

基于 AugModule 和 LightNet 的 轴承故障诊断方法研究

袁和金¹, 薛烁¹, 陈龙², 李金波², 马欢³

(1. 华北电力大学(保定)控制与计算机工程学院, 河北保定 071003;

2. 国网电力科学研究院有限公司, 南京 211106;

3. 国网安徽省电力有限公司超高压分公司, 合肥 230041)

摘要: 深度学习应用于轴承故障诊断时存在着样本稀缺和模型复杂的问题, 对此进行了研究并提出一种结合数据增广模块 AugModule 和轻量级分类网络 LightNet 的故障诊断方法; 采用 STFT 将设备的一维振动信号转换为二维时频图, 并通过引入谱归一化和 AdaptiveMix 改进的 WGAN 扩充数据集, 扩充后的数据集输入到 LightNet 网络进行图像分类; LightNet 基于 ShufflenetV2 改进, 使用 Sconv 提升精度, 采用 SMU 激活函数替代 Relu, 提出 Channel rate 和 Channel crossing 策略优化性能, 并重新设置每个 Stage 中的 Block 堆叠数目, 以实现准确率和效率的最佳平衡; 经实验测试, 该方法在减少模型参数的同时, 具有较高的准确率和实际工程意义。

关键词: 滚动轴承; 故障诊断; 深度学习; WGAN; LightNet

Bearing Fault Diagnosis Method Based on AugModule and LightNet

YUAN Hejin¹, XUE Shuo¹, CHEN Long², LI Jinbo², MA Huan³

(1. Computer department, North China Electric Power University, Baoding 071003, China;

2. State Grid Electric Power Research Institute, Nanjing 211106, China;

3. State Grid Anhui Ultra High Voltage Company, Hefei 230041, China)

Abstract: There are scarce samples and complex models with deep learning being applied in bearing fault diagnosis, this paper studied and proposed a fault diagnosis method combining the data augmentation module AugModule and the lightweight classification network LightNet. Short time Fourier transform (STFT) is used to convert one-dimensional vibration signal of the device into a two-dimensional time-frequency map, introduce spectral normalization and AdaptiveMix improved WGAN to expand the data set, input the expanded data set into the LightNet network for image classification, and improve the LightNet network based on ShufflenetV2 block. Using the Sconv module to improve accuracy, replacing the Relu with the SMU activation function, proposing the Channel rate and Channel crossing strategies to optimize its performance, and resetting the number of Block stacks in each Stage to achieve the best balance between accuracy and efficiency. Experimental results show that the proposed method has higher accuracy and practical engineering significance while reducing the parameters of the model.

Keywords: rolling bearing; fault diagnosis; deep learning; WGAN; LightNet

0 引言

在现代工业中, 机械设备的性能和安全至关重要, 开发智能高效的故障诊断方法对于提高机械设备的可靠性, 降低维护成本具有重要意义。

在旋转机械的轴承故障诊断领域, CWRU 数据集是故障诊断研究中常用的数据集, 利用该数据集进行故障诊断的方法有很多, 包括传统的信号处理方法, 基于机器学习方法以及基于深度学习的方法。传统做法是利用时域分析^[1-3]、频域分析^[4-6]和时频分析^[7-9]等方法来

收稿日期: 2024-05-15; 修回日期: 2024-07-03。

基金项目: 国家电网有限公司总部管理科技项目资助(5700-202340289A-1-1-ZN)。

作者简介: 袁和金(1977-), 男, 博士, 副教授。

引用格式: 袁和金, 薛烁, 陈龙, 等. 基于 AugModule 和 LightNet 的轴承故障诊断方法研究[J]. 计算机测量与控制, 2025, 33(7): 54-63.

提取振动信号的故障特征。这些方法依赖专家经验进行特征选择, 主观性强且效率不高, 难以捕捉复杂故障模式中的非线性和高阶特征。为解决这些问题, 有研究引入了支持向量机 SVM^[10] 和随机森林 RF^[11] 等机器学习方法, 尽管 SVM 在小样本和非线性问题中表现优异, 但处理大规模数据时效率较低; RF 在处理高维数据和样本不平衡方面表现出色, 但模型复杂且参数调整繁琐。

近年来, 深度学习在处理复杂故障特征时展现了更高的性能, 提供了更为有效的故障诊断解决方案, 其能够自动地从振动信号中学习和识别关键信息, 为故障诊断提供了新的可能性。比如, 文献 [12] 提出了一种基于改进的谱峭度图与多维融合 CNN 的故障诊断方法以减少干扰, 提高诊断的准确率。文献 [13] 基于 LeNet5 卷积神经网络对滚动轴承故障诊断方法进行了研究, 对轴承振动信号原始数据进行预处理后直接生成二维矩阵作为神经网络输入, 应用 sigmoid 函数进行分类。

然而, 上述深度学习网络在应用于实际故障诊断时存在一些挑战。首先, 大多数深度学习模型需要大量的标记数据进行训练, 这在实际工业场景中获取标记数据可能会面临困难。其次, 深度学习模型的训练和调优通常需要较长的时间和大量的计算资源, 这增加了部署成本和时间成本。因此高效的轻量化网络也是研究热点。目前, 经典的轻量级网络模型有 Xception^[14]、MobileNet^[15]、SqueezeNet^[16]、GhostNet^[17]、ShuffleNet^[18] 以及 FasterNet^[19] 等。

综上, 虽然深度学习方法在故障诊断中取得了显著的进展, 但仍然存在两个主要挑战: 首先, 由于故障样本数据的稀缺性, 传统的深度神经网络模型往往难以实现足够的诊断精度。其次, 现有的深度神经网络模型往往过于复杂, 难以在资源有限的移动设备上应用。这限制了深度学习在实际故障诊断工业场景中的广泛应用。

针对上述问题, 本文提出一种结合短时傅里叶变换、改进 WGAN 和 LightNet 的深度学习方法, 并将其应用于轴承故障诊断。

本文的工作内容如下:

- 1) 时频图的构建: 利用短时傅里叶变换来分析原始振动信号, 构建二维时频图;
- 2) 改进 WGAN 的应用: 对 WGAN 进行创新性的优化和改进, 有效地生成高质量的故障样本;
- 3) 高效轻量级网络的设计: 设计了一种更高效的新型轻量级网络 LightNet, 在保证故障分类准确率的同时降低计算复杂度, 实现模型轻量化。

1 基本原理

1.1 短时傅里叶变换 (STFT)

STFT 作为一种信号分析技术, 是傅立叶变换的一

个扩展, 其基本思想是首先将信号与窗口函数相乘, 进行一维傅立叶变换, 然后通过滑动窗口函数获得一系列时频信息, 将这些信息依次拼接, 形成最终时频图。其数学定义如下:

$$STFT\{x(t)\}(t, \omega) = \int_{-\infty}^{\infty} x(s)g(s-t)e^{-j\omega t} ds \quad (1)$$

