计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200400107

基于特征融合的文本到图像的生成

徐 泽 帅仁俊 刘开凯 马 力 吴梦麟
南京工业大学计算机科学与技术学院 南京 211816
(1401283266@qq.com)

摘 要 近年来,基于生成对抗网络(Generative Adversarial Network,GAN)从文本描述中合成图像这一具有挑战性的任务已经取得了令人鼓舞的结果。这些方法虽然可以生成具有一般形状和颜色的图像,但通常也会生成具有不自然的局部细节且扭曲的全局图像。这是因为卷积神经网络在捕获用于像素级别图像合成的高级语义信息时效率低下,以及处于粗略状态的生成器-鉴别器由于缺少详细信息生成了有缺陷的结果,而这个结果会作为输入促使最终结果的生成。因此,提出了一种基于特征融合的生成对抗网络。该网络通过嵌入残差块特征金字塔结构来引入多尺度特征融合,并通过自适应融合这些特征直接生成最后的精细图像,仅使用一个鉴别器就可以生成256 px×256 px的逼真图像。将所提方法在花类数据集Oxford-102和加利福尼亚理工学院鸟类数据库CUB上进行验证,使用Inception Score和FID评估生成图像的质量,结果表明,生成图像的质量明显优于以往若干经典的方法。

关键词:特征融合;鉴别器;残差块特征金字塔;生成对抗网络
中图法分类号 TP391

Generation of Realistic Image from Text Based on Feature Fusion

XU Ze,SHUAI Ren-jun,LIU Kai-kai,MA Li and WU Meng-lin
College of Computer Science and Technology,Nanjing Tech University,Nanjing 211816,China

Abstract Recent challenging task of synthesizing images from text descriptions based on the generative adversarial network (GAN) has shown encouraging results. These methods can produce images with general shapes and colors,but often produce global images with unnatural local details and distortions. This is due to the inefficiency of the convolutional neural network in capturing high-level semantic information for pixel-level image synthesis and the fact that the generator-discriminator in a rough state generates flawed results for lack of detail,which then serves as input to the final result. We propose a generative adversarial network based on feature fusion,which introduces multi-scale feature fusion by embedding residual block feature pyramid structure,generates the final fine image directly by adaptive fusion of these features,and produces a 256px×256px realistic image with only one discriminator. The proposed method is verified on the flower data set Oxford-102 and Caltech bird database CUB,and the quality of generated images is evaluated by using Inception Score and FID. The results show that the quality of the generated images produced by the proposed method is better than images produced by some classical methods.

Keywords Feature fusion,Discriminator,Residual block feature pyramid,Generative adversarial network

1 引言

从文本中生成逼真图像是计算机视觉中的一个重要问题,即输入一段文本描述以输出包含该文本语义信息的图像,具有广泛的应用。例如,为不同品种的鸟类配插图,帮助警察搜寻嫌犯等。生成对抗网络被广泛应用于文本到图像的合成。Reed等[1]最先提出了基于生成对抗网络的文本到图像合成,但是他们的生成图像很小(64 px×64 px)并且缺乏细节。Zhang等[2]提出了堆叠生成对抗网络(StackGAN),该网络附加一个额外的GAN,以生成由低到高分辨率的图像。这种方法虽能够生成更加细致的图像,但需要训练两个单独的GAN。后来,Zhang等将StackGAN进一步扩展为使用树状结构逐步生成3种尺寸的图像(64 px×64 px,128 px×128 px,256 px×256 px)的StackGAN++[3]。在StackGAN++的基础上,Xu等[4]提出了深度注意力生成对抗网络(AttnGAN),该网络充分利用了生成对抗网络中的单词层级信息以生成细粒度图像。无论是StackGAN++还是AttnGAN都需要训练多个生成器和鉴别器,例如,生成256 px×256 px的图像需要与3个生成阶段相对应的3个鉴别器,这导致计算效率低下,且容易产生伪像。这是因为处于粗略状态的生成器-鉴别

到稿日期:2020-04-23 返修日期:2020-09-07
基金项目:国家自然科学基金(61701222)
This work was supported by the National Natural Science Foundation of China(61701222).
通信作者:帅仁俊(srjwhy@sina.com)

