第 39 卷 第 3 期
2024 年 6 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 39 No. 3
Jun. 2024

DOI:10.13364/j.issn.1672-6510.20230084

基于十字形窗口的生成对抗网络模型

王 丹，王鹏程，张桉祺，王子涵
（天津科技大学人工智能学院，天津 300457）

摘 要：由于传统的生成对抗网络(generative adversarial network，GAN)都是以卷积神经网络(convolutional neural networks，CNN)作为基本框架，CNN 无法处理远程依赖关系，因此会导致图片特征分辨率低和精细细节损失的问题。CSWin Transformer 中的十字形窗口自注意力机制可以有效捕获图像组件之间的远程依赖关系，本文提出一种基于 CSWin Transformer 的生成对抗网络模型 CTGAN(CSWin Transformer GAN)，模型在 CIFAR-10 数据集和更高分辨率的 CelebA 数据集上进行测试，模型表现出了较好的生成效果，可以生成保真度高且细节丰富的图片。

关键词：生成对抗网络；CSWin Transformer；生成模型
中图分类号：TP181 **文献标志码**：A **文章编号**：1672-6510(2024)03-0064-08

Generative Adversarial Network Model Based on Cross-Shaped Window

WANG Dan，WANG Pengcheng，ZHANG Anqi，WANG Zihan
(College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China)

Abstract：Since traditional generative adversarial networks(GAN) are based on convolutional neural networks(CNN) as the basic framework，CNN cannot process remote dependency relationships. As a result，image feature resolution and fine detail loss will be caused. The cross-shaped window attention mechanism in CSWin Transformer can effectively capture remote dependencies between image components. Therefore，in this article we propose a generative adjoint network model CTGAN (CSWin Transformer GAN) based on CSWin Transformer. The model was tested on the CIFAR-10 datasets and the CelebA datasets with higher resolution，and it showed a good generation effect. Moreover，high fidelity and detailed images can be generated.

Key words：generative adversarial network；CSWin Transformer；generative model

引文格式：
王丹，王鹏程，张桉祺，等. 基于十字形窗口的生成对抗网络模型[J]. 天津科技大学学报，2024，39(3)：64–71.
WANG D，WANG P C，ZHANG A Q，et al. Generative adversarial network model based on cross-shaped window[J]. Journal of Tianjin university of science & technology，2024，39(3)：64–71.

近年来，图像生成领域取得了巨大进步，在各种生成模型中生成对抗网络(generative adversarial network，GAN)具有强大的竞争力，特别是基于 StyleGAN[1]的架构已经广泛应用于图像翻译、图像增强和图像编辑任务中。这些模型都是基于卷积神经网络(convolutional neural networks，CNN)，模型虽然有很多优点，但是依然有多种缺陷，例如模型利用卷积核或滤波器不断地提取抽象的高级特征，理论上其感受野应该能覆盖到全图，但许多研究表明其实际感受野远远小于理论感受野。这不利于研究人员充分利用上下文信息进行特征的捕获。模型虽然可以不断地堆叠更多的卷积层，但这显然会造成模型计算量急剧增加[2]，并且会导致丢失空间信息和图像精细细节损失的问题。

收稿日期：2023–04–04；**修回日期**：2023–09–11
基金项目：复杂电子系统仿真重点实验基金项目(DXZT–JC–ZZ–2020–013)；复杂能源系统智能计算教育部工程研究中心开放基金项目(ESIC202102)
作者简介：王 丹(1983—)，女，湖北黄冈人，讲师，wanghzc@163.com

GAN 在训练过程中很容易发生模式坍塌，所以训练不稳定一直是 GAN 最大的问题。WGAN 模型[3]引入 Wasserstein（最优传输）距离，很好地缓解了模型的不稳定性。InfoGAN 模型[4]在随机噪声中增加了可解释性的潜在编码，增加了对 GAN 模型的可解释性，能够控制图像的生成，但是增加了计算负担，造成生成图像多样性不足。LSGAN 模型[5]用最小二乘损失函数代替了 GAN 模型的交叉熵损失函数，解决了梯度消失的问题，可以生成高质量图像，但是会降低生成样本的多样性。BEGAN 模型[6]从 Wasserstein 距离得出的损失匹配自动编码器的损失分布，这不仅可以让模型收敛更快，而且判别器和生成器训练平衡，但是在超参数选取上有一定难度。BigGAN 模型[7]对生成器应用正交正则化可以使用简单的“截断技巧”进行训练，使模型训练更稳定，且生成的图像品质更好，但是模型参数大、计算成本高。StyleGAN 模型[1]采用逐级生成的方式提高图像生成的质量，但是这些模型都是卷积神经网络，其局部感受野使模型难以捕获图像信息，远距离依赖性和理解对象的全局结构使它在生成复杂场景方面还有待提升。

