doi:10.3969/j.issn.1672-4348.2023.03.009

基于改进 MobileNet 的咖啡豆缺陷检测

叶建华,唐辉,罗奋翔,徐欢,徐帅龙

(福建工程学院 机械与汽车工程学院,福建 福州 350118)

摘要:针对现有咖啡豆缺陷检测方法鲁棒性、实时性不高的问题,提出基于改进 MobileNet 的咖啡豆缺陷检测方法,构建以 MobileNet 为核心的轻量级检测网络。对咖啡豆检测样本进行采集和增强,构建包含5种类型的咖啡豆缺陷检测数据集。通过卷积通道数和卷积模块的优化调整,压缩模型参数以匹配分类任务和满足边缘设备的部署要求。引入 Mish 激活函数和学习率的自适应调整方法,提升模型的收敛性能。利用迁移学习的方式优化模型参数,进一步提升模型识别准确率。实验表明,改进模型在自建咖啡豆分类检测数据集上的平均准确率为 96.13%,较原模型平均准确率的 93.17%提升了 2.96%;模型参数则从 3.21×10⁶ 减少到了 0.15×10⁶,不到原模型的 5%。相比于 VGG16_bn、ResNet50、SqueezeNet 和 MobileNet,改进模型准确率分别高出 0.65%、1.39%、2.39%、2.96%。在提升精度的同时,该方法在参数量、内存占用量和浮点运算量上表现也为最优,能为相关农产品的缺陷检测提供参考。

关键词:图像识别;卷积神经网络;深度学习;MobileNet;缺陷检测;咖啡豆

中图分类号: TP391.4

文献标志码: A

文章编号: 1672-4348(2023)03-0257-07

Coffee bean defect detection based on improved MobileNet

YE Jianhua, TANG Hui, LUO Fenxiang, XU Huan, XU Shuailong (School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China)

Abstract: A coffee bean defect detection method based on improved MobileNet was proposed aiming at the problems of low robustness and poor real-time performance of existing coffee bean defect detection methods. A lightweight detection network based on MobileNet was constructed. A coffee bean data set with five defect types was constructed. The parameter numbers of the model were reduced by optimizing and adjusting its convolution channel number and convolution module. Therefore, it can match the classification task and meet the deployment requirements of edge devices. Mish activation function and adaptive adjustment method of learning rate was applied to improve the convergence performance of the model. Transfer learning was used to optimize the model parameters and further improve the model accuracy. Experiment results show that the average accuracy of the improved model on the self-built dataset was 96.13%, which was 2.96% higher than the 93.17% of the original model. The model parameters were reduced from 3.21×10⁶ to 0.15×10⁶, which was less than 5% of the original model. The accuracy of the improved model was 0.65%, 1.39%, 2.39% and 2.96% higher than that of VGG16_bn, ResNet50, SqueezeNet and MobileNet, respectively. While improving the accuracy, this method also has the best performance in terms of parameters, memory usage and floating-point operations. Therefore, it can provide a reference for the defect detection of related agricultural products.

Keywords: image recognition; convolutional neural network; deep learning; MobileNet; defect detection; coffee bean

