DOI:10.19651/j.cnki.emt.2416534

融合多尺度特征及注意力机制的食品图像识别*

李苗苗[1] 华才健[1,2] 谢 涛[3] 薛青霞[1,4]
(1.四川轻化工大学计算机科学与工程学院 宜宾 644000; 2.四川省大数据可视化分析技术工程实验室 宜宾 644000;
3.四川轻化工大学自动化与信息工程学院 宜宾 644000; 4.企业信息化与物联网测控技术四川省高等学校重点实验室 宜宾 644000)

摘 要: 针对食品图像中类间差异小、类内差异大以及结构复杂导致识别难度大等问题,提出了一种融合多尺度特征及注意力机制的食品图像识别方法。首先,采用特征提取能力更强的 ConvNeXt 模型作为主干网络,以更好地捕捉食品图像的细节特征;其次,引入改进的 ASPP 模块,扩展感受野并利用多尺度信息,增强模型对不同尺度特征的捕捉能力;最后,在每个卷积块后加入注意力机制,提高特征表达和上下文信息捕捉能力。实验结果表明,所提方法在 Vireo Food172 扩展数据集和 ETH Food101 数据集上的准确率分别达到 91.56%和 87.22%,相比原模型分别提高了 2.05%和 1.66%,验证了该方法的有效性。

关键词: 食品图像;ConvNeXt;多尺度特征;注意力机制
中图分类号: TP393;TN98 **文献标识码**: A **国家标准学科分类代码**: 520.6040

Integrating multi-scale features and attention mechanisms for food image recognition

Li Miaomiao[1] Hua Caijian[1,2] Xie Tao[3] Xue Qingxia[1,4]
(1. School of Computer Science and Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;
2. Sichuan Big Data Visualization Analysis Technology Engineering Laboratory, Yibin 644000, China;
3. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;
4. Key Laboratory of Measurement and Control Technology of Enterprise
Informatization and Internet of Things of Sichuan Universities, Yibin 644000, China)

Abstract: To address the challenges in food image recognition caused by small inter-class differences, large intra-class variations, and complex structures, this paper proposes a food image recognition method that integrates multi-scale features and an attention mechanism. First, the ConvNeXt model, which has stronger feature extraction capabilities, is used as the backbone network to better capture the detailed features of food images. Next, an improved ASPP module is introduced to expand the receptive field and utilize multi-scale information, enhancing the model's ability to capture features at different scales. Finally, an attention mechanism is added after each convolutional block to improve feature representation and the ability to capture contextual information. Experimental results show that the proposed method achieves accuracies of 91.56% and 87.22% on the extended Vireo Food172 dataset and the ETH Food101 dataset, respectively, which represents an improvement of 2.05% and 1.66% over the original model, thus verifying the effectiveness of the proposed method.

Keywords: food image;ConvNeXt;multi-scale features;attention mechanism

0 引 言

食物在人类生活中扮演着不可或缺的角色,不仅维持生存的基本需求,还在促进健康方面具有重要作用。与普通图像识别相比,食品图像识别面临形态复杂、颜色多样和目标体积较小等挑战;此外,同一种食物在不同条件下可能呈现出多种不同的特征,增加了识别难度。食品图像识别作为细粒度图像识别的一个重要分支,具有显著的研究意

收稿日期:2024-07-26
*基金项目:四川轻化工大学科研创新团队计划(SUSE652A006)、企业信息化与物联网测控技术四川省高校重点实验室(2022WYJ01)项目资助

