

# 基于 Swin Transformer 的沥青路面病害分类检测研究

郭晨<sup>1</sup>, 杨玉龙<sup>1</sup>, 左琛<sup>2</sup>, 杨冰鑫<sup>1</sup>

(1. 长安大学 信息工程学院, 西安 710064; 2. 长安大学 运输工程学院, 西安 710064)

**摘要:** 针对传统卷积神经网络模型在沥青路面病害检测中识别长距离裂缝结构能力不足以及面临的精度局限问题, 引入 Swin Transformer 模型进行沥青路面病害分类研究; 首先对于路面检测车采集到的沥青路面扫描图像对比度低的问题, 使用直方图均衡技术处理图像, 增加图像可视化效果; 其次, 选取 3 种经典卷积神经网络模型作为对比模型, 并在训练过程中采用更换损失函数, 调整预训练模型等手段解决过拟合问题; 并选用准确率、查全率、F1-score 作为评价指标; 在最终实验结果中 Swin Transformer 识别准确率达到 80.6%, F1-score 达到了 0.776, 不仅在整体分类准确率上超越了传统 CNN 模型, 并且对具有长距离特征结构的病害方面具有更高的识别准确率, 同时具有良好的可靠性。

**关键词:** Swin Transformer; 路面病害检测; 卷积神经网络; 图像分类; 图像处理

## Research on Classification and Detection of the Asphalt Pavement Diseases Based on Swin Transformer

GUO Chen<sup>1</sup>, YANG Yulong<sup>1</sup>, ZUO Chen<sup>2</sup>, YANG Bingxin<sup>1</sup>

(1. School of Information Engineering, Chang'an University, Xi'an 710064, China;

2. School of Transportation Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** Aiming at the problems of insufficient ability of identifying long-distance crack structure and accuracy limitation of the asphalt pavement disease detection in traditional convolutional neural network models, the Swin Transformer model is introduced to study the classification of the asphalt pavement disease. Firstly, for the problem of low contrast of the asphalt pavement scanning image collected by the road inspection vehicle, the histogram equalization technology is used to process the image, and increase the image visualization effect. Secondly, three classic convolutional neural network models are selected as comparison models, and the methods of replacing the loss function and adjusting the pre-training model are used to solve the over-fitting problem during the training process. And the accuracy rate, recall rate, F1-score are selected as the evaluation index. In the final experimental results, the recognition accuracy of the Swin Transformer reaches 80.6%, with the F1-score of 0.776, which not only surpass the traditional convolutional neural network (CNN) model in overall classification accuracy, but also has a higher recognition of diseases with long-distance characteristic structures and good reliability.

**Keywords:** Swin transformer; road damage detection; CNN; image classification; image processing

## 0 引言

沥青路面是当前城市交通中最常见的路面类型, 被广泛应用于城市道路、高速公路等交通基础设施。然而, 由于受到长期的车辆负荷、自然环境等因素的影响, 沥青路面容易出现各种病害<sup>[1]</sup>。交通运输部 2018 年 12 月发布的《公路技术状况评定标准》<sup>[2]</sup>, 将沥青路面损坏分为龟裂、块状裂缝、纵向裂缝、横向裂缝、沉陷、车辙、波浪拥包、坑槽、松散、泛油等 10 种类型。这些病害严重影响着沥青道路的正常使用寿命, 不仅导致路面使用寿命缩短, 更会导致交通事故风险增加<sup>[3]</sup>。其中裂缝损坏和坑槽是常见且对路

面正常使用影响较大的病害类型。裂缝不仅会破坏沥青路面结构, 导致路面整体强度降低, 并且会使得雨水和油污渗透到路面中, 加剧路面老化和破损, 进而使得裂缝发展成更为严重的病害。坑槽损坏使得行车舒适性下降, 严重的坑槽损坏直接增加了发生交通事故的风险。因此对于裂缝和坑槽的精准识别分类是十分重要的, 可以有效提高交通安全性, 延长路面使用寿命, 降低维护成本。

早期路面病害检测主要依赖于人工实地检查。这种方法虽然能够较为准确发现病害, 然而费时费力, 影响公路交通正常运行<sup>[4]</sup>。以灰度阈值分割、多尺度以及数字形态

收稿日期: 2023-08-04; 修回日期: 2023-08-18。

基金项目: 国家自然科学基金(41874140)。

作者简介: 郭晨(1984-), 女, 博士, 教授。

通讯作者: 左琛(1991-), 男, 博士, 讲师。

引用格式: 郭晨, 杨玉龙, 左琛, 等. 基于 Swin Transformer 的沥青路面病害分类检测研究[J]. 计算机测量与控制, 2024, 32(2): 114-121.

学为主的数字图像处理检测算法可以在图像中提取出路面病害特征, 但是这些算法泛化能力不足, 难以满足实际需求<sup>[5]</sup>。随着计算机技术的发展, 传统机器学习方法被引入到路面病害检测中。这些方法通常基于手工设计的特征提取和分类算法, 例如支持向量机 (SVM)、K 近邻 (KNN) 等<sup>[6]</sup>, 提高了检测效率, 并且通过合理的特征选择和算法优化, 可以实现一定程度的病害分类准确性。然而这些方法特征提取依赖于人工经验, 往往需要大量领域知识和实验调优, 工作量较大, 对于复杂多样的路面病害的泛化能力有限, 很难提供高度准确的检测结果<sup>[7]</sup>。

