DOI: 10.16210/j.cnki.1007-7561.2025.02.013
何伟婵, 杨志景, 秦景辉. 基于改进 MobileNetV3-Large 食物图像分类算法研究[J]. 粮油食品科技, 2025, 33(2): 90-96.
HE W C, YANG Z J, QIN J H. Research on food image classification algorithm based on improved MobileNetV3-Large[J]. Science and Technology of Cereals, Oils and Foods, 2025, 33(2): 90-96.

基于改进 MobileNetV3-Large 食物图像分类算法研究

何伟婵，杨志景，秦景辉✉

（广东工业大学 信息工程学院，广东 广州 510000）

摘　要：食物图像识别在食物安全监控、营养分析以及饮食推荐系统中发挥重要作用。然而，食物图像的多样性、复杂性以及光照等外部因素给识别任务带来了诸多难度和挑战。为了解决这些问题，提出了一种基于改进 MobileNetV3-Large 食物图像分类算法。在 MobileNetV3-Large 预训练模型基础上，引入 PReLu 激活函数和 NAM 注意力机制，通过捕捉图像中的非局部依赖关系来增强模型对关键特征的关注度；引入了多任务损失函数，通过同时优化多个相关任务来进一步提升分类性能；采用了 TrivialAugment 数据增强技术，通过扩展训练数据集的规模和多样性来增强模型的泛化能力。实验结果表明，通过这些改进，模型在 Food-101 数据集上的准确率从 66.9%提升至 84.2%，证明了所提方法的有效性。

关键词：MobileNetV3-Large；NAM 注意力机制；PReLu 激活函数；TrivialAugment 数据增强

中图分类号：TS201.1　　**文献标识码：**A　　**文章编号：**1007-7561(2025)02-0090-07

网络首发时间：2025-02-27 15:50:33

网络首发地址：https://link.cnki.net/urlid/11.3863.TS.20250227.1432.002

Research on Food Image Classification Algorithm based on Improved MobileNetV3-Large

HE Wei-chan, YANG Zhi-jing, QIN Jing-hui✉

(School of Information Engineering, Guangdong University of Technology, Guangzhou, Guangdong 510000, China)

Abstract: Food image recognition plays a crucial role in food safety monitoring, nutritional analysis, and dietary recommendation systems. However, the diversity, complexity, and external factors such as lighting conditions pose numerous difficulties and challenges to the recognition task. In order to address these issues,

收稿日期：2024-08-08；**修回日期：**2024-09-03；**录用日期：**2024-09-04

基金项目：国家自然科学基金-青年科学基金“知识增强的表示学习模型及其在应用题求解中的应用”（62206314）；广东省基础与应用基础研究基金“初等数学应用题求解关键技术研究”（2022A1515011835）

Supported by: The National Natural Science Foundation of China-Youth Science Fund “Knowledge-Augmented Representation Learning Models and Their Applications in Solving Applied Mathematical Problems” (No. 62206314); Guangdong Basic and Applied Basic Research Fund “Research on Key Technologies for Solving Elementary Mathematics Application Problems” (No. 2022A1515011835)

第一作者：何伟婵，女，1992 年出生，在读硕士生，研究方向为计算机视觉与人工智能，E-mail: 1015996335@qq.com

通信作者：秦景辉，男，1989 年出生，博士，讲师，研究方向为自然语言处理和计算机视觉，E-mail: qinjinghui@gdut.edu.cn

this paper proposed a food image classification algorithm based on improved MobileNetV3-Large. Firstly, building upon the pre-trained MobileNetV3-Large model, the PReLu activation function and NAM attention mechanism were introduced to enhance the model's focus on key features by capturing non-local dependencies in images. Subsequently, a multi-task loss function was incorporated to further improve the classification performance by simultaneously optimizing multiple related tasks. Finally, the TrivialAugment data augmentation technique was employed to enhance the model's generalization ability by expanding the scale and diversity of the training dataset. Experimental results demonstrated that through these improvements, the model's accuracy on the Food-101 dataset increased from 66.9% to 84.2%, demonstrating the effectiveness of the proposed approach.