其中: $x(t)$ 代表原始信号, $g(s-t)$ 是窗口函数, ω 表示频率。

T 和 F 为时间分辨率和频率分辨率, 如公式 (2) 和公式 (3) 所示, N_1 为样本长度, N_0 为窗口重叠宽度, N_2 为窗口宽度。

$$T = \left[\frac{N_1 - N_0}{N_2 - N_0} \right] \quad (2)$$

频域分辨率 F 根据窗口宽度的奇偶进行计算公式为:

$$F = \begin{cases} \frac{N_2}{2} + 1 & N_2 \text{ 为偶数} \\ \frac{N_2 + 1}{2} & N_2 \text{ 为奇数} \end{cases} \quad (3)$$

1.2 WGAN 的相关原理

生成对抗网络由生成器 (Generator) 和判别器 (Discriminator) 两个部分组成, 如图 1 所示, 生成器以随机噪声 z 作为输入, 目的是生成类似真实样本的合成数据。判别器则接收原始数据和生成器生成的数据, 其任务是区分真假。GAN 的核心目标是通过两个网络相互对抗, 让生成器生成逼真的数据, 并使判别器难以区分真实和生成的数据。

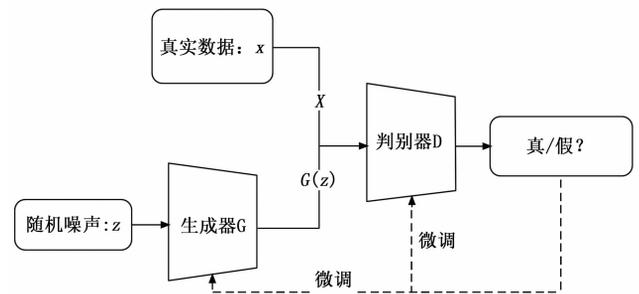


图 1 生成对抗网络 GAN 的基本结构

Wasserstein GAN^[20] (WGAN) 优化了传统 GAN 的训练流程, 主要是通过采用 Wasserstein 距离取代 Jensen-Shannon 散度来增强稳定性并提升了生成样本的质量。

1.3 ShufflenetV2 的相关原理

1.3.1 ShufflenetV2

ShufflenetV2 的基本模块结构如图 2 所示, 主要包括两种: 步长为 1 的普通模块 Block1 和步长为 2 的下采样模块 Block2。

普通模块 (a) 中, 输入特征首先按通道数进行平

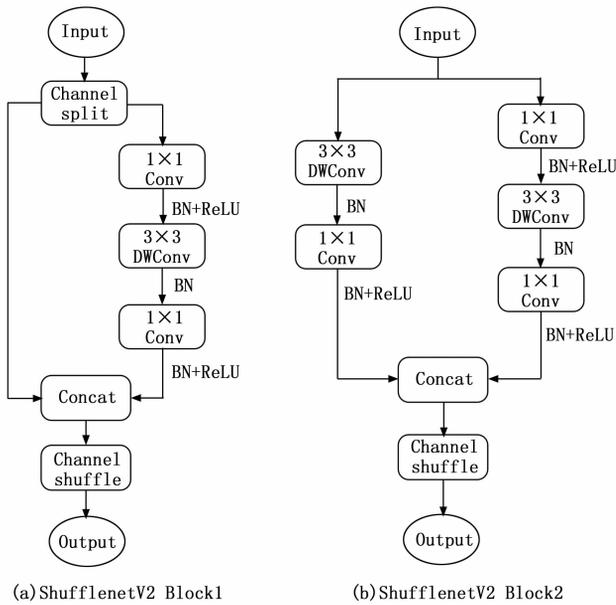


图 2 ShuffleNetV2 基本单元

均分割，分别进入左右两个分支。右分支依次进行、 1×1 逐点卷积和一次 3×3 深度可分离卷积操作，而左分支对输入特征不进行任何操作，最后两个分支进行通道级联和通道混洗。下采样模块 (b) 中，输入通道数也同样先一分为二。右侧分支采用和单元 (a) 基本相同的操作，而左侧支路进行 1 次 stride 为 2 的 3×3 深度可分离卷积和 1 次 1×1 逐点卷积，以完成特征的下采样。最终两个分支进行通道维度的拼接和通道混洗。

1.3.2 SCConv 卷积

如图 3 所示，SCConv^[21] 是一种新型卷积模块，它通过结合空间重建单元 (SRU) 和通道重建单元 (CRU) 两个组件，有效地限制了特征冗余。SRU 通过分离和重构的方式，抑制了空间维度上的冗余，增强了特征的表示能力。CRU 则利用分裂变换和融合策略，减少了通道维度上的冗余，降低了计算成本和存储需求。SCConv 模块不仅减少了模型参数和计算量，而且提高了特征表示的能力，为 CNN 的特征提取过程提供了新的视角。实验结果表明，嵌入 SCConv 模块的模型能够显著降低复杂性和计算成本，减少冗余特征，从而提高模型性能。

1.3.3 SMU 激活函数

ReLU 函数是深度学习中较为常见的一种激活函数，在原始 ShuffleNetV2 网络中被使用。但是 ReLU 也有缺点，LeakyReLU 的提出就是为了解决 ReLU 梯度消失的问题。

根据 Smith 等人的研究^[22]，将 SMU (Smoothed Rectified Linear Unit) 引入 ShuffleNet V2 模型中，以替代传统的 ReLU 激活函数，可以显著提升模型的性

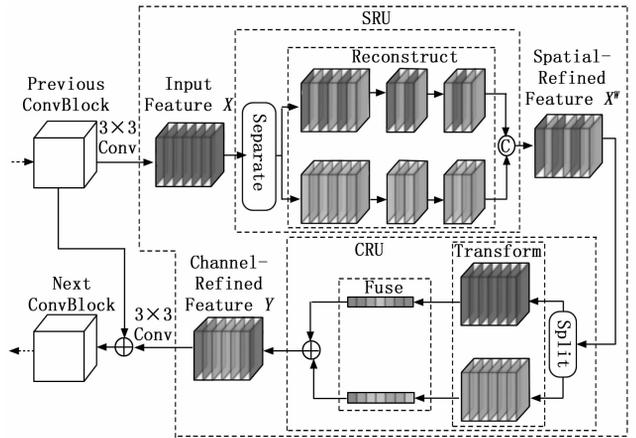


图 3 集成 SRU 和 CRU 的 SCConv 架构

能。SMU 是在已知激活函数 LeakyReLU 近似的基础上，提出的一种新的激活函数，其特点在于它的平滑性。公式如 (4) 所示：

$$f_1(x, \alpha; \mu) = \frac{(1 + \alpha)x + (1 - \alpha)x \operatorname{erf}[\mu(1 - \alpha)x]}{2} \quad (4)$$