随着计算机硬件的发展, 尤其是 GPU (图形处理单元) 的广泛应用以及深度学习算法的不断优化, 推动了基于深度学习的路面病害检测的发展<sup>[8]</sup>。卷积神经网络 (CNN, convolutional neural network) 架构于 1998 年由 LeCun<sup>[9]</sup> 等人提出, 卷积神经网络的优点在于可以从原始像素数据中学习抽象的特征表示, 实现对复杂图像的高级语义理解。2019 年, 沙爱民等<sup>[10]</sup> 提出了一种包含 3 个 CNN 的级联模型, 进行路面裂缝和坑槽检测。其中, 第一个 CNN 判定路面病害类型, 其余神经网络关注裂缝和坑槽的几何结构提取。实验结果表明, 级联 CNN 对路面裂缝与坑槽的识别与测量具有运算高效、结果精准等优势。2020 年, 陈泽斌等<sup>[11]</sup> 针对复杂背景下路面裂缝, 对 UNet 网络上采样过程的特征融合阶段进行改进, 进一步提升 CNN 的泛化能力和识别精度。为提升路面表观病害检测效率与精度, 2021 年晏班夫等人<sup>[12]</sup> 提出了基于 Faster R-CNN, CNN 与形态学的集成方法对路面病害进行快速识别, 相较于传统方法提升了效率。然而, 卷积神经网络主要依赖局部卷积操作来提取特征, 导致其本身缺乏全局依赖性, 对于长距离图像特征的提取准确率低下。另一方面 CNN 的性能依赖层级结构和参数设置, 导致其面临精度局限。

Transformer 模型在自然语言处理中的成功证明了注意力机制<sup>[13]</sup> 在建模长距离依赖和全局上下文方面的卓越能力。研究人员逐渐开始探索将其应用于计算机视觉领域。2020 年 Dosovitskiy 等人提出的 ViT (vision transformer) 模型<sup>[14]</sup> 取得了突破性成功, 在多个数据集的分类任务上表现出良好的性能。然而 ViT 全局注意力机制由于需要计算图像每个像素之间的关联性, 导致计算资源增加, 为处理高分辨率图像和大规模数据集带来挑战。在 ViT 此基础上, 研究人员设计了 Swin Transformer<sup>[15]</sup>。Swin Transformer 采用了层级注意力机制, 在保留长距离建模优势的基础上降低了计算资源的需求, 并且 Swin Transformer 还表现出强大的扩展能力。

2023 年 LUO 等人<sup>[16]</sup> 为了针对传统神经网络容易丢失边缘细节信息、对于不规则的路面裂缝形状检测效果不佳等问题, 提出了名为 STrans-YOLOX 的路面裂缝检测结构。该结构首先利用 CNN 进行特征提取, 然后通过引入 Swin Transformer 增加长距离建模能力。实验结果表明该结构在 RDD2020 数据集上的 mAP 值达到了 63.37%, 超越了最先

进的模型。

由此可见 Swin Transformer 在路面病害检测领域具有良好的应用前景, 为路面病害检测领域提供了新的思路 and 方向。然而目前基于 Swin Transformer 的路面病害检测所用的数据采集手段为车载相机拍摄或者手机拍摄<sup>[16-17]</sup>, 这样的图像采集方式成像质量不一, 且图像中非路面因素多, 对其进行分类或者分割任务相对比较困难。本文使用的数据集为路面检测车在我国高速公路的实际采集数据, 具有图像质量高, 没有与路面无关因素干扰等特点<sup>[18]</sup>。

文献 [19] 的研究成果初步表明, 与 CNN 结合的 ViT 混合模型, 在长距离路面裂缝的识别上具有一定的优势, 但是在纵向裂缝的识别能力上有待提高, 对于坑槽病害的检测能力尚未可知。并且纯 Transformer 模型对于路面病害的识别能力也有待研究。因此本文基于路面检测车采集图像引入纯 Swin Transformer 模型进行沥青路面病害分类检测研究。并与 3 种经典的 CNN 模型进行对比, 探索 Swin Transformer 在长距离路面病害结构上的识别优势。针对实验过程中出现过拟合问题, 采用更换损失函数、调整初始模型率等方法解决, 实验结果表明, Swin Transformer 不仅对于横向裂缝, 纵向裂缝等长距离裂缝结构具有良好的识别精度, 并且对于龟裂、坑槽等病害也有良好的识别效果, 整体的分类准确率也要优于 CNN 模型。通过计算各模型的 F1-score, 表明 Swin Transformer 相比于 CNN 模型具有更好的可靠性。

## 1 实验模型

### 1.1 Swin Transformer 模型

使用 Swin Transformer 模型进行路面病害图像分类如图 1 所示。Swin Transformer 的网络结构由 4 个阶段 (Stage) 组成, 每个 Stage 包含若干个 Swin Transformer Block, 本文所采用的 Swin\_base 模型在 4 个阶段分别有 2, 2, 18, 2 个 Swin Transformer Block。对于输入的路面病害图像, Swin Transformer 在图像分割层使用  $4 \times 4$  的窗口对图像进行分割, 分割后的窗口图像在通道方向展平<sup>[20]</sup>。原始图像的高和宽缩减为原来的  $1/4$ , 通道数变为原来的 16 倍。在首个 Stage 中通过线性映射将图像块转换为二维向量输入到 Swin Transformer Block 中。在后续每个 Stage 都通过图像块拼接层对图像块进行下采样。最后的输出经过一个全局池化层和一个全连接层得到病害的分类结果。

两个 Swin Transformer Block 结构如图 2 所示, Swin Transform Block 都是成对出现的。这是由于在第一个 Block 中 Swin Transformer 将原始 Transformer 中的多头自注意力 (MSA, multi-head self-attention) 模块替换成了窗口自注意力模块 (W-MSA, window multi-head self-attention)。W-MSA 不计算整幅图像中所有的像素之间的注意力, 而是将图像划分为若干个窗口, 对于每个窗口只计算窗口内部图像像素之间的注意力, 而不与其他窗口进行交互。单头窗口注意力的计算公式如下:

