[引用格式]叶雨婕,胡志超,刁龙,等. 基于改进 MobileNetV3-Small 网络的灰度图像分类方法[J]. 集美大学学报(自然科学版),2026,31(2):217-223. DOI:10.19715/j.jmuzr.2026.02.09.

基于改进 MobileNetV3-Small 网络的灰度图像分类方法

叶雨婕，胡志超，刁　龙，邹　蕾，王云超，林少滨

（集美大学海洋装备与机械工程学院，福建 厦门 361021）

[摘要] 针对 MobileNetV3 网络中 SE 模块的通道注意力局限性，提出了一种融合通道和空间注意力机制（SE-SAM Fusion）的图像分类方法。通过背景一致化、剪切和缩放等策略对原始图像进行数据扩增，开展了与 ResNet18、ShuffleNetV2 和 EfficientNet 等轻量化模型的对比分析，并执行了消融实验以验证改进模块的有效性。结果表明：改进的 MobileNetV3-Small 网络在自制的混合增强数据集上的平均准确率为 96.64%，损失值为 0.015 8，相较于初始模型，损失值降低 0.006 4；模型权重文件为 6.2 MB，分别约为同类轻量级模型 ResNet18 的 1/7，以及 EfficientNet 的 1/4。

[关键词] 图像分类；MobileNetV3-Small；注意力机制；迁移学习

[中图分类号]TP 391.41　[文章编号]1007－7405(2026)02－0217－07　DOI:10.19715/j.jmuzr.2026.02.09

Gray-Scale Image Classification Based on Improved MobileNetV3-Small

YE Yujie, HU Zhichao, DIAO Long, ZOU Lei, WANG Yunchao, LIN Shaobin

(College of Marine Equipment and Mechanical Engineering, Jimei University, Xiamen 361021, China)

Abstract: Aiming at the channel attention limitation of the SE module in the MobileNetV3 network, an image classification method based on SE-SAM Fusion attention mechanism is proposed. The original image was amplified using background uniformity, clipping, and scaling strategies, compared with lightweight models such as ResNet18, ShuffleNetV2, and EfficientNet, and ablation experiments were performed to verify the effectiveness of the improved module. The results indicate that the improved MobileNetV3-Small network achieves an average accurance of 96.64% on the mixed-enhanced dataset with a loss value of 0.015 8, marking a 0.006 4 reduction from its initial model. The model's weight file is 6.2 MB, which is approximately 1/7 of that of the similar lightweight model ResNet18 and 1/4 of that of EfficientNet, respectively.

Keywords: image classification; MobileNetV3-Small; attention mechanism; transfer learning

0　引言

深度学习技术的发展为视觉检测领域带来了自动化和精确性的革新。目前，许多工厂在进行图像分类和缺陷检测等视觉任务时，仍然依赖于传统的机器学习方法。例如，

[收稿日期] 2024－08－21
[基金项目] 国家自然科学基金面上项目子课题（52375046）；福建省自然科学基金项目（2021J01852）
[通信作者] 王云超，教授，从事多轴车辆的转向系统和悬架系统研究，E-mail:ychaowang@jmu.edu.cn

通过模板匹配等方法提取特征，然后使用支持向量机[1]、K-近邻[2]、朴素贝叶斯[3]等分类器进行分类。而这些传统方法由于误检率高、适应性差、特征提取依赖人工设计等局限性，已经逐渐被淘汰，而结合深度学习的卷积神经网络，如 GoogLeNet[4]和 VGG[5]等可在训练过程中自动学习图像特征，展现出了更好的适应性和泛化能力。

但是，在工业生产领域，获取大量标注数据集往往是具有挑战性的，数据集往往规模较小，难以满足传统卷积神经网络大量训练样本的需求。同时，随着网络规模的扩大，参数数量的增加也容易导致过拟合[6]，而迁移学习可以显著提高小数据集上模型的性能。于是，结合迁移学习和残差学习的深度网络成为有效的解决方案，如 ResNet[7]、MobileNet[8]、SqueezeNet[9]等。在此基础上，常江等[10]提出了一种基于改进生成对抗网络和 MobileNetV3 的带钢缺陷分类方法，通过调整卷积层和通道数，预测时间提高了 16.7%，准确率提高了 2.62%。胡名琪等[11]基于预训练 ResNet50 模型，采用迁移学习并重新建立全连接层的方法，对铜合金金相类别进行分类识别，克服了金相图像数据集小的问题。Zhu[12]等提出了一种具有迁移学习的改进型 MobileNetV3-Small 网络，解决了 PSPR 胶囊图像样本数量少导致的分类精度低、模型收敛慢问题。Woo 等[13]通过在通道和空间两个维度上引入注意力机制（CBAM），强调有意义的特征并抑制不必要的特征，可有效帮助信息在网络中流动。该研究还通过在 ImageNet-1K、MS COCO 检测和 VOC 2007 检测数据集上的广泛实验，证明了 CBAM 在多种模型上的分类和检测性能都有一致的改进，展示了 CBAM 的广泛适用性。上述研究针对小样本问题和模型轻量化实施了多种优化策略，提升了模型的整体性能。