Key words: MobileNetV3-Large; NAM attention mechanism; PReLu activation function; TrivialAugment data augmentation

随着信息技术和人工智能的快速发展，食物图像分类技术已成为计算机视觉领域的研究热点之一。食物图像分类不仅有助于提升食品安全监控的效率，还能为卡路里和营养含量的测量[1]等应用提供关键支持。然而，由于食物图像的多样性、复杂性以及拍摄环境的不可控性，使得食物图像分类任务面临着诸多挑战[2]。

为了克服这些挑战，研究者们不断探索新的方法和技术[3-4]。迁移学习作为一种有效的机器学习范式，通过利用已有领域的知识来解决新领域的问题，为食物图像分类提供了新的思路。本文选择 MobileNetV3[5]作为基础模型进行改进。首先，因为其具有轻量级和高效的特点，适合在资源受限的环境下进行部署。其次，其在计算机视觉领域已被广泛验证，在诸多图像分类任务中取得优异性能。相比其他模型，MobileNetV3 在保持高分类性能的同时，具有更小的模型体积和更快的推理速度，这使得它在食物图像分类任务中具有明显优势。因此，本文提出在 MobileNetV3 预训练模型的基础上，结合迁移学习，将其应用于食物图像分类任务。这一方法不仅能够加快模型训练速度，还能在一定程度上提高分类性能。

然而，仅仅依靠迁移学习和 MobileNetV3 并不足以解决食物图像分类中的所有问题。在实际应用中，我们还需要将同一类别下的不同细小特征进行区分和识别，从而更准确地识别不同种类的食物。

为此，本文在 MobileNetV3-Large 预训练模型基础上，引入了 NAM（Normalization-based attention module）注意力机制[6]和 PReLu 激活函数[7]，能够帮助模型自动学习并关注图像中的关键区域，从而提取出对分类任务更为重要的特征。通过多任务学习同时优化多个相关任务，使模型能够学习到更丰富的特征表示，提高食物图像分类的准确性和鲁棒性。此外，为了进一步提高模型的泛化能力和鲁棒性，本文还采用了 TrivialAugment 数据增强技术[8]。该技术通过对原始图像进行一系列随机变换，如旋转、缩放、裁剪等，生成更多的训练样本，从而增强模型的抗过拟合能力和对未知数据的适应能力。

接下来，本文将详细介绍该方法的实现过程，并通过实验验证其有效性和优越性。

1 MobileNetV3 模型

MobileNetV3 的整体架构基本沿用了 MobileNetV2[9]的设计，采用了轻量级的深度可分离卷积和残差块等结构，依然是由多个模块组成，但是每个模块得到了优化和升级，包括引入 SE（Squeeze-and-Excitation）模块[10]和 h-swish 激活函数等以提高模型精度。MobileNetV3 进一步被划分为 MobileNetV3-small 和 MobileNetV3-large 两个版本，分别针对轻量级和更大规模的部署需求，以满足不同场景下的计算资源和精度要求。图 1 为 MobileNetV3 块结构。

2 改进 MobileNetV3 模型

在 MobileNetV3-Large 预训练模型基础上，引

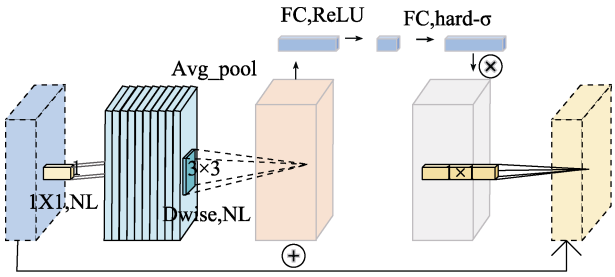

图 1 MobileNetV3 块结构

Fig.1 MobileNetV3 block structure

入 PReLu 激活函数和 NAM 注意力机制，将这一改进称为 NAM-MobileNetV3。NAM-MobileNetV3 旨在解决 MobileNetV3-Large 模型中存在的问题，并进一步提升模型性能。表 1 为 NAM-MobileNetV3 模型的基本结构。