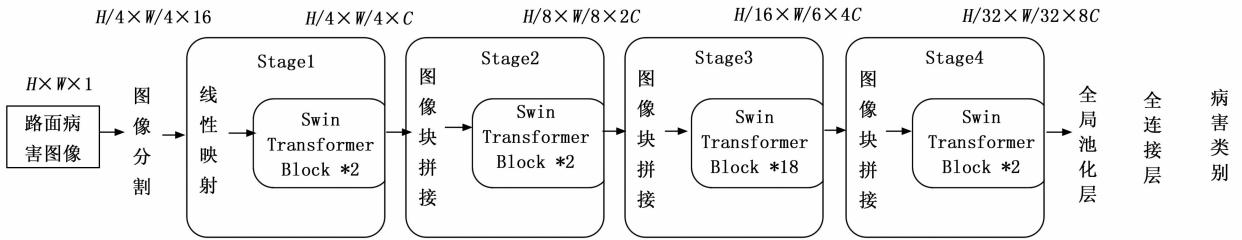


图 1 使用 Swin Transformer 进行路面病害图像分类

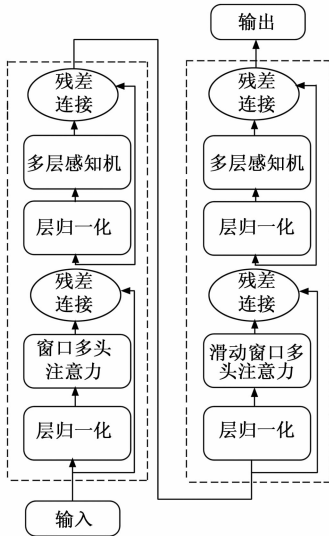


图 2 两个连续的 Swin Transformer Block 结构

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \quad (1)$$

式中， $Q$ 、 $K$ 、 $V$  分别代表查询矩阵  $Q$  (Query)、键矩阵  $K$  (Key) 和值矩阵  $V$  (value)。它们是由输入特征进行线性映射后得到的。假设输入特征  $X$  的形状为  $[B, H, W, C]$ 。 $B$  代表一次输入数量， $H, W, C$  分别代表输入特征的高、宽和通道数。则  $X$  分别与查询权重矩阵  $W^Q$ 、键权重矩阵  $W^K$  和值权重矩阵  $W^V$  相乘得到  $Q, K, V$ 。 $B$  代表相对位置编码，用来将注意力的计算范围限制在划定的窗口之内。 $d^k$  表示键矩阵  $K$  的维度，用来帮助模型在不同注意力头之间捕获不同层次特征关系。多头注意力机制是将多个单注意力模块级联起来，每个注意力头可以捕获不同特征之间的依赖关系，允许模型在多个抽象层次上对输入特征进行编码。多头窗口自注意力公式表示为：

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \dots, head_n)W^o \quad (2)$$

$$head_i = Attention(Q_i, K_i, V_i) \quad (3)$$

式 (2) 中， $W^o$  是一个可学习的权重矩阵，用于将注意力头的输出进行线性变换，使得最终的输出与输入特征具有相同的维度。

只在窗口之间计算注意力虽然降低了计算量，但是由于窗口和窗口之间没有信息的交互导致无法获取图像整体的信息。因此 Swin Transformer 在紧跟着下一个 Block 中采用滑动窗口注意力机制 (SW-MSA, shifted window based multi-head self attention)，SW-MSA 在 W-MSA 划分

的窗口的基础上将所有窗口通过循环移位的方式进行移动，在移动后的窗口内部再进行注意力计算，将各个窗口的信息重新组合为新的图像表示。使得在计算量不增加的情况下，每个窗口之间有了信息的交互，从而可以获得到图像的全局信息。因此在 Swin Transformer 中 W-MSA 和 SW-MSA 是两两交替出现，

除了注意力机制的不同，两个连续 Swin Transformer Block 中其他组件都是完全相同的，包括两个归一化层 (LN, layer normalization)，两个残差连接层和一个多层感知机层 (MLP, multilayer perceptron)。

### 1.2 对比模型

为了客观全面地评估 Swin Transformer 在路面病害图像分类任务上的表现。本文选择 ResNet101, Densenet161 和 EfficientNet\_b3 三个模型作为对比模型。

ResNet 由微软研究院的 He 等人<sup>[21]</sup>提出，是深度学习中的一种经典模型。其关键创新是引入了残差连接 (Residual Connection)，使得网络能够更容易地训练深层模型，有效地解决了传统深层网络中的梯度消失和梯度爆炸问题。ResNet101 是 ResNet 系列中较深、较复杂的一个模型，总共有 101 层。适用于计算机视觉任务中的图像分类、语义分割等任务。

DenseNet 是由 Huang<sup>[22]</sup>等人提出的一种深度学习架构。它的核心思想是在网络中引入密集连接，以解决深度神经网络中的梯度消失和梯度爆炸问题，同时改进了模型的训练效率和准确性。是深度学习领域重要模型之一。DenseNet161 是 DenseNet 的一个变体，每个 Dense Block 中包含 161 层。模型深度和参数量相对较大，在计算机视觉任务上具有强大的特征提取能力。

EfficientNet 是由谷歌研究人员 Tan 等人<sup>[23]</sup>提出的一种神经网络架构。它的核心思想是通过网络结构进行均衡缩放来实现高效且优秀的性能。已经在图像分类，目标检测和语义分割等计算机视觉任务上取得了显著成绩。EfficientNet\_b3 是其中一个较大规模的变体。

这些模型在多个数据集和任务上都经过充分验证，是深度学习领域经典且性能强大的模型，因此可以提供一个相对稳定的基准。

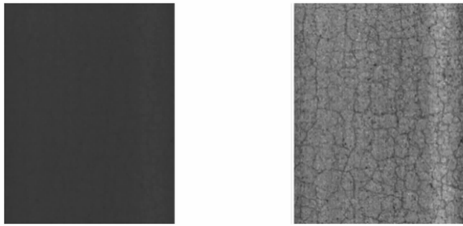
## 2 实验数据

本文数据采集地点位于我国辽宁省某段高速公路，由路面检测车进行采集。共采集沥青路面图像 2 000 张，图像

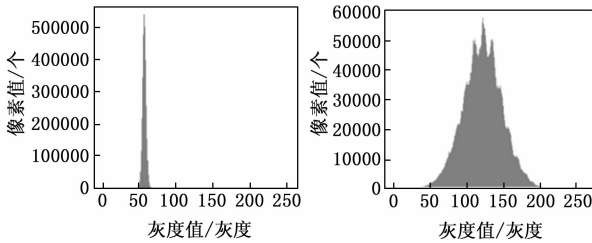
尺寸为 1 590×2 048。

### 2.1 图像预处理

路面检测车采集到的原始图像如图 3 (a) 所示, 由于采集设备能力的限制, 原始图像整体呈现为灰黑色, 对比度较低, 从视觉效果上很难分辨出路面病害的类型。其灰度直方图如图 3 (c) 所示, 灰度值主要集中在 50~60 之间。这样的图像对于深度学习模型的训练是十分不利的。因此需要一种能够增强图像对比度的方法来对原始图像进行处理, 增强图像可视化效果。



