冯正全 . 基于反绎学习的食品图像识别方法[J]. 智能计算机与应用,2026,16(1):198-202. DOI:10.20169/j.issn.2095-2163.24032701

基于反绎学习的食品图像识别方法

冯正全

(南京邮电大学 计算机学院、软件学院、网络空间安全学院, 南京 210023)

摘 要: 食品图像识别、包含食品类别识别和食材类别识别,是计算机视觉领域的一项重要任务,在近年来取得了可观的成果。然而,受限于有限的标注数据,使得很多优秀的模型未能发挥出最大的性能。与此同时,与其他视觉任务不同的是,食品和食材之间存在知识联系,这一点在以前的工作中并未得到充分利用。本文提出了一种新方法,即基于反绎学习的食品和食材识别(AFIRE),该方法利用反绎推理来优化感知模型的结果,并将修正后的标签用于训练感知模型,从而降低感知模型对于标注数据的依赖。在 3 个广泛使用的食品数据集上进行了仿真实验。结果表明,在使用相同标注数据量训练的情况下,与未使用 AFIRE 的感知模型相比,模型性能平均提升 2.43%。

关键词: 食品图像识别; 反绎学习; 知识推理

中图分类号: TP391.41 **文献标志码**: A **文章编号**: 2095-2163(2026)01-0198-06

Abductive learning-based food image recognition method

FENG Zhengquan

(School of Computer Science, School of Software, School of Cyberspace Security, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Food image recognition, encompassing both food recognition and ingredient recognition, is an important task in the field of computer vision that has achieved remarkable results in recent years. However, the performance of many excellent models is limited due to the scarcity of labeled data. Unlike other vision tasks, there is a knowledge connection between food and ingredients, and the connection has not been fully utilized in previous works. This paper proposes a new method, namely, Abductive learning for Food and Ingredient REcognition (AFIRE), which utilizes abductive reasoning to optimize the results of the perceptual model and uses the corrected labels to train the perceptual model, thereby reducing the perceptual model's dependence on labeled data. The paper conducts the experiments on three widely used food datasets, and the results show that with the same amount of annotated data for training, the model performance is improved by an average of 2.43% compared to the perceptual model without AFIRE.

Key words: food image recognition; abductive learning; knowledge reasoning

0 引 言

近年来,与食品相关的研究越来越受到关注[1-3]。从图像中识别食品和食材是一个重要的领域课题。和诸多计算机视觉任务一样,食品和食材识别效率近年来不断提升,并已取得了较高的性能[4]。然而,与其他计算机视觉任务相比,食品和食材识别面临着如下挑战:

(1)食品和食材识别属于细粒度图像识别[5],识别其各自的类别时捕捉关键特征尤其挑战。

(2)食品相关的标注数据集有限,这对需要此类数据进行训练的机器学习模型的开发构成挑战。

除上述挑战外,食品和食材识别还具有基于知识的关系特性,这种关系可以进一步用于开发更有效的食品和食材识别机器学习模型。

考虑到这些特点和挑战,研究提出了一种基于反绎学习的食品和食材识别方法(AFIRE)。该方法基于反绎学习(ABL)[6-7]框架,结合机器学习与一阶逻辑推理。与传统的反绎学习不同,本文提出的方法使用多任务感知模型同时识别食品和食材。感知模型的输出随后被输入到推理模型中,该模型执行一阶逻辑推理以纠正感知模型输出中的任何错误或不一致。通过将基于知识的推理纳入食品和食材识别过程,结合反绎学习框架中感知模型与推理模

作者简介: 冯正全(1990—),男,硕士研究生,主要研究方向:计算机视觉。Email:zq.feng@outlook.com。

收稿日期: 2024-03-27

型相互协作的特性，本文的方法可以在较少的标注数据集上展现出较高的性能。

1　相关工作

1.1　视觉增强注意力机

食品和食材识别主要有 2 种方法。一种是将 2 项任务视为独立的任务，另一种是考虑 2 项任务之间的关系，并对其同时加以解决。

目前，大多数研究工作主要关注单标签食品图像识别[8-12]，每张图像中只有一个食品对象。深度学习的发展使大多数研究[13-14]利用卷积神经网络（CNNs）来预测食品类别。一些研究[11,15]还为食品图像特性设计了特定的模型来预测食品类别。食材识别[16-18]是一项更细粒度的识别任务。例如，Bolaños 等学者[9]结合 Inceptionv3 和 ResNet-50 网络进行多标签食材预测。Pan 等学者[19]结合不同类型的 CNNs 进行食材识别。

