

# 食品图像识别方法综述

闵巍庆<sup>1),2),3)</sup> 刘林虎<sup>1),2),3)</sup> 刘宇昕<sup>1),2),3)</sup> 罗梦江<sup>1),2),3)</sup> 蒋树强<sup>1),2),3)</sup>

<sup>1)</sup>(中国科学院智能信息处理重点实验室 北京 100190)

<sup>2)</sup>(中国科学院计算技术研究所 北京 100190)

<sup>3)</sup>(中国科学院大学 北京 100049)

**摘要** 食品与人类的行为、健康和文化等密切相关。来自社交网络、移动网络和物联网等泛在网络产生的食品大数据及人工智能尤其是深度学习技术的快速发展催生了新的交叉研究领域食品计算。作为食品计算的核心任务之一,食品图像识别同时是计算机视觉领域中细粒度视觉识别的重要分支,因而具有重要的理论研究意义,并在智慧健康、食品智能装备、智慧餐饮、智能零售及智能家居等方面有着广泛的应用前景。为此,本文对食品图像识别领域进行了全面综述。该文首先从识别对象、视觉特征表示及学习方法等不同维度对目前的食品图像识别进行分类并深入阐述和分析了当前的研究进展,指出深度学习模型是当前食品图像识别的主流方法,融入各种上下文信息和外部知识是改进食品图像识别的重要手段。其次从食品数据规模、食品类型及是否包含上下文和多模态信息等方面介绍了现有的食品图像识别数据库,并给出了不同识别方法在主流数据库上的性能对比和分析。接下来总结了食品图像识别在饮食质量评估、食品自动结算、智能厨具、食品图像的组织检索和推荐等多个方面的应用情况。最后本文从面向食品特点的食品图像识别方法、大规模食品图像识别基准数据库构建和基于多传感器信息融合的食品识别等方面对食品图像识别未来可能的研究方向进行了总结和展望。

**关键词** 食品图像识别; 深度学习; 饮食管理; 食品计算

中图法分类号 DOI 号

## A Survey on Food Image Recognition

MIN Wei-Qing<sup>1)2)3)</sup> LIU Lin-Hu<sup>1)2)3)</sup> LIU Yu-Xin<sup>1)2)3)</sup>  
LUO Meng-Jiang<sup>1)2)3)</sup> JIANG Shu-Qiang<sup>1)2)3)</sup>

<sup>1)</sup>(Key Lab of Intelligent Information Processing, Chinese Academy of Sciences, Beijing 100190)

<sup>2)</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

<sup>3)</sup>(University of Chinese Academy of Sciences, Beijing 100049)

**Abstract** Food is closely related to human behavior, health and culture. The food big data generated from ubiquitous networks (e.g., social networks, mobile networks and the Internet of Things), and the rapid development of artificial intelligence, especially deep learning technology have brought a new multidisciplinary field called food computing. As one of the core tasks of food computing, food image recognition is an important branch of fine-grained visual classification (FGVC) in the field of computer vision. Therefore, it has important theoretical research significance and wide applications, such as smart health, food smart equipment, smart catering, smart retail and smart home. To this end, the paper proposes a comprehensive overview of food image recognition. The paper first classifies current food image recognition methods from different dimensions such as recognition objects, visual feature representation and learning methods, and deeply elaborates and analyzes

current research progress. The paper points out that deep learning models are the current mainstream method of food image recognition and integrating various contextual information and external knowledge is very helpful to improve the recognition performance. Second, the paper introduces existing food image recognition datasets from different dimensions such as the scale of dataset, food types, contextual information and multi-modal information, and compares and analyzes the performance of different recognition methods on mainstream datasets. Third, the paper summarizes the application of food image recognition in many aspects such as diet quality evaluation, automatic checkout service, smart kitchenware, food image organization, retrieval and recommendation. Finally, the paper summarizes and prospects future research directions of food image recognition such as food characteristic-oriented image recognition, construction of large-scale food image recognition benchmarks and multi-sensor information fusion-based food recognition.

