ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 5,Sep. 2025,pp. 1306－1321
DOI:10. 16337/j. 1004-9037. 2025. 05. 016

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于DID-AugGAN的小样本缺陷图像生成与数据增强算法

黄绿娥[1,2]，邓亚峰[1]，鄢化彪[3]，肖文祥[1]

(1. 江西理工大学电气工程与自动化学院,赣州 341000;2. 多维智能感知与控制江西省重点实验室,赣州 341000;3. 江西理工大学理学院,赣州 341000;)

摘　要：针对小样本条件下生成对抗网络(Generative adversarial network，GAN)生成缺陷图像质量低、不真实且多样性差的问题,提出一种缺陷图像生成算法(Defect image data augmentation GAN，DID-AugGAN),旨在实现小样本缺陷图像的数据增强。为解决传统卷积在有限数据集中难以有效学习图像中非刚性特征的问题,设计可学习偏移卷积,以提高模型对图像语义信息的学习能力;为避免关键缺陷特征丢失,提升局部特征之间的关联性,设计多尺度坐标注意力模块,重点关注缺陷位置信息;为提高网络对输入图像局部信息的判别能力,重新设计判别器网络架构,使其从传统的单一前馈网络转变为包含对称编码与解码路径的UNet-like结构;将DID-AugGAN与原算法在Rail-4c轨道扣件缺陷数据集上进行对比实验,并利用分类网络MobileNetV3进行验证。实验结果表明,改进后的方法显著提高了IS(Inception score),有效降低了FID(Fréchet inception distance)和LPIPS(Learned perceptual image patch similarity)指标,并且MobileNetV3分类准确率和F_1分数也得到提高。该算法能稳定生成高质量的缺陷图像,有效扩充缺陷数据样本,满足下游任务需求。

关键词：小样本学习;生成对抗网络;可学习偏移卷积;多尺度坐标注意力;UNet-like
中图分类号：TP391　　**文献标志码:**A

A Few-Shot Learning Algorithm for Defect Image Generation and Data Augmentation Based on DID-AugGAN

HUANG Lve[1,2], DENG Yafeng[1], YAN Huabiao[3], XIAO Wenxiang[1]

(1. School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China; 2. Jiangxi Province Key Laboratory of Multidimensional Intelligent Perception and Control, Ganzhou 341000, China; 3. School of Science, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract: To address the issues of low quality, lack of realism, and poor diversity in defect images generated by generative adversarial network (GAN) under small-sample conditions, this paper proposes a defect image generation algorithm, named defect image data augmentation GAN (DID-AugGAN), aiming at enhancing defect image data under limited sample conditions. First, to overcome the difficulty of traditional convolutional networks in effectively learning non-rigid features in images from limited datasets,

基金项目:国家自然科学基金(62363013);江西省自然科学基金(20224BAB202036);江西省教育厅科学技术重点研究项目(GJJ2200805);江西理工大学研究生创新计划资助项目(XY2023-S159)。
收稿日期:2024-07-13;**修订日期:**2024-09-22

we design a learnable offset convolution to improve the model's capability in capturing semantic information. Second, to prevent the loss of critical defect features and enhance the correlation among local features, we introduce a multi-scale coordinate attention module, which focuses on defect location information. Third, to enhance the discriminator's ability to distinguish local details in input images, we redesign its architecture, transforming it from a conventional feedforward network into a UNet-like structure with symmetric encoding and decoding pathways. Finally, we conduct comparative experiments between DID-AugGAN and the baseline algorithm on the Rail-4c track fastener defect dataset, and validate the generated images using the MobileNetV3 classification network. Experimental results demonstrate that the proposed method significantly improves inception score (IS) while effectively reducing Fréchet inception distance (FID) and learned perceptual image patch similarity (LPIPS). Moreover, the classification accuracy and F_1-score of MobileNetV3 are also improved. The proposed DID-AugGAN can stably generate high-quality defect images, effectively augment defect data samples, and meet the requirements of downstream tasks.

Key words: few-shot learning; generative adversarial network (GAN); learnable offset convolution (LOConv); multi-scale coordinate attention (MSCA); UNet-like

引 言

近年来,随着计算机视觉和深度学习技术的快速进步,基于图像的缺陷检测[1]和分类技术[2]引起了广泛关注,这些技术在制造业、交通运输和基础设施维护等行业展示出巨大的潜力。然而,要实现高精度的缺陷检测和分类往往需要庞大的缺陷数据集,但在许多实际应用中,缺陷图像数据却十分稀缺。这种稀缺性主要源于正常生产和运行过程中故障发生的概率较低,人工采集缺陷样本数据的难度很大,获取的缺陷图像种类也相对有限。因此,目前的研究面临一个重大挑战:缺陷图像样本不足会导致模型在训练过程中容易出现过拟合,严重影响缺陷检测和分类技术的智能化发展[3]。

为了解决缺陷样本稀缺问题,数据增强技术应运而生[4]。传统的数据增强方法包括对现有图像进行裁剪、旋转、像素插值、添加噪声等处理[5-7],以扩充已有数据集。这些方法的局限性是仅能在现有缺陷特征内进行变换,无法构造新特征,也不能充分表达具有复杂纹理背景的样本特征。生成对抗网络(Generative adversarial network,GAN)[8]作为一种强大的数据生成工具,为数据增强提供新思路和新方法,但GAN在训练时仍存在模式崩塌、梯度消失、爆炸等问题。在此基础之上,Radford等[9]提出使用深度卷积GAN(Deep convolutional GAN, DCGAN),采用转置卷积替代传统的上采样方法,并取消池化层,从而增强网络结构的稳定性,然而,DCGAN并未完全解决梯度消失问题。Arjovsky等[10]提出WGAN(Wasserstein GAN),通过引入Wasserstein距离替代GAN网络中的JS距离,成功克服梯度消失的难题。Gulrajani等[11]进一步提出WGAN-GP,以梯度惩罚替代传统的权重裁剪方式,取得更为出色的训练效果。此外,Karras等[12]提出的StyleGAN2-ada通过引入自适应判别器增强技术,有效缓解训练过程中的过拟合问题。学者们正积极采用GAN及其变种进行数据增强,不仅丰富了相敏光时域反射仪(Φ-OTDR)研究中的信号数据[13],而且在扩充医学图像[14]、红外图像[15]以及缺陷图像数据[16-19]等领域也展现出其独特优势。

尽管GAN在这些领域取得了一些进展,但仍然主要依赖于大量数据样本,在小样本条件下生成缺陷图像仍然面临诸多挑战。小样本数据增强是在训练数据极其有限的情况下训练模型生成新数据扩充训练集,由于在现实条件下,很多领域存在样本数量极少或不易获取的情况,小样本数据增强具有极高的现实需求和意义[20]。尽管有学者在小样本生成领域提出了一些算法,如文献[21]提出了AGE

(Attribute group editing)方法,旨在通过编辑类别无关属性,并且不需要重新训练GAN模型,只需利用少量样本模型就能达到较好的生成效果。文献[22]提出了基于扩散模型的少样本异常图像生成模型AnomalyDiffusion,通过嵌入自适应注意力权重机制,提高了异常图像的生成质量,从而增强下游异常检测任务的效果。文献[23]提出了基于条件变分自编码器(Conditional variational autoencoder, CVAE)的少样本学习策略,通过在基础类别上训练CVAE模型,同时移除基础训练集中的非代表性样本来指导CVAE严格生成代表性样本,增强了生成样本的代表性以提高少样本分类任务的性能。文献[24]提出了DCL (Dual contrastive learning)方法,通过利用最大化源域和目标域生成器之间的信息,减缓预训练的生成对抗网络在小样本数据集上生成图像多样性的退化,在保证生成质量的同时提高小样本图像生成的性能。但这些算法大多依赖于大规模的预训练数据集来初始化其模型参数,这在许多实际应用场景中并不总是可行的,尤其是在针对特定领域(如缺陷图像生成)时,其中缺陷图像通常包含多种缺陷类型,且这些缺陷的形状复杂多样,加之样本数量有限,导致GAN模型生成的缺陷图像往往质量不高、不够真实且缺乏多样性,这使得其难以满足后续图像处理任务的需求。因此,迫切需要一种能够直接在小样本条件下就能够生成高质量、多样化缺陷图像的生成算法。

