生成对抗网络及其在图像生成中的应用研究综述

陈佛计[1),2),3),4)] 朱 枫[1),2),4)] 吴清潇[1),2),4)] 郝颖明[1),2),4)]
王恩德[1),2),4)] 崔芸阁[1),2),3),4)]

[1)](中国科学院沈阳自动化研究所 沈阳 110016)
[2)](中国科学院机器人与智能制造创新研究院 沈阳 110016)
[3)](中国科学院大学 北京 100049)
[4)](中国科学院光电信息处理重点实验室 沈阳 110016)

摘 要 生成对抗网络(GAN)是无监督学习领域最近几年快速发展的一个研究方向,其主要特点是能够以一种间接的方式对一个未知分布进行建模.在计算机视觉研究领域中,生成对抗网络有着广泛的应用,特别是在图像生成方面,与其他的生成模型相比,生成对抗网络不仅可以避免复杂的计算,而且生成的图像质量也更好.因此,本文将对生成对抗网络及其在图像生成中的研究进展做一个小结和分析:本文首先从模型的架构、目标函数的设计、生成对抗网络在训练中存在的问题、以及如何处理模式崩溃问题等角度对生成对抗网络进行一个详细地总结和归纳;其次介绍生成对抗网络在图像生成中的两种方法;随后对一些典型的、用来评估生成图像质量和多样性的方法进行小结,并且对基于图像生成的应用进行详细分析;最后对生成对抗网络和图像生成进行总结,同时对其发展趋势进行一个展望.

关键词 生成模型;生成对抗网络;图像生成;生成图像质量评估
中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2021.00347

A Survey About Image Generation with Generative Adversarial Nets

CHEN Fo-Ji[1),2),3),4)] ZHU Feng[1),2),4)] WU Qing-Xiao[1),2),4)] HAO Ying-Ming[1),2),4)]
WANG En-De[1),2),4)] CUI Yun-Ge[1),2),3),4)]

[1)](*Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang* 110016)
[2)](*Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang* 110016)
[3)](*University of Chinese Academy of Sciences, Beijing* 100049)
[4)](*Key Laboratory of Opto-Electronic Information Process, Chinese Academy of Sciences, Shenyang* 110016)

Abstract In tasks of unsupervised learning, the generative model is one of the most critical techniques. The generative model consists of probability density estimation and sampling, which can learn data distribution by looking at existing samples and generate new samples that obey the same distribution as the original samples. For complex distributions in a high dimensional space, density estimation and sample generation are often hard to realize. Since high-dimensional random vectors are generally difficult to model directly, it is necessary to simplify the model with some condition independence hypothesis. Even given a complex distribution that has been modeled, there is a lack of effective sampling methods. With the rapid development of deep neural network technology, the generative model has made great progress. In the past few years, there has been

收稿日期:2019-08-28;在线发布日期:2020-05-19.本课题得到国家自然科学基金(U1713216)和机器人学重点实验室自主课题项目(2017-Z21)资助.**陈佛计**,硕士,主要研究方向为图像生成、机器学习、模式识别、视觉测量.E-mail:chenfoji@sia.cn.**朱 枫**(通信作者),博士,研究员,博士生导师,主要研究领域为机器人视觉、视觉测量、视觉检测、红外图像仿真、3D 物体识别.E-mail:1754208529@qq.com.**吴清潇**,博士,研究员,硕士生导师,主要研究领域为机器人视觉、机器视觉.**郝颖明**,博士,研究员,硕士生导师,主要研究领域为图像处理、空间视觉测量.**王恩德**,博士,研究员,硕士生导师,主要研究领域为小型飞行器控制、图像目标检测、识别与跟踪、微弱信号检测预处理.**崔芸阁**,硕士,主要研究方向为 SLAM 和图像生成.

a drastic growth of research in Generative Adversarial Network (GAN) which can model an unknown distribution in an indirect way and can avoid statistical and computational challenges. At the same time, generative adversarial networks are the latest and most successful technology among generative models. Especially in terms of image generation, compared with other generation models, generative adversarial networks can not only avoid complicated calculations, but also generate better quality images. Therefore, this paper will make a summary and analysis of generative adversarial networks and its applications in image generation. Firstly, from the theoretical aspect, the basic idea and working mechanism of generative adversarial networks are explained in detail; How to design the loss function of generative adversarial networks based on F-divergence or integral probability metric is introduced, and its advantages and disadvantages are summarized; From the two aspects of convolutional neural network structure and auto-encoder neural network structure, the model structure commonly used in generating adversarial networks is summarized; At the same time, the problems and corresponding solutions in the process of training generative adversarial networks are analyzed from both theoretical and practical perspectives; Secondly, based on the direct method and the integration method as the classification criteria, current methods of generating images based on generating adversarial networks are summarized, and the basic ideas of these methods are explained in details. Then, from the three aspects of image generation based on mutual information, image generation based on attention mechanism, and image generation based on a single image, the method of directly generating images based on random noise vectors is summarized. The current methods of generating images based on image translation are explained in details from the aspects of supervised and unsupervised methods. Later, from a qualitative and quantitative point of view, the existing methods used to evaluate the quality and diversity of generated images based on generative adversarial networks are analyzed, and contrasted. Finally, the application of generative adversarial networks in the field of small samples, data category imbalance, target detection and tracking, image attribute editing, and medical images processing is introduced in details. And some problems in theory and practice of generative adversarial networks and image generation are analyzed; The development trend of generative adversarial networks and the development trend of image generation are summarized and prospected.

Keywords enerative model; generative adversarial network; image generation; generate images quality assessment

1 引 言

生成模型是无监督学习任务中一类重要的方法.生成模型可以直接学习样本数据中的分布,然后从学到的分布中进行采样可以得到类似于样本数据、服从同一分布的样本.伴随着深度神经网络的快速发展,基于神经网络的生成模型取得了显著的成果.在神经网络兴起之前,生成模型主要是对数据的分布进行显式地建模,例如:基于有向图模型的赫姆霍兹机[1](Helmholtz machines)、变分自动编码器(Variational Auto-Encoder,VAE)[2]、深度信念网络[3](Deep Belief Network,DBN)等和基于无向图模型的受限玻尔兹曼机[4](Restricted Boltzmann Machines,RBM)、深度玻尔兹曼机[5](Deep Boltzmann Machines,DBM)等,以及自回归模型[6](AR 模型).由于被建模随机变量的高维度,学习十分困难.其主要体现在统计上的挑战和计算上的挑战,统计上的挑战就是这些生成模型不能很好地泛化生成的结果,计算上的挑战主要来自于执行难解的推断和归一化的分布.面对这些难以处理的计算,一种方法就是近似它们;另一种方法就是通过设计模型,完全避免这些难以处理的计算.基于这样的想法,研究者们提出了一系列新的模型,而由 Goodfellow 等人[7](2014)提出的生成对抗网络(Generative Adversarial Networks, GAN)是生成模型目前最好

的一种方法.

受博弈论中两人零和博弈思想的启发，GAN 主要由生成器和鉴别器两个部分组成. 生成器的目的是生成真实的样本去骗过鉴别器，而鉴别器是去区分真实的样本和生成的样本. 通过对抗训练来不断的提高各自的能力，最终达到一个纳什均衡的状态. 因为生成对抗网络在生成图像方面的能力超过了其他的方法，所以其成为了一个热门的研究方向. GAN 中的对抗学习思想逐渐与深度学习中的其他研究方向相互渗透，从而诞生了很多新的研究方向和应用. 相关综述性的文章包括：生成对抗网络教程[8](2016 NIPS)、Creswell 等人[9]的生成对抗网络综述、Kurach 等人[10]从损失函数、神经网络架构、正则化和归一化等角度做的综述、林懿伦等人[11]的生成式对抗网络、Zamorski 等人[12]的生成对抗网络的最新进展. 从这些文章中可以看出，关于生成对抗网络的研究主要是以下两个方面：

(1) 在理论研究方面，主要的工作是消除生成对抗网络的不稳定性和模式崩溃的问题. Goodfellow 在 NIPS 2016 会议期间做的一个关于 GAN 的报告中[8]，他阐述了生成模型的重要性，并且解释了生成对抗网络如何工作以及一些前沿的话题. Creswell 等人[9]在生成对抗网络的综述中，主要介绍了几种 GAN 的网络架构和 GAN 的应用，并且从信号处理的角度，除了确定训练和构造 GAN 的方法，还指出了在 GAN 的理论和实际应用中仍然存在的挑战. Kurach 等人[10]从损失函数、网络架构、正则化以及批标准化等角度对 GAN 的一些问题和可重复性进行了研究. 林懿伦等人[11]对 GAN 常见的网络结构、训练方法、集成方法、以及一些应用场景进行了介绍. Zamorski 等人[12]从学习隐空间表示的角度出发，对 GAN 最新的进展进行论述.

(2) 在应用方面，主要关注的是生成对抗网络在计算机视觉(CV)、自然语言处理(NLP)和其他领域的应用. 目前生成对抗网络在计算机视觉任务中已经有了很多的应用，例如：图像生成、语义分割、图像编辑、超分辨率、图像修复、域转换、视频生成和预测等；而生成对抗网络在自然语言处理中的应用也呈现日益增长的趋势，例如：从文本生成图像、字体的生成、对话生成、机器翻译等；同时生成对抗网络在语音生成方面也有一定应用. 生成对抗网络在视觉中的应用情况如表 1 所示. 在生成对抗网络的众多应用中，被研究最多的领域是图像生成，其目标是通过生成器来生成期望的图像.

表 1　GAN 在视觉任务中的应用

视觉任务	GAN 模型
图像转换	Pix2pix[13]，Cycle-GAN[14]，Disco-GAN[15] D2GAN[16]，ACGAN[17]
超分辨率	SRGAN[18]
属性编辑	SD-GAN[19]SL-GAN[20]，DR-GAN[21] AGE-GAN[22]，AttGAN[23]
目标检测	SeGAN[24]，Perceptual GAN [25]
视频生成	VGAN[26]，MoCoGAN[27]
图像修复	Generative Face Completion[28]
姿态估计	Pose Guided Person Generation Network (PG^2)[29]

本文首先介绍生成对抗网络的基本原理和存在的问题，以及针对存在问题做的改进. 其次对生成对抗网络在图像生成中应用，以及对生成图像的质量的评估进行探讨. 然后对基于图像生成的应用做一个详细介绍. 最后对生成对抗网络的发展趋势和其在图像生成领域中的应用进行展望.

2　GAN 的介绍

2.1　GAN 的工作机理

生成对抗网络由生成器(G)和鉴别器(D)两个部分组成，如图 1 所示. G 是由 θ 参数化的神经网络实现，G 的输入是一个服从于某一分布 p_z 的随机向量 $\boldsymbol{z}$，而 G 的输出可以看成是采样于某一分布 p_g 的一个样本 $G(\boldsymbol{z})$. 假设真实数据的分布为 p_{data}，在给定一定量真实数据集的条件下，对生成对抗网络进行训练，让 G 学到一个近似于真实数据分布的函数. GAN 中 G 的主要目的是生成类似于真实数据的样本以骗过 D，而 D 的输入由真实的样本和生成的样本两个部分组成，D 的目标就是判断输入的数据是来自于真实的样本还是来自于 G 生成的样本. G 和 D 经过对抗训练达到一个纳什平衡状态，即 D 判断不出其输入是来自于真实的样本，还是来自于 G 生成的样本，此时就可以认为 G 学习到了真实数据的分布. 在理论上，假设在生成对抗网络中真实数据分布为 $P_{\text{data}}(x)$，并且有一个被 θ 参数化的生成分布 $P_G(x;\theta)$. 如果想让真实数据分布和生成分布十分接近，首先从 $P_G(x;\theta)$ 随机采样数量为 m 的样

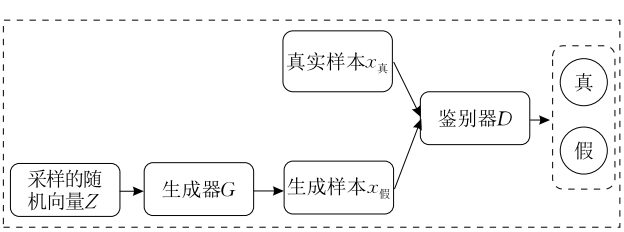


图 1　生成对抗网络结构

本，并且计算出 $P_G(x^i;\theta)$，最后通过最大似然函数：

$$L=\prod_{i=1}^{m}P_G(x^j,\theta) \tag{1}$$

来求出参数 θ；其结果如下：

$$\theta^*=\arg\min_{\theta}KL(P_{\text{data}}(x)\parallel P_G(x;\theta)) \tag{2}$$

知当 $D^*(x)=\dfrac{P_{\text{data}}(x)}{P_{\text{data}}(x)+P_G(x)}$ 时，似然函数的值最大；此时，将 $D^*(x)$ 代入到似然函数中就可以得到生成分布和真实数据分布之间的 Jensen Shannon 散度(JSD). 最终就可以将一个分布逼近另一个分布的问题转化为最小化两分布之间的 JSD. 基于这样的思想可以设计出 GAN 的目标函数如下式所示：

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

其中 $V(G,D)$ 是一个二分类的交叉熵函数，该损失函数的最终目标是最小化生成分布和真实分布之间的 KL 散度. 通过分析对抗网络的目标函数，并且从 D 的角度来看的话，如果 D 的输入是来自于真实样本，D 将会最大化输出；如果 D 的输入来自于 G 生成的样本，则 D 将会最小化输出；同时 G 想要去欺骗 D，那当 G 生成的样本作为 D 的输入的时候，必须最小化损失函数 $V(G,D)$. 但是当 D 被训练得非常好的时候，他将以很高的置信度直接将来自于 G 的样本判别为假. 此时 $\log(1-D(G(z)))$ 就会饱和，从而导致梯度为 0，最终参数得不到更新. 此时可以将 $\log(1-D(G(z)))$ 换成 $\log D(G(z))$. 尽管新的目标函数可以提供不同于原始损失函数的梯度，但是仍然存在梯度消失的问题；同时在理论上假设 D 和 G 具有充分的能力去对一个未知的分布进行建模，但实际上这种建模能力是有限的. 因此有很多学者尝试通过改变目标函数和神经网络结构等技巧来解决这些问题，接下来我们将分析这些 GAN 的变体，然后重点关注如何处理 GAN 训练中的存在问题以及模式崩溃.

