

融合大核注意力机制的土壤种类识别网络*

刘康^{1,2}, 曾绍华^{1,2}, 胡静³, 刘国一⁴, 李超⁵

(1. 重庆师范大学计算机与信息科学学院, 重庆 401331; 2. 重庆市数字农业服务工程技术研究中心, 重庆 401331;
3. 重庆市农村土地整治中心, 重庆 401121; 4. 西藏自治区农牧科学院农业资源与环境研究所, 拉萨 850032;
5. 重庆市铜梁区农业农村委员会, 重庆 铜梁 402560)

摘要:【目的】为解决因土壤图像纹理复杂、没有结构性特征导致的传统卷积神经网络模型难以提取其中关键性特征、识别准确率低的问题,提出了一种大核注意力(large kernel attention,LKA)机制模块与ResNet模型融合的土壤种类识别网络模型VAR(visual attention ResNet),以解决土壤种类样本不平衡和难分类样本造成的模型泛化能力弱的问题。【方法】以ResNet为主干网络,在主干中引入LKA机制,减少其中的残差块,构建土壤种类识别网络VAR,并改进网络的焦点损失函数(Focal Loss)。【结果】1)与传统模型ResNet18、ResNet34、VGG、GooleNet、VAN等相比,VAR模型在特定模型参数下对紫色土土壤图像数据集中土壤种类的识别精度更高;2)用3种不同大小VAR模型之一的VAR_small与以ResNet18为主干并嵌入传统注意力机制SE、CBAM、ECA 和 SK 的网络进行对比,实验结果显示LKA机制在土壤识别方面更加优秀;3)改进的Focal Loss 可让VAR更能注意到难分类的土壤图像样本。【结论】将LKA机制模块与ResNet模型融合的土壤种类识别网络模型VAR增强了网络提取土壤图像中关键性结构特征能力,同时还减少了网络参数,能更加有效地识别土壤种类。

关键词:土壤种类识别;VAR;大核注意力;焦点损失函数

中图分类号:TP391.41

文献标志码:A

文章编号:1672-6693(2023)03-0129-10

土壤是农作物生产的基础,农作物的种类选择、种植方式、田间管理等均受到土壤种类差异的影响,因而正确识别土壤种类对农业生产有一定的积极意义。土壤种类的传统识别方法为:由土壤学专业技术人员在待测土壤种类的田地里取土样,然后在掰开土样后观察它的自然断口,再根据相关专业知识来识别土壤种类。因此,传统的土壤种类识别方法不能满足一般人员对土壤种类识别的需求。当前,随着人工智能技术的发展,采用机器视觉方法识别土壤种类已成为可能,并可有效降低土壤种类识别成本,因而相关研究具有重要意义。

目前采用机器视觉的方法对土壤图像进行识别的研究较少,且在已有的文献中,研究的标准也各不相同,例如:Shima 等人^[1]利用颜色卡比对土壤颜色然后用KNN算法对土壤种类进行分类;Sruthitha 等人^[2]利用低通滤波、Gabor 滤波器和颜色量化技术提取土壤图像的纹理特征,然后用支持向量机分类。这些研究的方法均属于传统机器学习算法范畴。然而土壤图像纹理复杂,传统机器学习算法需要人工对图像进行预处理并手工提取图像特征,因此该方法存在操作复杂且可迁移性差的问题。随着深度学习技术不断发展,它在图像处理任务中逐渐崭露头角,在农业生产中亦得到了许多应用^[3-6]。深度学习^[7-8]有自动提取图像特征、泛化性强等优点,因此非常适合土壤图像分类任务。Inazumi 等人^[9]曾用两个卷积对不同类型的土壤质地进行特征提取并分类;曾莉^[10]改进了ResNet50模型,采用迁移学习的方法对4种土属进行分类。然而上述研究的研究对象均未细分,模型的泛化能力较弱,研究空间有待进一步拓展。

仔细分析土壤图像时可以发现,由于土壤图像纹理十分复杂,且在对它的种类进行直接判断时不易找到明显的结构性特征,因此传统网络模型很难快速提取到对土壤种类识别有用的信息。而在深度学习研究中,为网络添加注意力机制就能够有效提升识别准确率。基于上述背景,本研究对灰棕紫泥、暗紫泥、红棕紫泥等3个土属下的10个土壤种类进行了识别分析。首先对数据集预处理,采用已有研究的算法分割自然复杂背景,并以切

* 收稿日期:2022-08-25 修回日期:2023-05-06 网络出版时间:2023-06-16T10:09

资助项目:重庆市教育委员会科学技术研究重点项目(No. KJZD-K201900505);重庆市高校创新研究群体(No. CXQT20015);重庆师范大学研究生科研创新项目(No. YKC22016)

第一作者简介:刘康,男,研究方向为深度学习,E-mail:1914900130@qq.com;通信作者:曾绍华,男,教授,博士,E-mail:zsh_cqu@126.com

