光电技术及应用

DOI：10.16818/j.issn1001－5868.2024052602

文本生成图像的 DF-GAN 增强模型研究

魏依然*，易军凯，朱柯全，谭玲玲
（北京信息科技大学 自动化学院，北京 100192）

摘　要： 针对文本生成图像任务中生成图像的语义关联性低、细节较模糊以及结构完整性较差等问题，提出了一种结合自注意力机制的深度融合生成对抗网络（Deep Fusion Generative Adversarial Networks，DF-GAN）增强模型。首先，利用 BERT 模型挖掘文本上下文语义特征，同时结合语义深度融合模块，实现深层的文本语义与图像区域特征的匹配。其次，在模型架构层面引入一种自注意力模块作为卷积模块的补充，目的是更好地建立长距离、多层次依赖关系。实验结果显示，所提出的增强模型不仅加强了文本与图片之间的语义联系，而且保证了生成图像的细粒度和完整性。

关键词： 文本生成图像；BERT 预训练模型；生成对抗网络；自注意力机制；语义深度融合模块

中图分类号： TP183　**文章编号：** 1001－5868(2024)06－0960－06

Deep Fusion-GAN Enhancement Model for Text-to-Image

WEI Yiran, YI Junkai, ZHU Kequan, TAN Lingling
(College of Automation, Beijing Information Science and Technology University, Beijing 100192, CHN)

Abstract: A deep fusion generative adversarial network (DF-GAN) enhancement model combined with self-attention mechanism is proposed for low semantic relevance, fuzzy details, and inadequate structural integrity in text-to-image tasks. First, the bidirectional encoder representations from transformers (BERT) model is used to mine the semantic features of text context and combined with the deep text-image fusion block to realize the matching of deep text semantics and image regional features. Second, a self-attention mechanism module is introduced as a supplement to the convolution module at the model architecture level, aiming to enhance the establishment of long-distance and multilevel dependencies. The experimental results demonstrate that the proposed enhancement model not only strengthens the semantic relationship between the text and image but also ensures the inclusion of precise details and the overall integrity of the generated image.

Key words: text-to-image; BERT pre-trained model; generative adversarial networks; self-attention mechanism; deep text-image fusion block

0　引言

文本生成图像是一种利用自然语言描述生成相应图像的技术，属于计算机视觉和自然语言处理的交叉领域，具有重要的理论意义和研究价值。例如可以用于增强人机交互、提高视觉内容的可信度、创造新颖的艺术作品等[1]。文本生成图像领域的研究难点主要包括如何理解文本的语义信息、如何将语

收稿日期：2024－05－26.
基金项目：国家自然科学基金项目(U1636208).
* **通信作者：**魏依然

义信息转换为图像特征、如何生成高质量和多样性的图像等。目前，文本生成图像领域的研究方法主要基于生成对抗网络(GAN)[2]，利用其强大的表征能力和生成能力，可实现一定程度的文本到图像的映射。然而，文本生成图像领域仍然面临着许多挑战和问题，例如如何处理复杂和多样的文本输入、如何提高图像的清晰度和真实性、如何加强文本信息与图像之间的语义一致性等。因此，文本生成图像领域仍有很大的研究空间和发展潜力，值得进一步探索和创新。

充分发掘文本的语义特征是图像生成的关键之一。早期模型在提取文本信息上，大多使用循环神经网络(RNN)[3]和长短期记忆神经网络(LSTM)[4]作为文本编码器，产生文本嵌入向量。但由于 RNN 和 LSTM 具有一定的局限性，这种通过提取序列特征的方式在提取语义特征的过程中难免会丢失一些文本信息，导致生成的图像难以满足所有的语义条件。

近年来，基于 GAN 的网络模型越来越多，其能够被用来生成多元化的图像。Reed 等[5]在 CGAN[6]基础上提出了 GAN-INT-CLS 网络模型，实现了单阶段生成 64×64 低分辨率的图像。Dash 等[7]提出了 TAC-GAN 网络模型，将生成的图像置于文本描述而不是类标签进行调节，使得生成的图像更具多样性[8]。

为实现高分辨率图像的合成并确保模型训练的稳定性，Zhang 等[9]提出了 StackGAN 模型，该模型采用了两阶段的堆叠网络，最终得到 256×256 高分辨率的图像；随后，他们又提出 StackGAN＋＋[10]模型，该模型由多个生成器和鉴别器组成，每个阶段的生成器都把前一个阶段生成的图像和文本嵌入作为条件，进一步提升了生成图像的质量。

