文章编号:1674-2974(2026)04-0029-12 DOI:10.16339/j.cnki.hdxbzkb.2026264

基于轻量化线性自注意力反向知识蒸馏网络的TEDS图像缺陷检测研究

王登飞[1],苏宏升[1†],葛磊蛟[1],王少飞[2],殷文福[3]
(1. 兰州交通大学 自动化与电气工程学院,甘肃 兰州 730070;
2. 中国铁路呼和浩特局集团有限公司 机辆检测所,内蒙古 呼和浩特 010000;
3. 中国铁路青藏集团有限公司 西宁通信段,青海 格尔木 816000)

摘 要:动车组运行故障动态图像检测系统(trouble of moving electric multiple units dynamic image detection system, TEDS)需进行检测的部件形态多样、体积大小不一,导致既有的检测方法误报率、漏检率高,因此,本文提出一种伪缺陷多头深度可分离自注意力反向知识蒸馏网络进行TEDS图像的无监督缺陷检测.首先,通过深度可分离卷积取代矩阵生成自注意力头向量,并以聚焦函数调整相似度的尖锐分布,构建的多头深度可分离线性自注意力享有线性运算复杂度;其次,通过倒瓶颈残差模块和多头深度可分离线性自注意力模块构建以轻量级教师-学生模型为主干的反向知识蒸馏网络,在提高网络特征提取能力的同时,减少网络可训练参数量,加块检测速度;在教师网络各个模块后设置投影层,同时采用Simplex和随机裁剪伪缺陷机制来模拟训练过程中的伪缺陷样本,通过多重损失引导投影层从正常特征空间中推开缺陷信息,迫使投影层专注于探索正常特征的更深层表示,来限制缺陷信息流向学生网络,使得教师、学生网络对缺陷有更大的特征差异.研究表明,改进后的网络能有效提高TEDS图片的缺陷检测能力,评价指标Sample-Auroc、pixel-Auroc、Aupro分别达到94.6%、91.71%、80.1%,和其他算法对比,分别提高3.3、3.8、4个百分点;且能够取得0.37 s/张的TEDS缺陷检测速度,满足TEDS系统的实时性需求.

关键词:动车组运行故障动态图像检测系统;知识蒸馏;缺陷检测;多头深度可分离线性自注意力;伪缺陷
中图分类号:U279.3 **文献标志码**:A

Research on TEDS image defect detection based on lightweight linear self-attention reverse knowledge distillation network

WANG Dengfei[1], SU Hongsheng[1†], GE Leijiao[1], WANG Shaofei[2], YIN Wenfu[3]
(1. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Rolling Stock Inspection Station, China Railway Hohhot Group Co., Ltd., Hohhot 010000, China;
3. Xining Communication Department, China Railway Qinghai-Xizang Group Co., Ltd., Golmud 816000, China)

* **收稿日期**:2025-01-03
基金项目:甘肃省教育厅高校教师创新基金项目(2024B-056), University Teacher Innovation Fund of Gansu Provincial Department of Education (2024B-056);甘肃省科技厅科技重大专项(22ZD6GA063), Major Science and Technology Projects of Gansu Province(22ZD6GA063);兰州交通大学-西南交通大学联合创新基金(LH2024027), Lanzhou Jiaotong University-Southwest Jiaotong University Joint Innovation Fund(LH2024027)
作者简介:王登飞(1986—),男,甘肃高台人,兰州交通大学副教授,博士
† 通信联系人,E-mail: 2641433027@qq.com

Abstract: The trouble of moving electric multiple units dynamic image detection system (TEDS) needs to detect components with diverse shapes and sizes, which leads to high false positive and missed detection rates in the existing detection methods. Therefore, a pseudo anomaly multi-head depth separable self-attention reverse knowledge distillation network is proposed to achieve anomaly detection on TEDS images. Firstly, the self-attention head vector is generated by replacing the matrix with depthwise separable convolution, and the sharp distribution of similarity is adjusted with a focus function. The constructed multi-head depthwise separable linear self-attention enjoys linear computational complexity. Secondly, a lightweight attention teacher-student model based reverse knowledge distillation network is constructed using a bottleneck residual module and a multi-head depth separable linear self-attention module, which improves the network's feature extraction ability while reducing the number of trainable parameters, and accelerates the detection speed. Projection layers are set after each module of the teacher network. Meanwhile, the Simplex and random cropping pseudo-defect mechanisms are employed to simulate pseudo-defect samples during training. Through multi-loss guidance, the projection layers are pushed away from the normal feature space to exclude defect information, forcing them to focus on exploring deeper representations of normal features and restricting the flow of defect information to the student network, resulting in greater feature differences between the teacher and student networks for anomaly. Research shows that the improved network can effectively enhance the anomaly detection capability of TEDS images; the evaluation metrics of image-Auroc, pixel-Auroc, and Aupro reach 94.6%, 91.71%, 80.1%, respectively. Compared with other algorithms, these metrics show improvements of 3.3, 3.8, 4 percentage points, respectively. This method can achieve a detection speed of 0.37 s per sheet, meeting the real-time requirements of TEDS systems.

Key words: trouble of moving electric multiple units dynamic image detection system (TEDS); knowledge distillation; anomaly detection; multi-head depth separable linear self-attention; pseudo-defects

动车组运行故障动态图像检测系统(TEDS)拍摄运行中列车侧面和底部的图片,通过工业网络传输至检测室,采用图像识别技术对列车部件表面进行缺陷排查和故障检测.在实际应用中,这些缺陷将不同程度地影响部件的抗疲劳、耐磨、抗腐蚀以及电磁等特性,因而通过检测及时发现缺陷,提前维修以排除隐患,可有效保障列车运行安全.动车组表面部件结构复杂,导致缺陷形态多样,使得现有的基于特征匹配[1]的图像故障检测方式漏检率、误报率高.李萍等[2]采用深度网络对列车关键零部件进行故障检测,Sun等[3]和周等[4]采用神经网络对列车运行安全图像进行有监督缺陷检测,相较传统特征匹配类方法,识别精度有所提升,但受限于收集的列车运行缺陷,图像类别不全、数量与模型不匹配,导致未知故障漏检率、误报率高,使其不能满足铁路现场的应用需求,目前主要靠人工与本车次历史图像对比检测缺陷.随着我国铁路的快速发展,TEDS系统日监测列车组已突破1万列次,采集的一列车的图片量为两千多张,导致人工检测强度大幅增加,因此采取当前先进的图像识别技术自动排查并标注TEDS图像缺陷,是提高检测效率和确保列车稳定运行的关键.

目前用于工业图像缺陷检测的方法主要为基于传统方法和基于深度学习两大类型.早期传统的缺陷检测方法如基于稀疏表示[5]、ASIFT特征变换的模板匹配法[6]等,通过对比图像中的局部特征和模板图像特征之间的差异来实现缺陷检测,该类方法对于采集环境高度可控的场景有很高的检测精度,但不适用于多变场景中的检测目标.基于深度学习的缺陷检测方法相比于模板匹配法,大幅度地提高了检测的准确度,且能更好地适应复杂场景.基于有监督深度学习的表面缺陷检测方法[7-8]需要充足且已精确标注的缺陷图像,但实际场景中列车运行缺陷图像采集困难,使得现有缺陷图像数量与深度学习模型无法匹配,导致模型泛化能力差.随后,仅使用正常样本进行模型训练的一系列无监督深度学习缺陷检测方法成为研究的主流.在无监督学习模型中,

基于特征重建的图像缺陷检测方法[9-10]研究较多，其核心思想是对正常图像通过编码、解码进行重建，以最小化重建损失作为训练目标,学习正常图像的特征分布进行缺陷检测.如Schlegl等[11-12]采用生成对抗网络(generative adversarial network，GAN)重建图像，经判别器得到重建误差来定位缺陷.Wyatt等[13-14]采用扩散模型通过逐步添加噪声，随后消除噪声来重建输入图像,通过比较原始图像与重建图像的差异来识别缺陷区域.然而,上述两种方法对复杂场景的重建效果不甚理想,此外,重建和去噪过程导致网络计算量大,难以达到TEDS图像缺陷检测的实时性需求.