网络出版地址:https://kns.cnki.net/kcms2/detail/50.1165.N.20230615.1058.002.html

分子图的方式增强数据集,满足深度学习训练要求;并对当前比较先进的识别网络模型,选取最优的作为主干神经网络。然后,引入大核注意力(large kernel attention,LKA)机制解决模型难获取对土壤分类有作用特征的问题。最后,改进焦点损失函数(Focal Loss),从而解决由土壤图像样本不平衡和难分类样本造成的识别精度低、泛化能力弱问题。

1 实验数据及预处理过程

1.1 数据集获取

本研究所用数据集来自重庆市数字农业服务工程中心在重庆市璧山区自然场景下根据重庆市农业技术推广总站的土壤代码^[11]拍摄的紫色土心土图像数据集,包含 3 个土属 10 个土壤种类,共 202 幅图像(表 1)。土壤图像样本的获取方式为:用铁锹锹出土地耕层 20~35 cm 的耕层土壤,然后将土壤自然断口面放在距离摄像头 40 cm 左右处进行拍摄。每幅图像样本的像素规格为 3 648×2 736,部分图像样本如图 1 所示。

表 1 土壤图像样本数据集

Tab. 1 The data set of soil soil image samples

土属	土壤种类	样本量/幅	土属	土壤种类	样本量/幅
灰棕紫泥	石骨子土	17	暗紫泥	夹砂土	14
	砂土	9		油石骨子土	20
	半砂半泥土	20		红石骨子土	20
暗紫泥	大泥土	33	红棕紫泥	红紫砂泥土	25
	二泥土	20		红棕紫泥土	24

1.2 数据集预处理

由于不同种类土壤的地域分布散乱,且相关样本都在交通相对不便的农村田地采集,因此获取的数据量较少,不满足深度学习大数据集的要求。为解决该问题,本研究对土壤图像数据集进行了数据增强预处理。然而,常规数据增强算法如旋转、平移等方法对土壤图像数据进行增强后仍不满足深度学习的要求,而调整亮度、对比度等操作又会干扰网络对土壤种类的识别。因此,本研究采取切分子图的方式来增强数据集。首先,采用已有研究中基于自适应密度峰值聚类的野外紫色土彩色图像分割算法^[12]对自然界获取的土壤图像样本进行背景分割,获取只含土壤心土的白色背景图像;然后在从数据集中取 49 幅图像做验证集的基础上对其他 153 幅图像进行 4 等分,其中 3 份做训练集、1 份做测试集;最后,采用通过类似卷积核滑动窗口的方式,给定滑动步长与窗口大小,从土壤像素位置(0,0)处开始分别对训练集、测试集进行子图切割,选取子图中土壤覆盖率为 100% 的样本,得到土壤样本子图集。本文给定窗口像素大小为 224×224,步长为 50 个像素,共获取训练样本子图 12 629 幅,验证样本子图 3 958 幅。土壤样本切割子图预处理过程如图 2 所示。

2 现有模型存在的问题及相关研究

2.1 ResNet 模型

作为主干网络,网络的结构对提取图像特征的影响非常重要。He 等人^[13]提出的 ResNet 模型,是深度学习的研究的重大突破之一,他们所构建的模型解决了以往深度模型存在的梯度消失、梯度爆炸等问题,在图像识别、图像检测等方面取得重大成果。目前,已有很多研究者研究并改进 ResNet 模型,设计了多种基于 ResNet 模型的改进模型^[14-17]。

土壤图像纹理复杂,需要能够充分提取图像特征的网络结构作为主干网络。本研究在对比了目前几种先进的网络模型基础上选择采用 ResNet 模型的主干网络并加以改进后作为本研究的算法基础。传统 ResNet 模型由多个残差块堆叠构成,残差块的跳跃连接如图 3 所示,设 x 为特征输入,传统网络待拟合目标函数为 $H(x)$,残差块能使网络拟合残差映射 $F(x)=H(x)-x$,使之更易拟合。

2.2 LKA 机制

普通网络很难快速找到对土壤种类分类有用的关键信息,因而在使用深度学习识别目标时为主干网络添加注意力机制,网络能够更容易地对图像关键部分加以识别,从而提高网络的识别准确率。常见传统注意力机制如 SE 注意力机制^[18]、CBAM 注意力机制^[19]等将提取特征图通道或空间上的最大值或平均值作为该通道和空间的注意力权重,继而在原特征图中加权以获取对特征图关键信息的注意。然而对本研究而言上述方法过于片面,因为土壤子图数据集纹理十分复杂,1 个特征值不能完全代表它对应的通道或空间。