(a) 原始图像 (b) 直方图均衡后的图像



(c) 原图的灰度直方图 (d) 直方图均衡后图像灰度直方图

图 3 直方图

在图像处理领域常见的增强图像对比度方法有直方图均衡<sup>[24]</sup>、直方图拉伸<sup>[25]</sup>、直方图规定化<sup>[26]</sup>和小波变换<sup>[27]</sup>等。其中直方图拉伸需要手动调整参数、适用性窄, 并且容易导致图像亮度过于增强。直方图规定化需要目标直方图来进行匹配, 复杂度较高, 容易受噪声干扰。小波变换方法参数选择比较困难, 容易造成信息损失导致图像细节不够清晰。相对于以上 3 种图像处理方式, 直方图均衡技术具有能够自动增强图像对比度, 无需人工干预、避免像素浪费、可以更好地保持图像亮度平衡以及适用性广泛等优势。

因此, 选用直方图均衡技术对原始图像进行处理。它的核心思想是以像素灰度值的累积分布函数作为转换函数, 重新映射图像的灰度值, 使得图像的像素灰度值在整个灰

度范围内的分布更加均匀, 从而增加图像对比度。

经过直方图均衡后的图像如图 3 (b) 所示, 可以看出经过处理后的图像可视化效果得到很大增强。图 3 (d) 展示了其灰度直方图, 相较于原图, 处理后图像灰度值的分布更加均匀。

### 2.2 数据集划分

为了增加数据样本的多样性并扩充数据集, 使用图像裁剪将图像一分为四。并且为了契合在实际中的应用, 对于裁剪后的图像采用图像处理方法重新 resize 为原始图像的大小。然后进行数据清洗, 剔除掉不能使用的数据, 共有图像 7 650 张。按照《公路技术状况评定标准》中对于沥青路面损坏评定标准, 将 7 560 张图像分为横向裂缝、纵向裂缝、块状裂缝、龟裂、坑槽及无病害共 6 种类型。各病害图像如图 4 所示。

对于各病害类别图像按照 8 : 1 : 1 的方式划分训练集、验证集和测试集。划分后各数据集病害数量如表 1 所示。

表 1 各病害类别图像数量

	横向裂缝	纵向裂缝	块状裂缝	龟裂	坑槽	无病害
训练集	216	816	718	2 136	128	2 040
验证集	26	101	89	267	16	254
测试集	26	101	89	267	16	254

## 3 实验及评价指标

### 3.1 评价指标

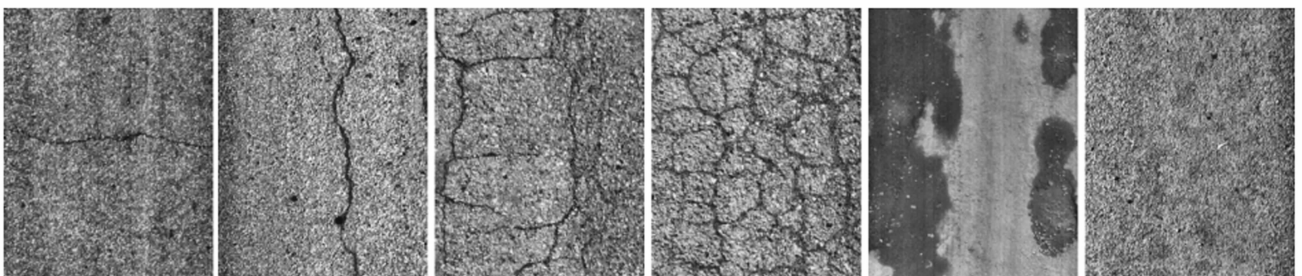
选用准确率 (accuracy), 查全率 (recall) 和 F1-score 这 3 种常用的图像分类评价指标来评估模型的性能。

准确率是最简单直观的评价指标, 代表模型正确分类的样本比例, 准确率越高表示模型在数据总体上的分类表现越好。计算公式表示为: 准确率 = 正确分类的样本数 / 所有的样本数。

查全率衡量模型正确预测为正例的样本占所有正例的比例, 强调模型正确捕捉到所有实际正例的能力, 查全率越高说明模型识别实际正例的能力越高。查全率的公式为:

$$Recall = TP / (TP + FN) \quad (4)$$

其中:  $TP$  和  $FN$  是混淆矩阵中的基本术语,  $TP$  表示真正例, 指模型正确地预测为正例的样本数;  $FN$  表示假反例, 指模型错误地预测为负例的样本数。此外还有  $TN$  表示真反例, 指模型正确地预测为负例的样本数;  $FP$  表示假