随着基于 Transformer（变换器）的预训练模型在自然语言处理（natural language processing，NLP）领域展示出了强大的性能，越来越多的工作将 Transformer 引入计算机视觉领域。ViT（vision transformer）是一种用于图像分类的 Transformer 架构，在视觉任务中展示出良好的性能[8]，Transformer 在广泛的判别任务中逐渐占据了主导地位。自从 CSWin Transformer[9]及其改进模型被提出后，模型中的核心部分十字形窗口自注意力机制具有更强的特征提取能力，在计算机视觉三大领域（图像分类、目标检测和语义分割）上表现出了强大的性能，并超越了 ViT 和 ResNet（residual neural network）。作为一个适合于视觉任务的模型，十字形窗口自注意力机制具有更大的感受野，可以捕获图像全局位置信息，解决长距离依赖问题。

在图像的生成模型方面，Transformer 还没有表现出与 CNN 相当的能力。与其他基于 Transformer 的视觉模型相比，仅使用 Transformer 构建 GAN 似乎更具挑战性。这是因为与分类等任务相比，真实图像生成的门槛更高，而且 GAN 训练本身具有较高的不稳定性。近年来，研究人员将生成对抗网络与 Transformer 模型结合，提出了一些可以提高生成质量的模型。Zhang 等[10]提出 SAGAN（self-attention GAN），在 GAN 中引入了自注意力机制，使生成器可以对图像中的每个像素点进行注意力加权，生成器由一系列的反卷积操作组成，并通过自注意力机制生成下一层的特征图，虽然在图像生成方面取得了一定的成功，但是模型的主体结构仍然是 CNN，依然不能很好地生成多样化的图像。Jiang 等[11]提出使用纯粹的 Transfomer 构建无卷积 GAN 模型 TransGAN。TransGAN 定制了一个基于 Transformer 的生成器和一个多尺度金字塔结构的判别器，配备网格自注意力机制，其中生成器采用逐级放大分辨率的方式减小计算量，每一个层级之间采用上采样模块提高分辨率。TransGAN 在 CIFAR-10 等常用数据集中取得了较好的成绩，并超过 StyleGAN-v2 等经典的生成模型，但是 TransGAN 模型中减少了全连接层（multilayer perceptron，MLP），因而无法合成高保真细节，并且模型需要大量的计算资源和训练时间。

本文提出一种基于 CSWin Transformer 的 GAN 模型 CTGAN（CSWin Transformer GAN），借鉴了基于样式生成器[1]的模型结构，以 CSWin Transformer 为模型的基本构建块。CSWin Transformer 中的十字形窗口自注意力有两组，其中一组在水平方向上做条纹自注意力，另外一组在垂直方向上做条纹自注意力，扩大了感受野，在不增大计算量的情况下有效提高生成器的容量。此外，本文将相对位置编码 LePE 和 SPE 引入模型中，增加模型生成图像的相对位置信息和全局位置信息。同时，为了解决 GAN 训练不稳定的问题，引入 Wasserstein 距离和梯度惩罚，解决 GAN 训练过程中的模式崩塌问题，模型在 CIFAR-10 数据集上取得很好的生成效果。当涉及更高分辨率的数据集 CelebA 的生成任务时，依然可以保持高保真度和合理纹理细节。

1　CTGAN 模型结构设计

1.1　模型整体架构

基于样式生成器的架构[1]被公认为是最先进的高质量图像生成器，包括 Mapping Network、AdaIN 等，使 StyleGAN 能够生成高质量的图像数据，被广泛应用于图像生成领域。CTGAN 模型生成器的网络结构如图 1 所示。模型借鉴了基于样式的体系结构，增强模型的生成能力，左边映射网络输入 512 维的噪声向量 $\boldsymbol{z}$ 经过全连接层进行空间映射到 $\boldsymbol{W}$。全连接层可以对特征进行解耦，将解耦后的特征输入模型右