义和应用前景[1]。

近年来,有许多学者对食品图像识别领域进行了广泛研究。Yang 等[2]提出了一种基于卷积神经网络分层结构的方法,通过多层次特征表示对菜品进行分类,这种方法能够提取较为深层的特征,但容易丢失细节信息;Min 等[3]提出了 SGLANet 网络,通过学习食品图像的局部和全局特征提升食品图像识别准确率,然而在处理大规模数据时可能存在计算效率的问题;姜枫等[4]基于自注意力机制,进一步引入局部注意力以捕捉细粒度特征,从而显著提高了识别准确率,但仍存在部分类别识别结果不理想;王海燕等[5]在 ResNet18 的基础上融合非对称卷积,增强局部信息学习,并引入注意力模块,提升了特征提取效率。然而,这种方法在复杂背景下的识别效果仍有待提升;邓志良等[6]提出了改进的残差网络 ENA-TL 模型,通过融合多尺度特征和注意力机制,提高了识别准确率,但在面对高噪声数据时表现不够理想。综上所述,尽管上述研究在食品图像识别领域取得了显著进展,但普遍存在对细节信息捕捉不够、对多尺度特征的处理不足以及对噪声干扰的抑制效果有限等问题。

为了解决这些问题并进一步提高识别的准确率,本文主要在以下两个方面做出贡献:1)针对食品图像呈现出多尺度和不规则特点的问题,引入改进的空洞空间金字塔池化模块(improved atrous spatial pyramid pooling, IASPP),有效捕获多尺度特征,增强了模型对不同尺度和复杂形态食品图像的识别能力;2)鉴于食品图像数据集中含有较多噪声,引入高效通道注意力机制模块(efficient channel attention, ECA),捕捉通道间的特征依赖关系,抑制噪声干扰,从而更好地保留食品图像的信息完整性。本文旨在通过以上改进,构建一种能够更有效处理复杂食品图像的模型,在多种食品图像数据集上验证其有效性,并在未来的实际应用中提供更优的解决方案。

1 研究方法

本文提出了一种融合多尺度特征及注意力机制的食品图像识别方法,该方法包括主干网络、IASPP 模块和 ECA 模块 3 部分。主干网络负责初步特征提取;IASPP 模块于扩展感受野,进一步捕捉多尺度特征;ECA 模块用于加强重要特征的表达并抑制噪声,其框架如图 1 所示。

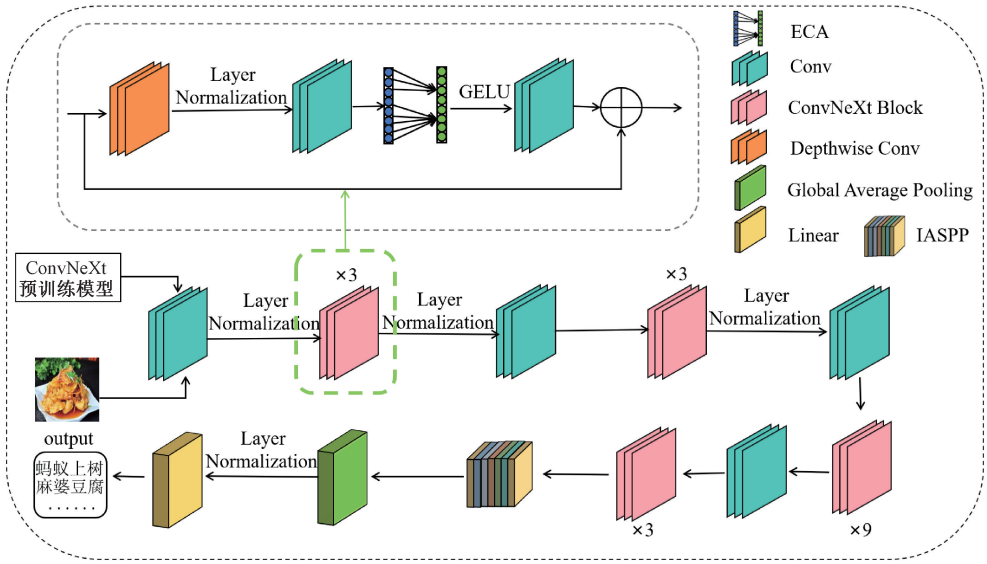


图 1 模型总体结构

Fig. 1 Overall model structure

1.1 ConvNeXt 模块

ConvNeXt 网络是由 Liu 等[7]引入 Transformer 网络思想到经典的 ResNet[8]网络中进行改进的,该网络具有高精度、高效率、可扩展性强和设计非常简单的特点,本文采用了 ConvNeXt-T 网络,其结构如图 2(a)所示。

