



### 一种基于知识蒸馏的轨道检测轻量化模型

汤文亮, 曾建杨, 何文晶

引用本文:

汤文亮, 曾建杨, 何文晶. 一种基于知识蒸馏的轨道检测轻量化模型[J]. 重庆理工大学学报(自然科学), 2023, 37(9): 173-179.

### 相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### 融合多信息的个性化推荐模型

Personalized Recommendation Model with Multiple Information Fusion

重庆理工大学学报(自然科学). 2021, 35(3): 128-138 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.03.017](https://doi.org/10.3969/j.issn.1674-8425(z).2021.03.017)

#### 改进样本加权K近邻分类器用于垃圾网页检测

Improved K Nearest Neighbor Classifier with Instance Weighting for Web Spam Detection

重庆理工大学学报(自然科学). 2021, 35(7): 283-290 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.07.035](https://doi.org/10.3969/j.issn.1674-8425(z).2021.07.035)

#### 跨社交网络用户身份关联技术

Matching User Profiles Across Multiple Online Social Networks

重庆理工大学学报(自然科学). 2021, 35(10): 152-162 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.10.020](https://doi.org/10.3969/j.issn.1674-8425(z).2021.10.020)

#### 一种深度置信提升网络集成分类模型

An Integrated Classification Model Based on Deep Belief Network and Gradient Promotion Strategy

重庆理工大学学报(自然科学). 2021, 35(1): 168-179 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.01.022](https://doi.org/10.3969/j.issn.1674-8425(z).2021.01.022)

#### 基于出行链数据的电动汽车充电需求预测模型

An Electric Vehicle Charging Demand Forecast Model Based on Trip Chain Data

重庆理工大学学报(自然科学). 2020, 34(6): 1-8 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2020.06.001](https://doi.org/10.3969/j.issn.1674-8425(z).2020.06.001)



# 一种基于知识蒸馏的轨道检测轻量化模型

汤文亮, 曾建杨, 何文晶

(华东交通大学 信息工程学院, 南昌 330013)

**摘 要:** 针对铁轨缺陷检测神经网络模型参数多、计算量大的问题, 提出了一种基于知识蒸馏的轨道检测轻量化模型及其训练方法, 该网络模型由六层卷积层和三层全连接层构成, 将训练好的 DenseNet 模型作为教师网络, 采用知识蒸馏的方法指导训练, 使得轻量级模型的训练更加简单, 也保证了准确性。在模型的训练阶段加入最小化锐度 SAM 优化算法, 提高了模型的泛化能力, 然后将 VggNet、ResNet、DenseNet 等模型当作对比实验, 评价模型优劣。经过知识蒸馏训练的自定义轻量级模型在铁轨检测数据集中的平均准确率为 97.3%, 且模型参数大小仅为 0.738 M, 均优于其他网络模型, 可部署在众多移动终端中。

**关键词:** 深度学习; 知识蒸馏; 图像增强; 图像识别; 优化算法

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1674-8425(2023)09-0173-07

## 0 引言

我国铁路、地铁里程巨大且逐年上升, 对交通运输、经济发展起着举足轻重的作用<sup>[1]</sup>。铁轨是铁路和地铁的重要组成部分, 在运行过程中, 因为天气、过重的运输负载等原因, 铁轨会随时间受到不同程度的损耗, 包括几何结构缺陷、铁轨部件缺陷、铁轨表面缺陷<sup>[2]</sup>。常见的铁轨表面缺陷包括疤痕、裂纹、波纹擦伤、褶皱、剥落等<sup>[3]</sup>。如果出现此类铁轨缺陷, 要及时进行排查修缮, 不然容易留下隐患, 引发重大事故。

目前, 铁轨检测多采用人工检测, 效率低、成本高, 而且还存在一定的安全隐患。还有一些传

统的轨道缺陷检测方法, 如磁粉法、电涡流法、超声检测法等<sup>[4]</sup>, 这些方法能满足部分的检测要求, 但是容易受到外界环境的干扰<sup>[5]</sup>, 准确率也不是很高。近些年, 随着人工智能的快速发展, 特别是在深度学习和计算机视觉<sup>[6]</sup>领域, 涌现出了很多识别速度快、准确率高的神经网络模型, 大量的学者基于这些模型做了应用和优化<sup>[7-8]</sup>。计算机视觉已经逐渐应用于铁轨缺陷检测领域, 相较于传统方法, 检测精度有了很大提升。戴鹏等<sup>[9]</sup>提出了一种无砟轨道扣件快速定位算法和识别扣件缺陷的半监督机器学习方法, 扣件缺失、弹条折断、弹条移位 3 类缺陷的检出率达到 95%。王子冠等<sup>[10]</sup>基于 MaskR-CNN<sup>[11]</sup>模型, 综合运用 Sobel 算