随着图像处理和机器视觉技术的快速发展, 大量学者将其应用于咖啡豆的缺陷检测,以克服 人工筛选的缺点。在特征明显咖啡豆的简单分类 场景下, Gunadi 等[1] 通过提取阈值分割后的咖啡 豆形状特征来区分正常咖啡豆和破损咖啡豆。 Arboleda 等[2]利用 RGB 空间的数值差异区分正 常咖啡豆和黑豆。García 等[3]则通过图像处理方 法提取咖啡豆中的颜色、圆度和面积特征,并将特 征输入 KNN 算法中对 6 种咖啡豆缺陷进行分类, 提升缺陷的识别能力。Oliveri等[4]则在暗箱环境 下通过高光谱成像拍摄咖啡豆,共采集120颗,利 用 K 近邻(K-nearest neighbor)算法识别 4 种颜色 进行品质分析,准确率最高为90.2%,最低为 50%,由于在该文献中实验数据较少,且需要在特 殊环境下进行拍照检测,因而可信度较低。Chen 等[5]利用推扫可见-近红外(push-broom visiblenear infrared) 高光谱传感器获取咖啡豆图像,选 择合适的波段和一个光谱特征输入支持向量机中 区分正常豆和虫蛀豆,总体检测准确率为95%。 Oliveira 等[6]通过从图像中提取相关特征,例如颜 色直方图和纹理分析等,将特征支持向量机输入 进行咖啡豆分类,总体检测准确率为96.5%。然 而,上述方法均依赖人工特征的设计,精确度、鲁 棒性不高,适用场景受限,难以满足高品质筛选要 求。近年来,深度学习技术由于其强大的特征自 提取能力,获得快速发展。伴随深度网络性能的 提升,模型的体积也越来越大,模型参数和计算量 十分庞大,通常只能在高算力的 GPU 上运行,难 以应用到运算能力受限的农业检测实际场景中。 刘阳等[7] 通过在 SqueezeNet 中调节卷积核数目 和移动 fire 模块位置对叶片病害模型进行改进, 在保持相近性能的情况下,大幅压缩了网络参数 和计算量。马本学等[8] 通过将 ResNet50 中的大 卷积核替换为小卷积核,修改下采样模块中的通 道数对大枣品质检测模型进行优化,将参数减少 约 6.9×10⁶ 日准确率高于原模型。

为此,本文在自建缺陷咖啡豆数据集的基础上,以 MobileNet^[9] 为基础网络,优化调整卷积通道数和模块以匹配分类任务,压缩参数规模满足边缘设备的要求;改进激活函数和学习策略,提升模型的收敛性能;引入迁移学习方法,通过微调优化模型参数,进一步提升模型识别准确率。

1 数据集构建

选取日晒法处理后的云南咖啡生豆作为研究对象,分类检测贝壳豆、破损豆和黑豆。考虑正常咖啡豆的正反面纹理、特征存在明显差异,将合格豆正面和背面做成不同的类别进行检测。检测数据集分为5个类别,部分样本如图1所示。



图 1 不同类别咖啡豆 Fig.1 Different types of coffee beans

图像由型号为 MER-630-60U3C 的彩色 CCD 相机采集,镜头焦距 8 mm,光源为 LED 白色光,采集分辨率为 3 088×2 064,图像保存格式为 BMP。每次拍照前在翻拍台上放入一批样本拍照采集,单粒咖啡豆由整幅图像裁剪获得,分辨率为 224×224。通过灰度变换,阈值分割获取咖啡豆轮廓,沿咖啡豆轮廓扩充并裁剪得到单粒咖啡豆图片集。为防止数据集过少导致模型训练不充分,对数据进行扩充,将数据集水平和垂直镜像翻转后分别旋转90°、180°、270°。扩充前后的数据量分别为 863 和10 356。扩充后数据见表 1 所示。

2 模型设计与优化

2.1 MobileNet 模型

MobileNet 网络的主要特性在于使用深度可分离卷积代替普通卷积,以大幅降低模型的计算量和参数量。MobileNet 网络结构如图 2 所示,其中,c表示常规卷积核个数,s表示卷积步长。第一层是常规的卷积方式,第二层开始以多个深度

可分离卷积模块进行堆叠。

表 1 咖啡豆缺陷检测数据集

Tab.1 Coffee bean defect detection data set

训练集	测试集	总数
1 104	684	1 788
1 056	756	1 812
2 160	1 440	3 600
1 128	756	1 884
924	348	1 272
6 372	3 984	10 356
	1 104 1 056 2 160 1 128 924	1 104 684 1 056 756 2 160 1 440 1 128 756 924 348

深度可分离卷积原理如图 3 所示。深度可分离卷积将卷积步骤分为深度卷积和逐点卷积。常规的卷积方式是以多层卷积核为一组将卷积核和网络上一层输出的特征图全部卷积,卷积核个数等于输出的特征图通道数。而深度卷积核以单层为一组与网络上一层输出的特征图分别卷积,卷积核个数等于上一层特征图的通道数。逐点卷积过程和常规卷积方式相同,只是卷积核大小为 1×1。每个输入通道通过单个卷积核的深度卷积公式如下:

$$G_{i,j,m} = \sum_{w,h} K_{w,h,m} \cdot X_{i+w,j+h,m}$$
 (1)