尽管现有研究已取得显著进展，但多数模型的数据集仍是三通道图像，针对单通道灰度图像的预训练模型和空间注意力机制的探索仍然不足。本文以工业生产时拍摄的背带扣注塑件灰度图像为例，提出了一种改进 MobileNetV3-Small 网络注意力机制的图像分类方法，以提高模型在小样本灰度图像数据集上的适应性。

1 混合背景数据集制作

本研究使用的相机为海康威视 MV-CS200-10gm 面阵网口工业相机，使用的镜头焦距为 12 mm，使用相机驱动采集的图像为 5472 像素 ×3648 像素，图片格式为 JPG。共收集了 357 张原始背带扣图像，根据工件的外形轮廓将其分为 8 类。

为了缓解数据不平衡、数据样本量小和背景阴影对数据集质量的影响，提高模型的自适应性和稳定性，采取以下措施：1）欠采样与过采样。对于数量较多的样品进行欠采样，而对于其他样品进行过采样，使每类样品图像数量达到 50 张。2）背景一致化。为了消除样品图像背景中的黑色阴影，抠出原图像中的工件并统一添加白色背景。3）数据扩增。采用随机旋转、剪切、随机缩放和随机改变亮度等数据增强策略扩充数据集至 10 000 张图像，训练集和验证集按 7:3 划分。

图像预处理及数据增强的变化效果如表 1 所示。

表 1　不同数据增强技术下的图像效果对比

Tab. 1　Comparison of image effects under different data augmentation techniques

样品	应用变化			
	无	随机改变亮度	随机旋转	随机缩放
1				
2				
3				
4				
5				
6				
7				
8				

2　SE-SAM Fusion 注意力机制

为了提高轻量化模型在小数据集上的分类性能，通过引入预训练权重参数和增加 dropout 层，提升模型的泛化能力，并用 SE-SAM Fusion 注意力模块替换单一的 SE 注意力机制[14]，强化了重要的空间区域，提升了模型的特征整合能力，而不会显著增加计算量或推理时间。改进后的注意力机制整体结构如图 1 所示。

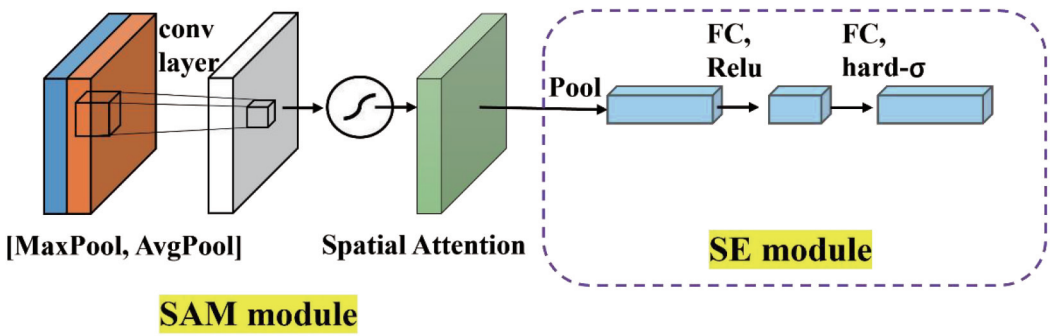


图 1　SE-SAM Fusion 注意力机制

Fig.1　SE-SAM Fusion attention mechanism

图 1 中 SE module 部分即为 MobileNetV3-Small 中采用的 SE 注意力机制，主要通过挤

压（squeeze）和激励（excitation）操作提取图片的通道特征。输入的特征图经过全局平均池化（GAP）将所有通道信息汇总成单一数值，提取每个通道强度分布，再将其送入两个全连接层 FC1 和 FC2，最后逐元素相乘应用至原始特征图中。尽管 SE 模块的计算量相对较小，但只考虑了通道维度上的注意力，无法捕捉空间维度上的特征。

