文章编号　1004-924X(2023)22-3357-14

基于Swin Transformer轻量化的TFT-LCD面板缺陷分类算法

夏　衍[1]，罗　晨[1*]，周怡君[1]，贾　磊[2]

(1. 东南大学 机械工程学院，江苏 南京 211189；
2. 无锡尚实电子科技有限公司，江苏 无锡 214174)

摘要：在TFT-LCD面板缺陷检测中，检测对象背景复杂、缺陷细微且种类繁多，而工业生产实时性要求高，传统的缺陷分类算法往往难以兼顾精度和速度要求，无法适用于实际生产应用。为均衡TFT-LCD面板缺陷分类的准确率和速率，提出一种基于Swin Transformer的轻量化深度学习图像分类模型。首先对模型每层输入的特征图进行Token融合以减少模型计算量，从而提高模型的轻量化水平。其次引入深度可分离卷积模块以帮助模型增加卷积归纳偏置，从而缓解模型对海量数据的依赖问题。最后使用知识蒸馏方法来克服模型轻量化导致的检测精度下降问题。在自制TFT-LCD面板缺陷分类数据集上的实验表明，本文提出的改进模型相比基线模型，FLOPs计算量降低了2.6 G，速度指标提升了17%，而Top-1 Acc精度仅损失1.3%，且与其他图像分类主流模型相比，在自制数据集和公开数据集上都具有更均衡的精度和速度。

关　键　词：TFT-LCD；Transformer；图像分类；计算机视觉

中图分类号：TP394.1　　**文献标识码**：A　　**doi**：10.37188/OPE.20233122.3357

A lightweight deep learning model for TFT-LCD circuits defect classification based on swin transformer

XIA Yan[1], LUO Chen[1*], ZHOU Yijun[1], JIA Lei[2]

(1. *School of Mechanical Engineering, Southeast University, Nanjing* 211189, *China*;
2. *Wuxi Shangshi-finevision Technology Co., Ltd, Wuxi* 214174, *China*)
* *Corresponding author, E-mail: chenluo@seu.edu.cn*

Abstract: Defect detection in thin film transistor-liquid crystal display (TFT-LCD) circuits is a challenging task because of the complex background setting, different types of defects involved, and real-time detection requirements from industry. Traditional methods have difficulties in satisfying the dual requirements of detection speed and accuracy. To address this challenge, in this study, a deep learning method is developed for image classification based on the Swin Transformer technique. First, token merging is used to reduce the computational complexity of each layer of the model, thus improving computation efficiency. Then, a depthwise separable convolution module is introduced to add convolutional bias to reduce the reli-

收稿日期：2023-03-30；**修订日期**：2023-05-10.

基金项目：国家自然科学基金资助项目(No. 51975119)；无锡市“太湖之光”科技攻关(产业前瞻及关键技术研发)项目资助(No. G20222011)

ance on massive data. Finally, a knowledge distillation method is applied to overcome the problem of reduced detection accuracy caused by the less-intensive computation design. Experimental results on the self-made dataset demonstrate that the proposed method achieves a 2.6 G FLOPs reduction and a 17% speed improvement compared to baseline models, with only a 1.3% Top-1 accuracy precision reduction. More importantly, the proposed model achieves better balance on accuracy and detection speed on both self-made and public datasets than existing mainstream models on image classification in the TFT-LCD manufacturing industry.

Key words: Thin Film Transistor Liquid Crystal Display (TFT-LCD); transformer; image classification; computer vision

1 引　言

伴随5G技术的出现，万物互联的发展趋势愈加明显，信息交互的增加进一步突出了智能显示的重要性，而作为智能显示的载体，显示面板行业也得到了极为迅速地发展，其中薄膜晶体管液晶显示器（Thin Film Transistor Liquid Crystal Display，TFT-LCD）充分满足了人们对于液晶显示面板性能、分辨率、价格等多方面的要求，逐渐成为了市场上的主流屏幕。

整个TFT-LCD面板制作过程极为复杂，其中不同工艺所产生的问题会在图像扫描的成像端上表现为各种形态不一的缺陷，而通过对这些缺陷进行是否为缺陷以及是何种缺陷的判定，可以帮助快速定位到导致缺陷产生的工艺，进而针对性地调整和改进工艺，因此TFT-LCD面板缺陷的准确快速分类对于整体生产过程具有重要意义。

随着液晶显示器内部结构更加精细，TFT-LCD缺陷逐渐缺乏固定的表现形式，不同种类的缺陷之间特征区别不明显，导致缺陷分类难度大大增加。同时，缺陷产生的位置也对缺陷分类产生了新的要求，由于不在线路区域的缺陷并不影响产品的正常使用，因此这类情况下即便存在缺陷也依然需要被判定为正常。此外，除了对分类准确率的高要求，分类算法还需要兼顾好对生产节拍时间的高要求。

目前TFT-LCD面板缺陷的分类算法主要为基于自动光学检测的传统方法，分为特征提取和分类器识别两阶段。特征提取能够帮助研究对象从图像矩阵格式转换为特征向量格式输入后续的分类器中，这些提取的特征剔除很多冗余信息后，显著提升了分类器的训练速度和精度。一般常用的特征提取方法包括灰度共生矩阵[1]、梯度方向直方图[2]和二值连通区域标记[3]等。分类器识别主要是将上一阶段提取到的缺陷特征输入以支持向量机[4]（Support Vector Machine，SVM）为主的各类分类器中完成缺陷分类。Kang等[5]定义并提取了24个LCD缺陷的关键特征，输入SVM后完成了对LCD生产流程中缺陷的分类。Huang等[6]提出了一种基于视觉词袋模型的TFT-LCD缺陷分类算法，引入颜色和SIFT（Scale-Invariant Feature Transform）特征描述缺陷区域后，利用基于多核卡方核函数的SVM获得了效果最佳的分类器。Kong等[7]使用带通滤波器和Sobel边缘检测算子对图像增强后获得了更为明显的缺陷特征，利用这些特征训练的SVM分类器完成了对TFT-LCD水渍缺陷的分类。但上述方法极为依赖人为选取的特征，因此对场景的变化比较敏感，缺乏对复杂背景下缺陷的分类能力。

近年来深度学习的快速发展为缺陷分类任务提供了新的思路[8]，如弱监督学习[9]、少样本学习[10]和抽象学习[11]等。Wen等[12]对于钢材表面缺陷提出了一种集成不同深度卷积神经网络的缺陷分类方法，通过数据增强分别训练三个不同的深度卷积神经网络以减少过拟合后，再将三个训练好的模型融合以提高分类准确性和鲁棒性。Fu等[13]采用预训练的SqueezeNet作为骨干网络，在只需要少量特定于缺陷的训练样本情况下，完成了钢材表面各类缺陷的高精度识别。Ihor等[14]基于ResNet50网络结构，结合注意力损失函数，实现了金属表面缺陷的多标签分类。He等[15]引入多尺度感受野提取多尺度特征，并利用自动编码器降低多尺度特征维度，进而提高了模型在训练

样本不足下的特征表达和泛化能力，最终实现了对热轧钢板表面缺陷的准确分类。Haselmanm 等[16]利用人工合成缺陷构建了更为准确的缺陷数据集，根据数据集训练好的深度卷积神经网络实现了塑料制品表面真实缺陷的准确分类。

尽管深度学习方法大大提高了对各类缺陷的分类能力，但其对于具有复杂纹理和模糊边界等特点的缺陷仍难以精确区分。同时，考虑到工业检测的实时性要求，深度学习模型还应当保证轻量化。因此，本文在高精度网络 Swin Transformer 基础上提出轻量化改进，在损失较小精度的情况下完成了最大化的速度提升，最终获得了具有 92.7% 精度和每秒检测 60.43 张图像速度的适用于工业生产的深度学习模型。本文的改进工作如下：

