DOI: 10.13382/j.jemi.B2306780

基于 CNN 融合 PGW-Attention 的金属表面缺陷识别方法*

赵云亮[1] 唐东林[1] 何媛媛[2] 丁 超[3] 杨 洲[1]

(1. 西南石油大学机电工程学院 成都 610500;2. 四川省特种设备检验研究院 成都 610061;
3. 成都工业学院智能制造学院 成都 611730)

摘 要:针对分散和细小的金属表面缺陷检测方法,卷积神经网络(CNN)缺乏全局特征捕捉能力,在识别氧化颗粒、裂纹和划痕等缺陷时易发生漏检和特征丢失,Transformer 能够捕捉图像全局信息,但全局计算导致较高的计算成本。为实现高效且精准的金属表面缺陷识别,将 CNN 的局部特征提取能力与 Transformer 的全局建模能力有效融合,提出了一种基于深度可分离卷积(DW-Conv)融合池化网格窗口注意力机制(PGW-Attention)的金属表面缺陷识别网络架构(DPG-Transformer)。在自建金属缺陷数据集(ST-DET)和公开金属缺陷数据集(NEU-CLS)上对该方法进行了实验验证,DPG-Transformer 的缺陷识别准确率分别为 99.3%和 99.6%,在准确率、计算量和浮点计算量等指标上优于多种经典网络。此外,在可视化实验中,DPG-Transformer 显示出比 CNN 模型更全面地腐蚀和氧化皮的缺陷特征提取能力,并能比 Transformer 模型更加精准地关注到细长裂纹和划痕的全局缺陷特征。实验结果表明,该方法可以降低 Transformer 模型的计算量和复杂度,同时能够更全面、精准地提取到金属表面缺陷特征,是一种更切合实际应用的金属表面缺陷检测方法。

关键词:金属表面缺陷;CNN;Transformer;深度卷积;PGW-Attention

中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:460.5040

Metal surface defect recognition method based on CNN with PGW-Attention

Zhao Yunliang[1] Tang Donglin[1] He Yuanyuan[2] Ding Chao[3] Yang Zhou[1]

(1. School of Electrical and Mechanical Engineering, Southwest Petroleum University, Chengdu 610500, China;
2. Sichuan Special Equipment Inspection Institute, Chengdu 610061, China; 3. School of Intelligent Manufacturing,
Chengdu Technological University, Chengdu 611730, China)

Abstract: To address the challenges in detecting dispersed and fine defects on metal surfaces, convolutional neural network (CNN) often fall short due to their limited ability to capture global features, leading to missed detections and loss of detail in identifying defects such as oxidation particles, cracks, and scratches. Although Transformers can capture comprehensive global information, the extensive computation required can be costly. In pursuit of an efficient and accurate method for metal surface defect detection, this study introduces a novel network architecture, the DPG-Transformer, which synergistically combines the local feature extraction capabilities of CNNs with the global modeling strengths of Transformer. This integration is facilitated through the use of depthwise separable convolutions (DW-Conv) and pooling grid window attention mechanisms (PGW-Attention). The effectiveness of the DPG-Transformer was validated on both a proprietary metal defect dataset (ST-DET) and a public dataset (NEU-CLS), achieving defect detection accuracies of 99.3% and 99.6%, respectively, and outperforming several classic networks in terms of accuracy, computational efficiency, and floating-point operations. Additionally, visualization experiments demonstrated that the DPG-Transformer more comprehensively extracts defect features associated with corrosion and scaling compared to CNN models, and more precisely focuses on the global features of elongated cracks and scratches than Transformer models. The results indicate that the DPG-Transformer not only reduces computational load and complexity but also enhances the comprehensive and precise detection of metal surface defects, making it

收稿日期:2023-07-27 Received Date: 2023-07-27

*基金项目:国家市场监督管理总局科技计划项目(2022MK115)、四川省市场监督管道局科技计划项目(CSCJZ2022007)资助

a highly suitable approach for practical applications in metal surface defect detection.

Keywords: metal surface defects; CNN; Transformer; deep convolution; PGW-Attention

0　引　言

在现代工业中,金属制品中的腐蚀、氧化和裂纹等缺陷不仅影响设备性能,还可能导致安全事故。因此,实现高精度和高效率的金属表面缺陷检测对保证产品质量、提升生产效率等至关重要。传统金属表面缺陷检测方法依赖人工视觉和无损检测,但由于检测速度慢和高度依赖专家经验,已无法满足工业自动化检测需求[1]。

随着深度学习方法的发展,对卷积神经网络(convolutional neural network,CNN)及其变体的研究大幅改善了金属表面缺陷检测性能[2-3]。如文献[4-6]通过改变 CNN 模型的宽度、深度和感受野使得模型非线性拟合能力变强,缺陷检测效果更佳。然而随着模型层数的加深,导致梯度消失问题的同时增加了模型参数量和复杂度。文献[7]采用 ResNext 模型结合涡流测试技术实现了对金属表面缺陷深度的有效评估,解决了梯度消失问题。文献[8]提出的紧凑型卷积神经网络减少了模型复杂度和计算量,在工业表面缺陷检测流水线上得到了有效应用。此外,基于 CNN 对金属表面缺陷检测的集成学习算法[9-11]和缺陷实时检测算法[12-13]也得到了较好的研究。然而 CNN 仍存在捕捉全局特征信息的局限性,无法通过增加模型的非线性拟合能力和降低模型复杂度等改进方法解决细小、分散的金属表面缺陷漏检问题。