(a) 横向裂缝 (b) 纵向裂缝 (c) 块状裂缝 (d) 龟裂 (e) 坑槽 (f) 无病害

图 4 各病害图像

正例，指模型错误地预测为正例的样本数。

$F1-score$  是查准率和查全率的调和平均值，用于综合评估模型的性能。 $F1-score$  越高，表明模型越可靠。 $F1-score$  公式为：

$$F1 - score = (2 * Recall * Precision) / (Recall + Precision) \quad (5)$$

式 (5) 中  $Precision$  代表查准率，计算公式为：

$$Precision = TP / (TP + FP) \quad (6)$$

式 (5) 的计算的是二分类问题的  $F1-score$ ，本文为六分类问题，所以本文的  $F1-score$  计算公式如式 (7) 所示，其中  $i$  代表类别序号。

$$F1 - score = \frac{\sum_{i=1}^6 F1 - score_i}{6} \quad (7)$$

### 3.2 实验方法

本文采用双阶段训练模型方法。实验分为初次实验和二次实验。二次实验在初次实验的基础上通过调整初始模型，修改训练参数来解决初次实验中出现的过拟合问题。过拟合通常是指深度学习模型在训练过程中对于训练数据拟合得过于好，以至于捕捉了数据中的噪声和随机性，导致在未见过的数据上表现较差。一般过拟合现象的表现形式为训练  $Loss$  在训练过程中不断下降，而验证  $Loss$  先下降再上升。

调整初始模型是将初次实验训练好的模型作为二次实验初始模型。这样的选择是考虑到初次实验完成的模型已经在初次实验参数条件下本文数据集有了较为均衡的拟合效果，故选择其作为初始模型可以对本文的数据具有更加针对性的学习，降低出现过拟合现象的风险。

修改训练参数包括更换损失函数，增加图像增强手段，降低学习率等。

修改损失函数是将交叉熵损失函数更换为 `MultiClassFocalLossWithAlpha ()` 损失函数。初步实验选择的交叉熵损失函数虽然可以帮助多分类模型判断预测的类别和真实类别之间的相似度。但是它并没有考虑数据量不平衡的情况。故而将损失函数更换为 `MultiClassFocalLossWithAlpha ()` 损失函数。这是一种扩展的损失函数，通过引入  $\alpha$  参数来调整不同类别在损失函数中的权重，以解决多类别分类问题中类别数量不平衡的问题。通过给横向裂缝，纵向裂缝，坑槽等数据量较少的类别增加权重提高在它们在模型训练过程中的影响力。

数据增强手段可以提升模型泛化性能，降低过拟合出现的风险，因此在初次实验的基础上增加随机灰度数据增强方式。学习率用于控制模型在优化过程中更新参数的步长，降低学习率可以使得模型在训练数据上的学习过程更为平缓，减缓模型对训练数据中噪声和异常样本的过度拟合<sup>[28]</sup>。选择将初次实验中的学习率降低十倍进行二次实验。

### 3.3 实验设置

本文所有实验均在 Linux 下，ubuntu 操作系统，版本

为 18.04。CPU 型号为 Intel Xeon Platinum 8255C，有 12 个虚拟 cpu 核心。GPU 型号为 RTX 3080，显存为 10 G。深度学习框架为 pytorch，版本为 1.1。python 版本为 3.8。

初次实验中，所有模型训练时 Batch 设置为 16，优化器使用随机梯度下降优化器 (SGD)，损失函数采用交叉熵损失函数 (CrossEntropyLoss)，学习率设置为 0.001，训练迭代次数设置为 100。同时为了在训练过程中增加数据的多样性，提高模型的泛化能力，选择两种数据增强手段，分别为随机裁剪缩放和随机水平翻转。选择训练过程中在验证集上准确率最高的模型作为最终模型。

二次实验损失函数选择为 `MultiClassFocalLossWithAlpha ()`，学习率设置为 0.000 1，数据增强手段增加随机灰度方式。仍然选择训练过程中验证精度最高的模型作为最终模型，训练  $Epoch$  设置为 50 轮。其他设置均与初次实验相同。

### 3.4 不同阶段实验对比

初次实验中 Swin Transformer 和 DenseNet 的训练  $Loss$  曲线分别如图 5 和图 6 所示。从它们的  $Loss$  曲线图可以看出符合过拟合现象的表现。

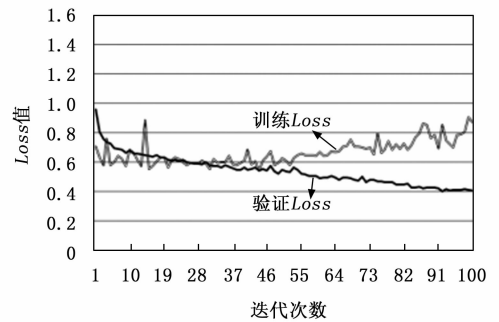


图 5 Swin Trasformer 训练过程  $Loss$  值

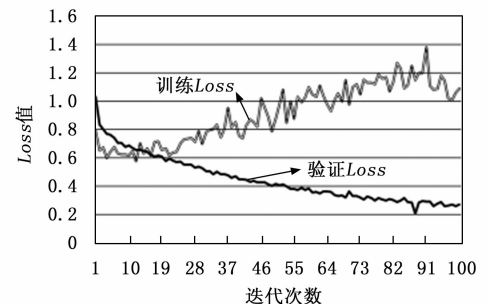


图 6 DenseNet 训练过程  $Loss$  值

Swin Transformer 和 DenseNet 二次实验最终模型分别在训练集、验证集和测试集上的分类准确率如图 7 所示。可以看出两个模型 3 个准确率比较接近，说明模型过拟合问题得到了妥善解决。

为了验证调整初始模型并在新的损失函数上训练是否使得分类模型在本文数据集上有更好的表现。选择初次实验使用的初始模型在新的损失函数上进行 100 个  $Epoch$  的

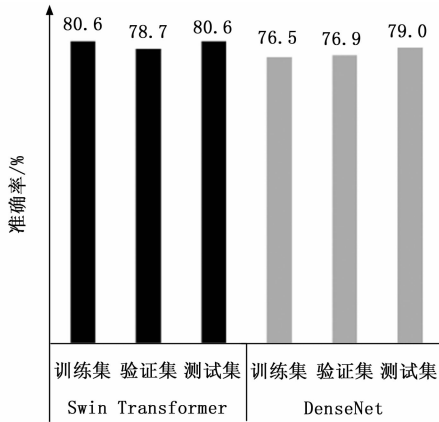


图 7 Swin Transformer 和 DenseNet 在不同数据集上的分类准确率

训练, 并与二次实验的结果进行对比。结果如图 8 所示, 实验 1 为不调整初始模型使用新的损失函数后, 最终模型在测试集上的测试结果。实验 2 为调整初始模型并使用新的损失函数在测试集上的测试结果。从图 8 可以看出, 实验 2 的所有模型分类准确率整体要比实验 1 所有模型分类准确率高, 证明所采用的调整初始模型方法可以使得训练模型更好地拟合本文数据集。