深度网络在追求高性能的同时面临参数规模大、计算复杂度高的问题.因此在机器视觉领域，轻量化网络开始被广泛研究，MobileNet系列模型采用深度可分离卷积、线性倒残差结构[15]、通道注意力模块和新型ReLU6激活函数[16]优化网络，在减少参数和计算量的同时,提升网络的特征提取能力.受轻量级网络的启发，如Liang等[17]以深度动态卷积(depth dynamic convolution，DDConv)为基本结构，构建了精细残差双向聚合网络,在增强特征聚合能力的同时,降低计算复杂度,满足镍镀冲压钢板缺陷检测对准确性和实时性的需求.

随着视觉Transformer的兴起,部分学者利用自注意力优化缺陷检测网络,如Guo等[18]针对钢材表面纹理复杂多变的特点,将自注意力用于双向特征金字塔网络实现检测精确率提升.如Tao等[19]设计基于细粒度特征重构的混合模型,通过金字塔架构与全局注意力结合来提取特征,增强上下文感知能力，对纹理数据进行缺陷检测.Guo等[20]基于卷积神经网络(convolutional neural network，CNN)和自注意力设计了并行特征提取网络来获取样本的多层次特征，在MVTec数据集上可获得高的检测准确度.然而上述方法中自注意力的平方复杂度严重影响网络运行速度.

基于教师-学生[21-22]模型(teacher-student model，T-S)的反向知识蒸馏网络(reverse distillation，RD)[23]已被证明在缺陷检测中是有效的，在该框架中,T、S分别采用编码器、反向解码器结构克服了传统知识蒸馏网络的弱点，使得缺陷检测和标注准确性得以明显提升.此外,伪缺陷被学者广泛用来引导网络扩大正常和缺陷样本的特征差异，如Li等[24]提出的CutPaste通过随机图片剪切，然后粘贴到其他位置来构造伪缺陷，Liu等[25]提出的SimpleNet采用高斯噪声生成伪缺陷区域来引导缺陷检测.Tien等[26]提出改进的反向知识蒸馏网络(reverse distillation++，RD++)，采用Perlin[27]中的Simplex伪缺陷机制来模拟训练过程中的伪缺陷样本，然后采用重建损失引导投影层从伪缺陷特征中重建正常特征空间，进一步提高了反向知识蒸馏模型缺陷检测的性能.

针对列车表面器件结构复杂、缺陷形态多样，且所收集的缺陷样本较少,无法获取所有可能缺陷类型的充足图像用于训练,进而致使当前所采用的TEDS图像检测方法精确度较低等问题,提出一种基于轻量化多头深度可分离线性自注意力反向知识蒸馏网络的TEDS图像实时缺陷检测方法,主要贡献包括:

1)采用深度可分离卷积优化自注意力头，减少产生映射参数$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$的矩阵参数量和计算复杂度，利用聚焦函数替代Softmax函数对特征权重进行尖锐分布调整,同时通过调整矩阵运算顺序来优化传统注意力，由此构造的多头深度可分离线性自注意力,具有线性计算复杂度.

2)采用倒瓶颈残差模块和多头深度可分离线性自注意力构建轻量级T-S反向知识蒸馏网络，以多级特征计算缺陷图，在提高特征提取和重建能力的同时,保证对TEDS图像缺陷检测的实时性需求.

3)采用随机裁剪噪声、Simplex噪声来模拟TEDS图像缺陷，通过和教师网络共享权重的方式构建伪缺陷特征提取网络，并在两个教师网络的每个特征提取模块后加入投影层，并以多重损失引导投影层抑制缺陷特征传入学生机，以加大教师和学生网络对缺陷的特征差异.

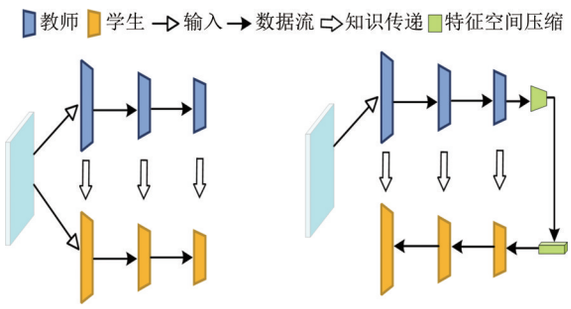
本文方法不需要标注缺陷样本，仅使用一组无缺陷样本训练就可识别并标注图像中的缺陷区域，研究表明,改进措施能有效提高TEDS系统的缺陷检测准确率，满足实时性检测要求.

1　反向知识蒸馏网络简介

对于非监督缺陷检测,定义$\boldsymbol{I}^{\mathrm{t}}=\{\boldsymbol{I}_1^{\mathrm{t}},\boldsymbol{I}_2^{\mathrm{t}},\cdots,\boldsymbol{I}_n^{\mathrm{t}}\}$为仅有正常图像构成的训练集,$\boldsymbol{I}^{\mathrm{q}}=\{\boldsymbol{I}_1^{\mathrm{q}},\boldsymbol{I}_2^{\mathrm{q}},\cdots,\boldsymbol{I}_m^{\mathrm{q}}\}$为同时包含正常和缺陷图像的测试集.在缺陷检测设置

中，I^{l}和I^{q}中的正常图像遵循相同的分布，其余被视为缺陷样本. 缺陷检测的任务即通过以上数据训练一个网络模型，来识别和定位测试集中的缺陷样本.

基于教师-学生模型的知识蒸馏架构的数据流如图1所示，知识蒸馏模型采用T-S模型对，让学生(S)从教师(T)处学习部分知识. 传统知识蒸馏架构[28]的数据流如图1(a)所示，T和S均采用编码器结构，同时送入T和S中对网络进行训练，经训练，当S和T输出特征相近时，说明S学会了T正常样本的特征表示. 测试过程中，输入为I^{q}，由于S在训练过程中只学习了正常图像的特征表示，S和T对正常图像输出的差异小，而对缺陷样本产生差异大的特征输出，通过差异的大小来判断输入图像是否有缺陷. 然而，T、S网络具有相似的或相同的架构，且在知识蒸馏期间具有相同的数据流有时会导致差异消失. 为了提高T-S模型在未知、分布外样本上的表示多样性，Deng等[23]提出如图1(b)所示的反向知识蒸馏网络架构，T采用编码器从输入图像提取特征，S采用对称的解码器以T的低维输出作为输入，反向生成和T相同的层次特征，在特征提取的过程中，S从T中获取知识. 对称性使S具有和T相同的表示维度，反向设计有助于消除S对缺陷的响应，由此S和T对缺陷区域可生成大的差异，提高检测准确性.



(a)传统知识蒸馏架构 (b)反向知识蒸馏架构

图1 基于教师-学生模型的知识蒸馏架构的数据流

Fig.1 Data flow of knowledge distillation architecture based on T-S model

2 用于TEDS图像缺陷检测的深度可分离线性自注意力RD网络

图2为实现TEDS图像实时缺陷检测的伪缺陷多头深度可分离线性自注意力RD网络结构. 采用倒瓶颈残差模块[23]和深度可分离多头线性自注意力构建以T-S为架构的反向知识蒸馏网络.

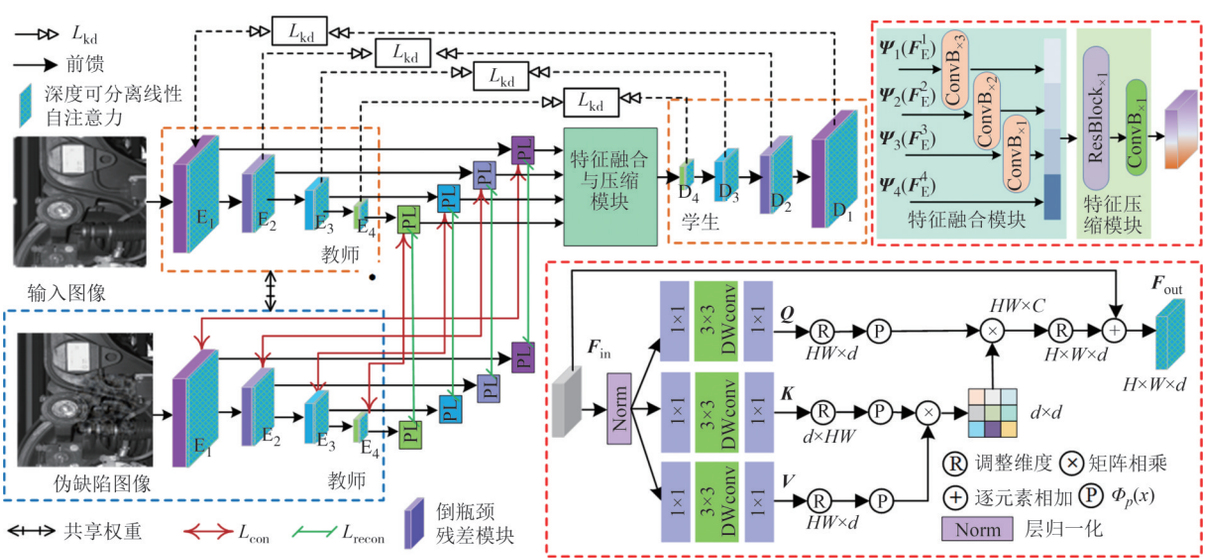


图2 伪缺陷多头深度可分离线性自注意力反向知识蒸馏网络结构

Fig. 2 Pseudo anomaly multi-head depth separable linear self-attention reverse knowledge distillation network structure