目前,Tansformer 算法因其优秀的全局信息建模能力,在图像识别领域得到了广泛应用。同时,已有研究将 Transformer 应用于金属表面缺陷的检测[14],如 Gao 等[15]提出的 Cas-Vswin 模型为表面缺陷的准确定位和分割任务提供了新的窗口自注意力计算方案。Ding 等[16]提出局部和整体混合窗口注意力提高了工业表面缺陷检测的精度。为使模型具备关注局部和整体信息的能力,Li 等[17]和 Uzen 等[18]提出的 CNN 融合 Transformer 模型更适合检测目标大小变化较大的任务,同时实现了对缺陷细节的更多关注,从而提升了检测准确率。然而,简单融合 CNN 与 Transformer 的模型具有极大的计算量和较高的模型复杂度,从而难以在实际工程应用中部署。

在此基础上,本文提出了一种新型的金属表面缺陷识别网络架构,旨在解决深度学习方法在识别细小、分散的氧化颗粒、裂纹与划痕等缺陷时易漏检和特征丢失的问题。该架构结合了深度可分离卷积和混合窗口注意力机制,通过深度可分离卷积逐层捕捉缺陷特征,同时利用窗口注意力和池化网格窗口注意力构建混合窗口稀疏注意力机制,以建立图像的全局特征关系。此外,引入 CoordAttention[19]聚焦于重要特征通道。该架构有效融合了 CNN 和 Transformer 的优点,提高了缺陷检测性能,同时通过更高效的特征提取和建模方式,减少了冗余特征信息和重复建模,降低了模型的计算量和复杂度。

1　本文方法

1.1　方法概述

DPG-Transformer 模型的优势是通过有效结合深度可分离卷积和混合池化网格窗口注意力机制(PGW-Attention),创建了一个既能关注局部也能捕捉全局信息,同时具有较低参数量和复杂度的金属表面缺陷检测模型。如图 1 所示,DPG-Transformer 模型通过标准卷积层对输入图像进行初步特征提取,随后在模型主体中通过 4 个重复的 stage 进一步加工处理。每个 stage 由一次深度可分离卷积、一次 CoordAttention 通道注意力机制和两次 PGW-Attention 组成,在不同分辨率下的特征图上实现了全局与局部空间信息的有效交互。最终,通过平均池化和全连接层完成图像的预测输出。DPG-Transformer 模型具体参数如表 1 所示。

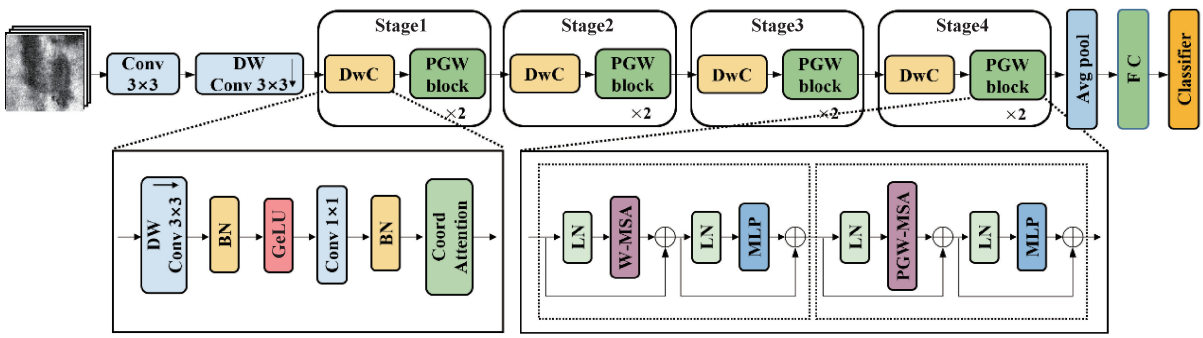

图 1　DPG-Transformer 模型结构

Fig. 1　Structure of the method DPG-Transformer model

表 1 DPG-Transformer 模型具体参数

Table 1 Specific parameters of DPG-Transformer

层号	操作	输出通道	标准化	激活函数	注意力头	卷积核	步距	层数
	Covn	24		GeLU	–	3	1	1
	DW_Conv	48		GeLU		3	2	1
Stage1	DW_Conv	48	BN	GeLU	–	3	2	1
	CoordAttention	48	BN	GeLU	–			
	PGW	48	LN	GeLU	3	–	–	2
Stage 2	DW_Conv	96	BN	GeLU	–	3	2	1
	CoordAttention	96	BN	GeLU	–			
	PGW	96	LN	GeLU	3	–	–	2
Stage 3	DW_Conv	192	BN	GeLU	–	3	2	1
	CoordAttention	192	BN	GeLU	–			
	PGW	192	LN	GeLU	3	–	–	2
Stage 4	DW_Conv	384	BN	GeLU	–	3	2	1
	CoordAttention	384	BN	GeLU	–			
	PGW	384	LN	GeLU	3	–	–	2
	AvgPool	384						1
	FC	384	–	–	–	–	–	1
	Classifier	Num_class						1

1.2 PGW-Attention

PGW-Attention 的核心设计思想是在 Swin Transformer 基础上引入一种更高效的跨窗口注意力机制,实现不同窗口间的信息交流,并减少冗余信息和重复特征。相比于 Swin Transformer 中的滑动窗口注意力机制,PGW-Attention 采用了更为稀疏的注意力计算方式构建全局信息关系。其设计原因是 DPG-Transformer 中 DW_Conv 层和每个 stage 中的 Window Attention 层都实现了图像局部信息的关系构建,因此完成跨窗口全局信息关系构建时,无需细化到特征层的每一个像素点。

其中,Window Attention 的本质是将输入特征图层划分为多个相同大小的窗口后,在窗口内进行自注意力计算,从而实现建立窗口内全局特征关系的同时大大降低计算量。自注意力机制计算过程如图 2 所示。