为解决上述问题,本文提出一种DID-AugGAN(Defect image data augmentation GAN)算法用于小样本缺陷图像的数据增强。主要贡献如下:

(1) 针对(如DCGAN[9]和WGAN-GP[11])传统卷积受限于固定的卷积核形状,难以在有限的数据集中捕捉图像中的非刚性特征的问题,设计可学习偏移卷积(Learnable offset convolution,LOConv)。该卷积通过动态调整卷积核偏移量,提升非刚性特征的捕捉能力,更好地学习图像语义信息,相较于文献[24]等的DCL方法减少了预训练依赖;

(2) 针对传统坐标注意力(Coordinate attention,CA)[25]因单尺度池化导致特征表达能力不足的问题,设计多尺度坐标注意力(Multi-scale CA,MSCA)模块,通过多尺度池化保留不同层次的缺陷细节,避免关键缺陷特征丢失,提升局部特征之间的关联性,较文献[22]的注意力机制更适应复杂纹理背景;

(3) 针对传统单一前馈结构的判别网络(如StyleGAN2-ada[12])对局部特征判别能力不足的问题,设计包含对称编码与解码路径的UNet-like判别器,以提升其对局部信息的判别能力,进而引导生成器提高图像生成质量,较文献[21]的AGE方法更直接提升生成图像的真实性。

1 DID-AugGAN算法

1.1 整体网络框架

为解决生成对抗网络在小样本条件下生成的缺陷图像质量较低、不够真实且多样性较差的问题,本文在StyleGAN2-ada模型的基础上,设计一种全新的缺陷图像生成模型:DID-AugGAN,其结构如图1所示。

DID-AugGAN网络由生成器和判别器两部分组成,生成器的任务是生成逼真的假图像,以欺骗判别器,而判别器负责准确地判断出输入图像是伪造的还是来自真实图像。

生成器网络由负责生成样式(Style)向量的样式映射网络和负责逐步增加图像分辨率的多个块组成,每个块(如Block 4×4)表示该块内的所有特征图分辨率为4×4,分辨率的增加通过上采样层(Upsample)实现。低分辨率层(如Block 4×4和Block 8×8)主要负责构建图像的基本结构和提取低级特征。在这些块的低分辨率风格化G-1层中,设计LOConv替代传统卷积(Conv)。LOConv能够动态调整卷积核形状,更好地适应缺陷图像特有的形状特征。这种调整使得模型在捕捉图像中的非刚性特征时更加精确,提升了其处理复杂图像信息的能力。然后,特征图经过Block 16×16和Block 32×32的上采样Upsample层和中分辨率风格化G-2层处理后,最终输入到Block 64×64中,生成64×64分辨率

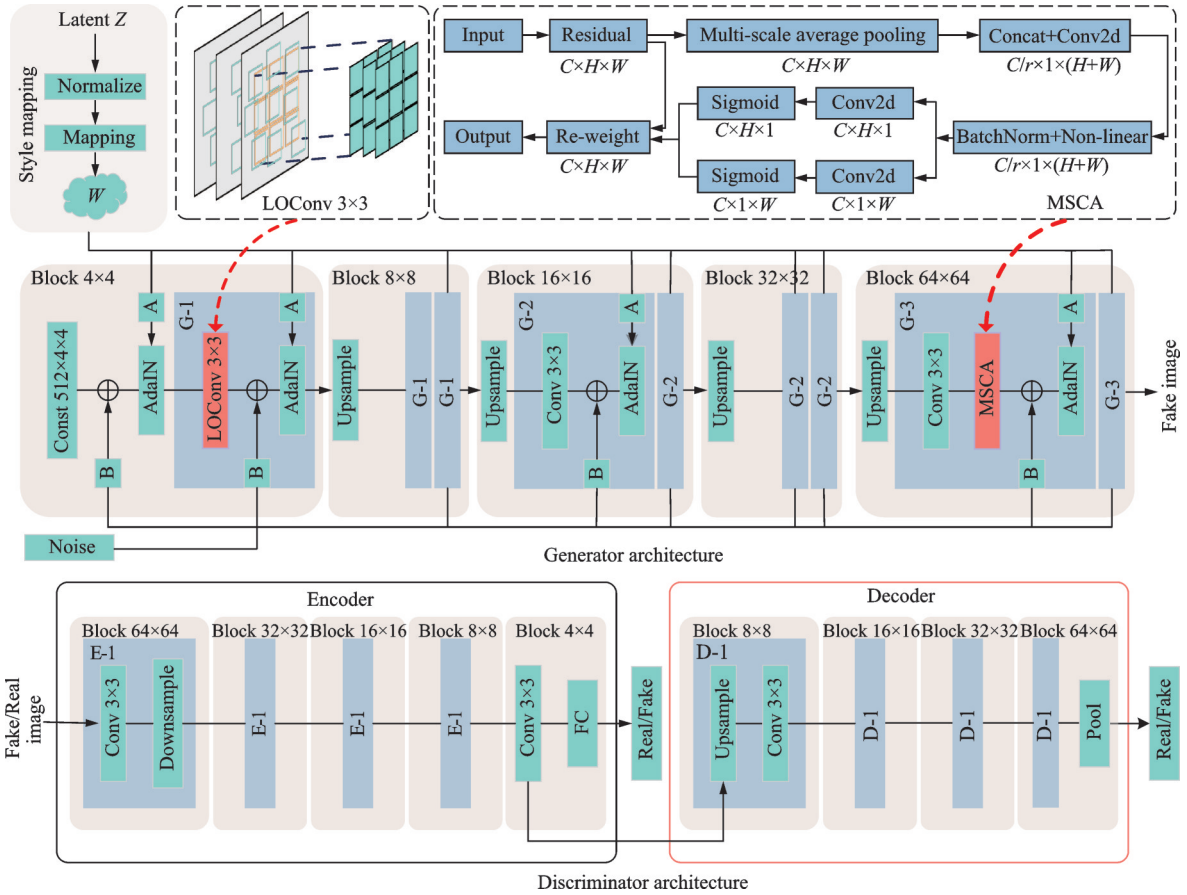


图1 DID-AugGAN生成器和判别器网络结构图

Fig.1 Network architecture diagram of the generator and discriminator for DID-AugGAN

的假图像。随着判别器网络层数的增加,为了避免在特征提取过程中丢失关键缺陷特征,在Block 64×64的高分辨率风格化G-3层中,设计MSCA模块嵌入到Conv层和实例归一化层(AdaIN)之间。MSCA模块能忽略不重要的信息,专注于学习缺陷图像的细节纹理,重点关注缺陷位置信息,提升局部特征之间的关联性,从而避免在生成图像时丢失关键缺陷特征,确保生成图像的质量。

判别器被调整为UNet-like结构,该结构由编码器和解码器两部分组成。在编码器部分,输入图像逐层(从Block 64×64到Block 8×8)通过多个编码块进行处理,每个编码块包含Conv层和下采样层(Downsample)。然后,特征图通过Block 4×4的全连接层(FC)输出整个图像的全局决策。在解码器部分,特征图从编码器的Block 4×4的Conv层输出,经过多个解码块(从Block 8×8到Block 64×64)解码,每个解码块包含Conv层和Upsample层,最后通过池化层(Pool)输出图像每个像素的局部决策。该判别器结构能够增强对输入图像局部信息的判别能力,从而鼓励生成器提升图像生成质量。

1.2 可学习偏移卷积设计

缺陷图像中存在大量非刚性特征,这些特征不同于刚性特征,会在刚性变换后发生形变。原模型的生成器结构中大量使用传统卷积,但由于传统卷积受限于固定的卷积核形状,在处理图像中不规则分布的非刚性特征时效果较差。尽管理论上可以通过堆叠多个卷积层并积累足够大的感受野来缓解该问题,但这样会导致模型复杂度骤增,并可能使网络不易收敛,在训练样本规模受限时,也极易出现