侧主体的合成网络中。模型的右边是模型合成的主体部分，采用逐级生成的方式，开始输入一个 512 维的常量，进入 CSWin Transformer Block 中进行特征提取，从分辨率 4×4 的图片开始进入上采样模块，不断增加生成图像的尺寸，逐级提高图片的分辨率，每个模块都会输入仿射变换的特征，用于对每个通道进行缩放、偏移，影响的方式称为 AdaIN（自适应实例归一化），从而实现对生成图片样式的控制，比如生成图像的角度、姿态。

为了扩大感受野，使模型尽可能把握图像的全局特征，克服卷积神经网络局部感受野难以捕捉远距离依赖性导致图片细节丢失的缺点。在合成网络的主体部分，以 CSWin Transformer Block 为基本的构建块进行图像生成，用通过一系列变化的 CSWin Transformer Block 模块进行特征提取，然后对特征图进行上采样，增加图片的尺寸。每个生成模块在标准化之后加入 MLP，用于调整特征图和网络的权重，实现对生成器输出图片的精细控制，增加图像生成的细节。

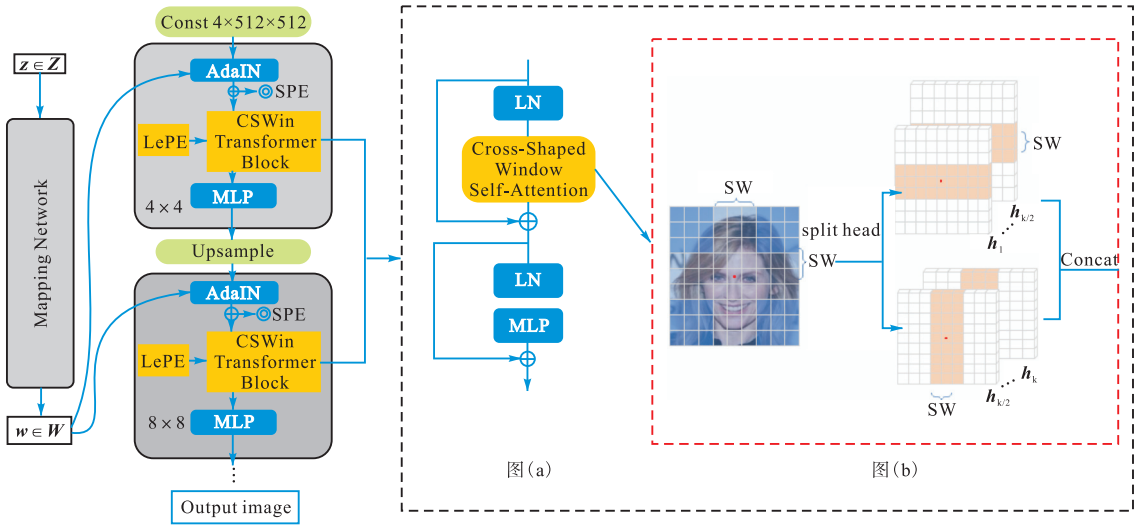

图 1　CTGAN 模型生成器的网络结构

Fig. 1　CTGAN generator network structural diagram

1.2　CSWin Self-Attention 模型的结构设计

CSWin Transformer Block 模块的核心部分是十字形窗口自注意力机制 CSWin Self-Attention（cross-shaped window self-attention）。由于 CNN 无法建立较远像素点之间的关联，因此无法获得全局感受野。注意力机制能够有效扩大感受野和图像建模能力，CSWin Self-Attention 将每个头均分成两组，一组在水平方向做条纹自注意力，另一组在垂直方向做条纹自注意力，完成后两个并行组拼接在一起，有效扩大了感受野，有利于生成图片的精细结构。

在生成图像的过程中，对于给定 l 层的特征图 $\boldsymbol{X}\in\boldsymbol{R}^{(H\times W)\times C}$，线性投影到 k 个头上，然后在水平或垂直条纹上执行局部注意力，如图 1(b)所示。对于水平条纹自注意力，生成的特征图均匀划分为不重叠的等宽的水平条纹[$\boldsymbol{X}^1$，…，$\boldsymbol{X}^M$]。假设第 k 个头的查询（Query）、键（Key）和值（Value）都有维度 d_k，则第 k 个头的水平条纹自注意力为

$$\text{H-Attention}_k(\boldsymbol{X})=\left[\boldsymbol{Y}_k^1,\boldsymbol{Y}_k^2,\cdots,\boldsymbol{Y}_k^M\right] \tag{1}$$