首先,通过 n 个不同的可训练权重 $\boldsymbol{w}^q$、$\boldsymbol{w}^k$、$\boldsymbol{w}^v$ 矩阵对输入特征图层 $\boldsymbol{x}$ 进行线性变换,映射出 n 组查询矩阵($\boldsymbol{Q}$)、匹配矩阵($\boldsymbol{K}$)、值矩阵($\boldsymbol{V}$)。

$$\boldsymbol{Q}=\boldsymbol{x}\boldsymbol{w}^q \quad \boldsymbol{K}=x\boldsymbol{w}^k \quad \boldsymbol{V}=x\boldsymbol{w}^v \tag{1}$$

然后对各组 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 矩阵进行 计算后得到的 n 个不同的 $\boldsymbol{Z}_n$ 矩阵,其中 *Softmax* 为归一化指数激活函数,d 值为查询矩阵维度。

$$Self\ \ Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}})\boldsymbol{V} \tag{2}$$

最后将 n 个不同的 $\boldsymbol{Z}_n$ 矩阵拼接后与权重矩阵 $\boldsymbol{w}^o$ 相乘得到包含了所有注意力头的效果矩阵 $\boldsymbol{Z}$,使得模型在不同子空间中具有关注不同位置关系的能力。

PGW-Attention 工作流程如图 3 所示,首先对经过窗口划分后的输入特征图层进行 Window Attention 操作,构建局部窗口自注意力。然后,将第 1 层 Window Attention 的输出特征层还原为窗口划分前的特征图层大小后,进行均值池化下采样操作,去除窗口内冗余信息。当特征图尺寸较大时池化尺寸可设置稍大,随着网络加深,特征图降采样倍数增加,池化尺寸适当减小。最后,通过 Grid Attention 操作将每个窗口中相同位置处的像素点提取出来,按照像素点的相对位置重新组成一个窗口特征图层进行 Window Attention 操作从而建立稀疏的全局注意力,然后通过上采样操作将全局重点关注信息传递回原特征层。

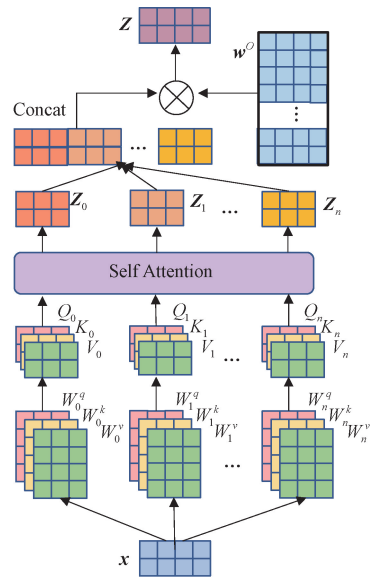

图 2 多头自注意力机制

Fig. 2 Multi head self attention mechanism

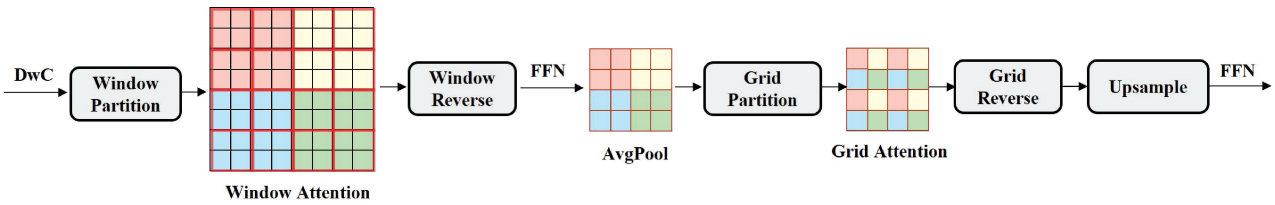

图 3　PGW-Attention 模型结构

Fig. 3　PGW-Attention structure diagram

PGW-Attention 能够建立与 Shift Window Attention 相同的全局特征关系，且整体计算量低于 Shift Window Attention，同时减少了制作掩码、循环位移等操作。MSA（multi-head Self-Attention）模块、W-MSA（Windows multi-head Self-Attention）模块与 PGW-Attention 计算复杂度对比如下，其中，h、w 和 C 分别表示特征图的高度、宽度和通道数，M 表示窗口大小。

$$\Omega(MSA) = 4hwC^2 + 2(hw)^2C \tag{3}$$

$$\Omega(W-MSA) = 4hwC^2 + 2M^2hwC \tag{4}$$

$$\Omega(PGM) = hwC^2 + \frac{M^2}{2}hwC \tag{5}$$

1.3　深度可分离卷积

深度可分离卷积[20]的主要作用是执行不同层次的缺陷特征提取和下采样操作。与传统卷积不同，深度可分离卷积将卷积过程分解为深度卷积和逐点卷积两个步骤。如图 4 所示，每个输入特征通道使用单独的卷积核进行单通道分离卷积，然后通过 1×1 的卷积核进行逐点卷积调整输出特征通道数。通过分步处理输入特征图的通道信息和位置信息，减少了模型的参数量和复杂度。

