

# 钢铁表面缺陷识别语义分割模型知识蒸馏优化

薛殿龙<sup>1</sup>, 李琳<sup>1</sup>, 周子杰<sup>1</sup>, 常永胜<sup>1</sup>, 向勇<sup>2</sup>, 陈德阳<sup>2</sup>

(1. 中国航空发动机集团有限公司 哈尔滨东安发动机有限公司, 哈尔滨 150000;

2. 重庆大学 机械与运载工程学院, 重庆 400044)

**摘要:** 针对钢铁表面缺陷识别语义分割模型尺寸和计算开销过大等问题, 提出一种以逐通道蒸馏的方式迁移改进复杂网络的缺陷识别知识到简单网络中的方法; 该方法以师生网络为基础, 在训练过程中让学生网络模型模仿教师网络模型的缺陷识别方式, 使用 KL 散度衡量差异, 共同优化后的学生网络模型具有轻量化的模型结构和较高的缺陷识别能力; 在钢铁表面缺陷数据集上的实验验证表明, 蒸馏优化后 ResUnet+ConvNeXt-T 语义分割基准网络对缺陷分割的平均相似系数 mDice 从 0.750 3 提高到 0.771 1, 学生网络模型的推理速度比教师网络模型提升了约 77%, 有效地提高了轻量化的学生网络模型对钢铁表面缺陷的识别能力。

**关键词:** 缺陷分割; ResUnet; ConvNeXt-T; 语义分割; 知识蒸馏

## Knowledge Distillation Optimization of Semantic Segmentation Model for Steel Surface Defect Recognition

XUE Dianlong<sup>1</sup>, LI Lin<sup>1</sup>, ZHOU Zijie<sup>1</sup>, CHANG Yongsheng<sup>1</sup>, XIANG Yong<sup>2</sup>, CHEN Deyang<sup>2</sup>

(1. Harbin Dongan Engine Co., Ltd., AECC, Harbin 150000, China;

2. College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China)

**Abstract:** To address issues such as the large size and computational cost of semantic segmentation models in steel surface defect recognition, a method is proposed to transfer and improve defect recognition knowledge from complex networks to simple networks via channel-wise distillation. Based on the teacher-student network, and during training process, the student network model imitates the defect recognition method of the teacher network model. The KL divergence is used to measure differences, and the jointly optimized student network model has a lightweight model structure and high defect recognition ability. Experimental verification on the steel surface defect dataset shows that after distillation optimization, the average similarity coefficient mDice of the ResUnet+ConvNeXt-T semantic segmentation benchmark network for defect segmentation increases from 0.750 3 to 0.771 1. Compared to the teacher network model, the student network model increases the inference speed by about 77%, effectively improving the recognition ability of lightweight student network models for steel surface defects.

**Keywords:** defect segmentation; ResUnet; ConvNeXt-T; semantic segmentation; knowledge distillation

## 0 引言

钢铁产品表面的缺陷如裂纹、划痕等会严重影响钢铁在使用过程中的性能, 因此, 对钢铁产品的缺陷进行检测识别至关重要<sup>[1-2]</sup>。近年来, 伴随着深度学习技术的发展, 越来越多的工业场景开始引入深度学习技术, 不少研究者使用深度学习代替传统的方法来检测钢铁表面缺陷。基于 U-Net 和 FPN 的编码器-解码器架构的

分割网络<sup>[3]</sup>, 实现图像像素级钢材表面缺陷的自动化检测, 该算法在时间成本和分割精度方面都具有不错的性能。语义分割可以将钢铁表面图像分割成不同的区域, 并将每个像素分配到不同的类别中(背景和各类缺陷), 这使得语义分割技术在钢铁表面缺陷的识别上具有很大的潜力。OCR-UNet 网络实现了金属表面缺陷准确分割<sup>[4]</sup>。一种钢铁表面缺陷分割算法<sup>[5]</sup>, 通过采用稠密连接策略、改进网络损失函数和训练数据增强等方法, 提

收稿日期:2024-08-26; 修回日期:2025-01-14。

作者简介:薛殿龙(1996-),男,硕士,工程师。

引用格式:薛殿龙,李琳,周子杰,等. 钢铁表面缺陷识别语义分割模型知识蒸馏优化[J]. 计算机测量与控制, 2025, 33(12): 182-188.