图 1 部分土壤图像样本

Fig. 1 The partial soil image samples

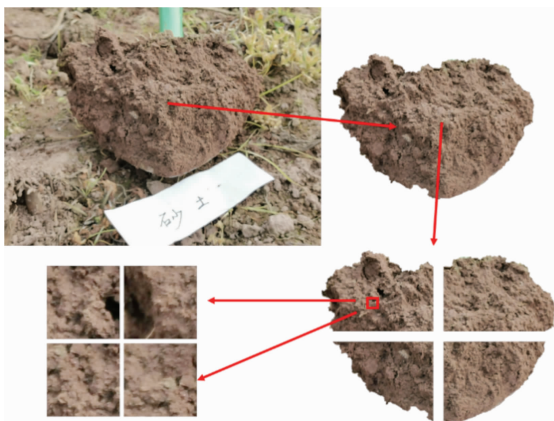


图 2 土壤图像样本预处理过程

Fig. 2 The pretreatment process of soil image sample

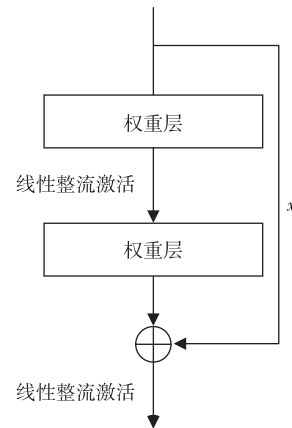


图 3 残差块

Fig. 3 Residual block

Guo 等人^[20]借鉴了自注意力机制^[21-23]的思想,提出了一种新机制即 LKA,它结合了自注意力机制和传统注意力机制的优点。LKA 的思想在于将一个大的卷积核分解成 3 个部分:1 个深度卷积、1 个深度扩张卷积和 1 个点态卷积。这样的分解法既保留了大核卷积具有大感受野的特性并做到了对图像局部区域的注意,也缩小了模型尺寸且使模型更轻量级。与传统注意力机制相比,LKA 关注到了特征图的每个像素点,并对每个像素点进行加权,针对土壤图像数据集,这是一个有效的注意力算法。LKA 的大核卷积分解过程如图 4 所示。

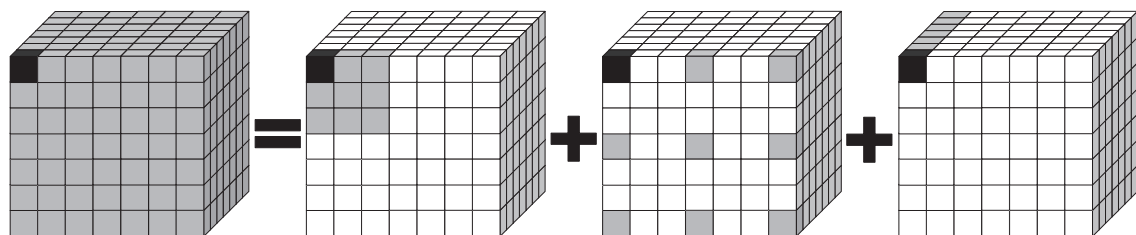


图 4 大核卷积分解过程

Fig. 4 The decomposition process of large convolution kernel

利用分解后的 LKA 机制计算注意力权重,具体过程可表示为下式:

$$k_{\text{attention}} = \text{Conv}_{1 \times 1}(\text{DW_D_Conv}(\text{DW_Conv}(F)))$$

其中: F 是该层的输入特征; DW_Conv 、 DW_D_Conv 和 $\text{Conv}_{1 \times 1}$ 分别表示深度可分离卷积、深度可分离扩展卷积以及点态卷积; $k_{\text{attention}}$ 是计算的注意力参数核,每个参数代表了 F 的每个特征的重要性。获得特征权重后即对特征加权,计算公式为:

$$F_{\text{output}} = k_{\text{attention}} \otimes F,$$

其中: F_{output} 表示该层的输出特征; \otimes 表示张量对应位置的乘积。LKA 机制的整个过程如图 5 所示。基于 LKA,Guo 等人^[20]提出了一种新的主干网络 VAN,在 ImageNet 数据集上取得了超越传统 ResNet 模型的良好效果。

3 融合 LKA 机制模块的 ResNet 模型

在对注意力机制的研究中,Hu 等人^[18]提出的 SE 机制、Woo 等人^[19]提出的 CBAM 机制都对 ResNet 模型做了改进。本研究利用 LKA 机制模块对 ResNet 模型进行改进。与以往传统注意力机制思想不同,本文没有在残差块中使用注意力机制,而是借鉴 VAN 的思想,将 LKA 机制模块作为主干网络块融合进 ResNet 模型中。将输入的土壤特征图由 LKA 机制模块加权,让该模块注意土壤纹理丰富的区域,然后将加权后的土壤特征图送给残差块,由残差块处理土壤特征图的局部纹理细节,将改进后的新网络命名为 VAR (visual attention ResNet)。借鉴 ResNet 模型的思想,将 VAR 设计为 4 个卷积阶段。每个卷积阶段如图 6 所示,其中:下采样层在第一阶段为 7×7 的标准卷积核,其他阶段为 3×3 的标准卷积核;BN 为批量归一化层;GELU 为高斯误差线性激活函数;Conv 表示标准卷积核;D_Conv 表示扩张卷积核; n 表示当前卷积阶段残差块个数。