（1）引入标记向量融合模块，通过融合相似的标记向量以减少每一层送入神经网络中的数据，从而完成对模型的轻量化操作；

（2）插入深度可分离卷积模块以帮助模型增加卷积归纳偏置，进而缓解模型中 Transformer 架构对训练数据的巨大需求并提高分类准确率；

（3）通过知识蒸馏策略将改进模型作为学生模型，从而保证模型轻量化的同时进一步提升模型的分类准确性。

2　Swin Transformer 算法

Swin Transformer 算法[17]由微软研究院在 2021 年的 ICCV 会议上提出，其在 Vision Transformer（ViT）模型[18]的基础上，引入类似卷积神经网络（Convolutional Neural Network，CNN）的层次化构建方法，通过对特征图的下采样实现了多尺度的检测能力。同时，模型使用窗口多头自注意力（Windows Multi-Head Self-Attention，W-MSA）机制以减少计算量，并使用移动窗口多头自注意力（Shifted Windows Multi-Head Self-Attention，SW-MSA）机制解决 W-MSA 导致的窗口间信息传递隔绝问题。具体参数量最少的 Swin Transformer Tiny（Swin-T）模型网络结构如图 1 所示。由图可知，整个 Swin Transformer 模型传播过程主要分为三个步骤：

（1）首先输入图片在图像块分割（Patch Partition）层进行分块，其中每 4×4 相邻像素为一个图像块（Patch）并沿通道方向展平为标记向量（Token），然后在线性嵌入层（Linear Embedding）对通道进行线性变换；

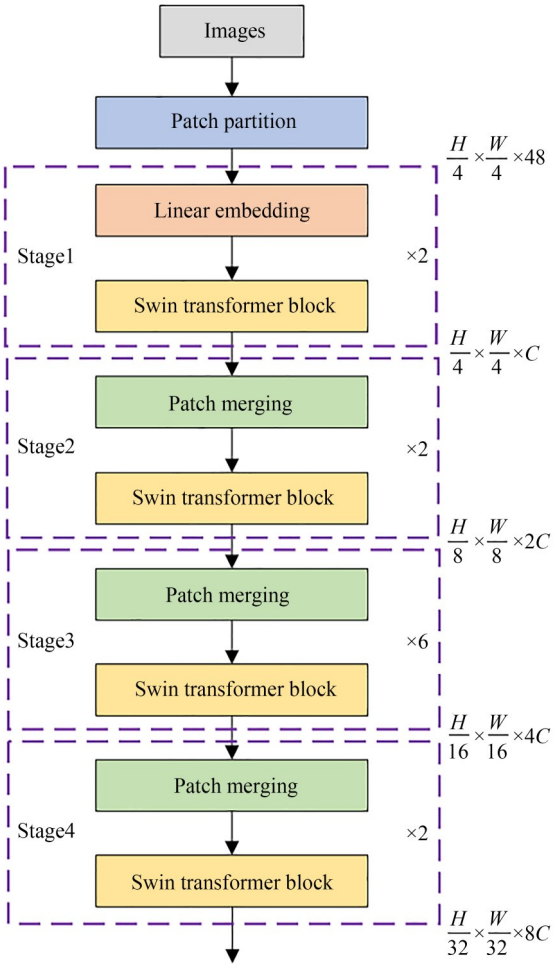


图 1　Swin-T 网络结构示意图

Fig. 1　Architecture of Swin-T model

（2）利用四个阶段构建不同大小特征图，其中后三个阶段中首先使用图像块融合（Patch Merging）进行下采样，然后重复堆叠 Swin Transformer Block 模块。值得注意的是，Block 模块中包含了图 2 显示的两种结构，前者使用 W-MSA 机制而后者使用 SW-MSA 机制，两种结构成对使用，因此堆叠的 Block 次数都为偶数；

（3）最后接上归一化（Layer Normalization，LN）层、全局池化层和全连接层完成图像分类任务。

Swin Transformer 目前在图像分类、目标检测和图像分割等视觉任务上均表现出优异的性能，但受限于其较大的计算开销，仍需要在保证

精度的情况下完成模型轻量化才能更好地应用于工业领域。

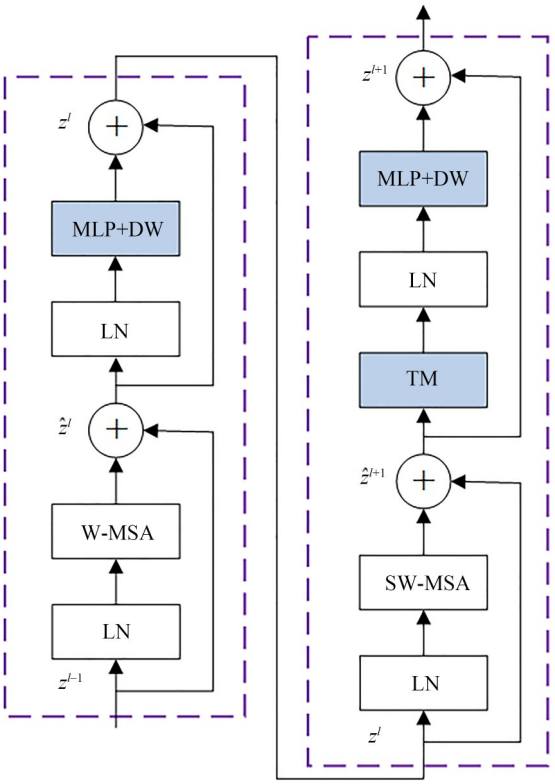


图 2　Swin Transformer Block 示意图

Fig. 2　Architecture of Swin Transformer Block

3　改进的 Swin Transformer 算法

针对 TFT-LCD 面板缺陷分类对速度和精度的高要求，本文在保证 Swin Transformer 精度的基础上进行了轻量化改进，具体改进部分在图 2 中以颜色标出。首先，在每个 Swin Transformer Block 内的 SW-MSA 和多层感知器（Multi-Layer Perception，MLP）之间插入 Token 融合（Token Merging，TM）模块[19]，从而完成对特征图数据和模型计算量的减少；然后，针对 Transformer 架构依赖巨大数据量的问题，在多层感知器中引入深度可分离卷积（Depthwise Separable Convolution，DW）模块[20]，帮助模型获得卷积归纳偏置以适应较小的数据量；最后将训练好的 Swin Transformer Base（Swin-B）大模型作为教师模型，改进后的轻量模型作为学生模型，通过知识蒸馏策略[21]在不改变模型参数量的情况下提高改进模型的分类精度。

3.1　Token 融合

输入图片被划分为众多图像块并展平为标记向量后，自注意力机制将计算每一个标记向量和其他所有标记向量之间的关系，因此模型复杂度和输入标记向量数量的平方成正比，而这种关系限制了 Transformer 架构对高分辨率图像的处理能力。从标记向量维度出发，目前一种常用的降低 Transformer 架构模型计算复杂度的方法是对标记向量进行剪枝，剪枝可以一定程度上保证精度并减少计算量，但也存在着极大的局限性：一方面剪枝需要引入额外的神经网络来计算每个标记向量的得分，进而判断哪些标记向量应当予以保留；另一方面剪枝可能造成重要信息的损失，因此需要确定好适当的剪枝比率。基于上述分析，相较于标记向量剪枝，选择引入无需训练并可以即插即用的 Token 融合模块对 Swin Transformer 进行轻量化操作。