首先,将尺寸为 224×224×3 的食品图像输入网络,经过卷积核大小为 4 和步长为 4 的卷积层处理,生成尺寸为 56×56×96 的特征图,随后对该特征图进行层归一化(layer normalization, LN)处理;接着,网络通过 3 个阶段,每个阶段包括 ConvNext Block 如图 2(b)所示,下采样如图 2(c)所示。下采样模块用于改变特征图的尺寸,以便更容易提取复杂特征。每个 ConvNeXt Block 模块由 3 个阶段组成,最后一个 ConvNeXt Block 模块的输出先经过全局平均池化层处理,然后通过 LN 层进行归一化,最后输入到线性层,输出最终的分类结果。

1.2 改进空洞空间金字塔池化模块(IASPP)

食品图像通常包含多种不同尺寸、形状和纹理的食物,卷积神经网络在处理这些复杂特征时可能存在不足。本文在 ConvNeXt 网络的最后一个 ConvNeXt 块之后引入改进后的空洞空间金字塔池化(atrous spatial pyramid

(a) ConvNeXt-T网络
(a) ConvNeXt-T network

(b) ConvNext Block
(b) ConvNext Block

(c) 下采样
(c) Down-sampling

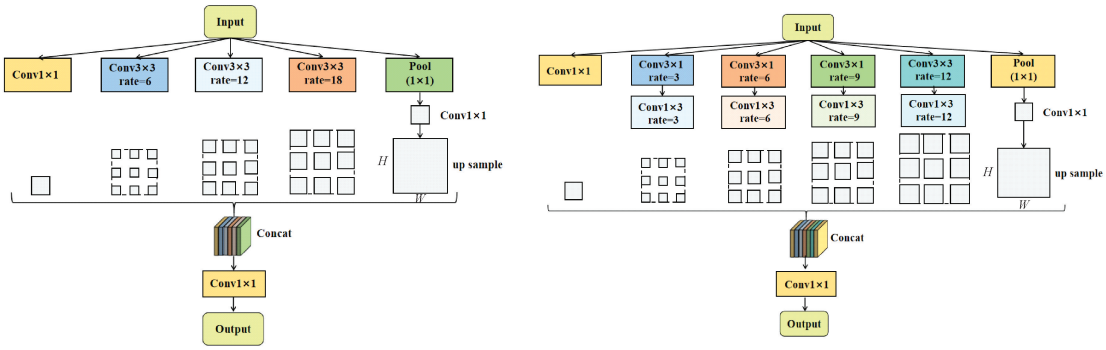
图 2　ConvNeXt 模型结构

Fig. 2　ConvNeXt model architecture

pooling，ASPP)模块。

传统 ASPP 模块[9]通过并行使用不同空洞率的卷积操作来增加感受野，捕捉多尺度特征。为了更好地适应食品图像的多样性和复杂性，故对 ASPP 模块进行改进。本文选择了 4 个不同大小的空洞率(3，6，9，12)的空洞卷积，从而更加全面地捕捉不同尺度的特征。此外，为了减少参数量并提高计算效率，将原有的 3 个并列的 3×3 空洞卷积替换为一组 3×1 和 1×3 的卷积，并增加了一个额外的分支，同样使用一组 3×1 和 1×3 的卷积。通过这种改进，不仅保留了 ASPP 模块在捕捉多尺度特征和上下文信息方面的优势，还显著降低了计算复杂度和参数量，提高了模型的效率。

之后，对不同尺度和大小的特征图进行池化和融合，以提高模型对不同尺度的分类能力，从而有效缓解尺度变化带来的问题，传统 ASPP 模块和 IASPP 模块结构如图 3 所示。