表 1 NAM-MobileNetV3 模型的基本结构

Table 1 Basic Structure of the NAM-MobileNetV3 Model

输入	类型	卷积核大小/步幅	非线性激活函数	注意力机制
224×224×3	Conv2d	2	Hwish	–
112×112×16	Bneck	3×3/1	PReLU	–
112×112×16	Bneck	3×3/2	PReLU	–
56×56×24	Bneck	3×3/1	PReLU	–
56×56×24	Bneck	5×5/2	PReLU	SE
28×28×40	Bneck	5×5/1	PReLU	SE
28×28×40	Bneck	5×5/1	PReLU	SE
28×28×40	Bneck	3×3/2	Hwish	–
14×14×80	Bneck	3×3/1	Hwish	–
14×14×80	Bneck	3×3/1	Hwish	–
14×14×80	Bneck	3×3/1	Hwish	–
14×14×80	Bneck	3×3/1	Hwish	NAM
14×14×112	Bneck	3×3/1	Hwish	NAM
14×14×112	Bneck	5×5/2	Hwish	NAM
7×7×160	Bneck	5×5/1	Hwish	NAM
7×7×160	Bneck	5×5/1	Hwish	NAM
7×7×160	Conv2d	1×1/1	Hwish	–
7×7×960	Avg_pool	7×7/1	–	–
1×1×960	Conv2d	1×1/1	Hwish	–
1×1×1 280	Conv2d	1×1/1	–	–

2.1 PReLu 激活函数

在 NAM-MobileNetV3 模型结构中，采用 PReLU 激活函数替代 ReLU 激活函数。尽管 ReLU 激活函数在神经网络中广泛应用，但它存在一些问题，例如可能导致神经元输出的稀疏性和对负数输入的完全抑制。为了解决这些问题，NAM-MobileNetV3 选择了 PReLU 激活函数。PReLu 激活函数具有一定的负值响应，可以缓解神经元“死亡”问题，并且具有更好的表示能力，有助于提升模型性能。PReLU 其数学公式和导数公式如式（1）和式（2）所示，图像如图 2 所示。

$$PReLu(x)=\left\{\begin{array}{l}x,x>=0\\ ax,x<0(a是可训练参数)\end{array}\right\} \quad 式（1）$$

$$PReLu'(x)=\left\{\begin{array}{l}1,x>=0\\ a,x<0(a是可训练参数)\end{array}\right\} \quad 式（2）$$

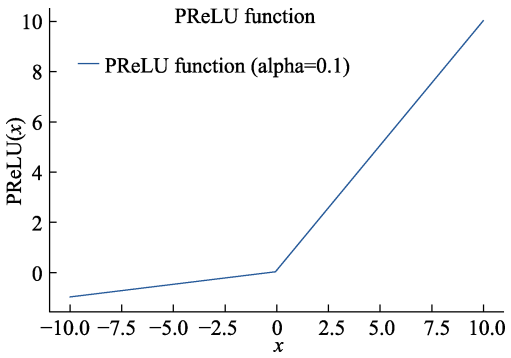

图 2 PReLu 图像

Fig.2 PReLu image

2.2 NAM 注意力机制

NAM-MobileNetV3 引入了 NAM 注意力机制，以取代 MobileNetV3-Large 模型中第 11 至第 15 层瓶颈结构的 SE 注意力机制。虽然 SE 注意力机制在一定程度上能够提升模型对特征通道重要性的识别能力，但它存在一些局限性。SE 注意力机制主要关注通道注意力，忽视了空间方面的注意力，这在处理与食物分类空间相关性较强的任务时，可能导致性能不佳。

相比之下，NAM 注意力机制结合了通道注意力和空间注意力的设计（如图 3~4 所示），通过调整模型对输入数据的关注度，使得模型能够更加关注对任务结果有重要影响的信息，同时忽略那些对结果影响较小的信息。这种设计采用了一种集成的方式，并利用权重的贡献因子来改善注意力机制，以提升模型对特征通道重要性的识别能力，同时考虑了空间特征敏感性和局部特征提取的问题。综合通道和空间注意力设计，使得 NAM-MobileNetV3 在保持轻量级和高效性能的同时，进一步提升了模型的准确性和特征识别能力，尤