考虑到自注意力机制通过对每个标记向量提取查询（Query，$\boldsymbol{Q}$）、键（Key，$\boldsymbol{K}$）和值（Value，$\boldsymbol{V}$）矩阵来完成对注意力权重的计算，其中键矩阵在转置后和查询矩阵进行点积以获得向量之间的相关性，其实质上已经包含了每个标记向量的信息，因此使用键矩阵均值的余弦距离来作为衡量标记向量之间相似度的标准。如图 3 所示，将 Token 融合模块插入在 SW-MSA 和 MLP 之间，一方面利用 SW-MSA 模块的自注意力计算来获得标记向量之间的相似性结果，另一方面保证融合后的标记向量可以被 MLP 继续传播。具体的 Token 融合模块中使用了二分软匹配算法（Bipartite Soft Matching）来快速匹配相似的标记向量，进而实现精确减少 r 个标记向量，具体方法步骤如下：

（1）将输入到模块内的所有标记向量均分为集合 A 和集合 B；

（2）对集合 A 中的每个标记向量，在集合 B 中找到和其最相似的标记向量并连线；

（3）保留所有连线中最相似的 r 条予以保留；

（4）对于仍然相连的标记向量取均值进行融合；

（5）最终输出上述集合的并集。

需要注意的是，由于融合后的标记向量包含了多个标记向量的信息，所以其所占的注意力权重也应当更大，因此在原注意力计算的公式内加

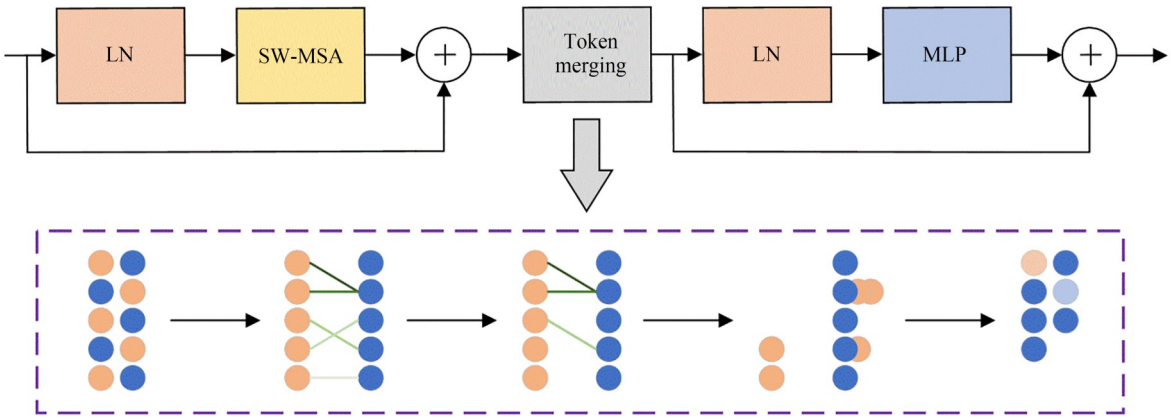


图 3　Token 融合模块示意图

Fig. 3　Token merging module

入包含每个标记向量尺寸的行向量 s 来帮助调整，计算公式如式(1)所示：

$$A=\mathrm{Softmax}\left(\frac{\boldsymbol{QK}^{\mathrm{T}}}{\sqrt{d_k}}+\log \boldsymbol{s}\right), \tag{1}$$

其中：$\boldsymbol{Q}$ 为查询矩阵，$\boldsymbol{K}$ 为键矩阵，d_k 为键的向量长度，取其平方根主要为了解耦点积结果方差与向量长度之间的关系，进而避免训练过程中梯度的消失。

考虑到 Swin Tansformer 模型的多层次网络设计，Token 融合模块中 r 值需要根据每一层的特征图大小进行动态调整。假设输入特征图宽高为 $H\times W$，如果将该特征图的宽高都减少 $2n$（取偶数以保证后续下采样的进行），则该层需要保留的标记向量数量应为：

$$r=2n(H+W)-4n^2. \tag{2}$$

同时，下一层图像块融合时的输入也应当同步调整为 $(H-2n)\times(W-2n)$。具体的 Token 融合操作如下：

(1) 自注意力机制模块中除了输出注意力矩阵外，同时输出对应的键矩阵均值；

(2) 判断是否为一个阶段内的最后一次 SW-MSA 计算，如果是，则对上述输出的键矩阵均值进行维度和通道数的变换，从而符合后续 Token 融合的输入要求；

(3) 确定对每个阶段输出的特征图宽高的减少值，从而计算出对应的每个阶段经过 Token 融合后保留的标记向量数量(r)；

(4) 利用键矩阵均值和二分软匹配算法对相似的标记向量进行融合，最终只保留设定的 r 个标记向量；

(5) 修改下一阶段下采样时的特征图宽高大小。

经过 Token 融合后的每一阶段输出的特征图都有所减小，减小后的特征图在进入下一阶段后即可有效降低下一阶段的计算量，最终完成模型轻量化。

此外，为了进一步显示 Token 融合的效果，对 Token 融合过程进行了可视化分析，如图 4 所示，相同背景块对应的标记向量得以融合，而不同背景块则不会进行融合，因此融合后的标记向量并不会对缺陷目标的识别造成过多的干扰。

3.2　深度可分离卷积

除了计算的巨大开销外，Transformer 架构模型往往还需要在超大数据集上进行预训练才能达到超越 CNN 的效果，而造成这种对训练数据巨大需求的原因在于其缺乏类似于 CNN 的归纳偏置。这些符合视觉任务属性的归纳偏置能够帮助 CNN 在较小的数据集上也能取得较好的检测效果，而 Transformer 架构的自注意力机制处理图像时考虑的是全局感受野，因此其虽然有效建模了全局信息但也损失了先验知识，进而需要大量数据进行学习。基于以上分析，在 Swin Tranformer 模型中加入了卷积操作以引入一定的归纳偏置和提升分类准确率，同时考虑到轻量化要求，加入的卷积为深度可分离卷积。

深度可分离卷积主要由逐通道卷积和逐点卷积组成：逐通道卷积中一个卷积核只负责一个通道，因此最终的输出通道数和输入通道数完全

相同;逐点卷积即为1×1卷积,其在遍历输入特征图每个位置的基础上在通道方向进行了加权组合,从而融合通道信息生成新的特征图,完成对特征图维度的改变。如图5所示,常规卷积的

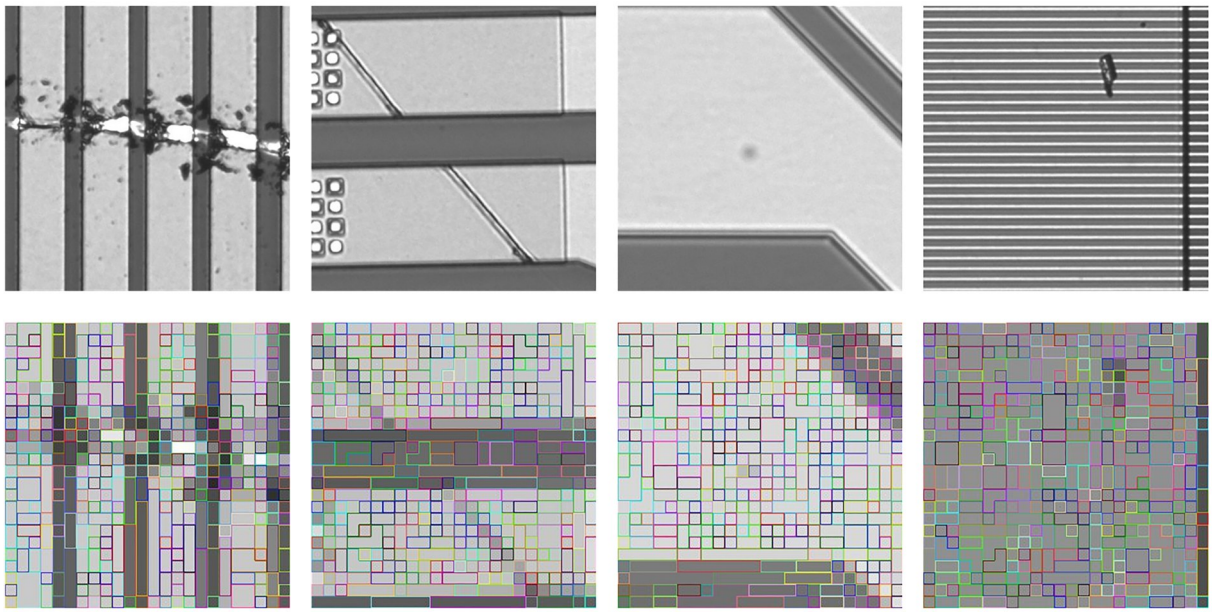
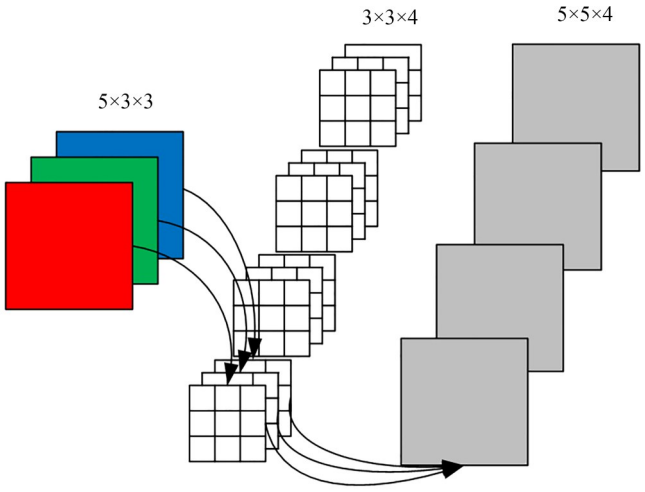