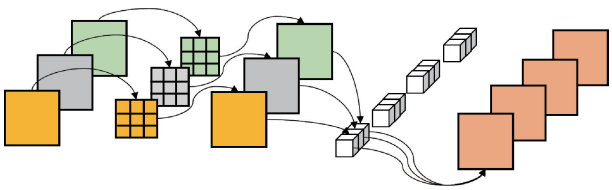

图 4　深度可分离卷积

Fig. 4　Depthwise separable convolution

假设卷积核尺寸为 K，输入特征图高宽为 H、W，输入、输出特征图深度分别为 M、N，通过比较普通卷积计算量 S_1 与 DW 卷积计算量 S_2 可得，普通卷积的计算量相对于 DW 卷积成倍增长，效率远低于 DW 卷积。

$$S_1 = K \cdot K \cdot M \cdot N \cdot H \cdot W \tag{6}$$

$$S_2 = (K \cdot K + N) \cdot M \cdot H \cdot W \tag{7}$$

$$\frac{S_2}{S_1} = \frac{1}{N} + \frac{1}{K^2} \tag{8}$$

1.4　CoordAttention 模块

通道注意力机制通过对输入特征图的全局空间信息进行压缩形成通道描述符，然后在特征学习的过程中优化通道描述符的赋值权重，实现突出重要特征通道、抑制冗余通道的作用。相比于常用通道注意力机制 SE[21]和 CBMA[22]，CoordAttention 保留了关键的通道特征位置信息，其具体工作流程如图 5 所示。

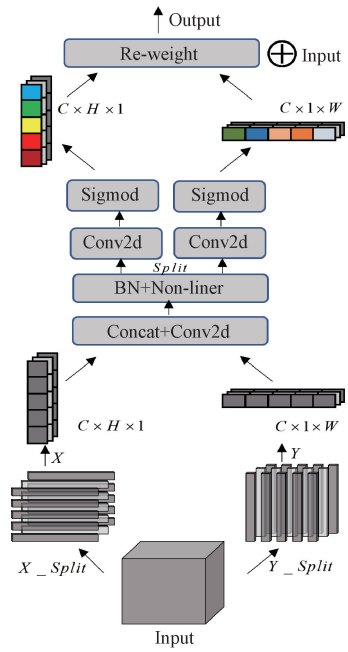

图 5　CoordAttention 通道注意力机制

Fig. 5　CoordAttention channel attention mechanism

CoordAttention 机制首先通过将全局池化操作分解为两个方向性池化核（H，1）和（1，W），分别对特征通道沿水平和垂直方向进行池化，实现全局空间特征的压缩及位置信息的编码。其次，将水平和垂直方向的特征图连接后，通过 1×1 卷积和非线性激活操作生成联合通道特征图。最后，将激活后的联合通道特征图分割成两部分，并通过两个 1×1 卷积重新转换为与原始输入同样通道数的特征图，完成特征通道权重的优化学习。

2 数据集和实验配置

2.1 数据集

为验证本文方法对金属表面缺陷的识别有效性和通用性,实验分别使用了自建金属罐体表面缺陷数据集 ST-DET 和钢材表面公开缺陷数据集 NEU-CLS。

ST-DET 中的缺陷图像由储罐表面缺陷图像截取出 480×480 pixels 大小的子图组成。数据集内含划痕(Sc)、氧化皮(Os)、表面腐蚀(Co)、表面裂纹(Cr)4 类缺陷,如图 6 所示为部分表面缺陷样本图片。

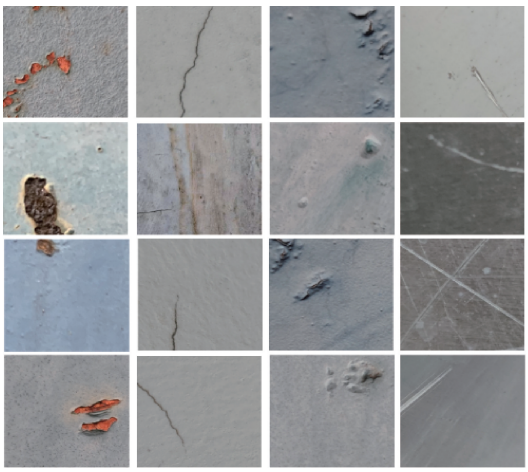

图 6 ST-DET 数据集部分缺陷样本

Fig. 6 Partial defect samples of the ST-DET dataset

NEU-CLS[23]是由东北大学制作的开放金属表面缺陷数据集,内含热轧钢带的 6 种典型表面缺陷图像,每类缺陷 300 张,大小为 200×200 pixels,共计 1 800 张,缺陷分别为开裂(Cr)、内含物(In)、斑块(Pa)、点蚀(Ps)、氧化皮(Rs)和划痕(Sc),如图 7 所示为部分缺陷样本图片。

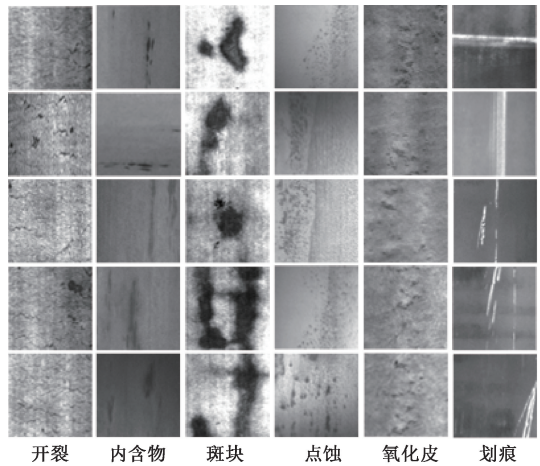

图 7 NEU-CLS 数据集部分缺陷样本

Fig. 7 Partial defect samples of NEU-CLS dataset

由于数据集样本数量较少,为确保模型充分训练,本文采用交叉验证方法进行 5 次独立实验,并取 5 次实验的结果平均值作为最终实验结果。在每次实验中,数据集的 80%用作训练集,剩余的 20%用作测试集。具体划分参数如表 2 所示。