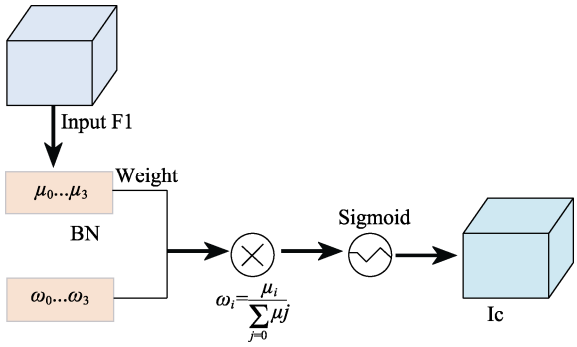

图 3 通道注意力

Fig.3 Channel attention

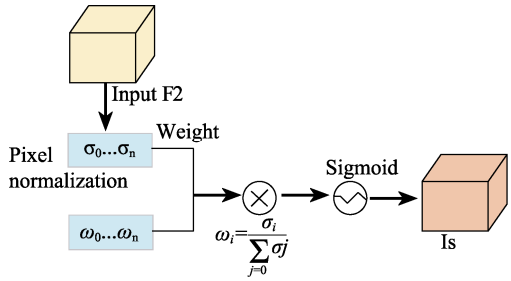

图 4 空间注意力

Fig.4 Spatial attention

其在处理食物图像分类任务中表现出色。

为了进一步验证不同注意力机制对食物图像分类任务的影响，还分别引入 CBAM（Convolutional block attention module）[11]注意力机制和 SA（Shuffle attention）[12]注意力机制，替代 MobileNetV3-Large 模型中第 11 至第 15 层瓶颈结构的 SE 注意力机制，并与 NAM-MobileNetV3 模型进行比较。对比实验结果如表 2 所示。通过表 2 可以看出，加入 CBAM 注意力机制后，模型的验证准确率达到了 82.1%，训练损失率为 0.224；而加入 SA 注意力机制后，验证准确率为 82%，训练损失率为 0.227。相比之下，加入 NAM 注意力机制后，模型的验证准确率提升至 82.7%，训练损失率降至 0.183。此外,原模型 MobileNetV3 在相同任务下的验证准确率仅为 66.9%，损失率为 0.968。

表 2 不同注意力机制下模型对比实验结果

Table 2 Experimental results of model comparison under different attention mechanisms

模型	验证准确率/%	训练损失率
MobileNetV3	66.9	0.968
CBAM-MobileNetV3	82.1	0.224
SA-MobileNetV3	82.0	0.227
NAM-MobileNetV3	82.7	0.183

这些结果表明，在处理食物图像分类任务时，NAM 注意力机制在空间注意方面更有效地弥补了 SE 注意力机制的不足，从而显著提升了模型的整体性能。通过引入 NAM 注意力机制，我们能够更好地捕捉到与食物图像相关的空间特征，同时保持对通道特征的敏感性，这使得该模型在特征识别和分类准确性上表现优异。因此，采用 NAM 注意力机制是提升食物图像分类性能的必要改进。图 5 为 NAM-MobileNetV3 的模型结构。

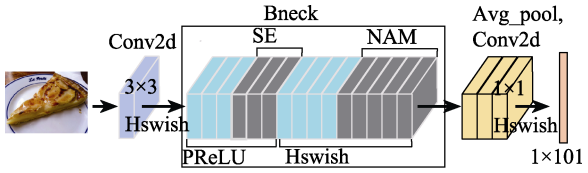

图 5 NAM-MobileNetV3 模型结构

Fig.5 Model structure NAM-MobileNetV3

2.3 多任务损失函数

在食物图片分类任务中，引入多任务学习可以进一步提升模型的性能。多任务学习是一种同时学习多个相关任务的方法，通过共享表示和互相促进来改进每个任务的学习。在本场景中，同时考虑食物的分类任务和边界框的回归任务，即将食物分类与食物定位任务相结合。通过分类损失函数和边界框回归损失函数的加权求和，来定义多任务损失函数，实现输入端到输出端的训练。

在进行分类和回归任务时，所采用的分类损失函数为交叉熵损失（Cross-entropy loss），衡量的是模型预测的概率分布与真实标签之间的“距离”，如公式 3 所示。所采用的回归损失函数为均方误差损失（Mean squared error loss，MSE），衡量的是模型预测值与真实值之间的平方差异的平均值，如公式 4 所示。本文定义该多任务损失函数中，分类任务的损失将被乘以 0.7，而回归任务的损失将被乘以 0.3，如式 5 所示。

$$L(x,y)=-\sum_{i=1}^{c}x_i\log y_i \qquad \text{式（3）}$$

其中，x 和 c 指将输入数据 x 分为 c 个不同的类别，x_i 表示真实标签的第 i 个元素，y_i 表示模型预测 x 属于第 i 个类别的概率。

$$MESloss(m,n)=\frac{1}{z.c}\sum_{i=1}^{z}\sum_{j=1}^{c}(m_{ij}-n_{ij})^2 \quad 式（4）$$