2.2 GAN 的目标函数

GAN 的主要目标就是去最小化真实数据分布 P_{data} 与生成数据分布 P_g 之间的距离，怎么样度量分布之间的距离对于 GAN 极其关键. 标准的对抗网络通过 JSD 来度量两分布之间的差异，然而这种度量方式存在很多缺陷. 针对这些问题，研究人员最近几年提出了不同的距离度量方式和散度度量方式来代替 JSD，以提高 GAN 的性能. 这节我们将讨论如何基于这些距离或者散度的度量方式来对分布 P_{data} 与 $P_G(x;\theta)$ 之间的差异进行准确地度量. 目前常见的度量方式分为以下几类，如表 2 所示.

表 2 分布之间距离的度量方式

Metric	度量方式	GAN 模型
F-divergence	KLD, JSD	标准 GAN[7]
	Pearson χ^2	LSGAN[30]
Integral Probability Metric (IPM)	Wasserstein Distance	WGAN[31]
		WGAN-GP[32]
	Maximum Mean Discrepancy (MMD)	GMMN[33]
		MMDGAN[34]

在接下来的小节中，将按照表 2 的分类方式，分别对每一种方法进行详细地分析.

2.2.1 F 散度(F-divergence)

F 散度[35]是用一种特殊的凸函数 f 来度量两分布之间差异的一种方法，基于两分布之间的比值，可以将两分布之间的 F-divergence 定义为如下的形式：

$$D_f(P_{\text{data}}\parallel P_g)=\int P_g(x)f\left(\frac{P_{\text{data}}(x)}{P_g(x)}\right)\mathrm{d}x \tag{4}$$

在采用式(4)对两分布之间的差异进行度量时，必须满足这样的前提条件：$f(1)=0$ 并且 f 是一个凸函数，即当两个分布是一致的时候，其比值为 1，而相应的散度应该为 0. 由于任意满足 $f(1)=0$ 条件的凸函数，都可以衍生一种 GAN 的目标函数，这样就在很大程度上拓展了标准 GAN. 但实际操作过程中，并不能准确地求出数据分布的函数形式，所以应该采用一种可以计算的方法将式(4)给估计出来. f-GAN 采用了变分估计的方法来估计模型的参数，首先求出凸函数 f(也叫作生成器函数)的共轭函数 f^*，也称为 Fenchel 共轭，其形式如式(5)所示：

$$f^*(t)=\sup_{u\in\operatorname{dom}f}\{ut-f(u)\} \tag{5}$$

由于 Fenchel 共轭是可逆，也可以将 f 表示为

$$f(u)=\sup_{u\in\operatorname{dom}f^*}\{ut-f^*(t)\} \tag{6}$$

将式(6)代入到式(4)中可以得到 f 的下界，如式(7)所示：

$$D_f(p_{\text{data}}\parallel p_g)=\int_{\chi}p_g(x)\sup_{t\in\operatorname{dom}f^*}\left(t\frac{p_{\text{data}}(x)}{p_g(x)}-f^*(t)\right)\mathrm{d}x \tag{7}$$

$$\geq\sup_{T\in\mathrm{T}}\left(\int_{\chi}T(x)p_{\text{data}}(x)-f^*(T(x))p_g(x)\right)\mathrm{d}x \tag{8}$$

$$=\sup_{T\in\mathrm{T}}(E_{x\sim p_{\text{data}}}[T(x)]-E_{x\sim p_g}[f^*(T(x))]) \tag{9}$$

其中 f^* 是凸函数 f 的 Fenchel 共轭函数，$\operatorname{dom}f^*$ 是 f^* 的域. 因为最大值的和大于其和的最大值，所以式(7)可以变为式(8)；在公式中 T 表示满足 $\chi\to\mathbb{R}$ 的一类函数，因此可以用 $T(x)$ 来代替公式中的 t；而且 $T(x)$ 可以用式(10)来表示：

$$T(x)=a(D_w(x)),\ a(\cdot):\mathbb{R}\to \mathrm{dom}\, f^*$$
$$D_w(x):\chi\to\mathbb{R} \tag{10}$$

在式(10)中，可以将 $T(x)$ 看成是带有一个特别激活函数 $a(\cdot)$ 的鉴别器；用不同的生成器函数 f 以及与其对应的激活函数 $a(\cdot)$ 可以导出很多 GAN 的变体. 与标准 GAN 一样，f-GAN 首先最大化等式(9)关于 $T(x)$ 的下界，然后最小化近似的散度，使得生成器学到的分布更加类似于真实数据的分布. KL 散度(KLD)、逆 KL 散度、JSD 以及一些其他的散度都可以由带有特殊生成器函数 f 的 f-GAN 架构衍生出来. 在这些衍生出来的 GAN 的变体中，LSGAN[29] 的性能是最好的一个；但是在标准的 GAN 中，当生成器学到的分布 P_g 和真实的数据分布 P_{data} 之间没有交集的时候，即 P_g 距离 P_{data} 还是很远时，他仍会以很高的置信度将生成器生成的样本判别为假，此时就会导致目标函数值是一个常量，反向传播的时候梯度为 0，最终导致梯度消失. 基于这些问题 LSGAN 采用最小二乘损失函数来代替原始 GAN 中的交叉熵-损失函数. 最小二乘损失函数相较于容易饱和的交叉熵-损失函数有一个优势，即只在某一个点是饱和的. 最小二乘损失函数不仅骗过鉴别器，而且还让生成器把距离决策边界比较远的样本拉向决策边界. 类似于等式(3)，LSGAN[30] 的损失函数如下所示：

$$\min_D J(D)=\min_D 0.5\times E_{x\sim p_{\text{data}}}[D(x)-a]^2+0.5\times E_{z\sim p_z}[D(G(z))-b]^2 \tag{11}$$
$$\min_G J(D)=\min_G 0.5\times E_{z\sim p_z}[D(G(z))-c]^2 \tag{12}$$

其中，$D(x)$ 表示鉴别器的输出、$G(z)$ 表示生成器生成的样本，z 表示服从某一分布的随机向量. 常数 a、b 分别是表示生成图像和真实图像的标记；c 是生成器为了让鉴别器判定生成的图像是真实数据而设定的一个阈值. 因此，与标准 GAN 目标函数不同的一点是，最小二乘损失函数不仅仅对真实样本和生成的样本进行分类，而且还迫使生成的样本数据更加靠近真实数据的分布. 我们总结 LSGAN 的优势如下，首先是稳定了训练，解决了标准 GAN 在训练过程中容易饱和的问题；其次是通过惩罚远离鉴别器的决策边界的生成样本来改善生成图像的质量.

2.2.2　Integral Probability Metric (IPM)

IPM[36] 是与散度相似的一种、用来对两个分布之间的差异进行度量的一种方式，并且在 IPM 中定义了属于某一个特殊函数类 $\mathcal{F}$ 的评价函数 f. 在一个空间中 $\chi\subset\mathcal{R}^d$，$P(\chi)$ 是定义在 χ 上的概率测度，基于这个测度，P_g 和 P_{data} 之间的 IPM 可以被定义为下边的形式：

$$\mathcal{M}_{f\in\mathcal{F}}(P_{\text{data}},P_g)=\sup_{f\in\mathcal{F}}|E_{p_{\text{data}}(x)}[f]-E_{p_{g(x)}}[f]| \tag{13}$$

在式(13)中，基于评价函数 f 的度量标准 IPM 决定了 P_g 和 P_{data} 之间的差异的大小. 在这里评价函数可以用一个被 $\omega(x)$ 参数化的神经网络和激活函数 v 的乘积来表示. 如式(14)所示：

$$\mathcal{F}_{v,w}=\{f(x)=\langle v,w(x)\rangle \mid v\in R^m, w(x):\chi\to R^m\} \tag{14}$$

类似于 F 散度，基于不同的评价函数就有 IPM 的不同的变体，典型的性能比较好的变体有 Wasserstein distance metric[31] 和 Maximum Mean Discrepancy[37] (MMD). 接下来，分别对这两种距离度量方法进行详细地分析.

首先对 Wasserstein 距离进行详细地讨论，WGAN[31] 采用最优传输理论中 Wasserstein 距离(也称作 Earth-mover (EM)距离)来度量两个分布 P_g 和 P_{data} 之间的差异. 并且 Wasserstein 距离被定义为如式(15)所示的形式：

$$W(P_g,P_{\text{data}})=\inf_{\gamma\sim\Pi(p_g,P_{\text{data}})}E_{(x,y)\sim\gamma}[\|x-y\|] \tag{15}$$

其中 $\Pi(P_g,P_{\text{data}})$ 是 p_g 和 p_{data} 组合起来的所有可能的联合分布的集合. 对于每一个联合分布 γ 而言，可以从联合分布中采样，从而得到一个真实的样本 y 和一个生成的样本 x，并且求出这两个样本之间的距离 $\|x-y\|$，然后计算在联合分布 γ 下的期望值 $E_{(x,y)}[\|x-y\|]$. 最后在所有可能的联合分布中求出期望值的下界，而此下界就定义为 Wasserstein 距离；直观上可以将 Wasserstein 距离理解为在最优路径规划下的最小能量消耗. 由于直接对式(15)进行求解是很困难的，WGAN 利用 Kantorovich-Rubinstein duality 的技巧将式(15)转换成以下形式：

$$W(P_g,P_{\text{data}})=\sup_{\|f\|_l\le k}E_{x\sim P_{\text{data}}}(f(x))-E_{x\sim P_g}(f(x)) \tag{16}$$

在式(16)中 sup 表示的是一个上确界，$\|f\|_l\le k$ 表示的是评价函数必须满足 k 利普希茨(Lipschitz)连续性约束；这里的 Lipschitz 连续性要求指的是，对于一个连续函数 f 施加一个限制，并且存在一个常数 $k\ge 0$ 使得定义域内的任何两个元素 x_1 和 x_2 都满足如式(17)：

$$|f(x_1)-f(x_2)|\le k|x_1-x_2| \tag{17}$$

式(17)中的 k 称为是函数 f 的 Lipschitz 常数，实际上该连续性约束是为了限制连续型函数最大局

部变动的幅度.式(16)中的 f 函数可以用一个用 ω 参数化的、最后一层神经网络不用非线性激活函数的多层神经网络 f_w 来实现(其实就是鉴别器神经网络D).在限制权值 ω 不超过某个范围的条件下,使得

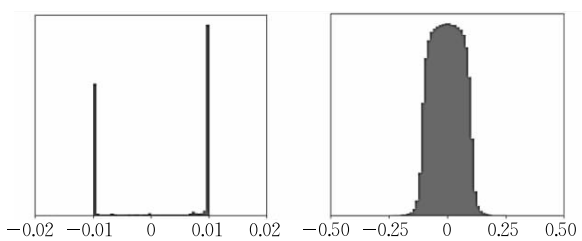
$$\mathcal{L}=E_{x\sim P_{\text{data}}}[f(x)]-E_{x\sim P_g}[f(x)] \quad (18)$$

尽可能最大,此时的$\mathcal{L}$就是近似真实分布和生成分布之间的 Wasserstein 距离.在实际实现的时候要注意,原始 GAN 的鉴别器做的是一个真假二分类的任务,所以最后一层需要添加一个非线性激活函数 sigmoid 函数.但是现在 WGAN 中的鉴别器是近似拟合 Wasserstein 距离,属于回归任务,神经网络的最后一层非线性激活函数要拿掉.我们的目标是要去最小化$\mathcal{L}$,因此基于式(18),可以设计出WGAN的损失函数如下所示:

G 的 Loss:$-E_{x\sim p_g}[f_w(x)]$ (19)

D 的 Loss:$E_{x\sim p_g}[f_w(x)]-E_{x\sim p_{\text{data}}}[f_w(x)]$ (20)

综上,采用 Wasserstein 距离来度量生成分布 P_g 和真实分布 P_{data} 之间差异的好处就是,当 P_g 和 P_{data} 之间没有交集或者是交集很小的时候,Wasserstein 距离不是一个常量,其仍然可以度量两分布之间的差异,所以很好地缓解了梯度消失的问题.但是在上述 WGAN 中,粗暴的权重裁剪会导致如下问题:在对抗网络中鉴别器的 Loss 是希望尽可能地拉大真假样本之间的差距,然后权重裁剪的策略又独立地限制每一个网络参数的取值范围,在这样的情况下就是让所有的参数走向极端,要么取最大值要么取最小值,导致参数值的分布很不均匀,如图 2(a)所示.针对这个问题,学者们又用梯度惩罚项来代替 WGAN 中的权重裁剪的技巧,通过限制鉴别器的梯度不超过 Lipschitz 常数 k 来构造梯度惩罚项.



(a) 梯度裁剪 (b) 梯度惩罚

图 2 WGAN[31]的权重值分布

改进后的损失函数,如式(21)所示:

$$\mathcal{L}=E_{x\sim p_g}[f_w(x)]-E_{x\sim p_{\text{data}}}[f_w(x)]+\lambda E_x[\|\nabla_x f_w(x)\|_p-1]^2 \quad (21)$$

通过图 2(b),可以观察到满足 k-Lipschitz 约束的梯度惩罚使得神经网络的参数分布得更加均匀.