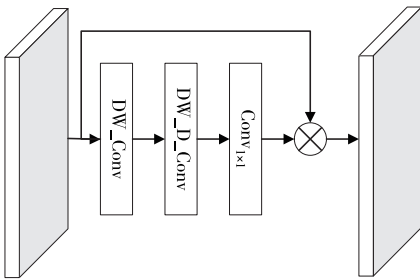


图 5 LKA 机制过程

Fig. 5 The process of LKA

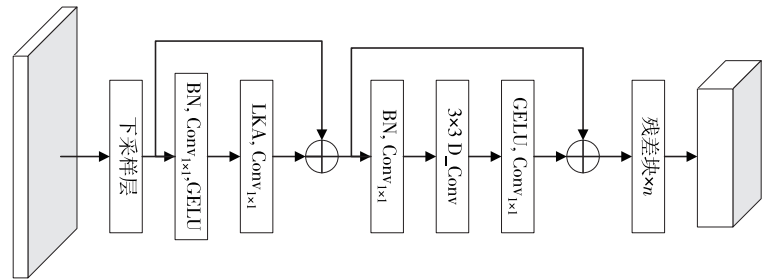


图 6 VAR 的一个卷积阶段

Fig. 6 One stage in VAR

4 加权交叉熵损失函数及改进 Focal Loss

4.1 加权交叉熵损失函数

由于土壤数据集样本之间数据存在不平衡的问题,会降低模型的鲁棒性。本文通过土壤图像样本量的反比为传统交叉熵损失函数加权,让网络多注意样本量少的土壤类别,提高网络的鲁棒性。传统的多分类交叉熵损失函数(J_{CE})计算公式如下:

$$J_{\text{CE}} = -(1/N) \sum_i^N (y_i \log(p_i))$$

其中: N 表示样本总量; y_i 、 p_i 分别表示第 i 个样本的正确分类标签、正确分类的概率。当待分类数据集均衡时,该函数往往能获得较好的效果。当分类样本量不均衡时,该函数训练得到的模型鲁棒性将会降低,将更加注意样本量多的类别。为解决该问题,本研究以样本量最大类别样本的样本量与其他类别样本量之比作为其他类别样本的类平衡因子。改进后的多分类交叉熵损失函数(J_{CE}^*)计算公式如下:

$$J_{\text{CE}}^* = -(1/N) \sum_i^N (y_i (M/O_j) \log(p_i))$$

其中: M 表示最大样本量类别样本的样本量; O_j 表示第 j 类样本的样本量, $j=0,1,\dots,9$ 。

4.2 Focal Loss 及有关改进

在神经网络训练中,对容易识别的样本而言,较好的训练效果往往很快就能达到,且此类样本占据数据集中的大多数。然而土壤图像纹理的复杂度较高,易存在难分类样本,这就降低了模型的训练速度与鲁棒性。Lin 等人^[24]通过引入调制因子对传统交叉熵损失函数进行重构,解决了目标检测问题中大量易分类背景样本损失对模型训练起到主导作用的问题。二分类问题的 Focal Loss(f_{FL})计算公式如下:

$$f_{FL} = -(1 - p_i)^\gamma \log(p_i), p_i = \begin{cases} p, & \text{如果 } y = 1 \\ 1 - p, & \text{否则} \end{cases}$$

其中: p_i 表示当前样本正确分类的概率, y 表示正确分类的标签, p 表示当前样本识别为正样本的概率,即识别为样本标签为 1 的概率, γ 表示聚焦参数,调整调制因子的加权程度。

在目标检测中,因为前后背景样本极端不平衡,接近 1 : 2 000, γ 通常设置为 2。虽然土壤图像数据集类间样本也十分不平衡,但相比目标检测任务则较为平衡,因此 γ 需重新选择。本研究以样本识别正确概率作为聚焦参数,自适应调节调制因子,并融合前文 4.1 节中的类平衡因子,以多分类为目标改写的 Focal Loss(f_{FL}^*)作为本研究训练模型的损失函数,具体公式为:

$$f_{FL}^* = -(1/N) \sum_i^N (y_i (M/O_j) (1 - p_i)^{p_i} \log(p_i))。$$

5 实验与分析

5.1 实验环境

本研究所涉实验均在 Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz、2.19 GHz(2 个处理器),内存 64 GB,显卡 NVIDIA TITAN V, Windows 10 教育版 64 位, PyCharm 2021.1.1 (Professional Edition) + Python 3.8.1 + Pytorch 1.9.0 环境下完成。

5.2 评价方法

要正确评价土壤种类识别模型的性能,不仅需要关注土壤子图测试集的识别准确率,同时也要关注土壤原尺寸图像数据集的识别准确率,一张像素规格为 3 648 × 2 736 的原尺寸土壤图像数据样本被正确分类才是研究的根本目的。因此在本研究中,首先将一张验证集图像从像素位置(0,0)处以滑动窗口切割子图的方式切割若干子图验证集;然后将所有子图样本送入训练好的神经网络模型中预测分类,并记录所有子图样本的预测结果并统计每个分类结果的分类数量;最后选出最大分类数量的结果用以表示该原尺寸土壤图像样本的分类结果。上述方法的具体计算方式如下:

$$N_{pre} = \max(\{N_j | j = 0, 1, \dots, 9\}),$$

其中: N_j 表示原尺寸土壤图像样本的子图样本中预测为第 j 类土壤种类的子图个数,则 N_{pre} 为预测结果的子图个数,它对应的编号即为原尺寸土壤图像样本预测结果。