其中，m 代表模型预测值，n 代表真实值，z 为批次中的样本数量，c 代表每个样本的输出维度，回归任务中通常为 1。

$$Loss=0.7L(x,y)+0.3MSEloss(m,n) \quad 式（5）$$

2.4 TA 数据增强

Trivial Augment（TA）是一种用于图像数据增强的方法，旨在通过对原始图像施加一系列简单但有效地变换，以增强数据的多样性，改善模型的泛化能力，并提高模型对于图像数据的鲁棒性。TA 方法包括一系列基本的图像变换操作，如随机裁剪、水平翻转、色彩抖动、缩放和平移等，如图 6 所示，它们在一定程度上改变了图像的外观，同时保持了图像的语义信息不变，从而改善模型在图像分类任务中的性能表现。

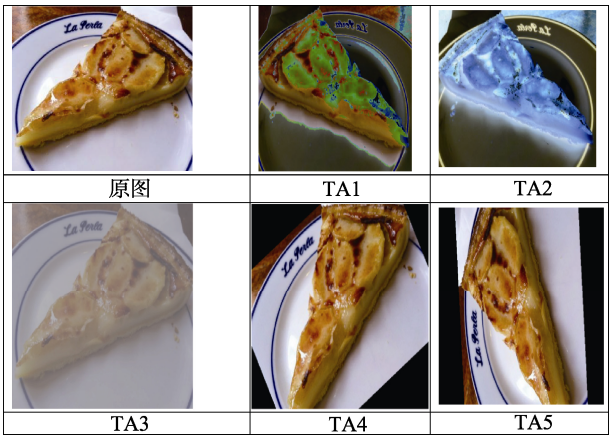

图 6 TA 数据增强示例图

Fig.6 TA data enhancement examples

3 结果与分析

3.1 实验数据集介绍

Food-101 数据集由斯坦福大学的研究人员于 2014 年发布，它包含 101 种不同的食物类别，每一种类别都包括 1 000 张图片，总共有约 101 000 张图像，这些图片包括各种各样的尺寸和质量。Food-101 数据集样例图如 7 所示。

图 7 Food-101 数据集样例图

Fig.7 Samples image of Food-101 dataset

3.2 实验环境

实验在 Windows11 操作系统上进行，采用 12th Gen Intel(R) Core(TM) i5-12450H 2.00 GHz CPU 和 NVIDIA GeForce RTX 3050 Lapton GPU。模型训练环境基于 Pytorch 2.0.1 框架，编程语言为 Python, CUDA 版本 11.8。

3.3 网络模型训练和验证流程

网络模型的设置是：训练 Epoch 数 180，批大小 64，优化器 adam，初始学习率是 0.001，每 30 轮衰减一半。在每次迭代之后，执行一个训练会话，然后更新模型参数的测试阶段。随后在验证集上对达到最佳训练性能的模型进行验证。

3.4 准确率指标分析

准确率（Accuracy）用于描述模型预测正确的样本数量占总样本数量的比例。Accuracy 的计算公式如式 6 所示。其中，真正例（True positives，TP）表示模型预测为正例，且实际也为正例的样本数量。假正例（False positives，FP）表示模型预测为正例，但实际为反例的样本数量。这通常被称为“误报”或“第一类错误”。真反例（True negatives，TN）表示模型预测为反例，且实际也为反例的样本数量。假反例（False negatives，FN）表示模型预测为反例，但实际为正例的样本数量。这通常被称为“漏报”或“第二类错误”。

$$Accuracy=\frac{TP+TN}{TP+FP+FN+TN} \quad 式（6）$$

3.5 模型对比实验

在 Food-101 数据集上的深入模型对比实验揭示了各模型在食物分类任务上的性能差异。如表 3 所示，本文模型 NAM-MobileNetV3 与基准算法

表 3 Food-101 数据集模型对比实验结果

Table 3 Results of model comparison experiment on Food-101 dataset

模型	准确率/%	Top3-acc/%	Top5-acc/%
MobileNetV3-Large	66.90	81.4	86.4
ResNet50	74.89	88.2	92.0
ViT	76.10	89.2	92.6
Swin transformer	77.41	89.7	93.0
CBAM-InceptionV3	82.01	91.8	94.6
NAM-MobileNetV3(ours)	84.20	92.9	95.4

MobileNetV3-Large[5]、ResNet50[13]、ViT[14]、Swin transformer[15]以及结合注意力机制的 CBAM-InceptionV3[16]等模型进行对比实验。