表 2 数据集划分参数

Table 2 Dataset partition parameter table

类别	ST-DET 数据集					NEU-CLS 数据集						
	Co	Cr	Os	Sc	Total	Rs	Pa	Cr	Ps	In	Sc	Total
训练集	480	480	480	480	1 920	210	210	210	210	210	210	1 260
测试集	120	120	120	120	480	90	90	90	90	90	90	540

2.2 实验配置和超参数设置

本文实验平台为 Pycharm,实验代码基于深度学习框架 Pytorch 开发实现,硬件设备采用 AMD EPYC 7R13 48-Core CPU@2.65 GHz、NVIDIA GeForce RTX 3060 (12 GB) GPU,模型训练参数设置如表 3 所示。

表 3 模型参数设置

Table 3 Model parameter setting

参数类型	ULFSL-DET	NEU-CLS
图像大小	3×480×480	1×200×200
学习率	0.000 1	0.000 1
注意力头数	[3,3,3,3,]	[3,3,3,3]
迭代次数	100	100
学习率调整	余弦退火 T = 80	余弦退火 T = 80
批大小	42	42
优化算法	AdamW	AdamW
损失函数	交叉熵	交叉熵

3 实验结果与分析

3.1 特征层可视化实验

为研究模型在各阶段提取缺陷特征情况和重点关注信息,在模型参数达到稳定状态后,对不同阶段提取的特征图进行了可视化实验。如图 8 所示为模型中第一卷积层、Stage1 层、Stage2 层、Stage3 层以及 Stage4 层的部分输

出特征层,图中颜色亮度越高的区域表示模型在该特征图中的重点关注位置。

由可视化结果可知,模型在第一卷积层和 Stage1 层中有效提取了缺陷的纹理、边缘等基本形态特征,不同特征层分别关注了缺陷内部、边缘和背景纹理。这表明初级阶段的卷积操作能够捕捉表征缺陷的基本语义特征。模型在 Stage2 层提取到的轮廓、纹理和边缘特征开始变得模糊和抽象,关注的特征区域变宽泛,全局特征逐渐明显,开始显示出 PGW-Attention 在构建特征图全局信息上的效果。Stage3 层的特征图与原图差异较大,能粗略显示缺陷位置,全局特征更为突出,说明模型提取到了更加本质和全面的特征信息。最后,Stage4 层作为预测前的最终注意力机制层,输出的特征图以高度抽象的语义特征呈现,关注重点分布全局,缺陷类型不易用肉眼分辨。

通过特征层可视化实验表明,本文模型 DPG-Transformer 在浅层型在浅层主要关注缺陷的轮廓、纹理、边缘、背景等基础特征,而在深层则转向抽象和本质的语义特征。且通过 PGW-Attention 模块,模型的关注点由局部逐步扩展到全局。实验结果验证了模型设计的有效性,证明了其在关注局部与全局特征信息方面的能力。

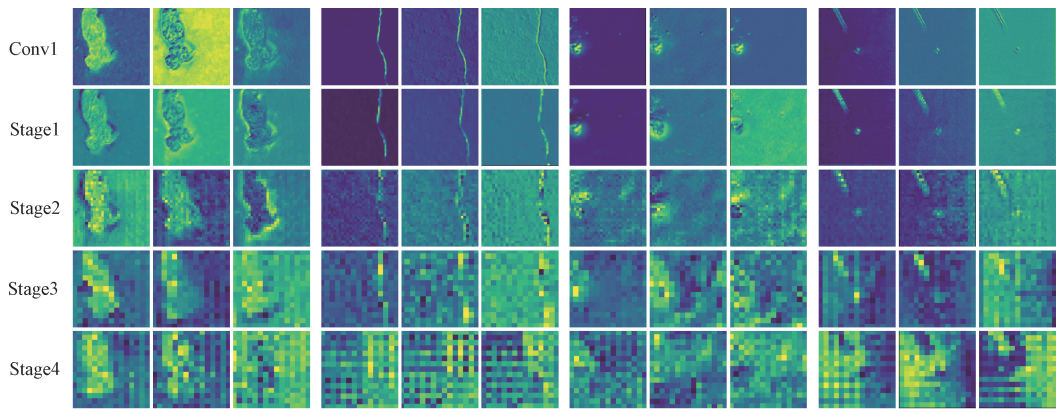

图 8　模型不同深度处部分输出特征图

Fig. 8　Visualization of output feature maps at different model layers

3.2　模型消融实验

为验证本文模型中不同模块的设计合理性及其效果,通过控制变量、替换相同功能模块的方式,分别对模型中自注意力机制、通道注意力机制和卷积方式进行了消融实验。实验同时在 ST-DET 数据集和 NEU-CLS 数据集上进行,通过模型参数量、浮点数和准确率 3 个指标进行评估,具体实验结果如表 4 所示。

表 4　消融实验

Table 4　Ablation experiment

模块	实验对象	ST-DET 数据集			NEU-CLS 数据集		
		Acc/%	FLOPs/G	Params/M	Acc/%	FLOPs/G	Params/M
Self-Attention	W-MSA+SW-MSA	98.5	3.65	4.86	98.9	0.86	4.86
	W-MSA+GW-MSA	98.7	3.65	4.86	99.2	0.86	4.86
	W-MSA+PGW-MSA	**99.3**	3.22	4.86	99.6	0.75	4.86
Channel-Attention	None	98.6	3.21	4.83	98.6	0.75	4.84
	SENet	98.9	3.22	4.85	98.9	0.75	4.86
	CBAM	98.9	3.22	4.85	98.8	0.75	4.86
	CoordAttention	99.3	3.22	4.86	99.6	0.75	4.86
Conv-Mode	None	97.8	3.12	5.15	98.5	0.68	5.12
	Traditional-Conv	99.5	4.30	5.65	99.8	0.94	5.65
	DW-Conv	**99.3**	**3.22**	**4.86**	99.6	0.75	4.86