随着网络层级的增加，教师网络特征图空间尺度减小而通道数增加，学生解码器D采用和教师编码器E相反且对称的网络结构. 在训练过程中，教师网络只接收正常图像，通过知识蒸馏，学生网络在知识蒸馏损失的引导下以教师模型的输出作为输入，采用反向网络重建图像特征学习教师模型的知识. 反向知识蒸馏在检测过程中有效的前提是学生被限制接收缺陷信息，故采用随机裁剪伪缺陷和Simplex伪缺陷两种缺陷机制来模拟训练过程中的伪缺陷样本，并采用文献[26]的策略，在教师网络的各个模块之后设置深度可分离卷积实现的投影层，利用重建损失引导投影层从伪缺陷特征中重建

正常特征空间，在训练过程中阻止缺陷信息传入学生网络.教师模型最后一层输出的具有冗余和高自由度的特征不利于学生只学习正常特征，随后将所有投影层的输出送入多尺度特征融合与压缩模块(feature fusion and compression module，FFCM)以进行深层正常特征重建，由此来抑制训练过程中学生网络接收缺陷信号，从而使得教师和学生网络减小对正常特征的差异、增大对缺陷特征的差异，提升缺陷检测的准确度.

投影层以各自教师块的特征为输入，通过L个由深度可分离卷积、实例归一化、LeakyReLU激活函数组成的卷积块堆叠而成，在将特征输入FFCM之前，将其投影到紧凑的特征表示中，此处L=4.

特征融合与压缩模块如图2所示，将投影层的输出特征$\boldsymbol{\psi}_k(\boldsymbol{F}_{\mathrm{E}}^k)$分别通过(4−$k$)次步长为2的bottleneck对浅层特征进行下采样、归一化和ReLU激活函数(这3步为ConvB操作，图中$\mathrm{ConvB}_{\times 3}$的下标表示卷积层重复次数)处理后形成堆叠特征，随后将特征送入一层ResBlock进行多级特征融合，并采用一层步长为1的ConvB模块进行通道数压缩和特征激活来获得丰富而紧凑的特征.由学生编码器E_1~E_4输出的四级特征$\boldsymbol{F}_{\mathrm{E}}^k$经投影层和特征融合与压缩模块处理，融合了四种尺度的特征信息，学生网络以此为输入进行特征重建，所生成特征包含多级特征信息，可有效提升多尺度目标检测的准确度.

2.1　改进的深度可分离多头线性自注意力描述

传统多头自注意力：首先将输入特征$\boldsymbol{F}_{\mathrm{in}}$逐通道调整为向量，得到$\boldsymbol{X}\in\mathbb{R}^{HW\times C}$，其中$HW$表示图像的高乘宽，$C$表示通道数.然后$\boldsymbol{X}$按通道被分为$N$个头，$\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n,\cdots,\boldsymbol{X}_N]$，其中$\boldsymbol{X}_n\in\mathbb{R}^{HW\times d}$，$d=C/N$，$n=1,2,\cdots,N$.对每一个注意力头，采用如式(1)所示的3个矩阵将$\boldsymbol{X}_n$变换为查询向量$\boldsymbol{Q}_n\in\mathbb{R}^{HW\times d}$、键向量$\boldsymbol{K}_n\in\mathbb{R}^{HW\times d}$和值向量$\boldsymbol{V}_n\in\mathbb{R}^{HW\times d}$.

$$\boldsymbol{Q}_n=\boldsymbol{W}_{Q_n}\boldsymbol{X}_n,\ \boldsymbol{K}_n=\boldsymbol{W}_{K_n}\boldsymbol{X}_n,\ \boldsymbol{V}_n=\boldsymbol{W}_{V_n}\boldsymbol{X}_n \tag{1}$$

则每个头的自注意力如式(2)所示：

$$A_n(\boldsymbol{Q}_n,\boldsymbol{K}_n,\boldsymbol{V}_n)=\mathrm{Sim}\left(\frac{\boldsymbol{Q}_n\boldsymbol{K}_n^{\mathrm{T}}}{\alpha_n}\right)\boldsymbol{V}_n \tag{2}$$

式中：Sim(·)表示相似度权重调整函数，当Sim(·)用Softmax(·)表示时，即传统多头自注意力，$\alpha_n=\sqrt{d}$为用于自适应地缩放矩阵乘法的参数.随后N个注意力头的输出堆叠后，经过一层卷积变换，生成输出$\boldsymbol{X}_{\mathrm{out}}\in\mathbb{R}^{HW\times C}$，最后调整$\boldsymbol{X}_{\mathrm{out}}$的维度，形成注意力特征$\boldsymbol{F}_{\mathrm{out}}\in\mathbb{R}^{HW\times C}$.

深度可分离多头线性注意力原理：通过式(2)进行注意力运算时的运算复杂度为$o[(WH)^2d]$，通常WH远大于d，平方复杂度为传统多头自注意力计算速度慢的重要原因.文献[29]将$\boldsymbol{Q}_n$、$\boldsymbol{K}_n$、$\boldsymbol{V}_n$的计算顺序由$(\boldsymbol{Q}_n\boldsymbol{K}_n^{\mathrm{T}})\boldsymbol{V}_n$调整为$\boldsymbol{V}_n(\boldsymbol{K}_n^{\mathrm{T}}\boldsymbol{Q}_n)$，使多头自注意力计算复杂度减小为$o[(WH)d^2]$，但其相似度矩阵$(\boldsymbol{K}_n^{\mathrm{T}}\boldsymbol{Q}_n)$的最大秩由$WH$减小为$d$，使得特征多样性大幅度下降.为此，本文设计了如图2(b)所示的多头倒瓶颈深度可分离线性注意力，旨在降低传统自注意力计算复杂度并保证特征多样性.

式(1)中的权重矩阵$\boldsymbol{W}_{Q_n}$、$\boldsymbol{W}_{K_n}$、$\boldsymbol{W}_{V_n}$的维度为$\mathbb{R}^{HW\times HW}$，导致网络可训练参数量过大，为提高自注意力的运算速度，如图2(b)所示，采用三组倒瓶颈深度可分离卷积产生$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，该过程用式(3)表示.

$$\begin{aligned}\boldsymbol{Q}_n&=\boldsymbol{W}_{Q_n}^{\mathrm{P}'}\boldsymbol{W}_{Q_n}^{\mathrm{D}}\boldsymbol{W}_{Q_n}^{\mathrm{P}}\boldsymbol{X}_n\\ \boldsymbol{K}_n&=\boldsymbol{W}_{K_n}^{\mathrm{P}'}\boldsymbol{W}_{K_n}^{\mathrm{D}}\boldsymbol{W}_{K_n}^{\mathrm{P}}\boldsymbol{X}_n\\ \boldsymbol{V}_n&=\boldsymbol{W}_{V_n}^{\mathrm{P}'}\boldsymbol{W}_{V_n}^{\mathrm{D}}\boldsymbol{W}_{V_n}^{\mathrm{P}}\boldsymbol{X}_n\end{aligned} \tag{3}$$