接下来对最大平均差异[37](MMD)做深入的讨论.最大平均差异被提出时最先被用于双样本检测问题,用于判断两个分布P和Q是否一样,其基本思想是:对于所有以分布生成的样本空间为输入的函数 f,如果两个分布P和Q生成足够多的样本,并且这些样本在函数 f 下值的均值相等,那么就可以认为这两个分布是同一个分布.首先介绍一下希尔伯特空间 H,希尔伯特空间是一个完备的线性空间,同时也是一个内积空间.核 k 被定义为 $k:\chi\times\chi\to\mathbb{R}$,$k(y,x)=k(x,y)$.对于任意一个给定的正定核 $k(\cdot,\cdot)$,都存在一个唯一的函数空间 $f:\chi\to\mathbb{R}$,由于其满足再生性,因此也叫做再生核希尔伯特空间(Reproducing Kernel Hilbert Space,RKHS).再生性指:H_k 是一个希尔伯特空间,并且其满足以下特性:

$$\langle f,k(\cdot,x)\rangle_{H_k}=f(x),\ \forall x\in\chi,\ \forall f\in H_k \quad (22)$$

假设有一个满足 P 分布规律的数据集 $X^s=[x_1^s,\cdots,x_n^s]$和一个满足Q分布的数据集 $X^t=[x_1^t,\cdots,x_m^t]$;并且存在一个RKHS和一个核函数 $\phi(\cdot)$:$X\to H$ 可以将原始数据 X 从原始空间映射到再生核希尔伯特空间.因此 MMD 可以被表示为如式(23):

$$M(X_s,X_t)=\left\|\frac{1}{n}\sum_{i=1}^{n}\phi(x_i^s)-\frac{1}{m}\sum_{i=1}^{m}\phi(x_1^t)\right\| \quad (23)$$

通过式(23)可以看出,其原理就是对每一个真实样本和生成的样本进行投影并求和、利用和的大小对 P_g 和 P_{data} 之间的差异进行度量.类似于IPM,MMD在参数空间是处处连续、可微的,同时在 IPM 的框架下 MMD 也可以被理解,此时其函数类是 $\mathcal{F}=\mathcal{H}_k$.

2.2.3 IPM 度量标准和 F 散度度量标准的比较

对于 F 散度来讲,在式(4)中被定义的、带有凸函数 f 的 f 散度函数族,当数据空间中的维数 d 逐渐增加的时候,f 散度是很难被估计的,并且两个分布的支撑集是没对齐的,又会导致散度的值趋向于无穷大.尽管等式(9)推导出了等式(4)的变分下界,但是在实践中不能保证变分下界对真实散度的收紧性,从而会导致不正确、甚至是有偏的估计.

Sriperumbudur等人[36]研究表明在 f-divergence 族和 IPM 族之间唯一的交集就是 Total Variation Distance;因此 IPM 族也没有继承 f-divergence 族的缺点;他们也证明了在使用独立同分布样本的情况下,IPM 估计器是在收敛性方面更加一致.

在实践应用中,更多采用 IPM 度量标准来对真实分布和生成分布之间的差异进行度量.与 F 散度度量标准相比,IPM 度量标准有以下优点:

(1) IPM 度量标准不会受到数据高维的影响;

(2) 始终可以反映两分布之间的真实距离，即使是两分布的支撑集没有相应的交集，IPM也不会发散.

2.3　GAN 的模型结构

目前在对抗网络中应用最为广泛的两种神经网络结构分别是卷积神经网络结构和自动编码神经网络结构.

基于卷积神经网络搭建的对抗网络，生成器由多层反卷积网络层构成，而鉴别器由多层卷积网络层构成. DCGAN[38] 是首先采用该结构的模型，其结构如图 3 所示，同时该模型也加入了批量正则化的技巧来帮助稳定 GAN 的训练；由于 DCGAN 良好的性能，基于此网络结构提出了很多新的方法；例如：pix2pix[13]、cycle-GAN[14] 等；同时受 DCGAN 思想的启发，Daniel 等人[39] 使用递归神经网络 (RNN) 去生成图像. 通过交替对抗训练，DCGAN 可以生成质量很高的图像.

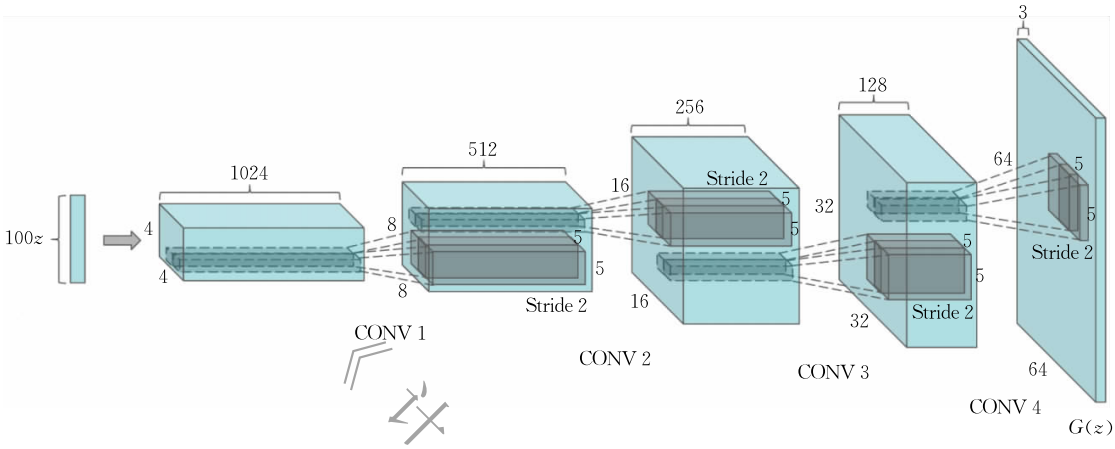
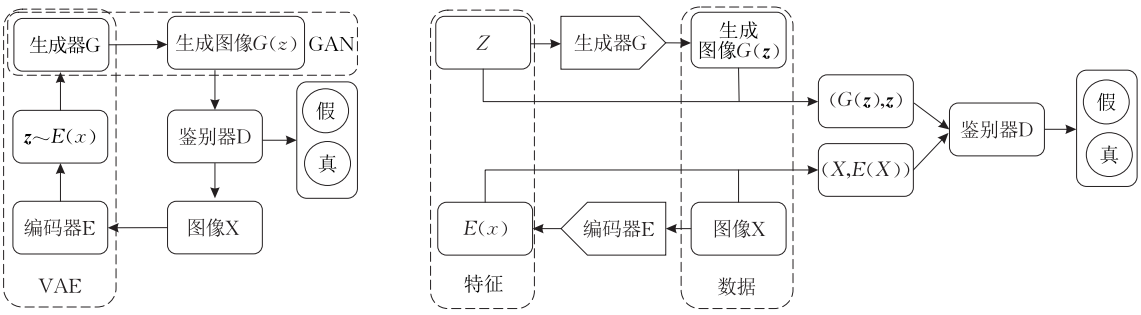


图 3　DCGAN 的网络结构[38]

自动编码网络[40]，如图 4，是一种用于无监督学习的自重构神经网络，并且将输入作为目标值，用自监督学习方法来进行训练. 自重构的目的是去学习输入数据的高维特征或者是压缩表示. VAE-GAN[41] 用鉴别器 D 来表示 VAE[2] 的重构损失，从而可以结合变分自编码器和 GAN 两者的优势去生成高质量的图像，最终该模型生成的图像要比单独用 VAE 或者是单独用GAN生成的图像质量要好. Bi-GAN[42] 模型用 Encoder- Decoder 结构来实现对抗网络中的生成器，该结构中的 Encoder，输入是真实图片，输出是一个隐编码；而与 Encoder 无关的 Decoder，其输入一段给定编码，输出是一张图片；同时有一个鉴别器，其输入是图像和隐编码组成的配对，它需要去判断这个配对是来自 Encoder 还是 Decoder. 该模型的目标就是让来自于 Encoder 配对的分布 $P(x, z)$ 和来自于 Decoder 配对的分布 $Q(x', z')$ 之间的距离越来越小，在 Bi-GAN[42] 中 Encoder 和 Decoder 就是一个互为逆运算的过程，从而更好地实现重构. EBGAN[43] 是一个由编码器、解码器和鉴别器三个部分组成的生成模型，其中鉴别器的作用是判断解码器对输入图像重构性的高低，而编码器、解码器组成 Autoencoder，该 Autoencoder 提前用真实图片进行预训练. 基于预训练好的 Autoencoder 网络和鉴别器网络，即可搭建该模型的对抗网络结构.



(a) VAE-GAN[41]的网络结构　　(b) Bi-GAN[42]的网络结构

图 4　自动编码生成对抗网络

采用卷积神经网络和转置卷积神经网络来搭建生成对抗网络是大多数人采用的方法. 如果想对隐变量空间进行一个很好的探索，例如对图像进行属性编辑，基于自动编码网络的 GAN 是一个最佳的选择.

2.4 训练 GAN 存在的问题以及应对策略

尽管 GAN 在某些方面取得了令人满意的效果，但是其在理论和实践中还是存在一些缺陷.

在理论方面，标准的对抗网络是用 KL 散度或者是 JSD 来度量真实数据分布和生成数据分布之间的差异. 由于这种度量方式在某些状态下是饱和的，梯度消失的问题就会发生，同时 KL 散度的不对称性使得对抗网络宁可丧失生成器生成模式的多样性，也不愿丧失鉴别器的准确性，最终导致模型的模式崩溃问题.

在实践过程中，对生成器生成图像质量的好坏的评估还没有一个统一的标准；并且在训练对抗网络的过程中无法根据损失函数的值来判断模型是否收敛；同时很难量化地判断生成器在什么样的条件下能够生成高质量的图像.

为了更好地生成图像，研究者们提出了相应的方法来解决上述在训练对抗网络中存在的问题. 比如采用替代损失函数的方法来改善梯度消失问题，Wasserstein GAN 提出用 EM 距离来替代标准 GAN 中的 JSD. 使用 EM 距离的优势在于，即使是真实数据分布和生成数据分布不相交，他也能很好地度量两者之间的差异. LSGAN 用的另一种方法是使用均方损失替代标准 GAN 中的对数损失，其目的是对距离决策边界较远的样本进行一个惩罚，使生成数据的分布更加接近于真实数据的分布.

针对模式崩溃的问题，DRAGAN[44] 采用梯度惩罚的方式来避免 GAN 的博弈达到一个局部平衡的状态，极大地增强 GAN 的稳定性，尽可能地减少模式崩溃问题的产生. Unrolled GANs 在更新参数的时候不是仅仅用当前的梯度值，而且是用前几次梯度值的加权和来对当前的参数值进行更新，从而以此方法来预防模式崩溃的问题. Pac-GAN[45] 将多个属于同一类的样本进行打包，然后传递给鉴别器，来减少模式崩溃现象的发生. 还有就是用集成的方法来处理模式崩溃的问题，一个 GAN 可能不足以有效地处理任务，因此学者们就提出用多个连续 GAN，其中每一个 GAN 解决任务中的一小块问题. STAR-GAN[46] 先独立地训练 N 对局部 GAN，然后基于局部 GAN 去训练全局 GAN，从而保证全局 GAN 生成模式的多样性. 在对抗网络的损失函数中加入感知正则化项，则在一定的程度上可以改善生成图像的质量问题. 而对 GAN 生成图像质量的评估方法，将在后边的章节中进行介绍.

2.5 生成对抗网络的优势和劣势

从上述讨论中可以知道，在目前生成模型的各种方法中，生成对抗网络相较于其他的方法有以下优势：

(1) GAN 通过一种间接的方式来对未知的分布进行建模，从而避免无监督学习中难解的推断、难解的归一化常数等问题；所以 GAN 不需要引入下界来近似似然.

(2) GAN 可以并行地生成数据，与自回归模型相比，GAN 生成数据的速度比较快；同时 GAN 生成的图像还比较清晰.

(3) 在理论上，只要是可微分的函数都能够用于构建生成器和判别器，因而 GAN 能够与深度神经网络结合来构建深度生成式模型.

但是生成对抗网络也存在着如下劣势：

(1) 可解释性比较差，因为最终生成器学到的数据分布只是一个端到端的、黑盒子一样的映射函数，而且没有显式的表达式.

(2) 在实际应用中 GAN 比较难以训练，由于 GAN 需要交替训练生成器和鉴别器两个模块，因此两者之间的优化需要很好地同步.

(3) 可能发生模式崩溃的现象，导致生成器学到的模式仅仅覆盖真实数据中的部分模式，使得生成样本的多样性变低.

(4) 训练不稳定，神经网络需要良好的初始化，否则可能找不到最优解，导致学到的分布距离真实数据的分布仍然很远，并且无法根据损失函数的值来判断模型的收敛性.

3 基于 GAN 做图像生成的一般方法

GAN 在计算机视觉任务中应用最多的是图像生成，各种模型可以按照是否有监督和直接法或是集成法的分类方式分为以下几类，如表 3 所示.

表 3 图像生成方法分类

模型	有监督	无监督
直接法	CGAN[47]、Pix2pix[13]、Texture-GAN[48]、G2-GAN[49]、Bicycle-GAN[50]、Contour2image[51]、SPADE[52]、PLDT[53]	GAN[10]、WGAN[31]、Least-square GAN[30]、WGAN-GP[32]、f-AN[35]、DCGAN[38]、Unrolled GAN[54]、Improved-GAN[55]、Info-GAN[56]、Loss Sensitive GAN[57]、DTN[58]、UNIT[59]、Self-Attention GAN[60]
集成法	Stack-GAN[61]、SS-GAN[62]	Dual-GAN[63]、Triangle GAN[64]、Star-GAN[46]、Combo-GAN[65]、XGAN[66]、LAP -GAN[67]、LR-GAN[68]、SGAN[69]、

本节将从直接方法和集成方法两个方面来对基于 GAN 做图像生成的方法做一个汇总，并且最后对图像生成方法进行一个小结.

3.1　直接法

如图 5(a)所示，在这种图像生成方法中，对抗网络只有一个生成器和一个鉴别器，生成器直接学习一个逼近真实数据分布的分布，从学习到的分布中采样来生成样本. 其中 DCGAN[38] 是最为典型的一个模型，其结构已经被很多模型作为一个基准，例如：Info-GAN[56]，Text-to-image[70]、IC-GAN[71] 等模型；DCGAN 中生成器和鉴别器的模块结构如图 6 所示，生成器的网络模块使用转置卷积-批量正则化-ReLU 激活函数，而鉴别器的网络模块使用卷积-批量正则化-LeakyReLU 激活函数层. 这种方法设计和实现起来通常比较直接.