本文设计的 PGW-MSA 在准确率和浮点数方面优于 SW-MSA[24] 和 GW-MSA[25],参数量三者相同。原因是 PGW-MSA 通过在网格注意力基础上引入池化操作构建更加稀疏的全局注意力,减少了计算量并扩大了重点关注区域。

此外,为增强对重要通道的关注并抑制冗余通道,模型引入了通道注意力机制。消融实验结果显示,SENet、CBAM 和 CoordAttention 3 种通道注意力机制均能有效提

升缺陷识别精度,其中 CoordAttention 参数量较高,但在提升识别准确率方面表现更佳。

对比 DW-Conv 与常规卷积方式(Traditional-Conv)可得,DW-Conv 由于少量的参数和计算量使得模型识别效率远高于 Traditional-Conv,但在识别准确率上略低于 Traditional-Conv,其原因是 DW-Conv 将卷积过程分解为深度卷积和逐点卷积两个步骤,其中深度卷积只关注通道内信息,导致特征捕获不够全面。本文模型在保证较高准确率的条件下选择了识别效率更高的 DW-Conv。

3.3 综合对比实验

为分析 DPG-Transformer 模型的综合性能,同时验证 DPG-Transformer 模型设计的有效性,选择经典的 CNN 模型 GoogLeNet[26] 和 ResNet34[27],以及优秀的 Transformer 模型 ViT、Swin-Tramsformer 进行综合对比实验。同时,为确保超参数设置不影响实验结果,所有模型在学习率、迭代次数、样本批大小、优化函数和损失函数采用相同设置,具体参数配置如表 3 所示。此外,仅将各模型最后一层的输出节点数调整为缺陷类别数,其他模型结构参数均保持不变。

首先,为验证 DPG-Transformer 模型捕捉缺陷图像的局部和全局特征的能力,采用 Grad-CAM[28] 类激活热力图展示各模型对缺陷图像的重点关注区域。如图 9 所示,在 ST-DET 数据集中选取了 4 类缺陷样本进行实验,图中由高亮颜色包围的区域为模型重点关注区域。其中,ViT 和 Swin_T 的通过自注意力机制对图像的整片区域进行了关注,但对边界不明显的氧化皮和划痕缺陷的特征提取能力较弱,主要表现为模型重点关注区域小于实际缺陷面积,其原因是缺少局部特征细化提取能力。ResNet34 和 GoogLeNet 模型对局部特征的关注能力较强,重点关注区域可覆盖缺陷周围区域,但因缺少全局信息关注能力,易漏掉相距较远的剩余缺陷特征。

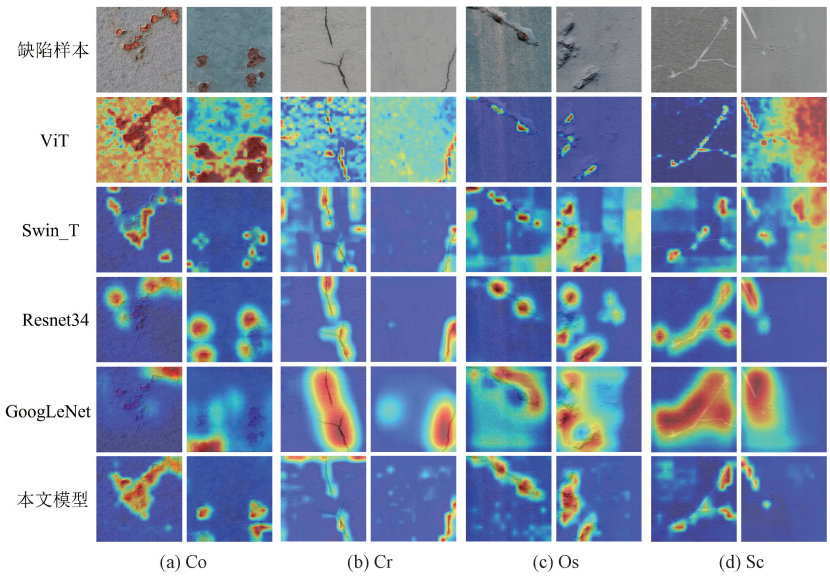

图 9 Grad-CAM 热力图综合对比实验

Fig. 9 Grad-CAM heatmaps activation experiment

相比之下,DPG-Transformer 能够比 CNN 模型更全面地提取腐蚀和氧化皮等区域集中的缺陷特征。同时,能够比 Transformer 模型更加精准地关注到细长裂纹和划痕的缺陷特征,且正确关注的缺陷面积更大。实验结果表明 DPG-Transformer 同时具有关注局部特征和全局特征的能力,能够正确筛选出重要的关注位置。证明了 DPG-Transformer 的能力符合模型设计的缺陷特征学习需求,且特征提取能力优于经典的 CNN 和 Transformer 模型。

其次,在 ST-DET 数据集和 NEU-CLS 数据集上,本次实验通过准确率(Acc)、参数量(Params)、浮点计算量(FLOPs)和推理时间(T)评估了各模型的复杂度和运行效率。如表 5 所示,DPG-Transformer 在准确率上表现最佳。得益于 PGW-Attention 的设计,该模型在参数量和浮点计算量上也优于其他模型。在推理时间上,DPG-Transformer 比 Transformer 模型快,但仍慢于 CNN 模型。ResNet34 与 GoogLeNet 虽然参数量、计算量较多但推理时间较快,其原因是 Transformer 模型涉及更多的序列计算,此外,Transformer 中的自注意力机制相较于 CNN 的卷积操作会引入更多的计算开销。