其中：$\boldsymbol{Y}_k^i=\text{Attention}\left(\boldsymbol{X}^i\boldsymbol{W}_k^Q,\boldsymbol{X}^i\boldsymbol{W}_k^K,\boldsymbol{X}^i\boldsymbol{W}_k^V\right)$，$\boldsymbol{Y}_k^i$ 表示使用 CSWin Self-Attention 得到的特征，$\boldsymbol{W}_k^Q\in\boldsymbol{R}^{C\times d_k}$、$\boldsymbol{W}_k^K\in\boldsymbol{R}^{C\times d_k}$、$\boldsymbol{W}_k^V\in\boldsymbol{R}^{C\times d_k}$ 分别表示第 k 个头的查询、键和值的投影矩阵；d_k 设为 C/K（K 为总头数），垂直条纹自注意力与此类似，其第 k 个头的输出表示为 $\text{V-Attention}_k(\boldsymbol{X})$。将 K 个注意力头平均分成两个平行组，每个组有 $K/2$ 个头部，K 通常是偶数，第一组的注意力头执行水平条纹自注意力，第二组的头执行垂直竖条纹自注意力，计算公式为

$$\boldsymbol{h}_k=\begin{cases}\text{H-Attention}_k(\boldsymbol{X}),\ k=1,\cdots,K/2\\ \text{V-Attention}_k(\boldsymbol{X}),\ k=(K/2)+1,\cdots,K\end{cases} \tag{2}$$

最后，将这两个并行组的输出连接，计算公式为

$$\text{CSWin-Attention}(\boldsymbol{X})=\text{Concat}\left(\boldsymbol{h}_1,\cdots,\boldsymbol{h}_K\right)\boldsymbol{W}^O \tag{3}$$

其中：$\boldsymbol{W}^O\in\boldsymbol{R}^{C\times C}$，是常用的投影矩阵，将自注意力结果投射到目标输出维度（默认设置为 C），用来调整特征图的通道数，可以将两个不同方向的自注意力特征融合。将多个头平均分成两个平行组，并相应地用不同的自注意力操作，使模块中的每个区域的注意力作用范围扩大。与之对比，普通的注意力每一个头的自

注意力都是一样的。

1.3　CSWin Transformer Block 模块

模型主体部分的 CSWin Transformer Block 的模型结构为

$$\hat{\boldsymbol{X}}^{l}=\text{CSWin-Attention}\left(\text{LN}\left(\boldsymbol{X}^{l-1}\right)\right)+\boldsymbol{X}^{l-1} \tag{4}$$

$$\boldsymbol{X}^{l}=\text{MLP}\left(\text{LN}\left(\hat{\boldsymbol{X}}^{l}\right)\right)+\hat{\boldsymbol{X}}^{l} \tag{5}$$

在生成图像的过程中，对于第 l 个 Transformer 块生成的特征图 $\boldsymbol{X}^{l}$，式(4)表示对特征图进行层归一化(LN)和 CSWin Self-Attention 操作，然后和输入的特征图拼接，有效扩大了感受野。式(5)表示将特征图进行 LN 和 MLP 操作，与输入的特征图进行特征融合后输出，不仅可以对特征图进行非线性变换和调整，而且可以调整线性层的权重。重新缩放 Transformer 模块内前馈网络(FFN)的权重，从而实现对特征更细粒度的控制和优化[12]。Transformer 模块的整体结构如图 1(a)所示。

1.4　局部全局位置编码

自注意力操作会忽略图像中重要的位置信息。为了将位置信息添加回来，在现有的 Transformer 中已经使用了不同的位置编码，在生成任务中同样需要对像素的相对位置进行编码，提供图片在上下文中的相对位置。相对位置编码是在对像素的相对位置进行编码，已被证明对判别任务至关重要[13]。LePE[5]是在每个 Transformer 模块内部增加的位置，计算方式为

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\boldsymbol{Q}\boldsymbol{K}^{\text{T}}/\sqrt{d}\right)\boldsymbol{V}+\text{DWConv}(\boldsymbol{V}) \tag{6}$$

其中：$\boldsymbol{Q}$ 代表查询向量，$\boldsymbol{K}$ 代表键向量，$\boldsymbol{V}$ 代表值向量。

对 $\boldsymbol{V}$ 进行深度卷积，加到 Softmax 之后的结果上。使用相对位置编码 LePE，一方面可以更好地处理局部位置信息，有利于提高生成图像的质量；另一方面，让生成器知道绝对位置，因为特定组件的合成高度依赖其空间坐标[14]。鉴于此，本文在每个尺度上引入正弦位置编码(SPE)[15]，在每个模块内部执行 AdaIN 操作以后，使用以下编码添加特征图。

$$[\underbrace{\sin(\omega_0 i),\cos(\omega_0 i),\cdots}_{\text{horizontal dimension}},\underbrace{\sin(\omega_0 j),\cos(\omega_0 j),\cdots}_{\text{vertical dimension}}]\in R^{C} \tag{7}$$