5.3 评价指标

为了评价本研究提出的土壤种类识别模型 VAR 的分类性能,选用子图识别准确率、原尺寸验证集识别准确率以及模型参数个数作为评价指标。在分类任务中,识别准确率($R_{accuracy}$)是最重要的评价指标,它以正确识别的样本量占测试总样本量的概率值来表示,即:

$$R_{accuracy} = (S_r / S_{all}) \times 100\%,$$

其中: S_r 表示土壤种类被正确识别样本量, S_{all} 表示土壤数据集的总样本量。

5.4 实验流程

实验的整个流程如图 7 所示:首先基于自适应密度峰值聚类的野外紫色土彩色图像分割算法分割自然背景,然后划分训练集与验证集并分别切割子图获取子图训练集与子图验证集,再用子图训练集训练卷积神经网络模型,最后用验证集评价模型优劣。

5.5 VAR 性能验证

为了证明所选择和改进的模型是有效的,本研究将 VAR 与本研究所借鉴的两种主干网络 ResNet、VAN 以及 VGG19^[25]、GooleNet^[26] 进行对比。对 ResNet、VAN 及其他先进网络,参照有关文献原文的模型大小进行设计。对照借鉴的算法,对 VAR 则设计 3 种大小的模型,即 VAR_tiny、VAR_small 和 VAR_base。VAR 的 3 种

模型卷积层参数设计如表 2 所示,其中使用者可根据自己的想法及应用场景重新设定超参数 C 和 n 以满足特定的需求。

为了正确识别土壤种类,在设计好以上卷积层后,为网络添加全连接层与 Softmax 层用于计算预测分类,在实验中,对所有网络模型的训练超参数进行均等设置,保证模型间的对比公平有效。选用 Adam 优化器,初始学习率设为 0.000 1;损失函数采用改进后的 Focal Loss;为了使模型更容易收敛,采用余弦退火算法调整学习率;网络模型训练迭代 80 次,数据批大小设置为 64。采用 Top-1 准确率作为评判标准。从表 3 可见:ResNet18、ResNet34 以较少的模型获得最好的精度,因而优于其他传统网络;而 VAN 随模型参数的增长精度反而下降。因此,选择 ResNet 作为本研究算法主干网络是最好的。此外从该表中还可以看到,不论是测试集子图识别还是验证集原尺寸图片识别,本研究提出的算法明显优于其他算法,证明 ResNet 结合 LKA 机制能够有效解决土壤图像纹理复杂的问题,后者能够让网络找到土壤图像中的关键信息,提高了识别精度。

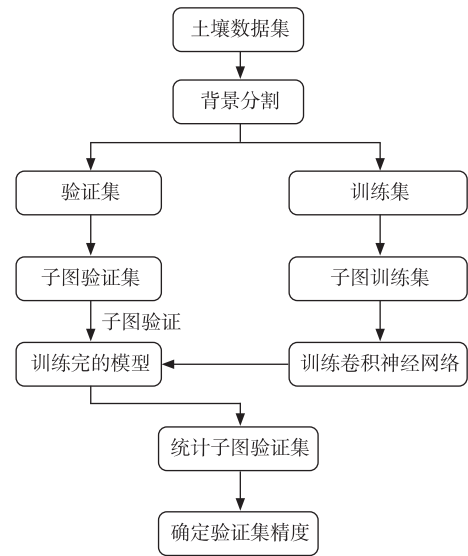


图 7 土壤种类识别流程

Fig. 7 The flow of soil species classification

表 2 VAR 模型参数

Tab. 2 The parameters of VAR

阶段 编号	当前卷积块操作 输出的特征图大小	VAR_tiny	VAR_small	VAR_base	阶段 编号	当前卷积块操作 输出的特征图大小	VAR_tiny	VAR_small	VAR_base
1	$(H/4) \times (W/4) \times C$	$C=32,$ $n=1$	$C=64,$ $n=1$	$C=64,$ $n=2$	3	$(H/16) \times (W/16) \times C$	$C=160,$ $n=1$	$C=256,$ $n=1$	$C=256,$ $n=2$
2	$(H/8) \times (W/8) \times C$	$C=64,$ $n=1$	$C=128,$ $n=1$	$C=128,$ $n=2$	4	$(H/32) \times (W/32) \times C$	$C=256,$ $n=1$	$C=512,$ $n=1$	$C=512,$ $n=2$

注: H 、 W 表示输入图像的高和宽(以像素个数计), C 表示当前卷积块操作输出的维度。

表 3 不同模型精度及参数大小

Tab. 3 The accuracy and parameter size of different models

模型	测试集 子图精度	验证集原尺寸 土壤图像精度	模型参数	模型	测试集 子图精度	验证集原尺寸 土壤图像精度	模型参数
Resnet18	91.36%	40/49	11.18E+06	VAN_small	92.42%	42/49	13.35E+06
Resnet34	93.20%	41/49	21.29E+06	VAN_base	90.37%	42/49	26.07E+06
VGG19	91.16%	39/49	139.60E+06	VAR_tiny	93.65%	44/49	2.55E+06
GooleNet	90.21%	40/49	10.31E+06	VAR_small	94.82%	44/49	8.96E+06
VAN_tiny	92.31%	43/49	3.85E+06	VAR_base	95.19%	45/49	15.23E+06