器因缺少详细信息而生成了有缺陷的结果，这个结果会作为输入促使最终结果的生成，从而导致生成了具有不自然的局部细节且扭曲的全局图像。

针对上述问题，我们提出了新型的端到端的框架，该框架是基于特征融合的生成对抗网络(Ff-GAN)，该网络通过嵌入残差块特征金字塔结构来引入多尺度特征融合，并通过融合这些特征直接生成最后的精细图像，使生成的图像更加逼真且具有更高的语义一致性。

我们在 CUB 鸟类[5]和 Oxford-102 花朵[6]这两个数据集上验证了所提方法。实验结果和分析证明，与现有技术水平相比，我们的方法是有效的，并且性能有很大的提高。

2 背景

过去几年中，基于深度生成网络的方法极大地推进了图像合成领域的发展。Kingma 等[7]使用随机反向传播来训练变分自动编码器(Variational Automatic Encoder，VAE)。Gregor 等[8]的 DRAW 模型通过具有卷积神经网络(Convolutional Neural Network，CNN)的注意力机制生成图像。此外，生成对抗网络[9]及其变体在图像建模(如图像合成、图像到图像[10]的转换)中取得了令人印象深刻的结果。最近，基于文本描述的高分辨率图像合成已成为 GAN 的一个有趣主题。Noh 等首先提出了结构化联合嵌入，接着 Reed 等[1]通过使用深度神经编码器生成的内部特征对其进行了改进。这些方法在生成图像时将整个编码的句子向量作为条件，采用 GAN 从文本描述中生成尺寸为 64×64 的令人印象深刻的图像。LAPGAN[11]专注于通过堆叠多个 GAN 来迭代细化图像。Zhang 等又通过共同训练多个生成器和鉴别器，进一步将 StackGAN[2]升级为 StackGAN++[3]。这种方法可以生成 3 种尺寸的引人注目的图像，并且比 StackGAN 更稳定。Zhang 等[12]又提出了分层嵌套的结构，以更好地适应多层次的鉴别器。为了生成详细的图像，这些高级模型都包含多个鉴别器以迭代优化图像，但是它们的计算效率不高，并且这些网络的不一致容易导致图像崩溃。

本文提出的 Ff-GAN 基于不同阶段进行特征融合可以很好地解决上述问题。我们采用 AttnGAN[4]中深度注意力多模态相似模型设计的文本编码器，将文本描述转换为句子条件和两个单词条件。我们在生成对抗网络的基础上引入了残差块特征金字塔注意模型，通过挖掘特征语义信息来准确定位每个特征区域对应的文本描述，进一步提高了目标图像细节合成的能力。

3 相关理论

3.1 生成对抗网络

Salimans 等[13]于 2014 年首次引入 GAN。基本 GAN 由两个网络组成：生成器和鉴别器，生成器接受随机噪声矢量 z，并学习数据分布以生成图像 $\boldsymbol{G}(z)$，而鉴别器区分 $\boldsymbol{G}(z)$ 是否为“真实”。在训练过程中，当生成器试图生成“真实”图像时，鉴别器旨在误导生成器。因此，这两个网络在 minmax 游戏中竞争，即：

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

其中，$D(t)$ 计算出 t 为“真”的概率，该概率应接近 1。生成的样本 t 的概率分布 P_g 收敛到训练输入的概率分布 P_{data}。鉴别器被训练用于最大化 V，而生成器的目的是最小化 V。

3.2 深度注意力文本条件

文本嵌入是为了学习文本和图像之间的对应关系，AttnGAN[4]提出了单词级注意力模型，该模型突出了更多细节。我们采用与 Xu 等类似的方法：预训练用于生成句子和单词向量的文本编码器，然后通过注意力模型将单词向量转换为文本条件，即：

$$\begin{aligned} H_0&=G_0(z,E)\\ H_j&=G_j(F_{j-1},A_j^{attn}(e,F_{j-1})) \end{aligned} \tag{2}$$