由于灰度图像只有一个颜色通道，通道特征相对较少。为加强模型对空间某些位置的特征提取，在原 SE 模块的基础上引入空间注意力机制 SAM。解决方案涉及的公式有：

$$\text{avgout} = \frac{1}{C}\sum_{c=1}^{C} I_c, \tag{1}$$

$$\text{maxout} = \max C_{c=1}\, I_c, \tag{2}$$

$$x_1 = \text{concat}(\text{avgout}, \text{maxout}), \tag{3}$$

$$x_2 = \text{Conv2}D(x_1, W), \tag{4}$$

$$x_3 = \sigma(x_2), \tag{5}$$

$$\text{out} = I \odot x_3。 \tag{6}$$

其中：I 是输入特征图；C 是通道数；I_c 是第 c 个通道的特征图；W 是卷积核；Conv2D 表示二维卷积操作；σ 是 Sigmoid 激活函数，用于将卷积层的输出归一化到 0 和 1 之间；⊙表示逐元素相乘。

该模块通过在最后一个维度（通道维度）上应用平均池化 avgout（式（1））和最大池化 maxout（式（2）），分别计算输入特征图的通道平均值和通道最大值，再将 avgout 和 maxout 在最后一个维度上进行 concat 得到式（3）中所示的 x_1，形成一个新的张量并进行 1 × 1 卷积操作（式（4））。最后，空间注意力权重将计算得到的注意力权重与原始输入特征图 inputs 进行逐元素相乘，以便从平均值和最大值中学习到空间注意力权重（式（6））。

3 实验结果与分析

3.1 测试环境和评估指标

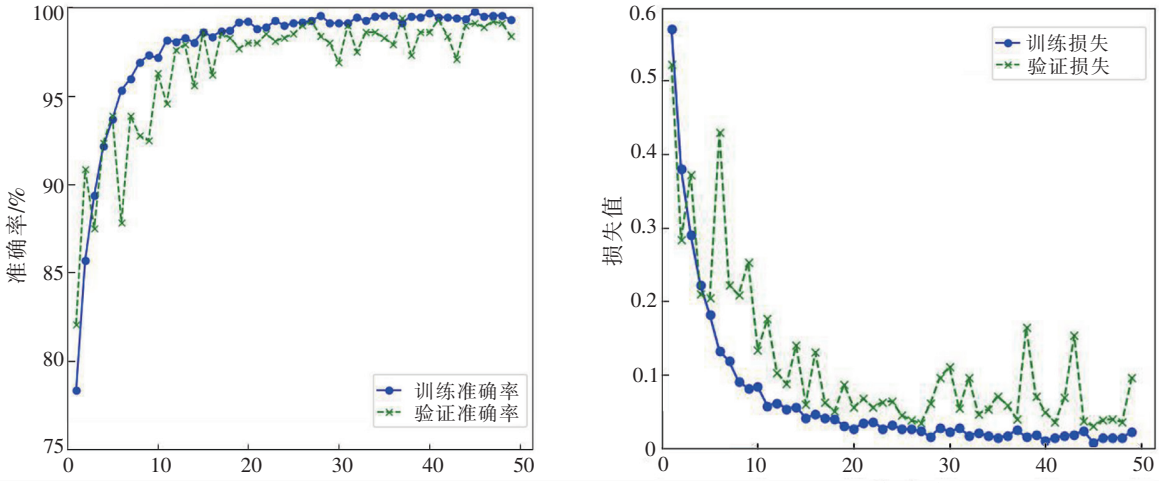
在本研究中，网络模型的训练采用了 TensorFlow2. 11. 0 深度学习框架。实验所用的硬件环境包括：CPU 为 Intel（R）Core（TM）i7-12700，GPU 为 NVIDIA GeForce RTX 3060。实验在 Ubuntu20. 04 操作系统上进行，使用的 Python 版本为 3. 8。在本次实验设置中，为寻求最优性能，对网络的超参数进行了细致的调整。模型训练的总轮次（epochs）定为 50 轮，每一轮训练的批次（batch size）为 16 次。经过多次实验确定学习率为 0. 000 1 时，模型表现最佳。

为了更直观、全面地评估模型的整体性能，使用了准确率（accuracy）、损失值（loss）、召回率（recall）、$F1$ 分数、AUC 曲线对改进前后的 MobileNetV3-Small 网络进行评估。