升了钢铁表面缺陷分割的精度, 但其对样本量较少、尺寸较小的缺陷分割效果有待提升。一种多任务模型用于钢铁表面缺陷分割和缺陷严重性评估<sup>[6]</sup>, 通过模型融合和过程参数的扩展, 提高了缺陷分割的准确性, 不过相较于单任务模型, 该模型相对更加复杂, 不利于实际生产中的实施。一种钢铁表面缺陷检测的链式空洞空间金字塔池化网络<sup>[7]</sup> (CASPPNet), 通过引入改进的全局注意力模块和残差边界细化模块有效提升了缺陷分割的精度。将语义分割应用于缺陷检测和分割任务中, 仍面临着许多挑战, 其中之一便是使用深度卷积神经网络建立的模型的尺寸和计算开销过大, 难以部署在嵌入式设备上运行, 尤其是移动设备等轻便型设备。针对这个问题, 学术界和工业界提出了各种模型压缩的方法, 如剪枝<sup>[8]</sup>、量化<sup>[9]</sup>、高效结构设计<sup>[10]</sup>以及知识蒸馏<sup>[11]</sup>等。知识蒸馏是直接在大容量的教师模型向小容量的学生模型迁移知识来提升性能, 其不需要改变网络结构或做内部操作, 实现起来比较灵活<sup>[12]</sup>。因此, 使用知识蒸馏通过从复杂但缺陷识别精度较高的模型中提取知识来训练简单且更小的模型, 从而减小模型的大小和计算复杂度。论文将首先阐述基于逐通道知识蒸馏的缺陷检测与分割方法, 然后利用钢铁表面缺陷数据集, 开展相关实验来验证本方法的有效性。

1 知识蒸馏方法

知识蒸馏 (Knowledge Distillation) 是一种通过从复杂的大型深度神经网络中提取出知识来训练简单的小型网络的方法, 通过将大型模型的知识迁移到小型模型中, 保证小型模型有接近大型模型的性能, 同时降低模型大小和计算开销, 有利于部署在移动便携式设备中。文献 [11] 提出在神经网络中使用知识蒸馏之后, 知识蒸馏广泛应用于各项深度学习的下游任务中, 典型的知识蒸馏方法如图 1 所示。

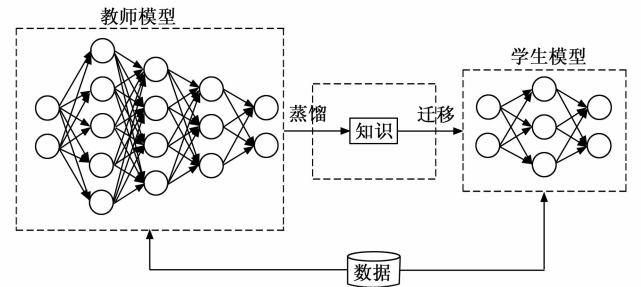


图 1 知识蒸馏方法

知识蒸馏能够将教师模型对数据的拟合能力进行蒸馏, 然后将蒸馏的对数据拟合更有效的知识迁移到学生模型中, 在不增加任何参数和计算开销的情况下增强学生模型对数据的拟合能力, 使学生模型有接近教师模型的性能。

2 缺陷检测语义分割模型逐通道知识蒸馏

2.1 改进的 ResUnet 表面缺陷语义分割模型

ResUnet<sup>[13]</sup>能较好地适应钢铁表面缺陷识别场景, 可用于钢铁表面缺陷的分割。相较于计算机视觉领域的自然图像分割, 工业领域的钢铁表面缺陷分割任务较为简单, 缺陷场景的背景复杂度低, 但缺陷的边缘等细节信息难以捕捉, 不同缺陷之间存在类内差异大且类间不明显的问题, 因此, 缺陷的高级语义信息和低级特征信息都很重要。使用 ConvNeXt<sup>[14]</sup>网络改进 ResUnet 网络的特征特征提取网络, 在解码阶段引入全局上下文模块, 增强网络的全局上下文建模能力, 同时在网络进入分割层之前, 融合解码阶段的多级特征, 增强高级语义信息和低级特征信息的复用度, 能有效提升网络对缺陷分割的性能。改进之后的 ResUnet 网络如图 2 所示。

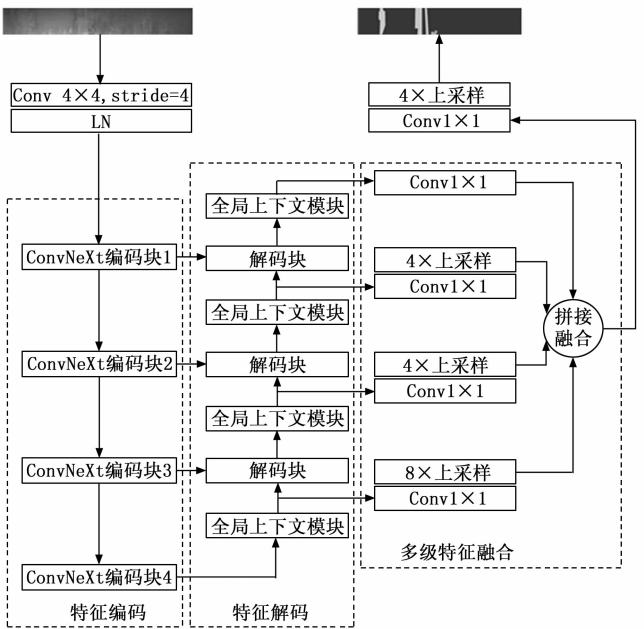


图 2 改进的 ResUnet 网络结构

提出的改进 ResUnet 网络首先使用步长为 4 的  $4 \times 4$  卷积将输入图像下采样 4 倍并使用层规范化 (LN) 操作对数据进行归一化, 这样能降低特征图的大小, 有效降低模型的计算量。紧接着特征图将经过特征编码、特征解码和多级特征融合 3 个阶段, 在特征编码阶段, 利用 ConvNeXt 编码块对特征进行初级编码以获得能够有效表达不同缺陷和背景的特征; 在特征解码阶段, 除了使用和 ResUnet 一样的残差块解码并融合特征之外, 还添加了全局上下文模块, 以增强缺陷分割网络的全局建模能力; 在多级特征融合阶段, 使用  $1 \times 1$  卷积调整各级特征的通道数至一致, 同时将低分辨率的高级语义特征上采样到与最大分辨率特征大小一致, 之后将各级特征通过通道维度拼接的方式组合在一起, 再经过一层