(a) 传统ASPP结构
(a) Traditional ASPP structure

(b) LASPP结构
(b) LASPP structure

图 3　ASPP 结构

Fig. 3　ASPP structure

1.3 高效通道注意力机制模块(ECA)

食品图像通常包含复杂的纹理和颜色特征,视觉外观多样且含有较多噪声,这些特性使得食品图像的识别和分类具有较大的挑战性。针对此特性,本文引入高效通道注意力机制模块 ECA[10],通过自适应调整各个通道的权重来突出关键特征,抑制冗余信息,从而增强网络对目标的关注能力。

大多数注意力机制通过增加模块复杂度来提升性能,但这也增加了模型的整体复杂性。ECA 注意力机制以 SENet 为基础进行构建,去掉了全连接层,引入自适应一维卷积以学习相邻通道之间的依赖关系,减少了计算复杂度和参数量,保持了模型的轻量化特性,从而增加通道间的信息交互,其中第 i 个通道 y_i 的权重 ρ_i 可以表示为:

$$\rho_i = \sigma\left(\sum_{j=1}^{k} \omega^j y_i^j\right), y_i^j \in \Omega_i^k \tag{1}$$

其中,Ω_i^k 为与 y_i 相邻的 k 个领域通道;σ 为激活函数;ρ_i 为通道包含的食品图像特征的重要程度,模型对第 i 个通道关注度越高,与该通道对应的 ρ_i 就越大,建模特征之间就越相关。

具体流程如下:首先,输入食品图像特征图大小为 $C \times H \times W$;然后,通过全局平均池化(global average pooling,GAP)操作减少模型参数量并提取全局特征;接着,应用一维卷积操作(大小为 k)以捕捉跨通道的交互信息,其中 k 表示局部通道交互的覆盖率;随后,采用 $Sigmoid$ 函数对权重进行归一化处理;最后,将生成的每个通道权重与输入特征图的对应通道逐一相乘,生成加权后的食品图像特征图。ECA 模块如图 4 所示。其中,C、H、W 分别为输入特征图的通道数、高度和宽度,$Sigmoid$ 为激活函数,⊗为矩阵相乘,$1 \times 1 \times C$ 为自适应一维卷积。

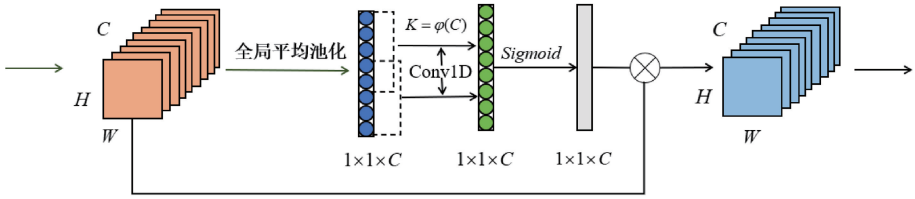


图 4 ECA 结构

Fig. 4 ECA structure

自适应一维卷积主要是通过参数 K 来调整输入通道 C 的大小。K 和 C 关系如式(2)所示。

$$C = \phi(K) = 2^{(\gamma * k - b)} \tag{2}$$

$$K = \varphi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{3}$$

其中,odd 表示 K 取奇数,C 表示通道数,本文中 γ 和 b 分别取 2 和 1。

2 实验结果与分析

2.1 实验环境

实验在 Intel(R) Xeon(R) Silver 4210R@2.40 GHz 处理器,NVIDIA RTX 3080(10 G)显卡,ubuntu 20.04.6 系统,Pytorch 1.13.0,Python 3.10.13 的环境下进行。

2.2 实验数据集

本文选取了两个数据集进行研究,分别是 Vireo Food172[11]和 ETH Food-101[12]。为了改善数据不平衡并提升识别准确度,本文通过网络爬取的方式补充了 Vireo Food172 中图像数量不足 500 张的类别,并删除了重复图像。最终,Vireo Food172 扩展数据集包含 117 967 张图像,172 类食品;ETH Fo-od101 数据集包含 101 000 张图像,101 类食品。所有图像分辨率均为 224×224。图 5 展示了数据集中一些食品类别的示例。数据集按照 8∶2 的比例划分为训练集和验证集。