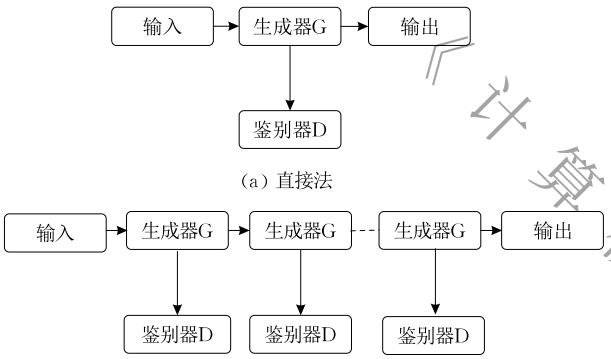


图 5　图像生成方法直接法和集成法的图示

图 6　搭建 DCGAN[38] 的网络模块

3.2　集成法

如图 5(b)所示，集成方法与直接进行图像生成的方法不同，集成方法模型的结构有以下几种形式：多个生成器一个鉴别器、一个生成器多个鉴别器、多个生成器和多个鉴别器. 集成方法的思想是去把视觉任务分成几个部分，然后每一个 GAN 去完成视觉任务的一部分，比如：可以用两个 GAN 去分别学习图像的内容和属性、前景和背景等，或者用多个 GAN 由粗到细、由小到大地去生成图像，而生成器之间的关系可以是迭代的，也可以是层次递进的. SS-GAN[62] 用了两个 GAN 来进行图像的生成，一个是结构 GAN，用来根据随机向量 $\boldsymbol{Z}$ 生成表面法线贴图；另一个是类型 GAN，以表面法线贴图和随机噪声 $\boldsymbol{Z}'$ 为输入来生成 2D 图像. 该方法首先生成图像的结构，然后基于图像结构再生成 2D 图像；结构 GAN 的实现方式采用和 DCGAN 一样的卷积模块，而类型 GAN 在实现方式上稍微有点不同. 类型 GAN 的生成器先让其输入量 $\boldsymbol{Z}'$ 和表面法线贴图先分别经过转置卷积层和卷积层的处理，最后将两个网络的输出合为一个向量，而合成后的向量作为类型 GAN 生成器的输入. 类型 GAN 的鉴别器以图像和图像表面法线向量在通道层面进行连接后的量作为输入. 在理想的情况下，生成器生成的图像和真实图像应该有相同的表面法线贴图；基于这一想法，SS-GAN 用一个全卷积神经网络来将生成图像再转变成表面法线贴图，并且基于此表面法线贴图构造一个重构损失作为损失函数的一个正则化项，从而约束生成器学到的分布向真实数据的分布靠近. LR-GAN[68] 的实现方法是使用不同的生成器去生成图像的前景内容和背景内容，而使用一个鉴别器来对图像进行判定；该模型通过实验证明了分别生成前景和背景内容，然后合成清晰的图像是实现图像生成的一种方法. 综上可知，SS-GAN 模型和 LR-GAN 模型都是集成了两个生成器，通过层级结构的方式来实现图像生成.

LAPGAN[67] 是用多个生成器由粗到细地来生成图像，底层的生成器以服从某一分布的随机向量作为输入，并且输出图像；其他的生成器都执行以下同样的功能：用前边生成器输出的图像和一个随机噪声向量作为输入，输出生成图像的细节，该细节可以被叠加到生成图像中，使得生成图像更加的清晰；除了底层生成器外，其他生成器唯一的不同之处是输入和输出维数的大小不一样. SGAN[69] 中集成了多种生成器，底层的生成器以随机噪声向量为输入，输出低层次特征向量；而中间层的生成器以低层次特征向量为输入，输出高层次特征向量；顶层的生成器以高层次特征向量为输入，输出生成图像. 并且 SGAN 在目标函数中加入了条件损失项和熵损失项，条件损失项可以帮助生成器有效地使用来自上一层的条件信息，熵损失项可以最大化生成器输出的条件熵的变分下界，这些约束项的加入可以很好地帮助生成器去生成图像. Stack-GAN[61] 有两个生成器，第一个生成器以随机噪声向量 $\boldsymbol{Z}$ 和类标签 C 组成的向量作为输入，输出是可以看出物体轮廓和模糊细节的模糊图像，而第二个生成器以第一个生

成器生成的图像和随机噪声向量以及类标签作为输入,然后生成一个逼真的图像.综上可知,LAPGAN、SGAN 和 Stack-GAN 都是集成了多个生成器,以迭代的方式来实现图像的生成.

与直接法相比,基于集成法来做图像生成,可以有效改善模式崩溃的问题,可以实现多个域之间的转换,同时还可以实现特征分离;但是基于集成法的模型训练起来会比较困难.

3.3 图像生成方法小结

基于 GAN 的图像生成主要考虑两个方面,分别是生成图像的质量和多样性.用标准 GAN 生成的图像,在质量和多样性方面存在着很多不足,所以针对这两个问题,很多方法基于 GAN 做出了改进:

(1)通过替代目标函数来改善生成图像的质量,例如:用 EMD 的距离度量方式来替代 JSD 或者是用均方损失函数替代对数损失函数.

(2)通过增加梯度惩罚项来改善生成图像的质量,该技巧不仅能缓解梯度消失或者是梯度惩罚的问题,而且可以极大地增强 GAN 的稳定性,尽可能地减少模式崩溃问题.类似的技巧还有谱归一化,该技巧比梯度惩罚更加高效.

(3)通过辅助信息来帮助改善生成图像的质量,例如:类标签信息等.

(4)通过搭建隐变量和观测数据之间的联系来改善生成图像的质量,比如:互信息等.

(5)在模型构建的时候,使用批量正则化的技巧,该技巧可以解决初始化差的问题,可以破坏原来的分布,在一定的程度上可以缓解过拟合.

(6)通过集成的方式来改善模式崩溃的问题,由于采用了多个生成器和判别器,它们之间有很多信息可以共享,从而可以提高生成器整体的学习能力.

4 基于随机向量生成图像

该方法的基本思想是用一个多层神经网络来实现一个非线性映射,该映射的功能是将一个服从某一分布的随机向量映射为采样于服从某一分布的图像.基于这样的思想,本小节将从基于互信息的图像生成,基于注意力机制的图像生成,以及基于单幅图像做图像生成三个方面来对这一图像生成的方法进行介绍.

(1)基于互信息的图像生成

在标准的对抗网络中,生成器的输入一般都是一段连续的单一的随机噪声向量;这样的情况下输入向量通常会被生成器进行过度地耦合处理,导致无法通过控制输入向量的某些维度来控制生成数据的语义特征.针对这一问题,Info-GAN 通过加入互信息正则化约束项来实现输入向量某些维度的可解释性,其模型结构如图 7 所示;该方法人为地将输入向量限制为随机噪声向量和隐向量两个部分,而且这些隐向量服从于某一先验的连续的或者离散的概率分布,用以表示生成数据的不同特征维度.

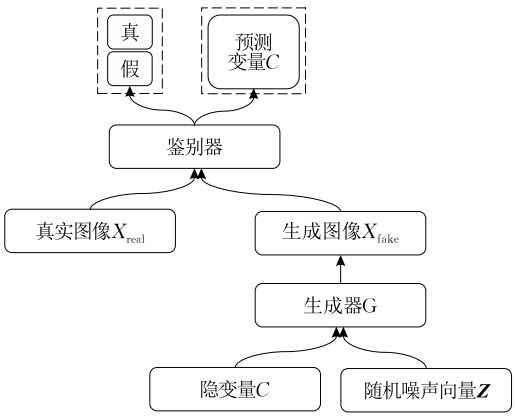


图 7 Info-GAN 的结构示意图

互信息是一种用来度量一个随机变量中包含的关于另一个随机变量的信息量,而该图像生成方法正是基于此度量方式来确定输出关于输入的信息量,从而实现对输入向量某些维度的可解释性.假设隐向量为 $\boldsymbol{C}$,而生成器的输出为 $G(\boldsymbol{Z},\boldsymbol{C})$,其中 $\boldsymbol{Z}$ 代表的是输入向量.因此输入和输出的互信息可以表示为如下形式:

$$I(\boldsymbol{C};G(\boldsymbol{Z},\boldsymbol{C}))=H(\boldsymbol{C})-H(\boldsymbol{C}\mid G(\boldsymbol{Z},\boldsymbol{C})) \tag{24}$$

式(24)中的 I 表示互信息,H 表示计算熵,并且基于互信息的图像生成方法的目标函数由对抗损失函数和互信息约束项两个部分组成,可以写成是如下形式:

$$\min_{G}\max_{D}V_I(D,G)=V(G,D)-\lambda I(\boldsymbol{C};G(\boldsymbol{Z},\boldsymbol{C})) \tag{25}$$

该模型的鉴别器有两个功能:一是辨别图像是否来自于真实数据分布;另一个是从图像中预测一个维度与输入隐向量相同的向量.在基于互信息正则项的约束下,该方法可以将隐向量中每一维度代表的特征信息学出来,因此该模型可以很好地解释输入向量中的隐变量.但是这种方法的可解释性仅仅局限在输入向量中人为添加的隐变量的那一小块,如果想对输入向量中所有维度代表的实际含义做出解释,这一方法就不适用了.

(2) 基于注意力机制的图像生成

注意力机制与人类对外界事物的观察机制很类

似，当人类观察外界事物的时候，一般不会把事物当成一个整体去看，往往倾向于根据需要选择性地去获取被观察事物的某些重要部分. 比如我们看到一个人的时候，往往是先注意到这个人的脸，然后再把不同区域的信息组合起来，形成一个对被观察事物的整体印象. 因此注意力机制可以帮助模型对输入向量的每一个部分赋予不同的权值，抽取出更加关键以及重要的信息，使模型做出更加准确的判断，同时不会对模型的计算和存储带来更大的开销.

在传统的 GAN 中，使用小的卷积核，导致难以发现图像中的依赖关系，使用大的卷积核，就会导致丧失了计算的效率，而注意力机制可以快速提取数据的重要特征，因此在 Self-Attention GAN[60] 中引入了自注意力机制；该机制的抽象数学模型如下，假定特征向量是 $\boldsymbol{X}$，首先该模型通过 1×1 的卷积分别对特征向量 $\boldsymbol{X}$ 做处理，从而得到 $f(x)$，$g(x)$，$h(x)$ 如果我们用 W_f，W_g，W_h 来表示网络的参数，则 $f(x)$，$g(x)$，$h(x)$ 可以被表示成式(26)：

$$f(x)=W_f x, g(x)=W_g x, h(x)=W_h x \quad (26)$$

通过式(27)来获得注意力权值：

$$\beta_{j,i}=\frac{\exp(s_{ij})}{\sum_{i=1}^{N}\exp(s_{ij})},\ s_{ij}=f(\boldsymbol{x}_i)^{\mathrm{T}}g(x_j) \quad (27)$$

基于注意力权值，进一步通过式(28)可以得到注意力特征映射：

$$o_j=\sum_{i=1}^{N}\beta_{j,i}h(x_i) \quad (28)$$

最后将式(28)融合到特征向量 $\boldsymbol{X}$ 中就得到带注意力机制的特征映射：$y_i=\gamma o_j+x_j$.

注意力机制可以将内部经验和外部感觉对齐，从而来增加对部分区域的观察精细度. 而自注意力机制是注意力机制的改进，其减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性. 并且基于自注意力机制的对抗网络允许图像生成任务中使用注意力驱动的、长距依赖模型，并且自注意力机制是对正常卷积操作的一个补充，全局信息也会被更好地利用去生成质量更好的图像.

为了更好地探索基于对抗网络生成的图片究竟可以精细到什么样的程度，基于 Self-Attention GAN 改进的 BigGAN 被提出. 该模型通过以下措施来提高模型生成图像的质量和多样性：

(1) 增大 Batch. 一个大的 Batch 可以让每个批次覆盖更多的内容，从而为生成器和鉴别器两个网络提供更好的梯度. 因此简单地增加 Batch，就可以实现性能上较好的提升，同时还可以在更短的时间内训练出更好的模型.

(2) 增大模型容量. 在合适的范围内通过增加每层网络的通道数来提高模型的容量.

(3) 共享嵌入. 将噪声向量 $\boldsymbol{Z}$ 等分成多块，然后将其和条件标签 C 连接后一起送入到生成网络的各个 BatchNorm 层.

(4) 分层潜在空间. 与传统模型直接将噪声向量 $\boldsymbol{Z}$ 嵌入生成网络初始层不同的是，BigGAN 将噪声向量 $\boldsymbol{Z}$ 输入模型的多个层，而不仅仅是初始层.

(5) 截断技巧. 在对先验分布 $\boldsymbol{Z}$ 采样的过程中，通过设置阈值的方式来截断 $\boldsymbol{Z}$ 的采样，其中超出范围的值会被重新采样以落入要求的范围内，该方法允许对样本多样性和保真度进行精细控制.

(6) 正交正则化. 该方法的目的是让生成网络的权重矩阵尽可能是一个正交矩阵，这样最大的好处就是权重系数彼此之间的干扰会非常得低.

在该模型设计过程中，增加 Batch 的大小会导致训练不稳定，因此在模型中采用谱正则化的技巧来改善训练模型时候的稳定性，从而抵消增加 Batch 对训练稳定性的影响. 最终这项工作表明通过上述技巧可以很好地改善生成网络的性能，但是该图像生成方法对计算力的要求很高.

(3) 基于单幅图像做图像生成

单幅图像中通常具有足够内部统计信息，可以使得网络学习到一个强大的生成模型. 基于这样的思想，Sin-GAN[72] 提出从单幅自然图像中去学习一个非条件生成模型，该模型可以以任意尺寸生成各种高质量的图像，同时也能够处理包含复杂结构和纹理的普通自然图像. 如图 8 所示，与常规 GAN 不同的是，该模型使用的训练样本是单幅图像不同尺度下采样的图像，而不是数据集中的整个图像样本.

该模型选择处理更一般的自然图像，使得模型具有生成纹理以外的其他功能. 为了更好地捕捉图像中目标的几何形状、位置信息、以及细节信息和纹理信息等图像属性，Sin-GAN[72] 采用了层级结构的对抗网络，由 N 对生成器 $\{G_N,\cdots,G_0\}$ 和鉴别器 $\{D_N,\cdots,D_0\}$ 组成，如图 8 所示. 其中，每个生成器负责生成不同尺度的图像，而相应的每个鉴别器负责捕捉图像不同尺度的分布. 从最粗到最细顺序的训练该模型的多尺度结构，当每个 GAN 被训练好以后，其参数就会被固定. 同时，模型中第 N 个 GAN 的损失函数由对抗损失 L_{adv} 和重建损失 L_{rec} 两个部分组成，如式(29)所示.