单纯的堆叠网络虽然改善了生成图像的质量，但是无法保证生成图片特征的细粒度。本文提出深度融合对抗网络(DF-GAN[11])增强网络模型，引入 BERT[12]预训练模型作为其文本编码器，利用其强大的的语义信息挖掘能力来确保文本生成图像的语义一致性。同时引入自注意力模块，与 AttnGAN[13]模型中对输入序列的单词嵌入使用注意力不同，本模型是将自注意力引入到模型架构层面[14]，即在卷积特征层后引入自注意力模块，结合所有位置的特征来生成细节，保证每个位置的细节都与图像远处部分的细节进行仔细协调，从而保证了生成图像的质量。

1　DF-GAN 增强模型

DF-GAN 是一种针对文本生成图像的对抗网络模型，其主要由三部分组成：一个生成器、一个鉴别器以及一个预训练的文本编码器。在 DF-GAN 中，生成器的作用是将文本编码器所编码的句子向量和从高斯分布中随机采样得到的噪声向量进行合并，从而生成一张与输入文本对应的图像。鉴别器则通过对比输入的图像和真实图像间的相似度来判断当前生成的图像是否真实可信。噪声向量经过一个全连接层后，通过一系列上采样模块来上采样图像特征。其中，上采样模块以语义图像深度融合模块 DFBlock(Deep Text-Image Fusion Block)为核心，并结合了上采样层和残差块，用于在图像生成过程中深度融合文本和图像特征，最后经过一个卷积层将特征转换为图像。

本文优化改进 DF-GAN 网络模型，引入预训练的 BERT 模型[15]作为其文本编码器，对文本信息进行深度挖掘。在生成器卷积特征层后引入自注意力模块进行补充，有助于对图像区域之间的长范围、多层次依赖关系进行建模。DF-GAN 增强模型结构如图 1 所示。

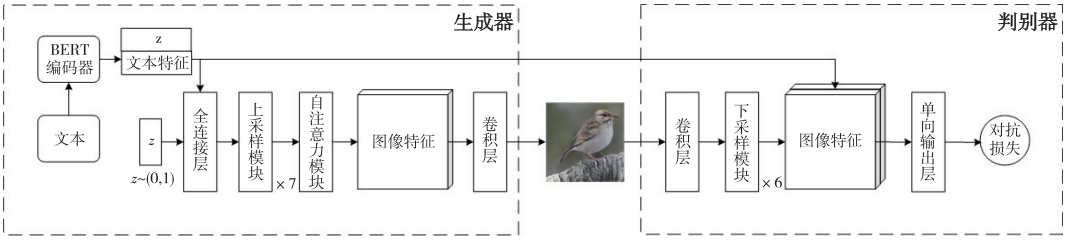


图 1　DF-GAN 增强模型

1.1　双向编码器 BERT

BERT 模型是一种基于 Transfomer[16]架构的预训练表征模型，通过在大量数据上进行训练，对下游任务的进行微调的方式，以提高模型的泛化能力

和性能，显著提高各项自然语言处理任务的准确率。BERT 模型能够通过上下文信息，从左到右和从右到左两个方向同时学习句子中每个词的表示，捕捉词语之间的复杂关系，增强语义理解和预测能力。由于 BERT 在大规模文本数据上进行了预训练，它能够捕获广泛的语言知识，成为一个强大的通用文本表示模型。这种预训练得到的语言表示可以用于各种下游任务，使 BERT 成为自然语言处理领域的重要突破之一。

BERT 的全称为 Bidirectional Encoder Representations from Transformer，有双向编码器表示的涵义。BERT 在预训练过程加入 MLM (Masked Language Model)和 NSP(Next Sentence Prediction)两个任务。GPT 等标准语言模型使用目标函数为：

$$P(\omega_i \mid \omega_1,\cdots,\omega_{i-1}) \tag{1}$$

BERT 使用的目标函数为：

$$P(\omega_i \mid \omega_1,\cdots,\omega_{i-1},\omega_{i+1},\cdots,\omega_n) \tag{2}$$

不同于 GPT 网络模型，BERT 使用能看到全局信息的目标函数进行训练，这是考虑到如果将自左向右编码和自右向左编码的单向编码器拼接起来，得到的双向编码器在性能、参数规模和效率等方面都不如直接使用深层的双向编码器。

