的隐空间中将分别来自两个集合的样本的翻译任务分解为翻译实例. DA-GAN 具有一定优越性, 并可广泛应用在其他领域, 例如数据增强、域翻译、姿态变形等. 在 DA-GANS-VHN-MNIST 数据集上进行的域自适应实验表明,

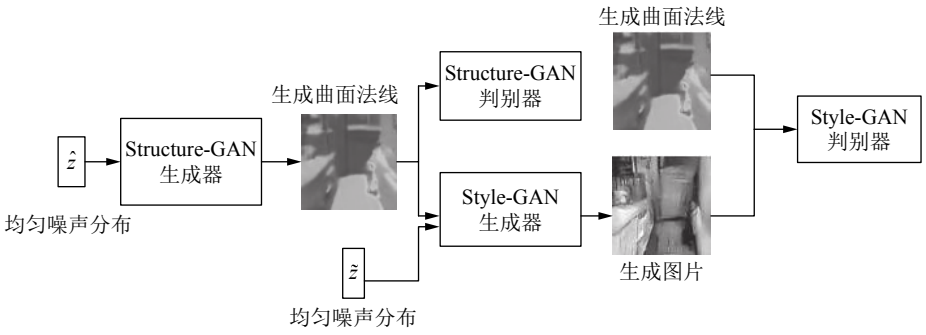


图 20　S²-GAN 结构

Fig. 20　The structure of S²-GAN

DA-GAN 与 UNIT (Unsupervised image-to-image translation)[51] 相比具有更高的准确率.

2.5 文本描述到图像生成

文本描述到图像的生成, 首先对图像进行人为的文本描述, 通过 GAN 生成出文本描述对应的图像. 若根据文本描述来生成一个图像, 产生的结果变化会非常大. 当文本中的某个词改变, 会导致生成的图像中大量像素发生改变. 这种变化是很难发现其中的关联性的. 但是, 文本描述到图像生成的反问题, 图像生成文本描述就没有这样严重的问题, 这是因为文本可以用语言模型建模. 所以利用 GAN 解决文本描述到图像生成的问题是很好的选择.

图像到文本 GAN (Text to image GAN)[52], 受到了 CGAN 和 DCGAN 的启发, 结构如图 21 所示.

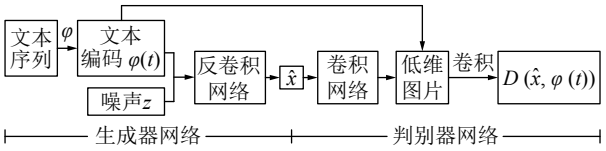


图 21 Text to image GAN 结构

Fig. 21 The structure of text to image GAN

生成器首先将文本序列通过 φ 进行编码并与噪声 z 进行串联作为卷积网络的输入, 由低维度变成高维度, 并生成图像 $\hat{x}$. 在判别器中, 生成图像通过卷积层由高维变为低维与原文本的编码 $\varphi(t)$ 进行串联, 再进行卷积并判别. 这个模型开发一种简单而有效的 GAN 体系结构和培训策略, Text to image GAN 能够将人为文字描述合成为逼真的鸟与花的图像. 不仅提出了文本合成模型, 还提出了流型插值 GAN (Learning with manifold interpolation, GAN-INT)、匹配感知 GAN 判别器 (Matching-aware discriminator, GAN-CLS) 以及反转风格转换的生成器.

1) GAN-INT

GAN-INT 是对训练文本编码进行简单插值, 从而能产生大量的文本编码. 由于通过深度网络学习到的特征表示具有可插值性, 换句话说, 当一对图像经过相同的层后, 对这些特征进行插值, 那么插值后的特征也是在数据所在流形附近. 尽管这样插值产生的数据可能并不存在, 但是这样不仅能够增加数据的多样性还不需要额外的标记成本, 只需在生成器处增加一个额外项

$$\mathrm{E}_{t_1,t_2\sim p_{\text{data}}}[\log(1-D(G(z,\beta t_1+(1-\beta)t_2)))] \quad (61)$$

其中, z 是从噪声分布中采样得到的, β 是在文本嵌入 t_1 和 t_2 之间进行插值.

由于对数据对的插值是伪造的, 判别器并没有对应的数据对来进行训练. 但是, 判别器能够学习到是否当前图像与文本相匹配.

2) GAN-CLS

当传统 GAN 对数据对进行训练时, 判别器的输入有两种: 真实的图像和与其匹配的文本; 以及生成图像和任意的文本. 与传统 GAN 不同, GAN-CLS 增加了第三种输入: 真实的图像和任意的文本. 这样判别器能够提供额外的信息给生成器.

3) 反转风格转换的生成器

如果文本编码 $\varphi(t)$ 捕获图像内容 (例如花朵形状和颜色), 则为了生成逼真的图像, 噪声样本 z 应该捕获诸如背景颜色和姿势的样式因素.

使用训练好的 GAN, 人们可能希望将查询图像的样式转换为特定文本描述的内容. 为了达到这个目的, 可以训练一个卷积网络来反演生成器从样本 $\hat{x}\leftarrow G(z,\varphi(t))$ 回归到 z 上. 使用简单的平方损失来训练样式编码器

$$L_{\text{style}}=\mathrm{E}_{t,z\sim N(0,1)}\|z-S(G(z,\varphi(t)))\|_2^2 \quad (62)$$

其中, S 是风格编码器网络. 使用训练有素的生成器和样式编码器, 从查询图像 x 到文本 t 的样式转换按如下方式进行:

$$s\leftarrow S(x),\hat{x}\leftarrow G(s,\varphi(t)) \quad (63)$$

其中, $\hat{x}$ 是结果图像, s 是预测风格.

将 GAN-INT 与 GAN-CLS 方法进行结合的新方法为 GAN-INT-CLS. 在文本描述到图像生成的实验中, GAN-INT-CLS 和 GAN-INT 比传统 GAN 实验结果更好.

GAWWN (Generative adversarial what-where network)[53] 能够对对象的位置姿势进行控制, 即给出了文字描述, 说明在图像的哪个位置绘制对象内容. GAWWN 与 GAN-INT-CLS 相比具有更加逼真的图像生成效果.

除此之外, 若将上一节的 DA-GAN 方法用于文本描述到图像生成过程, 将会发现 DA-GAN 与 GAWWN和 GAN-INT 相比, 具有更高的 IS 分数.

当学习任务为对图像进行文字描述, 可以看作是文本描述到图像的反问题, 即从图像来生成文本. 循环主题−过渡 GAN (Recurrent topic-transition GAN, RTT-GAN)[54] 能够通过推理局部语义区域和利用语言知识, 来得到语义丰富且连贯的段落描述, 能有效应用于图像文字描述任务.

2.6 图像语义分割

图像语义分割实现机器自动分割并识别图像中

的内容. 图像语义分割是 AI 和机器视觉技术中关于图像理解的重要一环. 在自动驾驶系统 (街景识别与理解)、无人机应用 (着陆点判断) 以及穿戴式设备应用中发挥着举足轻重的作用. 图像是由许多像素组成, 图像语义分割就是将像素按照图像中表达语义含义的不同进行分组或者分割.

Luc 等[55] 第一次将 GAN 应用到图像分割中. 它的结构如图 22 所示. 左侧是基于 CNN 的分割模型, 相当于生成器, 右侧是一个对抗网络的判别器. 对抗网络的输入有两种情况, 一是原始图像与参考标准; 二是原始图像与分割结果, 它的输出是一个分类值: 1 代表判断输入是第 1 种情况; 0 代表判断输入是第 2 种情况). 实验结果表明, 这种对抗性训练方法可以提高两个数据集 (Standford 背景数据集和 PASCAL VOC 2012 数据集)上的训练集和验证集上的预测准确率.

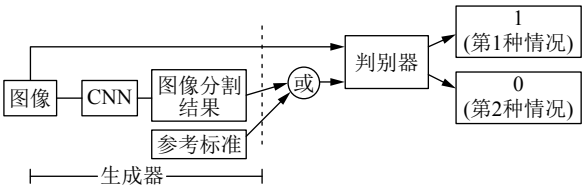


图 22　GAN 应用于图像语义分割

Fig. 22　GAN applied to image semantic segmentation

除此之外, Souly 等[56] 将 GAN 应用在基于半监督学习的图像语义图像分割的应用, 模型利用人工生成的数据和未标记的数据实现半监督基于图像语义的图像分割任务.

Liang 等[57] 提出了一个对比生成对抗网络 (Contrast-GAN) 用作图像到图像的语义操作 (将图像中的对象换成另一个对象), 其中用到了图像语义分割技术.

Contrast-GAN 受到 CycleGAN 的启发, 将 CycleGAN 中的循环一致损失函数 (L_{cyc}) 替换为对比对抗损失函数 (Adversarial contrasting loss). 在不成对的图像语义分割实验中, Contrast-GAN 与 CycleGAN 相比具有更好的翻译效果.

2.7　图像着色

图像自动着色是指对黑白图像进行自动着色的过程. 在视觉效果领域, 如果对一些黑白照片进行更好的渲染, 便可以使宝贵的历史影像资料变得更加具有生命力, 视觉效果更好. Liu 等[58] 提出使用 CGAN 来解决素描图到图像的合成 (自动着色) 问题. 文中称该自动着色模型为自动画家 (Auto-painter) 模型, 它能够兼容不同的颜色并能够绘制出合适的图像, 而且还可以根据用户的需要进行设定.

假定 x 是输入草图, y 是目标 (彩色的卡通图像), $G(x,z)$ 是生成器生成的图像, CGAN 模型作为着色模型. Auto-painter 的生成器的损失函数主要由以下几个部分组成:

$$L_G = \mathrm{E}_{\substack{x\sim p_{\text{data}}(x),\\ z\sim p_{\text{data}}(z)}}[\log(1-D(x,G(x,z)))] \tag{64}$$

其中, z 为高斯噪声.

此外, 引入 L_1 范数正则化项来约束模型中的像素损失

$$L_p = \mathrm{E}_{\substack{x,y\sim p_{\text{data}}(x,y),\\ z\sim p_{\text{data}}(z)}}[\|y-G(x,z)\|_1] \tag{65}$$

其中, L_p 表示生成图像与真实图片之间像素级的差异.

除了考虑像素损失, 还使用预先训练的 VGG 来提取图像中的抽象信息. 定义特征损失函数 L_f 为特征空间中的 L_2 范数在 x , y 以及高斯噪声上的联合概率分布上的数学期望.

$$L_f = \mathrm{E}_{\substack{x,y\sim p_{\text{data}}(x,y),\\ z\sim p_{\text{data}}(z)}}[\|\phi_j(y)-\phi_j(G(x,z))\|_2] \tag{66}$$

其中, ϕ_j 是在 ImageNet 数据集上预先训练的 16 层 VGG 网络 (VGG16) 的第 j 层激活函数的输出.

为了避免输出图像的颜色出现突变, 总的目标函数中还添加了全变差损失函数 L_{tv} . L_{tv} 可以限制生成结果中的像素变化程度, 并促使生成的彩色卡通图像更平滑.

$$L_{\text{tv}} = \sqrt{(y_{i+1,j}-y_{i,j})^2+(y_{i,j+1}-y_{i,j})^2} \tag{67}$$

总结以上结果, 得到总的目标函数为

$$L = w_pL_p + w_fL_f + w_GL_G + w_{\text{tv}}L_{\text{tv}} \tag{68}$$

如图 23 所示, 对自动画家 (Auto-painter) 模型与其他模型之间的生成效果进行比较. Auto-painter 通过主观评测方法对其生成的图片与 Pix2Pix 方法生成的图片相比, 取得了更高的支持率, 说明 Auto-painter 的填充效果更加逼真.

Koo[59] 应用 DCGAN 来处理黑白照片的自动着色问题. 虽然 CNN 能从图像中提取特征, 并进行快速着色. 但是使用 CNN 进行自动着色时, 倾向于将颜色不明确的物体生成棕褐色的色调. 除此之外, Suárez 等[60] 提出了采用 DCGAN 生成近红外显色 (Near infrared, NIR) 图像的方法.