表 5　不同模型的测试结果

Table 5　Test results of different models

模型	ST-DET 数据集				NEU-CLS 数据集			
	Acc/%	FLOPs/G	Params/M	T/s	Acc/%	FLOPs/G	Params/M	T/s
ViT	98.1	77.09	86.93	0.065 8	98.8	16.78	86.00	0.060 3
Swin-Transfomer	**99.3**	20.73	27.503	0.044 2	99.5	4.70	27.53	0.041 2
ResNet34	98.7	16.88	21.29	**0.036 6**	99.4	3.60	21.28	**0.032 0**
GoogLeNet	98.1	7.31	5.98	0.036 7	98.5	1.51	5.97	0.033 0
DPG-Transformer	**99.3**	**3.22**	**4.86**	0.042 7	**99.6**	**0.75**	**4.86**	0.038 9

为进一步分析不同模型的推理性能，在 ST-DET 数据集上使用混淆矩阵详细展示了各模型针对每一类别缺陷的识别性能。图 10 中横坐标表示真实类别，纵坐标表示预测类别，对角线上的数值代表正确识别的数量。在 ST-DET 测试集上，DPG-Transformer 模型在各类缺陷识别上表现出较高的准确率，总体达到了 99.3%。结合 Grad-CAM 热力图可得，DPG-Transformer 能有效识别各类缺陷的局部特征，并建立全局稀疏注意力，从而实现注意力集中覆盖在缺陷区域，同时减小背景信息的干扰。

相比于 DPG-Transformer，其余模型的缺陷误检数量更多，模型性能有所降低。ViT 由于注意力分散，对边界不明显的缺陷如划痕和氧化皮的识别能力较弱，其识别准确率最低。GoogLeNet 有效关注到了缺陷附近的区域，但其对全局信息的学习能力不足，未能充分捕捉缺陷特征，从而影响了模型的决策效果。ResNet34 利用其残差结构保留了更多原始特征信息，有效减少了识别腐蚀等集中缺陷的错误，但其对裂纹与划痕识别性能较差，原因是两类缺陷相识度较高，缺陷都呈细长状，对模型全局信息关系建立的能力有一定要求。Swin-Transfomer 得益于其窗口注意力机制，其局部特征提取能力优于 ViT，并具有良好的全局特征提取能力，故在 ST-DET 测试集上与 DPG-Transformer 持平。

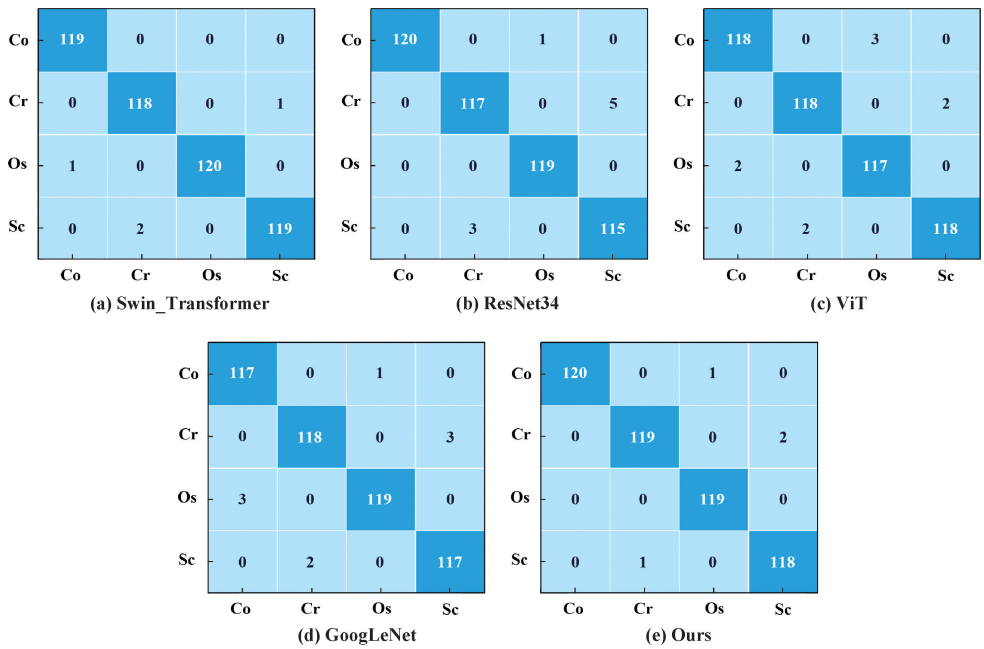

图 10　在 ST-DET 数据集上的识别混淆情况

Fig. 10　Identification confusion on ST-DET dataset

4　结　论

针对金属表面缺陷识别任务，本文提出了一种基于深度可分离卷积融合池化网格窗口注意力机制的网络架构 DPG-Transformer。通过在自建储罐缺陷数据集 ST-DET 和公开的钢材表面缺陷数据集 NEU-CLS 上的测试，验证了该模型的性能。

首先，DPG-Transformer 能够有效关注图像的局部与全局信息。其次，所提 PGW-Attention 在确保高准确率的同时，建立了更为稀疏的全局注意力机制，有效降低了模型复杂度。此外，相比于现有优秀模型，DPG-Transformer

在参数量、浮点计算量和准确率等多个指标上显示出优势,其关注缺陷区域相较于传统 Transformer 模型更全面,全局信息关注能力也超过常规 CNN 模型,有效减少了对细小分散缺陷的漏检。

实验结果表明,本文方法对金属表面缺陷高精度实时检测方法的实现具有良好的现实意义。进一步研究可以验证本文方法在目标检测领域的效果和应用到其他识别任务的通用性。

参考文献

[1] SUN M J, LIU T, CHENG X Z, et al. Nondestructive detecting method for metal material defects based on multimodal signals[J]. Acta Physica Sinica, 2016, 65(16): 167802.