MLM 任务是指掩码语言模型任务，此任务目的是让 BERT 模型能够学习到双向的上下文表示。MLM 任务的做法是随机地将输入序列中的一些词替换为特殊符号“MASK”，然后让模型预测被替换的词原本是什么。这样可以使模型利用左右两边的信息来推理被替换的词的含义，从而更好地捕捉单词之间的关系和语境。

NSP 任务是指下一句预测任务，此任务的目的是让 BERT 模型能够学习到句子级别的关系。NSP 任务的做法是给模型输入两个句子，并让模型来判断这两个句子在原文中的位置是否连续，这样可以增强模型对句子间相关性的把握能力。

BERT 模型通过对 Transformer 的编码器进行堆叠，利用 Transformer 较强的特征提取能力，来实现对多层特征信息的提取。Transformer 是一种基于自注意力机制的深度学习模型，自注意力的计算公式为：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Softmax\left(\frac{\boldsymbol{QK}^T}{\sqrt{\boldsymbol{d}_k}}\right)\boldsymbol{V} \tag{3}$$

式中，$\boldsymbol{Q} \in \boldsymbol{R}^{n\times d_k}$ 表示查询矩阵，$\boldsymbol{K} \in \boldsymbol{R}^{n\times d_k}$ 表示键矩阵，$\boldsymbol{V} \in \boldsymbol{R}^{n\times d_v}$ 表示值矩阵，$\boldsymbol{d}_k$ 和 $\boldsymbol{d}_v$ 分别为嵌入向量和 $\boldsymbol{V}$ 的维度，$\boldsymbol{QK}^T$ 为单词之间的相似度向量。

BERT 是一种基于多头注意力的神经网络架构，以 Transformer Enconder 为核心模块，它由编码器和解码器组成，可以实现端到端的序列到序列的映射。多头注意力机制的具体计算公式为：

$$MultiHead(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(\boldsymbol{head}_1,\boldsymbol{head}_2,\cdots,\boldsymbol{head}_h)\boldsymbol{W}^O \tag{4}$$

$$\boldsymbol{head}_i = Attention(\boldsymbol{QW}_i^Q,\boldsymbol{KW}_i^K,\boldsymbol{VW}_i^V) \tag{5}$$

式中，$\boldsymbol{W}_i^Q \in \boldsymbol{R}^{d\times d_k}$，$\boldsymbol{W}_i^K \in \boldsymbol{R}^{d\times d_k}$，$\boldsymbol{W}_i^V \in \boldsymbol{R}^{d\times d_v}$ 为第 i 个头部的投影参数矩阵；$\boldsymbol{W}^O \in \boldsymbol{R}^{d\times d_v}$ 为来自不同头部的信息输出投影矩阵。

1.2 语义深度融合模块

在跨模态领域中，传统模型不仅不能充分利用文本信息，同时计算成本还高，不能有效结合文本信息和图像特征。为解决这一问题，DF-GAN 模型提出一种语义图像深度融合模块(DFBlock)。

DFBlock 由 Affine 仿射变换层[17]和 ReLU 激活函数层组成。使用仿射变换层的目的是扩展生成器的条件表示空间。然而，仿射变换在本质上是线性变换，仅依靠基于每个通道的线性变换并不利于文本图像的融合。因此，为让融合过程具有非线性特性，将 ReLU 激活函数层作为补充引入到仿射变换层之间。这样图像特征可以根据文本语义信息被生成器映射到不同的表示中[5]。DFBlock 架构如图 2。仿射变换层是一个轻量级模块，其结构如图 3。

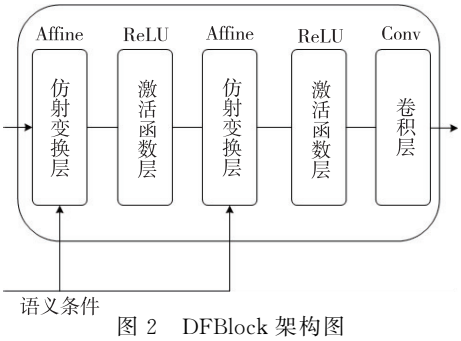


图 2 DFBlock 架构图

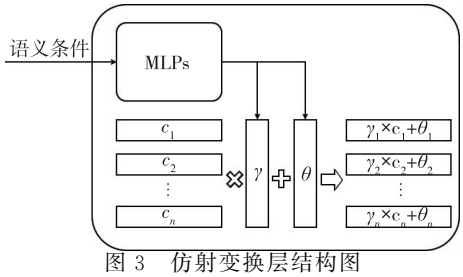


图 3 仿射变换层结构图

通过按通道缩放和移位操作来控制视觉特征图。采用多层感知器，从句子向量 $\boldsymbol{e}$ 来分别预测语言条件下的通道缩放参数 $\boldsymbol{\gamma}$ 和移位参数 $\boldsymbol{\theta}$：

$$\boldsymbol{\gamma}=MLP_1(\boldsymbol{e}),\boldsymbol{\theta}=MLP_2(\boldsymbol{e}) \tag{6}$$