2.8　GAN 的其他计算机视觉应用

2.8.1　视频预测

深度学习在视频预测领域有着非常好的前景,

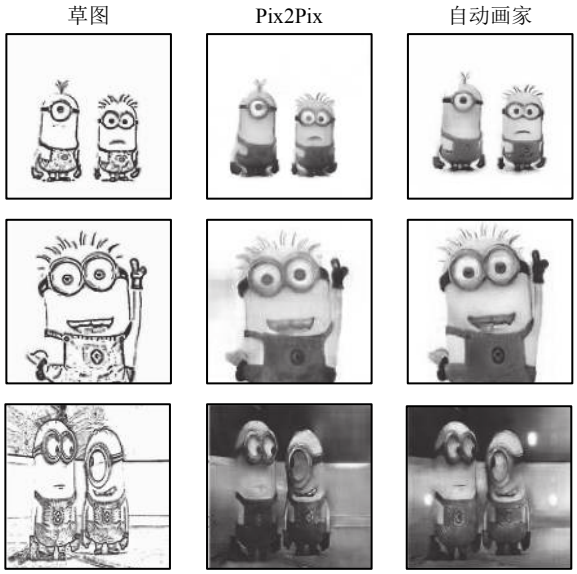


图 23 自动画家模型效果

Fig. 23 Experiment result of auto-painter

例如: 给出一个静态图像, 生成一段小视频. Vondrick 等[61] 首先提出将 GAN 应用于视频生成和视频预测, 实际上是对视频进行帧预测. 给出视频的前 4 帧, 生成器会生成视频的下一帧. 并通过判别器判别下一帧图像是否为参考图像.

另外, Vondrick 等[62] 利用 3D 卷积神经网络作为生成器, 并利用大量的无标签视频数据训练 GAN, 从而得到可以生成视频序列的模型.

Liang 等[63] 提出一个对偶运动 GAN (Dual motion GAN) 模型, 通过对偶学习机制强制使未来帧预测与视频中的像素流相一致, 在无监督视频表示学习中有显著优势. 对偶运动 GAN 通过在两个未来帧和未来流的生成器以及两个帧和流的判别器之间建立对偶对抗训练机制, 使得生成数据接近真实数据. 其中, 对偶学习机制通过相互审查来桥接未来帧和流预测之间的信息.

对偶运动 GAN 结构如图 24 所示. 对偶运动 GAN 将视频序列 $v = \{I_1, \cdots, I_t\}$ 作为输入, 通过融合未来帧预测 $\ddot{I}_{t+1}$ 和基于未来流的预测 $\bar{F}_{t+1}$ 来预测下一帧 $\hat{I}_{t+1}$. 模型采用简单的 1×1 卷积滤波器进行定影 (Fusing operation) 操作.

对偶运动 GAN 中的生成器由 5 个部分组成: 概率运动编码器 (E), 未来帧生成器 (G_I), 未来流生成器 (G_F), 流估计器 ($Q_{I\to F}$) 和流变形 (Flow-warping) 层 ($Q_{F\to I}$).

对偶运动 GAN 中的判别器由帧判别器 (D_I) 和流判别器 (D_F) 组成. 具体而言, E 首先将先前帧映射为隐编码 z. 然后 G_I 和 G_F 对 z 解码以分别预测未来帧 $\ddot{I}_{t+1}$ 和未来流 $\bar{F}_{t+1}$. $\ddot{I}_{t+1}$ 的保真度由 $\ddot{I}_{t+1}$ 欺骗 D_I 的程度以及 I_t 和 $\ddot{I}_{t+1}$ 之间的流 $\ddot{F}_{t+1}$ 计算得到的流估计器 $Q_{I\to F}$ 欺骗 D_F 的程度决定. 类似地, 未来流预测质量的好坏, 由 $\bar{F}_{t+1}$ 欺骗 D_F 的程度以及帧 I_t 与 $\bar{F}_{t+1}$ 计算得到的流估计器 $Q_{F\to I}$ 产生的变形帧 $\bar{I}_{t+1}$ 欺骗 D_I 的程度决定.

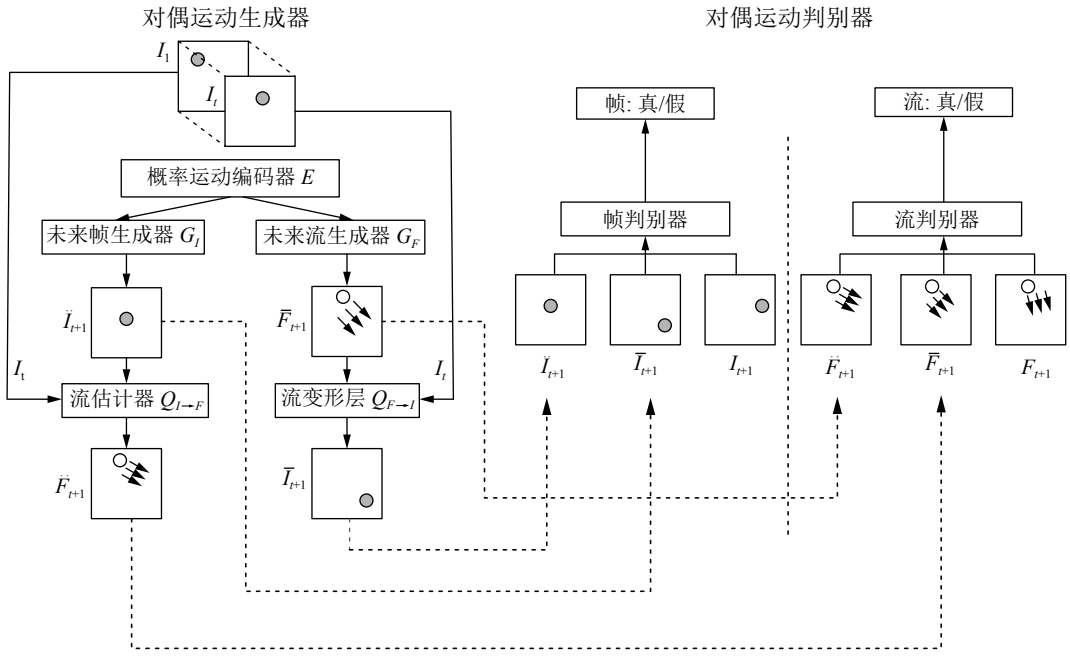


图 24 Dual motion GAN 结构

Fig. 24 The structure of dual motion GAN

综上所述, 对偶运动 GAN 的总目标函数定义为

$$L=\min_{\substack{E,G_I,G_F,\\ Q_{I\to F}}}\max_{D_I,D_F}\mathcal{L}_{VAE}(E,G_I,G_F,Q_{I\to F})+\lambda\mathcal{L}^I_{\text{GAN}}(G_I,G_F,D_I)\lambda\mathcal{L}^F_{\text{GAN}}(G_I,G_F,D_F,Q_{I\to F}) \tag{69}$$

对偶运动 GAN 的提出为 GAN 在无监督视频预测领域开拓了全新的领域. 概率运动编码器学习捕获空间运动不确定性, 而双向对抗判别器和生成器彼此发送反馈信号以生成更加逼真的流和帧. 视频帧预测、流预测和无监督视频表示学习的大量实验证明了该模型对运动编码和预测学习的有效性.

2.8.2　视觉显著性预测

视觉显著性预测是指预测人类的视觉的凝视点和眼球运动, 即估计图像中吸引人类注意力的位置, 视觉显著预测 GAN (Visual saliency prediction with GAN, SalGAN)[64] 利用 GAN 实现视觉显著性预测.

SalGAN 中的生成器遵循卷积编码器–解码器架构, 其中编码器部分包括减小特征映射大小的最大池层, 而解码器部分使用上采样层, 然后使用卷积滤波器构造与输入图片分辨率相同的输出. 生成器的编码器部分在架构上与 VGG-16 完全相同, 省略最后的池化层和全连接层. 生成器的解码器的架构与编码器的结构相同, 但层的顺序相反, 并且池化层由上采样层替换. 同样, 在所有卷积层中使用非线性 ReLU 激活函数, 并且添加具有 S 形 (sigmoid) 非线性的 1×1 卷积层以产生显著映射.

SalGAN 中的判别器由 6 个 3×3 的卷积层及嵌入其中的 3 个池化层和 3 个全连接层组成. 所有卷积层均使用 ReLU 激活函数, 而全连接层使用 tanh 激活函数 (最终层使用 S 形激活函数).

SalGAN 的内容损失函数使用基于二分类交叉熵损失函数 (Binary cross entropy, BCE). 采用 BCE 作为内容损失函数对初始化生成器和稳定训练都起到了较好的效果, 并在 SALICON和 MIT300 数据集上取得了较好的性能.

Fernando 等[65] 提出了一种新颖的显著性估计模型记忆增强 CGAN (Memory augmented conditional GAN, MC-GAN), 该模型利用 CGAN 的语义建模能力, 并结合捕获图像中人物 (主体) 的行为模式以及与任务相关因素的记忆体系结构 (主要通过长短期记忆网络 (Long short-term memory, LSTM) 完成). MC-GAN 不仅揭示了 GAN 的全新应用领域, 而且还强调了特定任务上显著性建模的重要性, 并证明了通过增强记忆架构充分捕捉轮廓的合理性.

2.8.3　图像密写

密写术是在非秘密信息 (“容器”) 内隐藏秘密信息 (“有效载荷”) 方法的集合. Volkhonskiy 等[66] 提出了一种基于 DCGAN 的图像密写 GAN (Steganographic GAN S-GAN).

生成器可以从噪声中产生逼真的图像. 判别器用于对图像是合成图像还是真实图像进行分类, 另一个判别网络 S 判别图像是否包含隐藏的秘密消息. 其具体结构如图 25 所示.

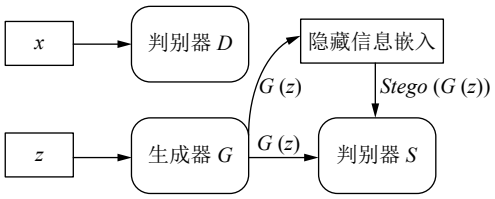


图 25　S-GAN 结构

Fig. 25　The structure of S-GAN

图 25 中, x 为真实图像, z 为噪声, $G(z)$ 为生成器生成的逼真样本. 如果用 $S(x)$ 表示 x 含有隐藏信息的概率, 其损失函数为

$$\begin{aligned}L=&\alpha(\mathrm{E}_{x\sim p_{\text{data}}(x)}[\log D(x)]+\\&\mathrm{E}_{z\sim p_{\text{noise}}(z)}[\log(1-D(G(z)))])+\\&(1-\alpha)\mathrm{E}_{z\sim p_{\text{noise}}(z)}[\log S(Stego(G(z)))+\\&\log(1-S(G(z)))]\end{aligned} \tag{70}$$

其中, $Stego(x)$ 表示将一些隐藏消息嵌入到容器 x 中的结果, α 为权衡参数, 并对这个目标函数进行极小极大游戏 $\min_G\max_D\max_S$. 实验结果表明, S-GAN 成功地欺骗了密写分析器, 并可用于密写应用, 为 GAN 开辟了一个全新的应用领域.

2.8.4　3D 图像生成

3D 图像生成一直是一个十分重要且具有前瞻性的课题. 如何从 2D 图像生成 3D 图像是一个难题. 受到 DCGAN 与 VAE-GAN 的启发, 3D-GAN[67] 恰好提出了一个全新的框架, 通过 GAN 学习 2D 图像到 3D 图像的映射, 它利用在 3D 生成模型上应用的体积卷积网络从概率空间生成三维图像. 生成器由内核为 $4\times4\times4$ 和步幅为 2 的 5 个体积全卷积层组成, 并在其中添加了批量归一化和 ReLU 层, 末端加入了 Sigmoid 层. 判别器基本上与生成器相同, 除了判别器使用了 Leaky ReLU 而不是 ReLU 层. 3D-GAN 的目标函数为

$$L\text{-3D}-\text{GAN}=\log D(x)+\log(1-D(G(z))) \tag{71}$$

其中, x 是 $64\times64\times64$ 空间中的 3 维实际对象,

$z \sim p(z)$ 是随机采样的噪声分布中的样本.