其中：$\omega_k=1/10000^{2k}$，(i,j) 表示 2D 位置。

LePE 和 SPE 一起使用，即在每个变换器中应用 LePE 提供本地上下文中的相对位置，而在每个尺度上引入的 SPE 则告知了全局位置。

1.5　模型损失函数

由于 GAN 在训练过程中很容易发生模式坍塌，所以训练不稳定一直是 GAN 最大的问题。本文引入了 Wasserstein 距离[3]和梯度惩罚训练 CTGAN。判别器用于测定图像样本的真实性，而生成器用于生成判别器错误地识别为真实样本的样本，CTGAN 的目标函数为

$$L(D)=-E_{x\sim P_{\text{r}}}\left[D(x)\right]+E_{x\sim P_{\text{g}}}\left[D(x)\right]+\lambda E_{x\sim P_{\tilde{x}}}\left[\|\nabla D(x)\|-1\right]^{2} \tag{8}$$

$$L(G)=-E_{x}\sim P_{\text{g}}\left[D(x)\right] \tag{9}$$

其中：$\tilde{x}=\varepsilon x_{\text{r}}+(1-\varepsilon)x_{\text{g}},x_{\text{r}}\sim P_{\text{r}},x_{\text{g}}\sim P_{\text{g}}$，$\varepsilon\sim$ Uniform [0,1]，P_{r} 代表真实的样本的分布，P_{g} 代表生成的样本的分布，γ 表示惩罚权重，∇ 代表梯度。

2　实　验

2.1　数据集和实验环境

2.1.1　数据集

CIFAR-10 数据集[16]是由 Geoffrey Hinton 收集，包含 60 000 张 32 像素 × 32 像素分辨率的彩色图片，其中有 50 000 张图片为训练集，10 000 张图片为测试集。数据集一共包含 10 个类别：飞机、汽车、鸟类、猫、鹿、狗、蛙类、马、船和卡车。

CelebA 数据集[17]是一个大规模的人脸数据集，包含 20 万张名人的图像。该数据集包含较大的姿势变化和杂乱的背景，包括 10 177 个身份、202 599 张人脸图片，原数据集为 178 像素 × 218 像素分辨率的图片。本文使用对齐和裁剪版本的数据集，并将分辨率大小调整为 64 像素 × 64 像素和 128 像素 × 128 像素。

2.1.2　实验环境

使用 PyTorch 1.1.0 框架、Ubuntu 18.04 操作系统，CPU 为 Xeon(R) Platinum 8255C，GPU 为 RTX 3090，内存大小为 30 GB，Python 3.7，加速环境为 Cuda 10.0。

2.2　评价指标

在图像生成任务中，对生成的图片进行评价需要从两个方面进行考量：一是生成的图片本身的效果，图片是否清晰和完整；二是生成图片的多样性，生成的图片应该与原数据集一样有不同的类别。使用图像生成领域最经典的评价指标 FID(Fréchet inception distance)和 IS(inception score)评估生成的效果。

FID 用来描述两个数据集之间的相似程度，FID 值越小，相似程度越高，说明模型的生成效果越好，计算公式为

$$\mathrm{FID}=\left\|\mu_{x_r}-\mu_{x_g}\right\|^2+\mathrm{tr}\left(\sum x_r+\sum x_g-2\left(\sum x_r\sum x_g\right)^{\frac{1}{2}}\right) \tag{10}$$

其中：x_r 和 x_g 表示真实图像和生成图像，μ_{x_r} 和 μ_{x_g} 表示真实图像和生成图像的均值，$\sum x_r$ 和 $\sum x_g$ 表示真实图像和生成图像的协方差矩阵，tr 代表表示矩阵的迹(矩阵对角元素之和)。

IS 从图像的生成质量和多样性两方面进行评估，IS 值越高，说明图像的多样性和生成质量越好，计算公式为

$$\mathrm{IS}(G)=\exp(E_{x\sim p_g}D_{\mathrm{KL}}(p(y\mid x)\,\|\,p(y))) \tag{11}$$