在这些模型中，本文提出的 NAM-MobileNetV3 表现突出，它在准确率上超越了所有对比算法，取得了最高的分类准确率。与结合注意力机制的 CBAM-InceptionV3 模型（其最高准确率为 82.01%）相比，NAM-MobileNetV3 展现出了显著优势，准确率提升了 2.19%。此外，在 Top-3 准确率和 Top-5 准确率方面，NAM-MobileNetV3 也表现优异，分别达到了 92.9%和 95.4%。与基准算法 MobileNetV3-Large 相比，NAM-MobileNetV3 的准确率大幅提升了 17.3%，达到了 84.2%，充分验证了其在食物分类任务上的高效性和准确性。

此外，为了验证模型的泛化能力，NAM-MobileNetV3 还在 CUB-200 数据集上进行了实验，取得了 82%的准确率，Top-3 准确率为 93.3%，Top-5 准确率为 95.9%。这些实验结果表明，NAM-MobileNetV3 在食物分类任务中展现了显著的性能优势，验证了所提方法的有效性，并突出了其在特征识别和分类准确性上较好的表现。

3.6 对比实验

此外，还使用 Grad-CAM 算法[17]，对改进后的 NAM-MobileNetV3 与基础模型 MobileNetV3-Large 网络的最后三层特征图进行可视化对比实验，实验结果如图 8 所示。从图 8 中可以看出，相比于基础模型 MobileNetV3-Large，改进后的 NAM-MobileNetV3 在最后三层特征图的对比实验中展现出了更加突出的特征激活，对特征图的关键区域关注度更高，显示出了对食物图像特征的更好捕捉和理解。这进一步证明了 NAM-

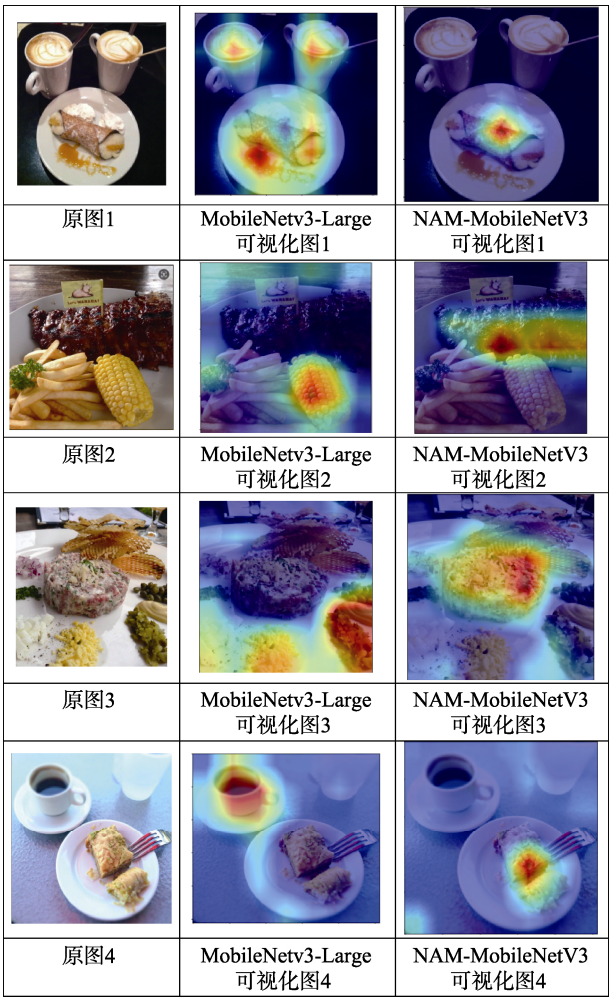

图 8 基于 Grad-CAM 的类激活热力图的可视化对比

Fig.8 Visual comparison of class activated heatmaps based on Grad CAM

MobileNetV3 在食物分类任务中具有更好的特征提取能力和区分能力，从而提高了分类准确率和模型性能。

为了进一步验证每个模块的必要性，进行了消融实验，结果如表 4 所示。原模型 MobileNetV3-Large 的准确率为 66.9%，损失率为 0.968；而改进后的 NAM-MobileNetV3 准确率提升至 82.7%，损失率降至 0.183。引入多任务损失后，NAM-MobileNetV3 的准确率进一步提高至 83.2%，损失率降至 0.134。最后，结合多任务损失和数据增强的 NAM-MobileNetV3 模型达到了最高准确率 84.2%，但损失率略微上升至 0.221。这些结果表明，各个模块的引入显著提升了模型的性能，验证了其必要性。