对于给定的输入特征图 $\boldsymbol{X}\in\boldsymbol{R}^{B\times C\times H\times W}$，分别利用缩放参数 $\boldsymbol{\gamma}$ 和位移参数 $\boldsymbol{\theta}$ 对特征 $\boldsymbol{X}$ 进行按通道缩放以及平移运算。该过程可以表示为：

$$AFF(\boldsymbol{x}_i\mid\boldsymbol{e})=\boldsymbol{\gamma}_i\cdot\boldsymbol{x}_i+\boldsymbol{\theta}_i \tag{7}$$

式中，AFF 表示仿射变换，$\boldsymbol{x}_i$ 是特征图的第 i 个通道，$\boldsymbol{\gamma}_i$ 是特征图相应通道下的的缩放参数，$\boldsymbol{\theta}_i$ 则是此通道下的移位参数。

1.3　自注意力模块

在图像生成领域，大多数 GAN 模型都是在卷积层的基础上构建的。它们难以有效捕捉图像中的长距离依赖关系，这是由于卷积层只能处理局部邻域的信息，从而降低了计算效率。同时以往的注意力机制仅仅应用于对输入序列的单词的嵌入使用，还没有应用到模型架构层面。为解决这一问题，本文将自注意力机制融入 GAN 的架构中，使得生成器和判别器都能够对不同空间区域之间的关系进行建模。自注意力模块的结构如图 4 所示。

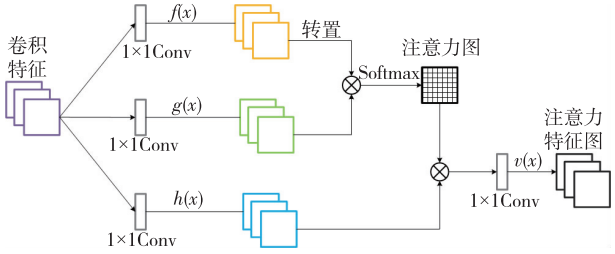


图 4　自注意力模块

首先将来自卷积层图像特征变换到两个特征空间中以计算注意力。

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{W}_{\mathrm{f}}\boldsymbol{x} \tag{8}$$

$$\boldsymbol{g}(\boldsymbol{x})=\boldsymbol{W}_{\mathrm{g}}\boldsymbol{x} \tag{9}$$

将 $\boldsymbol{f}(\boldsymbol{x})$ 的进行转置处理得到 $\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x})$，并与 $\boldsymbol{g}(\boldsymbol{x})$ 相乘，得到一个相关性矩阵 $\boldsymbol{S}$，即特征图上各个像素点之间的相关性。

$$\boldsymbol{s}_{ij}=\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_i)\boldsymbol{g}(\boldsymbol{x}_j) \tag{10}$$

再将相关性矩阵 $\boldsymbol{S}$ 经过 Softmax 归一化处理得到矩阵 $\boldsymbol{\beta}$，矩阵 $\boldsymbol{\beta}$ 可以看作一种注意力图，每一行向量代表了一种注意力的方式。$\beta_{j,i}$ 表示在合成第 j 个区域时，模型对第 i 个位置的影响程度。

$$\beta_{j,i}=\frac{\exp(s_{ij})}{\sum_{i=1}^{N}\exp(s_{ij})} \tag{11}$$

将注意力特征图与 $\boldsymbol{h}(\boldsymbol{x})$ 相乘，得到 N 个新的像素点作为输出 o。即每一个像素点都与整个特征图相关，相关性来自于 $\boldsymbol{\beta}$ 矩阵。

$$o_j=v\Big(\sum_{i=1}^{N}\beta_{j,i}h(x_i)\Big) \tag{12}$$

$$h(x_i)=\boldsymbol{W}_{\mathrm{h}}x_i \tag{13}$$

$$v(x_i)=\boldsymbol{W}_{\mathrm{v}}x_i \tag{14}$$

最后将注意力特征图与特征向量 $\boldsymbol{x}$ 融合，得到结合注意力的特征映射 $\boldsymbol{y}$，即：

$$y_i=\gamma o_i+x_i \tag{15}$$

式中，γ 为比例参数，$\boldsymbol{W}_{\mathrm{f}},\boldsymbol{W}_{\mathrm{g}},\boldsymbol{W}_{\mathrm{h}},\boldsymbol{W}_{\mathrm{v}}$ 是学习的权重矩阵。