其中：G 表示生成器，E 表示期望，$x\sim p_g$ 表示 x 是从 p_g 中生成的样本，D_{KL} 表示两分布间的 KL 散度，y 表示合成图像的预测标签。

2.3　在 CIFAR-10 数据集实验结果与分析

为了验证本文模型的有效性，将 CTGAN 模型与现有的生成模型在 CIFAR-10 数据集上进行对比，训练过程中 batch size 大小设置为 64，使用 Adam 优化器训练，其中 $\beta_1=0.9$，$\beta_2=0.99$，设置生成器和判别器的学习率分别为 1×10^{-4} 和 4×10^{-4}，对模型判别器进行可视化分析，如图 2 所示。随着迭代次数的增加，判别器的损失不断下降。

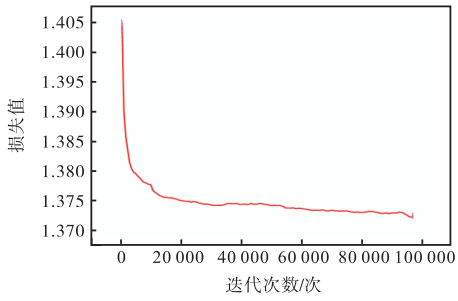

图 2　CIFAR-10 数据集判别器的损失

Fig. 2　Loss of CIFAR-10 dataset discriminator

本文模型和其他模型采用相同的计算方式，计算了训练集和 5 万张生成图像之间的 FID 值，实验结果见表 1。CTGAN 的 FID 指标超过基于 CNN 架构的经典模型 Progressive GAN[18]以及许多其他基于 CNN 的模型，如 SN-GAN[19]、AutoGAN[20] 和 AdversialNAS GAN[21]。本文模型的 FID 比最经典的 StyleGAN-v2[12]低了 4.02，和经过数据增强之后的 StyleGAN-V2+DiffAug[22]相比降低了 2.84，和最新的完全基于 Transformer 的 TransGAN[11]模型相比降低了 2.21。在 IS 指标方面，本文模型超越了之前的模型和最新的 TransGAN[11]，并且接近最经典的数据增强之后的 StyleGAN-V2 + DiffAug。

图 3 为 CTGAN 生成效果的展示图，经过 14 万次迭代之后，可以生成多种清晰且不同风格的分辨率为 32 像素 × 32 像素的彩色图片。

表 1　CIFAR-10 对比实验结果

Tab. 1　CIFAR-10 comparative experimental results

模型	IS	FID
WGAN-GP	6.49 ± 0.09	39.68
AutoGAN	8.55 ± 0.10	12.42
AdversarialNAS-GAN	8.74 ± 0.07	10.87
Progressive-GAN	8.80 ± 0.05	15.52
StyleGAN-V2	9.18	11.07
StyleGAN-V2+DiffAug	9.40	9.89
TransGAN	9.02 ± 0.12	9.26
CTAGN	9.20 ± 0.06	7.05

图 3　CIFAR-10 生成效果展示图

Fig. 3　Display diagram of CIFAR-10 generation effect

2.4　在 CelebA 数据集上的实验结果与分析

在 CelebA(64 × 64)数据集上进行了实验，训练过程中 batch size 设置为 32，迭代 9 000 次之后，计算了 5 万张生成图像和训练集 5 万张图片之间的 FID 分数，实验结果见表 2。

表 2　CelebA(64 × 64)对比实验结果

Tab. 2　CelebA (64 × 64) comparative experimental results

模型	训练集	测试集	FID
PAE	202 599	50 000	49.2
BEGAN-CS	202 599	50 000	34.14
PeerGAN	202 599	50 000	16.97
TransGAN	202 599	50 000	12.23
CTAGN	202 599	50 000	7.80

CTGAN 超越了近年来最经典的生成模型，包括 PAE[23]、BEGAN-CS[24]、PeerGAN[25]以及最新的纯 Transformer 构建的 TransGAN[11]模型。模型生成效果如图 4 所示，生成的人脸具有非常高的自然度和逼真度，并且在训练过程中没有发生模式崩坍，证明了本文引入 Wasserstein 距离和梯度惩罚训练 CTAGN 的有效性。

本文在 CelebA（128 × 128）数据集也进行了实验，batch size 设置为 8，迭代了 20 000 次之后，模型收敛到了最好的效果，生成效果对比如图 5 所示，在较高的分辨率下，CTGAN 生成图片和原数据集相似程度很高，可以合成细节丰富且色彩鲜艳的图片，证明了十字形窗口自注意力机制有助于模型捕捉图片组件之间的长距离依赖关系。