过拟合现象。为更精确地捕捉图像中的非刚性特征并提升模型处理复杂图像信息的能力，设计LOConv替换原模型中的传统卷积，进而更好地学习图像的语义信息，传统卷积与LOConv的区别如图2所示。

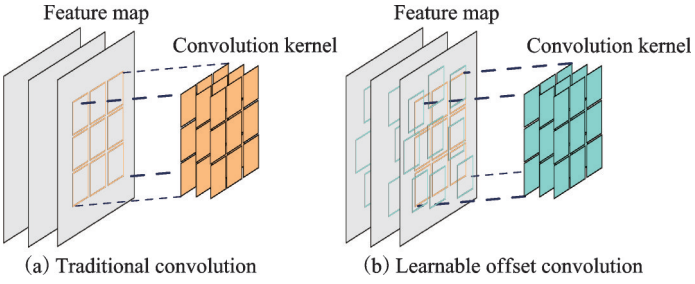


图2 传统卷积与可学习偏移卷积的区别

Fig.2 Difference between traditional convolution and learnable offset convolution

与传统卷积相比，LOConv带有可学习的偏移量，可以根据输入数据调整卷积核形状，完美解决传统卷积受限于固定卷积核形状的问题。具体来说，假设被卷积的中间特征为x，相应卷积后的输出为y，分别用$x(p)$和$y(p)$表示位置p处的特征。以3×3的卷积核为例，传统卷积操作是以位置p为中心，上下左右以及4个对角线一共9个位置采样特征来加权求和，在数学上可表示为

$$y(p)=\sum_{i=1}^{9} w_i \cdot x(p+p_i) \quad p_i \in [-1,0,1]^2 \tag{1}$$

式中：w_i为卷积核的第i个权重；p_i为固定的空间偏移。而LOConv的实现方式是在固定的空间偏移p_i上引入一个可学习的偏移量Δp。为了使加入的可学习偏移量更加依赖于训练数据，可以通过添加额外的卷积来预测偏移量，将额外卷积的输出维数设置为18，来实现预测18个标量以形成9对空间偏移量。而在StyleGAN2-ada中存在潜在向量z和调制卷积(ModConv)[12]，与以各种图像作为输入的学习任务不同，生成任务倾向于通过从高斯分布随机采样的潜在向量z作为输入，生成逼真的图像。由于相应的潜在向量z包含生成输出图像的所有语义信息，这能够确保预测出的偏移量Δp能够更精确地与特征对匹配。因此，LOConv通过利用潜在向量z和调制卷积来精确预测偏移量Δp，其计算式为

$$y(p)=\sum_{i=1}^{9} w_i \cdot x(p+p_i+\Delta p_i) \quad p_i \in [-1,0,1]^2 \tag{2}$$

$$\Delta p = \mathrm{ModConv}(x, z) \tag{3}$$

式中：Δp_i表示第i个位置的可学习偏移量；ModConv表示调制卷积，该卷积允许输入潜在向量z通过与卷积权重相乘(即$\boldsymbol{w}' = w \times z$)来动态调整权重。

1.3 多尺度坐标注意力模块设计

注意力机制在生成图像领域取得了广泛的应用，Hou等[25]提出的CA机制可以捕捉通道和空间位置之间的相关性，使模型更精确地关注图像中的重要区域，忽略不相关的信息。CA模块中特征增强主要通过在水平(X)和垂直(Y)两个坐标方向上对各通道进行全局平均池化来实现。然而，这种简单将特征压缩到一个平均值上的聚合方式，会过于强调正则化效果，从而忽略原始特征的结构信息，导致特征表达能力严重损失，进而在特征学习过程造成缺陷细节丢失的问题。为解决该问题，在CA模块的X方向和Y方向上引入不同尺度平均池化，设计出MSCA模块，MSCA在结构上有利于维持特征的丰富表示，以避免过度规范化对特征学习的不利影响，并且通过捕捉特征的内部相关性来减少对外部信息的依赖，其网络结构如图3所示。

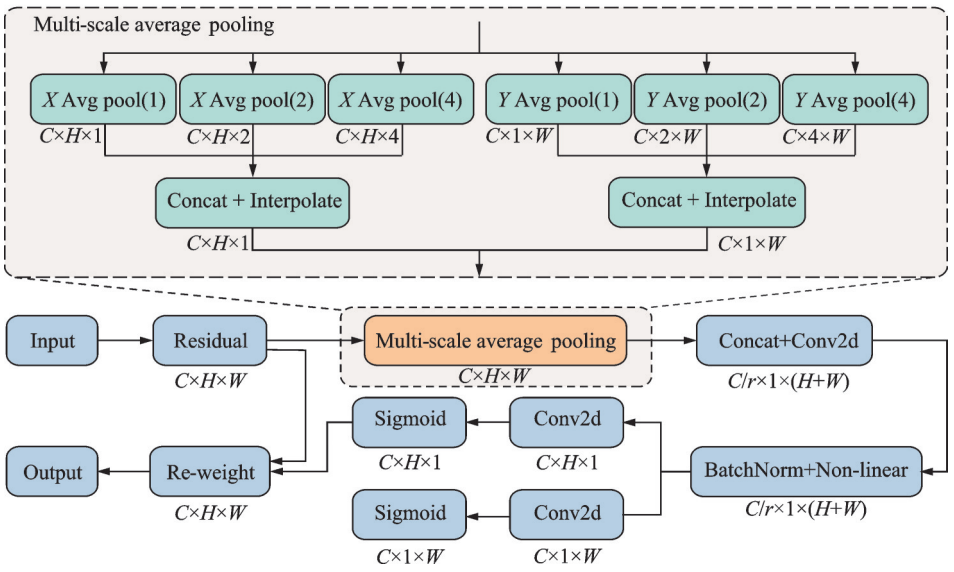


图3 多尺度坐标注意力模块

Fig.3 Multi-scale coordinate attention module

具体来说,在MSCA中,首先对输入的特征图x在X和Y两个坐标方向的各个通道进行多尺度平均池化操作,在两个方向均得到3个不同尺度的特征图。多尺度平均池化操作公式为

$$z_{ci}^{h}=\frac{i}{W}\sum_{0\leqslant j\leqslant W}x_c(H,j)\quad i\in[1,2,4] \tag{4}$$

$$z_{ci}^{w}=\frac{i}{H}\sum_{0\leqslant j\leqslant H}x_c(j,W)\quad i\in[1,2,4] \tag{5}$$

式中:c、h和w分别表示特征图的通道数、高度和宽度;i表示池化尺度。在两个坐标方向对各个通道平均池化为1是为了保留全局信息,确保全局上下文的联系。平均池化为2是为了捕捉中等尺度的特征信息,兼顾全局和局部特征,增强特征的多样性。平均池化为4目标是提取更加细粒度的局部特征,进一步细化特征表示,提升模型对细节的感知能力。然后将多尺度池化后的特征图在各自方向上进行拼接,并通过最近邻插值操作得到X和Y方向的特征图x^h和x^w,其数学表达式为

$$x^h=\mathrm{NNI}(\mathrm{Cat}[z_{c1}^{h},z_{c2}^{h},z_{c4}^{h}])\quad x^h\in\mathbf{R}^{C\times H\times 1} \tag{6}$$

$$x^w=\mathrm{NNI}(\mathrm{Cat}[z_{c1}^{w},z_{c2}^{w},z_{c4}^{w}])\quad x^w\in\mathbf{R}^{C\times 1\times W} \tag{7}$$