式中：$\boldsymbol{W}_{(\cdot)}^{\mathrm{P}}$为用于升维的1×1逐点卷积层(PW)，$\boldsymbol{W}_{(\cdot)}^{\mathrm{D}}$为3×3的深度可分离卷积层(DW)，$\boldsymbol{W}_{(\cdot)}^{\mathrm{P}'}$为用于降维的1×1逐点卷积层.

$$\mathrm{Sim}(\boldsymbol{Q}_n,\boldsymbol{K}_n)=\phi(\boldsymbol{Q}_n)\phi(\boldsymbol{K}_n)^{\mathrm{T}} \tag{4}$$

为进一步降低计算复杂度，用式(4)所示的核函数且令$\phi(\boldsymbol{x})=\boldsymbol{x}$计算式(2)中的相似度，即：

$$\mathrm{Sim}(\boldsymbol{Q}_n,\boldsymbol{K}_n)=\boldsymbol{Q}_n\boldsymbol{K}_n{}^{\mathrm{T}}$$

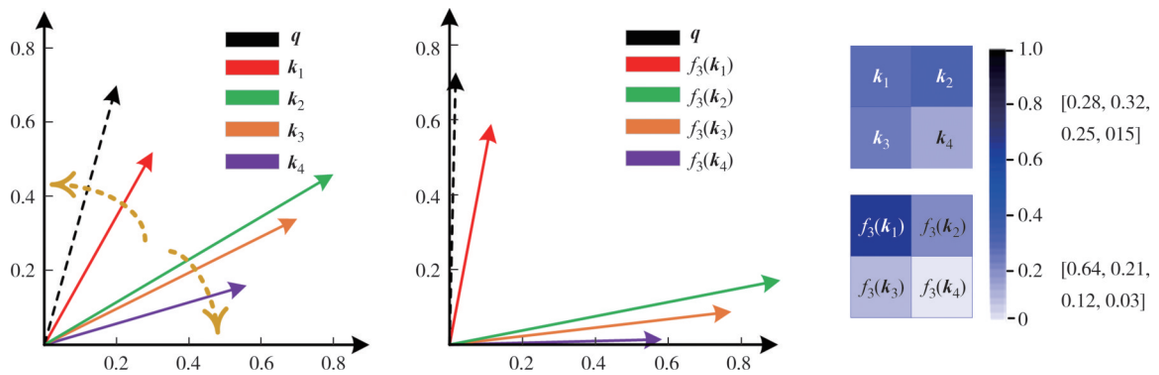
式(4)没有了Softmax(·)限制计算顺序，根据矩阵乘法的结合性质，可将$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$的计算顺序从$(\boldsymbol{Q}_n\boldsymbol{K}_n{}^{\mathrm{T}})\boldsymbol{V}_n$改为$\boldsymbol{Q}_n^{\mathrm{r}}(\boldsymbol{K}_n{}^{\mathrm{T}}\boldsymbol{V}_n)$，则形成线性注意力[30-31]，两者有相同的运算结果，但运算复杂度减小为$o(WHd^2)$.但上述线性注意力没有了Softmax函数的聚焦功能，导致注意力分布过于松散，性能较传统注意力明显下降.采用式(5)所示的聚焦函数调整权重，可使得式(4)所示的线性注意力达到近似于Softmax函数一样的尖锐分布.

$$\phi_p(\boldsymbol{x})=f_p(\mathrm{Relu}(\boldsymbol{x})),\qquad f_p(\boldsymbol{x})=\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}^{**p}\|}\boldsymbol{x}^{**p} \tag{5}$$

式中：$\boldsymbol{x}^{**p}$表示逐元素计算$\boldsymbol{x}$的p次方，可证明$\|\boldsymbol{x}\|=\|f_p(\boldsymbol{x})\|$，说明特征向量$\boldsymbol{x}$被$\phi_p(\boldsymbol{x})$变换后特征的范数不变，而方向得到了调整.

图3(a)和图3(b)给出了$\phi_p(\cdot)$将向量拉向靠近

它的坐标轴的调整效果，说明$\phi_p(\cdot)$作用于特征向量$\boldsymbol{q}$、$\boldsymbol{k}$时，将其按坐标轴分为若干组且将其拉向靠近它的坐标轴.如图3(c)所示，当$\phi_p(\cdot)$按式(4)进行相似度计算时，使得$\boldsymbol{q}$对更接近自己的特征$\boldsymbol{k}$有更高的相似度，并减小和它远离特征$\boldsymbol{k}$的相似度.图3中坐标轴为无量纲数据，取值范围为[0,1]，根据颜色深浅表示相似度.由此利用式(5)可聚焦相似的特征，实现相似和不相似特征间更尖锐的差异分布.



(a)调整前 (b)调整后 (c)$\phi_p(\cdot)$使相似度权重有更尖锐的分布

图3 $\phi_p(\cdot)$对特征向量和相似度权重的调整(以二维向量为例说明，取p=3)

Fig. 3 The resign of features and similarity weight by $\phi_p(\cdot)$(illustrated by two-dimensional vectors, p=3)

除了注意力聚焦功能，特征多样性也是限制注意力表示能力的重要因素，相似度矩阵的秩是特征多样性的重要体现.

$$\begin{aligned}&\operatorname{rank}\{\phi_p(\boldsymbol{Q}_n)\phi_p(\boldsymbol{K}_n)^{\mathrm{T}}\}\leqslant\\&\min\{\operatorname{rank}(\phi_p(\boldsymbol{Q}_n)),\operatorname{rank}(\phi_p(\boldsymbol{K}_n)^{\mathrm{T}})\}\leqslant\\&\min\{WH,d\}\end{aligned}\tag{6}$$

根据式(6)，线性注意力中相似度矩阵的秩被限制在较低的值，这表明注意力的许多行严重同质化.由于自注意力特征图是同一组$\boldsymbol{V}_n$的加权和，如式(7)所示，在注意力特征图后填加一组深度可分离卷积(DWC)来达到注意力特征增秩的目的.

$$\mathbf{Att}_n=\phi(\boldsymbol{Q}_n)\phi(\boldsymbol{K}_n)^{\mathrm{T}}\boldsymbol{V}_n+\mathrm{DWC}(\boldsymbol{V}_n)\tag{7}$$

$\boldsymbol{V}_n$由$\boldsymbol{F}_{\mathrm{in}}$得到，可用$\mathrm{DWC}(\boldsymbol{F}_{\mathrm{in}})$代替$\mathrm{DWC}(\boldsymbol{V}_n)$的转换作用.

$$\begin{aligned}\mathbf{Att}_n&=\phi(\boldsymbol{Q}_n)\phi(\boldsymbol{K}_n)^{\mathrm{T}}\boldsymbol{V}_n+\mathrm{DWC}(\boldsymbol{F}_{\mathrm{in}})=\\&\left[\phi(\boldsymbol{Q}_n)\phi(\boldsymbol{K}_n)^{\mathrm{T}}+\boldsymbol{M}_{\mathrm{DWC}}\right]\boldsymbol{V}_n\end{aligned}\tag{8}$$

可将DWC视为一种注意力，该注意力中每个$\boldsymbol{Q}_n$只关注特征空间中的若干局部特征，确保即使两个$\boldsymbol{Q}_n$对应的注意力值相同，仍可从不同的局部特征中获得不同的输出.$\mathrm{DWC}(\boldsymbol{F}_{\mathrm{in}})$有满秩的潜力，从而能更好地保证线性注意力特征的多样性.当$\mathrm{DWC}(\boldsymbol{F}_{\mathrm{in}})$用恒等映射实现时，即引入残差，式(8)可以表述为式(9)，由此构成多头深度可分离线性注意力.

$$\mathbf{Att}_n=\phi(\boldsymbol{Q}_n)\phi(\boldsymbol{K}_n)^{\mathrm{T}}\boldsymbol{V}_n+\boldsymbol{F}_{\mathrm{in}}\tag{9}$$

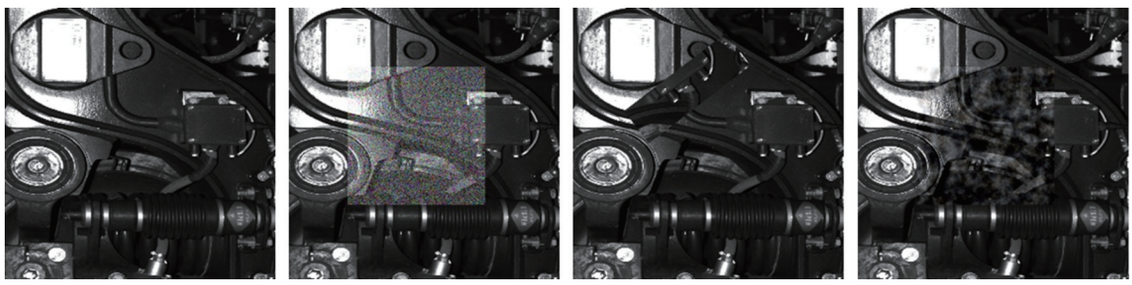
多头深度可分离线性注意力采用深度可分离卷积映射得出$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，通过式(5)来模仿Softmax函数对注意力的尖锐分布调整，在此基础上，引入从输入的映射解决线性注意力的低秩问题来恢复特征多样性.因此，本文利用深度可分离多头线性注意力优化RD网络，使其更好地捕捉到全局范围内的依赖关系，能够更好地理解目标的上下文信息，增强对不同尺度目标的特征表示，提高网络特征能力的同时大幅度提升检测速度.