$1 \times 1$  卷积融合这些特征, 获得用于缺陷分割预测的最终特征。最后使用  $1 \times 1$  卷积对特征图进行逐像素预测并上采样到原图大小, 获得与原图逐像素对应的缺陷分割输出图, 上采样方法为双线性插值法。

## 2.2 逐通道知识蒸馏

沿着通道维度对齐教师和学生之间的特征图有助于解决特征错位问题。计算每个通道中激活的平均值<sup>[15]</sup>, 并在分类中对齐每个通道的加权差。计算用于传递知识的所有空间位置 and 所有通道之间的成对关系<sup>[16]</sup>。提出每个通道中包含的信息是通用的<sup>[17]</sup>, 可以在不同的模式之间共享。缺陷检测与分割是典型的密集预测型任务, 逐通道知识蒸馏 (CWD, channel-wise knowledge distillation)<sup>[18]</sup> 是一种适用于密集预测任务的通道蒸馏方法, 它通过计算非对称 KL 散度 (KLD, kullback-leibler divergence) 来最小化教师网络和学生网络之间的差异, 以使简单的学生网络能具有接近教师网络的密集预测性能。KL 散度是衡量两个概率分布  $P$  和  $Q$  之间差异的一种统计度量, 可以用公式 (1) 表示:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \ln \frac{Q(i)}{P(i)} \quad (1)$$

式中,  $P$  和  $Q$  分别为两个概率分布,  $i$  为每一种可能的事件。

CWD 方法能够实现教师网络和学生网络之间对应通道激活图的软对齐。首先, 将教师网络和学生网络的特征图的激活图转换为分布概率图, 以便于使用 KL 散度计算概率距离度量以衡量师生网络之间的差异。使用  $T$  和  $S$  分别表示教师网络和学生网络, 教师网络和学生网络的激活图分别表示为  $y^T$  和  $y^S$ 。CWD 的损失函数通常如式 (2) 所示:

$$\varphi[\phi(y^T), \phi(y^S)] = \varphi[\phi(y^{T_c}), \phi(y_c^S)] \quad (2)$$

$$\phi(y_c) = \frac{\exp \frac{y_{c,i}}{\tau}}{\sum_{i=1}^{W \times H} \exp \frac{y_{c,i}}{\tau}} \quad (3)$$

式中,  $\phi(\cdot)$  用于将激活图转换为分布概率特征图, 转换公式如式 (3) 所示, 其中,  $c=1, 2, \dots, C$  表示特征图的通道,  $i$  表示某个通道特征图的空间位置,  $W$ 、 $H$  分别表示特征图的宽和高,  $\tau$  是一个超参数 (也叫做温度参数), 使用的温度参数越大, 则计算的概率越软, 意味着更关注每个通道更广的空间区域, 反之则更关注每个通道更窄但细节更立体的空间区域。

由于激活图转换为分布概率图使用了 softmax 归一化处理, 能够消除大型且复杂的教师网络和小型且紧凑的学生网络之间量级尺度不同所产生的影响。如果教师网络和学生网络的特征图通道数不匹配, 则需要首先对学生网络的特征图使用  $1 \times 1$  卷积将其特征通道数调整至与教师网络的特征通道数一致。该归一化有助于优化

提升知识蒸馏效果<sup>[19]</sup>。式中的  $\varphi(\cdot)$  函数是为了评估教师网络和学生网络特征通道分布之间的差异, 其计算指标为教师网络和学生网络之间的 KL 散度, 如式 (4) 所示:

$$\varphi(y^T, y^S) = \frac{\tau^2}{C} \sum_{c=1}^C \sum_{i=1}^{W \times H} \phi(y_{c,i}^T) \log \left[ \frac{\phi(y_{c,i}^T)}{\phi(y_{c,i}^S)} \right] \quad (4)$$

从上式可以看出, KL 散度是一种非对称指标, 如果  $\phi(y_{c,i}^T)$  很大, 为了最小化 KL 散度的值,  $\phi(y_{c,i}^S)$  就必须尽可能的大; 反之, 如果  $\phi(y_{c,i}^T)$  很小, 则 KL 散度不需要太注重最小化  $\phi(y_{c,i}^S)$ 。因此, 学生网络的激活图会倾向于生成和教师网络相似的前景显著性分布概率图, 而教师网络的激活图中的背景区域就很少指导学生网络的学习, KL 散度的这种不对称性特性更有利于密集预测任务的知识蒸馏学习。