式中:Cat表示拼接操作;NNI表示最近邻插值操作。接着将得到的两个特征图进行拼接和1×1卷积操作,再经过批归一化和非线性激活函数处理得到特征图f,将f沿着空间维度分割为两个独立的特征图f^h和f^w,之后经过1×1卷积和Sigmoid操作得到注意力权重g^h和g^w,最后将输入特征图与注意力权重相乘得到输出特征图。其具体数学表达式如下

$$f=\delta(\mathrm{BN}(\mathrm{Con}(\mathrm{Cat}[x^h,x^w])))\quad f\in\mathbf{R}^{C/r\times(H+W)} \tag{8}$$

$$g^h=S(\mathrm{Con}(f^h))\quad g^h\in\mathbf{R}^{C\times H\times 1} \tag{9}$$

$$g^w=S(\mathrm{Con}(f^w))\quad g^w\in\mathbf{R}^{C\times 1\times W} \tag{10}$$

$$y_c(i,j)=x_c(i,j)*g_c^h(i)*g_c^w(j)\quad y_c\in\mathbf{R}^{C\times H\times W} \tag{11}$$

式中:Con表示1×1卷积操作;BN表示BatchNorm操作;δ表示ReLU激活函数;r表示通道大小的缩减率;S表示Sigmoid操作。

1.4 改进的判别器网络架构

由于原StyleGAN2-ada模型的判别器只能对整个图像进行全局决策，缺乏对每个像素的局部决策能力。为提升判别器的判别能力，本文重新设计判别器网络架构，将其从传统的单一前馈网络改进为包含对称编码与解码双路径架构，将其命名为UNet-like结构，使其能够同时权衡全局和局部信息做出判断，改进后的判别器网络结构如图4所示。

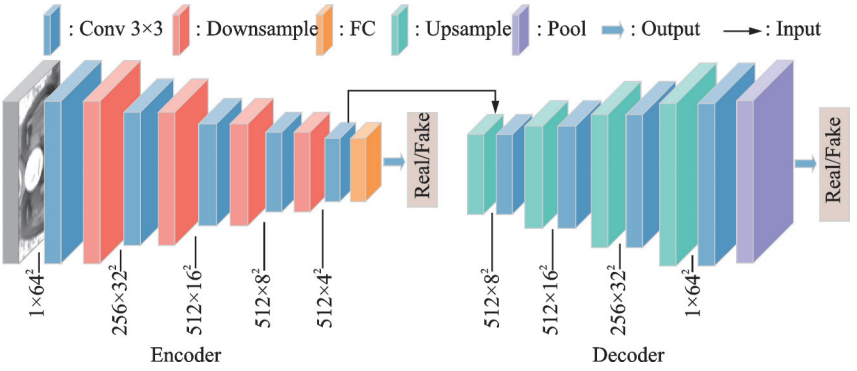


图4 改进后的判别器网络结构

Fig.4 Improved discriminator network structure

UNet-like架构的判别器分为编码器和解码器两部分。在编码器部分，仍采用StyleGAN2-ada的判别网络，由一系列卷积层和下采样层组成，负责逐层降低输入图像的分辨率，并同时增加特征图的通道数，以提取图像的高级语义特征。通过这种逐步抽象的过程，编码器能够捕捉到图像中不同层次的细节信息，包括边缘、纹理以及更复杂的结构特征执行全局决策，并通过FC层输出决策结果。解码器部分则与编码器相反，类似于生成器结构，以编码器末端的输出作为输入，通过多个上采样层和卷积层逐步增加特征图的分辨率，并恢复图像的细节信息，最后通过Pool层输出每个像素的局部决策结果。这种对称编码与解码双路径架构提升了判别器对输入图像局部信息的敏感度和判别能力，既鼓励生成器生成细节更丰富的图像，也使得生成的图像更难以欺骗判别器，从而提高生成样本的质量，最后通过改进损失函数为生成器提供反馈。

原模型的生成器和判别器的损失函数分别为

$$L_G=-E_{z\sim p_z(z)}\log D[G(z)] \tag{12}$$

$$L_D=-E_{x\sim p_{\text{data}}(x)}[\log D(x)]-E_{z\sim p_z(z)}\{\log\{1-D[G(z)]\}\} \tag{13}$$

由于改进后的判别器结构比改进前增加了一个解码器网络，因此改进后的损失函数也增加了一项用于衡量解码器对输入图像像素判断的损失，最终，改进后的生成器和判别器的损失函数为

$$L_G=-E_{z\sim p_z(z)}\{\log D_{\text{enc}}[G(z)]+\sum_{i=1}^{W}\sum_{j=1}^{H}\log D_{\text{dec}}[G(z)]_{ij}\} \tag{14}$$

$$L_{D_{\text{enc}}}=-E_{x\sim p_{\text{data}}(x)}[\log D_{\text{enc}}(x)]-E_{z\sim p_z(z)}\{\log\{1-D_{\text{enc}}[G(z)]\}\} \tag{15}$$

$$L_{D_{\text{dec}}}=-E_{x\sim p_{\text{data}}(x)}\{\sum_{i=1}^{W}\sum_{j=1}^{H}\log D_{\text{dec}}(x)_{ij}\}-E_{z\sim p_z(z)}\left\{\sum_{i=1}^{W}\sum_{j=1}^{H}\log\{1-D_{\text{dec}}[G(z)]_{ij}\}\right\} \tag{16}$$

$$L_D=L_{D_{\text{enc}}}+L_{D_{\text{dec}}} \tag{17}$$

式中：z表示噪声；p表示概率分布；E表示期望；G表示生成器；D_{enc}和D_{dec}分别表示判别器的编码器与解码器；i和j为判别模型解码器D_{dec}输出图像的像素位置，分别代表行和列。

2 算法有效性分析

本节旨在从理论机制上对DID-AugGAN中各改进模块的有效性进行深入分析,并通过新旧方法的对比与公式推导展示改进方案在捕捉复杂缺陷特征、保留不同层次细节和增强局部判别能力等方面的优势。下面分别从动态采样、多尺度特征融合及全局-局部判别3个方面进行分析。

2.1 动态采样与有效感受野扩展的理论推导

为便于后续公式推导,将式(1)和式(2)改写为一般化形式,即分别为式(18)和式(19)。式(1)和式(2)分别对应式(18)和式(19)在采样区域$R=\{p_i: p_i\in[-1,0,1]^2\}$(即3×3卷积核)情况下的特例,其中取$w(p_n)=w_i$。因此,式(1)和式(2)可以分别看作式(18)和式(19)在3×3卷积核条件下的特殊表达形式。

在一般化的表达式中,传统卷积层的输出由固定采样位置决定,其公式为

$$y(p_0)=\sum_{p_n\in R} w(p_n)x(p_0+p_n) \tag{18}$$

式中:R为预定义的采样位置集合,权重也表示为一个关于采样位置的函数$w(p_n)$。由于采样位置固定,传统卷积的有效感受野(Effective receptive field, ERF)仅由卷积核尺寸及网络层数决定,难以适应非刚性缺陷特征的空间变化。

而在LOConv中,通过引入动态偏移Δp_n(由潜在向量z预测),其输出变为

$$y(p_0)=\sum_{p_n\in R} w(p_n)x(p_0+p_n+\Delta p_n) \tag{19}$$

假设Δp_n服从零均值、方差为σ^2的分布,则采样位置的不确定性使得实际采样位置的分布具有更大的方差,从而有效地扩展了感受野。定义传统卷积的感受野为$\mathrm{ERF}_{\mathrm{Conv}}$,则近似有

$$\mathrm{ERF}_{\mathrm{LOConv}}\approx \mathrm{ERF}_{\mathrm{Conv}}+k\sigma \tag{20}$$

式中:k为与卷积核形状和网络结构相关的常数。式(20)表明,动态偏移机制能够使卷积层在捕捉远距离依赖时具有更大的灵活性,从而更好地提取因形变引起的非刚性特征。

为进一步理解这种动态采样的优势,考虑对LOConv中输出对偏移量的梯度进行分析。令损失函数为L,则对Δp_n的梯度可表示为

$$\frac{\partial L}{\partial \Delta p_n}=w(p_n)\frac{\partial x(p_0+p_n+\Delta p_n)}{\partial \Delta p_n} \tag{21}$$