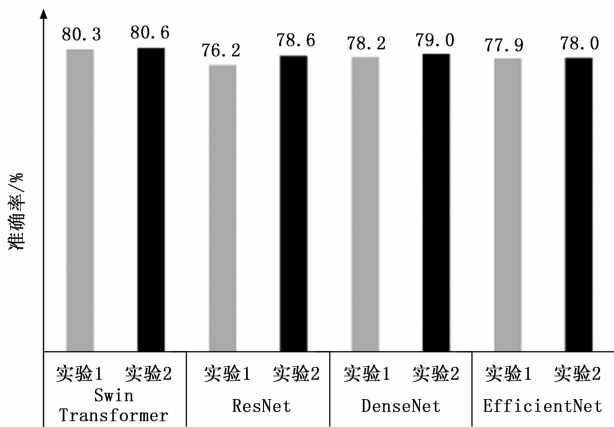


图 8 不同实验各模型在测试集上的分类准确率

### 4 实验结果分析

将二次实验完成后所有模型在测试集上的分类结果进行混淆矩阵图绘制, 如图 9~12 所示。

并通过混淆矩阵图计算出各模型分类准确率, 横向裂、纵向裂缝、块状裂缝的查全率以及 F1-score 如表 2 所示。

表 2 各模型评价指标对比

	准确率/%	横向裂缝查全率/%	纵向裂缝查全率/%	块状裂缝查全率/%	F1-score
Swin Transformer	<b>80.6</b>	<b>69</b>	<b>71</b>	<b>62</b>	<b>0.776</b>
ResNet	78.6	65	70	61	0.748
DenseNet	79.0	65	67	60	0.749
EfficientNet	78.0	61	65	59	0.728

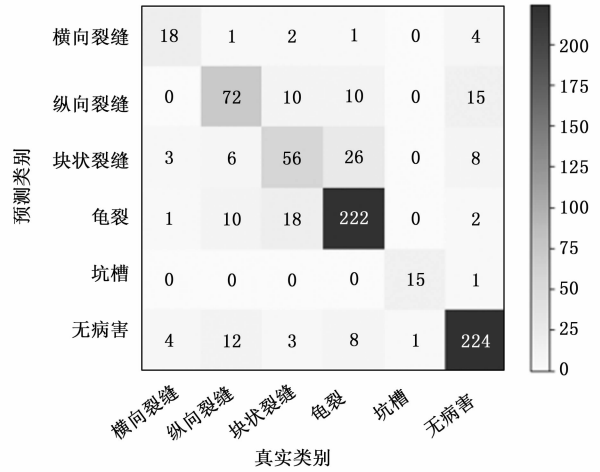


图 9 Swin Transformer 分类混淆矩阵图

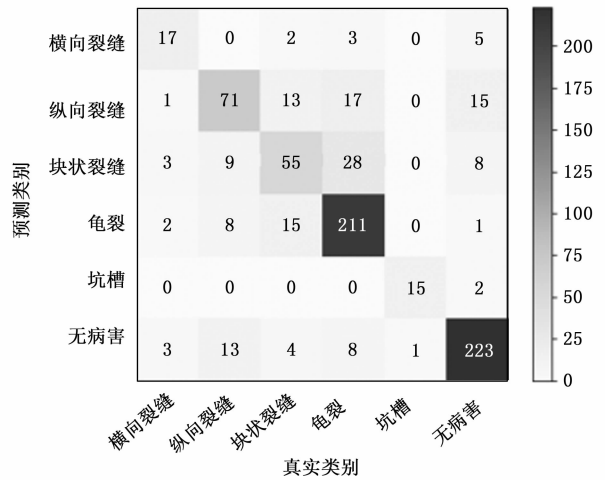


图 10 ResNet 分类混淆矩阵图

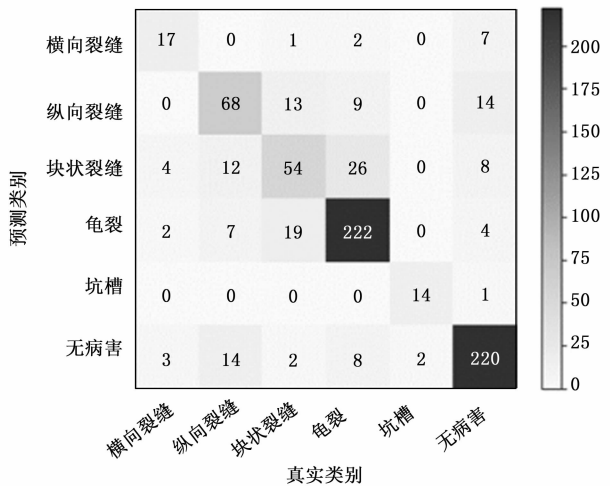


图 11 DenseNet 分类混淆矩阵图

从表 2 可得, Swin Transformer 相较于 CNN 模型表现出了更高的分类准确率。相较于 ResNet, Densenet 和 EfficientNet, 准确率分别提升了 2%, 1.6%, 和 2.6%。对于路面病害而言, 横向裂缝、纵向裂缝、块状裂缝数据长距