其中： x 表示输入到激活函数的值， α 和 μ 是 SMU 的参数， $\operatorname{erf}()$ 是误差函数。此公式结合了线性项和经误差函数调整的非线性项，通过参数 α 和 μ 控制激活函数的形状和平滑度。这种结构允许激活函数在保持 ReLU 的某些优点的同时，减少其缺点，有助于缓解梯度消失问题，提高了模型的训练稳定性。

2 基于 AugModule 和 LightNet 的故障诊断方法

2.1 总体流程

针对设备故障样本较少和网络模型效率不高的问题，本文提出一种结合数据增广模块 AugModule 和图像分类网络 LightNet 网络的故障诊断方法。

总体流程如图 4 所示，首先，通过对振动信号应用短时傅里叶变换，将一维信号转换为二维时频图。进而采用改进的 WGAN 对原始样本进行数据增强，生成器生成新的样本，而判别器则学习区分真实数据和生成数据，通过对抗博弈来不断优化生成的样本质量。随后，扩充后的数据集被用来训练 LightNet 网络，最终，训练好的模型将在测试集上进行故障分类。

2.2 数据增广模块 AugModule

传统数据增广技术如镜像翻转、旋转、缩放、平移等通过对原始数据进行简单变换或添加噪声的方式来增加训练数据的多样性，然而其局限性在于可能无法覆盖全部数据空间，且缺乏对数据生成过程的深入理解，可能引入不合理的噪声。为了解决传统数据增广方法存在的问题，本文提出了一种新的数据增广模块，称为 AugModule，其中包含两个关键部分：STFT 和改进

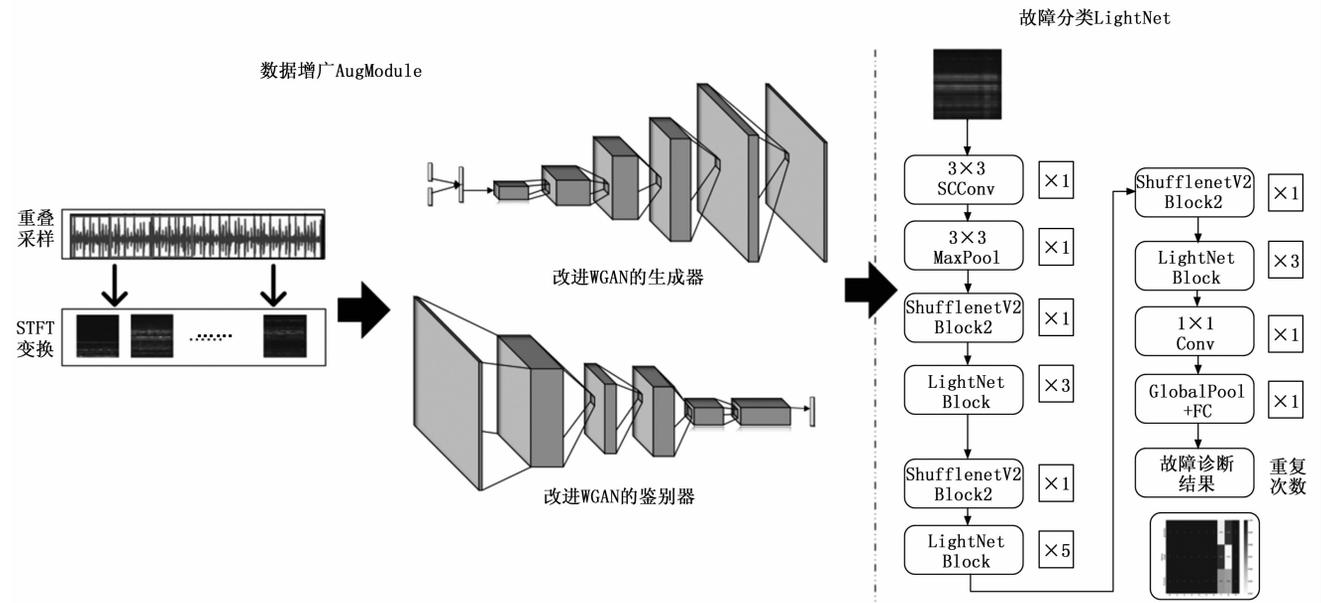


图 4 总体流程图

WGAN。

STFT 部分主要是信息增广, 将一维数据转换为二维时频图, 从而增加了数据的信息量和多样性。而改进 WGAN 部分主要是数量增广, 通过学习数据空间的分布特征, 更有效地生成高质量、多样性的时频图数据样本。

在大多数轴承故障监测系统中, 传感器捕获的原始振动信号是一维的、时变的、非稳态的, 如图 5 所示为原始振动信号。为了有效地利用 WGAN 和 CNN 在处理二维图像方面的优势, AugModule 模块先采用 STFT 将一维振动信号转换为二维时频图像。

在用 STFT 进行数据预处理的阶段, 首先将原始振动信号分割成多个 2 048 点的样本, 相邻样本之间有 128 点的重叠。接着, 对每个样本应用 STFT 变换, 并采用汉明窗作为窗函数, 窗宽设为 128, 重叠宽度为 64。将一维的原始振动信号经过上述处理之后得到二维的时频图像, 然后, 把得到图像按照 80% 和 20% 的比例随机划分成训练集和测试集。

AugModule 模块对 WGAN 网络的优化主要体现在以下两点:

1) 将判别器网络中的批量归一化替换为谱归一化^[23] (Spectral Normalization), 即对权重矩阵做奇异值分解, 然后用得到的最大的奇异值对权重矩阵来进行缩放, 保证 Lipschitz 连续性, 促进模型更快达到纳什平衡。

2) 设计新的损失函数, 引入 AdaptiveMix^[24], 通过降低特征提取器表示空间中难样本与其易样本之间的距离, 从而缩小判别器图像表示空间中训练数据的区