其中，z 是噪声矢量，$\boldsymbol{E}$ 表示从全局句子特征转换而成的条件向量。

3.3 空间金字塔

传统的图像金字塔任务从不同尺度的图像中提取特征，但这种方法增加了计算复杂度。空间金字塔网络已在许多研究中得到应用，但卷积展开可能导致局部信息丢失，造成特征图的局部不一致。例如，LD-Net[14]中提出的空间金字塔池化模块在不同尺度的池化操作中丢失像素定位，其网络模型如图 1 所示。

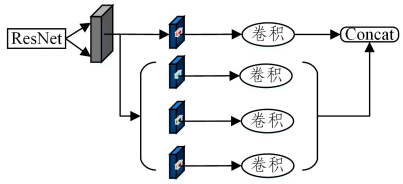

图 1 空间金字塔池化

Fig. 1 Space pyramid pooling

受空间金字塔和注意力机制的启发，我们的目标是从 CNN 提取的高级特征中准确提取像素级信息。

4 方法实现

4.1 特征金字塔注意模型(Feature Pyramid Attention Model，FPAM)

基于空间金字塔池化操作可以提高合成精度，我们在特征融合生成对抗网络的基础上引入了残差块特征金字塔注意模型(如图 2 所示)。通过多特征融合提高目标图像细节合成能力，同时实现高水平特征语义信息的挖掘，准确定位每个特征区域对应的句子信息。

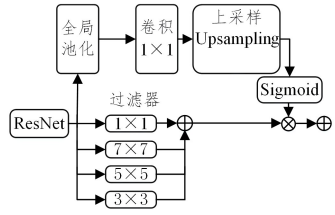

图 2 残差块特征金字塔模型

Fig. 2 Residual block feature pyramid model

两次向上采样后将文本描述输入到残差块中以合成低维

特征图，残差块特征金字塔模型可以通过 U 形网络结构将 4 个不同过滤器下的特征图融合在一起。为了获得较高的上下文信息质量，我们在残差块特征金字塔模型中引入了一种关注机制，以对更高级别的特征图产生更好的像素级别关注。该方法避免了使用像 LD-Net[14] 这样的信道级注意向量，避免了引入过多的计算。金字塔结构可以整合来自 4 个不同尺度的信息。此外，我们还引入了全局池化分支来捕获全局信息，使用 1×1 卷积核的目的是将残差网络的原始特征与金字塔注意特征相结合。

4.2　生成对抗网络

为了提高计算效率以生成高质量的图片，本文提出了一种基于端到端的架构，如图 3 所示。

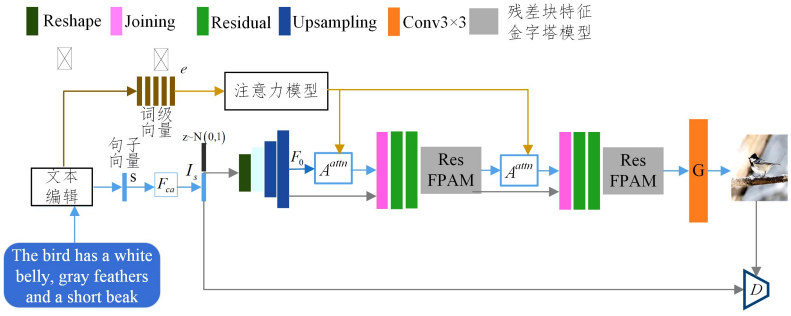

图 3　特征融合生成对抗网络

Fig. 3　Feature fusion generative adversarial network

通过自下而上的特征融合方法将低分辨率空间中的特征与高分辨率空间中的特征进行合并，所提出的特征融合生成对抗网络包含 3 个重要阶段，每个阶段具有多个输入和输出，即：

$$\boldsymbol{s}_i,\boldsymbol{e}_i=\varphi(d_i)$$

$$\boldsymbol{I}_s^i=F_{ca}(\boldsymbol{s}_i)$$

$$\boldsymbol{I}_w^{i,j}=A_j^{attn}(\boldsymbol{e}_i,F_j^i)$$

$$\hat{t}_i,(F_1^i,\cdots,F_{j+1}^i)=G(\boldsymbol{I}_s^i,\boldsymbol{I}_w^{i,j},z),j\in\{1,2\} \tag{3}$$