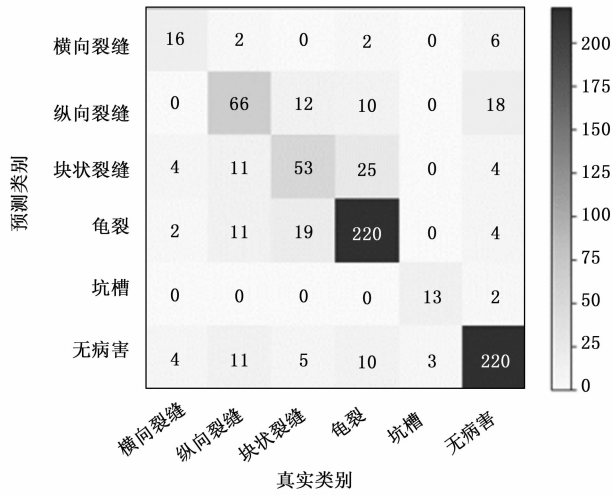


图 12 EfficientNet 分类混淆矩阵图

离裂缝信息，对于它们进行精准识别分类并做出及时修复，可以避免它们发展成更为严重的病害，有助于节约维护成本。从表 2 可以看出 Swin Transformer 在这 3 类病害上的查全率也都要优于 CNN 模型。说明 Swin Transformer 在长距离裂缝结构识别方面具有优势。

对于龟裂类型病害，图像中裂缝块更为密集，相较于长距离裂缝病害而言是更为严重的病害类型，同时也是图像数量最多的病害类型，对其精准识别也尤为重要。从各模型分类混淆矩阵图可以看出，Swin Transformer 对此类病害也具有较好的分类准确率，与 DenseNet 并列第一。同时对于坑槽型病害，直接影响着行车安全。对其进行精准识别并及时修复有助于减少安全事故的发生。从图中可以看出，在此类病害的识别上，Swin Transformer 也表现出了良好的分类性能，与 ResNet 并列第一。

### 5 结束语

为了解决传统 CNN 模型在路面病害检测领域对于长距离病害结构识别能力不足以及精度局限等问题，本文引入 Swin Transformer 进行沥青路面病害分类检测研究。

本文数据来自路面检测车采集数据，虽然具有图像清晰，无与路面无关因素干扰等优点。但是存在对比度低，数据量少等情况。并且由于是灰度图像，相对于常见数据具有一定的特殊性，因此在训练时容易出现过拟合现象。因此本文首先采用直方图均衡技术处理图像，增强对比度，并对数据进行裁剪扩充数据集。其次针对训练过程中模型过拟合问题，采用更换损失函数并调整初始模型的手段解决。不仅解决了过拟合问题，并且使得模型对本文数据的学习效果得到增加。最终实验结果证明，本文所采用的 Swin Trnsformer 在路面病害检测任务上相对于 CNN 模型具有更好的分类检测效果。整体分类准确率达到 80.6%，超越了所有对比模型。不仅在横向裂缝，纵向裂缝，块状裂缝等这些长距离病害结构识别上面具有优势并且对于非长距离病害结构也有很好的分类效果。F1-score 计算结果也证明训练好的 Swin Transformer 模型具有良好的可靠性，

能够满足实际工程中对沥青路面病害实现高效精准分类检测的需求。

### 参考文献:

- [1] 《中国公路学报》编辑部. 中国路面工程学术研究综述 [J]. 中国公路学报 2020, 33 (10): 1-66.
- [2] JTG 5210-2018, 公路技术状况评定标准 [S]. 北京: 人民交通出版社, 2019.
- [3] 梁 楦. 沥青路面裂缝智能化检测关键技术研究 [D]. 南京: 东南大学, 2021.
- [4] 任传亭, 任广海, 李森哲. 公路沥青路面检测技术发展综述 [J]. 公路交通技术, 2021, 37 (5): 29-34.
- [5] 李保险. 基于路面三维图像的沥青路面裂缝自动识别算法 [D]. 成都: 西南交通大学, 2019.
- [6] 郭 辰. 基于深度学习的沥青路面病害检测及评估方法研究 [D]. 石家庄: 石家庄铁道大学, 2022.
- [7] 王 磊. 基于机器视觉的路面裂缝分类与检测方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2019.
- [8] 邓 露, 褚鹤鸣, 龙砺芝, 等. 基于深度学习的土木基础设施裂缝检测综述 [J]. 中国公路学报, 2023, 36 (2): 1-21.
- [9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
- [10] 沙爱民, 童 峥, 高 杰. 基于卷积神经网络的路表病害识别与测量 [J]. 中国公路学报, 2018, 31 (1): 1-10.
- [11] 陈泽斌, 罗文婷, 李 林. 基于改进 U-net 模型的路面裂缝智能识别 [J]. 数据采集与理, 2020, 35 (2): 260-269.
- [12] 晏班夫, 徐观亚, 栾 健, 等. 基于 Faster R-CNN 与形态法的路面病害识别 [J]. 中国公路学报, 2021, 34 (9): 181-193.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] //Advances in Neural Information Processing Systems, 2017, 30.
- [14] DOSOVISKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J/OL]. Arxiv Preprint Arxiv: 2010.11929, 2020.
- [15] 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.
- [16] LUO H, LI J, CAI L, et al. STrans-YOLOX: fusing swin transformer and YOLOX for automatic pavement crack detection [J]. Applied Sciences, 2023, 13 (3): 1999.
- [17] GUO F. Pavement crack detection based on transformer network [J]. Automation in Construction, 2023, 145: 104646.
- [18] 马 建, 赵祥模, 贺拴海, 等. 路面检测技术综述 [J]. 交通运输工程学报, 2017, 17 (5): 121-137.
- [19] YULONG Y. A detection and classification method of asphalt pavement crack based on vision transformer [C] //International Conference on Frontiers of Traffic and Transportation

Engineering, 2023.

[20] 郑楚伟, 林 辉. 基于 Swin Transformer 的 YOLOv5 安全帽佩戴检测方法 [J]. 计算机测量与控制, 2023, 31 (3): 15-21.

[21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[22] HUANG G, LIU Z, VAN Der MAATEN L, et al. Densely connected convolutional networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[23] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C] // International Conference

on Machine Learning. PMLR, 2019: 6105-6114.

[24] 靳阳阳, 韩现伟, 周书宁, 等. 图像增强算法综述 [J]. 计算机系统应用, 2021, 30 (6): 18-27.