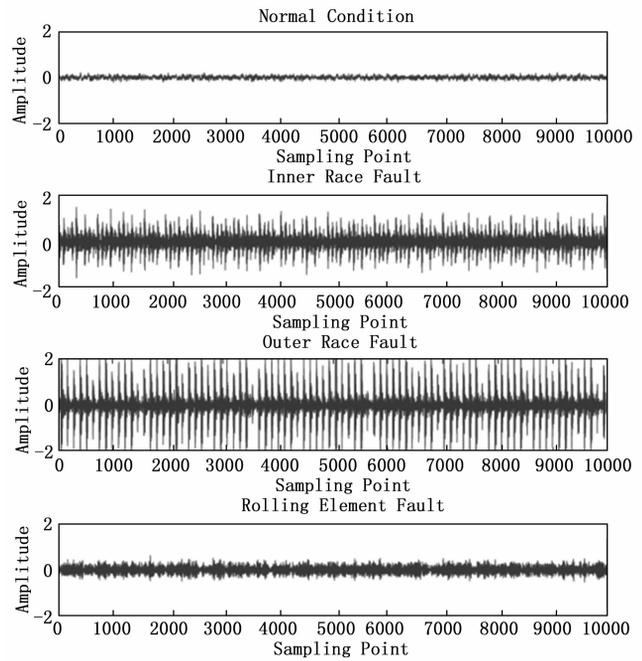


图 5 轴承振动信号波形图

域, 以提高 GAN 的训练效果。

其中, AdaptiveMix 是用 Mixup 的方法混合一对训练图像构建难样本, 通过在特征提取器 $F(\cdot)$ 表示的表示空间中减小难样本 x_j 与其对应的易样本 x_i 和 x_j 之间的距离, 从而缩小 x_i 或 x_j 所属类别的训练数据区域。如公式 (5) 所示, 通过设计新的损失函数 L_{ada} 来实现 AdaptiveMix 方法。

$$L_{ada} = \sum_i \sum_j D_V(\lambda F(x_i) + (1-\lambda)F(x_j), F(x^j)) + \sigma \tag{5}$$

通过在 WGAN 目标函数中引入 AdaptiveMix 模块, 优化 WGAN 的训练过程。如公式 (6) 所示, 这一目标函数融合了传统对抗训练以及上述 AdaptiveMix 损失。

$$\min_G \max_{F, J_{\rho, z}} IE[J(F(x))] - IE[J(F(G(z)))] + \min_{G, F_{\rho, z}, J_{\rho, z}} IE[L_{ada}] \quad (6)$$

改进 WGAN 的网络结构包括一个生成器和一个判别器。生成器的结构如图 6 所示, 其主体部分为反卷积层。输入是一个噪声向量 z 和一个条件向量 c , 其维度被 reshape 为 $4 \times 4 \times 256$ 。生成器首先通过 4 个 5×5 的反卷积层处理输入, 然后通过一个 3×3 的反卷积层, 最终生成 $64 \times 64 \times 3$ 的图像。生成器每层均使用 LeakyReLU 激活函数, 且每层均不用谱归一化处理。

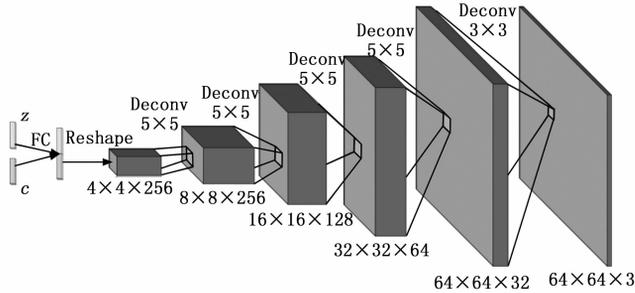


图 6 生成器网络结构

判别器的设计与生成器成对称, 具体网络结构如图 7 所示, 其主体部分为 4 层卷积核大小 3×3 的卷积网络。判别器通过卷积层处理输入, 将 $64 \times 64 \times 3$ 的输入图像转换为 $1 \times 1 \times 256$ 的特征张量。再连接全连接层, 最终得到对于输入的真假判断结果。判别器网络每层均使用 Leaky ReLU 激活函数, 中间 3 层加入谱归一化处理, 首尾层不用谱归一化处理。

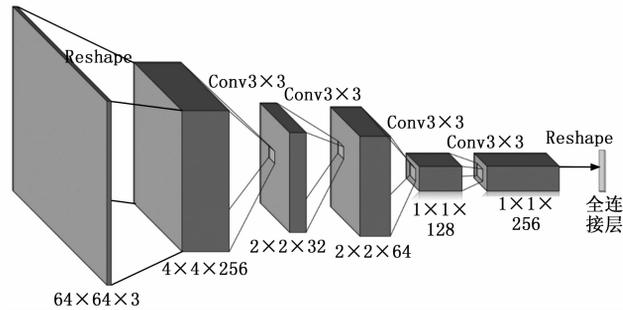


图 7 判别器网络结构

2.3 轻量化网络 LightNet

为了利用更少的网络参数实现更高的故障分类准确率, 本文设计了一种基于 ShufflenetV2 改进的轻量级神经网络 LightNet。

原始 ShufflenetV2 的 Shuffle Block 如图 8 所示, 存在以下不足之处: 由于模块中的三层卷积都要求保持输入输出通道数量一致, 模块计算分支容易产生通道冗

余; 中间信息层的信息并没有被有效利用, 模块的计算预算和控制能力较弱。

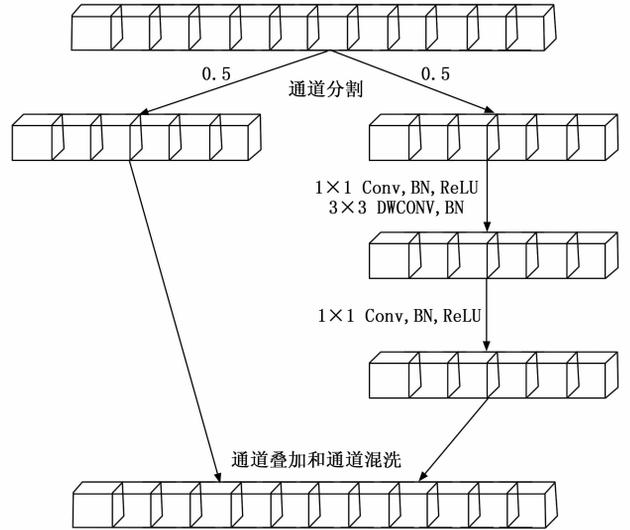


图 8 原始 Shufflenet Block

针对上述问题, 改进的 LightNet 基本单元如图 9 所示, 改进 LightNet Block 主要创新在于:

1) 设计通道分割比 Channel rate, 从而定义通过直接映射路径和通过卷积路径的通道数比率。右分支仍然可以保持足够的通道数目来学习复杂特征, 而辅助作用的左分支则可以用较少的通道数专注于提取更简单的特征。通过减少右分支的通道数, 可以降低卷积操作的计算量, 从而降低了整体模型的计算复杂度, 提升模型效率。

当 $rate > 0.5$ 时, LightNet Block 中的第一层和第二层会比原来的 Shuffle Block 更有效。为了保证在执行通道分割操作后, 两个分支都有整数通道, 设置, 便于实现和优化, 同时也减少因不对称的分割比率可能引入