**收稿日期:** 2022-11-11

**基金项目:** 国家自然科学基金项目(52062016); 江西省重点研发计划项目(20203BBE53034); 江西省教育厅科技重点项目(GJJ190296)

**作者简介:** 汤文亮, 男, 硕士, 教授, 主要从事计算机视觉、网络安全研究, E-mail: 535769575@qq.com。

**本文引用格式:** 汤文亮, 曾建杨, 何文晶. 一种基于知识蒸馏的轨道检测轻量化模型 [J]. 重庆理工大学学报(自然科学), 2023, 37(9): 173-179.

**Citation format:** TANG Wenliang, ZENG Jianyang, HE Wenjing. A lightweight model of track detection based on knowledge distillation [J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37(9): 173-179.

子和 Hough 变换识别图像中的铁轨,对轨道进行区域识别,但是存在识别速度较慢,边缘存在毛刺,影响识别效果的问题。侯博文等<sup>[12]</sup>运用迁移学习的方法,基于残差网络模型,实现对钢轨、扣件及支承块 3 种轨道结构的病害识别。

虽然目前铁轨缺陷检测的人工智能算法已经取得了一定的成果,但大多模型都是通过增加神经网络的层数和参数让模型获得更强的特征提取能力和学习能力,这样导致模型的参数存在巨大的冗余。因此,对网络模型进行压缩是非常有必要的,特别是在对检测时间有严格要求的铁轨检测邻域。Hinton 等<sup>[13]</sup>提出可以用一个大的教师模型去指导小的学生模型,从而达到压缩模型的目的,

称为知识蒸馏。针对上述问题,本文中提出了一个可用于铁轨检测的小型神经网络模型,并利用知识蒸馏的方法和 SAM 优化算法去训练该模型。该模型参数较小,可以在众多移动终端部署,快速地对铁轨进行缺陷检测。

## 1 相关工作

### 1.1 数据处理

铁轨缺陷数据集来自 kaggle 上的 Railway Track Fault Detection 数据集,其将钢轨磨损、扣件脱落、钢轨断裂、移位等病害分为 defective 类,将完好无损、没有缺陷的分为 no defective 类,如图 1 所示。



图 1 轨道缺陷检测数据集

数据集中虽然类别较少,但一个类别中包含几种病害类型,如裂缝、铁枕断裂、道床磨损、路基变形、扣件脱落等在本数据集中都有表现。数据集信息跨度较大,且各个图片的长宽比、像素大小也不一样,这样的数据集对于模型的特征提取能力和泛化能力有很高的要求。因此,在本数据集中,如何在不同病害且特征分布不一的图像中找出它们的共性,并提取出影响结果的关键共同特征,对于一个神经网络模型是巨大的挑战。由于铁路里程巨大,铁路检测的重点是要找出哪一段路线出现了问题,以便及时进行排除和维修,所以能够快速准确地检测出具体哪一段轨道存在安全隐患就显得尤为重要。

原始数据集中分为训练集、测试集、验证集,每一类各有 150、11、31 张图片,由于数据集较少,将它们放在一起进行处理。这些数据用来训练现在的网络模型容易产生过拟合的现象。Wong 等<sup>[14]</sup>提出数据增强对于提高性能和减少过度拟合有更大的好处。因此,使用增加亮度、增加锐

度、改变对比度、横移、翻转和 GAN<sup>[15]</sup> 等数据增强方法对数据集进行扩充。数据集从 192 分别扩充为原来的 10 倍,即 1 920 张图片,再将这些图片按 8:2 的比例分为训练集和测试集。所以有 1 536 张图片作为训练集,384 张图片作为测试集。由以上数据集可以得出,预处理过程中主要有 3 个难点:① 数据集较少,模型容易出现过拟合现象;② 图像的尺寸和拍摄的角度多样,关键特征的坐标位置难以确定,且图像是在自然光下进行获取的,因此出现很明显的阴影,影响模型识别;③ 每类样本中包含多类铁轨病害,图像特征之间有较大差别,如何找出这些图像共同的细小特征是一个巨大的挑战。