图 5 食品图片示例

Fig. 5 Sample food images

2.3 评价指标

本文采用多个评价指标衡量模型性能,包括 Top-1 准确率、Top-5 准确率[13]、精确率(Preci-sion)、召回率(Recall)以及 F1 值(F1-score)[14]。

Top-1 准确率指模型预测的概率向量中,最大概率对应的食品类别与实际食品类别匹配的比例;Top-5 准确率指预测概率向量中前 5 个食品类别包含实际食品类别的比例;Precision 指被预测为某一食品类别的样本中,实际为该食品类别的比例;Recall 表示实际为某一食品类别的样本中,被正确预测为该食品类别的比例;F1-score 是精确率和召回率的调和平均值,用于综合考虑分类器的精度和召回性能。

计算公式如下:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1-score=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{6}$$

其中，TP 是正确预测为该类别的样本数，FP 是错误预测为该类别的样本，FN 是实际为该类别但未被正确预测的样本数。

2.4　参数设置

实验采用交叉熵损失函数，并使用随机梯度下降(stochastic gradient descent，SGD)优化策略，训练时的批次大小设为 16，训练周期设为 50 个 epoch，动量参数设为 0.9。学习率初始值设为 5×10^{-4}，每 10 个 epoch 后将学习率衰减至原来的 1/10，所有主干网络均在 ImageNet1K 数据集上进行了预训练。

2.5　对比实验

1)不同 ASPP 模块对比

为了验证 IASPP 模块的有效性，通过对比原始 ASPP 模块以及本文使用的 IASPP 模块，并在验证数据集上进行对比，实验结果如表 1 所示。

实验结果表明，IASPP 模块相较于传统 ASPP 模块，Top-1 准确率提升了 0.23%，精准率、召回率和 F1 分别提升了 0.30%、0.26%和 0.21%；参数量和计算量分别显著降低了 1.17 M 和 0.06 G，这表明，通过在 ASPP 模块中更换卷积方式和扩张率，准确率不仅得到了提升，模型的计算复杂度和参数规模也减少了。因此，本文模型使用 IASPP 模块。

表 1　不同 ASPP 模块的性能比较

Table 1　Performance comparison of different ASPP modules

方法	ASPP 扩张率组合	Params/M	FLOPs/G	Top-1 准确率/%	Top-5 准确率/%	Precision/%	Recall/%	F1-score/%
ASPP	(6,12,18)	48.62	5.41	90.23	98.18	90.41	90.19	90.30
IASPP	(3,6,9,12)	47.45	5.35	90.46	98.12	90.71	90.45	90.51

2)模型对比

为了验证本文方法的有效性，在 Vireo Food172 扩展数据集上，与主流图像分类方法进行了对比实验。实验结果如表 2 所示。

表 2　在 Vireo Food172 扩展集上的性能比较

Table 2　Performance comparison on the Vireo Food172 extended set　%

方法	Top-1 准确率	Top-5 准确率	Precision	Recall	F1-score
VGG16[15]	86.42	96.68	86.66	86.67	86.59
Inception V3[16]	87.42	97.57	87.61	87.75	87.58
ResNet50[8]	88.13	97.90	88.40	88.39	88.27
DenseNet121[17]	88.79	98.01	89.01	89.05	88.94
SENet50[18]	89.46	98.08	89.47	89.46	89.47
ConvNeXt[7]	89.51	98.17	89.51	89.58	89.54
本文	91.56	98.69	91.55	91.49	91.54

表 2 的实验结果表明，本文所提出的模型在 Top-1 准确率、Top-5 准确率、精准率、召回率和 F1 值上均取得了最好的效果。将 ConvNeXt 作为主干网络后，模型的识别准确率相对较好，相对于 DenseNet121 和 SENet50 分别提升了 0.72%和 0.05%。与 ConvNeXt 相比，本文方法准确率提升了约 2.05%，精准率、召回率和 F1 值分别提升了 2.04%、1.91%和 2.00%。实验结果验证了多尺度特征与注意力机制的有效融合对提升食品图像识别准确率起到良好的效果。