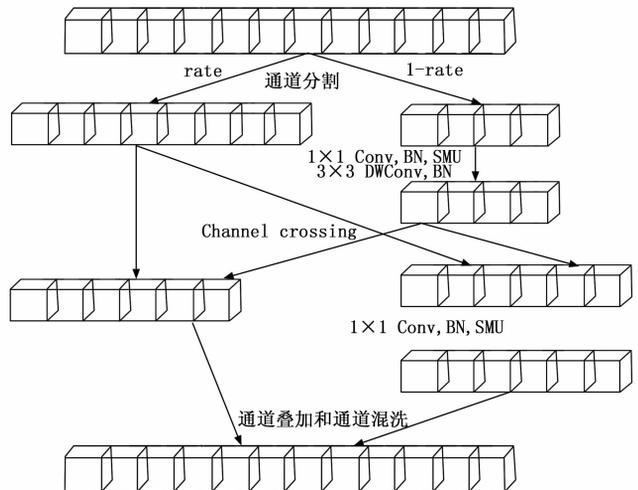


图 9 改进 LightNet Block

的计算复杂性。

2) 设计通道交叉 Channel crossing, 从而将输出特征和保留特征的一半进行交叉互换。这一操作不仅维持了每个通道的信息量, 而且确保了传递至下一层的特征同时包含当前和前一层的信息。这样的设计有效促进了跨层信息的交叉融合, 网络可以将深层的特征和浅层的特征进行有效整合, 从而避免了信息在网络中的过度浪费。

总的来说, 通过 Channel rate 以及 Channel crossing 可有效地改进 Shuffle Block 模块, 减少了通道冗余, 提升了信息的利用率, 使模型更加灵活高效。

在 ShufflenetV2 Block 和 LightNet Block 的基础上, 进一步设计 LightNet 网络, 其具体结构如图 10 所示, 其中 Stage2, 3, 4 部分是 LightNet Block, 数目代表对应模块的重复次数。

LightNet 网络结构和 ShufflenetV2 网络结构对比如图 10 所示, 改进主要体现在以下几点:

- 1) 采用 SCConv 卷积替换 3 个 stage 之前的常规 3×3 卷积, 在不增加参数量的同时降低冗余特征。
- 2) 用更加先进的 SMU 激活函数代替原 ReLU 激活函数, 以进一步提高模型准确率。
- 3) 引入通道分割比 Channel rate, 打破了传统固定 1:1 通道等量分割的模式, 通过调整 rate 的具体数值

控制计算开销, 减少冗余, 降低参数量。 $rate = \frac{x}{8}$ ($4 < x < 8$, 且 x 为整数), 实验测试了 $x = 5, 6, 7$ 三种不同设置, 发现 $x=5$ 时模型达到最高的性能指标。

4) 引入通道交叉 Channel crossing, 将输出特征和模块输入保留的特征通道交叉互换, 过交换部分通道, 促进了不同层之间信息的融合, 进一步整合信息, 无需增加额外的计算资源, 便提升了模型的表示能力和泛化能力。

5) 重新调整网络结构中的 stage2、stage3 和 stage4 中 Block 的堆叠数目, 以进一步降低模型的参数量和运算量。

3 实验结果与分析

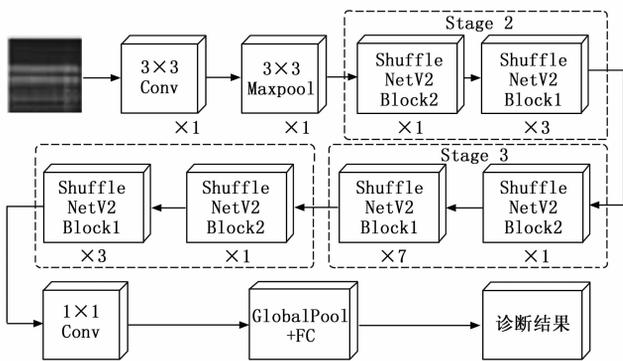
3.1 实验数据集

本文首先用验证故障诊断方法的经典数据集凯斯西储 CWRU^[25] 轴承数据集进行实验验证。选取 CWRU 数据集 12 kHz 的驱动端轴承振动数据进行试验, 所用由加速度传感器在不同负载条件 (0、1、2 和 3 hp) 下收集。

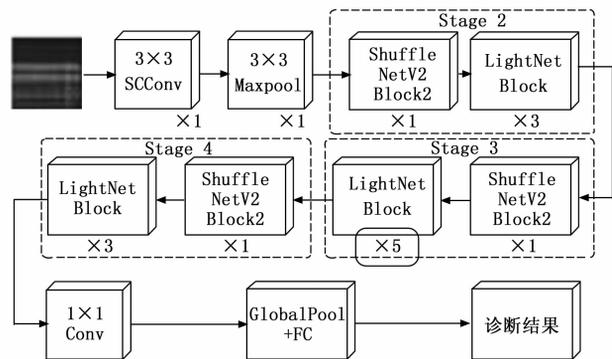
本文所用数据集包含 4 种工况下的 10 类数据, 详细描述如表 1 所示。并参考文献 [26] 选择数据集大小: 每类有 400 作为训练集, 100 作为测试集。

表 1 CWRU 数据集说明

类别标签	故障状态	故障尺寸/in	训练/测试样本数	负载/hp
0	正常	0	400/100	0,1,2,3
1	内圈故障	0.007	400/100	0,1,2,3
2	内圈故障	0.014	400/100	0,1,2,3
3	内圈故障	0.021	400/100	0,1,2,3
4	外圈故障	0.007	400/100	0,1,2,3
5	外圈故障	0.014	400/100	0,1,2,3
6	外圈故障	0.021	400/100	0,1,2,3
7	滚动体故障	0.007	400/100	0,1,2,3
8	滚动体故障	0.014	400/100	0,1,2,3
9	滚动体故障	0.021	400/100	0,1,2,3



(a) ShufflenetV2 的网络结构



(b) LightNet 的网络结构

图 10 ShufflenetV2 和 LightNet 的网络结构对比

3.2 数据增广实验结果

改进 WGAN 包含图 6 生成器网络和图 7 判别器网络, 采用 RMSProp 优化器更新网络参数, 生成器与判别器的学习率分别设置为 0.000 2 和 0.000 1。使用添加 AdaptiveMix 方法的新损失函数公式 (5) 来计算模型的损失, 模型训练的迭代次数为 2 000。

3.2.1 样本生成效果分析

将 STFT 图像原始训练集作为输入到改进 WGAN 模型的真实样本集, 模型训练完成后, 得到生成数据集, 在生成的各故障类型数据中随机选取一个与原始样本进行对比, 对比结果如表 2 所示。