在此基础上, 将 VAE-GAN 的思想加入其中, 3D-VAE-GAN 添加了一个额外的图像编码器 E, 它将 2D 图像 x 作为输入并输出隐表示矢量 z. 因此, 3D-VAE-GAN 由三个部分组成: 图像编码器 E, 解码器 (3D-GAN 中的生成器) 和判别器. E 由 5 个空间卷积层组成. 3D-VAE-GAN 的损失函数添加了另外两个损失函数:

1) KL 散度损失函数 L_{KL}

$$L_{\text{KL}} = D_{\text{KL}}(q(z|y)||p(z)) \tag{72}$$

其中, x 是来自 3D 数据集的样本, y 是其对应的 2D 图像, $q(z|y)$ 是隐变量 z 的变分分布. L_{KL} 的作用是限制编码器的输出分布.

2) 目标重建损失函数 L_{recon}

$$L_{\text{recon}} = \|G(E(y)) - x\|_2 \tag{73}$$

通过结合以上两个损失函数与 $L_{\text{3D-GAN}}$, 可以得到 3D-VAE-GAN 的目标损失函数

$$L = L_{\text{3D-GAN}} + \alpha_1 L_{\text{KL}} + \alpha_2 L_{\text{recon}} \tag{74}$$

其中, α_1 和 α_2 为权衡系数.

3D-VAE-GAN 与 AlexNet 和 T-L network 相比分类准确率更高. 3D-GAN 在不同的 3D 图像生成实验中均取得了不错成果, 为 GAN 在 3D 图像领域的应用做出了积极的贡献.

文献 [68] 提出了视觉对象网络 (Visual object networks, VON) 利用可解析的 3D 表示 (三个重要因素: 形状, 视点和纹理) 来合成 3D 图像. VON 合成图像的具体过程为:

1) 由于 3D-GAN 存在模式崩溃的问题, 为提高生成图像的质量以及多样性, 加入了 WGAN-GP 中的 Wasserstein 距离, 与 3D-GAN 进行结合生成 3D 形状.

2) 在采样视点 (Viewpoint) 的作用下, 从 3D 形状渲染对象的 2.5D 草图 (即轮廓和深度图).

3) 通过结合 CycleGAN, 为 2.5D 草图添加逼真的纹理, 生成逼真的图像.

VON 不仅可以生成比最先进的 2D 图像合成方法更逼真的图像, 而且可以实现许多 3D 操作, 例如改变生成图像的视点、形状, 实现纹理编辑、纹理和形状空间中的线性插值以及不同的对象和视点外观变换. VON 与 DCGAN、WGAN-GP 以及 LSGAN 在 ShapeNet 数据集上有更低的 FID 值, 以及较高的 IS 分数.

2.9 小结与分析

本节主要介绍了 GAN 在图像领域中的应用, 介绍了在人脸识别与图像生成应用中的模型: Age-cGAN 和 GLCA-GAN; 在多姿态人脸转正应用中的模型: TP-GAN, DRGAN; 在图像纹理合成应用中的模型: SGAN, MGAN 和 BigGAN; 在图像超分辨率应用中的模型: SRGAN 和 c-CycleGAN; 在图像复原与多视角图像生成应用中的模型: PG-GAN 和 VariGAN; 在图像转换应用中的模型: DualGAN, GeneGAN, S^2-GAN 和 DA-GAN; 在文本描述到图像生成应用中的模型: Text to image GAN, GAWWN 和 RTT-GAN; 在图像语义分割应用中的模型: 基于 GAN 的语义分割模型和 Contrast-GAN; 在图像着色应用中的模型: Auto-painter; 在视频预测领域应用中的模型: Dual motion GAN; 在视觉显著性预测应用中的模型: SalGAN 和 MC-GAN; 在图像密写应用中的模型: S-GAN; 在 3D 图像生成应用中的模型: 3D-GAN 和 VON. 表 2 汇总了本节中不同模型的分类情况.

表 2 GAN 在图像领域的应用

Table 2 GAN's application in the field of computer vision

内容	模型
人脸图像识别与图像生成	基于 CGAN 的人脸识别模型[28], Age-cGAN[29], GLCA-GAN[30], TP-GAN[31], DR-GAN[33], SGAN[34], MGAN[35], BigGAN[37]
图像超分辨率	SRGAN[38], c-CycleGAN[39]
图像复原与多视角图像生成	基于 GAN 的语义图像修复模型[41], PGGAN[42], VariGAN[45]
图像转换	DualGAN[47], GeneGAN[48], S^2-GAN[49], DA-GAN[50]
文本描述到图像生成	Text to image GAN[52], GAWWN[53], RTT-GAN[54]
图像语义分割	基于GAN的语义分割模型[55-56], Contrast-GAN[57]
图像着色	Auto-painter[58], DCGAN 用于图像着色[59]
视频预测	基于GAN的下帧图像生成模型[61], 利用 3D-CNN 作为生成器的 GAN[62], Dual motion GAN[63]
视觉显著性预测	SalGAN[64], MC-GAN[65]
图像密写	S-GAN[66]
3D 图像生成	3D-GAN[67], VON[68]

在本节的介绍中, 许多应用模型均采用 WGAN, CGAN, CycleGAN 和 DCGAN 等模型作为其基本模型, 并在网络结构以及损失函数或训练方法等方面进行改进. 利用 GAN 不仅在各个主要图像合成及识别领域具有强大的能力, 还对一些诸如视频预测、3D 图像生成的新兴领域做出了积极的贡献.

本节也总结了很多基于不同思想的应用模型, 例如, 利用双路径思想进行背景和纹理生成; 利用比较思想与对抗思想进行语义分割; 利用对偶学习思想与对抗思想进行图像转换等. 从这些方面不难看出, GAN 在图像领域已经具有更深入的进展, 而不仅仅是停留在增加罚项、修改网络结构以及更换 GAN 变种模型.

3 GAN 在语音与 NLP 领域的应用

3.1 GAN 在语音领域的应用

在语音领域, 深度学习发展迅猛, GAN 在语音领域近些年也取得了一定的成果. 本节将从语音去噪, 音乐、音符生成, 语音识别领域分别对 GAN 进行介绍.

3.1.1 语音增强

在人类的生产和生活环境中, 噪声可以说遍布了每个角落, 因此提高语音系统抗噪声性能是一个具有挑战性的任务, 而语音增强 (Speech enhancement) 是解决这个问题的有效技术之一. 语音增强是指提高被加性噪声污染的语音的清晰度和质量, 最主要的应用领域是提高噪声环境下的移动通信质量.

GAN 在计算机视觉领域生成逼真图像上取得了巨大成功, 可以生成像素级、复杂分布的图像. 这些在计算机视觉领域上的成功应用, 同时也推动了 GAN 在语音处理领域应用上的研究. Pascual 等[69] 将语音增强 GAN (Speech enhancement GAN, SEGAN) 应用于语音信号生成上. 在 SEGAN 之前, GAN 尚未应用于任何语音生成和增强任务. SEGAN是将对抗框架用于生成语音信号的第一种方法. 结合 CGAN和利用最小二乘函数代替 GAN 损失函数得到的最小二乘 GAN (Least squares GAN, LSGAN)[70], 并使用 $L1$ 范数稀疏正则化项, SEGAN 的判别器和生成器损失函数分别为

$$\begin{aligned}\min_D V_{\text{LSGAN}}(D) = &\frac{1}{2}\mathrm{E}_{x\sim p_{\text{data}}(x,x_c)}[(D(x,x_c)-1)^2]+\\ &\frac{1}{2}\mathrm{E}_{\substack{x_c\sim p_{\text{data}}(x_c),\\ z\sim p_{z(z)}}}[D(G(z,x_c))^2] \end{aligned} \tag{75}$$

$$\begin{aligned}\min_G V_{\text{LSGAN}}(G) = &\frac{1}{2}\mathrm{E}_{\substack{x_c\sim p_{\text{data}}(x_c),\\ z\sim p_{z(z)}}}[D(G(z,x_c))-1^2]+\\ &\lambda\|G(z,\bar{x})-x\|_1 \end{aligned} \tag{76}$$

其中, $x=G(z,x_c)$ 为生成语音, x_c 为纯净语音, z 为从正态分布中抽样的随机噪声.

SEGAN 主要优点有:

1) 提供了一种快速语音增强方法. 由于提出的模型不需要因果关系假设, 去掉了循环神经网络 (Recurrent neural network, RNN) 中的递归运算, 是可以直接处理原始音频的端到端方法, 不需要手工提取特征, 不需要对原始数据做出明显假设.

2) 从不同对话者和不同类型噪声中学习, 并将它们结合在一起形成相同的共享参数, 使得基于 SEGAN 的语音增强系统不仅结构简洁而且泛化能力较强.

实验结果表明, SEGAN 不仅可行, 而且还可以作为当前方法的有效替代方法 (SEGAN 可有效替代维纳滤波方法).

除此之外, 受到 SEGAN 的启发, Michelsantid 等[71] 将 Pix2Pix 的框架用于学习有噪音和纯净语音频谱图之间的映射关系.

3.1.2 音乐生成

生成音乐与生成图像和视频有一些明显的区别. 首先, 音乐可以说是随着时间变化的艺术, 需要时间模型. 其次, 音乐通常由多个音轨组成, 彼此之间具有紧密的关联性. 每个轨道都有自己的时间动态特性, 但总的来说, 它们会随着时间的推移而相互依赖地进行下去. MuseGAN (Multi-track sequential GAN)[72] 在研究多轨序列式音乐生成和伴奏时, 遵循了运用 GAN 和 CNN 生成音乐的思路, 并得到了一些新的研究方向, 将音乐分为乐段、乐句、小节、节拍和像素五个层级, 然后逐层进行生成, 值得借鉴.

3.1.3 语音识别

自动语音识别 (Automatic speech recognition, ASR) 正在逐渐走入人们的生活当中, 智能音箱、手机语音助手、智能电视等均需要 ASR 技术的支持. 虽然在 GAN 中, 语音识别方面的应用比较少, 但是这个领域将是未来这个方向的发展目标.

Sriram 等[73] 将 GAN 应用在语音识别领域, 提出了一个通用端到端且具有可扩展性的框架, 并通过实验证明该框架可在无需预处理的情况下, 提升远程语音识别性能. Shinohara 等[74] 引入了对抗多任务学习模式, 增强语音识别对噪音的鲁棒性.

文献 [75] 研究了语音生物识别系统中 GAN 合成欺骗攻击的能力. 通过对 WGAN 目标函数进行修改, 以合成足够逼真的类似于真实语音的数据. 该模型使用半监督学习方法实现目标攻击和非目标攻击, 在语音生物识别系统中提出了安全相关的

问题.

在语音识别中还有一个分支: 语音模仿. 语音模仿产生的声音必须令人信服, 不仅需要模仿语音的质量, 而且还应模仿目标说话者的风格. Gao 等[76] 提出利用 GAN 进行语音模仿 (Voice GAN, VoiceGAN). 实验表明, 该模型可以产生非常令人信服的模拟语音样本, 甚至能够有效地冒充不同性别的声音.

说话人验证 (Speaker verification, SV) 通过对语音的特征提取, 判断语音对话是否为说话人所说. Ding 等[77] 提出了多任务三元组生成对抗网络 (Multitasking triplet GAN, MTGAN). MTGAN 由一个具有条件信息的 GAN、一个用于提取特征的编码器以及一个分类器组成, 并将三重损失函数 (最小化类内距离, 同时最大化类间距离) 与生成对抗网络和多任务学习进行结合用于说话人验证任务, 取得了不错的效果. MTGAN 与利用 softmax 的模型、i-向量方法以及采用三联体系统的模型相比具有更低的等错误率 (Equal error rate, EER) 和更高的准确率.

在 ASR 领域中最常用的一个语音特征是梅西倒谱系数 (Mel freqency cepstral coefficients, MFCC). 文献 [78] 中利用残差 GAN (residual GAN) 组成的噪声模型, 为模拟的激励信号生成一个高频随机分量. 实验表明利用残差 GAN 的模型能够合成具有较高质量的语音样本.