3.2 数据集质量评估

图 2 为 MobileNetV3-Small 在混合增强数据集上测得的准确率和损失值。随着训练轮次的增加，图 2a 中的验证准确率不断向训练准确率靠拢且差值越来越小，最终没有超过训练准确率，这表明模型在训练时并没有出现过拟合的现象。图 2b 中的训练损失和验

证损失持续下降，尽管验证损失存在震荡现象，但整体仍呈下降趋势，表明模型正在有效学习，且具有进一步优化参数提升性能的潜力。



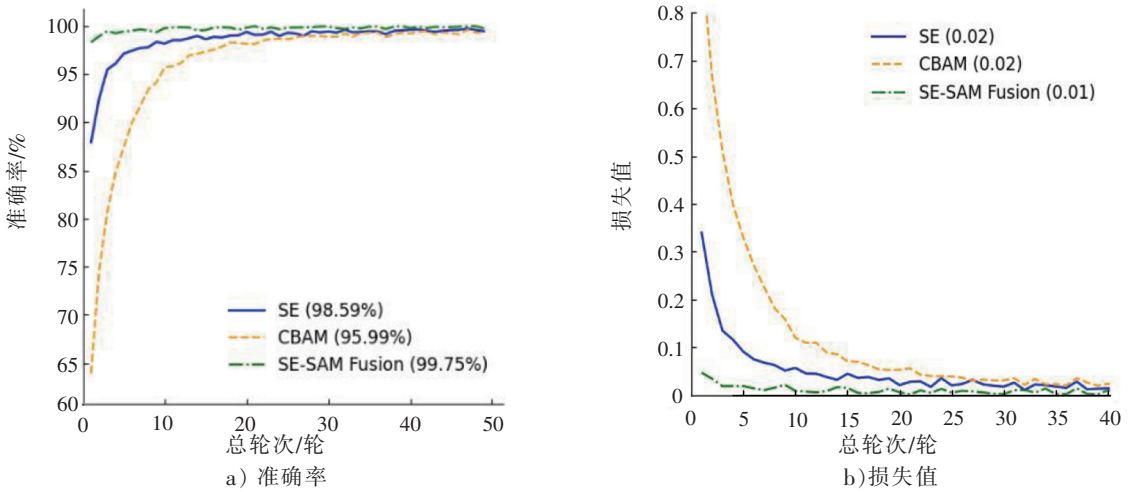
a）准确率　　b）损失值

图 2　MobileNetv3-Small 在混合增强数据集上的准确率和损失值

Fig.2　Accuracy and loss value of MobileNetv3-Small in hybrid enhanced dataset

3.3　注意力模块对比分析

本文提出的改进措施在特征提取的方法论上与现有研究是一致的。为评估所添加的注意力模块对于整个模型的贡献，将该 SE-SAM Fusion 与 SE 注意力模块和 CBAM 注意力模块的训练效果进行对比。CBAM 是一种轻量级注意力模块，可以根据中间特征图，沿着独立的（通道和空间）维度顺序推断注意力图，后进行自适应特征细化。图 3 展示了三种注意力模块训练结果的对比分析。图 3a 显示了训练准确率随着迭代次数的增加而稳步提升，其中 SE-SAM Fusion 模块在所有模块中实现了轮次的训练准确率 99.75%。图 3b 则显示了随着训练的深入，损失值持续降低，在最后一个轮次（epoch）内，SE-SAM Fusion 模块达到了最低损失值 0.01，两组图均表明 SE-SAM Fusion 注意力模块在特征提取效率和收敛速度上的优越性。



a）准确率　　b）损失值

图 3　三类注意力模块下模型的训练结果对比

Fig.3　Comparison of training results of three kinds of attention modules

在消融实验中，本文对 SE-SAM Fusion 注意力模块进行了评估。通过对比仅包含 SE 模块的对照组与集成了 SE-SAM 融合模块的完整模型，验证了新模块的有效性。从图 3 能看出添加 SAM 模块后模型的训练精度更高，收敛速度更快，训练值更接近于实际值。