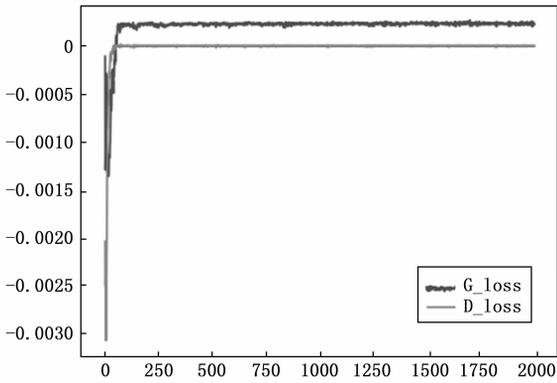
从表 2 可以看出, 生成的各故障类型图像与训练集

表2 真实样本与生成样本的对比

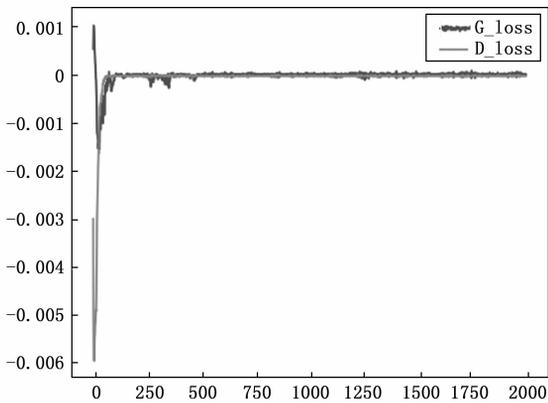
标签	0 正常状态	1 内圈故障	2 内圈故障	3 内圈故障	4 外圈故障	5 外圈故障	6 外圈故障	7 滚动体故障	8 滚动体故障	9 滚动体故障
真实										
生成										

的真实样本图像高度相似，说明改进 WAGN 模型训练优化程度良好。

图 11 为 2 000 批次训练后，改进前后 WGAN 的生成器和判别器的损失值变化曲线对比。由图可知：损失值经过前阶段的振荡后逐渐趋于平稳，达到纳什均衡。而 WGAN 在改进后的训练对比中表现出稳定的损失值下降，生成器和判别器的损失值迅速收敛并保持在较低的水平。这表明改进后的模型在训练过程中稳定性和收敛速度方面都有显著提升。



(a)改进前



(b)改进后

图 11 生成器和判别器损失值的变化曲线对比

为进一步分析，将改进 WGAN 生成样本和真实样

本之间的皮尔逊乘积矩相关系数 (PCC) 和余弦相似度 (CS) 作为评价指标来进一步评估生成样本的质量，各类别的分析结果如表 3 所示。

表 3 生成样本和真实样本之间的 PCC 和 CS

类别	PCC	CS
0	0.953 9	0.987 8
1	0.939 9	0.984 1
2	0.924 8	0.983 5
3	0.927 6	0.984 2
4	0.930 2	0.985 4
5	0.941 2	0.985 7
6	0.930 2	0.984 9
7	0.922 6	0.973 1
8	0.930 5	0.986 3
9	0.939 4	0.983 8

分析以上结果可知，每类样本的 PCC 值均大于 0.92 且 CS 值均大于 0.98，表明生成的样本与真实样本的数据分布高度相似，即改进 WGAN 模型生成的样本是有效的，可用于生成大量样本便于后续实验。

3.2.2 样本生成数与分类准确率的相关实验

为评估不同程度的数据增强对故障诊断模型性能的影响。用 AugModule 和 LightNet 模型在 CWRU 数据集上进行实验，比较了以下几种不同的实验设置，如表 4 所示。

表 4 不同数据集下的诊断模型准确率对比

实验序号	真实图片数目	生成图片数目	测试图片数目	准确率/%
1	400	0	100	95.80
2	400	100	100	96.70
3	400	200	100	97.40
4	400	300	100	97.70
5	400	400	100	99.20
6	400	500	100	98.60
7	400	600	100	97.50
8	400	700	100	96.30
9	400	800	100	95.00

1) 无数据增强 (实验 1): 仅使用原始真实数据进行训练。

2) 数据增强 1 (实验 2~5): 模型使用真实数据和生成数据的组合进行训练。随着逐步增加生成样本, 准确率稳步提高, 当生成样本的数量于真实样本的数量差不多时, 达到最高准确率。

3) 数据增强 2 (实验 6~9): 生成样本超过真实样本数目后继续增加, 准确率持续下降, 当生成的图像数量超过训练集的大小时, GAN 可能面临过拟合的风险。

总结来说, 在真实数据稀缺的情况下, 生成数据可以增强模型训练, 但过多的生成数据, GAN 可能面临过拟合的风险, 当真实数据有限时, 可能由于生成器模型引入的误差会超过增强的效果^[27]。实验结果证明, 提升分类准确率的最佳选择是补充 400 个生成样本到原始训练集中。

3.3 故障分类实验结果

LightNet 的训练集由真实样本和生成样本共同组成, 测试集仅包含真实样本。本文设置每个状态类的 400 个真实样本和 400 个生成样本作为 LightNet 的训练集。LightNet 模型使用 Adam 优化器更新参数, 批处理大小为 16, 初始学习率设置为 0.001, 学习率的衰减率为 0.09, 迭代次数为 50, 损失函数为交叉熵损失函数。

3.3.1 消融实验结果

为了探究使用 SCConv 卷积、LightNet Block 模块修改和 LightNet 的 Stage 结构调整, 3 种改进方式对 ShufflenetV2 带来的性能提升效果, 设置对照组, 采用控制变量的方式进行消融试验, 结果如表 5 所示。

表 5 LightNet 模型的消融实验

实验序号	修改卷积	修改模块	调整结构	ACC/%	Parameters/106
1	×	×	×	95.5	1.264
2	√	×	×	97.6	1.163
3	×	√	×	97.3	1.112
4	√	×	√	96.8	0.984
5	×	√	√	97.2	0.961
6	√	√	×	99.3	0.972
7	√	√	√	99.2	0.957

最终提出的 LightNet 轻量化模型的识别率为 99.2%, 较改进前的 ShufflenetV2 模型提高了 3.8 个百分点, 并且参数量降低了 2.4 个百分点, 有效实现了模型的更高精度和更高效率。

3.3.2 不同网络模型的对比分析

为进一步验证本文提出的改进模型进行故障诊断的可行性, 用两个数据集进行验证。其中数据集 1 为上述 CWRU 数据集, 数据集 2 选取 MFPT^[28] 数据中不同故障状态和负载下的轴承数据, 数据类型也包括正常状态、