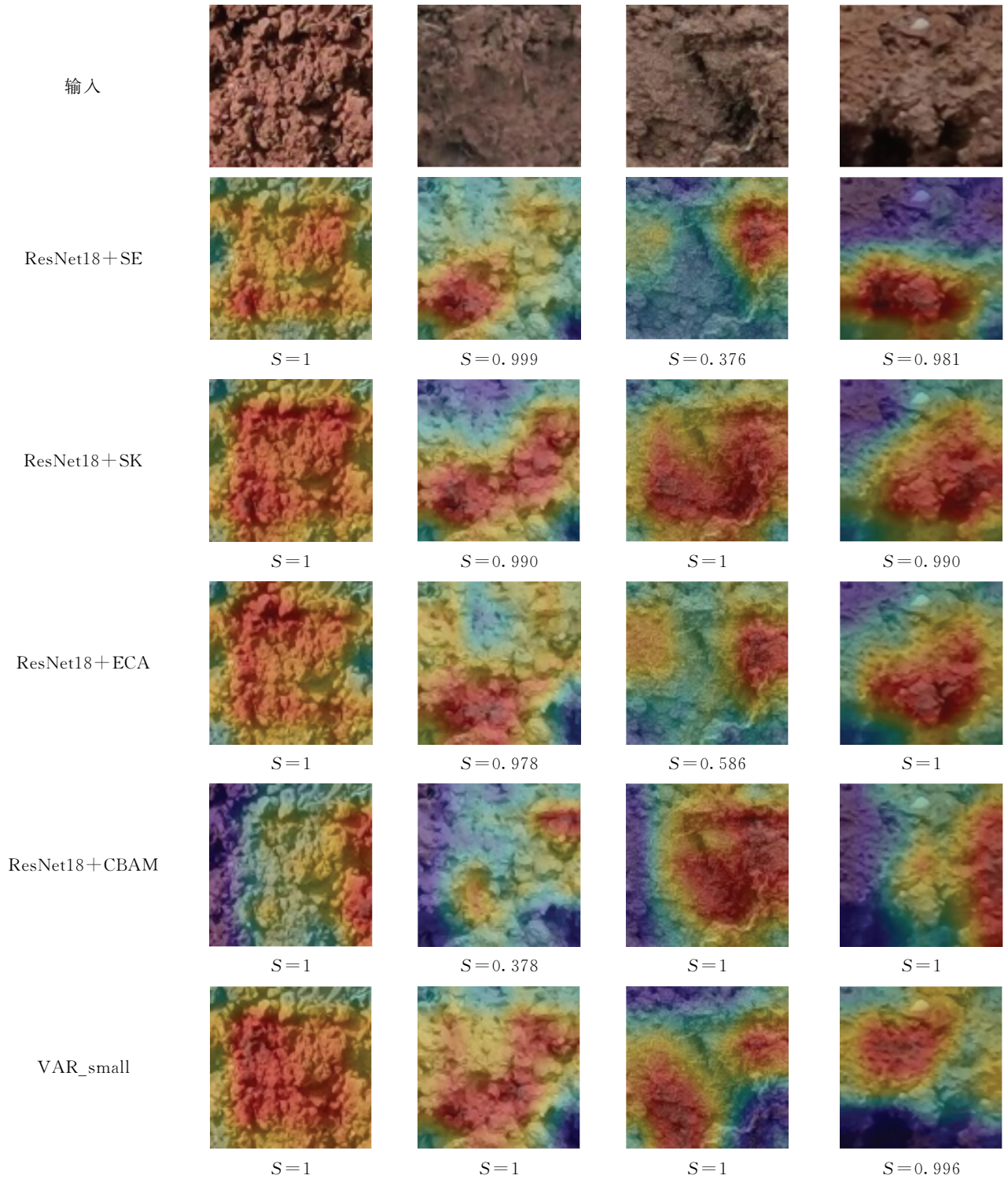
5.6 不同注意力机制对比

使用 ResNet18 融合如今比较先进的注意力机制如 SE、SK^[27]、ECA^[28] 和 CBAM 后,再与 VAR_small 模型对比。从表 4 可见,好的注意力机制能够有效提升 ResNet18 的性能。在土壤子图数据集训练中,SE 模块对 ResNet18 的测试集子图精度提升效果最好,但仍然无法与本文提出的 VAR 模型相比。这证明了 VAR 中的 LKA 机制在土壤图像数据集中识别方面的优越性。

本研究还利用 Grad-CAM^[29] 对注意力机制模型分类结果进行热力图分析并给出了对应的 Softmax 得分,结果发现:对于同一幅识别正确的土壤子图,VAR 中的 LKA 机制能够达到比其他注意力机制更好的效果,关注的位置更集中,Softmax 得分也更高,这也证明 LKA 机制在土壤识别方面优于其他注意力机制(图 8)。