式中, $G_{i,j,m}$ 为第 m 通道上的 i、j 坐标输出特征图; $K_{w,h,m}$ 为卷积核; $X_{i+w,j+h,m}$ 为输入特征图;w、h 为第 m 通道的卷积核权重元素坐标。

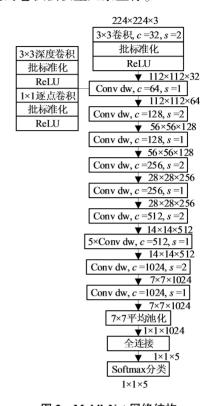


图 2 MobileNet 网络结构
Fig.2 MobileNet network structure

输入层 3×3 深度卷积 1×1 逐点卷积 输出层 $D_{\rm F}\times D_{\rm F}\times M \qquad (D_{\rm K}\times D_{\rm K}\times 1)\times N \qquad (1\times 1\times M)\times N \qquad D_{\rm o}\times D_{\rm o}\times N$

图 3 深度可分离卷积原理

Fig.3 Principle of deep separable convolution

深度可分离卷积与标准卷积的计算量对比 如下:

$$\frac{P_1}{P_2} = \frac{D_F^2 D_K^2 M + D_F^2 M N}{D_F^2 D_K^2 M N} = \frac{1}{N} + \frac{1}{D_K^2}$$
 (2)

式中, D_k^2 为输入特征图大小; D_k^2 为深度卷积卷积核大小;M 为卷积通道数;N 为特征图通道数; P_1 为深度可分离卷积计算量; P_2 为标准卷积计算量。

从式(2)可看出,深度可分离卷积计算量的减

少与卷积核大小和输出特征图通道数有关,在卷积核大小为3×3时,计算量是常规卷积的1/9。

2.2 模型优化

深度学习由于其强大的特征提取能力被广泛 应用于各种分类难度高的检测任务中,工业中的 缺陷检测任务往往背景单一,特征明显,一般的卷 积神经网络就能取得较好的分类效果,对网络结 构进行改进是一种常用的优化方法,如在网络中 嵌入注意力机制,但类似的网络结构改进会增加模型复杂性且对于与本文相近的检测任务(背景单一,特征明显)往往优化效果不明显,因此,本文将网络优化的重点放在常规优化手段和网络的压缩上。

2.2.1 网络模型简化

MobileNet 属于轻量化网络,在 ImageNet^[10]上的 top1 准确率达到 70.6%,本文中的分类任务 难度远小于 ImageNet,为适应在运算性能受限的 边缘设备部署的要求,模型参数有进一步的压缩 空间。拟采用的压缩方式有调整模型每层的输出 通道数和删除卷积层两种。具体做法是将每层的 通道数压缩为原来的四分之一并对模型的最后一个卷积层进行删减。压缩后的网络结构如表 2 所示,其中"dw 卷积"表示深度可分离卷积。

表 2 压缩后的网络结构

Tab.2 Compressed network structure

输入大小	操作	卷积核个数	卷积步长
224 ² ×3	普通卷积	8	2
$112^2 \times 8$	dw 卷积	16	1
$112^2 \times 16$	dw 卷积	32	2
$56^2 \times 32$	dw 卷积	32	1
$56^2 \times 32$	dw 卷积	64	2
$28^2 \times 64$	dw 卷积	64	1
$28^2 \times 64$	dw 卷积	128	2
$14^2 \times 128$	5×dw 卷积	128	1
$14^2 \times 128$	dw 卷积	256	2
$7^2 \times 256$	平均池化	_	_
1 ² ×256	全连接	_	

2.2.2 激活函数

激活函数可以使网络输出非线性,强化网络的学习能力,原 MobileNet 采用的是 ReLU^[11]激活函数, ReLU 激活函数由于其构造简单能减少模型计算量,且相较于 Sigmoid 和 Tanh 激活函数可有效减少梯度消失问题被广泛采用,但由于对负值的输入输出为 0,同样也有一定的梯度消失问题。选择 Mish^[12]激活函数改进网络, Mish 激活函数公式如式(3)。 Mish 是非单调且连续的光滑函数,在输入小于零时也能有一定输出。

$$Mish(x) = x \cdot Tanh(ln(1+e^{x}))$$
 (3)
式中, $Tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