3.4 模型性能对比分析

为了评估改进模型的性能，在统一实验条件下，对比了改进的模型与初始模型以及其他模型的整体性能指标。表 2 显示，在反映模型复杂度和预测效果的两项关键指标（权重文件大小（weights size）和损失值）上，MobileNetV3-Small 表现突出。特别是改进的 MobileNetV3-Small，在准确率、召回率、*F*1 分数维持稳定的前提下，损失值降低 0.006 4。相较于其他对比模型，改进的 MobileNetV3-Small 在保持较高准确率、精确率、召回率与 *F*1 分数的同时，权重文件仅 6.2 MB，显著小于 MobileNetV3-Large、EfficientNet、VGG16 及 ResNet18 等模型，轻量化优势明显；其损失值（0.0158）也处于较低水平，综合性能优异，更适用于背带扣图像分类任务。

表 2 不同模型下训练结果

Tab. 2 Training results under different models

模型	权重文件大小/MB	准确率/%	损失值	精确率/%	召回率/%	*F*1	AUC
MobileNetV3-Small	6.2	96.72	0.022 2	96	95	0.95	0.99
改进的 MobileNetV3-Small	6.2	96.64	0.015 8	96	95	0.95	0.99
MobileNetV3-Large	17.0	96.98	0.013 3	96	94	0.94	0.99
ShuffleNetV2	5.2	94.13	0.016 0	95	95	0.94	1.00
EfficientNet	26.4	85.91	0.024 1	90	89	0.88	0.98
VGG16	281.3	95.35	0.016 5	97	96	0.96	1.00
ResNet18	46.9	96.83	0.003 2	96	95	0.95	0.99

4 结论

针对工业生产中 8 类背带扣样品的灰度图像分类检测问题进行了研究，取得了以下结论：1）利用多种预处理和数据增强方式制作了混合增强数据集，在 MobileNetV3-Small 模型上训练效果稳定，收敛速度快，损失曲线下降较稳定，验证了混合增强数据集的有效性。2）改进了 MobileNetV3-Small 网络中的 SE 注意力模块，在 SE 机制前加入空间注意力机制，并在模型中添加 dropout 层和引入预训练权重，融合了三种独立训练策略的优化效果。训练结果显示，相较于基线模型，改进的 MobileNetV3-Small 网络的损失曲线震荡幅度明显减小，损失值最小达到了 0.015 8。3）与多种基于深度学习的图像分类网络相比，改进的 MobileNetV3-Small 模型权重文件小（6.2 MB），混淆矩阵误分类率低，凸显了其在精确识别和分类背带扣图像方面的高效性。

[参 考 文 献]

[1] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.

[2]SYRIOPOULOS P K,KALAMPALIKIS N G,KOTSIANTIS S B,et al. KNN classification:a review [J]. Annals of Mathematics and Artificial Intelligence,2023,93:43-75.

[3]RISH I. An empirical study of the naive Bayes classifier[C]//Proceedings of the IJCAI-01 Workshop on Empirical Methods in Artificial Intelligence. Seattle:IJCAI,2001:41-46.

[4]SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston:IEEE,2015:1-9.

[5]DING X,ZHANG X,MA N,et al. RepVGG:making VGG-style ConvNets great again[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville:IEEE,2021:13728-13737.

[6]CHEN L Y,LI S B,BAI Q,et al. Review of image classification algorithms based on convolutional neural networks [J]. Remote Sensing,2021,13(22):4712.

[7]GAO S H,CHENG M M,ZHAO K,et al. Res2Net:a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,43(2):652-666.

[8]HOWARD A G,ZHU M,CHEN B,et al. MobileNets:efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17)[2024-08-21]. https://doi.org/10.48550/arXiv.1704.04861.

[9]IANDOLA F N,HAN S,MOSKEWICZ M W,et al. SqueezeNet:AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24)[2024-08-21]. https://doi.org/10.48550/arXiv.1602.07360.

[10]常江,管声启,师红宇,等. 基于改进生成对抗网络和 MobileNetV3 的带钢缺陷分类 [J]. 激光与光电子学进展,2021,58(4):221-226.

[11]胡名琪,刘秋明,陈辉明,等. 基于改进残差网络与迁移学习的铜合金金相图分类方法 [J]. 有色金属科学与工程,2024,15(5):690-699.

[12]ZHU J,ZHANG C H,ZHANG C J. Papaver somniferum and papaver rhoeas classification based on visible capsule images using a modified MobileNetV3-Small network with transfer learning [J]. Entropy,2023,25(3):447.

[13]WOO S,PARK J,LEE J Y,et al. CBAM:convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Munich:Springer,2018:3-19.

[14]HU J,SHEN L,SUN G,et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011-2023.

(责任编辑　彭海滨　英文审校　黄振坤)