为验证本文方法在其他数据集上的泛化能力，在 ETH Food101 数据集上进行了对比实验，如表 3 所示。由表 3 可以看出本文方法同样取得相对较好的准确率。此外，本文方法在 2 个数据集上的精确率和召回率都达到接近准确率的水平，说明了准确率的有效性，也说明了该方法具有较好的鲁棒性。从 F1 值这个综合性指标也可以看出该方法具有较高的鲁棒性，都达到了较高的水平。图 6 展示了在 Vireo Food172 和 ETH Food101 验证集上不同模型的准确率对比情况，以图 6(a)为例，7 种网络均可以对食品图像进行有效识别，均达到 80%以上，而本文模型收敛速度更快，最高结果可达到 91.56%。综上所述，表明本文提出的方法可以提高食品识别准确度从而得到更好的识别效果。

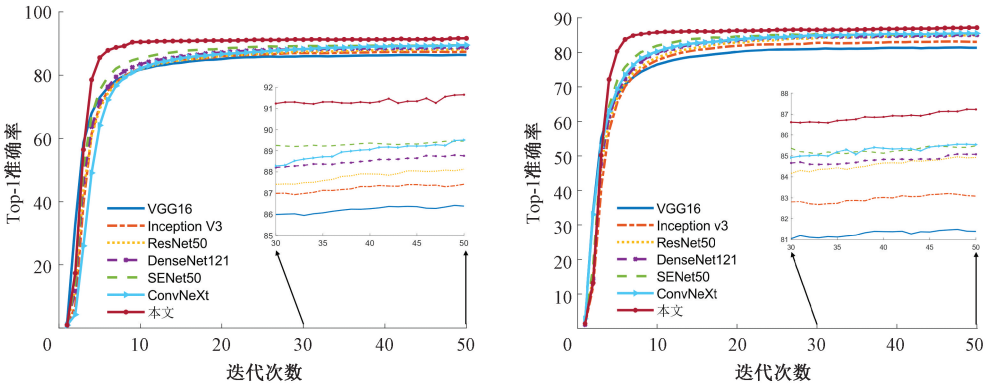
2.6　消融实验

为验证本文各个模块对原 ConvNeXt 模型性能的影响，在 Vireo Food172 扩展数据集上，加入各模块进行消融实验。ConvNeXt＋I 表示在 ConvNeXt 的基础上加上 IASPP 模块，ConvNeXt＋E 表示在 ConvNeXt 的基础上加上 ECA 模块，ConvNeXt＋I＋E 即本文所提方法。消融实验结果如表 4 所示。

表 3 在 ETH Food-101 上的性能对比

Table 3 Performance comparison on ETH Food-101 %

方法	Top-1 准确率	Top-5 准确率	Precision	Recall	F1-scoe
VGG16[15]	81.49	94.31	81.49	81.49	81.44
Inception V3[16]	83.21	94.95	83.16	83.21	83.12
ResNet50[8]	84.95	95.85	84.96	84.95	84.88
DenseNet121[17]	85.08	96.00	85.07	85.07	85.03
SENet50[18]	85.49	95.97	85.43	85.45	85.42
ConvNeXt[7]	85.61	96.38	85.68	85.65	85.63
本文	87.22	97.43	87.25	87.20	87.22



(a) Vireo Food172验证集各模型准确率
(a) Accuracy of each model in the Vireo Food172 validation set

(b) ETH Food101验证集各模型准确率
(b) Accuracy of each model in ETH Food101 validation set