在同时考虑食品和食材关系的任务中，多任务学习和知识图谱是 2 个主要方向。Zhang 等学者[20]提出了一种多任务学习方法，使用 CNNs 同时从菜肴图像中识别食品类型、食材和烹饪方法。Zhou 等学者[17]利用双分图结构来模拟食材和标签之间的丰富关系，然后将双分图与 CNNs 结合，实现多标签食材识别和食品识别。Wang 等学者[21]提出了一个多任务学习框架，用于同时预测食品类别和食材，其中构建了一个食材字典，用于食材相关区域的发现，并生成了一个以食材为中心的语义-视觉图卷积网络，用于学习食材关系。

在近期的工作中，由于模型如 Vision Transformer[22]和 Swin-Transformer[23]的发展，基于注意力机制[24]的方法也被用于食品和食材识别领域。当前细粒度识别的关键在于有效设计注意力结构，以更多地关注局部判别区域。因此，为食品图像识别定制的 Transformer 网络有望在食品识别中实现更高的性能。采用了 2 种代表性架构作为嵌入网络的 Progressive Self-Distillation（PSD）[25-26]，包括 CNN 和 Vision Transformer。Zhu 等学者[25]提出了一个基于监督的 Transformer 网络食品识别系统，可以在包含各种食品的冰箱环境中检测和识别食品类别。

1.2　反绎学习

反绎学习（ABL）[7]是一种创新的框架，将机器学习与一阶逻辑推理相结合，实现了两者之间的相互增益。ABL 的核心在于利用一阶逻辑规则对模型预测的标签进行修正，并以此修正后的标签作为训练数据集，进而优化机器学习模型。在 ABL 框架中，机器学习部分负责将原始数据转换为基本的逻辑事实，这些事实作为伪标签（$\hat{y}$）供逻辑推理使用。通过反绎过程，系统能够基于一阶逻辑规则推断出特定的事实，这些事实作为反绎标签（$\bar{y}$）被视作真值标签，用于指导和改进机器学习模型的训练过程。

反绎、作为一种基本的逻辑推理方式，旨在根据已知的前提或规则，寻找能够对观察到的现象做出最佳解释的假设。例如，如果已经掌握了“雨导致地面湿润”这一规则，当再次观察到地面湿润时，就可以推测可能是下雨了。ABL 正是基于这样的逻辑推理原理，通过机器学习模型与逻辑规则的相互作用，实现了对数据的更深层次理解和模型性能的提升。这种方法不仅增强了模型对数据的解释能力，也为处理复杂问题提供了新的视角，尤其是在需要逻辑推理和模式识别相结合的领域、如食品图像识别等，ABL 展现出了巨大的潜力和应用价值。

2　AFIRE

本节详细介绍研究提出的方法 AFIRE。AFIRE 框架设计如图 1 所示。AFIRE 遵循反绎学习框架的结构，包括 2 部分：感知模型和推理模型。其中，感知模型基于多任务学习框架，可以对食品识别和食材识别同时进行处理。推理模型则是基于知识图谱推理方法，负责纠正感知模型中的潜在错误。