2.2.3 学习率衰减策列

设定一个合适的学习率能够加快模型收敛速度以及使模型达到较高的准确率。相较于固定的学习率,在不同的训练阶段采用不同学习率往往能取得更好的效果。本文选取自适应调整学习率方法更新学习率,自适应调整学习率基于训练集误差动态衰减学习率,当模型的准确率不再上升时,降低学习率。除了学习率之外,还需设置学习率衰减系数和接受学习率不下降次数。衰减系数过小会导致训练速度变慢,甚至无法找到极值点,因此决定学习率衰减系数0.9,神经网络训练中有时会出现准确率不变,损失函数却在降低的情况,为使这种状况下损失值能充分减小,设定学习率不下降的次数为30。

2.2.4 迁移学习

迁移学习是将在大型数据集上预训练得到的网络参数引入自己的网络中,比起随机初始化,迁移学习初始化有着更好的特征提取能力。由于pytorch中没有提供 MobileNet 的预训练参数,选择 Kaggle 上的 ImageNet 数据集子集进行预训练,该数据集训练集总共有 13 万张图片。

3 结果和分析

3.1 实验环境和评价指标

使用的软件配置为 Pytorch 深度学习框架, Python3.8 编译环境、Anaconda 软件、CUDA 架构、cuDNN 开发库以及 Pycharm 编译器。硬件配置 CPU 为 Intel(R) Core(TM) i7-11800H @ 2.30 GHz, GPU 为 NVIDIA GeForce RTX 3060 Laptop,内存为 16 GB。

选择 Adam 优化器,批尺寸为 32,总迭代轮数 为 100 轮,损失函数选择交叉熵损失,学习率 0.001。

为评估改进后模型分类效果,引入准确率 A,精确率 P,召回率 R,F₁ 值作为评价指标,具体公式如下:

$$A = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{4}$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$R = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

式中,TP 为真实值为正且预测也为正的数量;TN 为真实值为负且预测也为负的数量;FP 为真实值为负但预测为正的数量;FN 为真实值为正但预测为负的数量。

3.2 消融实验

消融实验如表 3 所示。由于模型收敛后准确率会随着模型每轮迭代产生略微波动,为消除略微波动的随机性影响,选择迭代轮数最后十轮的平均准确率。

表 3 消融实验结果

Tab.3 Ablation experimental results

因素					A /07-	
e	c	a	l	t	$A_{\rm c}/\%$	$P_{\rm a}$
_	_	_	_	_	86.17	3.21×10 ⁶
\checkmark	_	_	_	_	93.17	3.21×10^{6}
\checkmark	$\sqrt{}$	_	_	_	92.69	0.15×10^6
$\sqrt{}$	_	\checkmark	_	_	93.43	3.21×10^{6}
$\sqrt{}$	_	_	$\sqrt{}$	_	94.44	3.21×10 ⁶
$\sqrt{}$	_	_	_	$\sqrt{}$	95.56	3.21×10^{6}
$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	96.13	0.15×10^6

注:e 为数据增强;c 为网络压缩;a 为 Mish 激活函数;l 为自适应调整学习率;t 为迁移学习; A_c 为测试集准确率; P_a 为模型参数量。

可以看出,相比原始数据集,数据增强将模型准确率提升了6.99%,这表明数据增强能使网络更充分地提取咖啡豆特征,为使模型各项改进不受数据集过小产生的不稳定性影响,其他改进方法均在数据增强的条件下进行;压缩后的模型在准确率下降0.47%的基础上大幅压缩了参数量,这表明针对咖啡豆分类任务,原始模型较为冗余,有进一步压缩的空间;将 Mish 激活函数替换ReLU 激活函数后,准确率提升了0.27%;自适应调整学习率使模型准确率提升1.28%,这表明在训练过程中,相较于固定的学习率,变化的学习率能取得更好的效果;迁移学习使模型准确率提升1.69%,迁移学习的参数保留了网络从大型数据

集中学习的特征,尽管用来迁移的数据集和本文的缺陷检测任务差异性较大,在数据量大的情况下,仍可以提升网络的特征提取能力。将各优化手段和网络改进方法综合应用到 MobileNet 中,最终准确率为 96.13%,相较于原始未改进只通过数据增强的网络,准确率提升 2.96%,将网络参数大幅压缩为原来参数量的 5%。