图 6 Top-1 准确率对比

Fig. 6 Comparison of Top-1 accuracy

表 4 消融实验结果

Table 4 Results of ablation experiments %

方法	Top-1 准确率	Top-5 准确率	Precision	Recall	F1-score
ConvNeXt	89.51	98.17	89.51	89.58	89.54
ConvNeXt+I	90.46	98.12	90.71	90.45	90.51
ConvNeXt+E	90.47	98.22	90.47	90.25	90.29
ConvNeXt+I+E	91.56	98.69	91.55	91.49	91.54

由表 4 的结果分析可知,在 ConvNeXt 模型中依次添加 IASPP 模块以及 ECA 模块,识别准确率分别比前者提升了 0.95%、0.96%。将两者结合后,图像识别的效果进一步提升,相较于 ConvNeXt 模型,Top-1 准确率提升了 2.05%,精准率、召回率和 F1 值都分别提升了 2.04%、1.91%和 2.00%,能够更精准地识别食品图像。

2.7 可视化实验

为了分析本文模型对网络提取特征能力的影响,使用 Grad-CAM 算法[19]对特征提取网络最后一层的特征图进行可视化。从数据集中随机选取两个样本图像,选择 VGG16、Inception V3、SENet50 和 ConvNeXt 网络进行特征图可视化,实验结果如图 7 所示。从图 7 中可以看出 SENet50 模型利用 SE 模块更能聚焦于图像中包含食品的区域,而本文模型同样具有较强的全局特征能力。以第 1 个数据集中的食品为例,本文模型相较于其他模型,更聚焦于食品的细节特征,表明了本文方法的有效性。以第 2 个数据集中的食品为例,本文方法可以捕捉到图像中出现的所有食物,而其他网络并未捕捉到图中上半部分出现的食物,同样表明了本文方法的有效性。

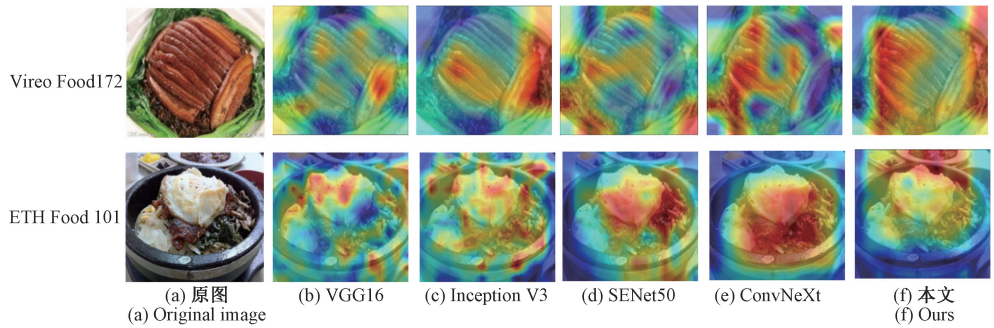


图 7　可视化实验结果

Fig. 7　Visualizing experimental results

3　结　　论

本文提出了一种融合多尺度特征及注意力机制的食品图像识别方法，旨在提高食品图像的识别准确度。为应对食品图像细粒度特点，本文引入了改进的空洞空间金字塔池化模块，以增强不同尺度特征之间的相互作用，并丰富浅层特征的语义信息，从而显著提升特征表达能力。同时，引入高效通道注意力机制，有效地减少了噪声干扰的影响。实验结果表明，本文方法在所使用的数据集上取得了良好的识别性能，相比 SENet 和 ConvNeXt 模型，准确率提高了 2.10%和 2.05%。未来的工作将结合食材成分进一步提升食品图像识别的准确率，继续提升模型的检测性能，同时降低模型的复杂度，通过轻量化设计来优化模型的效率。

参考文献

[1] 闵巍庆，刘林虎，刘宇昕，等. 食品图像识别方法综述[J]. 计算机学报，2022，45(3)：542-566.
MIN W Q，LIU L H，LIU Y X，et al. A survey on food image recognition [J]. Chinese Journal of Computers，2022，45(3)：542-566.

[2] YANG H，KANG S，PARK C，et al. A hierarchical deep model for food classification from photographs[J]. KSII Transactions on Internet and Information Systems(TIIS)，2020，14(4)：1704-1720.