2.2 伪缺陷机制

学生模型被限制接收缺陷信息是保证反向知识蒸馏在推理过程中有效的前提.仅靠FFCM无法严格限制学生模型接收缺陷信息.因此，如图2所示，采用伪缺陷机制来模拟训练过程中的伪缺陷样本，通过在教师网络中各个模块后面的投影层，同时采用特征重建损失引导投影层从伪缺陷特征中重建正常特征空间，配合FFCM限制缺陷信息流传入学生网络.图4所示为TEDS图像分别加入高斯噪声伪缺陷、随机裁剪伪缺陷、Simplex噪声伪缺陷的效果图，可看出Simplex噪声比高斯噪声产生更自然的缺陷，而随机裁剪能更好地模拟结构缺陷，因此本文采用Simplex和随机裁剪两种噪声模式模拟训练过程中的伪缺陷样本.

2.3 损失函数

在知识蒸馏损失L_{kd}的基础上采用重建特征损



(a)原图像　(b)高斯噪声伪缺陷　(c)随机裁剪伪缺陷　(d)Simplex噪声伪缺陷

图4　三种伪缺陷图像对比

Fig. 4　Comparison of three types of pseudo anomaly images

失L_{Recon}、对比度损失L_{contr}和自监督损失L_{ssot}以提高网络缺陷检测的效果，由此损失函数定义如式(10)所示：

$$L = L_{\text{kd}} + \alpha L_{\text{Recon}} + \beta L_{\text{contr}} + \gamma L_{\text{ssot}} \tag{10}$$

式中:α、β、γ为各损失的权重系数.

2.3.1　知识蒸馏损失

定义$\boldsymbol{\phi}$为从输入图像$\boldsymbol{I}$经教师T和FFCM运算后的特征输出，则T-S模型中的成对特征对应关系如式(11)所示.

$$\{\boldsymbol{F}_{\text{E}}^{k} = E^{k}(\boldsymbol{I}), \boldsymbol{F}_{\text{D}}^{k} = D^{k}(\boldsymbol{\phi})\} \tag{11}$$

其中:E^{k}、D^{k}分别表示T和S中的第k个编码和解码模块,k=1~4;$\boldsymbol{F}_{\text{E}}^{k}$、$\boldsymbol{F}_{\text{D}}^{k} \in \mathbb{R}^{C_k H_k W_k}$,分别表示第$k$个编码模块、解码模块的输出特征，$C_K$、$H_k$、$W_k$分别为第$k$个编码、解码模块输出特征的通道数、高、宽.针对特征$\boldsymbol{F}_{\text{E}}^{k}$、$\boldsymbol{F}_{\text{D}}^{k}$,以式(12)计算它们沿通道轴矢量的余弦相似性损失,得到二维通道缺陷图.

$$\boldsymbol{M}^{k}(w,h) = 1 - \frac{(\boldsymbol{F}_{\text{E}}^{k}(w,h)^{\text{T}} \cdot \boldsymbol{F}_{\text{D}}^{k}(w,h))}{\left\|\boldsymbol{F}_{\text{E}}^{k}(w,h)\right\| \cdot \left\|\boldsymbol{F}_{\text{D}}^{k}(w,h)\right\|} \tag{12}$$

式中:w、h分别为对应特征点的横、纵坐标;$\boldsymbol{M}^{k}(w,h)$值大,表明该位置缺陷严重.

知识蒸馏损失如式(13)所示,通过累积四级编码器、解码器产生的缺陷图获得缺陷值,用于FFCM和学生网络的优化，其中K为编码器个数,此处取K=4.

$$L_{\text{kd}} = \sum_{k=1}^{K}\left\{\frac{1}{H_k W_k}\sum_{h=1}^{H_k}\sum_{w=1}^{W_k}\boldsymbol{M}^{k}(w,h)\right\} \tag{13}$$

2.3.2　重建特征损失

定义$\boldsymbol{F}_{\text{E},i}^{k}\left[k=(1,\cdots,K)\right]$、$\tilde{\boldsymbol{F}}_{\text{E},i}^{k}$分别为输入正常图像$\boldsymbol{x}_i$、伪缺陷图像$\boldsymbol{\xi}(\boldsymbol{x}_i)$经第$k$个编码器模块作用后输出的特征.

$$\boldsymbol{F}_{\text{E},i}^{k} = E^{K}(\boldsymbol{x}_i),\ \tilde{\boldsymbol{F}}_{\text{E},i}^{k} = E^{K}(\xi(\boldsymbol{x}_i)) \tag{14}$$

设ψ_k为第k个教师编码器块的投影层，重建特征损失定义如下：

$$L_{\text{recon}} = \frac{1}{K}\sum_{k=1}^{K}(1 - \cos(\psi_k(\boldsymbol{F}_{\text{E},i}^{k}), \psi_k(\tilde{\boldsymbol{F}}_{\text{E},i}^{k}))) \tag{15}$$

式中:$\cos(\boldsymbol{a},\boldsymbol{b}) = \dfrac{\boldsymbol{a}\cdot\boldsymbol{b}^{\text{T}}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$表示余弦相似度.训练过程中，通过伪缺陷图像将缺陷信号注入特征空间，L_{recon}鼓励投影层学习如何从伪缺陷区域重建正常特征，以此增强投影层在检测过程中抑制缺陷信息的能力.

2.3.3　对比度损失

为了加强投影层在正常图像上的紧凑学习，采用如式(16)所示的具有裕度D的余弦相似度损失从投影层的正常空间中推开缺陷信息，迫使投影层专注于探索正常特征的更深层表示.

$$L_{\text{contr}} = \frac{1}{K}\sum_{k=1}^{K}\max(0, \cos(\psi_k(\boldsymbol{F}_{\text{E},i}^{k}), \tilde{\boldsymbol{F}}_{\text{E},i}^{k}) - D) \tag{16}$$

2.3.4　自监督损失

投影层从各自的教师编码器块接收正常特征，并期望将其投影到紧凑的特征空间中.故采用式(17)所示的自监督损失确保来自正常样本的投影特征表示彼此接近.

$$L_{\text{ssot}} = \frac{1}{Z}\frac{1}{K}\sum_{i,j=1}^{Z}\sum_{k=1}^{K}\cos(\delta(\psi_k(\boldsymbol{F}_{\text{E},i}^{k})), \delta(\psi_k(\tilde{\boldsymbol{F}}_{\text{E},i}^{k}))) \tag{17}$$

式中:δ为Softmax函数;Z为批量大小.

2.4　检测过程及缺陷分数

检测过程如图5所示，根据学生解码器对正常和缺陷样本产生的特征差异的大小,推断缺陷是否存在.输入图像经教师编码器生产四级特征$\boldsymbol{F}_{\text{E}}^{1}$~$\boldsymbol{F}_{\text{E}}^{4}$,逐层特征经投影层PL送入FFCM后经学生解码器生成特征$\boldsymbol{F}_{\text{D}}^{1}$~$\boldsymbol{F}_{\text{D}}^{4}$,两组对称的特征图通过式(12)生成缺陷图$\boldsymbol{M}^{k}(h,w)$,以此反映第$k$个特征图的逐点缺

陷.如式(18)所示,将$\boldsymbol{M}^k(h,w)$上采样到输入图像的大小,并将所有缺陷图按像素求和来获得推理图像中的缺陷分值$\boldsymbol{S}_{\mathrm{A}}$,由此网络可更好地适应多尺度目标的检测,取得比单级特征计算缺陷图更好的检测准确度.

$$\boldsymbol{S}_{\mathrm{A}}=\sum_{k=1}^{K}\lambda(\boldsymbol{M}^k) \tag{18}$$

式中:λ表示双线性上采样操作.为了去除缺陷分值中的噪声,使用高斯滤波器对$\boldsymbol{S}_{\mathrm{A}}$进行平滑,将滤波后$\boldsymbol{S}_{\mathrm{A}}$中的最大值定义为样本级缺陷得分.