3.2 GAN 在 NLP 领域的应用

自然语言处理 (Natural language processing, NLP) 是人工智能应用于语言领域的分支学科, 用于探讨如何处理及运用自然语言. 自然语言认知是指让电脑 "懂" 人类的语言. 自然语言生成系统将计算机数据转化为自然语言. 自然语言理解系统将自然语言转化为计算机程序更易于处理的形式. Goodfellow[2] 认为 GAN 只适用于连续型数据的生成, 对于离散型数据, 其效果不佳, 使得 GAN 在 NLP 领域一直无法超越 VAE 的性能. WGAN和 SeqGAN 的提出, 成功地将 GAN 应用于 NLP 领域.

本小节将 GAN 在自然语言处理领域的应用分为以下几类: 对话模型评估与生成、生成离散型序列、双语字典、文本分类与生成和语篇分析.

3.2.1 对话模型的评估与生成

1) 对话模型评估

在 NLP 领域, 对话模型评估仍然是一个重要的挑战, 受 GAN 成功应用在图像领域的启发, Kannan 等[79] 提出了一个对话模型评估. 生成器用的是 Seq2Seq 模型[80], 由 RNN 编码器和 RNN 解码器组成. 判别器也是一个 RNN, 但只有一个编码器和一个二元分类器. 采用这种生成对抗的方法, 可以减少对人类评估的需求, 同时更直接地对生成任务进行评估.

2) 对话模型生成

Li 等[81] 将对抗训练的思路应用于开放式对话问题并提出了对抗强化学习的方法. 但是, GAN 和 NLP 很难相互融合. Li 等参考了 SeqGAN, 利用强化学习 (RL) 来规避 GAN 在 NLP 中使用的难点. Li 等提出的思路中, 生成器用的是 Seq2Seq 模型, 由 RNN 编码器和 RNN 解码器组成. 判别器是一个二分类器, 将对话序列作为输入, 并输出一个标签来判断输入是由机器生成的还是人生成的对话. 这个方法的实验结果显示虽然没有明显的性能提升, 但是其对于人工生成序列与参考目标序列的概率分布差异比较大, 其学习任务是有效的.

3.2.2 生成离散型序列

当目标为离散型数据时, GAN 具有局限性. 将 GAN 应用于离散数据生成仍然是一个值得关注的开放性研究问题. 在变种模型中 SeqGAN 就是对生成离散型序列的一个很好的应用.

Kusner 等[82] 为了使 GAN 能够用于生成离散数据, 提出了基于 Gumbel-softmax 分布作为输出的 GAN (Gumbel-softmax GAN), 通过从 Gumbel-softmax 分布中采样得到可微的样本, 并采用擅长进行离散数据处理的长短期记忆神经网络 (LSTM) 作为生成器和判别器进行离散型序列生成. Gumbel-softmax 分布是对参数化 softmax 的多项分布的连续逼近. 经过 softmax 层后需要对变量进行 one-hot 采样

$$y = one\_hot(\arg\max_i (h_i + g_i)) \tag{77}$$

其中, g_i是服从 Gumbel 分布的独立随机变量. $h_i \sim \boldsymbol{h}$, $i = 1, \cdots, d$, $\boldsymbol{h}$ 是 d 维随机变量. 但是式 (77) 存在不可微分的问题 (argmax 不可微分). 因此用下式来对 y 进行逼近. 服从 Gumbel-softmax 分布的输出的计算式为

$$y = \text{soft}\max\left(\frac{1}{\tau(\boldsymbol{h} + \boldsymbol{g})}\right) \tag{78}$$

其中, τ 是一个反温度参数. 式 (78) 的概率分布是由 τ 和 $\boldsymbol{h}$ 决定, 称为 Gumbel-softmax 分布. 可以通过式 (78) 来训练离散数据上的 GAN, 从一些相对较大的 τ 开始, 然后在训练期间将其逐渐下降到零.

Gumbel-softmax GAN 在生成离散型序列任务中的性能评估表明, GAN 对处理离散型数据具

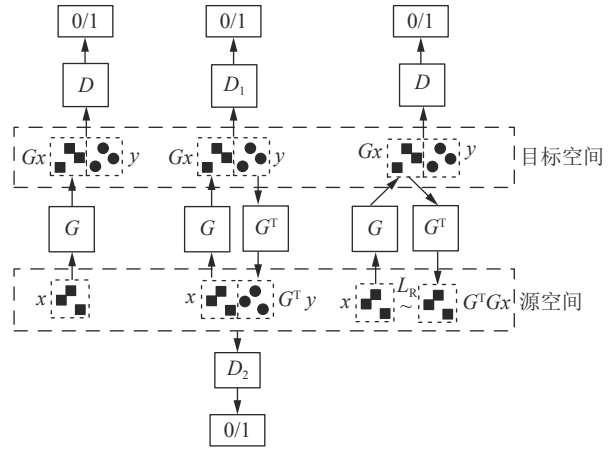
有有效性.

3.2.3　双语字典

在跨语言任务中, 利用非平行双语语料构建双语字典是一个得到持续关注的课题. 目前, 双语字典的实现过程一般需要跨语言信息作为监督信号, 但是 Zhang 等[83] 提出了三种 GAN 模型实现双语词典构建任务, 并取得了良好效果.

1) 单向转换模型

假设两种语言分别有两个词向量 x 和 y, x 服从 p_x 表示的概率分布, y 服从 p_y 表示的概率分布. 用生成器来代表映射 $f(x)$, 使 $f(x)$ 趋近于 y. 用判别器作为二分类器, 判别输入是属于 $f(x)$ 还是 y, 这样就形成了一个对抗网络, 如图 26(a) 所示. 将生成器参数化为转换矩阵 $G \in \mathbf{R}^{d\times d}$, 判别器为一个二分类器, 使用具有一个隐层的标准前馈神经网络来对判别器进行参数化, 并且其损失函数是交叉熵损失.



(a) 单向转换模型 (a) One way transformation model　(b) 双向转换模型 (b) Bidirectional transformation model　(c) 对抗自编码模型 (c) Adversarial auto-encoding model

图 26　双语字典 GAN 结构

Fig. 26　The structure of bilingual lexicon GAN

单向转换模型的判别器和生成器的损失函数为

$$L_D = -\log D(y) - \log(1 - D(Gx)) \tag{79}$$

$$L_G = -\log D(Gx) \tag{80}$$

单向转换模型主要存在的问题是模型很难训练, G 的参数搜索空间为 $\mathbf{R}^{d\times d}$ 太大, 需要对 G 进行参数正交化, 以减少模型的搜索空间.

2) 双向转换模型

单向转换模型的正交参数化过程计算复杂性高, 耗时多. 为了减少对参数正交化的依赖, 在单向模型的基础上增加了反向过程. G 是将源词嵌入空间转换为目标语言空间; 而 G^{T} 是将目标语言空间转换为源词嵌入空间. 这样就构成了一个双向转换模型, 如图 26(b) 所示.

此时生成器的损失函数为

$$L_G = -\log D_1(Gx) - \log D_2\left(G^{\mathrm{T}}y\right) \tag{81}$$

3) 对抗自编码模型

另一种减轻正交约束的方法是引入对抗自编码, 通过 G^{T} 映射回来的向量来重构词向量 x. 如图 26(c) 所示. 其中用余弦相似度度量重构的损失函数

$$L_R = -\cos(x, G^{\mathrm{T}}Gx) \tag{82}$$

若转换矩阵 G 为正交矩阵, L_R 将被最小化. 此时 L_G 损失函数将变为

$$L_G = -\log D(Gx) - \lambda\cos(x, G^{\mathrm{T}}Gx) \tag{83}$$

这里, λ 是一个权衡参数. $\lambda = 0$ 表示该模型将恢复为单向变换模型, 而更大的 λ 对应于强制执行更严格的正交约束.

在汉译英实验方面, 以上提出的三种模型都取得了非常突出的成绩, 表明 GAN 在双语字典领域中具有一定有效的应用. 在汉译英实验中, 与 TM (Translation matrix)[84] 和 IA (Isometric alignment)[85] 方法进行比较, 三种模型都能够获得更高的预测准确率.

3.2.4　文本分类与生成

1) 文本分类

文本分类属于自然语言处理中的一个重要分支. Liu 等[86] 提出多任务学习 (Multi-task learning, MTL) 来学习各任务之间的共享层, 从而得到各任务之间不变的特征. 提出的对抗多任务学习框架用来保证共享特征空间只包含共享和任务不变信息. Liu 等[86] 在 16 个不同的文本分类任务上进行了实验, 并证实了所提方法的有效性.

2) 文本生成

Press 等[87] 提出将 WGAN 应用于文本生成领域, 提出的对抗生成模型中生成器和判别器均使用 RNN. 初始时, 生成器可能只能生成带有空格的随机字母序列, 但是随着判别器的判别能力的提高, 生成器将生成单词, 进而生成器就可以生成更长、更连贯的文本序列. 判别器和生成器之间的这种相互作用有助于实现文本生成的增量学习, Press 等的研究工作证明了 GAN 在文本生成领域的可行性.

文献 [88] 提出了促进多样性 GAN (Diversity promoting GAN, DP-GAN) 用于解决现有文本生成方法倾向于生成重复且 “无聊” 的描述信息问题. DP-GAN 参考了 SeqGAN 中的奖励及梯度策略更

新方法, 对重复文本给予低奖励值, 对 “新颖”文本给予高奖励值, 从而鼓励生成器生成多样且信息丰富的文本. DP-GAN 中的生成器和判别器均采用 LSTM 结构.

DP-GAN 在 Yelp、Amazon 以及 Dialogue 数据集上与传统模型 MLE 以及 SeqGAN 模型相比取得了不错的性能提升.

3.2.5 语篇分析

语篇分析又称话语分析或篇章分析, 是对 “语篇”整体进行的分析, 包括语篇基本单元之间的关系、不同语篇单元的成份间关联以及语篇所含的信息等.

在语篇分析中, Chen 等[89] 提出了一个基于 GAN 的特征模仿框架, 称为对抗深度平均网络 (Adversarial deep averaging network, ADAN), 应用于隐性语篇关系分类中, 是第一个将对抗学习应用在语篇分析中的框架. ADAN 还能对情感进行分类, 情感分析是语篇分析的一个分支, 是对带有情感色彩 (褒义和贬义/正面情绪和负面情绪) 的主观性文本进行分析, 以确定该文本隐含的观点、喜好、情感倾向. ADAN 的基本结构如图 27 所示.

ADAN 是一个具有两个分支的前馈网络. 网络中主要有三个部分: 1) 将输入序列 x 映射到共享特征空间的联合特征提取器 F; 2) 在给定特征表示 $F(x)$ 下, 预测情感标签的情感分类器 P; 3) 语言记分器 Q, Q 计算出一个标量分数, 用来指示输入序列 x 是来自源域还是目标域, 假定

$$\begin{aligned} P_F^{\text{src}}&=P(F(x)|x\in SOURCE)\\ P_F^{\text{tgt}}&=P(F(x)|x\in TARGET) \end{aligned} \tag{84}$$

如上所述, ADAN 学习的目标是训练 F 使 $SOURCE$和 $TARGET$ 所服从的概率分布尽可能接近, 以学习语言不变特征, 以便更好地实现跨语言概括.

现有的训练 GAN 的方法等同于最小化 $SOURCE$ 和 $TARGET$ 所服从的概率分布的 Jensen-Shannon 距离, 但是 Jensen-Shannon 距离是不连续函数, 无法求导, 不能很好地进行训练. 由于 Wasserstein 距离函数为连续可微的, ADAN 选择最小化 P_F^{src} 与 P_F^{tgt} 之间的 Wasserstein 距离

$$W(P_F^{\text{src}},P_F^{\text{tgt}})=\sup_{\|g\|_L\le 1}\left(\mathop{\text{E}}_{f(x)\sim P_F^{\text{src}}}[g(f(x))]-\mathop{\text{E}}_{f(x')\sim P_F^{\text{tgt}}}[g(f(x'))]\right) \tag{85}$$

其中, 上确界 (最大值) 被限制在所有 Lipschitz 常数为 1 的函数 g 的集合上. 为了近似计算 $W(P_F^{\text{src}}, P_F^{\text{tgt}})$, 使用语言记分器 Q 作为式 (85) 的函数 g, 因而 ADAN 的目标是在式 (85) 中寻找上确界来估计 $W(P_F^{\text{src}}, P_F^{\text{tgt}})$. 为了使记分器 Q 为 Lipschitz 函数, 记分器 Q 的参数被剪裁到一个固定的范围内. 设 Q 由 θ_q 参数化, 则记分器 Q 的目标函数 J_q 变为

$$J_q(\theta_f)\equiv\max_{\theta_q}\left(\mathop{\text{E}}_{F(x)\sim P_F^{\text{src}}}[Q(F(x))]-\mathop{\text{E}}_{F(x')\sim P_F^{\text{tgt}}}[Q(F(x'))]\right) \tag{86}$$

直观地说, Q 会尝试输出源域实例的较高分数和目标域实例的较低分数.