## 2.3 缺陷检测语义分割模型的逐通道知识蒸馏

改进的表面缺陷分割方法通过在网络配置为 ResU-net+ConvNeXt-T 的基准网络上添加全局上下文模块和融合解码阶段特征的改进策略增强模型的缺陷识别能力。虽然改进策略能提升缺陷分割的精度, 但模型的运行速度相较基准网络有所降低, 因此, 采用 CWD 的方法将改进网络的知识迁移到基准网络中, 使基准网络在保持高图像分割速度的同时能实现与添加改进策略后网络的分割精度相近的性能。图 3 为用于钢铁表面缺陷语义分割模型逐通道蒸馏优化流程。

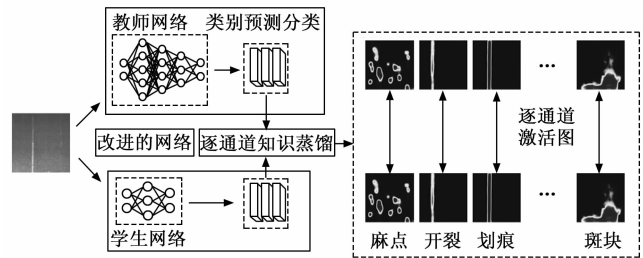


图 3 钢铁表面缺陷语义分割模型逐通道蒸馏优化流程

给定一张用于训练的输入图像, 一方面学生网络需要单独的与标签图像做对比, 以计算分割结果的损失, 另一方面训练好的教师网络能够输出一系列效果较好的特征图, 将学生网络生成的特征图与教师网络对应特征图进行逐通道的比较, 以计算 KL 散度的度量方式衡量两者之间的差异, 其差异即是蒸馏损失, 将此蒸馏损失与学生网络的分割结果的度量损失共同用于优化学生网络, 训练出具有教师网络的强大特征表示能力的学生网络, 采用语义分割模型最后的分类层所生成的特征图作为知识蒸馏时教师网络和学生网络差异对比的特征图。设输入图像为  $X \in R^{C \times W \times H}$ ,  $C$ 、 $W$ 、 $H$  分别表示输入图像的通道数、宽和高, 教师网络和学生网络分别进行前向传播之后生成各自的类别预测特征图  $Y^T$ ,  $Y^S \in$

$R^{Cla \times W \times H}$ ),  $Cla$  表示分类的类别数, 于是, 知识蒸馏的损失可由式 (4) 计算得出, 其中,  $\phi()$  为将激活图转变为分布概率图的函数, 计算公式由式 (5) 所示:

$$L_{KD} = \frac{\tau^2}{Cla} \sum_{c=1}^{Cla} \sum_{i=1}^{W \times H}$$
$$\phi(\text{softmax}(\phi(y_{c,i}^T)) \log \left[ \frac{\phi[\text{softmax}(y_{c,i}^T)]}{\phi[\text{softmax}(y_{c,i}^S)]} \right] \quad (5)$$

$Y_{c,i}^T, Y_{c,i}^S$  分别表示教师网络和学生网络在  $c$  类别相对的特征图位置  $i$  处的特征值。

此外, 学生网络对缺陷分割的预测结果同分割标签比较可计算出基于分割结果评价的使用加权的交叉熵损失  $L_w$ , 由于教师网络是预先训练好的性能较好的模型, 因此, 模型训练时仅更新学生网络的参数, 网络总损失由加权交叉熵损失和知识蒸馏的损失共同组成, 如式 (6) 所示, 以优化学生网络的训练, 使其达到更好的分割结果:

$$L = L_K D + L_w \quad (6)$$

3 实验验证结果与分析

验证实验在 Windows10 64 位的操作系统下进行, CPU 为 Intel Core i9-13900 K, 使用 3.8 版本的 Python 语言环境, 以 Pytorch 版本为 1.13.1 的深度学习框架搭建的网络模型, 后端使用 CUDA 11.8 软件开发工具包来使用 GPU 加速, GPU 加速器使用的是 GeForce RTX 3090, 显存为 24 G。实验数据集使用 Severstal 公司提供在 Kaggle 平台上的钢铁缺陷数据集<sup>[20]</sup>, 将其中有缺陷图像以 8:2 的比例分为训练集和测试集。设置温度参数  $\tau$  为 1, 使用 AdamW 优化函数训练网络, 初始学习率设为 0.000 1, 训练总轮次设为 80 个 epoch, 批处理大小为 8, 训练时通过随机裁剪、水平翻转、垂直翻转、亮度和对比度变换的方式对数据进行增强, 以降低缺陷数据样本较少所带来的影响。本文设置了 3 个验证实验来检验语义分割模型逐通道知识蒸馏优化效果。

3.1 验证实验一

实验一的目的是验证本方法是否能够有效地将改进网络的知识蒸馏出来并迁移到原来的基准网络。实验一中学生网络使用 ResUnet+ConvNeXt-T 的基准网络, 教师网络使用训练好的 ResUnet+ConvNeXt-T+GC+FF 的复杂网络, 蒸馏特征层为对各类别分割预测的最后层特征。图 4 为模型训练过程中蒸馏损失  $L_{KD}$  和总损失  $L$  的变化情况, 可以看到, 蒸馏损失在训练了 20 个 epoch 之后基本趋于稳定, 而总损失则在训练了 50 个 epoch 之后也趋于稳定, 最终网络的蒸馏损失与总损失稳定在 0.2 与 0.4 附近。