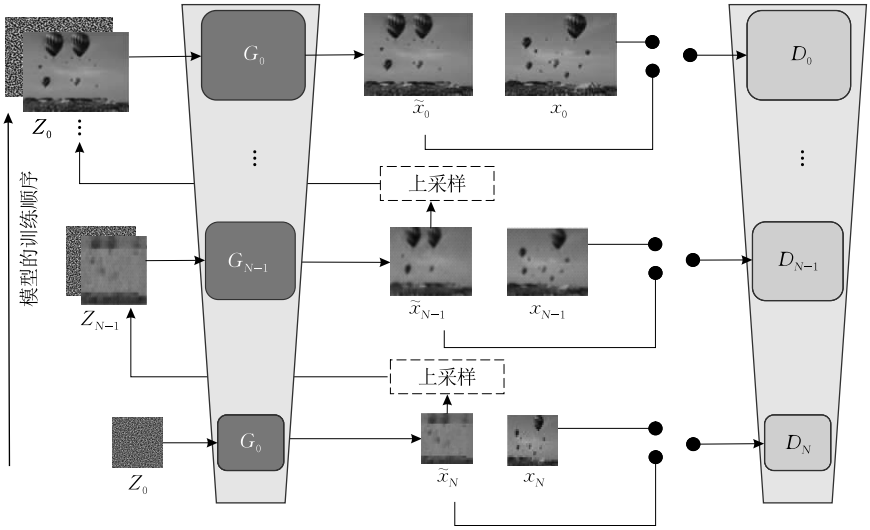


图 8 Sin-GAN 模型的多尺度网络架构[72]

$$\min_{G_n}\max_{D_n} L_{\text{adv}}(G_n,D_n)+\alpha L_{\text{rec}}(G_n) \tag{29}$$

式(29)中的 α 是超参数,为了增加训练的稳定性,对抗损失采用 WGAN-GP 损失,重建损失是为了确保使模型存在可以生成原始图像 x 的特定噪声图谱集合,如式(30)所示.

$$L_{\text{rec}} = \| G_n(0,(\tilde{x}_{n+1}^{\text{rec}})\uparrow)-x_n \|^2 \tag{30}$$

式(30)中,$(\tilde{x}_{n+1}^{\text{rec}})\uparrow$表示上个尺度生成图像上采样后的结果,$x_n$指的是尺度 N 下的真实图像.这项工作不仅仅具有生成纹理的能力,而且还具有为复杂自然图像生成各种逼真样本的能力.因此,其为多种图像处理任务提供了其强大的工具,但是,该方法在语义多样性方面存在固有的限制.

5 基于图像转换生成图像

在这一节,图像到图像的转换将被从有监督和无监督两个方面来进行介绍,而无监督方面,将从以下四个方面来进行讨论,分别是:基于自重构损失约束的图像转换、基于辅助分类器的图像转换、基于特征分离的图像转换、图像多域之间的转换.

5.1 基于有监督方式的图像转换

Pix2pix 可以将一种类型的图像转换到另一种类型,例如:黑夜转成白天、猫素描图转成猫的真实图片等,如图 9 所示.用于构建该模型生成器的架构是编码器-解码器网络结构的一种——U-Net[73] 网络,并且在网络中允许编码器到对称解码器之间的跳跃连接,基于这样的操作可以共享一些低层的信息.由于数据采用成对的图像,Pix2pix 的目标函数被设计成两个部分:一个是对抗损失函数;第二个是 $\mathcal{L}_1$ 正则化项,即生成图像与Ground-truth的差的模;由于 $\mathcal{L}_2$ 正则化使得生成的图像比较模糊,因此采用的是 $\mathcal{L}_1$ 正则化项.该模型的目标函数如式(31)所示:

图 9 图像到图像转换示意图[13]

$$G_{\text{target}}=\arg\min_G\max_D[E_{x,y}[\log D(x,y)]+E_{x,z}[1-\log D(x,G(x,z))]]+\lambda I_{L1}(G)] \tag{31}$$

I_{L1}代表的是$\mathcal{L}_1$正则化项,其实现方式如式(32):

$$I_{L1}(G) = E_{x,y,z}\ [\| y-G(x,z)\|] \tag{32}$$

在式(32)中 y 代表真实图像;x 代表类标签;z 表示彩色图像;G 表示生成器;D 表示鉴别器.该模

型在训练过程中，鉴别器的输入不是整幅图像，而是将整幅图像分割后的很多小块；用小块图像训练该模型，可以让模型很好地去捕捉局部细节或者是高频信息，同时正则化可以让模型更好地学到低频信息. 虽然这种方法可以生成高质量的图像，但是唯一的缺点就是必须使用成对的图像.

Pix2pixHD[74]在 Pix2pix 模型的基础之上，基于实例分割图像，使用多尺度的生成器以及鉴别器来生成高分辨率的图像. 该模型的生成器由 G_1 和 G_2 两个部分组成，其中 G_1 是一个端到端的 U-Net 网络结构. G_2 被分割成两个部分，G_2 的一部分用于提取特征，并且将该特征和 G_1 输出层的前一层特征进行相加融合. 最后将融合后的特征作为 G_2 另一部分的输入来生成高分辨率的图像. 而该模型的鉴别器是一个多尺度鉴别器，判别的三个尺度分别是：原图、原图的 1/2 降采样、原图的 1/4 降采样；并且对最后的判别结果取平均作为最终的结果. 多尺度判别的主要目的是让模型更好地保持内容一致性，而细节性的东西则由网络自己去学. 该模型的损失函数除了对抗损失函数以外，还加入了由特征提取器对生成样本和真实样本提取特征后构建的特征匹配损失. 同时该模型生成图像多样性的方法不同于 Pix2pix 模型，其在模型的输入端加入类标签信息，通过学习隐变量的方式来达到控制图像颜色、纹理风格信息的目的，从而来增加生成样本的多样性.

PLDT[53]通过在对抗网络的基础上增加一个用来判断来自不相同域的图像对是否相关的鉴别器来实现有监督的图像到图像转换，该鉴别器能够约束模型去保持生成图像和 Ground-truth 图像之间的一致性；该模型的生成器采用基于卷积的编码-解码结构网络来实现，而鉴别器都是采用全卷积网络来实现. 利用该模型来做图像到图像的转换，在保持不相同域中相关图像纹理一致性的同时，也可以修改图像中物体的形状.

基于有监督的方式做图像生成，就是要将模型的输入和输出联系起来，而基于这种联系就可以构造一个相应的约束项，在进行网络参数更新的时候，就可以约束网络拟合的分布向着真实数据的分布逼近；同时，以有监督方式生成的图像的质量一般都比较好.

5.2　基于无监督方式的图像转换

在无监督学习方法中有很多技巧被应用到了图像到图像转换的视觉任务中，例如：自重构损失、辅助分类器、距离约束、以及多域之间的图像转换等等；这一小节将对这些方法进行深入地讨论.

(1) 基于自重构损失约束的图像转换

自重构损失也被称为是自我一致性约束，目的是经过一个循环变换以后让输出和输入保持一致性. 如图 10 所示，在 Cycle-GAN[14]中有 2 个生成器，分别是 G_{xy}、G_{yx} 和 2 个鉴别器，分别是 D_x、D_y；其中 G_{xy} 生成器的目的是将源域 X 中的图像转换到目标域 Y，而 G_{yx} 执行相反的变换；而鉴别器 D_x 和 D_y 预测输入的图像是否属于相应的域.

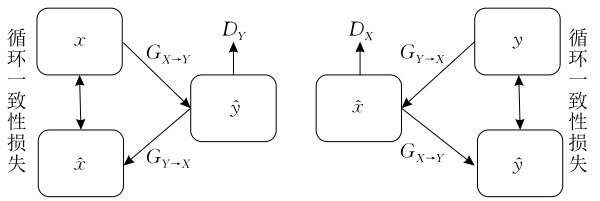


图 10　循环生成对抗网络[13]

对于 G_{xy} 和 D_y 组成的对抗网络的损失函数如式(33)所示：

$$\mathcal{L}_{\mathrm{GAN}}(G_{XY},D_Y)=E_{y\sim p_{\mathrm{data}(y)}}[\log D_Y(y)]+E_{x\sim p_{\mathrm{data}}(x)}[\log[1-D_Y(G_{XY}(x))]] \tag{33}$$

而由 G_{yx} 和 D_x 组成的对抗网络的损失函数可以表示成式(34)：

$$\mathcal{L}_{\mathrm{GAN}}(G_{YX},D_X)=E_{x\sim p_{\mathrm{data}(x)}}[\log D_X(x)]+E_{y\sim p_{\mathrm{data}}(y)}[\log(1-D_X(G_{YX}(y)))] \tag{34}$$

同时，在该模型中自重构损失是通过最小化重构误差来实现的，具体指将一幅图像 X 转换到另一个域 Y，再将得到的结果从 Y 域反向转换到 X 域后得到 $\hat{X}$，而用 X 和 $\hat{X}$ 做差的 1 模构建重构误差，上述过程如下：$X\to G_{XY}(x)\to G_{YX}(G_{XY}(x))\approx\hat{X}$；该重构误差的实现方式如式(35)所示：

$$\mathcal{L}_{\mathrm{cyc}}(G_{xy},G_{yx})=E_{x\sim p_{\mathrm{data}(x)}}[\|x-G_{YX}(G_{XY}(x))\|_1]+E_{y\sim p_{\mathrm{data}(y)}}[\|y-G_{XY}(G_{YX}(y))\|_1] \tag{35}$$

最终将上述三种损失函数组合起来可以得到 Cycle-GAN 的损失函数，如式(36)所示：

$$\mathcal{L}(G_{xy},G_{yx},D_x,D_y)=\mathcal{L}_{\mathrm{GAN}}(G_{XY},D_Y)+\mathcal{L}_{\mathrm{GAN}}(G_{YX},D_X)+\lambda\mathcal{L}_{\mathrm{cyc}}(G_{xy},G_{yx}) \tag{36}$$

在式(36)中，λ 是一个超参数，其大小可以决定 $\mathcal{L}_{\mathrm{cyc}}$ 在训练过程中起到的作用的大小. 模型最终的目标是优化式(37)：

$$G_{xy}^*,G_{yx}^*=\arg\min_{G_{xy},G_{yx}}\max_{D_Y,D_X}\mathcal{L}(G_{xy},G_{yx},D_x,D_y) \tag{37}$$

Dual-GAN[63]也采用了自重构损失约束项，其模型架构采用了和 Cycle-GAN 一样的结构，但是其目标函数用的是式(38)所示的最小二乘损失函数，在训练过程中最小二乘损失函数在稳定训练和解决梯

度消失问题上有一定的优势.

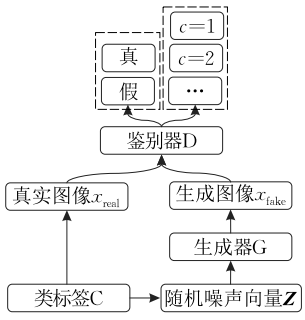
$$\mathcal{L}_{\text{TARGET}}(G_{X,Y},D_Y)=E_{y\sim p_{\text{data}(y)}}[(D_Y(y)-1)^2]+E_{x\sim p_{\text{data}(x)}}[D_Y(G_{XY}(x))^2] \tag{38}$$

Dual-GAN 在训练更新鉴别器参数的时候不仅仅用当前生成器生成的图像,而且还会用到以前生成器生成的图像,有点类似于 Unrolled GAN[45] 的训练策略,这种技巧在一定的程度上可以缓解模式崩溃的问题.

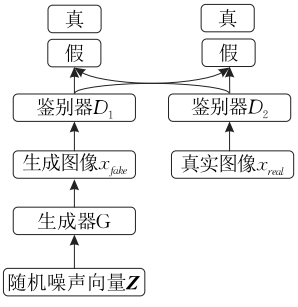
虽然基于自重构损失的无监督的图像转换模型可以生成高质量的图像,但是相较于 Pix2pix 等有监督的图像转换模型合成的图像还是有点模糊. 尽管通过实验证明了基于不成对的图像和自重构损失可以生成质量比较好的图像,但是在实际中该方法并不是适用于所有的情况,当图像转换中存在几何转换的时候,该方法的效果会变的较差. 在数据集中的图像是不成对的或者是不同风格的情况下,自重构损失约束的技巧是一个比较好的选择.

(2) 基于辅助分类器的图像转换

该方法是通过在 GAN 模型结构的隐空间中增加更多的网络结构,并且在目标函数中增加相应的约束项来提高生成图像的质量和增加生成图像的多样性. 在 ACGAN[17] 中,如图 11(a),鉴别器不仅仅判别输入图像是来自于生成数据的分布还是来自于真实数据分布,而且还会对输入图像的类别做一个预测. 鉴别器会给出域概率分布和类标签概率分布,$[P(S|X),P(C|X)]=D(X)$;而该模型的目标函数



(a) ACGAN[17] 的网络结构



(b) D2GAN[16] 的网络结构

图 11 辅助分类对抗网络

有两个部分:一部分是域损失函数 L_S,另一部分是类损失函数 L_C,如式(39)和(40)所示:

$$L_S=E[\log P(S=\text{real}\,|X_{\text{real}})]+E[\log P(S=\text{fake}\,|X_{\text{fake}})] \tag{39}$$

$$L_C=E[\log P(C=c\,|X_{\text{real}})]+E[\log P(C=c\,|X_{\text{fake}})] \tag{40}$$

目标函数中的 X_{real} 和 X_{fake} 分别表示真实数据样本和生成数据样本,而 C 表示样本的类别标签;在该模型中,D 的目标是最大化 L_S+L_C,而 G 的目标是最大化 L_C-L_S;基于这样的策略,该模型不仅仅可以提高生成图像的质量,而且还能够稳定 GAN 的训练.