表 4 不同注意力机制精度
Tab. 4 The accuracy of different attention models

模型	测试集子图精度	模型	测试集子图精度	模型	测试集子图精度
Resnet18+SE	93.91%	Resnet18+ECA	93.58%	VAR_small	94.82%
Resnet18+SK	90.32%	Resnet18+CBAM	91.96%		



注:S 为 Softmax 得分。

图 8 不同注意力机制热激活图

Fig. 8 The thermal activation diagram of different attention models

5.7 基于改进 Focal Loss 的模型验证

不同的土壤类别图像样本量的极度不平衡影响到网络的拟合及模型的鲁棒性,本研究引入了多分类的 Focal Loss 并加以改进,让模型投入更多的关注给图像样本较少的土壤种类以及难分类的土壤图像样本。实验以 VAR_small 作主干网络,不同损失函数精度对比结果显示:对损失函数加权能够有效提高识别精度,证明加权平衡土壤种类类别有效;以正确标签的识别概率作为聚焦参数的改进 Focal Loss 精度有明显的提升(表 6)。因此,改进的 Focal Loss 能让 VAR 更能注意到难分类的土壤图像样本。

表 6 不同损失函数下的精度

Tab. 6 The accuracy of different loss function

损失函数	精度	损失函数	精度
交叉熵损失函数	92.61%	Focal Loss ($\gamma=2$)	93.58%
加权交叉熵损失函数	93.95%	Focal Loss ($\gamma=5$)	93.63%
Focal Loss ($\gamma=0.5$)	94.15%	Focal Loss ($\gamma=p$)	94.82%
Focal Loss ($\gamma=1$)	93.73%		

6 结束语

本研究构建了一种基于 ResNet 主干网络 VAR 模型,它在 ResNet 主干中融合了 LKA 机制以提升网络特征提取能力,并减少主干中的残差块,减少模型参数,适合机器视觉识别土壤种类。同时,本研究还改进了 Focal Loss,以最大样本土壤种类图像的样本量和其他土壤种类图像样本量的比值作为加权因子解决了土壤图像样本不平衡的问题,并以正确标签的识别概率做聚焦参数,用自适应调节调制因子解决难分类样本问题。验证实验结果表明:VAR 模型能有效识别土壤种类,与 ResNet18、ResNet34、VGG、GooleNet、VAN_tiny、VAN_small、VAN_base 等传统网络模型以及同 SE、CBAM、ECA、SK 等传统注意力机制结合的 ResNet18 相比,VAR 以更少的模型参数得到了更高精度的模型。需要注意的是,本研究构建的 VAR 模型目前仅能对紫色土图像进行识别,对其他非紫色土土壤种类的识别尚需进一步研究。

参考文献:

- [1] MANIYATH S R, HEBBAR R, AKSHATHA K N, et al. Soil color detection using KNN classifier[C]//2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C). Bangalore: IEEE, 2018: 52-55.
- [2] SRUNITHA K, PADMAVATHI S. Performance of SVM classifier for image based soil classification[C]//2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs). Paralakhemundi: IEEE, 2016: 411-415.
- [3] 王美华, 吴振鑫, 周祖光. 基于注意力改进 CBAM 的农作物病虫害细粒度识别研究[J]. 农业机械学报, 2021, 52(4): 239-247.
WANG M H, WU Z X, ZHOU Z G. Fine-grained identification research of crop pests and diseases. based on improved CBAM via attention[J]. Transactions of the Chinese Society of Agricultural Machinery, 2021, 52(4): 239-247.
- [4] 甘雨, 郭庆文, 王春桃等. 基于改进 EfficientNet 模型的作物害虫识别[J]. 农业工程学报, 2022, 38(1): 203-211.
GAN Y, GUO Q W, WANG C T, et al. Recognizing crop pests using an improved EfficientNet model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(1): 203-211.
- [5] 徐建鹏, 王杰, 徐祥等. 基于 RAdam 卷积神经网络的水稻生育期图像识别[J]. 农业工程学报, 2021, 37(8): 143-150.
XU J P, WANG J, XU X, et al. Image recognition for different developmental stages of rice by RAdam deep convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(8): 143-150.
- [6] 王金鹏, 高凯, 姜洪喆等. 基于改进的轻量化卷积神经网络火龙果检测方法(英文)[J]. 农业工程学报, 2020, 36(20): 218-225.
WANG J P, GAO K, JIANG H Z, et al. Method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(20): 218-225.
- [7] 张珂, 冯晓晗, 郭玉荣等. 图像分类的深度卷积神经网络模型综述[J]. 中国图象图形学报, 2021, 26(10): 2305-2325.
ZHANG K, FENG X H, GUO Y R, et al. Overview of deep convolutional neural networks for image classification[J]. Journal of Image and Graphics, 2021, 26(10): 2305-2325.