表 1 为使用知识蒸馏的策略之后, 网络配置为 ResUnet+ConvNeXt-T 的学生网络在蒸馏前后对钢铁表

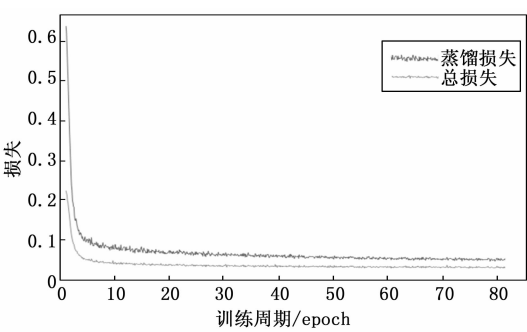


图 4 实验一模型训练时的蒸馏损失和总损失

面缺陷分割的平均相似系数的对比, 以及教师网络和学生网络的参数量与模型推理速度 (FPS 指标) 对比, 从表中对比可以看到, 使用 CWD 方法蒸馏复杂教师模型的知识并迁移到简单的学生模型中能有效提升学生模型的缺陷分割能力, 即使学生模型不包含 GC 和 FF 两种改进策略, 也能通过教师网络有效增强学生模型的上下文建模能力和多尺度缺陷识别能力, 最终使用 CWD 方法增强了学生模型对各缺陷类别分割预测的最终层的表达能力, 其平均相似系数  $mDice$  最终达到 0.763 3, 较蒸馏前的学生模型提升了 1.3 个百分点。此外, 学生网络比教师网络的模型运行速度提升了约 22%。

表 1 实验一模型蒸馏前后平均相似系数( $mDice$ )

网络	$mDice$		参数量/M	FPS
	蒸馏前	蒸馏后		
ResUnet+ConvNeXt-T+GC+FF (教师网络)	0.778 4		36.02	68.0
ResUnet+ConvNeXt-T (学生网络)	0.750 3	0.763 3	33.91	82.7

对是否使用知识蒸馏策略的学生模型对各缺陷分割的相似系数对比如图 5 所示, 由图可以看出, 使用知识蒸馏策略之后, 模型对麻点、开裂、划痕、斑块 4 种缺陷分割的相似系数 ( $Dice$ ) 均有所提升, 证明了 CWD 方法对复杂网络所含改进策略的知识蒸馏的有效性。

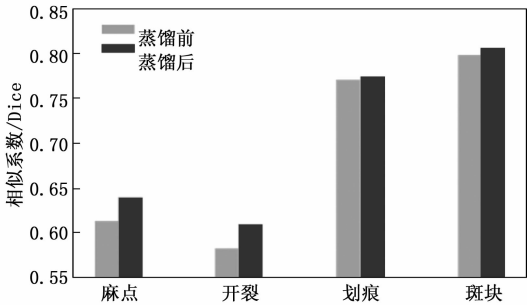


图 5 实验一模型蒸馏前后各缺陷值

3.2 验证实验二

实验二的目的是为了验证 CWD 方法是否能够将更深的骨干网络的特征提取能力蒸馏并迁移知识到较浅的

骨干网络中。实验二中学生网络使用与教师网络分别使用 ResUnet+ConvNeXt-T+GC+FF 和 ResUnet+ConvNeXt-B+GC+FF 的网络配置,即使用改进网络作为学生网络,而将骨干网络改为 ConvNeXt-B 作为教师网络,相较于 ConvNeXt-T 骨干网络,ConvNeXt-B 具有更深的网络层与更多的参数量,其 4 个阶段的 Block 堆叠次数分别为 3、3、27、3,输出特征维度分别为 128、256、512、1 024,其提取的特征具有更好的缺陷描述能力。

实验二的学生模型蒸馏前后的平均相似系数 ( $mDice$ ) 对比如表 2 所示。对实验二而言,蒸馏前的学生模型为改进模型,即便不使用知识蒸馏的策略,其对缺陷分割的指标也能达到 0.778 4,在使用知识蒸馏策略之后,分割性能得到进一步提升, $mDice$  指标达到 0.783 5,证明 CWD 方法通过蒸馏语义分割模型的分层特征的知识能有效的将更深骨干网络的特征表达能力给蒸馏出来传递给骨干网络更浅的学生模型。此外,实验二的学生网络比教师网络的推理速度提高了 46%。

表 2 实验二模型蒸馏前后平均相似系数 ( $mDice$ )

网络	$mDice$		参数量/M	FPS
	蒸馏前	蒸馏后		
ResUnet+ConvNeXt-B+GC+FF (教师网络)	0.785 9		100.98	46.7
ResUnet+ConvNeXt-T+GC+FF (学生网络)	0.778 4	0.783 5	36.02	68.0