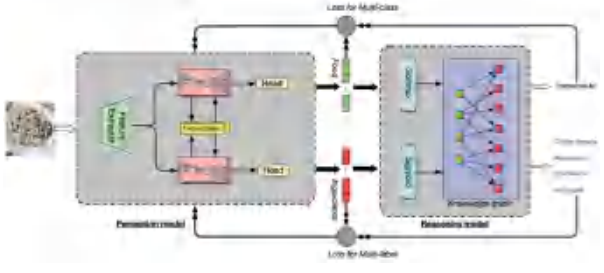

图 1　AFIRE 框架

Fig. 1　AFIRE framework

2.1　多任务感知模型

在 AFIRE 框架中，感知模型发挥着至关重要的作用，负责执行目标任务。本文设计该感知模型为多任务架构，以便带来双重优势：一方面，能够通过并行学习食品及其成分的特征来高效提取关键信息；另一方面，该模型能同时输出食品和成分的识别结果，为逻辑推理过程提供必要的前提条件。

在多任务感知模型（MTPM）中，研究用到的公

式具体如下：

$$f_x = \text{Embedding}(x) \tag{1}$$

利用嵌入层从输入图像 x 中提取共享特征 f_x。该过程实现了从图像空间到共享特征空间的转换，为后续任务提供统一的特征表示。本研究中的嵌入层采用预训练的 Swin-Transformer v2 模型，通过层次化的特征转换机制，有效地捕获了图像的通用视觉特征。

在模型的前向传播阶段，用如下公式进行定义：

$$\hat{f}_{x_{\text{food}}}, \hat{f}_{x_{\text{food}}} = \Delta(\text{Decoder}_{\text{food}}(f_x), \text{Decoder}_{\text{ing}}(f_x)) \tag{2}$$

共享特征被分别传递至 2 个任务特定的解码器 $\text{Decoder}_{\text{food}}$ 和 $\text{Decoder}_{\text{ing}}$。此处，代表了本文采用的特征融合策略，旨在整合各任务的特征表征。通过这种融合机制，得到了针对每个任务优化的特征嵌入 $\hat{f}_{x_{\text{food}}}$ 和 $\hat{f}_{x_{\text{ing}}}$。

在此基础上，进一步推得：

$$\text{logits}_{\text{food}} = \text{HEAD}_{\text{food}}(\hat{f}_{x_{\text{food}}}) \tag{3}$$

$$\text{logits}_{\text{ing}} = \text{HEAD}_{\text{ing}}(\hat{f}_{x_{\text{ing}}}) \tag{4}$$

其中，$\text{HEAD}_{\text{food}}$ 和 HEAD_{ing} 分别表示任务特定的输出头，通过接受经过优化的特征嵌入 $\hat{f}_{x_{\text{food}}}$ 和 $\hat{f}_{x_{\text{ing}}}$ 作为输入，分别生成食品和成分识别的预测结果 $\text{logits}_{\text{food}}$ 和 $\text{logits}_{\text{ing}}$。

模型在反向传播过程中，采用融合的损失函数来统一更新梯度，增强了不同任务间的学习协同效应。总损失函数定义如下：

$$L_{\text{total}} = \sum_{t=1}^{T} \alpha_t L_t \tag{5}$$

其中，T 表示任务总数；L_t 表示第 t 个任务的损失函数；α_t 表示对应的任务权重，确保模型在多任务学习中平衡各个任务的重要性。

2.2 推理模型

在 AFIRE 中，使用了知识图谱推理[27]来实现反绎的过程。具体过程包含知识图谱构建和推理方法实现两部分。

2.2.1 构建知识图谱 FIKG

世界上的大多数食品都是根据现有食谱准备的。尽管一些食品可能略微偏离标准食谱，但通常不会有明显的差异。最近，几个高质量的数据集，如 Vireo Food-172[6] 和 Ingredient-101[9]，已经被标记了食品类别和其中使用的食材。研究中可以使用标记数据为基础，构建与每个数据集相对应的知识图谱。AFIRE 将使用标记食品数据集作为构建 FIKG 的数据基础。

所描述的知识图谱，可以表示为：

$$\text{FIKG G} = (V, E, X) \tag{6}$$

其中，包含 2 种不同类型的节点，即 $\text{Node}_{\text{food}} \subset V$ 和 $\text{Node}_{\text{ing}} \subset V$，以及表示食品实体与食材实体之间关联的边 E。图中的每条边都附带一个频率属性 X，该属性捕获了相应成分与对应食品结合使用的概率。值得注意的是，由于烹饪过程中的变化，某些食谱中列出的食材可能不会出现在最终菜品中。因此，并不是菜谱中所有的食材都会出现在所构建的知识图谱中。