### 1.2 SAM 最小化锐度优化算法

对于现在的主流模型,训练损失一般来说是非凸的,有多个局部甚至是全局的极小值,不同的极小值模型的泛化性能也有显著的差异。主流模型大多参数巨大,损失函数在迭代训练的过程中,对模型的泛化能力的保障是很小的。在实际研究

中,如果只追求最小的损失值,很容易导致模型的质量不佳。Foret 等<sup>[16]</sup>提出将损失景观的几何图形和一般化联系起来,引入锐度 sharpness 概念,在训练模型时同时最小化损失函数和损失锐度。SAM 函数可以高效、简便地在具有均匀低损耗值的邻域内寻找参数,极大地提高了模型的泛化能力。

根据 PAC 定理,生成的训练集从分布上看极大概率满足式(1):

$$L_T(w) \leq \max_{\|\xi\|_2 \leq \rho} L_S(w + \xi) + h(\|w\|_2^2 / \rho^2) \quad (1)$$

式中: $h$  是一个单调递增函数。将不等式右边展开为:

$$\left[ \max_{\|\xi\|_2 \leq \rho} L_S(w + \xi) - L_S(w) \right] + L_S(w) + h(\|w\|_2^2 / \rho^2) \quad (2)$$

其中,中括号内就是训练集损失的 sharpness(记为  $L_S^{SAM}(w)$ ),可以衡量从  $w$  移动到相邻参数的差值和训练损失值的增加速度。然后只需要最小化 sharpness,可以通过对内部最大化求微分得到  $L_S^{SAM}(w)$  梯度的近似值,如式(3)所示。

$$\nabla_w L_S^{SAM}(w) \approx \nabla_w L_S(w) |_{w+\xi(w)} \quad (3)$$

最后,将 sharpness 的梯度值进行迭代即可最小化锐度。

### 1.3 知识蒸馏方法

随着大数据和深度学习的快速发展,神经网络模型越来越趋向于深度化。已经有不少学者证明神经网络越深越大,该模型就越容易训练<sup>[17]</sup>。但是在实际部署到用户设备和终端中时,设备可

以存储的空间是有限的,拥有的计算资源也是有限的,参数量巨大的模型很难部署到终端,而且模型计算速度也难以得到保障。模型的压缩引起大家的关注,知识蒸馏是模型压缩的一个分支。知识蒸馏可以通过教师模型来指导小型学生模型的方式,让学生模型在教师模型提取的特征和先验知识的基础上进行迭代训练。学生模型通过知识蒸馏可以很好地解决小型模型难以训练的问题,一般性能可以达到甚至超过教师模型,而且网络结构更加简单,参数更少,模型的计算速度也更快。

知识蒸馏一般需要构建 2 个神经网络模型,即教师模型和学生模型,教师模型是层数深、参数量大、结构复杂的神经网络模型,学生模型是层数浅、参数量小、结构精简的神经网络模型。教师模型一般易训练且表现效果好,而学生模型计算速度快且占用存储资源小。

将数据集输入到教师模型中迭代训练,得到一个较好的结果,再将数据集输入到训练好的教师模型中,将最后一个全连接层输出的向量除以温度( $T$ )后再作 softmax 计算,得到软标签值。将软标签值和真实标签值进行对比,得到 loss1;再将数据集输入学生模型,将得到的预测值和真实值进行对比,得到 loss2,将 2 部分的损失相加,得到知识蒸馏的总损失,对总损失进行梯度下降,调整参数,训练学生模型,流程如图 2 所示。