表 3 为实验二的学生模型蒸馏前后对各缺陷分割的相似系数 ( $Dice$ ) 的结果对比。由表 3 可以看出,使用网络层数更深的 ConvNeXt-B 作为分割网络的骨干网络,并将此分割网络作为教师网络,原网络作为学生网络做知识蒸馏,能使模型对缺陷数据样本量较少的开裂缺陷的识别有较大提升,使用知识蒸馏的策略之后,模型对开裂缺陷分割的  $Dice$  指标提升了 2.12%,除此之外,该蒸馏策略也能少量的提升模型对划痕缺陷和斑块缺陷分割的值,虽然对于麻点缺陷分割的  $Dice$  值有轻微的下降,但整体上来说,使用所提知识蒸馏策略之后,模型对缺陷的分割性能有较明显的提升。

表 3 实验二模型蒸馏前后各缺陷分割的相似系数 ( $Dice$ )

是否蒸馏	$Dice$			
	麻点	开裂	划痕	斑块
蒸馏前	0.648 4	0.669 2	0.787 6	0.799 8
蒸馏后	0.647 7	0.690 4	0.791 9	0.799 9

3.3 验证实验三

实验三的目的是为了验证 CWD 方法是否能够同时蒸馏较深骨干网络的特征提取能力和网络改进模块的特征表达能力并迁移到简单网络,实验三的学生网络使用

ResUnet+ConvNeXt-T,而教师网络使用 ResUnet+ConvNeXt-B+GC+FF 的网络配置,其余与实验二相同。

实验三的学生模型蒸馏前后的平均相似系数 对比如表 4 所示。观察实验三结果可以看到,使用同时具有改进模块和更深骨干网络的教师网络,能够通过 CWD 方法将学生模型 (网络配置为 ResUnet+ConvNeXt-T) 对缺陷分割的  $mDice$  指标从 0.750 3 提升到 0.771 1,说明本方法可以同时蒸馏更深骨干网络的特征提取能力和网络改进模块的特征表达能力。此外,实验三的学生网络比教师网络的推理速度提高了 77%。综合比较实验三与实验一的结果,可以看出,对相同的学生模型,教师网络越强大,学生网络能获得的蒸馏知识对缺陷的表达能力越强,也即是缺陷分割的  $mDice$  指标越高。

表 4 实验三模型蒸馏前后平均相似系数 ( $mDice$ )

网络	$mDice$		参数量/M	FPS
	蒸馏前	蒸馏后		
ResUnet+ConvNeXt-B+GC+FF (教师网络)	0.785 9		100.98	46.7
ResUnet+ConvNeXt-T (学生网络)	0.750 3	0.771 1	33.91	82.7

表 5 为实验三的学生模型蒸馏前后对各缺陷分割的相似系数 ( $Dice$ ) 的结果对比。由表 5 中结果可以看出,使用以 ConvNeXt-B 为骨干网络,且含改进模块的分割网络模型作为教师网络,蒸馏知识到学生网络 ResUnet+ConvNeXt-T 中,不仅能将更深层网络的特征提取能力传递给学生网络,从而提高学生网络的特征提取能力,还能将改进模块的特征表达能力转化为知识传递给学生网络,进而使网络层数更少,参数量较少,且网络结构更简单的学生网络具备强大的缺陷分割能力,其中,对麻点和开裂缺陷分割的  $Dice$  提升最为明显,蒸馏后分别较蒸馏前提升了 3.32%和 6.57%,划痕缺陷也有 0.52%的少量提升。

表 5 实验三模型蒸馏前后各缺陷分割的相似系数 ( $Dice$ )

是否蒸馏	$Dice$			
	麻点	开裂	划痕	斑块
蒸馏前	0.613 5	0.582 9	0.771 0	0.798 6
蒸馏后	0.6467	0.648 6	0.776 2	0.797 6

3.4 验证实验模型的缺陷分割效果

图 6 为 3 个实验的各个模型对钢铁表面的 4 种缺陷 (麻点、开裂、划痕、斑块) 的分割结果,其中,对比图 (c) (d) (e) 三组分割结果,可以看到未蒸馏的学生模型分割结果较差,出现了断裂的分割结果 (如麻点缺陷),而采用蒸馏策略之后能有效改善这种情况,提

升分割效果；对比 (f) (g) 两组分割结果，可以看到，使用知识蒸馏策略能改善蒸馏前模型对缺陷分割的断裂和空洞现象（如斑块缺陷），使缺陷分割结果与真实情况更接近，总之，采用知识蒸馏策略，可以将复杂模型的缺陷识别能力迁移到简单模型中，提升简单模型的缺陷分割精度。