2.5　消融实验

为了验证本文模型的有效性，分别在 CIFAR-10 数据集和 CelebA（64 × 64）数据集上进行消融实验，实验结果见表 3。在基础模型中引入 SPE 和 LePE 可以有效降低 FID，证明 SPE 和 LePE 可以提供图像的局部和全局位置信息，提高生成图像的质量。将特征图进行 LN 和 MLP 操作，与输入的特征图进行特征融合后输出，可以降低 FID，实现对特征的更细粒度控制和优化。

图 4　CelebA（64 × 64）生成效果展示图

Fig. 4　Display diagram of CelebA（64 × 64）generation effect

（a）原数据集图片　　（b）CTGAN生成图片

图 5　CelebA（128 × 128）生成效果可视化对比展示图

Fig. 5　Display diagram of CelebA（128 × 128）generation effect

表 3　消融实验结果

Tab. 3　Results of ablation experiment

数据集	模型	训练集	测试集	FID
CIFAR-10	Base	50 000	50 000	10.08
	+LePE	50 000	50 000	9.06
	+SPE	50 000	50 000	7.44
	+MLP	50 000	50 000	7.05
CelebA（64 × 64）	Base	202 599	50 000	11.30
	+LePE	202 599	50 000	10.02
	+SPE	202 599	50 000	8.16
	+MLP	202 599	50 000	7.80

3　结　语

本文提出了一种基于 CSWin Transformer 的生成对抗网络模型 CTGAN，解决了基于 CNN 构建的 GAN 模型局部感受野难以捕获远距离依赖性导致图片细节丢失的问题。模型的主体部分采用 CSWin Transformer Block 作为基本的构建模块进行生成，同时引入了相对位置编码 LePE 和正弦位置编码 SPE 提供位置信息。在 CIFAR-10 数据集和 CelebA（64 × 64）数据集上测试，模型在定性指标 FID 上超过了很多经典的模型和最新的 TransGAN，在更高分辨率和更复杂的生成场景人脸数据集 CelebA（128 × 128）上，仍然可以合成具有高保真度的图片，证明了 CTGAN 模型具有较好的生成效果。

参考文献：

[1]　KARRAS T，LAINE S，AILA T. A style-based generator architecture for generative adversarial networks[C]// IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. New York：IEEE，2019：04948.

[2]　田平荣. 基于深度学习的管廊视觉异常检测方法研究

[D]. 哈尔滨:哈尔滨工业大学,2021.

[3] ADLER J, LUNZ S. Banach Wasserstein GAN[C]// NIPS. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Denver: NIPS, 2018: 6755–6764.

[4] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[J]. Advances in neural information processing systems, 2016, 29: 2180–2188.

[5] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2794–2802.

[6] BERTHELOT D, SCHUMM T, METZ L. BEGAN: boundary equilibrium generative adversarial networks [EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.1703.10717.

[7] BROCK A, DONAHUE J, SIMONYAN K. Large scale GAN training for high fidelity natural image synthesis [EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.1809.11096.

[8] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: transformers for image recognition at scale[EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.2010.11929.

[9] DONG X, BAO J, CHEN D, et al. CSWin Transformer: a general vision transformer backbone with cross-shaped windows[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 12124–12134.

[10] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]// PMLR. International Conference on Machine Learning. New York: PMLR, 2019: 7354–7363.

[11] JIANG Y, CHANG S, WANG Z. TransGAN: two pure transformers can make one strong GAN, and that can scale up[J]. Advances in neural information processing systems, 2021, 34: 14745–14758.

[12] KARRAS T, LAINE S, AITTALA M, et al. Analyzing and improving the image quality of StyleGAN[C]// IEEE. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE, 2020: 8110–8119.

[13] DAI Z, LIU H, LE Q V, et al. CoAtNet: marrying convolution and attention for all data sizes[J]. Advances in neural information processing systems, 2021, 34: 3965–3977.

[14] LIN C H, CHANG C C, CHEN Y S, et al. Coco-GAN: generation by parts via conditional coordinating[C]// IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 4512–4521.

[15] CHOI J, LEE J, JEONG Y, et al. Toward spatially unbiased generative models[EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.2108.01285.

[16] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[J]. Handbook of systemic autoimmune diseases, 2009, 1(4): 1–60.

[17] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015: 3730–3738.

[18] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.1710.10196.