3.3 混淆矩阵可视化

在测试集上对改进后的网络性能进行测试,生成的混淆矩阵如表 4 所示,根据混淆矩阵算出的精确率、召回率、F₁ 值和准确率如表 5 所示。从表中可以看出,模型对黑豆的识别准确率最高,全部的黑豆都能正确识别,这是因为黑豆比其他类别有较大的颜色差异,容易区分。对于合格咖啡豆的正反面都有较高的精确率,分别为 96.88%和 98.43%。破损豆和贝壳豆由于形状不规则存在一定相似性,对这两种缺陷的分类容易混淆,有10.2%的破损豆被分类为贝壳豆,10.9%的贝壳豆被分类为破损豆,召回率也较低,分别为 85.58%和 88.51%。模型总的准确率为 96.13%,在仅用于区分合格品和不合格品时,准确率为 99.9%。以上结果表明,本文提出的改进模型能较好地识别咖啡豆中各类缺陷。

表 4 混淆矩阵结果

Tab.4 Confusion matrix results

	真实值				
预测值	黑豆 破损豆		合格豆 正面	合格豆 背面	贝壳豆
黑豆	1 440	0	0	0	0
破损豆	0	647	0	2	38
合格豆正面	0	20	682	0	2
合格豆背面	0	12	0	754	0
贝壳豆	0	77	2	0	308

3.4 与其他网络对比

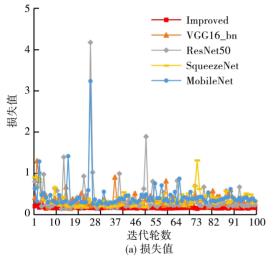
将模型和 VGG16_bn、ResNet50、SqueezeNet 以及未改进的 MobileNet 进行对比,其中 VGG 和 ResNet 是传统的大型网络, SqueezeNet 和 MobileNet 是轻量化网络。其他网络均在迁移学习条件下用相同数据集训练测试,各模型在测试集上准确率以及损失函数对比如图 4 所示,图中

Improved 表示本文改进模型。

表 5 各类别精确率、召回率、 F_1 值和准确率

Tab.5 Accuracy rate, recall rate, F_1 value and accuracy rate of each category

类别	精确率	召回率	F ₁ 值	准确率
黑豆	1	1	1	0.961 3
破损豆	0.941 8	0.855 8	0.896 7	0.961 3
合格豆正面	0.968 8	0.997 1	0.982 7	0.961 3
合格豆背面	0.984 3	0.997 4	0.990 8	0.961 3
贝壳豆	0.795 9	0.885 1	0.838 1	0.961 3



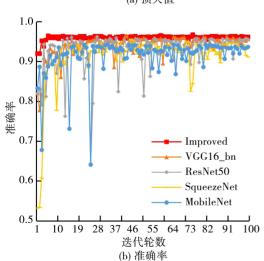


图 4 不同模型的损失值和准确率对比图

Fig.4 Comparison of loss and accuracy of different models

从准确率曲线图中可以看出,所有网络均能 快速达到收敛,相较于其他网络,改进的网络模型 最先达到收敛,收敛后准确率波动范围小且高于 其他网络。损失函数图中,改进模型损失函数值 为所有模型中最低。分析原因,改进模型采用了 自适应调整学习率方法,网络在前期以较大的学 习率学习,使准确率快速提升,当准确率不再提升 时,转而使用较小学习率,从而收敛后的准确率也 波动较小;迁移学习对网络进行预训练,加快了网 络的收敛时间。

各模型的准确率、参数量、内存占用量和浮点运算量如表 6 所示。改进模型的准确率为96.13%,参数量为 0.15×10⁶,内存占用量为14.14 MB,浮点运算量为 39.94MFlops (million floating-point operations per second,每秒百万个浮点操作),相比于 VGG16_bn、ResNet50、SqueezeNet和MobileNet,准确率分别高出 0.65%、1.39%、2.39%、2.96%。参数量、内存占用量和浮点运算量占其他模型最优结果的 20%、41%和7%,在4种指标中均超过其他模型和未改进的MobileNet原模型。以上结果表明,本文提出的改进算法在检测性能和网络规模上都有一定优势,更适用于运算性能受限设备上的部署。