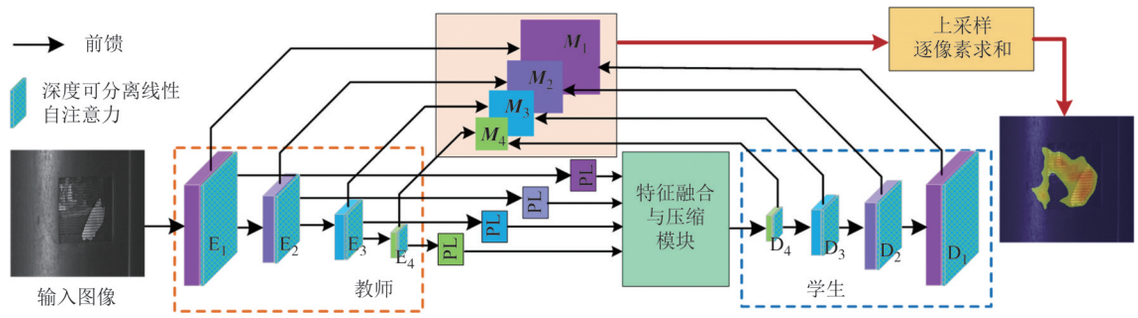


图5 伪缺陷多头深度可分离自注意力反向知识蒸馏网络缺陷检测过程

Fig. 5 Pseudo anomaly multi-head depth separable self-attention reverse knowledge distillation network detection process

2.5 评价指标

为验证所用方法的性能,采用Auroc和Aupro对所采用方法的性能进行评价.Auroc的定义如式(19)所示,通过绘制不同阈值下的真正例率(TPR)和假正例率(FPR)曲线来评估模型的分类能力.

$$\mathrm{Auroc}=\sum_{n=1}^{N}\frac{(\mathrm{FPR}_n-\mathrm{FPR}_{n-1})(\mathrm{TPR}_n+\mathrm{TPR}_{n-1})}{2} \tag{19}$$

式中:N是评估为缺陷的阈值数量;$\mathrm{TPR}_n=\frac{\mathrm{TP}_n}{\mathrm{TP}_n+\mathrm{FN}_n}$,$\mathrm{FPR}_n=\frac{\mathrm{FP}_n}{\mathrm{FP}_n+\mathrm{TN}_n}$分别表示第$n$个缺陷阈值下的真正例率和假正例率;TP即模型正确预测为缺陷的样本数;FN即模型错误预测为正常的缺陷样本数;FP即模型错误预测为缺陷的正常样本数;TN即模型正确预测为正常的样本数.

Aupro通过计算预测的缺陷区域与真实缺陷区域之间的重叠度来评估模型的性能.

$$\mathrm{Aupro}=\sum_{l=1}^{L}\frac{(\mathrm{Recall}_l-\mathrm{Recall}_{l-1})(\mathrm{Precision}_l+\mathrm{Precision}_{l-1})}{2} \tag{20}$$

式中:L是评估的区域总数;$\mathrm{Recall}_l=\frac{\mathrm{TP}_l}{\mathrm{TP}_l+\mathrm{FN}_l}$、$\mathrm{Precision}_k=\frac{\mathrm{TP}_l}{\mathrm{TP}_l+\mathrm{FP}_l}$分别表示第$l$个召回率阈值下的召回率和精确率;TP表示预测区域与真实区域重叠的部分;FN表示真实为缺陷但预测为正常的部分;FP表示预测为缺陷但实际不是缺陷的部分.Aupro分数越高,表示模型定位缺陷区域的性能越好.

3 实验与分析

3.1 训练数据与训练策略

本文将某局TEDS系统获得的3趟列车运行时图像,经人工筛选滤除重复度较高的图像,在剩余1 892张图像中随机抽取3/4作为训练集,将其余1/4的正常图像和收集的划痕、挂物、油迹、螺栓丢失、天线变形、盖板打开等多类缺陷共131张图像作为验证集,对缺陷图像采用PixelAnnotationTool进行了缺陷标注.为提高检测方法对因暗光条件下引起的图像质量退化、列车运行和相机震动引起的运行模糊的适应性,根据前期的研究对图像进行了去运动模糊和暗光增强处理,并将此图像合并入原图像集进行网络的训练.硬件平台使用RTX4070Ti 16G显卡、i7处理器,采用Ubuntu18.04的操作系统,以PyTorch作为深度学习框架进行网络的训练.训练过程优化学生网络、特征融合与压缩模块和投影层的参数.将TEDS图片经随机缩放、随机旋转、随机裁剪等数据增强方式调整为448×448大小进行模型训练,训练100个轮次.教师机采用在ImageNet上预训练好的参数,在训练过程中冻结教师编码器参数.

3.2 实验结果与分析

图6所示为本文所用网络模型训练过程中Sample Auroc、Pixel Auroc、Aupro三个评价指标的变

化曲线.模型在整个训练过程中曲线波动幅度小,在训练30轮次后基本趋于稳定.为最大化检测模型在TEDS缺陷检测任务中的表现,通过调整式(18)所示缺陷值S_A来平衡模型的漏检率和误检率,最终将其值大于0.62的像素点识别为缺陷点.

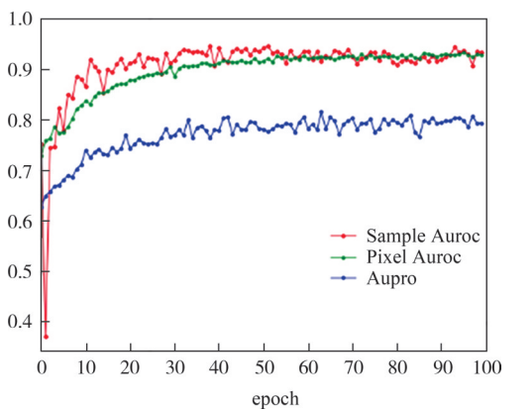


图6　本文方法网络训练过程评价指标曲线

Fig. 6　Evaluation indicators during network training process

在此设置下,采用图7所示的混淆矩阵展示模型在测试数据集上的表现,模型的假正例率为5.8%,召回率为96.1%(漏检率为3.9%),这表明虽然模型能够识别出大部分缺陷样本,但也有部分缺陷样本被识别为正常样本,经分析,漏检为小的螺帽丢失和天线变形两类故障.