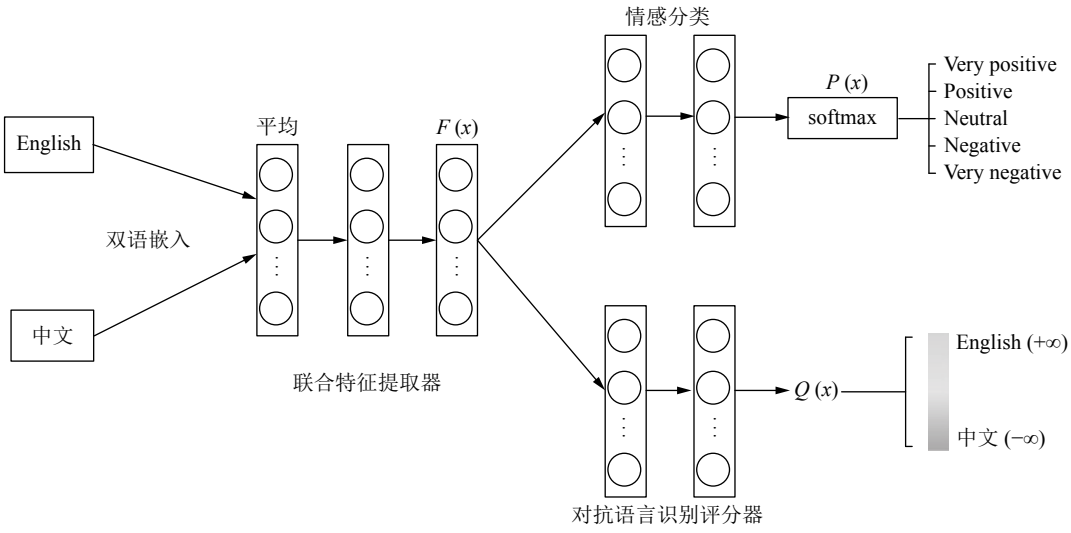


图 27 ADAN 结构

Fig. 27 The structure of ADAN

$$J_p(\theta_f) = \min_{\theta_p} \mathop{\mathrm{E}}_{(x,y)} [L_p(P(F(x)), y)] \tag{87}$$

$$J_f \equiv \min_{\theta_f} J_p(\theta_f) + \lambda J_q(\theta_f) \tag{88}$$

对于由 θ_p 参数化的情感分类器 P, 使用传统的交叉熵损失 $L_p(\hat{y}, y)$ 作为情感分类器 P 的目标函数. 因此 ADAN 的一个主要特点是应用了对抗思想以及应用了一些 WGAN 中解决训练不稳定的方法. ADAN 在域自适应学习、机器翻译等实验中, 大幅领先深度平均网络 (Deep averaging network, DAN)[90].

3.2.6 机器翻译

机器翻译是 NLP 领域中的一个重要且具有挑战性的任务, 它的目标是将源语言的句子自动翻译成相应的目标语言的句子. 与传统机器翻译相比, 神经机器翻译 (Neural machine translation, NMT)[91] 的提出似乎将机器翻译领域提升到了一个新高度.

在神经机器翻译领域中, 文献 [92] 受到 SeqGAN 的启发将 GAN 应用于 NMT 中, 提出了 BR-CSGAN (BLEU reinforced conditional sequence GAN), 其中, 增加了一项 BLEU 静态损失函数对生成器进行强化训练.

除此之外, 文献 [93] 提出了富有创新性的 Multi-CSGAN-NMT (Muti-conditional sequence GAN for neural machine translation), 该模型可利用多个生成器和判别器完成机器翻译, 并对每个生成器增添一个信息生成器模块用于生成器之间的信息传递, 从而提升翻译效果.

另外, 还有利用 CNN 作为判别器、NMT 模型作为生成器的 Adversarial-NMT 模型[94], 并具有比传统 NMT 模型更好的翻译效果.

文献 [95] 为了解决 GAN 在 NMT 任务中训练不稳定问题, 提出了 BGAN-NMT (Bidirectional GAN for neural machine translation), 该模型是一个双向网络模型, 利用生成器模型作为判别器的方法, 使判别器能够更好地观察整个翻译空间, 从而减轻训练不充分问题. 该过程需要用到两个 GAN 网络, 一个用于 NMT 任务的 GAN 以及一个辅助 GAN 网络, 其中辅助 GAN 网络的生成器和判别器为执行 NMT 任务的 GAN 网络的判别器和生成器.

在该领域中, 四个模型通过双语互译质量评估辅助工具 (Bilingual evaluation understudy, BLEU), 在汉译英和德译英的实验中与传统方法相比, BGAN-NMT 和 Muti-CSGAN-NMT 都具有更好的翻译效果.

3.3 小结与分析

本节主要介绍了 GAN 在语音和 NLP 领域的应用. 首先在语音领域, 介绍了在语音增强应用中的模型: SEGAN; 在音乐与音符生成应用中的模型: MuseGAN; 在语音识别应用中的模型: 基于 WGAN 在 ASR 中的应用, VoiceGAN, MTGAN, Residual GAN 等.

GAN 在语音领域的应用还不是非常丰富, 主要依托于 CGAN 的基本架构, 条件信息的加入可使模型在语音合成、音乐合成方面具有可控性.

在 NLP 领域, 介绍了在对话模型的评估与生成应用中的模型: 基于 Seq2Seq 与 RL 的 GAN 模型; 在生成离散型序列应用中的模型: Gumbel-softmax GAN; 在双语字典应用中的模型: 单向 GAN 模型, 双向 GAN 模型和对抗自编码 GAN 模型; 在文本生成与分类应用中的模型: DP-GAN, 基于 WGAN 与 RNN 的 GAN 模型; 在语篇分析应用中的模型: ADAN; 在机器翻译应用中的模型: BR-CSGAN, Multi-CSGAN-NMT, BGAN-NMT 和 Adversarial-NMT.

GAN 最开始应用于生成连续数据, 但在 NLP 中, 由于生成离散型序列使得生成器与判别器在进行梯度训练时遇到问题, 因为梯度训练需要生成器与判别器损失函数都是可微的. 另外 GAN只能对完整的序列进行评分. 自从 SeqGAN 提出后, 为 GAN 在 NLP 领域的发展提供了动力. 大部分 NLP 领域的 GAN 都是采用 LSTM 和具有记忆功能的 RNN 神经网络, 确保在进行对抗训练时, 能够使模型更多地记住过去学到的信息. 表 3 汇总了本节不同模型的分类情况.

4 GAN 在其他领域的应用

本文已经介绍了 GAN 在图像、语音和 NLP 领域的部分应用. 本节将讨论 GAN 在日常生活中其他领域的应用. 具体地, 将从人类姿态估计、恶意软件检测、数据集标记、物理应用、医学数据、隐私保护等领域对 GAN 的应用进行介绍.

4.1 人体姿态估计

在计算机视觉领域中, 人体姿态估计是通过图像观察来检测出人体各个部分的位置、方向以及大小等信息. 近年来, 人体姿态估计已经从静态估计发展为动态估计.

Merel 等[96] 发现使用强化学习 (RL) 方法进行姿态估计时, 往往会产生不够人性化以及肢体过于僵硬的移动行为. 所以将生成对抗的思想与强化学

表 3 GAN 在语音与 NLP 领域的应用
Table 3 GAN's application in the field of speech and NLP

内容	模型
语音增强	SEGAN[69], 基于 Pix2Pix 的语音增强模型[71]
音乐生成	MuseGAN[72]
语音识别	基于 GAN 的语音识别模型[73], 基于多任务对抗学习模式的语音识别模型[74], WGAN 用于语音识别[75], VoiceGAN[76], MTGAN[77], Residual GAN[78]
对话模型的评估与生成	基于 SeqGAN 的对话评估模型[79], 基于 SeqGAN 的对话生成模型[80]
生成离散序列	Gumbel-softmax GAN[82]
双语字典	基于 GAN 的双语字典模型[83]
文本分类与生成	对抗多任务学习模型[86], 基于 WGAN 的文本生成模型[87], DP-GAN[88]
语篇分析	ADAN[89]
机器翻译	BR-CSGAN[92], Multi-CSGAN-NMT[93], Adversarial-NMT[94], BGAN-NMT[95]

习相结合, 使模仿效果更加细腻逼真. Chou 等[97] 利用生成器作为人体姿态估计器, 判别器与生成器具有相同的结构, 判别器执行判别任务. 实验结果表明, 所提出的 GAN 模型能有效提升人体姿态估计预测的准确性.

文献 [98] 提出了一种通过语音驱动进行头部运动生成的方法, 但是需要学习语音与头部运动之间的关系. 因此通过引入双向 LSTM (Bidirectional long-short term memory, BLSTM) 的 CGAN, 从条件分布采样中为每个语音片段生成多个头部运动图像. 实验表明, 与传统动态贝叶斯网络 (Dynamic Bayesian network, DBN) 和 BLSTM 相比具有更好的性能.

4.2 恶意软件检测

在软件安全领域中, 随着个人信息重要性的日益提高, 防止恶意软件攻击成为一个严峻的课题. 近年来, 机器学习被应用于恶意软件检测中. 文献 [99] 提出了一种基于 GAN 的恶意代码生成器 MalGAN 来生成具有对抗性的恶意代码. 实验结果表明, 恶意代码能够有效地绕过黑盒检测器, MalGAN 通过提出有效的攻击手段, 促进信息安全的防御能力提高.

4.3 数据集标记与数据增强

1) 数据集标记

对大数据集进行标注是一个十分昂贵且耗时的过程, 有一些数据集需要人亲自采集过后才能标注, 所以人工合成带有标签的逼真的图像数据是一个很有价值的工作. 随着图形技术的快速发展, 在人工合成图像上训练模型变得更易处理, 但是人工合成图像与真实图像分布之间存在差异, 往往无法达到期望性能.

文献 [100] 中引入了基于 GAN 的仿真无监督学习框架 (Simulated + Unsupervised learning with GAN), 确保神经网络在生成逼真图像的同时, 保留人工合成图像中的标注信息. 应用这种方法来生成训练样本对人脸识别的很多任务都有极大的参考价值, 可以用来生成不同姿势、光照的人工合成图像数据. 文献 [101] 提出了 RenderGAN (Render generative adversarial network) 框架, RenderGAN结合 3D 模型和 GAN 框架生成大量逼真的标记图像.

2) 数据增强

数据增强也叫数据扩充 (Data augmentation). 当训练集样本不够多或者是某一类数据比较少, 以及为防止模型出现过拟合, 让模型具有鲁棒性等学习场景, 数据增强是一个不错的方法. 以往的数据增强只是简单地对图像进行旋转、翻转、缩放、裁剪及平移等.

相比于这种简单的数据增强方式, 文献 [102] 提出了一个数据增强生成对抗网络 (Data augmentation GAN, DAGAN). 通过对抗训练并结合 WGAN 与 CGAN来学习源域的图像数据的方法是一种高级的数据增强方法. 其中, DAGAN 使用了 U-net 与 ResNet的结合得到的网络 UResNet 作为生成器, 判别器则是采用了 DenseNet 网络[103]. 实验证明, DAGAN能够比传统的数据增强方法更具优越性.