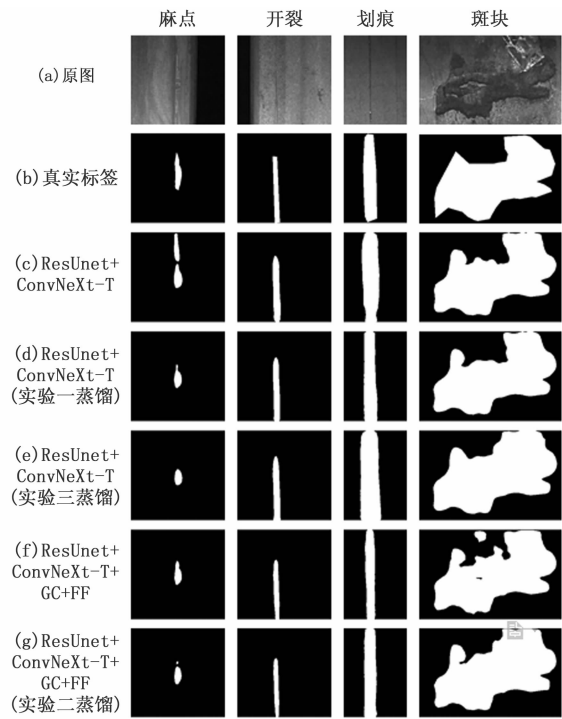


图 6 各模型对 4 种缺陷的分割效果

4 结束语

提出了一种语义分割模型逐通道蒸馏的方式迁移改进复杂网络的缺陷识别知识到简单网络中的方法，从而提高简单基准模型的缺陷检测与分割精度，该方法以师生网络为基础，在训练过程中让学生网络模型模仿教师网络模型的缺陷识别方式，使用 KL 散度来衡量师生网络对应特征层的概率图差异，并使用蒸馏损失和学生网络模型自身训练的损失相结合，共同优化学生网络模型，获得的学生网络模型不仅具有较高的缺陷识别能力，还因为其轻量化的模型结构，使其在识别钢铁表面缺陷时速度更快，更利于工业制造领域的生产和加工。在钢铁表面缺陷数据集上的实验验证了所提方法的有效性。对语义分割模型进行知识蒸馏，有效提高了简单的学生网络模型对钢铁表面缺陷的识别能力，且教师网络越强大，蒸馏之后的学生网络模型的缺陷识别能力通常会更强，最终将配置为 ResUnet+ConvNeXt-T 的语义分割基准网络对缺陷分割的平均相似系数从 0.750 3 提高到 0.771 1，且学生网络模型的推理速度比教师网络模型的推理速度提升了约 77%，有效的提高了结构简

单的学生网络模型对钢铁表面缺陷的识别能力，提高了运行效率，减少了计算资源的消耗。

参考文献：

[1] CAO J G, YANG G T, YANG X Y. A pixel-Level segmentation convolutional neural network based on deep feature fusion for surface defect detection [J]. IEEE Transactions on Instrumentation and Measurement, 2021: 70: 1-12.

[2] YANG J, LI S B, WANG Z, et al. Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges [J]. Materials, 2020, 13 (24): 5755.

[3] QIAN K. Automated detection of steel defects via machine learning based on real-time semantic segmentation [C] // Proceedings of the 3rd International Conference on Video and Image Processing, 2019: 42-46

[4] 陈德阳, 唐智, 何牧耕. 基于 OCR-UNet 的金属表面缺陷分割 [J]. 组合机床与自动化加工技术, 2023 (11): 169-173.

[5] 李原, 李燕君, 刘进超, 等. 基于改进 Res-UNet 网络的钢铁表面缺陷图像分割研究 [J]. 电子与信息学报, 2022, 44 (5): 1513-1520.

[6] NEVEN R, GOEDEME T. A multi-branch U-Net for steel surface defect type and severity segmentation [J]. Metals, 2021, 11 (6): 870.

[7] ZHENG Z, HU Y, ZHANG Y, et al. Casppnet: a chained atrous spatial pyramid pooling network for steel defect detection [J]. Measurement Science and Technology, 2022, 33 (8): 85403.

[8] HAN S, MAO H, DALLY W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding [J]. Arxiv Preprint Arxiv: 1510.00149, 2015.

[9] ZAFRIR O, BOUDOUKH G, IZSAK P, et al. Q8bert: quantized 8bit Bert [C] //2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS), IEEE, 2019: 36-39.

[10] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C] //Proceedings of the European Conference on Computer Vision (ECCV). 2018: 116-131.

[11] HINTON G. Distilling the knowledge in a neural network [J]. Arxiv Preprint Arxiv: 1503.02531, 2015.

[12] 邵仁荣, 刘宇昂, 张伟, 等. 深度学习中知识蒸馏研究综述 [J]. 计算机学报, 2022, 45 (8): 1638-1673.

[13] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-net [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15 (5): 749-753.