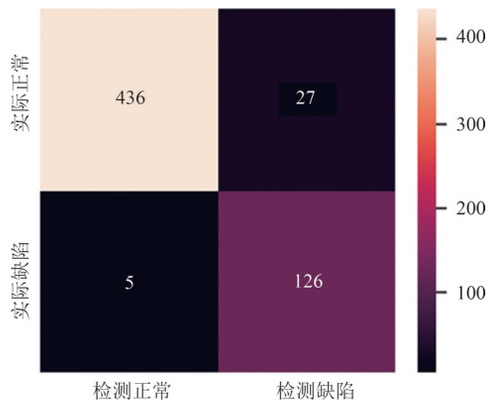
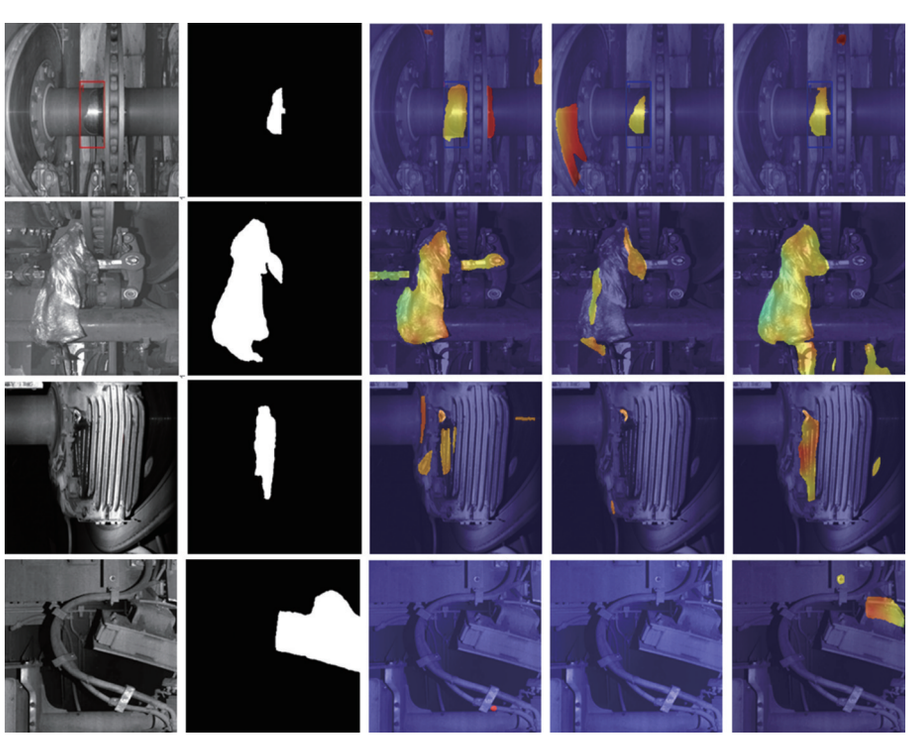


图7　本文方法在TEDS验证集上的混淆矩阵

Fig.7　Confusion matrix on TEDS validation set

为验证本文算法的有效性和先进性,将本文算法与CutPaste[24]、RRD[26]进行对比,图8所示为上述三种算法对TEDS现场图像的缺陷标注效果,第1~3行分别为划痕、挂物和油迹三类结构性缺陷的标注,可看出CutPaste标注超出实际缺陷区域明显,RRD部分区域漏标,部分区域错误的标注为缺陷,而本文所采用方法明显有更少的漏标和错标区域.第4行所示为对天线错位的标注,CutPaste错误的判断为正常,RRD和本文方法错标漏标明显,说明三类方法对错位这类逻辑性缺陷检验普遍效果不佳.



(a)原图　(b)真实标注　(c)CutPaste　(d)RRD　(e)本文方法

图8　各算法对TEDS图像的缺陷标注对比

Fig. 8　Comparison of three algorithms for anomaly annotation on TEDS images

表1列出了图8所示4张图像中各算法标注区域与实际缺陷区域的交并比.本文采用image-Auroc、pixel-Auroc、Aupro和检测速度四种评价指标对本文算法的缺陷检测进行分析,各算法的评价指标如表2所示,本文算法的image-Auroc为94.6%,pixel-Auroc为91.71%,Aupro为80.1%,三种评价指标均明显高于CutPaste和RRD.此外,改进的轻量级网络缺陷检测速度可达0.37 s/张,是RRD的3.1倍.能满足TEDS对计算机硬件成本和检测实时性的要求.

如表2所示,本文算法在MVTec hazelnut数据集上的评价指标均高于RRD,可见本文算法可以获得准确度最好、运算速度更快的网络模型,且可适用于更广义的缺陷检测任务,更加实用.从表2数据可看出三种算法对TEDS的缺陷检测评价指标普遍低于hazelnet,原因在于TEDS图像采集于实际的工业应用场景,检测对象结构复杂,图片采集来源于列车运行中,且受到图像采集时光线的影响,图片存在运动模糊和暗光模糊.

表1 图8所示图像标注区域与实际缺陷区域的交并比

Tab.1 Intersection over Union between the annotated region and the actual anomaly area in Fig. 8

缺陷检测	交并比/%		
	CutPaste	RRD	本文方法
划痕	46	84	92
挂物	73	56	87
油迹	3	64	91
形变	0	0	17

表2 本文算法与其他两种算法缺陷检测的评价指标对比

Tab.2 Comparison of evaluation indicators for various anomaly detection algorithms

算法	image-Auroc/%		pixel-Auroc/%		Aupro/%		TEDS检测速度/(s·张$^{-1}$)
	TEDS	hazelnut	TEDS	hazelnut	TEDS	hazelnut	
CutPaste	84.8	—	76.74	—	64.2	—	1.47
RRD	91.3	99.44	87.92	98.25	76.1	94.99	1.14
本文方法	94.6	99.44	91.71	98.36	80.1	95.11	0.37

为验证各投影层对多级特征图中信息处理的有效性,分别在E_1~E_4各级特征以及更高级特征上加入投影层进行实验.经主干网络提取特征,E_1特征图大小为224×224,经后续模块作用后,特征图大小逐级减半.表3展示了投影层对检测性能的影响.

表3 投影层位置及数量对检测性能的提升

Tab.3 The improvement of detection performance by the position and quantity of projection layers

编号	E_1	E_2	E_3	E_4	pixel-Auroc/%
1					85.32
2				√	86.33
3			√	√	87.43
4		√	√		91.02
5		√		√	90.31
6	√			√	89.92
7		√	√	√	91. 68
8	√	√	√	√	**91.71**

若所有特征图后不加投影层,网络从正常样本中学习特征,则无法有效抑制缺陷信息传入学生网络,此时pixel-Auroc为85.32%,在E_4后加入投影层,并采用L_{recon}鼓励网络从伪缺陷区域重建正常特征,pixel-Auroc上升1.01个百分点,达到86.33%,在E_1~E_3中选两级加入投影层,发现在E_2、E_3这一组合后加入,pixel-Auroc提升幅度最大为5.7%,在E_1~E_4四级特征后均加入投影层,较在E_1~E_3三级特征后加入提升0.03个百分点,最终达到91.71%.此外,在E_4之后的更高级特征上加入投影层进行实验,未观察到明显的性能提升,这表明在中间特征上加入投影层,能有效提升缺陷的标注准确度.在低级特征和高级特征后加入投影层,网络性能提升有限.考虑到加入投影层对网络的检测速度影响不大,本文方法加入了四级投影层,以此鼓励网络在训练中从伪缺陷中推开异常信息,关注重建正常特征.

为验证改进措施的有效性,采用逐步增加改进措施的方式进行如表4所示的消融实验.首先采用MobileNetV2取代原RRD中的ResNet作为主干网络,image-Auroc、pixel-Auroc、Aupro较RRD(ResNet)均有所下降,说明MobileNetV2作为主干网络,参数量大幅度减少,导致网络的表示能力有轻微下降.在Simplex伪缺陷的基础上引入随机裁剪缺陷生成后网络的特性有所提升,说明随机裁剪缺陷生成策略可作为Simplex伪缺陷的补充,进一步提高RRD网络训练过程中对缺陷的抑制能力.在MobileNetV2主干网络的基础上加入传统多头自注意力后,三项指标参数均得到大幅度提升,其中pixel-Auroc提高了2.16个

百分点，Aupro提高2.48个百分点，说明加入自注意力后像素点的误报率、漏报率都有所下降，但加入传统多头自注意力后，平方计算复杂度导致模型检测速率为1.67 s·张$^{-1}$，而在MobileNetV2的基础上加入本文设计的多头深度可分离线性自注意力后，三项指标均与加入传统多头自注意力基本一致，image-Auroc达到94.6%，pixel-Auroc达到91.71%，Aupro达到80.1%，说明多头深度可分离线性自注意力在大幅度提高网络检测速度的同时可取得与传统多头自注意力相当的特征提取能力，提高模型的检测准确度.

可见，各改进措施均能有效提升本文方法的性能.