[19] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [EB/OL]. [2023–03–01]. https://arxiv.org/pdf/1802.05957.pdf.

[20] BUTHGAMUMUDALIGE V U, WIRASINGHA T. Neural architecture search for generative adversarial networks: a review[C]//IEEE. 2021 10th International Conference on Information and Automation for Sustainability. New York: IEEE, 2021: 246–251.

[21] GAO C, CHEN Y, LIU S, et al. AdversarialNAS: adversarial neural architecture search for GANs[C]// IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 5680–5689.

[22] ZHAO S, LIU Z, LIN J, et al. Differentiable augmentation for data-efficient GAN training[J]. Advances in neural information processing systems, 2020, 33: 7559–7570.

[23] BÖHM V, SELJAK U. Probabilistic auto-encoder [EB/OL]. [2023–03–01]. https://doi.org/10.48550/arXiv.2006.05479.

［24］ CHANG C C，LIN C H，LEE C R，et al. Escaping from collapsing modes in a constrained space［C］//IEEE. Proceedings of the European Conference on Computer Vision（ECCV）. New York：IEEE，2018：204–219.

［25］ WEI J，LIU M，LUO J，et al. PeerGAN：generative adversarial networks with a competing peer discriminator［EB/OL］.［2023–03–01］. https://doi.org/10.48550/arXiv.2101.07524.

责任编辑：郎婧

（上接第 27 页）

74.

［13］ ZHENG Y，WANG Q，HUANG J，et al. Hypoglycemic effect of dietary fibers from bamboo shoot shell：an in vitro and in vivo study［J］. Food and chemical toxicology，2019，127：120–126.

［14］ LIU Y，ZHANG H，YI C，et al. Chemical composition，structure，physicochemical and functional properties of rice bran dietary fiber modified by cellulase treatment［J］. Food chemistry，2021，342：128352.

［15］ HABERMEYER M，ROTH A，GUTH S，et al. Nitrate and nitrite in the diet：how to assess their benefit and risk for human health［J］. Molecular nutrition and food research，2015，59（1）：106–128.

［16］ 王旭. 米糠膳食纤维的改性制备及其特性研究［D］. 北京：中国农业大学，2018.

［17］ ZHOU Y，XIE F，ZHOU X，et al. Effects of Maillard reaction on flavor and safety of Chinese traditional food-roast duck［J］. Journal of the science of food and agriculture，2016，96（6）：1915–1922.

［18］ WU L，SUN H，HAO Y，et al. Chemical structure and inhibition on α-glucosidase of the polysaccharides from *Cordyceps militaris* with different developmental stages［J］. International journal of biological macromolecules，2020，148：722–736.

［19］ LIU X，SUO K，WANG P，et al. Modification of wheat bran insoluble and soluble dietary fibers with snail enzyme［J］. Food science and human wellness，2021，10：356–361.

［20］ WEN Y，NIU M，ZHANG B，et al. Structural characteristics and functional properties of rice bran dietary fiber modified by enzymatic and enzyme-micronization treatments［J］. LWT-Food science and technology，2017，75：344–351.

［21］ GE S，WU Y，PENG W，et al. High-pressure CO_2 hydrothermal pretreatment of peanut shells for enzymatic hydrolysis conversion into glucose［J］. Chemical engineering journal，2020，385：123949.

［22］ ULLAH I，YIN T，XIONG S，et al. Structural characteristics and physicochemical properties of okara（soybean residue）insoluble dietary fiber modified by high-energy wet media milling［J］. LWT-Food science and technology，2017，82：15–22.

［23］ SUN C，WU X，CHEN X，et al. Production and characterization of okara dietary fiber produced by fermentation with *Monascus anka*［J］. Food chemistry，2020，316：126243.

［24］ JING Q，YUE L，KINGSLEY G M，et al. The effect of chemical treatment on the in vitro hypoglycemic properties of rice bran insoluble dietary fiber［J］. Food hydrocolloids，2016，52：699–706.

［25］ REN F，FENG Y，ZHANG H，et al. Effects of modification methods on microstructural and physicochemical characteristics of defatted rice bran dietary fiber［J］. LWT-Food science and technology，2021，151：112161.

［26］ BRIONES-LABARCA V，MUOZ C，MAUREIRA H. Effect of high hydrostatic pressure on antioxidant capacity，mineral and starch bio-accessibility of a nonconventional food：*Prosopis chilensis* seed［J］. Food research international，2011，44（4）：875–883.

责任编辑：郎婧