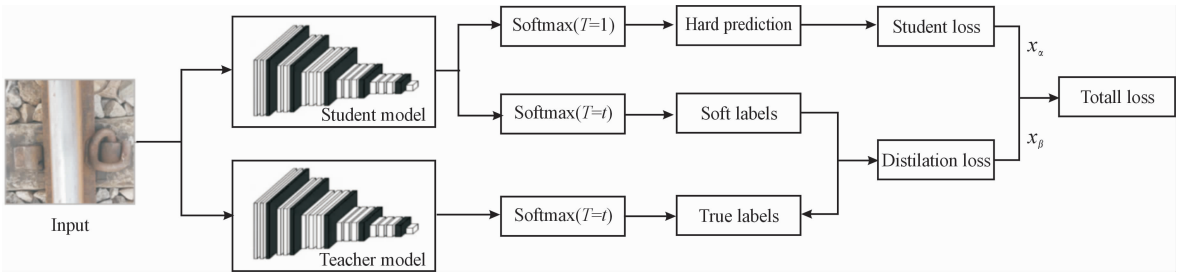


图 2 知识蒸馏训练流程

神经网络模型在进行分类任务时,最后一层的神经元输出信息都会经过一个 softmax 函数,如式(4)所示,它将神经元的输出直接映射成该图像类别的概率分布。在知识蒸馏中,这样经过 softmax 输出的概率信息称为硬标签信息。

$$P_i = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (4)$$

式中: $P_i$ 是教师网络模型输出该图像为第  $i$  类的概率; $x_i$ 是最后一层神经元输出的信息。

由于 softmax 函数只输出概率分布的独热编

码,如果直接用教师模型的 softmax 输出指导学生模型训练,显然会损失教师模型在计算中得到的各类别概率的丰富信息。基于此,知识蒸馏对其进行修改,增加了温度系数  $T$ ,如式(5)所示,使得输出一个“软化”后更加平滑的概率分布向量。

$$P_i = \frac{\exp(x_i/T)}{\sum_j \exp(x_j/T)} \quad (5)$$

式中: $T$ 为温度系数, $T$ 越大,产生的概率就越平滑; $P_i$ 是教师网络模型输出表示该图像为第  $i$  类的概率; $x_i$ 是最后一层神经元输出的信息。

在学生模型训练时用到的总的损失函数是 KLDivLoss,即计算真实标签、教师模型输出的软标签和学生输出硬标签的 KL 散度。KL 散度又叫相对熵,可以很好地衡量 2 个随机分布之间的距离,当 2 个随机分布的差别增大时,它们的相对熵也会增大,文中用到的具体操作如式(6)所示。

$$Loss_{KL}(p, q) = \sum_{i=1}^n p(x_i) (\log p(x_i) - \log q(x_i)) \quad (6)$$

式中: $p$ 是真实标签的处理值; $q$ 是教师模型加学生模型的输出,如式(7)、式(8)所示

$$p = \log \left( \frac{\exp(x_i)}{\sum_j \exp(x_j)} \right) \quad (7)$$

$$q = \alpha * \text{SoftMax} \left( \frac{x_1}{T} \right) + \beta * \text{CrossEntropy}(x_2, y) \quad (8)$$

式中: $\alpha$ 、 $\beta$ 是线性系数; $x_1$ 是教师模型输出; $T$ 是温度系数; $x_2$ 是学生模型输出; $y$ 是真实标签值。

#### 1.4 缺陷检测轻量模型

该模型结构简单,参数量小,称为 SconNet6,包含 6 个卷积层,用于提取图片特征信息,3 个全连接层把特征信息映射到具体的类别,具体模型结构和卷积层参数如图 3、表 1 所示。

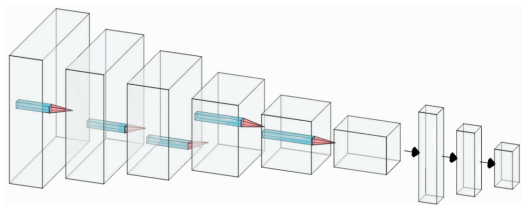


图 3 SconNet6 模型结构

表 1 SconNet6 各卷积层的详细信息

	ker_size	stride	padding	out_channels
Input	3 × 3	1	1	3 × 224 × 224
Conv1	3 × 3	1	1	16 × 112 × 112
Conv2	3 × 3	1	1	24 × 56 × 56
Conv3	3 × 3	1	1	32 × 28 × 28
Conv4	3 × 3	1	1	48 × 14 × 14
Conv5	3 × 3	1	1	64 × 7 × 7
Conv6	3 × 3	1	1	80 × 4 × 4