图4　Token融合可视化

Fig. 4　Token merging for visualization

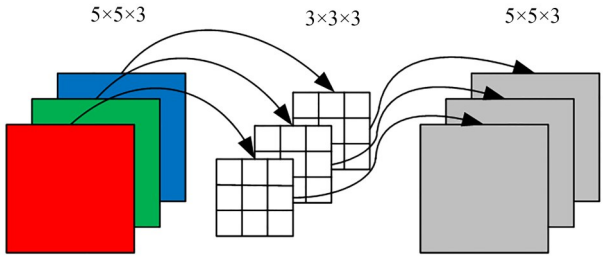


Params=3×3×3×4=108

FLOPs=3×3×(5-3+2×1+1)×(5-3+2×1+1)×3×4=2 700

(a) 常规卷积

(a) Conventional convolution

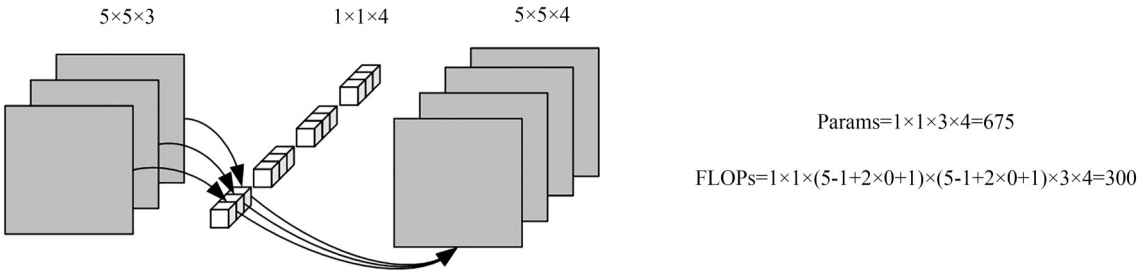


Params=3×3×3×1=27

FLOPs=3×3×(5-3+2×1+1)×(5-3+2×1+1)×3×1=675

(b) 逐通道卷积

(b) Depthwise convolution



(c) 逐点卷积
(c) Pointwise convolution

图5　常规卷积和深度可分离卷积示意图

Fig. 5　Conventional convolution and depthwise separable convolution

参数量(Params)为108,浮点数计算量(FLOPs)为2 700,逐通道卷积卷积的参数量为27,计算量为675,逐点卷积的参数量为12,计算量为300。因此,深度可分离卷积的整体参数量为27+12=39,整体计算量为675+300=975,与常规卷积108的参数量和2 700的计算量相比,减少了约2/3。因此,对于同样的输入和输出,深度可分离卷积更加满足轻量化要求。

为了加入深度可分离卷积模块,首先将Swin Transformer模型MLP模块的输入数据由展平的一维序列调整为二维的特征图格式,然后依次经过1×1卷积(特征图维度升降)、GELU激活函数和Dropout操作,接着加入深度可分离卷积来帮助模型引入归纳偏置,最后再经过1×1卷积和Dropout操作得到输出并调整回一维序列格式。

3.3　知识蒸馏策略

知识蒸馏(Knowledge Distillation, KD)作为一种常用的模型压缩方法,其主要利用从具有良好性能的大模型中学到的知识来指导训练小模型,从而帮助小模型在参数量大幅减少的同时保证和大模型相当的性能,最终实现模型的压缩。知识蒸馏方法首先训练一个复杂大模型,然后将该复杂模型的输出与数据真实标签相结合去训练新的轻量小模型,其中复杂大模型也被称为教师模型(Teacher),轻量小模型被称为学生模型(Student),来自教师模型的输出信息即为知识(Knowledge),学生模型学习教师模型信息的过程即为蒸馏(Distillation)。

一般分类任务中的标签往往为独热编码(one-hot)形式,即除了正标签为1,其他的负标签值都为0,例如一张内容为老虎的图像,其老虎的标签为1,猫和鸭子的标签都是0,则此时猫和鸭子的概率相同,而老虎和猫之间的潜在关系被忽略了。因此,为了提供更多的信息量,软目标概念被提出,其实质为引入蒸馏温度T后教师模型在归一化指数层(Softmax)输出的各个类别的概率,即每个类别都分配有一定的概率,即便同样为负标签,但某些负标签的概率远大于其他负标签,说明教师模型的推理中该标签与样本存在一定相似性。所谓蒸馏温度,其主要是在原始Softmax的基础上除以T,具体计算公式如式(3)所示:

$$q_i=\frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}, \tag{3}$$

其中:q_i为每个类别最终输出的概率,z_i为每个类别的得分,T为蒸馏温度。由公式可得,$T=1$即为原始Softmax,当$T>1$时,Softmax的输出将趋于平滑,即原本某些概率被抑制为0的类别也将获得较小的概率,但当T过大时,类别概率将趋于相同,此时网络将丧失区分能力。因此,合适的T可以适当放大负标签携带的信息,使模型在训练时也会关注到负标签。

在实际运用中,将训练好的Swin-B作为教师模型,改进后的轻量模型作为学生模型,其中Swin-B模型作为Swin Transformer的基础模型,参数量高达88 M,整体复杂度约为Swin-T的三倍。首先教师模型和学生模型在相同的蒸馏温度$T=3$下分别获得软目标并计算交叉熵损失得到软目标损失,然后学生模型在$T=1$的情况下获得硬目标并和原始标签计算交叉熵损失得到