[3] MIN W Q，LIU L H，WANG ZH L，et al. ISIA Food-500：A dataset for large-scale food recognition via stacked global-local attention network[C]. 28th ACM International Conference on Multi-media，2020：393-401.

[4] 姜枫，周莉莉. 改进注意力模型的食品图像识别方法[J]. 计算机工程与应用，2024，60(12)：153-159.
JIANG F，ZHOU L L. Method for recognition of food images based on improved attention model [J]. Computer Engineering and Application，2024，60(12)：153-159.

[5] 王海燕，张渺，刘虎林，等. 基于改进的 ResNet 网络的中餐图像识别方法[J]. 陕西科技大学学报，2022，40(1)：154-160.
WANG H Y，ZHANG M，LIU H L，et al. Chinese food image recognition method based on improved ResNet[J]. Journal of Shaanxi University of Science & Technology，2022，40(1)：154-160.

[6] 邓志良，李磊. 基于改进残差网络的中式菜品识别模型[J]. 激光与光电子学进展，2021，58(6)：264-272.
DENG ZH L，LI L. Chinese food recognition model based on improved residual network[J]. Laser & Optoelectronics Progress，2021，58(6)：264-272.

[7] LIU ZH，MAO H Z，WU CH Y，et al. A ConvNet for the 2020s[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)，2022：11966-11976.

[8] HE K M，ZHANG X Y，REN SH Q，et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition，2016：770-778.

[9] 王蒙蒙，刘秀清，张衡，等. 基于双通道特征融合编解码网络的极化 SAR 图像分类[J]. 国外电子测量技术，2023，42(1)：187-196.
WANG M M，LIU X Q，ZHANG H，et al. PolSAR image classification based on dual-channel feature fusion encoder-decoder network[J]. Foreign Electronic Measurement Technology，2023，42(1)：187-196.

[10] 侯艳丽，王娟. 基于 Efficientnet 的红外目标检测算法[J]. 电子测量技术，2023，46(16)：64-72.
HOU Y L，WANG J. Infrared target detection algorithm based on Efficientnet [J]. Electronic Measurement Technology，2023，46(16)：64-72.

[11] CHEN J J，NGO C W. Deep-based ingredient recognition for cooking recipe retrieval[C]. 24th ACM International Conference on Multimedia，2016.

[12] BOSSARD L，GUILLAUMIN M，GOOL L V. Food101-mining discriminative components with

random forests [C]. European Conference on Computer Vision, 2014: 446-461.

[13] 刘雨萌, 桑海峰. 基于关键帧定位的人体异常行为识别[J]. 电子测量与仪器学报, 2024, 38(3): 104-111.
LIU Y M, SANG H F. Human abnormal behavior recognition based on keyframes localization [J]. Journal of Electronic Measurement and Instrument, 2024, 38(3): 104-111.

[14] 周道先, 张吟龙, 徐高飞, 等. 基于形变卷积和深层聚合网络的水下文物检测[J]. 仪器仪表学报, 2023, 44(11): 185-195.
ZHOU D X, ZHANG Y L, XU G F, et al. Underwater cultural artifact detection based on deformable convolution and deep layer aggregation network[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 185-195.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv preprint arXiv:1409.1556, 2014.

[16] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.

[17] HUANG G, LIU ZH, LAURENS V D M, et al. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2261-2269.

[18] HU J, LI SH, SUN G. Squeeze and excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[19] SELVARAJU R R, MICHAEL C, ABHISHEK D, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 618-626.

作者简介

李苗苗,硕士研究生,主要研究方向为计算机视觉。
E-mail:1930769273@qq.com

华才健(通信作者),博士,副教授,主要研究方向为计算机视觉、大数据分析。
E-mail: hwacj@suse.edu.cn

谢涛,硕士研究生,主要研究方向为智能算法与智能控制。

薛青霞,硕士,主要研究方向为计算机视觉。