D2GAN[16],如图 11(b),与标准 GAN 不同的是,D2GAN[16] 基于集成法来构建模型的架构,该模型有两个鉴别器,这两个鉴别器仍然是与一个生成器进行极小极大的博弈,其中一个鉴别器会给符合分布数据样本高的概率值,而另一个则相反,该模型的生成器要同时欺骗两个鉴别器. 理论分析表明,优化 D2GAN 的生成器可以让原始数据分布和生成数据分布之间的 KL 散度和反 KL 散度同时最小化,从而有效地避免模式崩溃的问题. 该模型的目标函数如式(41)所示:

$$\min_G \max_{D_1,D_2} \mathcal{J}(G,D_1,D_2)=\alpha\times E_{x\sim p_{\text{data}}}[\log D_1(x)]+E_{z\sim p(z)}[-\log D_1(G(z))]+E_{x\sim p_{\text{data}}}[-D_2(x)]+\beta\times E_{z\sim p(z)}[\log D_2(G(z))] \tag{41}$$

该目标函数中的 $D(x)$ 和 $D(G(z))$ 表示鉴别器其输入分别为真实样本和生成样本下的输出,其中超参数 α,β 的取值范围是 $(0,1]$,超参数的目的有两个,第一个是稳定 GAN 的训练,由于两个鉴别器的输出都是正的,$D_1(G(z))$ 和 $D_2(x)$ 可能会变得比 $\log D_1(x)$ 和 $\log D_2(G(z))$ 大,最终可能影响学习的稳定性. 为了克服这个问题,可以降低 α,β 的值. 第二个目的就是控制 KL 散度和反 KL 散度对优化的影响. 与标准 GAN 一样,通过交替对抗训练来对 D2GAN 进行训练.

综上可知,基于辅助分类器的方法来做图像生成是通过增加网络结构的功能或者是集成更多的鉴别器等方法来实现的. 虽然基于此方法可以生成高质量的图像,但是带标签的数据在实际中是很难获取到的,同时此方法可以有效地避免模式崩溃的问题.

(3) 基于特征分离的图像转换

在无监督学习中,因为缺乏对齐的、成对的训练

图像，所以研究者们就提出通过特征分离的技巧来实现两个域之间的图像转换，特征分离的基本思想是在隐空间中将图像的内容和属性分别学习，然后将图像的属性和内容任意组合来生成带有期望属性的图像.

DRIT[75]模型提出基于分离表示(Disentangled representation)的方法来做图像转换，如图 12 所示. 该方法将图像嵌入到两个空间：一个是域不变的内容空间，另一个是特定域的属性空间. 该模型通过编码器从给定的输入图像中提取内容特征和属性特征. 如图12所示该模型的输入来自两个不相同域的不同图像X,Y；该模型首先通过编码器来分别提取两幅图像的内容特征$z_x^a=E_x^a$，$z_y^a=E_y^a$和属性特征$z_x^c=E_x^c$，$z_y^c=E_y^c$；然后交换两幅图像的属性编码，将(E_x^a,E_y^c)和(E_y^a,E_x^c)作为生成器G_x,G_y的输入，假定生成器的输出分别为u,v，基于这些生成的图像，再次用编码器提取内容特征$z_u^a=E_u^a$，$z_v^a=E_v^a$和属性特征$z_u^c=E_u^c$，$z_v^c=E_v^c$，然后像前边一样交换两幅图像的属性特征，将得到的总的特征向量作为生成器G_x,G_y的输入，生成器的输出是X',Y'；最终的期望是输出的图像和该模型输入的图像是一致的，即$X=X',Y=Y'$；由于在模型中会将一个域中图像的内容表示和另一个域中图像的属性表示组合起来，该模型基于此提出了一个交叉循环一致性约束，为了执行此约束，将其实现为式(42)：

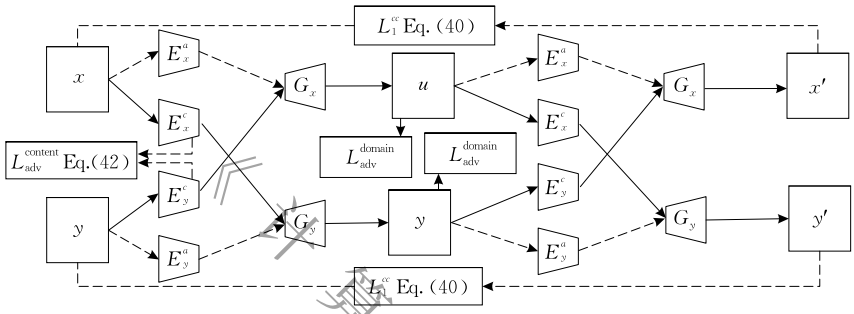


图 12　特征分离示意图[75]

$$\mathcal{L}_1^{cc}(G_x,G_y,E_x^c,E_y^c,E_x^a,E_y^a)=\mathbb{E}_{x,y}[\|G_x(E_y^c(v),E_x^a(u))-x\|_1+G_y(E_x^c(u),E_y^a(v))-y\|_1] \tag{42}$$

其中$u=G_x(E_y^c(v),E_x^a(u))$，$v=G_y(E_x^c(u),E_y^a(v))$生成器不仅仅可以实现不同域图像的属性交换，而且还能基于一幅图像的内容特征和属性特征重构出原始的图像，这一约束可以通过式(43)来实现：

$$\mathcal{L}_1^{rec}=\mathbb{E}_{x,y}[\|G_x(E_x^c(x),E_x^a(x))-X\|_1+\|G_y(E_y^c(x),E_y^a(y))-Y\|_1] \tag{43}$$

由于不同域中图片的内容信息不包含特征信息，所以应该是不可区分的；在这样的前提下，两个内容编码器的最后一层网络的参数应该共享，保证内容分布一致，同时两生成器第一层网络参数也要共享，并且还得让内容鉴别器D^c辨别不出两个内容特征是属于哪一类. 这一目标通过在目标函数中加入如式(44)所示的损失项实现.

$$\mathcal{L}_{adv}^{content}(E_x^c,E_y^c,D^c)=\mathbb{E}_x\left[\frac{1}{2}\log D^c(E_x^c(x))+\frac{1}{2}\log(1-D^c(E_x^c(x)))\right]+\mathbb{E}_y\left[\frac{1}{2}\log D^c(E_y^c(y))+\frac{1}{2}\log(1-D^c(E_y^c(y)))\right] \tag{44}$$

为了在测试的时候进行随机采样，可以通过在目标函数中加入一个 KL 散度约束项来让属性表示向量逼近一个先验高斯分布，该约束项的实现方式如式(45)所示：

$$\mathcal{L}_{KL}=\mathbb{E}[D_{KL}((z_a)\|N(0,1))] \tag{45}$$

为了实现图像和隐变量空间的可逆映射，该模型通过在目标函数中加入隐变量回归损失$\mathcal{L}_1^{latent}$约束项来实现. 如果从某一个高斯先验分布中随机采样一个隐向量z作为属性特征向量，必须能够用式(46)实现重构它.

$$z'=E_x^a(G_x(E_x^c(x),z)),z'=E_y^a(G_y(E_y^c(y),z)) \tag{46}$$

与标准 GAN 一样，该模型还有一个用来判断生成图像是来自于哪个域的域对抗损失项$\mathcal{L}_{adv}^{domain}$；所以该模型的损失函数由内容对抗损失、交叉循环损失、$\mathcal{L}_1$正则化项、域对抗损失、一个对噪声的约束项和隐变量回归损失组成，其形式如式(47)所示：

$$\mathcal{L}=\min_{G,E^c,E^a}\max_{D,D^c}[\lambda_{adv}^{content}\mathcal{L}_{adv}^{content}+\lambda_1^{rec}\mathcal{L}_1^{rec}+\lambda_1^{cc}\mathcal{L}_1^{cc}+\lambda_{adv}^{domain}\mathcal{L}_{adv}^{domain}+\lambda_1^{latent}\mathcal{L}_1^{latent}+\lambda_{KL}L_{KL}] \tag{47}$$

其中λ是超参数，用来控制每一项的重要性. 虽然可以通过这种方法来实现图像域的转换，但是当图像之间的域有很大的差别的时候，该方法实现的效果不是很好，同时由于训练数据量的限制，导致属性空间不能被完全覆盖. 同时该模型的目标函数比较复杂，不好训练；当图像域之间的差别不是很大的时

候，基于此方法来做图像转换是一个不错的选择.

(4)多域图像之间的转换

之前讨论的模型，大多都是在两个域之间进行转换图像. 如果想要在多个域之间相互转换，就必须在每两个域之间单独训练一个对抗网络，然而这样做的效率很低，而且每次训练特别的耗时. 为了解决这一问题，Star-GAN[46]使用一个生成器来实现域之间的相互转换. 与标准 GAN 不同的是，Star-GAN 用图像和目标域的类标签作为输入，将输入图像转换到由类标签指明的域；同时允许该模型在具有不相同域的多个数据集上进行训练，为了预测生成图像所在的域，类似于 ACGAN[17]、DAAC[76]，该模型的鉴别器增加了辅助分类器，用来预测输入图像的域，也就是说，鉴别器不仅仅要判断输入图像是生成图像还是真实图像，而且还得输出域标签的概率分布，$D: x \rightarrow \{D_{\mathrm{src}}(x), D_{\mathrm{cls}}(x)\}$. 为了使得生成的图像不同于真实的图像，该模型采用如式(48)所示的对抗损失：

$$\mathcal{L}_{\mathrm{adv}} = \mathbb{E}_x[\log D_{\mathrm{src}}(x)] + \mathbb{E}_{x,c}[\log(1 - D_{\mathrm{src}}(G(x,c)))] \tag{48}$$

其中 $G(x,c)$表示生成器基于输入图像 x 和目标域标签 c 生成的图像. 同时为了将输入图像转换后的图像分类到目标域 c 中，该模型在目标函数中增加了一个域分类损失项；并且将该损失分为了真实图像的域分类损失 $\mathcal{L}_{\mathrm{cls}}^r$ 和生成图像的域分类损失 $\mathcal{L}_{\mathrm{cls}}^f$；这两个损失函数分别被定义为

$$\mathcal{L}_{\mathrm{cls}}^r = \mathbb{E}_{x,c'}[-\log D_{\mathrm{cls}}(c' \mid x)] \tag{49}$$

$$\mathcal{L}_{\mathrm{cls}}^f = \mathbb{E}_{x,c}[-\log D_{\mathrm{cls}}(c \mid G(x,c))] \tag{50}$$

其中 $D_{\mathrm{cls}}(c' \mid x)$表示鉴别器输出的域标签的概率分布. 通过最小化 $\mathcal{L}_{\mathrm{cls}}^r$，鉴别器学着去将真实图像分类到相应的原始域 c'；生成器去最小化 $\mathcal{L}_{\mathrm{cls}}^f$，使得生成的图像被分类到目标域 c. 由于训练集不是成对的图像，因此模型使用循环一致性损失 $\mathcal{L}_{\mathrm{rec}}$ 来保证生成图像和输入图像内容上的一致性.

$$\mathcal{L}_{\mathrm{rec}} = \mathbb{E}_{x,c,c'}[\| x - G(G(x,c),c') \|_1] \tag{51}$$

最终优化生成器和鉴别器的目标函数可以被写成如式(52)和(53)：

$$\mathcal{L}_D = -\mathcal{L}_{\mathrm{adv}} + \lambda_{\mathrm{cls}} \mathcal{L}_{\mathrm{cls}}^r \tag{52}$$

$$\mathcal{L}_G = \mathcal{L}_{\mathrm{adv}} + \lambda_{\mathrm{cls}} \mathcal{L}_{\mathrm{cls}}^f + \lambda_{\mathrm{rec}} \mathcal{L}_{\mathrm{rec}} \tag{53}$$

其中，λ_{cls} 和 λ_{rec} 是控制域分类损失和重构损失重要性的超参数；该模型在人脸属性转换中取得了很好的效果.

采用多域图像转换的模型，可以解决用一对一图像转换模型训练低效和训练效果有限的问题；用来做多域图像转换的模型，可以利用其他领域的数据来增强模型的泛化能力，从而可以生成质量更高的图像，同时其扩展性也比较好.

6 生成图像质量和多样性的评估方法

在目前的研究中，对生成图像的质量进行合理地评估主要是从定性评估和定量评估两个方面进行. 定性的评估一般在众包平台靠人完成；而定量的评估方法有 Inception score、Fréchet Inception Distance、Mode Score 等方法. 定性的评估方式将在这一节首先被讨论.

6.1 定性评估

该方法主要还是靠人的眼睛来进行判断. 一般的做法是将真实图像和生成图像对上传到众包平台上让人来判断图像的真假，并且给出两者的相似程度，最后根据打分的结果统计一个最终的指标.

在实践中由于人的主观性是很强的，每个人的标准是不一致的，导致定性评估不是一个通用的标准；视觉检查在评估一个模型对数据的拟合程度时，在低维度数据的情况下可以工作得很好，但是在高维度数据的情况下，这种直觉性可能会导致误导. 如图 13 所示，从标准 GAN 到最近 BIGGAN 模型架构的变化，以及生成图像的结果来看，图像生成模型

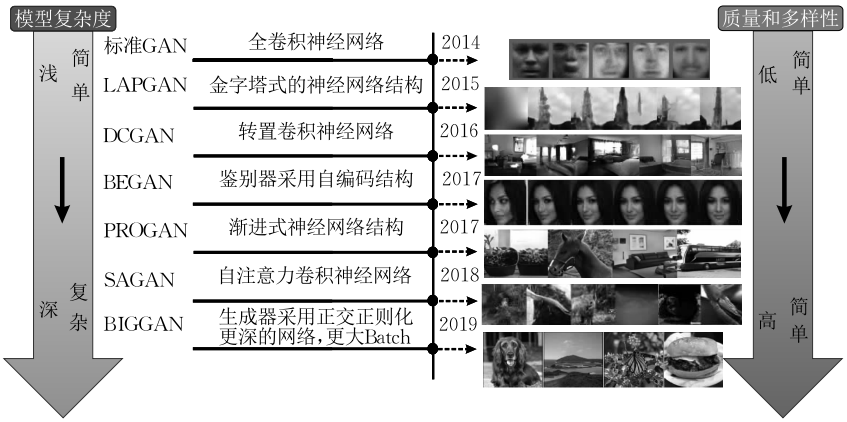


图 13 生成图像质量和多样性，以及模型复杂度的变化流图[77]

的复杂度和计算量在不断地增加；同时，生成图像的逼真度越来越好，多样性也越来越好.