表 6 各模型的准确率、参数量、内存占用量和浮点运算量 Tab.6 Accuracy, parameters, memory occupation and floating-point operation of each model

		_		
模型	准确率/ %	参数量	内存/ 氵 MB	デ点运算量/ MFlops
Improved	96.13	0.15×10 ⁶	14.14	39.94
VGG16_bn	95.48	17.23×10 ⁶	161.07	1.5×10 ⁴
ResNet50	94.74	23.50×10 ⁶	109.68	4.1×10^3
SqueezeNet	93.74	0.74×10^6	34.32	743.45
MobileNet	93.17	3.21×10^6	57.71	582.90

4 结论

1)在数据增强和迁移学习条件下,通过选用 Mish 激活函数,自适应调整学习率方法,减少网 络通道数和删除最后一层卷积模块对网络进行改 进,平均准确率为 96.13%,相较于未改进的 MobileNet,准确率提升了 2.96%,参数量占原模 型 5%。

2)对网络测试集结果进行混淆矩阵可视化, 准确率为96.16%。5种咖啡豆类别中有3种咖 啡豆有较高的 F_1 值,在仅用于检测咖啡豆中的合格品和非合格品时,准确率为 99.9%,证明了网络具有较高的分类性能。

3) 将改进模型和 VGG16_bn、ResNet50、SqueezeNet 以及未改进的 MobileNet 进行对比,改

进模型在准确率最高的情况下,参数量、内存占用量和浮点运算量占其他模型最优结果的 20%、41%和7%,进一步证明了改进模型的性能优势以及更适用于运算性能受限设备上的部署。

参考文献:

- [1] GUNADI I A, ARTHA I K, CHRISTYADITAMA I P, et al. Detection of coffee bean damage in the roasting process based on shape features analysis [J]. Journal of Physics: Conference Series, 2020, 1503(1):012001.
- [2] ARBOLEDA E R, FAJARDO A C, MEDINA R P. An image processing technique for coffee black beans identification [C] // 2018 IEEE International Conference on Innovative Research and Development (ICIRD). Bangkok, Thailand: IEEE, 2018:1-5.
- [3] GARCÍA, CANDELO-BECERRA, HOYOS. Quality and defect inspection of green coffee beans using a computer vision system [J]. Applied Sciences, 2019, 9(19);4195.
- [4] OLIVERI P, MALEGORI C, CASALE M, et al. An innovative multivariate strategy for HSI-NIR images to automatically detect defects in green coffee [J]. Talanta, 2019, 199;270-276.
- [5] CHEN S Y, CHANG C Y, OU C S, et al. Detection of insect damage in green coffee beans using VIS-NIR hyperspectral imaging [J]. Remote Sensing, 2020, 12(15):2348.
- [6] DE OLIVEIRA E M, LEME D S, BARBOSA B H G, et al. A computer vision system for coffee beans classification based on computational intelligence techniques [J]. Journal of Food Engineering, 2016, 171;22-27.
- [7] 刘阳,高国琴. 采用改进的 SqueezeNet 模型识别多类叶片病害[J]. 农业工程学报,2021,37(2):187-195.
- [8] 马本学,李聪,李玉洁,等. 基于残差网络和图像处理的干制哈密大枣外部品质检测[J]. 农业机械学报,2021,52 (11);358-366.
- [9] HOWARD A G,ZHU M L,CHEN B, et al. MobileNets; efficient convolutional neural networks for mobile vision applications [EB/OL]. 2017; arXiv;1704.04861. https://arxiv.org/abs/1704.04861.
- [10] DENG J, DONG W, SOCHER R, et al. ImageNet; a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA; IEEE, 2009; 248-255.
- [11] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C] // Proceedings of the 27th International Conference on International Conference on Machine Learning. New York; ACM, 2010; 807-814.
- [12] MISRA D. Mish; a self regularized non-monotonic activation function [EB/OL]. 2019; arXiv; 1908.08681. https://arxiv.org/abs/1908.08681.

(责任编辑:陈雯)