**Key words** food image recognition; deep learning; dietary management; food computing

## 1 引言

食物不仅对人类的健康营养等生活方面产生深远影响，而且在定义我们的身份、社会地位和文化等方面同样起着重要作用<sup>[1]</sup>，正如法国美食家布里亚·萨瓦兰所说的“告诉我你吃什么，我就告诉你你是谁”。因此，与食物相关的研究<sup>[2-3]</sup>一直是一个研究热点。来自不同研究领域的研究人员已经从不同角度进行了与食物相关的研究，包括食物选择<sup>[4]</sup>、食物感知<sup>[5]</sup>、食品消费<sup>[6]</sup>、食品安全<sup>[7]</sup>和食品文化<sup>[8]</sup>等。由于食品相关研究涉及到多个研究领域，因此研究的较为分散，缺乏一个系统整体。2019年，Min 等人<sup>[9]</sup>系统提出了食品计算框架，指出食品计算包括食品感知、识别、检索和推荐等多个任务，并服务于医学、生物学、农业、食品工业和营养健康等多个领域。其中食品图像识别是其基础且核心任务之一。

从计算机视觉领域来看，食品图像识别属于细粒度视觉识别的重要分支<sup>[10-13]</sup>，具有重要的研究价值。当前各种便携式设备（如手机和相机等）及可穿戴设备（如可穿戴相机）的广泛普及和人工智能技术的迅速发展使得食品图像识别同时具有广泛的应用前景。例如，我们一旦识别出菜品的类别、食材或其他属性信息，可以进行各种健康相关的营养分析和饮食习惯分析，从而实现自动高效的多媒体饮食日志记录和饮食质量评估，实现个人的健康监管和疾病防控。食品图像识别可通过识别客户用餐、生鲜果蔬和包装食品等实现食品的自动结算，应用在无人餐厅、无人超市及食品工业中。此外，通过食品图像识别还可以进一步实现食品推荐和

社交网站的食品图像组织检索。正因为如此，食品图像识别已逐渐成为计算机视觉<sup>[14-15]</sup>、多媒体<sup>[16-19]</sup>、工业信息学<sup>[20-21]</sup>、医学与健康信息学<sup>[22-24]</sup>、农业和生物工程<sup>[25-28]</sup>等多个领域的研究热点。



图 1 来自食品数据 ETHZ Food-101<sup>[15]</sup>, Vireo Food-172<sup>[4]</sup>, ISIA Food-200<sup>[18]</sup>, VegFru<sup>[13]</sup> 和 Grocery<sup>[29]</sup> 的一些实例图像

食品图像识别属于细粒度图像识别。细粒度图像识别是指对同一类物体下的不同子类进行识别，如识别不同种类的鸟或者不同类型的车。食品图像识别的主要任务是针对含有食品的图像利用计算机技术识别出食品的种类或其他不同粒度的语义信息（如食材（Ingredient）和菜系等）。根据《中华人民共和国食品安全法》，食品被定义为各种供人食用或者饮用的成品和原料以及按照传统既是食品又是药品的物品。因此食品的种类丰富多样，当前主流的食品图像识别主要是针对菜品的识别，但也包括对生鲜果蔬、食材和包装食品的识别等。图 1 展示了来自当前食品数据库的一些实例图像。现有食品图像识别主要针对某大类食品下的分类，如菜品识别<sup>[15]</sup>、果蔬识别<sup>[20]</sup>、食材识别<sup>[21]</sup>和包装食

品识别<sup>[29]</sup>等。其识别结果可进一步支持各种应用,如饮食热量估计和自动结算等。

食品图像识别的研究可以追溯到 1977 年, Parrish 等人<sup>[30]</sup>最先开展基于视觉的果蔬识别研究,应用于水果采摘。1996 年, Bolle 等人<sup>[31]</sup>提出了果蔬识别系统 VeggieVision, 用于超市等场景的果蔬结算。在 2000 年之前, 食品图像识别主要用于与生产链和市场食品质量评估相关的工程应用<sup>[32]</sup>。之后由于菜品图像识别的广泛应用价值而受到越来越多的关注。2008 年, Kitamura 等人<sup>[33]</sup>提出了多媒体饮食日志系统。该系统通过菜品图像和非菜品图像的分类得到用户的菜品图像, 然后对其进行成分和热量分析以提供饮食建议。2009 年, Chen 等人<sup>[34]</sup>发布了第一个快餐食品数据库 PFID, 之后于 2010 年提出了一种联合视觉词袋模型和颜色直方图的特征提取方法在该数据库上进行了测试<sup>[35]</sup>。同年