对$x(p_0+p_n+\Delta p_n)$泰勒近似展开,得到

$$x(p_0+p_n+\Delta p_n)\approx x(p_0+p_n)+\Delta p_n\cdot\nabla x(p_0+p_n) \tag{22}$$

则有

$$\frac{\partial L}{\partial \Delta p_n}\approx w(p_n)\cdot\nabla x(p_0+p_n) \tag{23}$$

式(23)表明,通过动态偏移,额外的梯度信号直接引导偏移量的更新,从而使网络在训练过程中能够更灵敏地捕捉因非刚性特征引起的局部结构变化。

2.2 多尺度特征融合与局部细节保留的理论推导

传统的单尺度注意力机制(如传统坐标注意力)仅依赖于全局平均池化,容易导致关键信息因全局平均化而被抹平,进而丢失关键特征。为解决这一问题,MSCA模块通过在X、Y两个方向上分别采用多尺度池化(例如尺度$s=1,2,4$)提取特征,获得不同层次的特征信息,并进行拼接融合。

设在尺度s下提取到的特征为$F^{(s)}$,通过平均池化获得的局部特征表达为

$$F^{(s)}=\frac{1}{N_s}\sum_{i=1}^{N_s} x_i^{(s)} \tag{24}$$

式中:N_s为尺度s下参与池化的像素数量,$x_i^{(s)}$表示对应像素的局部特征。不同尺度对应的信息粒度不

同:尺度$s=1$侧重全局信息,而尺度$s=4$更关注细粒度局部特征。将各尺度特征通过拼接操作融合,可得到综合特征表示

$$F_{\text{agg}}=\text{Concat}\{F^{(1)},F^{(2)},F^{(4)}\} \tag{25}$$

为说明融合后的特征如何更好地保留关键信息,从统计角度考察融合前后信号的方差。假设各尺度特征的均值与方差分别为μ_s和σ_s^2,并引入权重β_s(满足$\sum_s \beta_s=1$)来反映各尺度在融合中的贡献,则融合后特征的方差可写为

$$\sigma_{\text{agg}}^2=\sum_s \beta_s^2\sigma_s^2+\sum_{s\neq s'}\beta_s\beta_{s'}\text{Cov}(F^{(s)},F^{(s')}) \tag{26}$$

当各尺度特征在捕捉关键信息方面具有互补性时,其协方差项较小,融合后的方差能够较好地保留各尺度中的有效信息,而不会因单一尺度的局限性而丢失细节。

接下来,通过1×1卷积及非线性激活函数,对融合后的特征进行变换,得到注意力权重A,即

$$A=\sigma\left(\boldsymbol{W}_1\cdot\text{ReLU}(\boldsymbol{W}_0\cdot F_{\text{agg}})\right) \tag{27}$$

式中:$\boldsymbol{W}_0$和$\boldsymbol{W}_1$为可训练参数矩阵,$\sigma(\cdot)$表示Sigmoid激活函数。得到的注意力权重A能自适应地调节各局部区域的重要性,使得关键缺陷特征在后续处理中得到强化。

2.3 全局与局部判别信息协同的理论推导

传统GAN中判别器通常采用单一前馈结构,其损失函数仅反映了全局判别信息。这种设计在数据样本有限的场景下,容易忽略图像中细小、关键的局部缺陷信息,从而为生成器提供的反馈不足以引导其生成高质量细节。在本文工作中,改进后的判别器采用了UNet-like架构,设计上在保持全局判别能力的同时,引入了解码器路径,通过局部判别损失对图像细节进行监督。判别器改进前后损失函数分别为式(13)和式(17),为便于推导,本节将其简化为L_D^{pre}和式(28)。

$$L_D=L_D^{\text{enc}}+L_D^{\text{dec}} \tag{28}$$

式中:L_D^{enc}为编码器部分(即全局判别)的损失,L_D^{dec}为解码器部分(即局部判别)的损失。

为了从理论上说明这一改进的有效性,从梯度传递的角度进行分析。设判别器整体参数为θ,则改进后的梯度更新可表示为

$$\nabla_\theta L_D=\nabla_\theta L_D^{\text{enc}}+\nabla_\theta L_D^{\text{dec}} \tag{29}$$

在传统判别器中,仅存在$\nabla_\theta L_D^{\text{enc}}$项,这意味着生成器得到的梯度反馈主要来源于图像的整体(全局)信息。当图像存在细微的局部缺陷时,全局判别往往难以捕捉这些细节,从而导致生成器在局部细节上的优化不足。

引入$\nabla_\theta L_D^{\text{dec}}$项后,判别器在局部区域上能够提供更细粒度的监督。具体来说,可以认为解码器部分输出的局部判别结果对应于图像中每个局部区域的真实与否判断,其误差梯度可近似描述为

$$\delta_{\text{local}}=\frac{\partial L_D^{\text{dec}}}{\partial x_{\text{local}}} \tag{30}$$

式中:x_{local}表示图像中某一局部区域的特征表示。相较于仅依赖全局特征产生的梯度

$$\delta_{\text{global}}=\frac{\partial L_D^{\text{enc}}}{\partial x_{\text{global}}} \tag{31}$$

局部梯度δ_{local}能更精确地反映细粒度信息。故整体梯度更新为

$$\nabla_\theta L_D=\delta_{\text{global}}+\delta_{\text{local}} \tag{32}$$

这一组合使得生成器在训练过程中,不仅能够获得关于整体图像真实性的反馈,还能获得关于局部细节表现的指导。换而言之,通过额外的局部判别损失,生成器在梯度更新时会更关注细节处的误差,从而在生成过程中提高对局部缺陷特征的刻画能力。这从理论上说明了改进后的判别器能为生成器提供更丰富、更细致的反馈,进而有助于生成图像质量的提升。

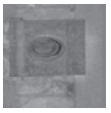
3 实验结果及分析

3.1 实验环境与数据集

实验环境与参数设置如下:操作系统为Ubuntu 20.04,内核为Linux,深度学习框架为PyTorch1.8.1,CPU为Intel(R) Xeon(R) Gold 6248R,GPU为NVIDIA Tesla A800,优化器为Adam,初始学习率为0.002,batchsize为32,kimg设置为1 000。本次实验使用的数据集为自制的Rail-4c轨道扣件缺陷数据集,包括无缺陷、扣件断裂、扣件移位和扣件缺失4类,每类各400张图像,分辨率为64×64,全为单通道灰度图。每类数据集被划分为300张图像的训练集和100张图像的测试集。训练集用于DID-AugGAN生成模型和MobileNetV3[26]分类模型的训练,测试集用于评估MobileNetV3分类模型的性能,详细的数据集构成如表1所示。

表1 实验样本数据集展示

Table 1 Presentation of experimental sample datasets

数据集类型	无缺陷	断裂	移位	缺失
图像展示				
训练集数量	300	300	300	300
测试集数量	100	100	100	100

3.2 评价指标

为评估生成图像质量,选用IS(Inception score)[27]、FID(Fréchet inception distance)[28]和LPIPS(Learned perceptual image patch similarity)[29]三种评价指标。IS用来评价生成图像的清晰度和多样性,值越大越好,表达式为

$$\mathrm{IS}(G)=\exp\left\{E_{x\sim p_g}D_{\mathrm{KL}}[P(y|x)\|P(y)]\right\} \tag{33}$$

式中:$x\sim p_g$表示生成器生成图像;D_{KL}表示对$P(y|x)$和$P(y)$求KL散度。FID用于评价生成的图像与真实图像在特征空间中的分布相似程度,值越小越好,表达式为

$$\mathrm{FID}(X,Y)=\|\boldsymbol{\mu}_X-\boldsymbol{\mu}_Y\|^2+\mathrm{tr}(\boldsymbol{C}_X+\boldsymbol{C}_Y-2\sqrt{\boldsymbol{C}_X\boldsymbol{C}_Y}) \tag{34}$$