内圈故障和外圈故障。然后分别选取轻量级网络 ShuffleNetV1、FasterNet、ShufflenetV2 和 MobileNetV2 在相同实验条件下进行对比, 实验结果如表 6 所示。

表 6 不同网络模型的性能对比

网络模型	数据集 1 ACC/%	数据集 2 ACC/%	Parameters/ 10 ⁶
ShuffleNetV1	84.3	80.4	1.825
FasterNet	88.9	85.9	3.840
MobileNetV2	92.3	88.6	3.504
ShufflenetV2	95.5	92.7	1.264
LightNet	99.2	96.3	0.957

从上述结果可以看出, 不同网络架构的识别效果不同, 在 CWRU 数据集上, LightNet 的识别率达到了 99.2%, 比 ShufflenetV2 高 3.7%, 参数量少 307 000。对比来看, LightNet 准确率最高而且参数量小于 ShuffleNetV1、FasterNet、ShufflenetV2 和 MobileNetV2。关于检测效率, LightNet 在 CWRU 数据集上的平均检测时间为 0.02 秒/样本, 相比于改进前的 ShuffleNetV 的 0.023 秒/样本, LightNet 在检测时间上具有显著优势。因此改进 LightNet 在故障诊断的准确率和轻量化之间取得了良好的平衡, 在参数量和性能上都优于其他网络模型。

图 12 展示了 ShufflenetV2 和 LightNet 在 CWRU 数据集上的混淆矩阵对比。混淆矩阵的每个单元格显示了预测类别与真实类别的匹配情况, 其中横轴代表预测类别, 纵轴代表实际类别。改进前, 类别 0、2、5 和 9 存在一些错误分类。改进后, 绝大部分类别的分类准确度显著提高, 只有类别 8 的识别准确率有所下降。分析原因可能是在模型改进过程中对类别 8 的特征提取和分类器参数进行了调整, 导致类别 7 和类别 8 之间的区分度下降, 从而引起误分类。此外, 改进前的模型可能对类别 8 存在过拟合现象, 改进后的模型在泛化能力上有所提升, 虽然在类别 8 的识别上表现有所下降, 但整体模型的适用性得到了增强。未来的工作将进一步优化特征选择和分类器参数调整, 确保各类故障样本的均衡性, 以提高模型的整体性能。通过对比, 可以得出: LightNet 网络的改进有效提高了模型整体的分类准确性。

在用 LightNet 进行分类时, 测试集上每个类别的精度、召回率和 f_1 分数结果如图 13 所示: 可以看出, 大多数类别的这些指标都接近或达到了 1.0, 这意味着模型在这些类别上的性能十分优异。总体来说, 实验结果显示模型的良好的性能, 能够高精度高效率地识别不同故障类别。

4 结束语

本文研究了一种基于 AugModule 和 LightNet 的轻

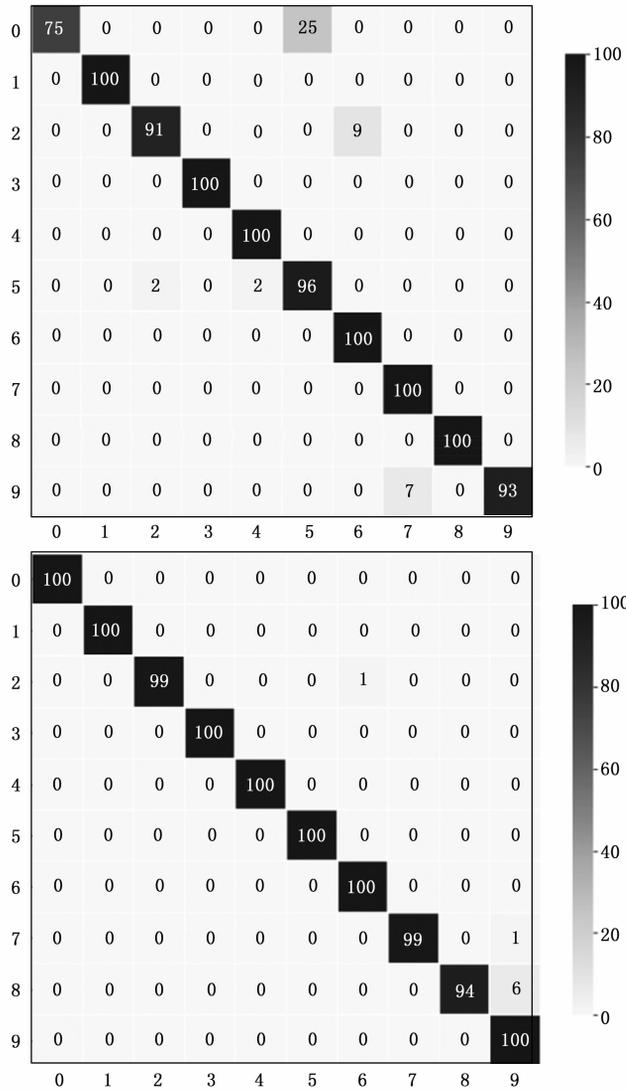


图 12 混淆矩阵对比

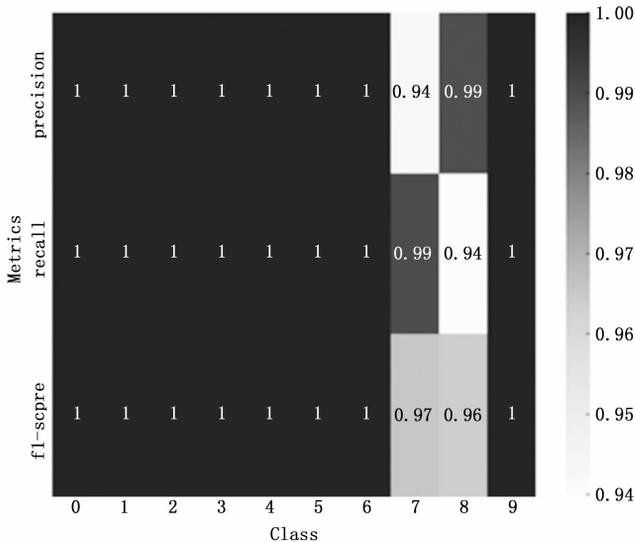


图 13 LightNet 分类时的参数

量化设备智能故障诊断方法，通过 STFT 将一维振动信号转换为二维时频图，并利用改进的 WGAN 进行数据扩充，有效缓解了故障样本数据稀缺问题，提升了模型训练效果。设计的 LightNet 网络在减少模型参数量和计算复杂度的同时，保证了高分类准确率。LightNet 在多个数据集上表现出优异性能，整体模型的泛化能力得到增强。

然而，尽管整体性能优异，模型在类别 7 和类别 8 之间仍存在一定的混淆问题，并且可能由于改进前模型对类别 8 的过拟合，导致改进后模型在该类别上的识别准确率有所下降。未来的研究将致力于进一步分类网络参数并在更多实际应用场景中进行验证，以进一步提高模型的鲁棒性和实用性。总之，本文方法在实际工业场景中具有显著的应用价值，有助于实现高效、智能的故障诊断。

参考文献：

[1] LIU Y. One-level stationary wavelet packet transform & hilbert transform based rolling bearing fault diagnosis [C] //2018 IEEE International Conference on Information and Automation (ICIA). IEEE, 2018: 1475 - 1479.