2　实验及结果分析

2.1　实验设置

本次实验设置如表 1 所示。

表 1　实验设置

名称	配置
操作系统	Linux
CPU	Intel i9
GPU	RTX 4090
CUDA	11.4
Python	3.7
PyTorch	1.11.0

2.2　数据集

采用由加州理工学院在 2010 年提出 CUB 鸟类数据集，该数据集共有 11 788 张鸟类图像，包含 200 种鸟类子类，每张图片配以 10 句文本描述，包含了鸟的关键部位信息以及鸟的属性等。

2.3　评估指标

IS（Inception Score）[18] 和 FID（Fréchet Inception Distance）[19] 主要用于评价生成对抗网络生成的图片的质量和多样性。它们都利用了预训练的 Inception V3 网络，这是一个在 ImageNet 数据集上训练的图片分类器。

IS 评估指标通过测量边缘分布和条件分布之间的 KL 散度，表示两个分布的差异。KL 散度越大，说明两个分布越不相似，生成的图像越符合要求。KL 散度又称为相对熵或 KL 距离。

$$IS=\exp(E_xD_{\mathrm{KL}}(p(y\mid x)\|p(y))) \tag{16}$$

式中，x 为数据特征，y 为网络模型预测的标签，$p(y\mid x)$是给定图片 x 时的类别概率分布，$p(y)$是边缘概率分布，D_{KL} 是 KL 散度。IS 的数值越大，

表示生成图片既清晰又多样。

FID 用于描述两个数据集的相似程度，计算两个数据集合的均值和协方差矩阵之间的 Fréchet 距离，表示两个数据集合之间的距离。

$$FID = \|\mu_{x_1} - \mu_{x_2}\|_2^2 + \mathrm{Tr}\left(\boldsymbol{\Sigma}_{x_1} + \boldsymbol{\Sigma}_{x_2} - 2(\boldsymbol{\Sigma}_{x_1}\boldsymbol{\Sigma}_{x_2})^{\frac{1}{2}}\right) \tag{17}$$

式中，x_1，x_2 分别代表真实图像和生成图像，μ_{x_1}，μ_{x_2} 表示两种图像的均值，$\boldsymbol{\Sigma}_{x_1}$，$\boldsymbol{\Sigma}_{x_2}$ 表示两种图像的协方差矩阵。*FID* 值越小，表示生成图像与真实图像之间的数据越接近，生成图像的质量越高。

2.4 结果分析

2.4.1 定量分析

本文引入一些在文本生成图像领域中较为先进的模型来进行对比，结果如表 2 所示。显然，与其他模型相比，本文提出的增强模型在性能指标上有明显改善。本文模型的 *IS* 得分为 4.98±0.03，在原 DF-GAN 模型上提升了 0.12。*FID* 得分为 15.77，在原 DF-GAN 模型上降低了 3.47。由此可见，本文方法大大改善了模型的性能，有效地提高了生成图片的质量。

表 2 不同模型间的性能指标对比

方法	评价指标	
	IS	*FID*
GAN-INT-CLS	2.88±0.04	68.79
StackGAN++	3.84±0.06	30.30
AttnGAN	4.36±0.03	23.98
MirrorGAN	4.56±0.05	18.34
DF-GAN	4.86±0.04	19.24
本文方法	4.98±0.03	15.77

2.4.2 定性分析

本次实验引入两种具有代表性的模型 AttnGAN 和 DF-GAN 与本文方法进行对比。

对比结果如图 5 所示。显然，AttnGAN 生成的图像细节较为模糊，且语义一致性较差，例如第一张图片错误地将鸟的头绘制成黑色，与所给定的语义条件不符。而第二张图片并没有体现出语义中的“黑色的头”，发生了语义缺失，而且出现了“树枝长在鸟身上”这种不合理的现象。DF-GAN 的表现整体上差强人意，但在构图细节上还是表现还是有待提高，比如第三张图像中生成的鸟缺失了一条腿。