式中:X和Y分别表示生成图像和真实图像的特征向量的数据分布;$\boldsymbol{\mu}_X$和$\boldsymbol{\mu}_Y$分别表示X和Y的特征向量均值;$\boldsymbol{C}_X$和$\boldsymbol{C}_Y$分别表示X和Y的特征向量协方差矩阵;$\mathrm{tr}(\boldsymbol{A})$表示矩阵$\boldsymbol{A}$的迹;$\|\cdot\|^2$表示向量的L2范数的平方。LPIPS用于评估生成图像与真实图像的视觉相似性,值越小越好,表达式为

$$\mathrm{LPIPS}(A,B)=\sum_i w_i\cdot\|\phi_i(A)-\phi_i(B)\|_2 \tag{35}$$

式中:A和B分别表示生成图像和真实图像;$\phi_i(A)$和$\phi_i(B)$分别为图像A和B在不同网络层次i上的特征表示;$\|\cdot\|_2$表示L2范数;w_i表示在不同层次i上的感知权重。

3.3 对比实验分析

为验证DID-AugGAN算法性能的优越性,实验选取DCGAN、WGAN、WGAN-GP和StyleGAN2-ada四种生成算法进行对比。这4个对比算法在所研究的问题领域中表现优异,是目前被广泛应用和引用的代表性算法。为保证对比的公平性,所有对比模型的参数均设置为最优状态,实验结果如表2所示,其中加粗字体为最优值。

表2 不同网络模型实验对比结果

Table 2 Experimental comparison results of different network models

评价指标	类型	模型				
		DCGAN	WGAN	WGAN-GP	StyleGAN2-ada	DID-AugGAN
IS ↑	无缺陷	1.51	1.88	2.03	2.82	**2.97**
	断裂	1.73	2.19	2.46	2.74	**2.87**
	移位	1.43	1.76	2.11	2.70	**2.85**
	缺失	1.95	2.20	2.29	2.79	**2.91**
	平均值	1.66	2.01	2.22	2.76	**2.90**
FID ↓	无缺陷	284.30	215.55	182.34	39.30	**28.02**
	断裂	248.73	136.95	70.32	45.41	**37.38**
	移位	297.71	240.74	162.17	48.33	**42.74**
	缺失	197.26	133.02	105.97	43.79	**34.45**
	平均值	257.00	181.57	130.20	44.21	**35.65**
LPIPS ↓	无缺陷	281.19	249.92	244.45	225.06	**212.34**
	断裂	291.65	222.71	209.58	178.53	**166.95**
	移位	354.63	310.29	292.34	236.32	**223.03**
	缺失	356.47	345.73	329.51	234.28	**220.89**
	平均值	320.99	282.16	268.97	218.55	**205.80**

注：表格中的LPIPS值是乘以10^3之后的结果。

根据表2的实验结果可知，与改进前StyleGAN2-ada算法相比，DID-AugGAN在所有类型图像生成上均展现出性能上的提升。在“无缺陷”类型中，IS提升5.32%，FID降低28.70%，LPIPS降低5.65%；在“断裂”缺陷类型中，IS提升4.47%，FID降低17.68%，LPIPS降低6.49%；在“移位”缺陷类型中，IS提升5.56%，FID降低11.57%，LPIPS降低5.62%；在“缺失”缺陷类型中，IS提升4.30%，FID降低21.33%，LPIPS降低5.71%。综合4类图像来看，IS平均提升5.07%，FID和LPIPS平均分别下降19.36%和5.83%。根据3个指标的对比可以看出，DID-AugGAN算法在小样本条件下的扣件图像生成任务中展现出显著的优越性，该算法不仅能够生成多样且清晰的图像，而且在特征和感知上与真实图像高度相似，充分满足数据增强和平衡分布多样化的需求。

3.4 可视化对比

改进前StyleGAN2-ada算法与改进后DID-AugGAN算法在不同阶段生成效果如图5所示，其中图5(a～c)和图5(d～f)分别代表改进前和改进后模型训练100 kimg、500 kimg和1 000 kimg训练阶段的效果图(“kimg”为“千张图像”代表训练进度)，第1列到第4列分别代表无缺陷、断裂、移位和缺失4种类型。从图5可以看出，DID-AugGAN在同一阶段下生成的不同类型图像都要比StyleGAN2-ada更加清晰，且在最终收敛阶段，改进后算法生成的图像更接近原图。

DID-AugGAN与4种对比算法生成效果展示如图6所示，其中图6(a～d)分别对应无缺陷、断裂、移位、缺失4种扣件类型。每种类型均展示4张，6列代表6个来源，从左到右依次是真实图像、DCGAN、WGAN、WGAN-GP、StyleGAN2-ada、DID-AugGAN算法生成的图像。从图6可以看出，DCGAN、WGAN、WGAN-GP、StyleGAN2-ada生成的扣件图像存在不同程度的模糊、不清晰、噪声多等问题，尤其在扣件边界与螺帽处的细节不如DID-AugGAN。这是因为DID-AugGAN使用LOConv代替传统卷积，专注于学习图像中的非刚性特征，如扣件的形变和螺帽的形状。通过增强对这些复杂特征的捕捉