6.2　定量评估

(1) Inception score

Inception score[78] (IS)是一种用类概率分布来对生成图像进行评估的方法. 该方法使用一个预先在 ImageNet 数据集上训练的 Inception V3 网络，然后以生成的图像 x 作为输入，并且输出 $p(y|x)$；如果一幅生成图像的质量越好，那条件概率分布 $p(y|x)$ 的熵就越低，也就意味着分类器以很高的置信度将图像分到某一类. 生成器生成的图像应该具有多样性，因此边缘分布 $p(y)=\int p(y|x=G(z))\mathrm{d}z$ 应该有很高的熵；基于这样的条件，IS 可以通过式(54)来计算：

$$\mathrm{IS}=\exp(E_{x\sim G(z)}D_{\mathrm{KL}}(p(y\mid x)\parallel p(y)) \tag{54}$$

式(54)中的 E 表示计算期望值，D_{KL} 表示计算两分布之间的 KL 散度；Lucic 等人[79]指出 Inception score 对类标签的先验分布和距离的度量方式都是不敏感的；由于生成模型只需在每一个类中生成一个质量很好的图像就可以得到高的 IS 值，因此该方法也面临无法判别模式崩溃的问题. 并且 IS 可以展现出生成图像的质量和多样性之间的合理关联，所以该评估指标在实践应用中被广泛采用.

如表4所示，基于 ImageNET 数据集和 CIFAR-10 数据集，用定量评估指标 Inception Score 对多个模型生成的图像的质量和多样性进行了定量计算. 可以发现 BigGAN 模型是当前性能最好的模型，但是通过比较也发现，该评估指标并不适合评估与 ImageNET 数据集差别较大的图像数据.

表 4　定量指标 Inception score 下图像生成模型的实验结果

数据集	模型	Inception Score(IS)
ImageNET 128×128	AC-GAN	28.50
	Projection Discriminator	36.80
	SAGAN	52.52
	S3 GAN	83.10
	BigGAN	166.30
	BigGAN-DEEP	166.50
CIFAR-10	AC-GAN	8.25
	BEGAN	5.62
	Auto-GAN	8.55
	BigGAN	9.22
	DCGAN	6.58
	PGGAN	8.80
	SGAN	8.59
	Improved GAN	6.86

总的来讲，Inception Score 是用来衡量生成模型个体特征和整体特征的方法. 个体特征指生成的图像要清晰，质量要好. 整体特征指生成的图像要有多样性，即使他们属于同一类别，他们的输出的向量还是应该有差别.

(2) Fréchet Inception Distance

Fréchet Inception Distance[80] (FID)的基本思想是用 Inception 网络的卷积特征层作为一个特征函数 φ，并且用特征函数将真实数据分布 P_r 和生成数据分布 P_g 建模为两个多元高斯随机变量. 这样就可以计算多元高斯分布的均值 μ_x，μ_g 和方差 $\sum_x$，$\sum_g$. 基于这些信息，生成图像的质量可以通过式(55)由两个高斯分布之间的 Fréchet 距离来计算.

$$\mathrm{FID}(X,G)=\parallel\mu_x-\mu_g\parallel_2^2+\mathrm{Tr}\left(\sum_x+\sum_g-2\left(\sum_x\sum_g\right)^{\frac{1}{2}}\right) \tag{55}$$

FID 度量方式的思想和人类判断是一致的，该评价指标值越小，表示生成的图像越接近真实图像，生成的图片质量越好. FID 和生成图像的质量之间有很强的负相关性；该度量方式的优势在于其对噪声不是很敏感，而且可以检测出类内的模式崩溃的问题. 如表 5 所示，基于 ImageNET 数据集和 CIFAR-10 数据集，对一些图像生成模型在 FID 下的性能进行了汇总. 通过分析和比较，可以发现 BigGAN-DEEP 模型在生成图像的质量和多样性方面都是最好的.

表 5　定量指标 FID 下图像生成模型的实验结果

数据集	模型	FID
CIFAR-10	WGAN-GP	29.30
	WGAN-GP+TTUR	24.80
	RSGAN-GP	25.60
	SN-GANS	21.70
	DIST-GAN	17.61
	Auto-GAN	12.42
ImageNET 128×128	Projection Discriminator	27.62
	SAGAN	18.65
	S3 GAN	7.70
	BigGAN	9.60
	BigGAN-DEEP	7.40

(3) Mode Score

Mode Score[81] 可以看成是 Inception score 的一个改进版本，其被定义为如下形式：

$$MS(\mathbb{P}_g)=e^{E_{x\sim p_g}[KL(p_M(y|x)\parallel p_M(y))-KL(p_M(y)\parallel p_M(y^*))]} \tag{56}$$

其中 $p_M(y^*)=\int_x p_M(y|x)\mathrm{d}\mathbb{P}_r$ 是真实数据分布中样本的边缘标签分布；与 IS 不同的是，Mode Score 可以通过 $KL(p_M(y)\parallel p_M(y^*))$ 测量真实数据分布 P_r 和生成数据分布 P_g 之间的非相似性. 由于该模型是基于 IS 的一种评估指标，因此其沿袭了 IS 的固有缺陷，一些简单的扰动就有可能导致彻底地欺骗该

评估指标,从到导致该评估该方法也无法判别模式崩溃的问题.

(4)1-最近邻双样本检验

在双样本检验中,1-最近邻分类器被使用去评估两个分布是否完全相同;给定两个样本集,分别是真实数据样本集$S_r\sim P_r^n$和生成数据样本集$S_g\sim P_g^m$,并且将样本集S_r全部标注为正样本,而样本集S_g全部被标注为负样本.基于正负样本集,可以训练一个1-最近邻分类器,并且可以计算1-最近邻分类器的留1(LOO)准确率.此准确率是一个统计量,当样本的数量足够大的时候,并且两个数据集的分布是一致的时候,该留1准确率的值应该是0.5;当生成器学到的数据分布P_g过拟合真实数据分布P_r时,留1准确率的值应该小于0.5;反之则该值大于0.5.

该评估方法在理论上存在一个极端的情况,如果生成器仅仅是简单地记住真实数据集S_r中的每一个样本,并且可以精确地重新生成每一个样本的时候,导致S_r中的每一个样本在S_g中都有一个距离为0的最近邻,所以LOO准确率将变为0;原则上分类器可以采用任意的二分类器,但是该方法只考虑1-NN分类器,因为该分类器不需要特殊的训练,并且只有很少需要调整的超参数.

模式崩溃问题出现时,真实图像和生成图像的主要最近邻都是生成的图像;由于真实数据分布的模式通常都可以被生成器捕捉,就会导致S_r中大多数真实样本的周围都是生成的样本,这就会导致较低的LOO准确率.而生成样本倾向于聚集到少量的模式中心,而这些模式一般都是相同类别的生成样本包围,因此会导致较高的LOO准确率.所以1-最近邻双样本检验的评估方式在保证可以很好地鉴别真与假的同时,还可以很好地鉴别模式崩溃的问题,并且该方法有很高的计算效率.

6.3 评估方法小结

(1)尽管上述评估方法在不同的任务中展现了有效性,但是在什么样的场景用什么样的评估方法或者是在什么样的场景下用那个评估方法容易导致误解目前是不清晰的;一种评估方式是否合适,只有在实际应用的上下文中才能知晓.

(2)不同的评估方法适合于不同的模型,所以根据自己的任务选择与任务相匹配的评估方式相当重要.

(3)目前的评估方式都是基于样本来度量的,大多数现有的方法都试图展现其与人类评估的相关性来证明自身的正确性.但是人往往只注重图像的质量,而会忽视对于无监督学习很重要的整体分布特征,用人来做评估评估还容易受主观因素的影响,因此人的评估是有偏的.所以不要以人的标准来看图像.在本小节最后,将通过表6对上述四种评估方式做一个汇总和比较.

表6 评估标准的比较

评估方法	优点	缺点
Inception score Mode Score	可以很好地展现质量和多样性之间的关联	无法检测到过拟合和模式崩溃的问题;对扰动比较敏感;不能用于和ImageNET差别比较大的数据集
Fréchet Inception Distance(FID)	判别力、鲁棒性、效率方面表现良好	无法捕捉细微的变化;无法断定高的FIS值是由什么原因导致的;
1-最近邻双样本检验	判别力好;鲁棒性强;对模式崩溃敏感;计算效率高	—

7 图像生成的应用

基于生成对抗网络强大的隐式建模能力,目前可以生成十分清晰的图像,而且在实践过程中不需要知道真实样本数据的显式分布,同时也不需要假设更多的数学条件.这些优势使得基于对抗网络的图像生成可以被应用到很多学术和工程领域.

(1)小样本问题

在目前的工作中,深度学习的良好表现很大程度上依赖于大的数据量和计算力的提高.但是很多实际的项目难以有充足的数据来完成任务,而要保证很好地完成任务,就得必须寻找很多的数据或者是用无监督学习的其他方法来完成.目前比较常用的方法是在已有的数据上,用几何变换类方法和颜色变换类方法来获取更多的数据,但是这种方法没有实质性的增加数据.而基于GAN的图像生成方法是解决这一问题的一个很好的思路.通过图像生成的方式可以生成与真实数据分布一致的很多样本,对小样本集进行一个扩充,然后将混合样本集作为其他视觉任务的数据集,从而达到增强模型学习效果的目的.

(2)数据类别不平衡

数据类别不平衡指的是数据集中各个类别的样本数量有很大的差别.目前针对这一问题比较常用的方法是随机采样,该方法从数据角度出发来解决这一问题.随机采样又分为上采样和下采样,上采样方法指从少数类的样本中进行随机采样来增加新的

样本，而下采样方法是从多数类样本中随机选择少量样本，再合并原有少数类样本作为新的训练数据集.但是如果采用上采样的方法，上采样后的数据集中会反复出现一些样本，训练出来的模型会有一定的过拟合；而下采样的缺点是最终的训练集丢失了数据，模型只学到了总体模式的一部分.而基于图像生成的方法来解决数据类别不平衡，则可以避免上述缺点，更好地扩充数量少的数据.

(3) 超分辨率

基于图像生成能够以低分辨率图像为输入，然后输出带有清晰细节信息的超分辨率图像.SRGAN采用基于残差块构建的生成器和基于全卷积网络构建的鉴别器来做单幅图像的超分辨率.该模型的损失函数除了对抗损失外，还组合了像素级别的MSE损失、感知损失和正则化损失，并且该模型可以生成质量很好的图像.

(4) 目标检测和跟踪中的应用

受益于图像生成在超分辨率领域中的应用，在目标检测任务中，对一幅图像中的小目标进行检测经常会遇到目标对象是低分辨率的情况.因此，Li等人[82]试着将低分辨率的小物体转换成高分辨率的大物体，从而提高物体的可判别性；在该模型中鉴别器被分成了两个部分：对抗部分和感知部分.对抗部分的作用是与生成网络进行对抗训练，使得生成网络可以是生成高分辨率大尺度的目标；而感知部分的作用是确保生成的大尺度目标对检测任务是有用的.Wang等人[83]提出通过对抗网络生成带有遮挡和变形的图片样本来训练检测网络，从而提高检测网络的性能.

由于基于生成对抗网络做图像生成可以保持图像的细节纹理特征.因此，Orest等人[84]提出DeblurGAN来实现对运动图像的去模糊化.图像模糊是视觉任务中经常遇到的一个问题，比如：图像数据采集过程中由于物体运动导致的模糊，目标跟踪中相机的运动导致的模糊等.而基于DeblurGAN的去模糊方法则为处理模糊问题提供了一个很好的途径.同时基于生成对抗网络强大的生成能力，VGAN被提出去生成视频，而生成的视频可以为目标跟踪任务提供更多的运动信息.

(5) 图像属性编辑

图像属性编辑指通过对抗网络学习一个映射，该映射不仅具有生成图像的功能，而且还具备根据属性信息向量修改图像属性的能力.IcGAN[85]提出通过学习两个独立的编码器 E_z 和 E_y，其中 E_z 的作用是将一幅图像映射成一个隐向量 $\boldsymbol{Z}$，而 E_y 的作用是学习一个属性信息向量 $\boldsymbol{y}$.属性编辑操作是通过调整属性信息向量 $\boldsymbol{y}$，并将其和隐向量 $\boldsymbol{Z}$ 链接后，一起送入生成器来实现的.而AttGAN[86]实现了在保留原图像细节信息的同时，编辑人脸图像的单个或多个属性，生成带有新属性的人脸图像.该模型基于编码器-解码器架构，通过解码以期望属性为条件的给定面部的潜在表示来实现面部属性的编辑.而这些图像属性编辑的方法在图像编辑软件，以及一些娱乐软件中将会有很好的应用前景.

(6) 医学图像领域中的应用

在医学领域，由于过度辐射，会对人体造成一定的伤害，而降低辐射剂量已经被作为一种有效的解决方案.但是，剂量的减少会增加医学图像的噪声水平，这就会导致一些信息的丢失.目前，基于卷积网络的去噪声方法的主要问题是在优化中使用了均方误差，导致预测的图像比较模糊，无法提供常规剂量下图像的那种高质量的纹理.因此，可以使用图像转换的方法建立噪声图像和去噪图像之间的映射来消除这个问题，并且生成高质量的图像.在获取一些医学图像的过程中，由于运动而使得一些器官的关键信息丢失，而基于图像生成的办法则可以在有信息丢失和完全采样的图像之间建立一个映射，帮助更好地采集图像.

8　总结和展望

生成对抗网络作为一种概率生成模型，其已经被应用于很多视觉任务中，特别是在图像生成方向的优良表现.本文首先从工作机理、目标函数、模型结构、和训练GAN存在的问题以及应对策略等角度对GAN进行了一个详细地讨论.其次，本文按照直接法和集成法的分类方式对基于GAN做图像生成的方法进行了一个汇总；然后根据输入向量形式的不同，对图像生成进行了详细地探讨.并且对图像生成的应用做了详细介绍.最后，本文对目前工作中对生成图像进行质量评估的方法做了详细地汇总和分析.

通过以上论述，总体来看，基于生成对抗网络来做图像生成的方法相较于2014年提出来的GAN，其做出的改进主要集中在以下几方面：生成器和鉴别器的神经网络架构、损失函数的设计、改善模型训练时候的稳定性，以及改善模式崩溃.