其中，$\varphi(\cdots)$是 Xu 等在 AttnGAN[4] 中设计的文本编码器；F_{ca} 代表条件增强，它将句子向量$\boldsymbol{s}_i$转化为全局条件句子向量$\boldsymbol{I}_s^i$；$\boldsymbol{I}_w^{i,j}$是由注意力模型计算得到的条件单词向量，它有两个输入：单词向量$\boldsymbol{e}_i$和图像特征 F_j^i。我们的方法中，一个图像文本对(t_i,d_i)就是一个训练样本。生成器生成相对应的合成图像 $\hat{t}_i$。全局条件句子向量$\boldsymbol{I}_s^i$首先与噪声向量$\boldsymbol{z}^i$结合以提取初始浅层特征，然后在接下来的两个阶段与条件单词向量$\boldsymbol{I}_w^{i,j}$连接以用于更高级别的特征合成，最后，$F_1^i,\cdots,F_{j+1}^i$是由生成网络在不同阶段生成的特征图，每一个特征图F_j都与下一个特征图F_{j+1}合并以进行全局特征融合。

判别器 D 有两个训练目标：1）评估输入图像是真实的还是伪造的；2）判断图像文本条件对是否匹配。我们的鉴别损失函数是条件损失和无条件损失的组合，其中，无条件损失用于确定图像是真实的还是伪造的；条件损失用于确定图像和句子是否匹配。我们为鉴别器提供了 5 种类型的输入形式。

（1）t_i：第 i 个真实图像。

（2）$\hat{t}_i$：第 i 个生成图像。

（3）$(\boldsymbol{t}_i|\boldsymbol{I}_s^i)$：第 i 个真实图像与相对应的匹配文本。

（4）$(\hat{t}_i|\boldsymbol{I}_s^i)$：第 i 个生成图像与匹配文本。

（5）$(t_i|\hat{\boldsymbol{I}}_s^i)$：第 i 个真实图像与不匹配文本。

鉴别器的损失函数为：

$$L_D=-\frac{1}{2}E_{t_i\sim p_{data}(t_i)}-\frac{1}{2}E_{\hat{t}_i\sim p_g(\hat{t}_i)}[\log(1-D(\hat{t}_i))]-$$

$$\frac{1}{3}E_{t_i\sim p_{data}(t_i)}-\frac{1}{3}E_{\hat{t}_i\sim p_g(\hat{t}_i)}[\log(1-D(\hat{t}_i|\boldsymbol{I}_s^i))]-$$

$$\frac{1}{3}E_{t_i\sim p_{data}(t_i)}[\log(1-D(t_i|\hat{\boldsymbol{I}}_s^i))] \tag{4}$$

其中，前两项为非条件损失，后三项为条件损失。

生成器的损失函数为：

$$L_G=-\frac{1}{2}E_{\hat{t}_i\sim p_g(t_i)}[\log(D(G(\boldsymbol{I}_s^i,\boldsymbol{I}_w^{i,j},z^i)))]-$$

$$\frac{1}{2}E_{\hat{t}_i\sim p_g(t_i)}[\log(D(G(\boldsymbol{I}_s^i,\boldsymbol{I}_w^{i,j},z^i),\boldsymbol{I}_s^i))] \tag{5}$$

其中，前一项为非条件损失，后一项为条件损失。

4.3　特征融合

较低分辨率的特征图虽然在空间上较粗糙，但在语义上更为准确。它们反映了图像的整体结构和颜色分布。本文在横向路径中对这些特征进行采样，并将其与通过主路径生成的特征图合并，这有助于收集不同的特征。加法和串联是两种常用的特征融合方法，在每个密集块中，每一层都与后续层相联以保留局部特征信息。当前密集块的输入和输出通过身份映射融合在一起，因为它们具有相同的尺寸大小，所以跨不同等级阶段的每个残差学习均被视为加权全局融合。每个阶段的输出特征图附加一个上采样层和 1×1 卷积以匹配其他阶段中的特征图的尺寸，从而形成多特征附加融合层，进而支持连续的状态转换。