- [14] WANG G H, GE Y, WU J. Distilling knowledge by mimicking features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44 (11): 8183–8195.
- [15] ZHOU Z, ZHUGE C, GUAN X, et al. Channel distillation: channel-wise attention for knowledge distillation [J]. Arxiv Preprint Arxiv: 2006.01683, 2020.
- [16] PARK S, HEO Y S. Knowledge distillation for semantic segmentation using channel and spatial correlations and adaptive cross entropy [J]. Sensors, 2020, 20 (16): 4616.
- [17] WANG Y, HUANG W, SUN F, et al. Deep multimodal fusion by channel exchanging [J]. Advances in Neural Information Processing Systems, 2020, 33: 4835–4845.
- [18] SHU C, LIU Y, GAO J, et al. Channel-wise knowledge distillation for dense prediction [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5311–5320.
- [19] LIU X, XU K, ZHOU D, et al. Improved contourlet transform construction and its application to surface defect recognition of metals [J]. Multidimensional Systems and Signal Processing, 2020, 31: 951–964.
- [20] HUANG L, YUAN Y, GUO J, et al. Interlaced sparse self-attention for semantic segmentation [J]. Arxiv Preprint Arxiv: 1907.12273, 2019.
- ~~~~~
- (上接第 173 页)
- [4] 李宜轩, 夏凡, 杨宗谕, 等. 基于深度学习的 HL-2A 装置 ELM 实时控制系统的研制 [J]. 核聚变与等离子体物理, 2023, 43 (4): 380–385.
- [5] 王保华, 范天峰, 荀盼盼, 等. 基于改进型前馈控制的轻型高炮随动控制策略研究 [J]. 火炮发射与控制学报, 2023, 44 (2): 33–38.
- [6] 史琼艳, 杨风波, 张江华. 双缸液压驱动系统灰色预测模糊 PID 控制研究 [J]. 中国工程机械学报, 2024, 22 (5): 572–576.
- [7] 耿少航, 李忠于, 邹玉蓉, 等. 导弹发射车起竖系统建模与仿真分析 [J]. 火控雷达技术, 2024, 53 (1): 103–107.
- [8] 于子翔, 傅中, 胡丹. 采用不对称功率器件布局的双三相直流偏置型同步电机功率变换器设计 [J]. 微特电机, 2023, 51 (3): 7–13.
- [9] 王亚军, 杨立波, 马斌, 等. 虚拟同步机惯量及阻尼系数协调优化方法 [J]. 电力系统保护与控制, 2022, 50 (19): 88–98.
- [10] 甄然, 梁栋伟, 武晓晶, 等. 基于预定性能滑模的四旋翼无人机容错控制 [J]. 计算机仿真, 2023, 40 (12): 99–107.
- [11] 张中琳, 朱明亮, 刘波, 等. 基于伺服电机的辊锻机双轧辊同步控制电气系统设计 [J]. 锻压技术, 2023, 48 (6): 156–161.
- [12] 韩晓庆, 姜媛媛, 刘延彬. 基于凸优化平滑动态窗口法的运动规划方法 [J]. 电子测量与仪器学报, 2023, 37 (8): 234–240.
- [13] 洪学峰, 李承宇, 王勒铭, 等. 多举缸液压起竖系统的同步升举设计 [J]. 计算机仿真, 2023, 40 (6): 321–326.
- [14] 何雅娟, 王恒, 郑圆圆, 等. 海工装备液压缸多重组合式密封结构设计与仿真分析 [J]. 润滑与密封, 2023, 48 (3): 90–95.
- [15] 张翰林, 职彦锋, 张立勇, 等. 考虑负载及误差激励下高速齿轮传动的动力学仿真分析 [J]. 机械传动, 2024, 48 (6): 89–94.
- [16] 刘玄, 方宗德, 雷凤杰, 等. 不对称齿距偏差人字齿轮大周期的三维振动特性分析 [J]. 振动与冲击, 2024, 43 (19): 83–93.
- [17] 何泽银, 易锋, 杨震, 等. 计入冲击与受载齿轮副时变啮合刚度解析算法 [J]. 湖南大学学报 (自然科学版), 2024, 51 (12): 38–46.
- [18] 周琪, 范伟军, 郭斌, 等. 汽车电动助力转向系统负载阻力矩特性多自由度建模仿真分析 [J]. 计量学报, 2023, 44 (6): 907–916.
- [19] 赵宏, 李珈瑞, 刘静. 基于局部时空的多峰优化算法及其在 PID 控制中的应用 [J]. 计算机学报, 2024, 47 (6): 1323–1340.
- [20] 周娟, 林加顺, 吴乃豪, 等. 一种联合 PID 控制与扩展卡尔曼滤波的磷酸铁锂电池荷电状态估算方法 [J]. 电网技术, 2023, 47 (4): 1623–1632.
- [21] 毛志宇, 李培强, 郭思源. 基于自适应时间尺度小波包和模糊控制的复合储能控制策略 [J]. 电力系统自动化, 2023, 47 (9): 158–165.
- [22] 张庭森, 彭忆强, 赵玉友, 等. 前级可变车载充电机双闭环区间二型模糊控制研究 [J]. 控制工程, 2024, 31 (8): 1455–1467.
- [23] 雷佳成, 杨鑫, 董盼, 等. 基于均衡应力分布的高压电缆接头保护装置安全裕度提升方法 [J]. 南方电网技术, 2023, 17 (7): 27–36.
- [24] 陈永淑, 洪海生, 唐娟, 等. 考虑安全裕度的多方竞争电网扩展规划方法研究 [J]. 电测与仪表, 2025, 62 (2): 99–106.
- [25] 刘瑞, 朱姿娜, 赖磊捷, 等. 基于改进蜻蜓算法的磁齿轮复合电机控制研究 [J]. 工程设计学报, 2024, 31 (4): 502–510.
- [26] 孔健, 韩江, 夏链. 基于电子齿轮箱的内啮合强力珩齿同步误差补偿方法研究 [J]. 制造技术与机床, 2024 (2): 105–110.