Pornpanomchai 等人<sup>[36]</sup>融合多种手工特征和分类器进行果蔬识别。2014 年 Bossard 等人<sup>[37]</sup>发布了第一个大规模西方菜品图像数据库 Food-101, 并较早将深度学习用于食品图像识别, 同年 Yanai 等人<sup>[38]</sup>也尝试将深度学习用于菜品图像识别。随着深度学习技术的快速发展和大规模食品图像数据库的不断增多, 食品图像识别的相关研究逐渐多了起来。2015 年, 微软等研究机构<sup>[39]</sup>利用传统的视觉词袋模型提取菜品视觉特征用于识别菜品, 并进一步实现菜品的热量估计。同年 Google 提出了一个 Im2Calories 系统<sup>[14]</sup>, 该系统通过训练深度学习网络提取特征进行菜品识别, 然后基于识别的食材估计菜品热量。Xu 等人<sup>[40]</sup>尝试将餐馆位置信息融入到识别模型以改进菜品图像识别性能。之后多个研究团队<sup>[12,18]</sup>挖掘食材等多模态信息, 结合深度学习模

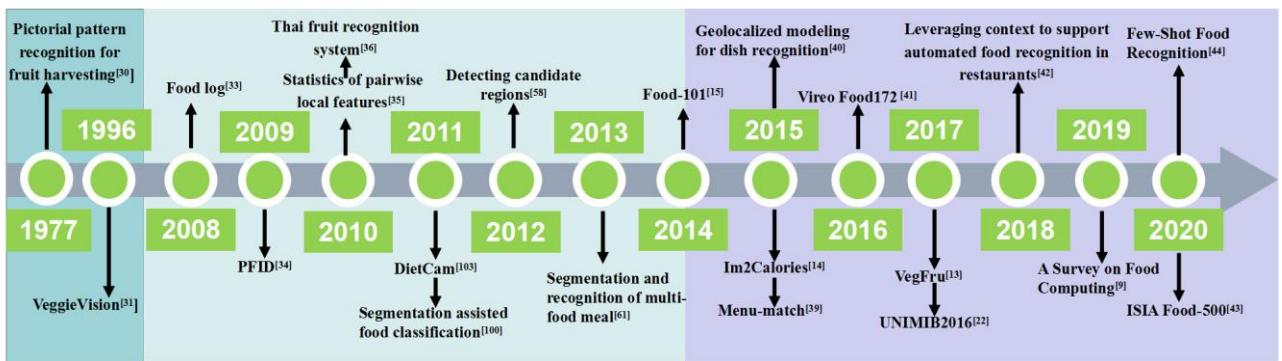


图 2 各时间节点的代表性食品识别工作

型, 通过多任务学习实现菜品和食材的同时识别。2016 年, Chen 等人<sup>[41]</sup>发布了中餐菜品数据库 Vireo Food-172。2017 年 Ciocca 等人<sup>[22]</sup>发布了面向餐厅的托盘菜品识别数据库 UNIMIB2016。同年 Hou 等人<sup>[29]</sup>发布了果蔬数据库 VegFru, 包括 292 类果蔬和超过 160000 果蔬图片。2018 年, Aguilar 等人<sup>[42]</sup>提出了一种面向餐厅托盘的菜品检测和识别方法, 并在 UNIMIB2016 进行了测试。2019 年 Min 等人<sup>[9]</sup>系统提出食品计算框架, 并指出食品图像识别是其核心任务之一, 之后于 2020 年发布了新的大规模菜品数据库 ISIA Food-500<sup>[43]</sup>。该数据库包括中西方菜品种类 500 类, 共 399726 张菜品图片, 是目前菜品数据规模最大的菜品识别数据集, 有望推动食品图像识别领域的进一步发展。近来一些研究学者尝试将小样本学习<sup>[44]</sup>应用到菜品图像识别任务中。与此同时, 由于深度学习技术的发展使得果蔬识别技术在农业相关领域也实现了快速发展<sup>[45]</sup>。按照其

发展脉络, 图 2 展示了各时间节点的代表性食品识别工作, 可以看出 2014 年是一个重要的分界线, 一方面以 AlexNet 网络为代表的深度学习方法<sup>[46]</sup>在图像识别等领域获得了巨大成功; 另一方面是 Food-101 及之后多个较大规模食品图像识别基准数据集的发布推动了食品图像识别技术的迅速发展。