如图 4 所示，我们提出的生成对抗网络大致可以分为 3 个阶段。阶段 1 用于从句子矢量$\boldsymbol{I}_s$和随机噪声$\boldsymbol{z}$的串联中提取浅层特征图F_1，即：

$$F_1=H_1(\boldsymbol{z},\boldsymbol{I}_s)$$

$$F_{d,1}=[F_1,\boldsymbol{I}_s,\boldsymbol{I}_w^1] \tag{6}$$

其中，$F_{d,1}$表示文本条件和特征图的F_1串联，其被应用于深层次的特征融合，即：

$$F_2=H_2(H_{dense}(F_{d,1})+F_{d,1})$$

$$F_{2,f}=F_2+F_1{}' \tag{7}$$

$$F_{d,2}=[F_{2,f},\boldsymbol{I}_s,\boldsymbol{I}_w^2]$$

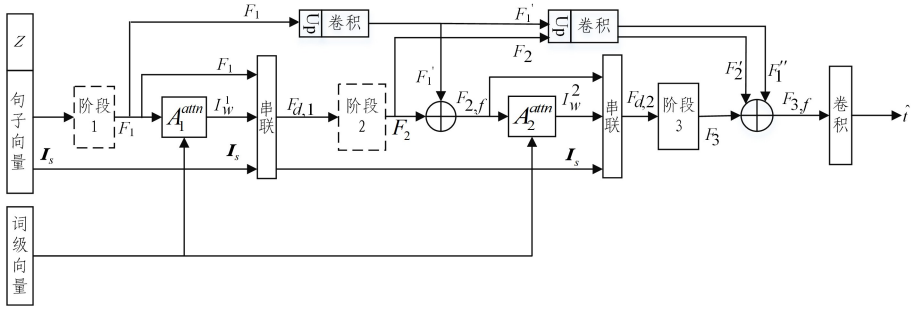

图 4 特征融合示意图

Fig. 4 Feature fusion diagram

根据 Huang 等[15]的研究，我们在接下来的两个阶段采用了密集层，这样做的优势在于不仅能在一定程度上解决梯度消失问题，还能更好地利用不同层的特征信息。$F_{2,f}$ 是通过全局残差学习实现的全局融合特征图之一，$F_{d,2}$ 是文本条件和融合特征图 $F_{2,f}$ 串联的结果。

阶段 3 和阶段 2 有相同的结构，因此我们可以采用相同的方式获得 F_3，即：

$$F_3 = H_3(H_{dense}(F_{d,2}) + F_{d,2}) \tag{8}$$

为了以全局方式利用各阶段特征，我们的生成器还在最后阶段的末尾采用全局特征融合，以自适应地融合来自不同层次的特征，即：

$$F_{3,f} = F_3 + F_2' + F_1'' \tag{9}$$

其中，

$$\begin{aligned} F_1'' &= H_{L2}(H_{L1}(F_1)) \\ F_2' &= H_{L2}(F_2) \end{aligned} \tag{10}$$

通过残差学习以加权相加的方式直接融合特征，该方法与连接方式相比，可以减少一半的参数和计算成本。

5 实验结果及分析

5.1 实验环境

本文的实验环境为 Python3.6，处理器为 i7-6800K，内存为 32GB，Linux 操作系统，同时配备 GTX1080Ti 显卡，研究方案主要基于开源的深度学习框架 Tensorflow。

5.2 数据集及评估指标

数据集：为了验证模型的有效性，分别在 CUB 和 Oxford-102 数据集上进行了实验，数据集如表 1 所列。我们使用文献[4]提供的预训练文本编码器将每个句子编码为句子嵌入向量和词嵌入向量。

表 1 实验数据

Table 1 Experimental data

图像数量	CUB[7]		Oxford-102[6]	
	训练集	测试集	训练集	测试集
	8855	2933	7034	1155
图像描述数量	10	10	10	10

评价指标：

(1)我们使用 Inception Score[14]来评估我们的方法，Stack-GAN 为 CUB 和 Oxford-102 提供了预训练的初始模型，其计算公式为：

$$I = \exp(E_x D_{KL}(p(y|x) \parallel p(y))) \tag{11}$$

其中，x 为生成样本；y 为 Inception model 预测的标签。好的生成模型应该生成多样且有意义的图像，因此，边缘分布 $p(y)$ 和条件分布 $p(y|x)$ 的 KL 散度越大越好。