2.2.2 基于静态知识图谱推理的 PLR 方法

知识图谱推理通过利用知识来修改感知模型的预测标签，以最大化其与给定知识库的一致性。在食品识别任务的背景下，Top-5 的识别精度通常处于较高水平。因此，研究提出一种新颖的方法。该方法结合预测的食材结果和知识推理，从 Top-5 结果中识别出与知识库最高一致性的食品和食材结果。这是通过引入 PLR 方法实现的。

在本文中，研究使用变量来跟踪感知模型的食品和成分精度。最初，在感知模型中计算 FIP，表示为 $\text{FIP}_{\text{food}} \subset \mathbb{R}^{C_{\text{food}}}$ 和 $\text{FIP}_{\text{ing}} \subset \mathbb{R}^{C_{\text{ing}}}$，其中 C_{food} 表示食品类别的总数，C_{ing} 表示成分类别的总数。

食材 $\text{Node}_{\text{ing}}[j]$ 被用于菜品 $\text{Node}_{\text{food}}[i]$ 的概率表示为 P_{ij}，根据 FIKG 被菜品 i 使用的食材集合表示为 $J_i = \{\exists i \rightarrow j,\ i \in \text{Node}_{\text{food}},\ j \in \text{Node}_{\text{ing}}\}$。

在 PLR 过程中，为 Top-5 预测结果中的每个菜品候选项计算排名分数。对此可以表示为：

$$\text{Scores}[i] = \text{FIP}_{\text{food}}[i] \cdot \text{Softmax}(\text{logits}_{\text{food}})[i] + \beta \cdot \sum_{j}^{j \in J_i} (\text{FIP}_{\text{ing}}[j] \cdot \text{Sigmoid}(\text{logits}_{\text{ing}})[j] \cdot P_{ij}) \tag{7}$$

排名分数综合考虑了感知模型预测的食品及其相关成分的置信度，以及知识图谱中的成分-食品关联性。

具体来说，$\text{Scores}[i]$ 汇总了菜品识别精度 ($\text{FIP}_{\text{food}}[i]$)、菜品预测概率 ($\text{Softmax}(\text{logits}_{\text{food}})[i]$)、食材识别精度 ($\text{FIP}_{\text{ing}}[j]$)、食材预测概率 ($\text{Sigmoid}(\text{logits}_{\text{ing}})[j]$)，以及基于知识图谱知识，食材 j 与菜品 i 关联的先验概率 P_{ij}。超参数 β 控制了与食材相关项的影响权重。研究中通过结果分数对 Top-5 候选者进行排名，PLR 能够纠正预测结果。其数学模型如下：

$$\bar{y}_{\text{food}} = \text{Argmax}(\text{Scores}) \tag{8}$$

$$\bar{y}_{\text{ing}} = J_{\bar{y}_{\text{food}}} \tag{9}$$

其中，$\bar{y}_{\text{food}}$ 和 $\bar{y}_{\text{ing}}$ 表示通过 PLR 修改后的标签。

算法 1 展示了 AFIRE 的工作流程，其中 $\bar{y}$ 是经过 PLR 修正后的标签，$\bar{y}$ 会作为目标标签，与输入 x 一起用于更新 MTPM。

算法 1　AFIRE

输入　未标注数据集 D，标注数据集 D'，在标注数据上训练好的感知模型 MTPM，食品和食材知识图谱 FIKG，迭代次数 E，推理方法 PLR，食品和食材精度 IP，更新 FIP 间隔的迭代数 Z

输出　更新后的感知模型 MTPM

```
for e = 1 to E do
  STEPS=0
  reset FIP
  for x in D do
    if STEPS% Z == 0 do
      updateFIP
    logits=MTPM (x)
    ȳ = PLR(logits,FIKG)
    update MTPM via (x,ȳ)
    STEPS++
  endfor
endfor
return MTPM
```