彩色铁轨图片的数据经过处理,得到一个  $3 \times 224 \times 224$  的张量作为输入。6 层卷积层提取不同维度的特征,全连接层根据特征预测结果。在每个 Conv 中都包含 4 项操作:卷积层、批量归一化处理、激活函数和池化层。

先对输入进行卷积处理,卷积层可以提取图像特征信息,经过卷积后,图像的通道数和尺寸大小都会相应变化。之后对数据进行批量归一化处理,可以将离散的数据变成一个方差为 1 的正态分布,有效加快模型训练时的训练收敛速度,同时防止梯度消失和梯度爆炸,具体操作如式(9)所示。

$$y = \frac{x - E(x)}{\sqrt{\text{Var}(x) + \ell}} * \gamma + \beta \quad (9)$$

将数据输入到激活函数中,激活函数可以加入非线性因素,提高神经网络模型的表达能力。本模型中用到的激活函数是 ReLU 函数,如式(10)所示。

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (10)$$

最后,用最大池化的方法将  $n * n$  个网格中的最大值当作该大网格的值,可以有效提取关键特征和压缩模型参数。

## 2 实验

### 2.1 实验设定

本实验用到的操作系统是 Ubuntu,显卡是显存为 12 GB 的 NVIDIA GeForce RTX 2080Ti,实验中用的语言和主要库版本为 python3.8.5,用python语言实现运行,版本为 3.8.5、pyTorch 1.8.0 + cu111。

## 2.2 模型训练与结果分析

将之前准备好的训练集导入到 VggNet16、ResNet18、ResNet50、DenseNet121 模型进行训练,测试集用来验证模型的性能,以上模型当作对照实验。在训练过程中,将 epochs 设为 100,初始学习率设为 0.001 且随着训练而慢慢减小, batch size 设置为 16。在该实验中使用的损失函数是交叉熵损失函数,优化器是对 Adagrad 算法进行扩展的 Adadelta 优化器。以上 4 个模型训练的测试结果如图 4 所示。

由图 4 可知,4 种经典的网络都在本数据集中表现出不错的识别准确度。除 VggNet16 外,其他模型都在训练到 40 多个 epochs 时开始收敛,识别准确率达到 90% 以上。

但是训练时的 average accuracy 曲线的抖动变

化较大,有可能是训练的学习率过大、batch size 过小,或者是在模型梯度下降时优化算法并没有很好的效果。为此,在上述模型训练时加入 SAM 最小化锐度优化算法,结果如图 5 所示。