由于食品相关研究的跨学科性, 食品图像识别的研究分散在计算机视觉、多媒体、农业、生物、医学和营养健康等不同的研究领域中。目前尽管有一些食品图像识别的综述, 但是这些工作主要是针对某一个特定领域的总结。例如 Khurram Hameed 等人<sup>[45]</sup>主要面向农业领域系统地阐述了果蔬识别技术, 并总结了果蔬识别的相关应用和发展前景。Lo 等人<sup>[24]</sup>在生物医学与健康信息学领域重点对面向饮食质量评估的菜品图像识别和检测算法给予了系统综述。此外, Min 等人<sup>[9]</sup>在国际上第一次统

提出了食品计算框架,对包括食品图像识别在内的多个食品计算任务进行了较为系统总结。不同于已有综述,本文基于食品定义,以更宽广的视角对食品图像识别进行系统综述。首先从识别对象、视觉特征学习和学习方法等不同维度对食品图像识别技术进行了系统梳理和全面总结,其次介绍了现有

的食品图像识别数据库,并给出了不同识别方法在主流数据库上的性能对比和分析,然后总结了食品图像识别在多个领域中的具体应用情况,最后展望了该领域未来研究的可能发展方向。作为阅读指导,图 3 给出了本综述各部分内容之间的总体路线图。

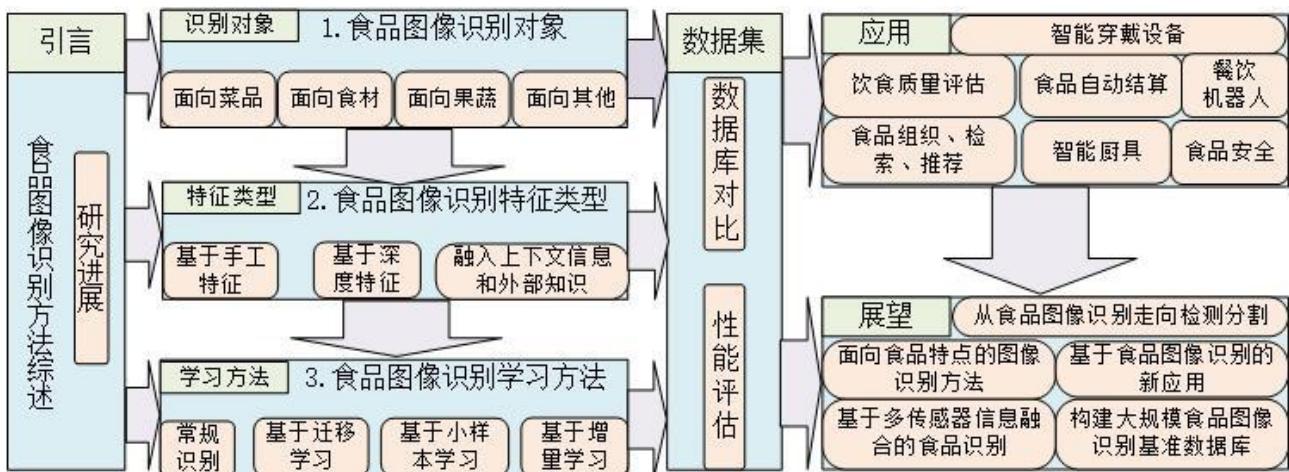


图 3 本文各部分内容的总体路线图

## 2 研究进展

目前已有大量的食品图像识别方法,这些工作发表在计算机视觉、多媒体、医学、营养与健康等不同的领域。考虑到食品种类的多样性,食品图像识别的对象也多种多样,包括菜品、果蔬、食材和包装食品等。针对某一食品类型的食品图像识别系统,我们首先需要考虑应该提取的视觉特征类型,其次根据数据的统计分布等情况考虑所采用的机器学习方法。基于此,本文对现有食品图像识别方法从(1)识别对象;(2)特征类型和(3)学习方法等三个维度分别进行阐述。

### 2.1 识别对象

根据食品的定义,食品图像识别的类型多种多样,包括菜品、食材和果蔬等,除此之外还有一些和食品图像识别相关的任务,如烹饪属性识别、食物状态识别和饮食场景识别等。具体来说,本节从食品图像识别的不同类型出发,将主要从以下四个方面进行回顾和总结:(1)面向菜品的识别;(2)面向食材的识别;(3)面向果蔬的识别;(4)面向其它食品相关类型的识别。