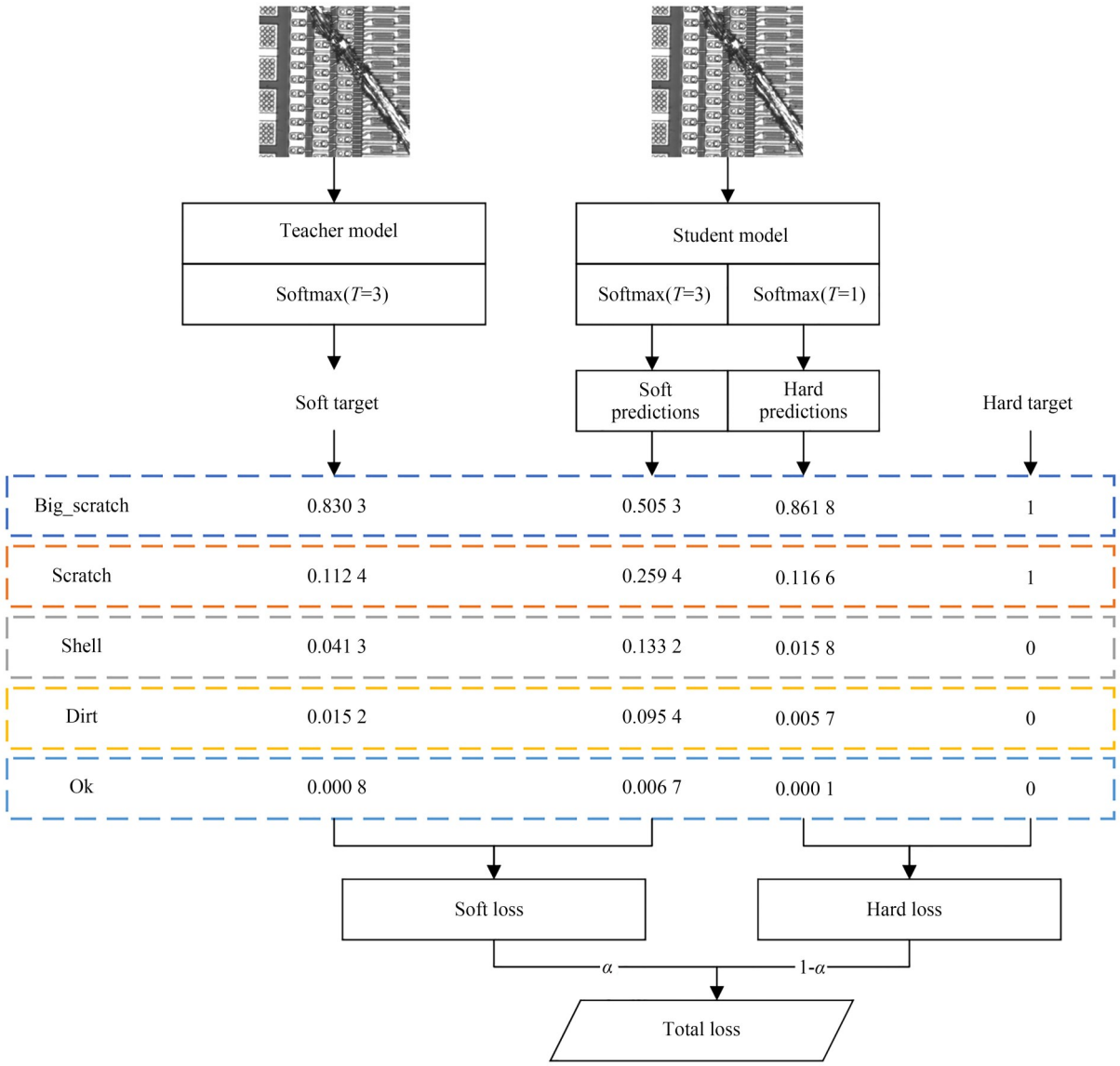


图 6　知识蒸馏流程示意图

Fig. 6　Knowledge distillation flow

硬目标损失，最后按一定的系数将软目标损失和硬目标损失加权得到最终损失，具体知识蒸馏流程如图 6 所示，其中损失函数计算如式(4)所示：

$$\begin{aligned} L_{\text{soft}} &= -\sum_{i=1}^{N} \boldsymbol{y}_i^{\mathrm{T}} \log(\boldsymbol{q}_i^{\mathrm{T}}) \\ L_{\text{hard}} &= -\sum_{i=1}^{N} c_i \log(q_i^1) \\ L_{\text{loss}} &= \alpha L_{\text{hard}} + (1-\alpha) L_{\text{soft}} \end{aligned} \quad , \tag{4}$$

其中：$\boldsymbol{y}_i^{\mathrm{T}}$ 为教师模型在 T 温度下预测出的为 i 类的概率(0-1 之间)，$\boldsymbol{q}_i^{\mathrm{T}}$ 为学生模型在 T 温度下预测出的为 i 类的概率(0～1 之间)，c_i 为在 i 类上的原始标签值(非 0 即 1)，q_i^1 为学生模型在 T=1 温度下预测出的为 i 类的概率(0～1 之间)，α 为平衡软目标和硬目标的权重系数。

4　实　验

4.1　数据集和评价指标

目前没有 TFT-LCD 面板缺陷分类的公开数据集，所以本文从工业现场收集了包含大划伤(big_scratch，240 张)、划伤(scratch，200 张)、贝壳(shell，668 张)、脏点(dirt，210 张)和无缺陷(ok，672 张)5 个种类 1 990 张 448×448 图像的分类数据集，其中训练集 1 393 张，测试集 398 张，验证集

199 张，每张图像仅有一个标签，即每张缺陷图像内仅包含一类缺陷，并不涉及多标签分类任务。

对于图像分类任务的评估，评价指标主要针对模型分类的准确率和计算量，其中准确率常用 Top-1 准确率（Top-1 Acc）、精准率（Precision）、召回率（Recall）和 F_1 得分，计算量常用 FLOPs。Top-1 Acc 即取最终输出中概率向量最大的一个作为预测结果时的分类准确性。Precision 表示被检测为正样本的缺陷中实际为正样本的比例。Recall 表示被检测为正样本的缺陷占总体正样本的比例。F_1 得分为精确率和召回率的调和平均数。FLOPs 则主要通过统计模型结果中加法和乘法计算的次数来衡量模型整体计算的复杂度。

4.2　实验环境

本实验所用计算机处理器型号为 Intel Core i7-9700@3.00 GHz，显卡型号为 NVIDIA GeForce RTX 3090 GPU，内存为 16 GB。操作系统为 Windows10 64 位，神经网络部分在 Pytorch 框架下搭建，软件编程环境为 python3.9，使用 CUDA 11.7 并行架构来提高计算能力。训练过程中，由于 GPU 显存限制，批处理大小为 8，输入尺寸为 448×448，训练轮数为 300，采用 Adam 优化器，初始学习率为 0.001 并使用余弦退火学习率策略进行调整，Dropout 为 0.1。此外，使用在大型数据集 ImageNet 上训练好的参数作为教师模型（Swin-B）训练时的初始化参数，从而保证小规模数据集在深度神经网络下训练的精度。实验过程中模型损失如图 7 所示，随着训练轮数的不断增加，loss 曲线逐渐趋于平稳，在 150 轮后逐渐收敛，训练过程未出现过拟合现象。

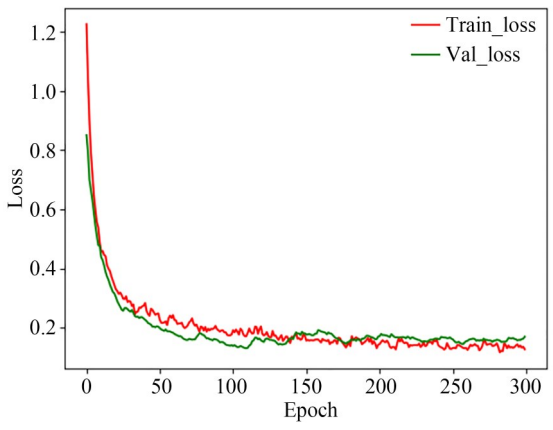


图 7　Loss 曲线

Fig. 7　Loss curve

4.3　参数 n 的影响分析

参数 n 值作为确定随不同层级动态变化的 r 值的变量，可以控制 Token 融合过程中每一层去除的标记向量数量，越大的 n 值去除越多的标记向量，模型也就更加轻量化，但检测精度也会对应有所下降，不同 n 值情况下的模型检测结果如表 1 所示。随着 n 值的增大，模型 Top-1 Acc 逐渐下降，同时 FLOPs 显著下降，每秒检测的图片数量（i m/s）则显著增加，在均衡模型精度和速度的前提下，n=2 时对应的 r 值更加符合工业实时检测的要求。需要注意的是，伴随着模型各阶段的下采样操作，每一层的特征图宽高在不断变小，因此 n 值不能过大以避免后续无法继续下采样，由于当前模型的输入尺寸为 448×448，因此 n 值应不大于 2。综合以上分析，在后续实验中，取 n=2 时对应的 r 值作为每一层的标记向量减少数量。