综上所述，通过可视化对比实验和消融实验的分析结果，可以看出 NAM-MobileNetV3 在食

表 4 Food-101 数据集模型消融实验结果

Table 4 Ablation experiment results of the model on the Food-101 dataset

模型	最佳验证轮次	验证准确/%	训练损失率
MobileNetV3-Large	152	66.9	0.968
NAM-MobileNetV3	154	82.7	0.183
NAM-MobileNetV3-多任务损失	142	83.2	0.134
NAM-MobileNetV3-多任务损失-数据增强	159	84.2	0.221

物图像分类任务中相较于基础模型 MobileNetV3-Large 取得较好的改进。这些结构的引入进一步验证了 NAM-MobileNetV3 模型在食物分类任务上的较好性能。

4 结论

针对食物图像分类任务中存在的复杂性和多样性导致难以准确识别、网络增加深度时可能出现性能退化、模型泛化能力有限以及单一任务学习无法充分利用图像中的多种信息等问题，提出了一个基于改进 MobileNetV3-Large 食物分类算法 NAM-MobileNetV3。NAM-MobileNetV3 算法在 MobileNetV3-Large 的基础上，引入 NAM 注意力机制和 PReLu 激活函数，使得模型能够捕捉图像中的非局部依赖关系，关注到关键特征，从而提高对复杂食物图像的识别准确率。此外，还采用多任务损失函数使得模型能够同时优化多个相关任务，学习到更丰富的特征表示，提高了食物图像分类的准确性和鲁棒性。最后，通过数据增强技术扩展了数据集的规模和多样性，提高了模型的泛化能力。结果表明，NAM-MobileNetV3 模型在 Food-101 数据集中准确率取得了较好的提升，相较于原始的 MobileNetV3-Large 模型，最高准确率提高了 17.3%，达到了 84.2%。综上所述，NAM-MobileNetV3 算法在解决食物图像分类任务中展现出了较好的性能，能保持较高的分类准确率，可以为相关领域的研究提供有价值的参考。

参考文献：

[1] PRAVEENA S, SUDHARSAN R, SWETHA S, et al. Survey on customized diet assisted system based on food recognition[C]// 2023 Second International Conference on Electronics and Renewable Systems (ICEARS). Tuticorin, India: IEEE, 2023: 1577-1584.

[2] ARSLAN B, MEMIŞ S, SÖNMEZ E B, et al. Fine-grained food classification methods on the UEC FOOD-100 database[J]. IEEE Transactions on Artificial Intelligence, 2022, 3(2): 238-243.

[3] ZHANG Y, DENG L, ZHU H, et al. Deep learning in food category recognition[J]. Information Fusion, 2023, 101859.

[4] TARANNUM S, JALAL M S, HUDA M N. HALALCheck: a multi-faceted approach for intelligent halal packaged food recognition and analysis[J]. IEEE Access, 2024, 12: 28462-28474.

[5] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South), 2019: 1314-1324.

[6] LIU Y, SHAO Z, TENG Y, et al. NAM: normalization-based attention module[R]. arXiv: 2111.12419, 2021.

[7] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [J]. IEEE Computer Society, 2015.

[8] MÜLLER S G, HUTTER F. TrivialAugment: tuning-free yet state-of-the-art data augmentation[J]. arXiv preprint, 2021.

[9] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4510-4520.

[10] HU J, SHEN L, SAMUEL A, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

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

[12] ZHANG Q L, YANG Y B. SA-Net: shuffle attention for deep convolutional neural networks[C]// ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada: IEEE, 2021: 2235-2239.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. IEEE, 2016.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is Worth 16x16 Words: Transformers for image recognition at scale [C]//International Conference on Learning Representations. 2021.

[15] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[R]. arXiv preprint, 2021. arXiv: 2103. 14030.

[16] 杜慧江, 崔潇以, 王艺蒙, 等. 基于 CBAM-InceptionV3 迁移学习的食品图像分类[J]. 粮油食品科技, 2024, 32(1): 91-98.
DU H J, CUI X Y, WANG Y M, et al. Food image classification based on CBAM-Inception V3 transfer learning[J]. Science and Technology of Cereals, Oils and Foods, 2024, 32(1): 91-98.

[17] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via Gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2):336-359.

备注：本文的彩色图表可从本刊官网（http://lyspkj.ijournal.cn）、中国知网、万方、维普、超星等数据库下载获取。