表4　消融实验分析

Tab.4　Ablation experiments analysis

改进策略	image-Auroc/%	pixel-Auroc/%	Aupro/%	TEDS检测速度/(s·张$^{-1}$)
RRD(ResNet)	91.3	87.93	76.13	1.21
RRD(MobileNetV2)	89.2	84.77	75.55	0.16
随机裁剪缺陷生成	92.1	89.21	77.64	—
传统多头自注意力	94.6	**91.37**	80.12	**1.67**
多头深度可分离线性自注意力	**94.6**	91.71	**80.10**	**0.34**

4　结 论

TEDS现存的缺陷样本少、部件结构复杂且缺陷形态多样，导致既有的故障检测算法误报率高，为进一步提高TEDS系统缺陷检测的性能，设计了多头深度可分离自注意力反向知识蒸馏无监督网络对TEDS图像进行实时缺陷检测研究，对网络的改进如下：

1)采用倒瓶颈残差模块和多头深度可分离线性自注意力模块两种轻量级组件构建以教师-学生架构为骨干网络的反向知识蒸馏模型，有效提高网络检测速度的同时提高网络的表示能力，反向蒸馏结构可有效提升教师、学生网络对缺陷的差异表示.

2)采用Simplex、随机裁剪伪缺陷机制来模拟训练过程中的伪缺陷样本，然后采用重建损失和对比度损失引导投影层从伪缺陷特征中重建正常特征空间，配合FFCM来限制缺陷信息流向学生网络，进一步加大教师、学生网络对缺陷检测的特征差异.

3) 通过多特征融合与压缩模块、多级特征图计算缺陷图、多头深度可分离自注意力，可有效提升TEDS多尺度部件的缺陷检测准确度.

4)改进后的网络image-Auroc、pixel-Auroc、Aupro三个评价指标较其他算法有所提升，可实现对TEDS图像高精度、低误报率的缺陷检测，且可取得较其他算法多倍的速度提升，能满足TEDS系统的实时性检测需求，提升检测人员的工作效率，更好地保障铁路安全运行.

TEDS现存的故障、缺陷样本少，部件结构复杂且缺陷形态多样，图片质量差，导致目前TEDS系统缺陷检测算法误报率高，后续可研究集图像去运动模糊和暗光增强的一体化网络，提升图像清晰度，以提高网络缺陷检测的能力.

参考文献

[1] 黄粤豫，周航，陈业泓，等. 借助弱纹理匹配的TEDS车底故障区域定位算法[J]. 智能系统学报，2024，19(3)：670-678.
HUANG Y Y，ZHOU H，CHEN Y H，et al. TEDS underbody fault location algorithm in virtue of weak texture matching[J]. CAAI Transactions on Intelligent Systems，2024，19(3)：670-678.(in Chinese)

[2] 李萍，吴斌方，刘默耘，等. 基于深度学习的铁路列车关键零部件图像故障检测[J]. 铁道科学与工程学报，2019，16(12)：3119-3125.
LI P，WU B F，LIU M Y，et al. Vision-based fault detection for key components of railway train based on deep learning[J]. Journal of Railway Science and Engineering，2019，16(12)：3119-3125.(in Chinese)

[3] SUN J H，XIAO Z W，XIE Y X. Automatic multi-fault recognition in TFDS based on convolutional neural network[J]. Neurocomputing，2017，222：127-136.

[4] 周雯，史天运，李平，等. 基于卷积神经网络的动车组行车安全图像缺陷检测与分割[J]. 铁道学报，2019，41(10)：76-83.
ZHOU W，SHI T Y，LI P，et al. Defects detection and segmentation of operation safety image of EMU based on convolutional neural network[J]. Journal of the China Railway Society，2019，41(10)：76-83.(In Chinese)

[5] 周李洪，龚金科，李兵. 基于稀疏表示的车用带钢表面图像信息修复[J]. 湖南大学学报(自然科学版)，2021，48(8)：141-148.
ZHOU L H，GONG J K，LI B. Image information restoration of automotive strip steel surface based on sparse representation. Journal of Hunan University(Natural Sciences)，2021，48(8)：141-148.(in Chinese)

[6] MOREL J M, YU G S. ASIFT: a new framework for fully affine invariant image comparison [J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469.

[7] 贾晓芬, 江再亮, 赵佰亭. 裂缝小目标缺陷的轻量化检测方法[J]. 湖南大学学报(自然科学版), 2024, 51(6): 52-62.
JIA X F, JIANG Z L, ZHAO B T. Lightweight detection method for small crack target defects [J]. Journal of Hunan University (Natural Sciences), 2024, 51(6):52-62. (in Chinese)

[8] ZHANG H, SONG Y N, CHEN Y R, et al. MRSDI-CNN: multi-model rail surface defect inspection system based on convolutional neural networks [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 11162-11177.

[9] ZAVRTANIK V, KRISTAN M, SKOČAJ D. Reconstruction by inpainting for visual anomaly detection [J]. Pattern Recognition, 2021,112:107706.

[10] RISTEA N C, MADAN N, IONESCU R T, et al. Self-supervised predictive convolutional attentive block for anomaly detection[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022:13566-13576.

[11] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]//Information Processing in Medical Imaging. Cham:Springer, 2017:146-157.

[12] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks [J]. Medical Image Analysis, 2019, 54: 30-44.

[13] WYATT J, LEACH A, SCHMON S M, et al. AnoDDPM: anomaly detection with denoising diffusion probabilistic models using simplex noise [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-20, 2022, New Orleans, LA, USA. IEEE, 2022: 649-655.

[14] ZHANG H, WANG Z, ZENG D, et al. DiffusionAD: norm-guided one-step denoising diffusion for anomaly detection [EB/OL]. 2023:arXiv:2303.08730. https://arxiv.org/abs/2303.08730.

[15] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018:4510-4520.

[16] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019. Seoul, Korea. IEEE, 2019:1314-1324.

[17] LIANG Y C, LI J Q, ZHU J, et al. A lightweight network for defect detection in nickel-plated punched steel strip images[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72:3505515.

[18] GUO Z X, WANG C S, YANG G, et al. MSFT-YOLO:improved YOLOv5 based on transformer for detecting defects of steel surface [J]. Sensors, 2022, 22(9):03467.

[19] TAO X, ADAK C, CHUN P J, et al. ViTALnet: anomaly on industrial textured surfaces with hybrid transformer [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 5009013.

[20] GUO Y H, JIANG M, HUANG Q H, et al. MLDFR:a multilevel features restoration method based on damaged images for anomaly detection and localization [J]. IEEE Transactions on Industrial Informatics, 2024, 20(2): 2477-2486.

[21] WANG G D, HAN S M, DING E R, et al. Student-teacher feature pyramid matching for anomaly detection [C]//Proceedings of the British Machine Vision Conference 2021. British Machine Vision Association, 2021: 349.

[22] YAMADA S、HOTTA K. Reconstruction student with attention for student-teacher pyramid matching. [EB/OL]. 2021: arXiv: 2111.15376. https://doi.org/10. 48550/ arXiv.2111.15376.

[23] DENG H Q, LI X Y. Anomaly detection via reverse distillation from one-class embedding [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 9727-9736.

[24] LI C L, SOHN K, YOON J, et al. CutPaste: self-supervised learning for anomaly detection and localization [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 9659-9669.

[25] LIU Z K, ZHOU Y M, XU Y S, et al. SimpleNet: a simple network for image anomaly detection and localization [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 20402-20411.

[26] TIEN T D, NGUYEN A T, TRAN N H, et al. Revisiting reverse distillation for anomaly detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023:24511-24520.

[27] PERLIN K. Improving noise[C]//Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, July 23 - 26, 2002, San Antonio, Texas. ACM, 2002:681-682.

[28] CHEN P G, LIU S, ZHAO H S, et al. Distilling knowledge via knowledge review [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021. Nashville, TN, USA. IEEE, 2021: 5006-5015.

[29] CAI Y H, BIAN H, LIN J, et al. Retinexformer: one-stage retinex-based transformer for low-light image enhancement [C]// 2023 IEEE/CVF International Conference on Computer Vision (ICCV), October 1-6, 2023, Paris, France. IEEE, 2024: 12470-12479.

[30] HAN D C, PAN X R, HAN Y Z, et al. FLatten transformer: vision transformer using focused linear attention [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV), October 1-6, 2023, Paris, France. IEEE, 2024:5938-5948.

[31] 王登飞, 苏宏升, 陈光武, 等. 基于聚焦线性注意力Retinexformer的TEDS图像实时暗光增强方法研究[J]. 铁道科学与工程学报, 2024, 21(11): 4840-4850.
WANG D F, SU H S, CHEN G W, et al. Real time low light enhancement method of TEDS images based on focused linear attention Retinexformer [J]. Journal of Railway Science and Engineering, 2024, 21(11): 4840-4850. (in Chinese)