表 1　参数 n 影响实验结果

Tab. 1　Results of parameter n affects experiments

n	Top-1 Acc	FLOPs/G	i/(m·s^{-1})
0	94.0	17.5	51.42
1	92.1	16.3	54.49
2	90.6	14.7	61.92

4.4　参数 T 和 α 的影响分析

参数 T 值影响知识蒸馏过程中教师和学生模型最终输出的软目标结果，越大的 T 值导致 Softmax 输出的概率更加平滑，模型对负标签的关注更高，但过大的 T 值也会导致类别概率之间缺乏区分度。如图 8 所示，不同 T 值对模型整体

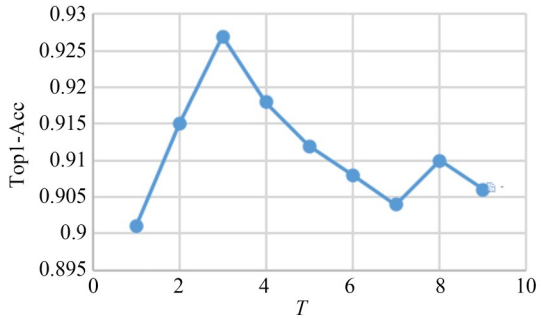


图 8　不同参数 T 对模型准确率影响

Fig. 8　Different effect of the parameters T on the model accuracy

精度的影响相差不大，模型的FLOPs也并不受影响。当T=1时，模型准确率最低，说明模型需要关注到一定的负标签信息；T=3时，模型准确率达到最高，因此可以在该值的基准上对α进行调整。

在T=3的基础上进一步验证权重参数α对模型结果的影响，α越小则软目标损失占比更高，模型更倾向于向教师模型学习相关信息。如图9所示，相对于α值取0.5，其偏向于0或1一侧时的模型精度更高，其中α取0.8时模型精度虽然在周围表现最高，但仍低于α取0.3时的检测精度，因此最终选择α=0.3。

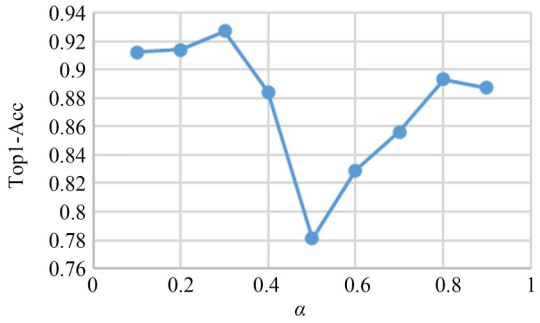


图9　T=3时不同参数α对模型准确率影响

Fig. 9　Different effect of the parameters α on the model accuracy at T=3

4.5　消融实验

为分析各改进部分对模型性能的影响，共设计五组实验对不同改进模块进行分析，每组实验均在相同训练参数，不同模型内容上进行测试。模型性能检测结果如表2所示，其中各模块依次叠加。对比Swin-T和Swin-T＋Token融合的性能，加入Token融合后FLOPs提升了16%，速度指标提升约20%，表明Token融合模块可以一定程度上加速模型，帮助减少模型计算量。虽然Top-1 Acc相较基线模型下降了3.4%，F_1得分相较基线模型下降了0.034 5，但也符合预期在牺牲一部分精度的情况下对模型进行轻量化。深度可分离卷积的引入为系统提高了0.6%的Top-1 Acc和0.013 6的Precision，其原因在于加入了一定的归纳偏置，但同时模型的FLOPs略增了1.4%，Recall和速度指标也略有下降。对改进模型进行知识蒸馏后，在不改变FLOPs和速度的情况下提升了1.5%的Top-1 Acc和0.008 6的F_1得分，说明改进模型作为学生模型成功从教师模型(Swin-B)中学到了知识。以上改进策略最终在Top-1 Acc仅损失1.3%和F_1得分仅下降0.024 5的情况下完成了对FLOPs指标14%的降低和速度指标17%的提升，证明了三种改进方法在速度和精度均衡方向上的有效性。

此外，对于实际工业生产中的TFT-LCD缺陷分类任务，面阵相机拍摄的图像分辨率为1 280×1 024，检测时间要求为100 ms，分类精度要求为90%。在将模型部署到实际工业现场的C＋＋平台后，模型每张图像的推理时间约为15ms，将待测图像分割为6张512×512分辨率的小图像送入模型中(在模型中压缩为448×448分辨率)，最终的检测时间约为90 ms，符合检测时间的需求。同时，改进模型的分类Top-1准确率达到92.7%，符合检测精度的需求。