(2)Inception Score 的缺点是输出样本不能与真实图像进行比较。它不能反映生成的图像是否更接近真实图像。因此，我们引入了另一个评估指标 Frechet Inception Distance (FID)，它是一个更加基于规则的、全面的度量标准，在评估生成样本的真实性和可变性方面，已经被证明与人类的评估更一致。众所周知，经过预处理的神经网络的顶层可以提取图像的高级信息，因此，这在一定程度上可以反映图像的本质。FID 的计算公式如下所示：

$$FID = \| \mu_r - \mu_g \|^2 + T_r(\boldsymbol{\Sigma}_r - \boldsymbol{\Sigma}_g - 2(\boldsymbol{\Sigma}_r \boldsymbol{\Sigma}_g)^{1/2}) \tag{12}$$

其中，μ_r 和 μ_g 分别是真实和合成图像特征的均值，$\boldsymbol{\Sigma}_r$ 和 $\boldsymbol{\Sigma}_g$ 分别是真实和合成图像特征的协方差矩阵。FID 值越低，合成数据分布与实际数据分布之间的距离越近。

5.3 结果与比较

图 5 给出了在 CUB 数据集上通过输入文本描述生成的精细图像，以及中间层融合特征图的可视化结果。因此，我们的方法产生了与输入文本一致的逼真的结果。

图 5 在 CUB 数据集上训练的模型合成的结果

Fig. 5 Model synthesis results trained on CUB data sets

将所提方法与最新的基于输入文本的 ResFPA-GAN[5]在 CUB 测试集上进行对比(见图 6)，本文方法生成的图像在颜色和细节上处理得更加细腻，也更加清晰。相比之下，Res-FPA-GAN[5]生成的鸟类外形过胖(如图 6(a)所示)，身材比例不协调(如图 6(b)所示)，眼睛和喙(如图 6(c)所示)等细节部位比较模糊，尾巴与文本不符以及出现头部伪像(如图 6(d)所示)。

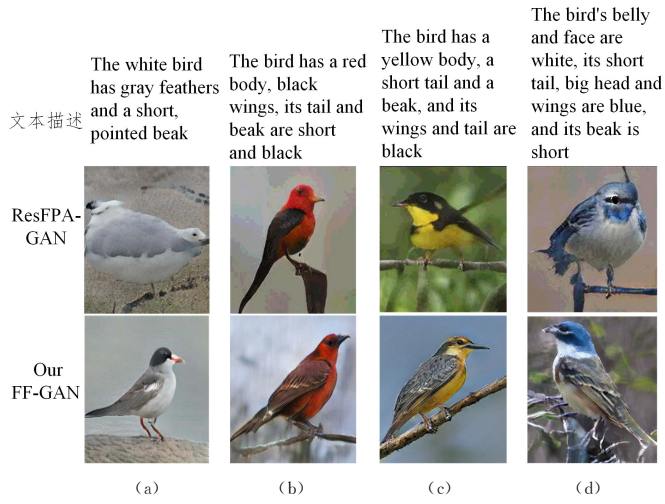

图 6　在 CUB 测试集上生成的图像示例

Fig. 6　Generated image example on the test set

图 7 给出了我们的方法与 StackGAN++[4] 在 Oxford-102 花类测试集上的实验对比。由图 7 可知，我们的方法生成的花朵的形状和颜色更加真实，StackGAN++生成的花朵很模糊，花瓣之间没有鲜明的层次感，花蕊部分存在形状不清晰、颜色错误等问题，不能充分展示文本描述中花瓣以及花蕊的特点。