[2] YADAY O P, PAHUJA G L. An automatic approach to diagnose bearing defects using time-domain analysis of vibration signal [C] // International Conference on Advances in Electrical and Computer Technologies. Singapore: Springer Nature Singapore, 2020: 1305 - 1319.

[3] KUMAR J P, CHAUHAN P S, PANDIT PP. Time domain vibration analysis technique for condition monitoring of rolling element bearing: A review [J]. Materials Today: Proceedings, 2022, 62: 6336 - 6340.

[4] LI D, CAI Z, QIN B, et al. Signal frequency domain analysis and sensor fault diagnosis based on artificial intelligence [J]. Computer Communications, 2020, 160: 71 - 80.

[5] 陈 辉, 李伟光, 林 鑫, 等. 谱峭度方法在柔性薄壁轴承故障特征频率提取中的应用 [J]. 振动与冲击, 2020, 39 (1): 131 - 139.

[6] ZHANG X, WAN S, HE Y, et al. Teager energy spectral kurtosis of wavelet packet transform and its application in locating the sound source of fault bearing of belt conveyor [J]. Measurement, 2021, 173: 108367.

[7] CHEN F, YE Y, CHEN W, et al. Fault diagnosis of motorized spindle via modified empirical wavelet transform-kernel PCA and optimized support vector machine [J]. Journal of Vibroengineering, 2017, 19 (4): 2611 - 2631.

[8] XIE J, LIN M, YANG B, et al. A novel bearing fault diagnosis method under small samples using time-frequency multi-scale convolution layer and hybrid attention mechanism module [J]. Measurement Science and Technology,

- 2023, 34 (9): 095121.
- [9] ZHANG Q, HUO R, ZHENG H, et al. A fault diagnosis method with bi task-based time and frequency domain feature learning [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [10] 许迪, 葛江华, 王亚萍, 等. 量子遗传算法优化的 SVM 滚动轴承故障诊断 [J]. *振动. 测试与诊断*, 2018, 38 (4): 843 - 851.
- [11] 李兵, 韩睿, 何怡刚, 等. 改进随机森林算法在电机轴承故障诊断中的应用 [J]. *中国电机工程学报*, 2020, 40 (4): 1310 - 1319.
- [12] 楼伟, 陈曦晖, 赵伟恒. 基于改进谱峭度图与多维融合 CNN 的轴承故障诊断方法 [J]. *电子测量技术*, 2023, 46 (5): 185 - 191.
- [13] 刘林密, 崔伟成, 李浩然, 等. 基于卷积神经网络的滚动轴承故障诊断方法 [J]. *计算机测量与控制*, 2023, 31 (9): 9 - 15.
- [14] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 1251 - 1258.
- [15] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [J]. *ArXiv Preprint ArXiv*: 1704.04861, 2017.
- [16] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: alexNet-level accuracy with 50x fewer parameters and <0.5 MB Model Size [J]. *ArXiv Preprint ArXiv*: 1602.07360, 2016.
- [17] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1580 - 1589.
- [18] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 6848
- [19] CHEN J, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021 - 12031.
- [20] ARJOVSKY M, CHINTALA S, BOTTOUL. Wasserstein generative adversarial networks [C] // *International Conference on Machine Learning*, PMLR, 2017: 214 - 223.
- [21] LI J, WEN Y, HE L. SCConv: spatial and channel reconstruction convolution for feature redundancy [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 6153 - 6162.
- [22] BISWAS K, KUMAR S, BANERJEE S, et al. SMU: smooth activation function for deep networks using smoothing maximum technique [J]. *ArXiv Preprint ArXiv*: 2111.04682, 2021.
- [23] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [J]. *ArXiv Preprint ArXiv*: 1802.05957, 2018.
- [24] LIU H, ZHANG W, LI B, et al. Improving gan training via feature space shrinkage [J]. *ArXiv Preprint ArXiv*: 2303.01559, 2023.
- [25] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study [J]. *Mechanical Systems and Signal Processing*, 2015, 64: 100 - 131.
- [26] DEVECI B U, CELTIKOGLU M, ALBAYRAK O, et al. Transfer learning enabled bearing fault detection methods based on image representations of single-dimensional signals [J]. *Information Systems Frontiers*, 2023: 1 - 53.
- [27] 杨青, 陆见光, 唐向红, 等. 基于 infoGAN 和 CNN 的轴承小样本故障诊断 [J]. *兵器装备工程学报*, 2021, 42 (11): 235 - 240.
- [28] HEJAZI S Z, PACKIANATHER M, LIUY. A novel customised load adaptive framework for induction motor fault classification utilising MFPT bearing dataset [J]. *Machines*, 2024, 12 (1): 44.
- [29] LIU YUN HAI, WU YU HUAN, SUN GUOLEI, et al. Vision Transformers with Hierarchical Attention, 2021.
- [30] ZHENG ZHAOHUI, WANG PING, LIU WEI, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *AAAI Conference on Artificial Intelligence*, 2019.
- [31] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39 (6): 1137 - 1149.
- [32] 蔡聪艺. 改进 PSO-YOLOv5 的金属表面缺陷检测方法 [J]. *廊坊师范学院学报 (自然科学版)*, 2022, 22 (4): 37 - 41.
- [33] 张宇杰, 蔡乐才, 成奎, 等. 改进 YOLOv5 金属表面缺陷检测方法 [J]. *四川轻化工大学学报 (自然科学版)*, 2022, 35 (4): 32 - 41.
- [34] 徐静萍, 王芳. 基于改进的 S-ReLU 激活函数的图像分类方法 [J]. *科学技术与工程*, 2022, 22 (29): 12963 - 12968.

(上接第 53 页)

- [18] 刘涛, 张涛. 基于 GhostNet-YOLOv4 算法的印刷电路板缺陷检测 [J]. *电子测量技术*, 2022, 45 (16): 61 - 70.
- [19] LIU YUN HAI, WU YU HUAN, SUN GUOLEI, et al. Vision Transformers with Hierarchical Attention, 2021.
- [20] ZHENG ZHAOHUI, WANG PING, LIU WEI, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *AAAI Conference on Artificial Intelligence*, 2019.
- [21] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time object detection with region proposal net-