[2] 李秀艳,夏琦琦,王琦,等. 基于深度学习与多频电磁阵列检测的金属板材表面和内部缺陷识别方法[J]. 电子测量技术, 2021, 44(4): 118-125.
LI X Y, XIA Q Q, W Q, et al. A recognition method for sheet metal surface and internal defects based on deep learning and multi-frequency electromagnetic array detection[J]. Electronic Measurement Technology, 2021, 44(4): 118-125.

[3] 刘铭璇,唐东林,何媛媛,等. CNN 集成机器学习的金属缺陷少样本分类方法[J]. 电子测量与仪器学报, 2023, 37(3): 86-94.
LIU M X, TANG D L, HE Y Y, et al. Classification of metal defects with few-shot based on CNN integrated machine learningl[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 86-94.

[4] HE D, XU K, WANG D. Design of multi-scale receptive field convolutional neural network for surface inspection of hot rolled steels[J]. Image and Vision Computing, 2019, 89: 12-20.

[5] BALCIOGLU Y S, SEZEN B, GOK M S, et al. Image processing with deep learning: surface defect detection of metal gears through deep learning[J]. Materials Evaluation, 2022, 80(2): 44-53.

[6] DENG H, CHENG Y, FENG Y, et al. Industrial laser welding defect detection and image defect recognition based on deep learning model developed[J]. Symmetry, 2021, 13(9): 1731.

[7] MENG T, TAO Y, CHEN Z, et al. Depth evaluation for metal surface defects by eddy current testing using deep residual convolutional neural networks[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 2515413.

[8] ZHOU X, NIE Y, WANG Y, et al. A real-time and high-efficiency surface defect detection method for metal sheets based on compact CNN[C]. 2020 13th International Symposium on Computational Intelligence and Design (ISCID). IEEE, 2020: 259-264.

[9] CHEN H Y, LIN C C, HORNG M H, et al. Deep learning applied to defect detection in powder spreading process of magnetic material additive manufacturing[J]. Materials, 2022, 15(16): 5662.

[10] HUANG Y C, HUNG K C, LIN J C. Automated machine learning system for defect detection on cylindrical metal surfaces[J]. Sensors, 2022, 22(24): 9783.

[11] HU C, WANG Y. An efficient convolutional neural network model based on object-level attention mechanism for casting defect detection on radiography images[J]. IEEE Transactions on Industrial Electronics, 2020, 67(12): 10922-10930.

[12] XIE Q, ZHOU W, TAN H, et al. Surface defect recognition in steel plates based on impoved Faster R-CNN[C]. 2022 41st Chinese Control Conference (CCC), IEEE, 2022: 6759-6764.

[13] 杨珂, 方诚, 段黎明. 基于深度学习模型融合的铸件缺陷自动检测[J]. 仪器仪表学报, 2021, 42(11): 150-159.
YANG K, FANG CH, DUAN L M. Automatic detection of casting defects based on deep learning model fusion[J]. Chinese Journal of Scientific Instrument, 2021, 42(11): 150-159.

[14] 慕长平, 耿丽清, 董建, 等. 基于多层级 Swin-Transformer 的低剂量 CT 重建[J]. 国外电子测量技术, 2022, 41(12): 156-163.
MU CH P, GENG L Q, DONG J, et al. Low-dose CT reconstruction based on multi-layer Swin-Transformer[J]. Foreign Electronic Measurement Technology, 2022, 41(12): 156-163.

[15] GAO L, ZHANG J, YANG C, et al. Cas-VSwin transformer: A variant swin transformer for surface-defect detection[J]. Computers in Industry, 2022, 140: 103689.

[16] DING C, TANG D, ZHENG X, et al. DHT: Dynamic vision transformer using hybrid window attention for industrial defect images classification[J]. IEEE Instrumentation & Measurement Magazine, 2023, 26(2): 19-28.

[17] LI Y, XIANG Y, GUO H, et al. Swin transformer combined with convolution neural network for surface defect detection[J]. Machines, 2022, 10(11): 1083.

[18] UZEN H, TURKOGLU M, YANIKOGLU B, et al. Swin-MFINet: Swin transformer based multi-feature integration network for detection of pixel-level surface defects[J].

Expert Systems with Applications, 2022, 209: 118269.
[19] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
[20] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
[21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
[22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
[23] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69 (4): 1493-1504.
[24] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
[25] TU Z, TALEBI H, ZHANG H, et al. Maxvit: Multi-axis vision transformer[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 459-479.
[26] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
[27] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
[28] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 618-626.

作者简介

赵云亮，2021 年于四川轻化工大学获得学士学位，现为西南石油大学硕士研究生，主要研究方向为模式识别与图像处理。
E-mail: zhaoyl_0313@163.com

Zhao Yunliang received his B. Sc. degree from Sichuan University of Science and Engineering in 2021. Now he is a M. Sc. candidate at Southwest Petroleum University. His main research interests include pattern recognition and image processing.

唐东林（通信作者），2006 年于天津大学获得博士学位，现为西南石油大学教授，博士生导师，主要研究方向为模式识别、光机电一体化技术。
E-mail: tdl840451816@163.com

Tang Donglin (Corresponding author) received his Ph. D. degree from Tianjin University in 2006. Now he is a professor and doctoral supervisor at Southwest Petroleum University. His main research interests include pattern recognition and optoelectro-mechanical integration technology.