4.4 物理应用

GAN 不仅在计算机视觉、语音和 NLP 领域有所建树. 在物理学领域也有相应的应用. de Oliveira 等在文献 [104] 中首次将机器学习中的生成建模与高能粒子物理中的模拟物理过程之间建立了联系, 将新的 GAN 框架应用于喷射图像 (来自粒子与量热计相互作用的能量沉积的 2D 表示) 的生成. 这项工作为进一步探索 GAN 在物理学领域中的应用提供了基础, 并表明 GAN 可在高能粒子物理学图像生成中实现更加便捷的模拟过程.

4.5 医学领域

GAN 在医学领域的主要应用集中在图像重建上, 例如压缩感知磁共振图像 (Compressed sensing magnetic resonance imaging, CS-MRI) 以及眼底图像的视网膜血管分割. 除此之外, 也应用于电子健康记录 (Electronic health record, EHR) 数据上.

文献 [105] 提出 RefineGAN (Refine generative adversarial network), 将其用于实现快速准确的 CS-MRI 重建, 并证明了该方法在运行时间和图像质量方面的表现优于目前最先进的 CS-MRI 方法.

除此之外, 由于多对比度的磁共振图像 (Multi contrast magnetic resonance imaging, Multi-contrast MRI) 虽然可以获得更多的诊断信息, 但是受到扫描时间限制导致无法获取某些对比度条件下的 MRI 图像, 或是某些对比度的 MRI 图像可能被噪声及伪影破坏. 在这种情况下, 需要通过从剩余的对比度 MRI 图像中合成未获得或损坏的图像, 从而提高诊断效果. 文献 [106] 中利用 CGAN 生成多对比度 MRI 图像. 实验证明该方法与现有方法相比具有一定优势.

EHR 数据促进了医学研究的进展. 但是由于这些 EHR 数据往往都是隐私的, 所以生成人工合成的 EHR 数据是很有必要的. Choi 等[107] 通过将 VAE 和 GAN 相结合提出了 MedGAN (Medical generative adversarial network), 用于生成高维多标签离散样本. 实验结果表明, MedGAN 生成的电子健康记录 (EHR) 具有良好的应用效果.

视网膜血管分割是利用眼底图像自动检测视网膜疾病的必不可少的步骤, 但是现有方法存在一些问题. Son 等[108] 提出了一种方法, 通过 GAN 生成精确的视网膜血管图, 能够描绘出足够的细节以及更加清晰的线条, 并在两个数据集 (DRIVE 和 STARE) 上获得了最好的学习性能.

文献 [109] 利用 WGAN 组成的生成器和判别器对血管的几何形状进行合成, 并将合成冠状动脉血管几何形状应用于心脏 CT 血管造影 (Cardiac CT angiography, CCTA).

4.6 隐私保护

GAN 可以作为一种对隐私进行攻击的攻击手段, 从而我们可以提出相应的防御手段. Hitaj 等[110] 利用 GAN 作为一种攻击手段, 在遵从协同学习协议的参数服务器中对其他用户的信息进行学习, 从而得到他人隐私. 把 GAN 作为攻击者, 取得的 “成果”, 为协同学习在安全领域的应用指明了方向.

4.7 域自适应学习领域

将 GAN 引入迁移学习中, 是迁移学习目前取得突破的研究方向之一. 其中, 域自适应学习[111] 指的是在迁移学习中, 源域和目标域的数据分布不同, 但两者任务相同. 用基于 GAN 结构的无监督转换网络 (Unsupervised pixel-level domain adaptation method, PixelDA)[112] 来解决无监督域自适应学习问题, 该方法不仅产生合理的样本, 而且在许多无监督的域自适应学习场景中也取得了骄人的学习结果.

Zhang 等在文献 [113] 中提出了基于 GAN 结构的神经网络, 实现了域自适应分类任务. 在视觉领域的域自适应学习问题上, Sankaranarayanan 等[114] 利用 GAN 结构学习源域与目标域的联合嵌入特征空间, 并在无监督域自适应学习问题上取得了显著的效果.

4.8 自动驾驶

自动驾驶是近期人工智能研究中最有前途的应用之一. 自动驾驶要求在复制人类的驾驶行为的同时牢记安全问题. Ghosh 等[115] 使用 GAN 对驾驶场景进行预测, 并在视频游戏 Road rash 上训练模型, 测试其驾驶场景预测的准确性. Santana 等[116]利用 VAE 和 GAN 实现对未来路况预测.

4.9 小结与分析

本节主要介绍了 GAN 在一些目前大热以及比较冷门领域的发展. 介绍了在人体姿态估计应用中的模型: 基于 RL 与 GAN 的模型, 引入 BLSTM 的 CGAN 模型; 在恶意软件检测应用中的模型: MalGAN; 在数据集标记与数据增强应用中的模型: 基于 GAN 的仿真无监督学习框架, RenderGAN, DAGAN; 在物理应用中的模型: 将 GAN 应用于高能粒子物理学中喷射图像的生成; 在医学领域应用中的模型: RefineGAN, 基于 CGAN 生成 Multi-contrast MRI 图像的模型, MedGAN, 基于 GAN 生成视网膜血管图, 基于 WGAN 生成冠状动脉血管图应用于 CCTA; 在隐私保护应用中的模型: 基于 GAN 的攻击手段以提高安全性能; 在域自适应学习领域应用中的模型: PixelDA, 基于 GAN 学习源域与目标域的联合嵌入空间特征; 在自动驾驶应用中的模型: 基于 GAN 的驾驶场景预测以及基于 VAE 与 GAN 的未来路况预测. 表 4 汇总了本节中不同模型的分类情况.

这些领域主要还是集中在图像的生成方面, 将生成图像作为样本并参与到相应的领域应用中, 具有不错的效果. GAN 的提出不仅只是停留在理论与计算机领域, 其应用更会直接影响到人们的日常生活, 如 GAN 在医学、自动驾驶领域的实际应用.

表 4 GAN 在其他领域的应用
Table 4 GAN's application in other fields

内容	模型
人体姿态估计	基于 RL 与 GAN 的姿态估计模型[96], 基于 GAN 的姿态估计模型[97], 基于双向 LSTM 的 CGAN 模型[98]
恶意软件检测	MalGAN[99]
数据集标记与数据增强	基于 GAN 的仿真无监督学习框架[100], RenderGAN[101], DAGAN[102]
物理应用	基于 GAN 的高能粒子物理图像生成模型[104]
医学领域	RefineGAN[105], 基于 CGAN 的多对比度 MRI 图像生成模型[106], MedGAN[107], 基于 GAN 的视网膜血管图像生成模型[108], 基于 WGAN 的 CCTA 模型[109]
隐私保护	基于 GAN 的用户信息攻击模型[110]
域适应学习领域	PixelDA[112], 基于 GAN 的域自适应分类任务[113], 基于 GAN 的域间联合嵌入特征空间模型[114]
自动驾驶	基于 GAN 的驾驶场景预测模型[115], 基于 VAE 与 GAN 的路况预测模型[116]

5 总结及未来趋势与展望

5.1 总结

鉴于 GAN 的理论研究意义和实际应用价值, 本文对 GAN 在图像领域、语音领域、NLP 领域进行了系统综述. 本文首先介绍了 GAN 的基本理论及其训练方式, 并提出了 GAN 面临的三大问题: 训练过程难收敛、模式崩溃以及没有可靠的快速收敛信息. 根据这三大问题, 详细介绍了 9 种主流 GAN 变种模型: CGAN, DCGAN, WGAN, WGAN-GP, InfoGAN, SeqGAN, Pix2Pix, CycleGAN, Augmented CycleGAN, 并将这 9 种 GAN 变种模型进行了分类. 主要可以分为三类: 损失函数改进型 GAN、网络结构改进型 GAN 以及前两者混合型 GAN.

在图像应用领域, 主要分为 13 类应用, 在人脸识别与图像生成应用中介绍了 Age-GAN 和 GLCA-GAN; 在多姿态人脸转正应用中介绍了 TP-GAN 和 DRGAN; 在图像纹理合成应用中介绍了 SGAN、MGAN 和 BigGAN. 在图像超分辨率应用中介绍了 SRGAN 和基于条件信息的 c-CycleGAN; 在图像复原与多视角图像生成应用中介绍了 PGGAN 和 VariGAN; 在图像转换应用中介绍了基于对偶学习的 DualGAN、GeneGAN、S^2-GAN 和 DA-GAN; 在文本描述到图像生成应用中介绍了 Text to image GAN、GAWWN 和 RTT-GAN; 在图像语义分割应用中介绍了基于 GAN 的语义分割模型 Contrast-GAN; 在图像着色应用中介绍了 Auto-painter; 在视频预测领域应用中介绍了 Dual motion GAN; 在视觉显著性预测应用中介绍了 SalGAN 和 MC-GAN; 在图像密写应用中介绍了 S-GAN; 在 3D 图像生成应用中介绍了 3D-GAN 和 VON.

在语音与 NLP 应用领域, 主要分为 9 类应用, 在语音增强应用中介绍了 SEGAN; 在音乐与音符生成应用中介绍了 MuseGAN; 在语音识别应用中介绍了基于 WGAN 的 ASR 中的应用, Voice-GAN、MTGAN 以及 Residual GAN 等; 在对话模型的评估与生成应用中介绍了基于 Seq2Seq 与 RL 的 GAN模型; 在生成离散型序列应用中介绍了 Gumbel-softmax GAN; 在双语字典应用中介绍了单向 GAN 模型、双向 GAN 模型, 以及对抗自编码 GAN 模型; 在文本生成与分类应用中介绍了 DP-GAN, 以及基于 WGAN 与 RNN 的 GAN 模型; 在语篇分析应用中介绍了 ADAN; 在机器翻译应用中介绍了 BR-CSGAN、Multi-CSGAN-NMT、Adversarial-NMT 和 BGAN-NMT.

另外, 对 GAN 在其他应用领域中的模型进行了梳理, 主要分为 8 类, 在人体姿态估计应用中介绍了基于 RL 与 GAN 的模型, 以及引入 BLSTM 的 CGAN 模型; 在恶意软件检测应用中介绍了 MalGAN; 在数据集标记与数据增强应用中介绍了基于 GAN 的仿真无监督学习框架 RenderGAN, 以及 DAGAN; 在物理应用中介绍了将 GAN 应用于高能粒子物理学中喷射图像的生成; 在医学领域应用中介绍了 RefineGAN, 基于 CGAN 生成 Multi-contrast MRI 图像的模型 MedGAN, 基于 GAN 生成视网膜血管图, 以及基于 WGAN 生成冠状动脉血管图应用于 CCTA; 在隐私保护应用中介绍了基于 GAN 的攻击手段以提高安全性能; 在域自适应学习应用中介绍了 PixelDA, 以及基于 GAN 学习源域与目标域的联合嵌入空间特征; 在自动驾驶应用中介绍了基于 GAN 的驾驶场景预测以及基于 VAE 与 GAN 的未来路况预测.

5.2 未来趋势与展望

GAN 在各领域中的应用取得了令人鼓舞的成

就, 但是 GAN 本身的模型、GAN 的神经网络结构, 以及 GAN 的训练算法均存在值得研究的问题, GAN 在各实际领域的应用也只是刚刚开始, 还有许多值得探索的未知领域. GAN 本身的模型、GAN 的神经网络结构, 以及 GAN 的训练算法的改进, 必然带来应用效果的提升; GAN 在各实际领域的应用遇到的问题, 也促使 GAN 本身理论、模型以及训练算法的发展. 以下本文尝试给出一些值得探索的问题.

1) GAN 仍存在着一些问题, 如上文所提到的, 主要有生成器与判别器难收敛、模式崩溃以及训练不稳定等问题. 尽管 WGAN 对梯度消失和训练不稳定问题的研究有了一定进展, 但这仍是目前需要解决的问题, 所以提出新的 GAN 模型, 设计新的网络结构, 修改训练算法仍是需要研究的方向之一. 若 GAN 不在训练技巧和训练方法上下功夫, 会导致 GAN 被别的更好的模型所取代.

2) GAN 模型对超参数十分敏感, 轻微超参数变化会对训练结果产生很大的影响, 当确定模型以及编写代码实现后, 剩下主要的工作就是调整超参数, 如何确定一个适合特定应用领域的模型的超参数是 GAN 需要亟待解决的问题之一. 实现自动确定 GAN 的网络结构, 用训练数据和学习目标函数自动学习数据依赖的、任务依赖的 GAN 网络结构和模型的超参数是值得探讨的问题.