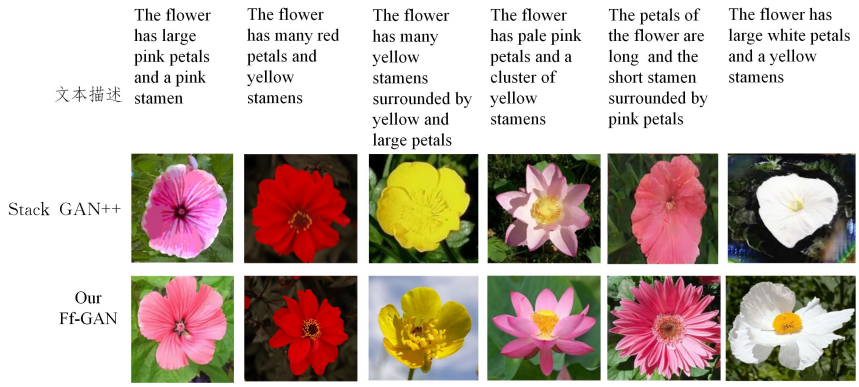

图 7　在 Oxford-102 测试集上生成的图像示例

Fig. 7　Generated image example on the flower class test set

使用文献[2]在 ImageNet 数据集上预训练的 Inception model 进行评估，得到 Inception Score(IS)值。表 2 列出了不同方法之间的 IS 值与 FID 值的比较，我们提出的方法在两个数据集上的 IS 值分别为 4.06 和 4.48，高于以往文献中提出的很多模型。FID 值的不断减小也说明了基于我们的方法所生成的样本更接近于真实图像。

表 2　各模型分别在 Oxford-102 和 CUB 测试集上的 IS 值和 FID 值

Table 2　IS and FID values of each model on the oxford-102 and CUB test sets

Method	Datasdet			
	Oxford-102		CUB	
	IS	*FID*	*IS*	*FID*
GAN-INT-CLS	2.66±0.03	—	2.88±0.04	—
StackGAN	3.20±0.01	55.28	3.70±0.04	32.12
StackGAN++	3.26±0.01	48.68	3.84±0.06	18.35
ResFPA-GAN	3.75±0.02	43.17	4.15±0.03	16.48
Our Ff-GAN	4.06±0.01	38.76	4.48±0.04	13.04

结束语　本文提出了一种基于特征融合的生成对抗网络，并将它成功地应用于文本合成图像这一任务中。Ff-GAN 的生成网络通过嵌入残差块特征金字塔结构来引入多尺度特征融合，与之前的模型相比，我们的 Ff-GAN 在生成一致的、高质量的图像方面做得更好，这是因为层次特征图的融合能够充分提取和利用局部和全局特征。此外，我们的端到端路径生成与专用鉴别器可以有效地解决以前方法中存在的网络不一致的问题。实验结果表明，我们的模型生成的图像不仅在色彩方面表现得更加丰富和细腻，在一些局部细微处也表现更出色。对于场景比较复杂的数据集，结合视觉对话语义以进一步提高生成图像的质量会是以后的研究工作。

参 考 文 献

[1] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis[J]. arXiv: 1605.05396, 2016.

[2] ZHANG H, XU T, LI H, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks [C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 5907-5915.

[3] ZHANG H, XU T, LI H, et al. Stackgan++: Realistic image

synthesis with stacked generative adversarial networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(8): 1947-1962.

[4] XU T, ZHANG P, HUANG Q, et al. Attngan: Fine-grained text to image generation with attentional generative adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1316-1324.

[5] WAH C, BRANSONS, WELINDERP, et al. The caltech-ucsd birds-200-2011 dataset. : CNS-TR-20111-001[R]. State of California: California Institute of Technology, 2011.

[6] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes[C]// 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing. IEEE, 2008: 722-729.

[7] KINGMA D P, WELLING M. Auto-encoding variational bayes [J]. arXiv: 1312. 6114, 2013.

[8] GREGOR K, DANIHELKA I, GRAVES A, et al. Draw: A recurrent neural network for image generation[J]. arXiv: 1502. 04623, 2015.

[9] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27: 2672-2680.

[10] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.

[11] DENTON E L, CHINTALA S, FERGUS R. Deep generative image models using a laplacian pyramid of adversarial networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 1486-1494.

[12] ZHANG Z, XIE Y, YANG L. Photographic text-to-image synthesis with a hierarchically-nested adversarial network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6199-6208.

[13] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training gans[J]. arXiv: 1606. 03498, 2016.

[14] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1520-1528.

[15] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected onvolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.

XU Ze, born in 1994, postgraduate. His main research interests include image processing and machine learning.

SHUAI Ren-jun, born in 1962, postgraduate, associate professor. His main research interests include artificial intelligence and intelligent medical care.