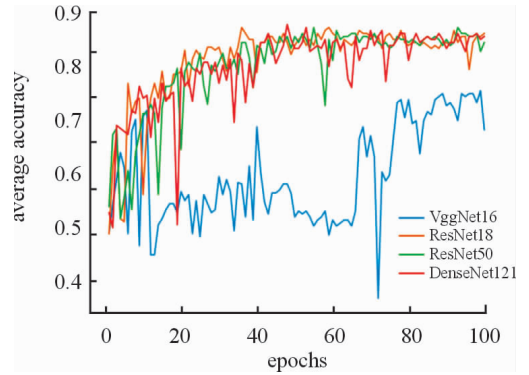


图 4 各模型实验结果曲线

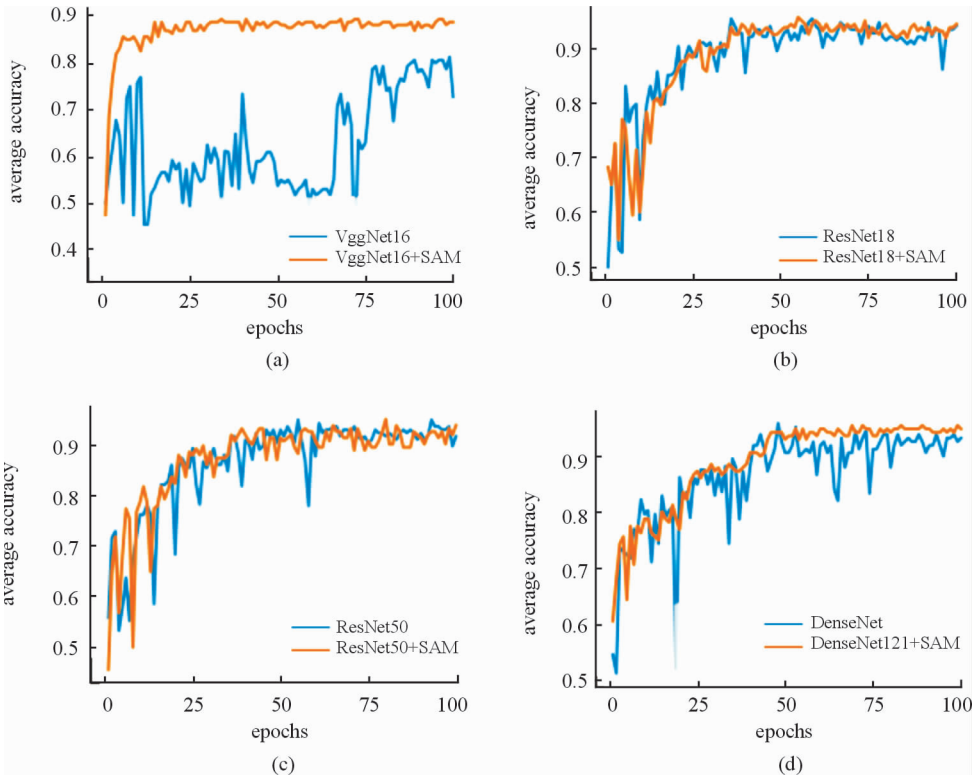


图 5 SAM 算法优化实验结果曲线

由图 5 可知,在加入 SAM 优化算法后,4 个模型相较于传统的优化算法, average accuracy 曲线都变得更加平滑了,说明模型更加稳定,泛化能力也更强,并且在所有模型上的识别准确率也有一些提高。尤其是在 VggNet16 模型的实验结果上,

average accuracy 曲线变得平滑了很多,而且准确率提升了近 10 个百分点。SAM 优化算法在数据集中能够很好地优化梯度下降的过程,提高模型的稳定性和准确率。所以在下文 SconNet6 网络的训练中将使用 SAM 优化算法进行训练。

接下来用上述已经训练好的模型当作教师模型,开始对学生模型 SconNet6 进行训练。先对 SconNet6 进行单独的训练,当作对比实验,再对教师模型做知识蒸馏指导训练,结果如图 6 所示。在本实验中,知识蒸馏的温度系数( $T$ )为 5,全能因子( $\alpha$ )为 0.7,其他设置和之前实验一样。

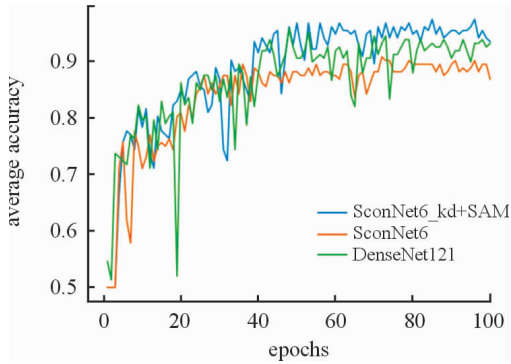


图 6 实验对比结果曲线

如图 6 所示,选取在实验中表现最好的模型 DenseNet121 当作对比,在 SconNet6 单独进行模型训练时,DenseNet121 表现较好,平均准确率高出 5 个百分点,但在 SconNet6 进行训练时,在经过 40 多个 epochs 后模型开始收敛,识别平均准确率超过 DenseNet121。

各模型的实验结果如表 2 所示,在所有的模型中,除了 VggNet16 以外,所有的模型识别平均准确率都超过了 95%,其中采用知识蒸馏方法和 SAM 算法训练的 SonNet6 模型表现最好,且模型大小仅为 0.738 M,仅为 DenseNet 网络的 1/4。

表 2 模型的识别准确率和参数

模型类别	平均准确率/%	模型大小/M
DenseNet16	81.2	512.4
ResNet18	95.1	42.8
ResNet50	95.3	90.7
DenseNet121	95.7	27.1
SonNet6_kd + SAM	97.3	0.738