3) 从本文中可以看到, 很多与 GAN 结合的生成式模型, 例如 VAE 与 GAN 进行结合, 利用强化学习的 SeqGAN, 与卷积神经网络结合的 DCGAN 等. 这些模型的提出, 不仅完善了 GAN, 而且在这些模型上的应用也大量出现. 所以将生成式模型与 GAN 进行融合创新, 不仅能提高 GAN 理论完备程度, 还会丰富应用种类和领域.

4) 如何使得 GAN 生成的图像、文本和语音具有多样性是一个值得研究的问题. 度量多样性最基本的标准是熵, 因而把 GAN 的训练目标函数变为与熵有关的函数, 如最大互信息; 考虑输入数据的结构, 把训练数据看作服从多个概率分布的子数据集; 把数据看作不同噪声混合后的随机变量, 提取不同噪声级别的特征表示, 得到不同层次的特征表示; 在训练目标函数里显式地引入不同的归纳偏置, 得到多个模型, 都可以尝试用来生成具有多样性的图像、文本和语音.

5) 机器学习理论认为好的模型应具有更好的泛化能力. 重新思考深度学习的泛化能力, 从模型复杂性、偏差–方差权衡等观点出发, 理论上讨论生成对抗网络的学习机制, 给出生成对抗网络之所以成功的理论基础, 从而真正确立生成对抗网络在深度学习中的显著地位, 是值得思考的问题.

6) GAN 在计算机视觉领域发展迅猛, 但 GAN 的图像处理能力还十分依赖计算性能, 计算能力的增强将会使得 GAN 具有更加出色甚至完美的生成效果, 例如: BigGAN 可以产生超逼真图片. 但这是一把双刃剑, 需要十分昂贵的计算机硬件设备, 以及长时间的训练. 这需要改进 GAN 网络结构, 提出维持训练过程稳定的手段, 对硬件设施进行特殊优化, 提高速度并节约硬件设施及学习时间. 所以对 GAN 的训练进行改进并更加高效地与硬件设施进行结合也是未来发展方向之一.

7) 在 NLP 领域, 由于 GAN 不善于处理离散型序列及文本数据, 导致 GAN 在这个应用领域不是特别占优势. 但是也有很多优秀的应用, 例如 SeqGAN的提出使得 GAN 在 NLP 领域崭露头角; 机器翻译领域利用 NMT 与 GAN 进行结合的思想, 能够让 GAN 充分发挥作用. 所以研究 GAN 在 NLP 领域的特定模型, 也是发展方向之一.

8) 随着时代的快速发展, 身边的智能助手越来越多, 如智能家居物联网、语音助手、聊天机器人、自动驾驶等领域. GAN 在自动驾驶领域可能刚刚被拓展, GAN 在这些领域的应用还不是很多, 如何将 GAN 的思想以及成果运用在目前生活中最常见的场景中, 智能家居物联网、语音助手等都是日常使用频率最高的应用场景, 如何加速 GAN 与这些领域的融合, 让智能家居及手机拥有自己的生成对抗模型, 需要发展智能助手与 GAN 的融合模型, 所以这也是 GAN 的发展方向之一.

9) GAN 也会与 RNN, LSTM 等对信息具有记忆能力的网络进行结合, 或许可以尝试将一些更加新颖的记忆网络, 例如, 动态记忆网络 (Dynamic memory networks, DMN)[117]、主动长期记忆网络 (Active long term memory networks, A-LTM)[118] 与 GAN 进行结合. 确保 GAN 在训练时, 对过去的信息具有记忆性, 使其保留更多可用信息, 提升 GAN 在 NLP 领域的效果.

10) 随着 GAN 越来越多地应用在不同的应用领域, 以及被引入到非传统机器学习研究领域, 其技术中难免会出现漏洞或者缺陷. 需要从实际角度挖掘及理解 GAN 当前的模型及相应变种. 例如, Kurach 等[119] 在实际应用中对各个 GAN 模型进行验证, 分析了 GAN 应用于大型数据集时采用相应手段的有效性, 以及常见的陷阱、GAN 生成模型的可重复性问题和需要实际考虑的因素等. 这项研究

得到 Goodfellow 等一些 AI 界的权威人士的赞同. 当在非传统机器学习任务中应用 GAN 时, 难免性能会不如传统学习方法. 这时需要改变思路, 不能盲目使用 GAN 方法. 因此, GAN 在实际问题上的验证评估也是未来方向之一.

11) 若想让 GAN 模型的理论在正确的方向上发展, 需要良好的评估系统对 GAN 变种模型进行评估. 所以需要提出不同应用领域中有效的评估指标, 组成实用的评估系统. 这样才能确保 GAN 具有正确的研究方向.

12) 目前 GAN 仅仅被引用于传统机器学习和人工智能专属领域, 对于工业生产领域的应用, 还比较鲜见. GAN 显然可以用于工业流程过程的缺失值补全, GAN 也可以用来构造软仪表. 对于工业领域普遍关注的多采样率和多尺度系统, GAN 显然可以用来补全慢时变过程所缺失的测量值, 以便与快时变过程相适应.

从以上的总结可以看出, GAN 的应用领域越来越广泛, 并且各领域的模型种类丰富多样, 需要我们尽快地跟踪和总结已有的研究内容, 为以后的 GAN 研究提供依据. 尽管 GAN 存在着问题和挑战, 但是不可否认, 随着 GAN 理论研究的深入和应用领域的拓展, GAN 将会成为未来人工智能领域中的主流机器学习技术.

References

1 Silver D, Huang A J, Maddison C J, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016, **529**(7587): 484−489

2 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 2014, **3**: 2672−2680

3 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: The state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321−332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. 自动化学报, 2017, **43**(3): 321−332)

4 Kingma D P, Welling M. Auto-encoding variational Bayes. arXiv: 1312.6114v10, 2013.

5 Ratliff L J, Burden S A, Sastry S S. Characterization and computation of local Nash equilibria in continuous games. In: Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing. Monticello, IL, USA: IEEE, 2013. 917−924

6 Larsen A B L, S\onderby S K, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York City, USA: ACM, 2016. 1558−1566

7 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico, 2015.

8 Salimans T, Goodfellow I J, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proceedings of the 2016 Advances in Neural Information Processing Systems. Barcelona, Spain, 2016. 2226−2234

9 Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks. arXiv: 1701.04862v1, 2017.

10 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv: 1411.1784, 2014.

11 Krizhevsky A, Sutskever I, Hinton G E. ImagEnet classification with deep convolutional neural networks. In: Proceedings of the 2012 Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA: Curran Associates Inc., 2012. 1106−1114

12 Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXiv: 1701.07875v3, 2017.

13 Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 2012, **4**: 26−31

14 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A C. Improved training of Wasserstein GANs. In: Proceedings of the 2017 Advances in Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates, Inc., 2017. 5767−5777

15 Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In: Proceedings of the 2016 Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016. Barcelona, Spain: Curran Associates Inc., 2016. 2172−2180

16 Yu L T, Zhang W N, Wang J, Yu Y. SeqGAN: Sequence generative adversarial nets with policy gradient. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI, 2017. 2852−2858

17 Liu Quan, Zhai Jian-Wei, Zhang Zong-Zhang, Zhong Shan, Zhou Qian, Zhang Peng, et al. A survey on deep reinforcement learning. *Chinese Journal of Computers*, 2018, **41**(1): 1−27
(刘全, 翟建伟, 章宗长, 钟珊, 周倩, 章鹏, 等. 深度强化学习综述. 计算机学报, 2018, **41**(1): 1−27)

18 Isola P, Zhu J Y, Zhou T H, Efros A A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 5967−5976

19 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015. 234−241

20 Zhu J Y, Park T, Isola P, Efros A A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2242−2251

21 He D, Xia Y C, Qin T, Wang L W, Yu N H, Liu T Y, et al. Dual learning for machine translation. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 820−828

22 Almahairi A, Rajeswar S, Sordoni A, Bachman P, Courville A C. Augmented CycleGAN: Learning many-to-many mappings from unpaired data. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 195−204

23 Che T, Li Y R, Jacob A P, Bengio Y, Li W J. Mode regularized generative adversarial networks. arXiv: 1612.02136, 2016.

24 Gretton A, Borgwardt K M, Rasch M J, Schölkopf B, Smola A J. A kernel method for the two-sample-problem. In: Proceedings of the 2007 Advances in Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2007. 513−520

25 Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Proceedings of the 2017 Advances in Neural Information Processing Systems. Long Beach, CA, USA, 2017. 6626−6637

26 Lopez-Paz D, Oquab M. Revisiting classifier two-sample tests. arXiv: 1610.06545, 2016.

27 Xu Q T, Huang G, Yuan Y, Guo C, Sun Y, Wu F, et al. An empirical study on evaluation metrics of generative adversarial networks. arXiv: 1806.07755, 2018.

28 Gauthier J. Conditional generative adversarial nets for convolutional face generation. In: Proceedings of the 2014 Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition. 2014. 2

29 Antipov G, Baccouche M, Dugelay J L. Face aging with conditional generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE, 2017. 2089−2093

30 Li P P, Hu Y B, Li Q, He R, Sun Z N. Global and local consistent age generative adversarial networks. In: Proceedings of the 24th International Conference on Pattern Recognition. Beijing, China: IEEE, 2018. 1073−1078

31 Huang R, Zhang S, Li T Y, He R. Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2458−2467

32 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016. 694−711

33 Tran L, Yin X, Liu X M. Disentangled representation learning GAN for pose-invariant face recognition. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 1283−1292

34 Jetchev N, Bergmann U, Vollgraf R. Texture synthesis with spatial generative adversarial networks. arXiv: 1611.08207, 2016.

35 Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016. 702−716

36 Ulyanov D, Lebedev V, Vedaldi A, Lempitsky V. Texture networks: Feed-forward synthesis of textures and stylized images. In: Proceedings of the 33rd International Conference on Machine Learning. New York City, USA: ACM, 2016. 1349−1357

37 Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. arXiv: 1809.11096, 2018.

38 Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 105−114

39 Lu Y Y, Tai Y W, Tang C K. Attribute-guided face generation using conditional CycleGAN. arXiv: 1705.09966, 2017.

40 Chen Z M, Tong Y G. Face super-resolution through wasserstein GANs. arXiv: 1705.02438v1, 2017.

41 Yeh R A, Chen C, Lim T Y, Schwing A G, Hasegawa-Johnson M, Do M N. Semantic image inpainting with deep generative models. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 6882−6890

42 Demir U, Unal G. Patch-based image Inpainting with generative adversarial networks. arXiv: 1803.07422, 2018.

43 Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 2017, **36**(4): Article No. 107

44 Pathak D, Krähenbühl P, Donahue J, Darrell T, Efros A A. Context encoders: Feature learning by inpainting. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 2536−2544

45 Zhao B, Wu X, Cheng Z Q, Liu H, Jie Z Q, Feng J S. Multi-view image generation from a single-view. arXiv: 1704.04886, 2017.

46 Sohn K, Lee H, Yan X C. Learning structured output representation using deep conditional generative models. In: Proceedings of the 2015 Advances in Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015. 3483−3491

47 Yi Z L, Zhang H, Tan P, Gong M L. DualGAN: Unsupervised dual learning for image-to-image translation. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2868−2876

48 Zhou S C, Xiao T H, Yang Y, Feng D Q, He Q Y, He W R. GeneGAN: Learning object transfiguration and attribute subspace from unpaired data. arXiv: 1705.04932, 2017.