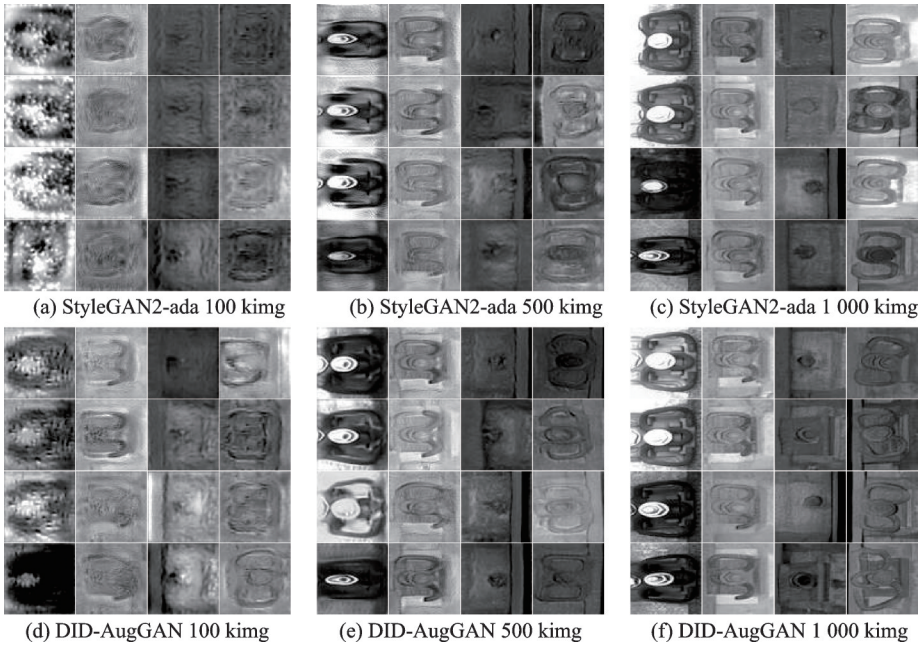


图5 模型改进前后不同阶段的生成效果图

Fig.5 Generation results at different stages before and after model improvement

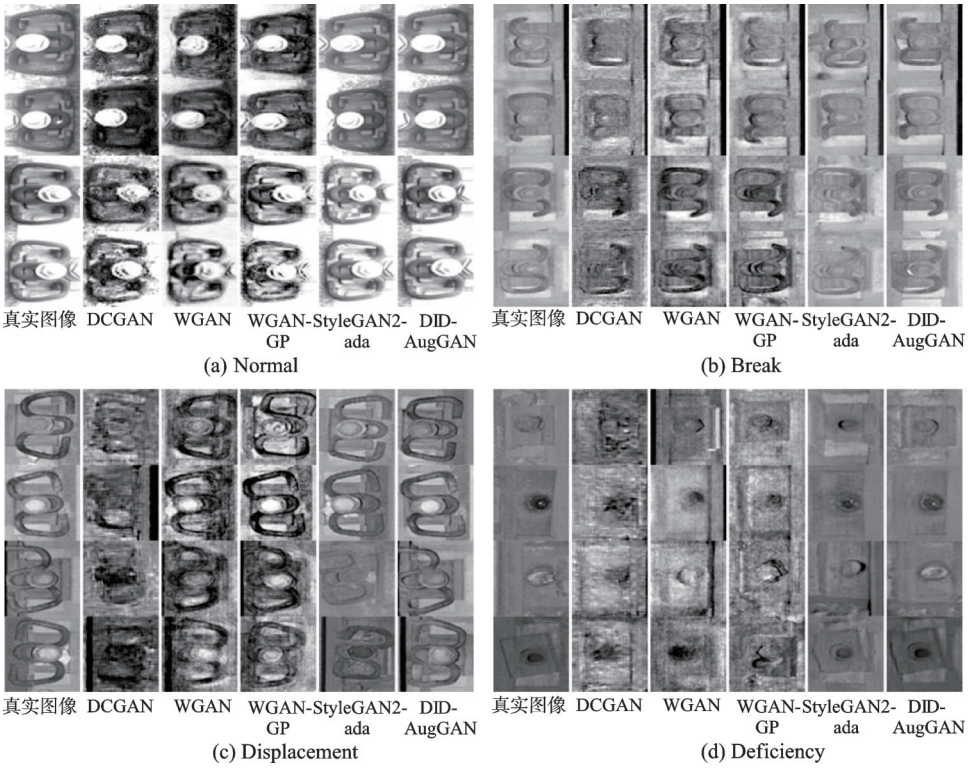


图6 多种GAN模型生成的效果对比图

Fig.6 Comparison chart of generation effects of various GAN models

能力，DID-AugGAN能够更精细地学习图像的语义信息，从而在生成图像时保留更多的细节和清晰度。其次，在模型中嵌入了MSCA模块，重点关注关键缺陷特征，防止模型在提取特征时丢失重要信息。最后，将判别器改进为UNet-like结构，解决了原判别器对局部信息判别能力不足的问题，进一步推动生成器生成出更加清晰、真实且细节丰富的扣件缺陷图像。通过上述改进，DID-AugGAN在生成图像的质量和细节上明显优于对比方法，理论分析和实验结果均证明了所提方案的有效性。

为验证本文所设计的LOConv对提高模型捕获非刚性特征能力的有效性，设计如下实验方案：在模型训练完成后，通过将学习到的偏移量设置为零来禁用LOConv，使其失效。最后，通过比较禁用前后的生成图像的差异来进行验证。如图7所示，第1行展示了DID-AugGAN生成的图像，第2行展示了将学习好的偏移量设置为零后模型生成的图像。值得注意的是，当将可学习的偏移量置零时，生成过程似乎出现了崩溃，生成的扣件形状发生了形变和扭曲。这是由于LOConv通过可学习的偏移量拟合非刚性特征，由此证明，本文提出的LOConv提高了非刚性特征的捕捉能力。

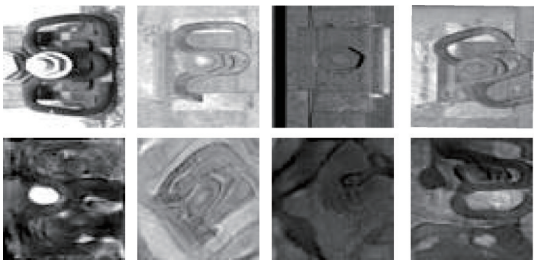


图7 禁用LOConv前后的生成效果图

Fig.7 Generation results before and after disabling LOConv

3.5 分类网络验证实验

为验证生成的轨道扣件缺陷图像及生成图像与真实图像混合对分类网络准确率的影响，本文设计11组分类实验，采用不同组成的训练集对MobileNetV3分类网络进行训练，具体实验内容如表3所示。

表3 分类训练集分布

Table 3 Distribution of classification training sets

实验编号	训练集组成	模型	生成图像数量	真实图像数量	训练集∶测试集
1	仅生成图像	DCGAN	300×4	0	3∶1
2	仅生成图像	WGAN	300×4	0	3∶1
3	仅生成图像	WGAN-GP	300×4	0	3∶1
4	仅生成图像	StyleGAN2-ada	300×4	0	3∶1
5	仅生成图像	DID-AugGAN	300×4	0	3∶1
6	仅真实图像	—	0	300×4	3∶1
7	混合图像	DCGAN	300×4	300×4	6∶1
8	混合图像	WGAN	300×4	300×4	6∶1
9	混合图像	WGAN-GP	300×4	300×4	6∶1
10	混合图像	StyleGAN2-ada	300×4	300×4	6∶1
11	混合图像	DID-AugGAN	300×4	300×4	6∶1

实验1～6的训练集均为单一生成图像或真实图像，训练集与测试集比例为3∶1，实验7～11的训练集是生成样本和真实样本的混合图像，分别将每类300张生成样本与300张真实样本混合，总计2 400张图像，训练集与测试集比例为6∶1。实验1～11未划分验证集，使用的测试集为同一样本，均为真实扣件图像，包括4类，每类100张，总计400张图像。MobileNetV3分类网络的训练轮次设置为50，并采用F_1分数作为评价指标，综合反映分类器的精确率和召回率，最终实验结果如表4所示，其中加粗字体为最优值。由表4可知，在仅使用单一生成数据集或真实数据集训练的情况下（实验1～6），

DID-AugGAN生成的数据集在分类准确率、精确度、召回率和F_1分数上均表现出优异性能,与真实数据集的结果非常接近。这表明该模型能够较好地保留真实数据的特征,使得生成图像在分类任务中具有较高的可用性。当生成图像与真实图像混合形成训练集时(实验7~11),混合数据集在分类性能上普遍优于单一数据集,这表明生成图像与真实图像的混合使用能够增强模型的泛化能力,提高分类准确率。此外,在混合数据集中,DID-AugGAN生成的数据集与真实图像的混合数据集依然表现最佳,这表明在真实数据的存在下,DID-AugGAN生成的图像依靠高质量的特征与真实数据相辅相成,共同提升分类网络性能。在缺陷扣件数据增强方面,DID-AugGAN具有巨大潜力,能够有效缓解真实数据不足的问题。

表4 分类实验结果

Table 4 Classification experiment results

实验编号	准确率/%	精确率/%	召回率/%	F_1分数
1	69.71	70.31	69.67	0.70
2	72.40	74.40	73.23	0.74
3	80.14	81.98	80.25	0.81
4	84.72	87.40	84.78	0.86
5	87.27	88.37	87.23	0.88
6	87.69	88.23	87.76	0.88
7	86.74	88.88	86.78	0.88
8	87.63	89.14	87.58	0.88
9	88.02	89.79	88.11	0.89
10	89.12	91.41	89.25	0.90
11	**91.33**	**92.10**	**91.26**	**0.92**

3.6 消融实验

为验证每个改进点的有效性,设计消融实验,通过在基准模型StyleGAN2-ada中逐一加入各改进模块,观察对生成图像的FID、KID、LPIPS指标的影响来评估改进效果。其中,M1为基准模型、M2为在M1基础上添加CA、M3为在M1基础上添加MSCA、M4为在M3基础上引入LOConv、M5为在M4基础上改进判别器结构,即为本文所提的DID-AugGAN算法。实验结果如表5所示,其中加粗字体为最优值。由表5可知,引入CA模块带来的提升不如引入MSCA模块显著,表明MSCA模块在CA的基础上增强了模型对细节和细微特征的处理能力。引入LOConv后,IS值进一步提升,FID和LPIPS值显著降低,说明LOConv在捕捉图像中的非刚性特征方面发挥了重要作用,增强模型对复杂图像结构的理解。最后将判别器改进为UNet-like结构后,即为DID-AugGAN模型,所有指标均达到最优,这证明UNet-like结构的判别器显著提升了模型对局部特征的判别能力,从而引导生成器生成更高质量的图像。消融实验结果表明,每一步的改进都对最终生成的图像质量产生了积极影响。随着改进措施的逐步集成,生成的图像质量不断提升,真实感和多样性也显著增强。