#### 2.1.1 面向菜品的识别

由于菜品图像识别在饮食质量评估、智能厨具

等多个领域的重要应用价值,当前针对菜品图像的识别受到了诸多学者的广泛关注。大多数研究工作主要研究单标签的菜品图像识别<sup>[35,47-56]</sup>,即一张图像中仅含有一个菜品对象。深度学习的发展使得当前大部分工作<sup>[38,50-53]</sup>均采用卷积神经网络(Convolutional Neural Networks, CNN)预测菜品类别。一些工作<sup>[10,54-56]</sup>则针对菜品图像的特性设计专门的模型预测菜品类别。然而在餐饮等许多场景中,菜品图像中经常包含多个菜品类别,此时需要对该图像中的多种菜品同时进行预测和识别。目前多标签菜品图像识别的相关工作还不是太多。第一个多标签菜品识别的工作<sup>[57]</sup>使用了菜品之间的共现关系信息并采用流行排序的方法实现多种菜品的识别。之后大部分方法的思路是先对菜品图像中不同的菜品区域进行检测或分割<sup>[58-62]</sup>,然后对其区域进行识别。例如 Aguilar 等人<sup>[60]</sup>微调物体检测算法 YOLOv2<sup>[64]</sup>来进行多种食物检测和识别。苏等人<sup>[63]</sup>考虑到了中餐食品检测数据集的匮乏,采用了多目标分割算法识别图像中的食品区域,并提出了一种双流 CNN 模型识别食品种类。

#### 2.1.2 面向食材的识别

除了基于食品类别的识别,有一些更细粒度的识别工作,即食材识别<sup>[11-12,65-69]</sup>。例如 Bolaños 等

人<sup>[67]</sup>结合 Inception v3 和 Resnet-50 网络进行多标签食材预测。Zhang 等人<sup>[11]</sup>通过嵌入食材结构, 设计了一个多任务学习框架实现多标签食材识别。Zhang 等人<sup>[68]</sup>提出了一种多任务学习方法, 该方法可以使用 CNN 从菜品图像中同时识别菜肴类型、食材和烹饪方式。Zhou 等人<sup>[12]</sup>提出使用二分图 (Bipartite-graph) 结构建模丰富的食材和标签之间的关系, 然后将二分图和 CNN 结合, 实现多标签食材识别和菜品识别。Pan 等人<sup>[69]</sup>组合不同类型的 CNN 进行食材识别。以上工作均属于针对菜品的多标签食材识别。近来一些工作则直接针对原始的食材进行识别以实现其自动结算服务。例如 Xiao 等人<sup>[21]</sup>针对不同场景下训练数据不匹配的问题, 通过迁移学习并采用 CNN 实现原材料的识别和自动结算。

### 2.1.3 面向果蔬的识别

水果和蔬菜是一类特殊的食品对象, 其视觉外观与其他的烹饪菜品有较大的差异, 与其他食品相比其类内差异性和类间相似性更为明显。果蔬识别由于在食品工业、农业和零售业等领域的重要应用价值得到了广泛关注, 这些研究大多集中在农业和健康等相关领域<sup>[27,31,36,70-78]</sup>。例如 Ruud 等人<sup>[31]</sup>提出了一种农产品自动识别系统 Veggie Vision, 该系统融合颜色和纹理等特征预测果蔬类别。Shiv 等人<sup>[70]</sup>提出了一种果蔬识别框架。该框架包括图像分割、特征提取及分类器训练实现果蔬识别。一些工作<sup>[27,75]</sup>针对超市应用场景提出了一种果蔬识别方法, 该方法融合了不同类型的视觉特征, 并结合不同的分类器进行识别。Susovan 等人<sup>[72]</sup>提出了一种水果采摘环境下的水果分类和收获方法。近年来, CNN 被广泛应用于果蔬识别任务中<sup>[20,25-26,79-84]</sup>。例如 Katarzyna 等人<sup>[81]</sup>提出了两个相同架构的 9 层 CNN 以进行零售水果分类。两个 CNN 分别使用背景图像和包含单个水果的图像进行水果图像分类。Steinbrener 等人<sup>[26]</sup>使用改良的 GoogLeNet 对水果和蔬菜进行分类。Patino 等人<sup>[82]</sup>提出了一种改进型的 Fruit-AlexNet 模型用于热带水果分类。Horea 等人<sup>[85]</sup>提出了一个种类最为多样的水果数据集 Fruits-360, 并评估了常见 CNN 在该数据集上的水果识别性能。Hou 等人<sup>[13]</sup>提出了一个用于细粒度图像分类的水果蔬菜数据集 VegFru。他们还提出了一个果蔬识别的框架 HybridNet, 该框架由两个 CNN 组成, 能够分别处理分层标签以提取多个粒度特征, 然后通过双线性池化等方式融合这些特征以得到