表2　改进策略消融实验

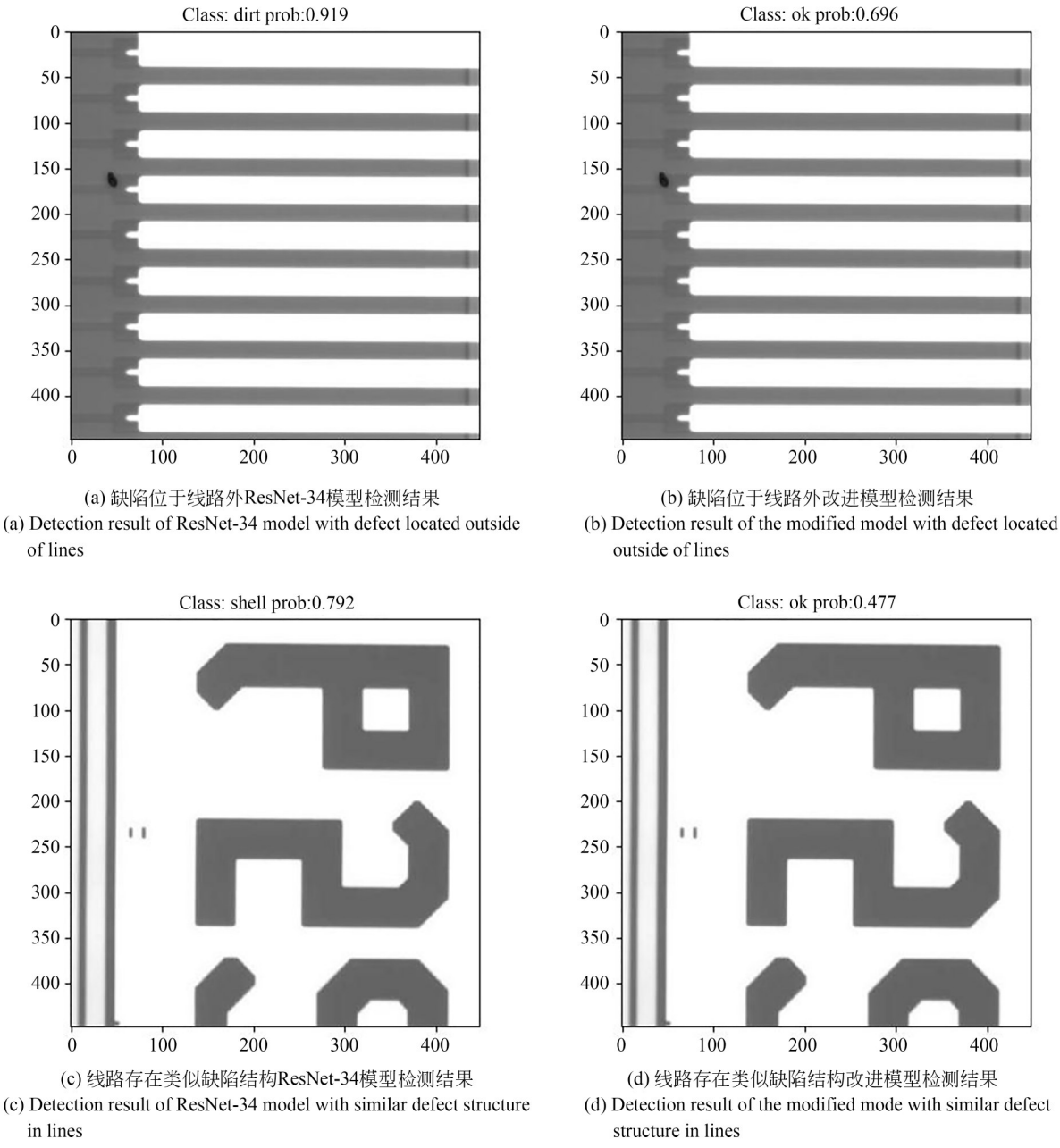
Tab. 2　Results of ablation experiments

Model	Top-1 Acc/%	Precision	Recall	F_1	FLOPs/G	i/(m•s^{-1})
Swin-T	94.0	0.935 4	0.935 6	0.935 5	17.5	51.42
＋Token merging	90.6	0.900 2	0.901 8	0.901 0	14.7	61.92
＋DW	91.2	0.913 8	0.891 3	0.902 4	14.9	60.43
＋KD	92.7	0.916 5	0.905 6	0.911 0	14.9	60.43

4.6　对比实验

为验证改进模型的检测性能，将其与主流图像分类模型做对比实验，包括多种经典常用图像分类网络(ResNet系列[22]、DenseNet系列[23]、EfficientNet系列[24]和Vision Transformer系列[25])以及图像分类最新成果(ConvNext系列[26-27]和EVA系列[28-29])，实验结果如表3所示，其中输入图像尺寸均为448×448。由表可知，

虽然卷积神经网络 ResNet-34 比改进模型具有更小的 FLOPs 和更快的速度，但精度相差 5.3%。如图 10(a)和 10(b)所示，对于存在缺陷但位于线路外的情况，ResNet-34 依旧将其分类为缺陷(类别为脏点，置信度在 0～1 之间达到了 0.919)，而改进模型则成功将其分类为无缺陷(类别为无缺陷，置信度为 0.696)。图 10(c)和(d)则反映了线路上存在类似缺陷结构时的检测情况，可以发现，ResNet-34 将其误判为缺陷(类别为贝壳，置信度为 0.792)，而改进模型则成功将其判断为无缺陷(类别为无缺陷，置信度为 0.477)。对于 DenseNet169 和 EffecientNet-v2 模型，虽然两者都具有较小的 FLOPs，但在精度和速度上均不及改进模型。MobileViT-v2 模型取得了最高的检测速度，但在精度上仍具有一定劣势。最后，虽然改进模型相较 ConvNext-T 和 EVA02 模型在 Top-1 Acc 指标上下降了 2.4% 和 3.1%，但速度提升了约 16.6% 和



(a) 缺陷位于线路外ResNet-34模型检测结果
(a) Detection result of ResNet-34 model with defect located outside of lines

(b) 缺陷位于线路外改进模型检测结果
(b) Detection result of the modified model with defect located outside of lines

(c) 线路存在类似缺陷结构ResNet-34模型检测结果
(c) Detection result of ResNet-34 model with similar defect structure in lines

(d) 线路存在类似缺陷结构改进模型检测结果
(d) Detection result of the modified mode with similar defect structure in lines

图 10　ResNet-34 模型与改进模型检测效果对比

Fig. 10　Comparison of detection performance between ResNet-34 model and the improved model

表 3　主流图像分类模型性能对比实验

Tab. 3　Results of comparison experiments

Model	Top-1 Acc/%	FLOPs/G	i/(m·s^{-1})
Resnet-34	87.4	14.6	68.97
DenseNet169	91.1	13.7	32.72
EffecientNet-v2	90.6	11.6	28.57
MobileViT-v2	90.5	9.8	73.95
ConvNext-T	95.1	17.8	50.36
EVA02	95.8	87.5	10.62
Ours	92.7	14.9	60.43

469%，更加满足工业实时检测要求。因此，从检测精度和速度综合评价，改进模型在自制数据集上较现阶段图像分类主流模型具有更加均衡的精度和速度。

表 4 显示了在图像分类领域常用公开数据集 ImageNet-1k 上的对比实验结果。由表可知，在类似的输入尺寸下，与高精度大模型相比，改进模型在参数量和计算量上都有明显下降，与轻量级模型相比，改进模型则在精度上有一定优势。因此，改进模型在 ImageNet-1k 公开数据集上也对精度和速度具有良好的适应性。

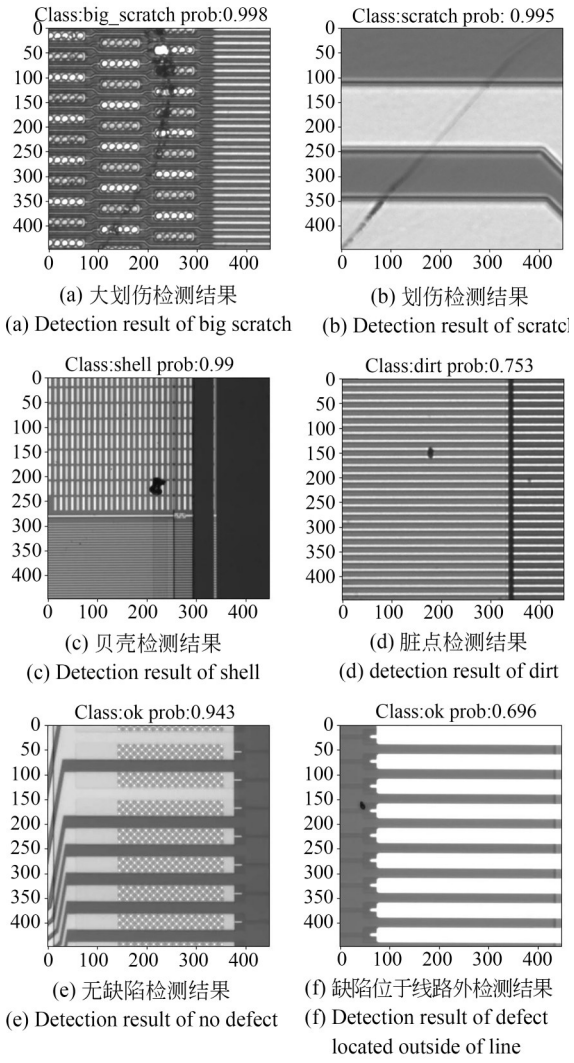
表 4　公开数据集对比实验

Tab. 4　Results of comparison experiments on public dataset

Model	Size	Top-1 Acc/%	Params/M	FLOPs/G
EVA-G/14	336	89.5	1013.01	445.56
ViT-H/14	336	88.6	632.46	363.64
ConvNext-XL	384	87.7	350.20	179.03
Swin-L	384	87.1	196.74	100.28
ResNet-101d	320	83.0	44.57	23.82
MobileViT-v2	384	82.9	14.25	12.35
EffecientNet-v2	288	82.2	13.65	8.16
Ours	384	83.3	27.50	13.89