3　实验

3.1　数据集

实验中，使用的数据集详情如下。

(1) ETH Food-101[8]。是一个典型的西式食品数据集，包含来自 101 个食品类别的 101 000 张图像。每个类别有 1 000 张图像，包括 750 张训练图像和 250 张测试图像。

(2) Ingredient-101[9]。包括 ETH Food-101 数据集中包含的 101 个食品类别的最常见成分列表，共有 446 种独特成分(每个食谱平均 9 种)。该数据集被划分为训练集、验证集和测试集，以确保 101 种食品类型平衡。

(3) Vireo Food-172[6]。是一个大规模的中国菜食品数据集，包含 172 个食品类别。由 110 241 张图像组成，每个类别平均有 640 张图像。这些图像来自互联网和专业摄影师，并且标记有食品类别和成分信息。该数据集包括 353 种独特的成分，由专业人士标注。

3.2　实验设置

实验设置中，将每个数据集划分为 3 个子集：训练集、验证集和测试集。对于训练集，进一步将其分成 2 部分，使用 3 个预定义的比例。一个是标记数据子集，另一个是未标记数据子集。标记数据子集首先用于感知模型的预训练，预训练完成后，会和未标记数据子集一起用于 AFIRE 方法中进行训练。具体的数据分割方案见表 1。

表 1　数据集划分

Table 1　Datasets setting　%

实验方法	训练集		验证集	测试集
	标注数据	未标注数据		
D_1	54	6	10	30
D_2	48	12	10	30
D_3	42	18	10	30

为了评估食品识别的性能，采用了 Top-1 准确率和 Top-5 准确率作为主要评价指标。对于食材识别性能的评估，则使用宏观 $F1$ 分数来评估每个成分类别的平均性能，以及微观 $F1$ 分数来评估所有样本的整体性能。

多任务感知模型(MTPM)的实现过程中，为了保持开发简洁性，研究中基于开源的 LibMTL[28] 框架进行编码，并进一步开发了 MTPM 所需的模块。MTPM 的每个任务的损失函数不同。对于食品识别任务，将其视为多类分类问题，并使用交叉熵损失函数。对于成分识别任务，将其视为多标签问题，使用二值交叉熵损失作为损失函数。通过固定权重简单地融合了 2 个损失函数，每个任务的损失函数权重设置为 $\frac{1}{2}$，其中 2 表示任务数量。

推理模型中，将 β 值设置为 0.4，这表示推理结果最终的得分中，食材所占的权重为 0.4。

3.3　结果分析

ETH Food-101 和 Ingredient-101 数据集的实验结果见表 2，Vireo Food-172 数据集的实验结果见表 3。在表 2 和表 3 中，R_{label} 指的是在 D_i 的标记数据上训练的多任务感知模型，R_{afire} 指的是使用 AFIRE 在 D_i 的标记数据和未标记数据上训练的多任务感知模型。整体上来看，R_{afire} 对比 R_{label} 都有了明显的提升，这表明在标注数据大小一致的情况下，AFIRE 可以有效地利用无标注数据提升感知模型的性能。

此外，由表 2 中结果可以发现，AFIRE 方法对于食材识别性能的提升远超过对食品识别的影响。究其原因就在于，相比于食材，食品识别的性能本身就

相对更优,因此在食品识别方面的提升空间相对较小。

表 2 ETH Food-101 和 Ingredient-101 数据集的实验结果

Table 2 Experimental results on the ETH Food-101 and Ingredient-101 datasets

实验方法		食品识别		成分识别	
		Top-1 准确率/%	Top-5 准确率/%	微观 $F1$ 分数	宏观 $F1$ 分数
D_1	R_{label}	88.98	96.21	87.16	86.68
	R_{afire}	**89.03**	-	**88.87**	**88.36**
D_2	R_{label}	88.24	96.18	86.67	84.60
	R_{afire}	**88.26**	-	**88.21**	**86.33**
D_3	R_{label}	87.45	95.62	85.47	84.19
	R_{afire}	**87.66**	-	**87.74**	**85.17**