图 5 可视化对比

相比之下，所提出的增强模型无论是在语义一致性上，还是在生成图片的细粒度和完整性上都具有明显优势，尤其是在较复杂语境“这只鸟有红色的喙，黑色的翅膀，蓝色的羽毛和白色的胸脯”下，本文模型生成的图片不仅满足了语义的全部细节，而且兼具了图片在细节构图上的完整性。

2.5 消融实验

为进一步探究本文方法的有效性，将继续在 CUB 数据集上，以 DF-GAN 模型为基准模型，对引入的 BERT 编码器和自注意力模块进行消融实验。

实验结果如表 3 所示。

表 3　消融实验结果

方法	实验结果	
	IS	*FID*
DF-GAN	4.86±0.04	19.24
＋Bert	4.92±0.03	17.89
＋SA	4.90±0.05	17.36
＋Bert＋SA	4.98±0.03	15.77

由表 3 可以看出，引入 BERT 预训练模型和自注意力模块显著提高了模型的性能。BERT 模型对文本信息有强大的挖掘能力，这有助于更好地理解文本和图像之间的关联性。在网络框架中引入自注意力模块，可以使用来自所有特征位置的线索来生成细节，以此保证图像的细节与结构完整性。

3　结论

本文没有采用传统的堆叠架构模型，而选择以单阶段框架模型 DF-GAN 为基准，避免了堆叠结构中不同生成器之间的特征纠缠。引入 BERT 预训练模型作为文本编码器，结合语义深度融合模块 DFBlock，在语义一致性上取得了较好的实验结果。并在模型架构层面引入自注意力机制，这使得模型在 IS，FID 指标和实际生成图片的结果中都有明显改善，大大提高了生成图像的质量。实验结果表明，所提出的增强模型在文本图像生成领域有着良好的效果，与当前最先进的方法相比，在合成逼真图像和文本匹配图像方面更简单并且更高效，有效地改善了以往模型在语义一致性、生成图像的完整性以及图像细粒度等方面的问题。

参考文献：

[1] Li D, Wang S, Zou J, et al. Paint4Poem: A dataset for artistic visualization of classical Chinese poems[J]. arXiv preprint arXiv: 2109.11682, 2021.

[2] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.

[3] Elman J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.

[4] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[5] Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis[C]// ICML'16: Proc. of the 33rd International Conference on Machine Learning, 2016: 1060-1069.

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

[7] Dash A, Gamboa J C B, Ahmed S, et al. Tac-GAN text conditioned auxiliary classifier generative adversarial network[J]. arXiv preprint arXiv: 1703.06412, 2017.

[8] 赖丽娜，米　瑜，周龙龙，等. 生成对抗网络与文本图像生成方法综述[J]. 计算机工程与应用，2023，59(19)：21-39.
Lai Li'na, Mi Yu, Zhou Longlong, et al. Survey about generative adversarial network and text-to-image synthesis[J]. Computer Engineering and Applications, 2023, 59(19): 21-39.

[9] Zhang H, Xu T, Li H, et al. StackGAN: Text to photorealistic image synthesis with stacked generative adversarial networks[C]// Proc. of the IEEE International Conference on Computer Vision, 2017: 5907-5915.

[10] 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.

[11] Tao M, Tang H, Wu F, et al. DF-GAN: A simple and effective baseline for text-to-image synthesis[J]. arXiv preprint arXiv: 2008.05865, 2020.

[12] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv: 1810.04805, 2018.

[13] Xu T, Zhang P, Huang Q, et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks[J]. arXiv preprint arXiv: 1711.10485, 2017.

[14] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[J]. arXiv preprint arXiv: 1805.08318, 2018.

[15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv: 1810.04805, 2018.

[16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// 31st Conference on Neural Information Processing Systems (NIPS), 2017: 1-15.

[17] Perez E, Strub F, De Vries H, et al. FiLM: Visual reasoning with a general conditioning layer[J]. arXiv preprint arXiv: 1709.07871, 2017.

[18] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs[J]. arXiv preprint arXiv: 1606.03498, 2016.

[19] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a nash equilibrium[J]. arXiv preprint arXiv: 1706.08500, 2017.

作者简介：

魏依然(1999—)，男，天津市人，硕士研究生，主要研究领域为机器学习。