### 3 结论

针对铁轨缺陷检测模型过大、参数冗余导致

检测时间长和难以部署到移动终端的问题,提出使用知识蒸馏的技术和 SAM 优化算法,训练一个小型的铁轨缺陷检测模型。经过实验证明,通过知识蒸馏的 SonNet6 模型在铁轨数据集中的准确率和各项性能都远超过单独训练时的模型,分类的平均准确率和运算速度超过了其他模型,极小的模型参数也大幅度降低了对计算资源的需求,可以很好地部署在众多存储空间不大的终端上,有很好的应用价值。

### 参考文献:

- [1] 郑拓.我国高速铁路与经济发展研究[J].铁道学报,2020,42(7):34-41.
- [2] 陈良江,周勇政.我国高速铁路桥梁技术的发展与实践[J].高速铁路技术,2020,11(2):27-32.
- [3] 姚盼盼.铁轨表面缺陷机器视觉检测技术研究[D].长沙:湖南大学,2020.
- [4] 刘雄祥.基于卷积神经网络的铁轨表面缺陷识别研究[D].绵阳:西南科技大学,2018.
- [5] 陈建琪,孙勇,张晓敏.高速铁路环境振动影响综述与实测研究[J].工业安全与环保,2019,45(3):73-78.
- [6] HUANG Yuwen, YANG Gongping, WANG Kuikui, et al. Multi-Scale deep cascade bi-forest for electrocardiogram biometric recognition [J]. Journal of Computer Science and Technology, 2021, 36(3): 617-632.
- [7] WANG Xue, LI Zhanshan, KANG Heng, et al. Medical image segmentation using PCNN based on multi-feature Grey Wolf optimizer bionic algorithm [J]. Journal of Bionic Engineering, 2021, 18(3): 711-720.
- [8] ZENG Zheng, WANG Lu, WANG Beibei, et al. Denoising stochastic progressive photon mapping renderings using a multi-residual network [J]. Journal of Computer Science and Technology, 2020, 35(3): 506-521.
- [9] 戴鹏,王胜春,杜馨瑜,等.基于半监督深度学习的无砟轨道扣件缺陷图像识别方法[J].中国铁道科学,2018,39(4):43-49.
- [10] 王子冠,爰国华.基于传统图像处理算法和深度学习的轨道区域识别研究[J].电气自动化,2019,41(4):111-114.
- [11] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2980-2988.

- [12] 侯博文,杨晓,高亮,等. 基于深度残差网络的轨道结构病害识别[J]. 铁道学报,2020,42(8):100-106.
- [13] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science,2015,14(7):38-39.
- [14] WONG S C, GATT A, STAMATESCU V, et al. Understanding data augmentation for classification: when to warp? [C]//2016 International Conference on Digital Image Computing Techniques and Applications. Australia:IEEE,2016:1-6.
- [15] 程显毅,谢璐,朱建新,等. 生成对抗网络 GAN 综述[J]. 计算机科学,2019,46(3):74-81.
- [16] FORET P, KLEINER A, MOBAHI H, et al. Sharpness-aware minimization for efficiently improving generalization [C]//International Conference on Learning Representations. 2021.
- [17] LI Qimai, HAN Zhichao, WU Xiaoming. Deeper insights into graph convolutional networks for semi-supervised learning [C]//AAAI Conference on Artificial Intelligence. New Orleans:AAAI press,2018:3538-3545.

## A lightweight model of track detection based on knowledge distillation

TANG Wenliang, ZENG Jianyang, HE Wenjing

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Aiming at the problems of many parameters and large calculation of the neural network model for rail defect detection, a lightweight model for rail detection based on knowledge distillation and its training method are proposed. The network model is composed of six layers of convolution layer and three layers of full connection layer. The trained DenseNet model is used as the teacher network to guide the training with the method of knowledge distillation, which makes the training of lightweight model simpler and ensures its accuracy. In the training phase of the model, SAM optimization algorithm with minimum sharpness is added to greatly improve the generalization ability of the model. Then VggNet, ResNet, DenseNet and other models are used as comparative experiments to evaluate the model. The average accuracy of the customized lightweight model trained by knowledge distillation in the rail detection data set is 97.3%, and the model parameter size is only 7.38 M, which is superior to other network models and can be deployed in many mobile terminals.

**Key words:** deep learning; knowledge distillation; image enhancement; image recognition; optimization algorithm

(责任编辑 王欢)