49 Wang X L, Gupta A. Generative image modeling using style and structure adversarial networks. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016. 318−335

50 Ma S, Fu J L, Chen C W, Mei T. DA-GAN: Instance-level image translation by deep attention generative adversarial networks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018. 5657−5666

51 Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks. In: Proceedings of the 2017 Advances in Neural Information Processing Systems. Long Beach, CA, USA:

Curran Associates Inc., 2017. 700−708

52 Reed S, Akata Z, Yan X C, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. In: Proceedings of the 33rd International Conference on Machine Learning. New York City, USA: ACM, 2016. 1060−1069

53 Reed S E, Akata Z, Mohan S, Tenka S, Schiele B, Lee H. Learning what and where to draw. In: Proceedings of the 2016 Advances in Neural Information Processing Systems. Barcelona, Spain, 2016. 217−225

54 Liang X D, Hu Z T, Zhang H, Gan C, Xing E P. Recurrent topic-transition GAN for visual paragraph generation. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3382−3391

55 Luc P, Couprie C, Chintala S, Verbeek J. Semantic segmentation using adversarial networks. arXiv: 1611.08408, 2016.

56 Souly N, Spampinato C, Shah M. Semi and weakly supervised semantic segmentation using generative adversarial network. arXiv: 1703.09695v1, 2017.

57 Liang X D, Zhang H, Xing E P. Generative semantic manipulation with contrasting GAN. arXiv: 1708.00315, 2017.

58 Liu Y F, Qin Z C, Wan T, Luo Z B. Auto-painter: Cartoon image generation from sketch by using conditional Wasserstein generative adversarial networks. *Neurocomputing*, 2018, **311**: 78−87

59 Koo S. Automatic colorization with deep convolutional generative adversarial networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE, 2017. 212−217

60 Suárez P L, Sappa A D, Vintimilla B X. Infrared image colorization based on a triplet DCGAN architecture. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE, 2017. 212−217

61 Vondrick C, Torralba A. Generating the future with adversarial transformers. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 2992−3000

62 Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics. In: Proceedings of the 2016 Advances in Neural Information Processing Systems. Barcelona, Spain, 2016. 613−621

63 Liang X D, Lee L S, Dai W, Xing E P. Dual motion GAN for future-flow embedded video prediction. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 1762−1770

64 Pan J T, Ferrer C C, McGuinness K, O'Connor N E, Torres J, Sayrol E, et al. Salgan: Visual saliency prediction with generativse adversarial networks. arXiv: 1701.01081, 2017.

65 Fernando T, Denman S, Sridharan S, Fookes C. Task specific visual saliency prediction with memory augmented conditional generative adversarial networks. In: Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, USA: IEEE, 2018. 1539−1548

66 Volkhonskiy D, Borisenko B, Burnaev E. *Generative Adversarial Networks for Image Steganography*. 2016.

67 Wu J J, Zhang C K, Xue T F, Freeman W T, Tenenbaum J B. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016. 82−90

68 Zhu J Y, Zhang Z T, Zhang C K, Wu J J, Torralba A, Tenenbaum J B. Visual object networks: Image generation with disentangled 3D representation. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: ACM, 2018. 118−129

69 Pascual S, Bonafonte A, Serrà J. SEGAN: Speech enhancement generative adversarial network. In: Proceedings of the 18th Annual Conference of the International Speech Communication Association. Stockholm, Sweden: ISCA, 2017. 3642−3646

70 Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z, Smolley A P. Least squares generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2813−2821

71 Michelsanti D, Tan Z H. Conditional generative adversarial networks for speech enhancement and noise-robust speaker verification. In: Proceedings of the 18th Annual Conference of the International Speech Communication Association. Stockholm, Sweden: ISCA, 2017. 2008−2012

72 Dong H W, Hsiao W Y, Yang L C, Yang Y H. MuseGAN: Symbolic-domain music generation and accompaniment with multitrack sequential generative adversarial networks. arXiv: 1709.06298v1, 2017.

73 Sriram A, Jun H, Gaur Y, Satheesh S. Robust speech recognition using generative adversarial networks. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada: IEEE, 2018. 5639−5643

74 Shinohara Y. Adversarial multi-task learning of deep neural networks for robust speech recognition. In: Proceedings of the 17th Annual Conference of the International Speech Communication Association. San Francisco, CA, USA: ISCA, 2016. 2369−2372

75 Cai W, Doshi A, Valle R. Attacking speaker recognition with deep generative models. arXiv: 1801.02384, 2018.

76 Gao Y, Singh R, Raj B. Voice impersonation using generative adversarial networks. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada: IEEE, 2018. 2506−2510

77 Ding W H, He L. MTGAN: Speaker verification through multitasking triplet generative adversarial networks. In: Proceedings of the 19th Annual Conference of the International Speech Communication Association. Hyderabad, India: ISCA, 2018. 3633−3637

78 Juvela L, Bollepalli B, Wang X, Kameoka H, Airaksinen M, Yamagishi J, et al. Speech waveform synthesis from MFCC sequences with generative adversarial networks. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada: IEEE, 2018. 5679−5683

79 Kannan A, Vinyals O. Adversarial evaluation of dialogue models. arXiv: 1701.08198, 2017.

80 Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning

with neural networks. In: Proceedings of the 2014 Advances in Neural Information Processing Systems. Montreal, Quebec, Canada, 2014. 3104−3112

81 Li J W, Monroe W, Shi T L, Jean S, Ritter A, Jurafsky D. Adversarial learning for neural dialogue generation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, 2017. 2157−2169

82 Kusner M J, Hernández-Lobato J M. GANS for sequences of discrete elements with the gumbel-softmax distribution. arXiv: 1611.04051, 2016.

83 Zhang M, Liu Y, Luan H B, Sun M S. Adversarial training for unsupervised bilingual lexicon induction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017. 1959−1970

84 Mikolov T, Le Q V, Sutskever I. Exploiting similarities among languages for machine translation. arXiv: 1309.4168, 2013.

85 Zhang Y, Gaddy D, Barzilay R, Jaakkola T. Ten pairs to tag-multilingual POS tagging via coarse mapping between embeddings. In: Proceedings of the 2016 Association for Computational Linguistics. San Diego California, USA: Association for Computational Linguistics, 2016. 1307−1317

86 Liu P F, Qiu X P, Huang X J. Adversarial multi-task learning for text classification. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017. 1−10

87 Press O, Bar A, Bogin B, Berant J, Wolf L. Language generation with recurrent generative adversarial networks without pre-training. arXiv: 1706.01399, 2017.

88 Xu J J, Ren X C, Lin J Y, Sun X. DP-GAN: Diversity-promoting generative adversarial network for generating informative and diversified text. arXiv: 1802.01345, 2018.

89 Chen X L, Sun Y, Athiwaratkun B, Cardie C, Weinberger K. Adversarial deep averaging networks for cross-lingual sentiment classification. arXiv: 1606.01614, 2016.

90 Iyyer M, Manjunatha V, Boyd-Graber J, Daumé III H. Deep unordered composition rivals syntactic methods for text classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. Beijing, China: Association for Computational Linguistics, 2015. 1681−1691

91 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.

92 Yang Z, Chen W, Wang F, Xu B. Improving neural machine translation with conditional sequence generative adversarial nets. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana, USA: Association for Computational Linguistics, 2018. 1346−1355

93 Yang Z, Chen W, Wang F, Xu B. Generative adversarial training for neural machine translation. *Neurocomputing*, 2018, **321**: 146−155

94 Wu L J, Xia Y C, Zhao L, Qin T, Lai J H, Liu T Y. Adversarial neural machine translation. In: Proceedings of the 10th Asian Conference on Machine Learning. Beijing, China: PMLR, 2018. 534−549

95 Zhang Z R, Liu S J, Li M, Chen E H. Bidirectional generative adversarial networks for neural machine translation. In: Proceedings of the 22nd Conference on Computational Natural Language Learning. Brussels, Belgium: Association for Computational Linguistics, 2018. 190−199

96 Merel J, Tassa Y, TB D, Srinivasan S, Lemmon J, Wang Z Y, et al. Learning human behaviors from motion capture by adversarial imitation. arXiv: 1707.02201, 2017.

97 Chou C J, Chien J T, Chen H T. Self adversarial training for human pose estimation. arXiv: 1707.02439, 2017.

98 Sadoughi N, Busso C. Novel realizations of speech-driven head movements with generative adversarial networks. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada: IEEE, 2018. 6169−6173

99 Hu W W, Tan Y. Generating adversarial malware examples for black-box attacks based on GAN. arXiv: 1702.05983, 2017.

100 Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W D, Webb R. Learning from simulated and unsupervised images through adversarial training. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 2242−2251

101 Sixt L, Wild B, Landgraf T. RenderGAN: Generating realistic labeled data. *Frontiers in Robotics and AI*, 2018, **5**: 66

102 Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks. arXiv: 1711.04340, 2017.

103 Huang G, Liu Z, Van Der Maaten L, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 2261−2269

104 de Oliveira L, Paganini M, Nachman B. Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis. *Computing and Software for Big Science*, 2017, **1**(1): 4

105 Quan T M, Nguyen-Duc T, Jeong W K. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. arXiv: 1709.00753, 2017.

106 Dar S U H, Yurt M, Karacan L, Erdem A, Erdem E, Çukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging*, 2019, **38**(10): 2375−2388

107 Choi E, Biswal S, Malin B, Duke J, Stewart W F, Sun J M. Generating multi-label discrete patient records using generative adversarial networks. arXiv: 1703.06490, 2017.

108 Son J, Park S J, Jung K H. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. arXiv: 1706.09318, 2017.

109 Wolterink J M, Leiner T, Isgum I. Blood vessel geometry synthesis using generative adversarial networks. arXiv: 1804.04381, 2018.

110 Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: Information leakage from collaborative deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas, TX, USA: ACM, 2017. 603−618

111 Liu Jian-Wei, Sun Zheng-Kang, Luo Xiong-Lin. Review and research development on domain adaptation learning. *Acta Automatica Sinica*, 2014, **40**(8): 1576−1600
(刘建伟, 孙正康, 罗雄麟. 域自适应学习研究进展. 自动化学报, 2014, **40**(8): 1576−1600)

112 Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 95−104

113 Zhang Y, Barzilay R, Jaakkola T S. Aspect-augmented adversarial networks for domain adaptation. *Transactions of the Association for Computational Linguistics*, 2017, **5**: 515−528

114 Sankaranarayanan S, Balaji Y, Castillo C D, Chellappa R. Generate to adapt: Aligning domains using generative adversarial networks. arXiv: 1704.01705, 2017.

115 Ghosh A, Bhattacharya B, Chowdhury S B R. Sad-GAN: Synthetic autonomous driving using generative adversarial networks. arXiv: 1611.08788, 2016.

116 Santana E, Hotz G. Learning a driving simulator. arXiv: 1608.01230, 2016.

117 Furlanello T, Zhao J P, Saxe A M, Itti L, Tjan B S. Active long term memory networks. arXiv: 1606.02355, 2016.

118 Kumar A, Irsoy O, Ondruska P, Iyyer M, Bradbury J, Gulrajani I, et al. Ask me anything: Dynamic memory networks for natural language processing. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York City, NY, USA: ACM, 2016. 1378−1387

119 Kurach K, Lucic M, Zhai X H, Michalski M, Gelly S. A large-scale study on regularization and normalization in GANs. arXiv: 1807.04720, 2018.

刘建伟 博士, 中国石油大学 (北京) 副研究员. 主要研究方向为智能信息处理, 机器学习, 复杂系统分析, 预测与控制, 算法分析与设计. 本文通信作者. E-mail: liujw@cup.edu.cn
(**LIU Jian-Wei** Ph.D., associate professor in the Department of Automation, College of Information Science and Engineering, China University of Petroleum (Beijing). His research interest covers intelligent information processing, machine learning, analysis, prediction, controlling of complicated nonlinear system, and analysis of the algorithm and the designing. Corresponding author of this paper.)

谢浩杰 中国石油大学 (北京) 信息科学与工程学院硕士研究生. 主要研究方向为机器学习.
E-mail: xhj19941116@163.com
(**XIE Hao-Jie** Master student in the Department of Automation, College of Information Science and Engineering, China University of Petroleum (Beijing). His main research interest is machine learning.)

罗雄麟 博士, 中国石油大学 (北京) 教授. 主要研究方向为智能控制和复杂系统分析, 预测与控制.
E-mail: luoxl@cup.edu.cn
(**LUO Xiong-Lin** Ph.D., professor in the Department of Automation, College of Information Science and Engineering, China University of Petroleum (Beijing). His research interest covers intelligent control, and analysis, prediction, controlling of complicated nonlinear system.)