表 3 Vireo Food-172 数据集的实验结果

Table 3 Experimental results on the Vireo Food-172 dataset

实验方法		食品识别		成分识别	
		Top-1 准确率/%	Top-5 准确率/%	微观 $F1$ 分数	宏观 $F1$ 分数
D_1	R_{label}	84.93	96.59	80.14	54.94
	R_{afire}	**86.13**	-	**82.99**	**56.04**
D_2	R_{label}	83.27	95.52	78.54	53.24
	R_{afire}	**84.53**	-	**79.17**	**53.50**
D_3	R_{label}	82.75	94.96	77.13	**51.99**
	R_{afire}	**83.15**	-	**78.12**	51.59

表 3 展示了 Vireo Food-172 数据集在使用 AFIRE 后,食材识别的整体性能有了显著提升。然而,针对 D_3 数据集的宏观 $F1$ 分数却低于未使用 AFIRE 的情况。这一现象可能源于 Vireo Food-172 数据集中食材样本的分布不均[29],特别是某些食材样本量极少。这种不均衡分布导致这些少数样本的识别性能在计算宏观 $F1$ 分数时具有较大的权重,进而影响了整个系统的综合表现。这一发现强调了在处理具有不均衡样本分布的数据集时,需要采用更细致的方法来优化模型的性能,确保即便在样本稀缺的情况下也能保持较高的识别准确率。

4 结束语

本文提出一种基于反绎学习的食品图像识别方法 AFIRE,该方法为食品图像识别领域提供了一种新的解决方案。首次将反绎学习应用于食品和食材识别领域,并提出使用多任务模型作为反绎学习的感知模型,同时识别食品和食材,生成适合推理的结果。本文在 3 个流行的食品数据集上进行了对比实验:2 个西方食品数据集(ETH Food-101[8] 和 Ingredient-101[9]),以及 1 个中国食品数据集 Vireo Food-172[6]。研究结果表明,本文所提的方法在所有 3 个数据集上都不同程度地降低了感知模型性能对标注数据的依赖。研究中利用外部知识来弥补感知模型的不足,降低了感知模型对大量标注数据的依赖,为未来的研究和应用开辟了新的可能性。

参考文献

[1] TIAN Yijun, ZHANG Chuxu, GUO Zhichun, et al. Recipe2Vec: Multi-modal recipe representation learning with graph neural networks[J]. arXiv preprint arXiv,2205.12396,2022.

[2] HAUSSMAN N S, SENEVIRATNE O, CHEN Yu, et al. FoodKG: A semantics-driven knowledge graph for food recommendation[C]// Proceedings of Lecture Notes in Computer Science. Cham: Springer, 2019: 146-162.

[3] CHEN Jingjing, ZHU Bin, NGO C W, et al. A study of multi-task and region-wise deep learning for food ingredient recognition [J]. IEEE Transactions on Image Processing, 2021: 1514-1526.

[4] 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.

[5] PENG Yuxin, HE Xiangteng, ZHAO Junjie. Object-part attention model for fine-grained image classification[J]. IEEE Transactions on Image Processing, 2018,27(3): 1487-1500.

[6] CHEN Jingjing, NGO C W. Deep-based ingredient recognition for cooking recipe retrieval[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM, 2016: 32-41.

[7] DAI Wangzhou, XU Qiuling, YU Yang, et al. Bridging machine learning and logical reasoning by abductive learning [C]// Proceedings of 33rd Conference on Neural Information Processing Systems. Vancouver, Canada: NIPS Foundation, 2019: 2811-2822.

[8] BOSSARD L, GUILLAUMIN M, VAN GOOL L. Food-101-mining discriminative components with random forests [M]// FLEET D, PAJDLA T, SCHIELE B, et al. Computer Vision, Lecture Notes in Computer Science. Cham: Springer, 2014: 446-461.

[9] BOLAÑOS M, AINA F, RADEVA P. Food ingredients recognition through multi-label learning [M]//BATTIATO S, FARINELLA G, LEO M, et al. New Trends in Image Analysis and Processing. Lecture Notes in Computer Science. Cham: Springer, 2017:394-402.