[25] 祝中秋, 李 斌. 基于直方图拉伸的图像增强算法及其实现 [J]. 信息技术, 2009, 33 (5): 54-57.

[26] 张燕红, 孔 波, 邢翠芳, 等. 直方图规定化在图像增强中的应用及实现 [J]. 电脑知识与技术, 2009, 5 (34): 9814-9815.

[27] 王 浩, 张 叶, 沈宏海, 等. 图像增强算法综述 [J]. 中国光学, 2017, 10 (4): 438-448.

[28] 刘建伟, 刘 媛, 罗雄麟. 深度学习研究进展 [J]. 计算机应用研究, 2014, 31 (7): 1921-1930.

船舶设计通讯, 2019 (2): 16-20.

参考文献:

[1] ZHANG Z, GUAN C, CHEN H, et al. Adaptive privacy-preserving federated learning for fault diagnosis in internet of ships [J]. Institute of Electrical and Electronics Engineers (IEEE), 2022, 9 (9): 6844-6854.

[2] 杜德锋, 周文进, 何江贤, 等. 面向海量数据的全船振动分析管理系统研究 [J]. 计算机测量与控制, 2022, 30 (10): 202-208.

[3] SONG X, LIAO Z, WANG H, et al. Incrementally accumulated holographic SDP characteristic fusion method in ship propulsion shaft bearing fault diagnosis [J]. IOP Publishing, 2022, 33 (4): 045011.

[4] ZHIMING C, HOULIN L, RUNAN H, et al. Research on multi-fault identification of marine vertical centrifugal pump based on multi-domain characteristic parameters [J]. Mdpi Ag, 2023, 11 (3): 551.

[5] 饶鑫平, 张 新, 杨粉蝶, 等. 船舶双推进轴系振动分析软件开发研究 [J]. 计算机测量与控制, 2015, 23 (8): 2900-2903.

[6] 侯立巍, 顾为柏, 杜阳阳, 等. 反应堆冷却剂泵调试对中异常分析及改进 [J]. 中国机械, 2023, (19): 95-99.

[7] 李伟通, 于 雷, 陈玉清. 船用反应堆主冷却剂泵建模研究与仿真 [J]. 舰船科学技术, 2016, 38 (7): 77-80.

[8] 严新平. 智能船舶的研究现状与发展趋势 [J]. 交通与港航, 2016, 3 (1): 25-28.

[9] 孙燕妮, 白晓军. 《中国制造 2025》——中国特色的强国战略 [J]. 智能制造, 2020 (10): 43-45.

[10] 王蒙云, 刘桂香, 王 玲. 船舶智能制造技术的应用和发展 [J]. 科学技术创新, 2021 (34): 142-144.

[11] 陈 震, 甄希金, 王体涛. 船舶智能制造 [J]. 船舶工程, 2022, 45 (6): 5.

[12] 秦 时. 全球首艘智能 VLOC 在上海交付 中国智能船舶全面迈入 1.0 新时代 [J]. 船舶设计通讯, 2018 (2): 107.

[13] 曹 阳, 张淇鑫, 朱 蕾. 智能船舶 1.0 研发专项申报及管理 [J]. 船舶设计通讯, 2019 (2): 11-15.

[14] 朱 兵, 翁 爽. DNV GL 智能船舶入级指南要点 [J]. 船

舶设计通讯, 2019 (2): 16-20.

[15] 我国研制的全球首艘智能船舶交付 [J]. 交通企业管理, 2018, 33 (1): 23-23.

[16] 谢 珉, 刘洪亮. 风电运维母船智能系统的研究与应用 [J]. 船舶工程, 2023, 45 (4): 58-60, 65.

[17] 郑 兰. 基于大数据分析的运行预测与智能管理系统设计 [J]. 集成电路应用, 2023, 40 (8): 338-339.

[18] 洪 祥, 房淑华, 王学永. 基于机器学习算法的船舶电气故障分类与诊断方法 [J]. 舰船科学技术, 2023, 45 (15): 143-146.

[19] 潘荣军, 张 伟. 基于机器学习的自主航行技术研究: 探索人工智能在航海领域的应用 [J]. 中国水运 (下半月), 2023, 23 (7): 46-49.

[20] 牛旭彤, 范洪博. 浅析机器学习在智能航运的潜力与未来 [J]. 中国水运, 2023 (10): 96-97.

[21] 方 家. 基于实测数据的地铁钢轨打磨与振动相关性分析 [J]. 交通科技与管理, 2023, 4 (20): 86-88.

[22] 董德亿. 基于输出信号相关性分析的轴承故障诊断方法 [D]. 杭州: 杭州电子科技大学, 2018.

[23] 蒋 慧. 船舶及海岛鸟类声信号特征提取与识别研究 [D]. 舟山: 浙江海洋大学, 2020.

[24] 袁明新, 戴现令, 刘 超, 等. 基于空间位置和轮廓线距离的船舶焊缝特征参数提取 [J]. 焊接学报, 2023, 44 (1): 84-92.

[25] 王洪彬, 何 荷, 黄睿灵, 等. 基于人工智能的智能变电站保护系统故障诊断 [J]. 电工电能新技术, 2023, 42 (9): 97-104.

[26] 鞠萍华, 蒋德轩, 冉 琰. 故障树和人工神经网络在可靠性再分配中的应用 [J]. 重庆大学学报, 2018, 41 (4): 11-19.

[27] 岳 毅. 基于迁移学习的轴向往塞泵故障诊断方法研究 [D]. 秦皇岛: 燕山大学, 2022.

[28] 田 森, 苏晓明, 陈长征, 等. 基于领域自适应的风力机发电机轴承故障诊断方法研究 [J]. 太阳能学报, 2023, 44 (11): 310-317.

[29] 李振宝. 基于深度学习的典型液压泵智能故障诊断与寿命预测方法 [D]. 秦皇岛: 燕山大学, 2022.

[30] 唐宏宾, 杨 婧, 唐 一. 轴向往塞泵疲劳损伤分析及寿命预测 [J]. 机床与液压, 2023, 51 (16): 165-171.