最终的特征表示。马<sup>[86]</sup>提出了一个基于安卓手机的移动端果蔬检测与分类系统, 并提出了一种基于改进的 Faster R-CNN 的果蔬分类方法。

此外, 还有一些研究工作主要对果蔬的成熟状态进行识别<sup>[87-89]</sup>。例如 Wajid 等人<sup>[87]</sup>提出了一种快速区分橙子状态(成熟、未成熟、结垢或腐烂)的方法。Wan 等人<sup>[88]</sup>提出了一种识别番茄成熟度(绿色、橙色和红色)的方法。

### 2.1.4 面向其它食品对象的识别

除了对食品本身的对象识别, 近期还有一些针对食品相关对象的识别工作, 如对食品的烹饪类型(如印度菜肴和意大利菜肴等)、食物状态(如块状、片状和糊状等)和饮食场景(如快餐店和咖啡厅等)的识别。例如 Zhang 等人<sup>[90]</sup>将食材信息作为属性信息进行菜品图像的烹饪类型分类。Su 等人<sup>[91]</sup>将食材视为特征, 并构建不同的分类器来预测食谱的烹饪类型。Min 等人<sup>[92]</sup>使用多模态深度玻尔兹曼机融合食材和图像信息进行食谱(Recipe)分类。Druck 等人<sup>[93]</sup>使用了食谱中的类别、配料以及做法, 来预测食谱属性(例如口味和烹饪类型)。Ciocca 等人<sup>[94]</sup>围绕食物的不同状态(如块状、片状和糊状等)利用 CNN 进行识别。由于用户饮食场景的识别对发现用户饮食习惯的重要性, 近来一些工作<sup>[23], [95]</sup>采集第一视角饮食场景数据然后通过 CNN 实现对用户饮食场景的识别。Bolaños 等人<sup>[96]</sup>则研究基于食品图像的餐厅菜单识别问题。

### 2.1.5 小结

食品识别对象丰富多样, 包括菜品、食材、果蔬和其他食品相关的对象。由于菜品图像识别在智慧健康、智慧餐饮、智能零售及智能家居等多个方面的应用前景而受到了广泛关注。然而菜品本身的非刚性特点及大规模菜品数据集的缺失使得当前的菜品识别性能还未能达到最优。在菜品图像类别识别的基础上, 为了进一步深入分析菜品的营养成分或者热量, 菜品的食材识别是必不可少的一步。由于菜品的食材识别和种类识别任务的相关性, 通常采用多任务学习方法同时学习和预测。然而由于很多食材在菜品图像中的不可见性, 现有的食材识别主要针对视觉可见的食材进行识别。除了菜品, 果蔬由于在人类日常饮食中的重要地位及在农业(如自动采摘)和食品工业(如果蔬的自动结算)等领域的重要应用价值使得果蔬识别的研究也较为广泛。果蔬识别相比于菜品识别, 其本身的挑战性有所降低, 但更进一步精细化的识别如果蔬的成

熟度或者质量预测则变得较为困难, 原因在于普通的 RGB 视觉信息只能捕获果蔬的表观信息而无法得到其内部信息。在这种情况下, 基于高光谱和多光谱的果蔬识别提供了一种可行的方法<sup>[26]</sup>。除了基本的菜品、果蔬和食材识别, 最近出现了一些和食品识别相关的新工作, 比如饮食场景识别和食物状态识别等。基于食物对人类的基础性地位, 随着食品数据的不断丰富和识别技术的不断成熟, 未来新的食品相关对象识别和分析的研究将会不断涌现。

## 2.2 特征类型

判别性视觉特征的提取是食品图像识别的关键。按照特征类型可以将食品图像识别分为 (1) 基于手工特征的食品图像识别; (2) 基于深度特征的食品图像识别; 和 (3) 融入上下文信息和外部知识的食品图像识别。

### 2.2.1 基于手工特征的食品图像识别

传统的食品图像识别包括两个步骤: (1) 食品图像特征提取和 (2) 分类模型训练。其中图像特征提取和选择是食品图像识别的关键。在本小节中, 这里的特征主要指的是手工特征, 包括从简单的颜色、纹理、形状等特征到 SIFT<sup>[97]</sup>和 HoG<sup>[98]</sup>等。基于手工特征的食品图像识别方法可进一步分成两种, 一种是仅仅使用单一手工特征来进行识别, 另一种则是融合不同手工特征进行识别。