- [8] 郑远攀,李广阳,李晔.深度学习在图像识别中的应用研究综述[J].计算机工程与应用,2019,55(12):20-36.
ZHENG Y P, LI G Y, LI Y. Survey of application of deep learning in image recognition[J]. Computer Engineering and Applications, 2019, 55(12): 20-36.
- [9] INAZUMI S, INTUI S, JOTISANKASA A, et al. Artificial intelligence system for supporting soil classification[J]. Results in Engineering, 2020, 8: 100188.
- [10] 曾莉.基于ResNet50的紫色土图像分类[J].现代计算机,2019(31):28-32.
ZENG L. Purple soil image classification based on ResNet50[J]. Modern Computer, 2019(31): 28-32.
- [11] 重庆市农业技术推广总站,西南大学,重庆土壤学会.重庆市地方标准 DB50/T 796—2017:重庆土壤分类与代码[S].重庆:重庆市质量技术监督局,2017:9.
Chongqing Agricultural Technology Extension Station, Southwest University, Chongqing Soil Society. Chongqing local standard DB50/T 796—2017: Classification and code for Chongqing soil[S]. Chongqing: Chongqing Municipal Bureau of Quality and Technical Supervision, 2017: 9.
- [12] 曾绍华,唐文密,詹林庆,等.基于自适应密度峰值聚类的野外紫色土彩色图像分割[J].农业工程学报,2019,35(19):200-208.
ZENG S H, TANG W M, ZHAN L Q, et al. Color image segmentation of field purple soil based on adaptive density peaks clustering[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(19): 200-208.
- [13] 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), Las Vegas: IEEE, 2016: 770-778.
- [14] IANDOLA F, MOSKEWICZ M, KARAYEV S, et al. Densenet: implementing efficient convnet descriptor pyramids[DB/OL]. (2014-04-07)[2022-08-25]. <https://arxiv.org/pdf/1404.1869.pdf>.
- [15] BELLO I, FEDUS W, DU X Z, et al. Revisiting resnets: improved training and scaling strategies[DB/OL]. (2021-03-13)[2022-08-25]. <https://arxiv.org/pdf/2103.07579.pdf>.
- [16] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 5987-5995.
- [17] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[DB/OL](2021-03-13)[2022-08-25]. <https://arxiv.org/pdf/1602.07261.pdf>.
- [18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City: IEEE, 2018: 7132-7141.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//ECCV 2018: European Conference on Computer Vision. Munich: ECCV, 2018: 3-19.
- [20] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[DB/OL](2022-02-20)[2022-08-25]. <https://arxiv.org/pdf/2202.09741.pdf>.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL](2017-12-06)[2022-08-25]. <https://arxiv.org/pdf/1706.03762v5.pdf>.
- [22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[DB/OL](2017-12-06)[2022-08-25]. <https://arxiv.org/pdf/2010.11929v1.pdf>.
- [23] 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). Montreal: IEEE, 2021: 9992-10002.
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999-3007.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL](2015-04-10)[2022-08-25]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [26] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston: IEEE, 2015: 1-9.
- [27] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 510-519.
- [28] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 11531-11539.
- [29] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 618-626.

Soil Species Classification Network Incorporating Large Kernel Attention

LIU Kang^{1,2}, ZENG Shaohua^{1,2}, HU Jing³, LIU Guoyi⁴, LI Chao⁵

- (1. College of Computer and Information Science, Chongqing Normal University, Chongqing 401331;
2. Chongqing Research Center on Engineer Technology of Digital Agricultural & Services, Chongqing 401331;
3. Chongqing Rural Land Consolidation Center, Chongqing 401121; 4. Institute of Agricultural Resources and Environment, Tibet Academy of Agriculture and Animal Husbandry Sciences, Lhasa 850032;
5. Agriculture and Rural Affairs Committee of Chongqing Tongliang District, Tongliang Chongqing 402560, China)

Abstract: [Purposes] Because there are complex textures and no structural characteristic in soil image, it is difficult to extract key features from a soil image with traditional convolutional neural network. A new classification network named visual attention ResNet (VAR), which combined ResNet with large kernel attention (LKA) module, is proposed to solve the low generalization ability that is caused by imbalance datasets and hardly classified samples and it involves VAR_tiny, VAR_small and VAR_base. [Methods] To construct VAR network for soil classification, the LKA is embedded into ResNet as a part of the backbone and the residual modules is partially removed from ResNet, and an improved focal loss function (Focal Loss) is introduced to VAR. [Results] 1) Compared with the traditional models ResNet18, ResNet34, VGG, GooleNet, VAN, etc., VAR model has a higher identification accuracy for images data set of purple soil species under specific model parameters. 2) VAR_small, one of the three VAR models with different sizes, was compared with ResNet18 networks embedded with traditional attention mechanisms such as SE, CBAM, ECA and SK. The experimental results showed that LKA mechanism was better in soil recognition. 3) The improved Focal Loss enables VAR to pay more attention to soil image samples that are difficult to classify. [Conclusions] The soil species identification network VAR, which integrates the LKA module and ResNet, enhances the ability of the network to extract the key structural features from soil images, makes it identify soil species more effectively. And it also reduces the network parameters.

Keywords: soil species classification; VAR; large kernel attention; focal loss function

(责任编辑 方 兴)