5　结　论

为了更好更快地解决 TFT-LCD 面板缺陷分类问题，本文在 Swin Transformer 算法的基础上提出了一种改进模型。首先插入 Token 融合模块使模型更加轻量化，然后利用深度可分离卷积模块引入归纳偏置和提高模型分类能力，最后使用知识蒸馏策略来进一步提升模型分类的准确率。实验证明，该改进模型相对于基线模型，在精度仅下降 1.3% 的情况下大大减少了模型的复杂度，从而提升了 17% 的速度，适用于对 TFT-LCD 面板缺陷进行分类的工业任务，其最终检测结果如图 11 所示，其中各类缺陷均被正确分出，大划伤、划伤、贝壳和无缺陷四种类型置信度达到了 0.9 以上，脏点类型由于容易误检所以置信度相对较低，但仍然达到了 0.753。在后续研究中将继续完善模型算法来进一步提高检测精度和检测速度，并补充和完善 TFT-LCD 面板缺陷分类数据集。



(a) 大划伤检测结果　(b) 划伤检测结果

(a) Detection result of big scratch　(b) Detection result of scratch

(c) 贝壳检测结果　(d) 脏点检测结果

(c) Detection result of shell　(d) detection result of dirt

(e) 无缺陷检测结果　(f) 缺陷位于线路外检测结果

(e) Detection result of no defect　(f) Detection result of defect located outside of line

图 11　改进模型检测效果

Fig. 11　Detection results of the modified model

参考文献:

[1] HARALICK R M, SHANMUGAM K, DINSTEIN I. Textural features for image classification [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3(6): 610-621.

[2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR'*05). 20-25, 2005, *San Diego, CA, USA*. IEEE, 2005: 886-893.

[3] STOCKMAN G, SHAPIRO LG. *Computer Vision*[M]. New Jersey: Prentice Hall, 2002: 69-73.

[4] PLATT J C. Sequential minimal optimization: a fast algorithm for training support vector machines [S]. *Microsoft Research Technical Report*, 1998.

[5] KANG S B, LEE J H, SONG K Y, *et al*. Automatic defect classification of TFT-LCD panels using machine learning [C]. 2009 *IEEE International Symposium on Industrial Electronics*. 5-8, 2009, *Seoul, Korea* (*South*). IEEE, 2009: 2175-2177.

[6] HUANG W, LU H T. Defect Classification of TFT-LCD with bag of visual words approach[C]. 2012 *IEEE 6th International Conference on Information and Automation for Sustainability*. 27-29, 2012, *Beijing*. IEEE, 2013: 167-170.

[7] KONG L F, SHEN J, HU Z L, *et al*. Detection of Water-Stains defects in TFT-LCD based on machine vision[C]. 2018 11*th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (*CISP-BMEI*). 13-15, 2018, *Beijing, China*. IEEE, 2019: 1-5.

[8] 肖术明, 王绍举, 常琳, 等. 面向手写数字图像的压缩感知快速分类[J]. 光学 精密工程, 2021, 29(7): 1709-1719.
XIAO S M, WANG S J, CHANG L, *et al*. Compressive sensing fast classification for handwritten digital images[J]. *Opt. Precision Eng.*, 2021, 29(7): 1709-1719. (in Chinese)

[9] 苗传开, 娄树理, 李婷, 等. 基于弱监督学习的多标签红外图像分类算法[J]. 光学 精密工程, 2022, 30(20): 2501-2509.
MIAO C K, LOU S L, LI T, *et al*. Multi-label infrared image classification algorithm based on weakly supervised learning [J]. *Opt. Precision Eng.*, 2022, 30(20): 2501-2509. (in Chinese)

[10] CHIKONTWE P, KIM S, PARK S H. CAD: Co-Adapting discriminative features for improved Few-Shot classification [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 18-24, 2022, *New Orleans, LA, USA*. IEEE, 2022: 14534-14543.

[11] ZHANG Z, XUE Z, CHEN Y, *et al*. *Boosting Verified Training for Robust Image Classifications via Abstraction* [EB/OL]. 2023: *arXiv*: 2303.11552. https://arxiv.org/abs/2303.11552.pdf

[12] CHEN W, GAO Y, GAO L, *et al*. A new ensemble approach based on deep convolutional neural networks for steel surface defect classification [J]. *Procedia CIRP*, 2018, 72: 1069-1072.

[13] FU G, SUN P, ZHU W, *et al*. A deep-learning-based approach for fast and robust steel surface defects classification[J]. *Optics and Lasers in Engineering*, 2019, 121: 397-405.

[14] KONOVALENKO I, MARUSCHAK P, BREZINOVÁ J, *et al*. Steel surface defect classification using deep residual neural network[J]. *Metals*, 2020, 10(6): 846.

[15] HE D, XU K, WANG D. Design of multi-scale receptive field convolutional neural network for surface inspection of hot rolled steels[J]. *Image and Vision Computing*, 2019, 89: 12-20.

[16] HASELMANN M, GRUBER D. Supervised machine learning based surface inspection by synthetizing artificial defects[C]. 2017 16*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*). 18-21, 2017, *Cancun, Mexico*. IEEE, 2018: 390-395.

[17] LIU Z, LIN Y T, CAO Y, *et al*. Swin transformer: hierarchical vision transformer using shifted windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). 10-17, 2021, *Montreal, QC, Canada*. IEEE, 2022: 9992-10002.

[18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. *An image is worth* 16*x*16 *words: transformers for image recognition at scale*[EB/OL]. 2020: *arXiv*: 2010.11929. https://arxiv.org/abs/2010.11929.pdf

[19] BOLYA D, FU C, DAI X, *et al*. *Token Merging: your ViT but Faster*[EB/OL]. 2022: *arXiv*: 2210.09461. https://arxiv.org/abs/2210.09461.pdf

[20] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 1800-1807.

[21] HINTON G, VINYALS O, DEAN J. *Distilling The Knowledge in a Neural Network*[EB/OL]. 2015: *arXiv*: 1503.02531. https://arxiv.org/abs/1503.02531.pdf

[22] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 770-778.

[23] HUANG G, LIU Z, VAN DER MAATEN L, *et al*. Densely connected convolutional networks[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 2261-2269.

[24] TAN M, LE Q V. *EfficientNetV2: Smaller Models and Faster Training*[EB/OL]. 2021: *arXiv*: 2104.00298. https://arxiv.org/abs/2104.00298.pdf.

[25] MEHTA S, RASTEGARI M. *Separable Self-Attention for Mobile Vision Transformers*[EB/OL]. 2022: *arXiv*: 2206.02680. https://arxiv.org/abs/2206.02680.pdf.

[26] LIU Z, MAO H Z, WU C Y, *et al*. A ConvNet for the 2020s[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 18-24, 2022, *New Orleans*, *LA*, *USA*. IEEE, 2022: 11966-11976.

[27] WOO S, DEBNATH S, HU R H, *et al*. ConvNeXt V2: Co-Designing and scaling convnets with masked autoencoders[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 17-24, 2023, *Vancouver*, *BC*, *Canada*. IEEE, 2023: 16133-16142.

[28] FANG Y X, WANG W, XIE B H, *et al*. EVA: exploring the limits of masked visual representation learning at scale[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 17-24, 2023, *Vancouver*, *BC*, *Canada*. IEEE, 2023: 19358-19369.

[29] FANG Y, SUN Q, WANG X, *et al*. *EVA*-02: *a Visual Representation for Neon Genesis*[EB/OL]. 2023: *arXiv*: 2303.11331. https://arxiv.org/abs/2303.11331.pdf.

作者简介：

夏　衍(1998－)，男，安徽安庆人，硕士，2020年于东北大学获得学士学位，2023年于东南大学获得硕士学位，主要从事图像处理、深度学习的研究。E-mail:xiayan@wxautowell.com

罗　晨(1980－)，女，江苏扬州人，博士，副教授，博士生导师，2002年于东南大学获得学士学位，2005年于上海交通大学获得硕士学位，2010年于上海交通大学获得博士学位，主要从事机器视觉、三维测量、机器人等方面的研究。E-mail:chenluo@seu.edu.cn