表5 消融实验对比结果

Table 5 Comparison results of ablation experiments

类型	模型	IS ↑	FID ↓	LPIPS ↓
无缺陷	M1	2.82	39.30	225.06
	M2	2.84	37.12	223.87
	M3	2.86	36.86	222.73
	M4	2.93	30.78	215.65
	M5	**2.97**	**28.02**	**212.34**
断裂	M1	2.74	45.41	178.53
	M2	2.76	44.09	175.74
	M3	2.77	43.72	174.92
	M4	2.82	39.76	169.23
	M5	**2.87**	**37.38**	**166.95**
移位	M1	2.70	48.33	236.32
	M2	2.73	47.86	234.98
	M3	2.74	46.92	234.13
	M4	2.81	43.15	226.38
	M5	**2.85**	**42.74**	**223.03**
缺失	M1	2.79	43.79	234.28
	M2	2.81	41.65	232.18
	M3	2.83	40.28	230.91
	M4	2.88	37.84	224.72
	M5	**2.91**	**34.45**	**220.89**

注:表格中的LPIPS值是乘以10^3之后的结果。

4 结束语

本文提出的DID-AugGAN小样本缺陷图像生成与数据增强算法，成功解决生成对抗网络在小样本条件下生成缺陷图像质量较低、不够真实且多样性差的问题。该算法首先在生成器中设计LOConv替代传统卷积，精确地捕捉图像中非刚性特征，聚焦于学习图像中的非刚性特征，更好地学习图像语义信息；其次设计MSCA模块，目的在于重点关注关键缺陷特征，解决在提取特征时发生特征丢失的问题；最后将判别器改进为UNet-like结构，解决原判别器对局部信息判别能力不足的问题，进而引导生成器生成高质量图像。实验表明，与其他流行生成算法对比，本文算法在小样本条件下的缺陷图像生成任务中展现出显著的优越性，不仅能够生成多样且清晰的图像，而且在特征和感知上与真实图像高度相似，充分满足数据增强和平衡分布多样化的需求，并改善因缺少缺陷数据样本造成下游任务效果不佳的情况。在未来的工作中，计划将DID-AugGAN算法拓展至其他数据稀缺领域，如医学图像、遥感图像等，以解决这些领域数据获取困难的难题。

参考文献：

[1] QIU S, CAI B, WANG W, et al. Automated detection of railway defective fasteners based on YOLOv8-FAM and synthetic data using style transfer[J]. Automation in Construction, 2024, 162: 105363.

[2] LI S F, WU C X, XIONG N X. Hybrid architecture based on CNN and transformer for strip steel surface defect classification [J]. Electronics, 2022, 11 (8): 1200.

[3] SANTOS C, PAPA J. Avoiding overfitting: A survey on regularization methods for convolutional neural networks[J]. ACM Computing Surveys (CSUR), 2022, 54(10): 1-25.

[4] HYUNKYU S, YONGHAN A, SUNGHO T, et al. Enhancement of multi-class structural defect recognition using generative adversarial network [J]. Sustainability, 2021, 13 (22): 12682.

[5] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: Learning augmentation strategies from data[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 113-123.

[6] REIPSCHLAGER P, FLEMISCH T, DACHSELT R. Personal augmented reality for information visualization on large interactive displays[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(2): 1182-1192.

[7] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.

[8] GOODFELLOW I, POUGETABADIE-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014, 3(2): 2672-2680.

[9] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2016-01-07)[2024-04-08]. http://arxiv.org/abs/1511.06434.

[10] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. [S.l.]: PMLR, 2017: 214-223.

[11] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs[C]//Proceedings of 2017 Advances in Neural Information Processing Systems. Long Beach: ACM, 2017: 5769-5779.

[12] KARRAS T, AITTALA M, HELLSTEN J, et al. Training generative adversarial networks with limited data[J]. Advances in Neural Information Processing Systems, 2020, 33: 12104-12114.

[13] 张印，胡挺，李猷兴，等. 基于条件对抗生成网络数据增强的相敏光时域反射仪模式识别[J]. 光学学报，2024，44(1): 0106026.
ZHANG Yin, HU Ting, LI Youxing, et al. Pattern recognition of phase-sensitive optical time-domain reflectometer based on conditional generative adversarial network data augmentation[J]. Acta Optica Sinica, 2024, 44(1): 0106026.

[14] 曹国刚，刘顺堃，毛红东，等. 基于优化循环生成对抗网络的医学图像合成方法[J]. 数据采集与处理，2022，37(1): 155-163.
CAO Guogang, LIU Shunkun, MAO Hongdong, et al. Medical image synthesis based on optimized cycle-generative adversarial networks[J]. Journal of Data Acquisition and Processing, 2022, 37(1): 155-163.

[15] 蔡伟，姜波，蒋昕昊，等. 非成对训练样本条件下的红外图像生成[J]. 光学精密工程，2023，31(24): 3651-3661.
CAI Wei, JIANG Bo, JIANG Xinhao, et al. Infrared image generation with unpaired training samples[J]. Optics and Precision

Engineering, 2023, 31(24): 3651-3661.

[16] 李宝平, 戚恒熠, 王满利, 等. 联合3D建模与改进CycleGAN的故障数据集扩增方法[J]. 光学精密工程, 2023, 31(16): 2406-2417.
LI Baoping, QI Hengyi, WANG Manli, et al. Equipment fault dataset amplification method combine 3D model with improved CycleGAN[J]. Optics and Precision Engineering, 2023, 31(16): 2406-2417.

[17] ZHANG G, CUI K, HUNG T Y, et al. Defect-GAN: High-fidelity defect synthesis for automated defect inspection[C]//Proceedings of 2021 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2021: 2524-2534.

[18] HANG R, ZHOU F, LIU Q, et al. Classification of hyperspectral images via multitask generative adversarial networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2): 1424-1436.

[19] DUAN Y, HONG Y, NIU L, et al. Few-shot defect image generation via defect-aware feature manipulation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(1): 571-578.

[20] 徐惠灵,尚政国,董胜波,等. 面向深度神经网络应用的小样本学习技术研究[J]. 南京航空航天大学学报, 2022, 54 (S1): 80-86.
XU Huiling, SHANG Zhengguo, DONG Shengbo, et al. Review on few-shot learning for DNN applications[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54 (S1): 80-86.

[21] DING G, HAN X, WANG S, et al. Attribute group editing for reliable few-shot image generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New Orleans: IEEE, 2022: 11194-11203.

[22] HU T, ZHANG J, YI R, et al. AnomalyDiffusion: Few-shot anomaly image generation with diffusion model[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(8): 8526-8534.

[23] XU J, LE H. Generating representative samples for few-shot classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New Orleans: IEEE, 2022: 9003-9013.

[24] ZHAO Y, DING H, HUANG H, et al. A closer look at few-shot image generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New Orleans: IEEE, 2022: 9140-9150.

[25] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13708-13717.

[26] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul: IEEE, 2019: 1314-1324.

[27] BARRATT S, SHARMA R. A note on the inception score[EB/OL]. (2018-01-06) [2024-04-07]. http://arxiv.org/abs/1801.01973.

[28] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in Neural Information Processing Systems, 2017, 6: 30.

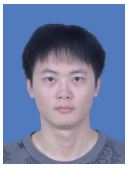
[29] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City: IEEE, 2018: 586-595.

作者简介:

黄绿娥(1981-),女,博士,副教授,硕士生导师,研究方向:图像处理、深度学习,E-mail:9320080310@jxust.edu.cn。

邓亚峰(1999-),男,硕士研究生,研究方向:深度学习、图像生成,E-mail:6720220767@mail.jxust.edu.cn。

鄢化彪(1978-),**通信作者**,男,副教授,硕士生导师,研究方向:深度学习、复杂系统建模,E-mail:yanhuabiao@jxust.edu.cn。



肖文祥(2000-),男,硕士研究生,研究方向:图像处理、深度学习,E-mail:6720220748@mail.jxust.edu.cn。

(编辑:夏道家)