对基于单一手工特征的食品图像识别, 颜色和纹理等视觉特征被应用到食品图像识别。早期的工作<sup>[30,99]</sup>采用一些简单的图像处理方法提取颜色、纹理或形状等特征进行食品图像识别。Zhu 等人<sup>[100]</sup>开发了一种移动菜品图像识别方法。他们首先通过图像分割方法定位图像中食物所在区域, 然后提取并融合区域的颜色和纹理特征进行菜品图像识别。Giovanni 等人<sup>[101]</sup>将词袋模型 (BOW) 和纹理特征相结合得到 Bag of Textons, 通过菜品图像的 Bag of Textons 的视觉词分布来表示图像的视觉特征。Oo 等人<sup>[102]</sup>提取形状等特征实现水果识别。相比于颜色和纹理等特征, SIFT 描述子具有尺度和旋转不变性等优点, 可以有效应对食品比如菜品图像的形变特性, 因而被广泛应用到菜品图像识别任务中, 例如 Wu 等人<sup>[48]</sup>提取菜品图像的 SIFT 描述子, 再将特征点用矢量表示成词袋 (BOW) 作为视觉特征进行菜品图像识别。Kong 等人<sup>[103]</sup>提出了一个菜品识别系统 DietCam, 该系统考虑到光照、尺度和仿射不变性等特点提取 SIFT 特征作为菜品图像特征, 但是考虑到遮挡等问题, 该系统一次从三种不同拍

摄角度得到三张菜品图像, 然后基于这三张菜品图像进行更为鲁棒的识别。Anthimopoulos 等人<sup>[49]</sup>在自己收集的菜品图像数据集上对颜色、纹理、SIFT 和其变种等不同类型的特征分别进行了性能评估。

考虑到不同手工特征之间的互补性, 大多数方法融合不同类型的手工特征来改进食品图像识别性能。各种集成学习方法被广泛应用于不同特征之间的融合。例如一些工作<sup>[104-105]</sup>通过多核学习融合 SIFT、Gabor 纹理和颜色直方图等各种类型的图像特征来进行食品图像识别。Nguyen 等人<sup>[106]</sup>使用 Non-Redundant Local Binary Pattern (NRLBP) 提取菜品的局部表观信息, 采用提出的形状描述子提取全局结构信息, 最后融合这两种特征进行食品图像识别。Rocha 等人<sup>[27]</sup>从果蔬的图像中提取各种特征描述子包括颜色和边缘等, 通过简单的串接进行特征融合, 然后采用多分类器融合的方法实现最终的果蔬识别。Oliveira 等人<sup>[107]</sup>则提出了一种移动菜品识别系统, 该系统通过融合颜色和纹理等特征进行食品图像分割, 然后通过提出的多排序框架选出最具有判别性的分割区域作为其视觉特征。Martinel 等人<sup>[108-109]</sup>提出基于委员会的学习 (Committee-based Learning) 方法从大量提取的手工特征 (GIST、LBP、BGP、LCP 和 HOG 等) 中选择最佳的特征子集作为最终特征用于菜品图像识别。Ravi 等人<sup>[110]</sup>提出了一个面向能量消耗估算的实时菜品识别平台。该系统提取图像的颜色、纹理 LBP 和 HOG 三种类型的特征, 并基于不同类型特征的组合实现特征层次构造, 最后通过基于 Fisher Vector 的层次特征实现食品图像识别。类似的, Kawano 等人<sup>[111]</sup>提出了一种移动菜品识别系统 FoodCam。实验表明结合 HOG 描述子和颜色描述子并通过 Fisher vector 进行视觉特征表示可以取得较好的识别性能。

针对菜品图像, 考虑到菜品均是由食材组成, 因此捕捉这种部件或者属性区域及他们之间的空间关系对菜品图像识别有很大的改进。Yang 等人<sup>[35]</sup>将食品纹理直方图扩展到成对纹理分布用于挖掘原材料之间的空间关系, 此类方法只适用于标准化菜品。Bossard 等人<sup>[15]</sup>采用随机森林 (Random Forest) 挖掘所有类共享的判别性区域并提取区域的菜品图像特征。He 等人<sup>[112]</sup>提出了一种自动的菜品图像识别方法 DietCam, 该方法首先通过融合纹理特征和可变形的组件模型 DPM (Deformable Part Model) 进行菜品成分检测, 然后基于菜品成分的