[10] YANG Shulin, CHEN Mei, POMERLEAU D, et al. Food recognition using statistics of pairwise local features [C]// Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2010: 2249-2256.

[11] MARTINE L N, FORESTI G L, MICHELONI C. Wide-slice residual networks for food recognition[J]. arXiv preprint arXiv, 1612.06543, 2016.

[12] MING Zhaoyan, CHEN Jingjing, CAO Yu, et al. Food photo recognition for dietary tracking: system and experiment [M]// SCHOEFFMANN K, et al. MultiMedia Modeling. Lecture Notes in Computer Science. Cham: Springer, 2018: 129-141.

[13] YANAI K, KAWANO Y. Food image recognition using deep convolutional network with pre-training and fine-tuning[C]//Proceedings of 2015 IEEE International Conference on Multimedia& Expo Workshops (ICMEW). Piscataway, NJ: IEEE, 2015: 1-6.

[14] KAGAYA H, AIZAWA K, OGAWA M. Food detection and recognition using convolutional neural network[C]//Proceedings of the 22nd ACM International Conference on Multimedia. New York: ACM, 2014: 1085 - 1088.

[15] QIU Jianing, LO F P W, SUN Yingnan, et al. Mining discriminative food regions for accurate food recognition[J]. arXiv preprint arXiv, 2207.03692, 2019.

[16] ZHANG Xiaofan, ZHOU Feng, LIN Yuanqing, et al. Embedding label structures for fine-grained feature representation[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 1114-1123.

[17] ZHOU Feng, LIN Yuanqing. Fine-grained image classification by exploring bipartite-graph labels[J]. arXiv preprint arXiv, 1512.02665, 2015.

[18] CHEN Jingjing, ZHU Bin, NGO C W, et al. A study of multi-task and region-wise deep learning for food ingredient recognition [J]. IEEE Transactions on Image Processing, 2020, 30: 1514-1526.

[19] PAN Lili, LI Cong, POUYANFAR S, et al. A novel combinational convolutional neural network for automatic food-ingredient classification[J]. Computers, Materials and Continua, 2019, 61(3): 731-746.

[20] ZHANG Xijin, LU Yifan, ZHANG Songhai. Multi-task learning for food identification and analysis with deep convolutional neural networks[J]. Journal of Computer Science & Technology, 2016 (3): 489-500.

[21] WANG Zhiling, MIN Weiqing, LI Zhuo, et al. Ingredient-guided region discovery and relationship modeling for food category-ingredient prediction[J]. IEEE Transactions on Image Processing, 2022, 31: 5214-5226.

[22] LIU Ze, LIN Yutong, CAO Yue, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2021: 9992-10002.

[23] ZHU Yaohui, LIU Linhu, TIAN J. Learn more for food recognition via progressive self-distillation[J]. arXiv preprint arXiv, 2303.05073, 2023.

[24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image is Worth 16×16 Words: Transformers for image recognition at scale[J]. arXiv preprint arXiv, 2010.11929, 2020.

[25] ZHU Yousong, ZHAO Xu, ZHAO Chaoyang, et al. Food det: Detecting foods in refrigerator with supervised transformer network [J]. Neurocomputing, 2020, 379: 62-171.

[26] LIANG Ke, MENG Lingyuan, LIU Meng, et al. A survey of knowledge graph reasoning on graph types: Static, Dynamic, and Multimodal[J]. arXiv preprint arXiv, 2212.05767, 2023.

[27] LIN Baijiong, ZHANG Yu. LibMTL: A Python library for multi-task learning[J]. arXiv preprint arXiv, 2203.14338 , 2021.

[28] HUANG Yi, GILEDERELI B, KÖKSAL A, et al. Balancing methods for multi-label text classification with long-tailed class distribution[J]. arXiv preprint arXiv, 2109.04712, 2021.

[29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of Neural Information Processing Systems. Long Beach, USA: NIPS Foundation, 2017: 5998-6008.