虽然这些改进后的模型在业界取得了一系列成果，但是模式崩塌问题仍然是做图像生成过程中一个严重的问题.在模型训练过程中，生成网络有选择

地学习了某些模式,同时又放弃了某些模式.针对这一问题,目前的方法只是通过修改目标函数、改变训练方式等来改善这一问题,而导致这一问题的原因目前还尚不清楚.因此,在理论上对这一问题的研究有待进一步的突破.

基于生成对抗网络来做图像生成,其良好性能很大程度上还是依赖于神经网络强大的拟合能力.所以生成图像质量和多样性的好坏,与神经网络中的架构有着直接的关系.但是,现在还没有成功的理论可以根据环境来优化神经网络的结构,或者评价修改神经网络结构对生成模型性能的影响.而只有针对实际问题进行彻底的实验研究,才能得到满意的效果.因此,针对设计的神经网络架构不一定是最优架构这一问题,基于神经网络架构搜索找最优神经网络架构可能是一个很好的解决方案.该方法通过定义一个合适的搜索空间,设计一个合适的搜索策略,在合适的性能指标下找到一个最佳模型.

除了上述问题,仍然有很多问题在制约图像生成的发展,最为突出的是模型的可解释性.要想将图像生成方法成功落地,可解释性是必不可少的一个环节,并且有关 GAN 收敛性的数学分析仍有待建立.因此目前图像生成的研究主要是建立在深度学习积累的经验之上.其次,对 GAN 模型生成图像质量和多样性的评估,目前还没有一个统一的、适用于所有模型的方式,因此在实际操作中,只能根据实际要解决的问题来选择一个合理的评估方式,并且目前存在的对生成图像进行评估的方式,都有一定的局限性.最后就是目前的图像生成方法,对计算力的要求都很高,如果将模型进行部署的话,对模型的大小定会提出新的要求.因此,如何建立起有关分析基于 GAN 做图像生成的机制,以及基于这些机制如何对模型进行优化和压缩,如何拓展图像生成的应用范围,这些问题都有待研究者们进一步地研究.

随着人工智能技术的发展,多模态融合是一个必然的发展趋势,通过改进神经的架构和算法,基于语音和文字生成语义一致的图像是一个很好的研究方向.由于基于有监督方式生成图像的质量比较好,但是实际中大量带标签的数据是很难去获得的,而少量带标签的数据很容易得到,因此,探索如何组合 GAN 和半监督学习去更好地做图像生成也是一个很有希望的研究方向.

在神经网络安全领域,图像生成将会有很大的用处.目前深度神经网络虽然精确度越来越高,但是也发现它们极其容易被攻击和影响.如果对样本做轻微的扰动,而神经网络就会以很高的置信度,做出错误的分类或者是预测,这一现象就是对抗攻击.同时,深度神经网络对于对抗攻击鲁棒性差是一个非常普遍的现象.因此,为了增强网络抗攻击的能力,可以通过图像生成来生成对抗样本,基于对抗样本来训练网络,不断地提高深度神经网络的鲁棒性,使其性能有更大的提升.所以将图像生成用于提高网络的鲁棒性是一个非常需要研究者们去研究的方向.

参 考 文 献

[1] Dayan P. Helmholtz machines and wake-sleep learning. Handbook of Brain Theory and Neural Network. Cambridge, USA: MIT Press, 2000

[2] Kingma, Diederik P, Max W. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013

[3] Hinton G E. Deep belief networks. Scholarpedia, 2009, 4(5): 5947

[4] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering//Proceedings of the 24th International Conference on Machine Learning. New York, USA, 2007: 791-798

[5] Salakhutdinov R, Hinton G. Deep Boltzmann machines//Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS). Clearwater, USA, 2009: 448-455

[6] Oord, van den A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759, 2016

[7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Montrél, Canada, 2014: 2672-2680

[8] Goodfellow, Ian. NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2017

[9] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 2018, 35(1): 53-65

[10] Kurach K, et al. The GAN landscape: Losses, architectures, regularization, and normalization. arXiv preprint arXiv: 1807.04720, 2018

[11] Lin Yi-Lun, Dai Xing-Yuan, Li Li, et al. The new frontier of AI research: Generative adversarial networks. Acta Automatica Sinica, 2018, 44(5): 775-792(in Chinese)
(林懿伦, 戴星原, 李力等. 人工智能研究的新前线: 生成式对抗网络. 自动化学报, 2018, 44(5): 775-792)

[12] Zamorski M, Zdobylak A, Zięba M, et al. Generative adversarial networks: Recent developments//Proceedings of the International Conference on Artificial Intelligence and Soft Computing. Cham, USA, 2019: 248-258

[13] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1125-1134

[14] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2223-2232

[15] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1857-1865

[16] Nguyen T, Le T, Vu H, et al. Dual discriminator generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 2670-2680

[17] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 2642-2651

[18] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 4681-4690

[19] Donahue C, et al. Semantically decomposing the latent spaces of generative adversarial networks. arXiv preprint arXiv:1705.07904, 2017

[20] Yin Wei-Dong, et al. Semi-latent GAN: Learning to generate and modify facial images from attributes. arXiv preprint arXiv: 1704.02166, 2017

[21] Tran L, Yin X, Liu X. Representation learning by rotating your faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(12): 3007-3021

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

[23] He Z, Zuo W, Kan M, et al. AttGAN: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing, 2019, 28(11): 5464-5478

[24] Ehsani K, Mottaghi R, Farhadi A. SeGAN: Segmenting and generating the invisible//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 6144-6153

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

[26] Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 613-621

[27] Tulyakov S, Liu M Y, Yang X, et al. MocoGAN: Decomposing motion and content for video generation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 1526-1535

[28] Li Y, Liu S, Yang J, et al. Generative face completion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 3911-3919

[29] Ma L, Jia X, Sun Q, et al. Pose guided person image generation //Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 406-416

[30] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2794-2802

[31] Arjovsky M, Soumith C, Léon B. Wasserstein GAN. arXiv preprint arXiv:1701.07875, 2017

[32] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 5767-5777

[33] Li Y, Swersky K, Zemel R. Generative moment matching networks//Proceedings of the International Conference on Machine Learning. Lille, France, 2015: 1718-1727

[34] Li C L, Chang W C, Cheng Y, et al. MMD GAN: Towards deeper understanding of moment matching network//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 2203-2213

[35] Nowozin S, Cseke B, Tomioka R. F-GAN: Training generative neural samplers using variational divergence minimization//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 271-279

[36] Sriperumbudur B K, et al. On integral probability metrics, φ-divergences and binary classification. arXiv preprint arXiv: 0901.2698, 2009

[37] Xu Qian-Tong, et al. An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint arXiv:1806.07755, 2018

[38] Radford A, Luke M, Soumith C. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015

[39] Im D J, et al. Generating images with recurrent adversarial net-works. arXiv preprint arXiv:1602.05110, 2016

[40] Makhzani A, et al. Adversarial autoencoders. arXiv preprint arXiv:1511.05644, 2015

[41] Larsen A B L, et al. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300, 2015

[42] Donahue J, Philipp K, Trevor D. Adversarial feature learning. arXiv preprint arXiv: 1605.09782, 2016

[43] Zhao Jun-Bo, Michael M, Yann Le-Cun. Energy-based generative adversarial network. arXiv preprint arXiv: 1609.03126, 2016

[44] Kodali N, et al. How to train your DRAGAN. arXiv preprint arXiv:1705-07215, 2017

[45] Lin Z, Khetan A, Fanti G, et al. PacGAN: The power of two samples in generative adversarial networks//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 1498-1507

[46] Choi Y, Choi M, Kim M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8789-8797

[47] Mirza M, Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014

[48] Xian W, Sangkloy P, Agrawal V, et al. TextureGAN: Controlling deep image synthesis with texture patches//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8456-8465

[49] Song L, Lu Z, He R, et al. Geometry guided adversarial facial expression synthesis//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea, 2018: 627-635

[50] Zhu J Y, Zhang R, Pathak D, et al. Toward multimodal image-to-image translation//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 465-476

[51] Dekel T, et al. Smart, sparse contours to represent and edit images. arXiv preprint arXiv:1712.08232, 2017

[52] Park T, Liu M Y, Wang T C, et al. Semantic image synthesis with spatially-adaptive normalization//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 2337-2346

[53] Yoo D, Kim N, Park S, et al. Pixel-level domain transfer//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 517-532

[54] Metz L, et al. Unrolled generative adversarial networks. arXiv preprint arXiv:1611.02163, 2016

[55] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 2234-2242

[56] Chen X, Duan Y, Houthooft R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 2172-2180

[57] Qi Guo-Jun. Loss-sensitive generative adversarial networks on lipschitz densities. arXiv preprint arXiv: 1701.06264, 2017

[58] Taigman Y, Adam P, Lior W. Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200, 2016

[59] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 700-708

[60] Zhang Han, et al. Self-attention generative adversarial networks. arXiv preprint arXiv:1805.08318, 2018

[61] Zhang H, Xu T, Li H, et al. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 5907-5915

[62] Denton E L, Chintala S, Fergus R. Deep generative image models using a laplacian pyramid of adversarial networks//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2015: 1486-1494

[63] Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised dual learning for image-to-image translation//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2849-2857

[64] Gan Z, Chen L, Wang W, et al. Triangle generative adversarial networks//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 5247-5256

[65] Anoosheh A, Agustsson E, Timofte R, et al. ComboGAN: Unrestrained scalability for image domain translation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA, 2018: 783-790

[66] Royer A, et al. XGAN: Unsupervised image-to-image translation for many-to-many mappings. arXiv preprint arXiv: 1711.05139, 2017

[67] Wang X, Gupta A. Generative image modeling using style and structure adversarial networks//Proceedings of the European Conference on Computer Vision. Cham, Netherlands: Springer, 2016: 318-335

[68] Yang Jian-Wei, et al. LR-GAN: Layered recursive generative adversarial net-works for image generation. arXiv preprint arXiv: 1703.01560, 2017

[69] Huang X, Li Y, Poursaeed O, et al. Stacked generative adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 5077-5086

[70] Reed S, et al. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396, 2016

[71] Perarnau G, et al. Invertible conditional GANs for image editing. arXiv preprint arXiv:1611.06355, 2016

[72] Shaham T R, Dekel T, Michaeli T. SinGAN: Learning a generative model from a single natural image//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea, 2019: 4570-4580

[73] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany, 2015: 234-241

[74] Wang Ting-Chun, Liu Ming-Yu, Zhu Jun-Yan, et al. High-resolution image synthesis and semantic manipulation with conditional GANs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA, 2018: 8798-8807

[75] Lee H Y, Tseng H Y, Huang J B, et al. Diverse image-to-image translation via disentangled representations//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 35-51

[76] Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition. Hawaii, USA, 2017: 3722-3731

[77] Wang Zheng-Wei, Qi She, Tomas E W. Generative adversarial networks: A survey and taxonomy. arXiv preprint arXiv: 1906.01529, 2019

[78] Barratt S, Rishi S. A note on the inception score. arXiv preprint arXiv:1801.01973, 2018

[79] Lucic M, Kurach K, Michalski M, et al. Are GANs created equal? A large-scale study//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 700-709

[80] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 6626-6637

[81] Che Tong, et al. Mode regularized generative adversarial networks. arXiv preprint arXiv:1612.02136, 2016

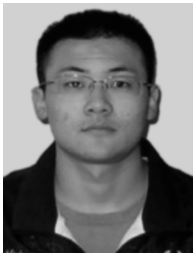
[82] Li J, Liang X, Wei Y, et al. Perceptual generative adversarial networks for small object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1222-1230

[83] Wang X, Shrivastava A, Gupta A. A-fast-RCNN: Hard positive generation via adversary for object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 2606-2615

[84] Kupyn O, Budzan V, Mykhailych M, et al. Deblur-GAN: Blind motion deblurring using conditional adversarial networks //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8183-8192

[85] Perarnau G, et al. Invertible conditional GANs for image editing. arXiv preprint arXiv:1611.06355, 2016

[86] He Z, Zuo W, Kan M, et al. AttGAN: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing, 2019, 28(11): 5464-5478



CHEN Fo-Ji, M. S. His research interests include image generation, machine learning, pattern recognition, visual measurement.

ZHU Feng, Ph. D., professor, Ph. D. supervisor. His research interests include robot vision, visual measurement, visual detection, infrared image simulation, and 3-D object recognition.

WU Qing-Xiao, Ph. D., professor. His research interests include robot vision and machine vision.

HAO Ying-Ming, Ph. D., professor. Her main research interests include image processing and spatial vision measurement.

WANG En-De, Ph. D., professor, M. S. supervisor. His research interests include small aircraft control, image detection, recognition and tracking and weak signal detection and preprocessing.

CUI Yun-Ge, M. S. His research interests include image generation and SLAM.

Background

Deep learning-based methods have achieved excellent performance in many vision tasks in recent years. But the good results always rely on large amounts of data with labels and powerful computing power. As the labeled data is hard to collect or even impossible to collect, which causes that fewer models are learned by the model and the generalization ability of the model is not well. Therefore, the application of methods based on deep learning to practical problems is difficult.

To efficiently complete vision tasks, it is necessary to collect more labeled data. For the excellent performance in the field of image generation, the generative adversarial networks have received a lot of attention. The generative adversarial networks model an unknown distribution in indirectly way and avoid computational difficulties. Compared with other methods in generative models, images generated by generative adversarial networks are high-quality. Therefore, it is a good idea to do infrared images augmentation based on images generation with generative adversarial networks.

To provide a comprehensive and systematic understanding of image generation based on generative adversarial networks for researchers who want to work on this field, it is necessary to carry out an investigation into the basic theory, model architecture, objective function, and some related tricks. In this paper, how the generative adversarial networks work and how to construct a model are introduced firstly. And then methods about images generation are discussed in details; At the same time, the fundamental theory and existing problems of current methods are discussed. A summary and analysis of methods which are used to do evaluation of generated images generated by generative adversarial networks is done. Finally, in theory, the existing problems and challenges are discussed; Meanwhile, some tricks that are employed to improve the performance of generative adversarial networks in practical applications are introduced and summarized. In practical application, doing images set augmentation which based on generative adversarial networks and guided by prior knowledge is a promising research direction. Meantime, it is hoped that images generation with generative adversarial networks can be applied to a wider range of areas.