直方图分布通过多视角多核 SVM 实现菜品图像分类。

### 2.2.2 基于深度特征的食品图像识别

CNN 能够逐层学习图像特征, 其底层是具有普遍性的(General)特征, 如图像的边缘和纹理等; 高层特征是低层特征的组合, 是针对特定任务有针对性的(Specific)特征<sup>[113]</sup>。逐层特征学习模拟了人脑分层处理信息机制, 能够直接从原始像素自动学习视觉特征。由于 CNN 强大的表达能力, 很快也被应用到食品图像识别领域中, 具体的使用方法主要包含以下几种:

(1)在预训练好的网络上直接提取特征 该类方法中, 预训练好的 CNN 模型可以直接用来当特征提取器, 提取的特征之后可用做分类器的训练。此类研究工作主要使用不同类型的预训练 CNN 网络 (如 AlexNet<sup>[46]</sup>, VGG<sup>[84]</sup>, ResNet<sup>[114]</sup>, GoogleNet<sup>[115]</sup>, DenseNet<sup>[116]</sup> 和 SENet<sup>[117]</sup>) 直接提取视觉特征以进行食品图像识别。由于食品图像识别属于物体识别, 所以所用的预训练 CNN 均是在 ImageNet 数据集上训练好的模型。例如 Ming 等人<sup>[52]</sup>使用 ResNet 网络直接提取视觉特征进行菜品图像识别。Tanno 等人<sup>[118]</sup>从 Network-In-Networks 深度网络提取视觉特征。Pouladzadeh 等人<sup>[119]</sup>提出了一种移动菜品识别系统。在训练阶段, 该系统首先使用区域建议算法 (Region Proposal Algorithms) 生成候选区域, 并通过 CNN 提取所有区域的特征。然后使用提出的子模优化方法 (Submodular Optimization Method), 利用最大覆盖度进行区域挖掘, 为每个菜品类别选择正区域。McAllister 等人<sup>[120]</sup>直接从预训练的 ResNet-152 和 GoogleNet 网络提取视觉特征, 使用融合的深度特征来训练分类器进行菜品图像识别。

(2)在食品图像数据集上对现有深度网络的“精细化”调整 (Fine-tuning) 在特定大规模数据集上训练好的模型有很强的泛化性能, 但是 Fine-tuning 能够进一步提升分类性能。Fine-tuning 是在目标菜品图像数据集上重新调整网络参数, 从而使深度模型能够捕获针对目标任务更具有区分性的特征<sup>[121]</sup>。例如最早将深度学习网络应用到菜品图像识别的工作<sup>[50]</sup>通过微调 AlexNet 网络来提取图像特征。Yanai 等人<sup>[38]</sup>采用了相似的方法。Zhu 等人<sup>[25]</sup>通过微调 Alexnet 网络提取视觉特征实现果蔬识别。Zeng 等人<sup>[83]</sup>则通过微调 VGG 网络提取视觉特征进行果蔬识别。Hassannejad 等人<sup>[51]</sup>微调 Inception V3 网络来进行识别。一些研究工作进一步通过微

调和融合来自不同神经网络的特征来提高识别性能。Klasson 等人<sup>[29]</sup>则在零售食品数据集上比较了直接使用和微调现有的 CNN 网络预测其零售食品类别。例如 Pandey 等人<sup>[122]</sup>首先微调 AlexNet, GoogLeNet 和 ResNet 等三种 CNN 网络, 然后基于微调的网络提取和融合来自不同网络的视觉特征通过某种集成学习方法实现菜品图像识别。相比于直接从预训练的神经网络中提取特征, 通过微调 CNN 能从模型中提取针对目标任务更区分性的视觉特征, 因而进一步改进了食品图像识别的性能。

(3)重新设计针对食品图像识别任务的深度神经网络 以上方法均直接使用现有的 CNN 网络来提取食品图像特征, 未能针对食品图像的特点设计针对性的 CNN 网络结构, 因此识别性能未能达到最优。最近一些研究开始着手设计针对食品图像的专用网络, 例如, Yang 等人<sup>[53]</sup>根据菜品图像的类别特性, 提出了一个 CNN 分层体系结构来进行菜品图像识别。Metwalli 等人<sup>[123]</sup>则基于 DenseNet 网络提出了一个专注于菜品图像的 DenseFood 模型。该模型在训练过程中结合 Softmax 损失函数和中心损失函数联合优化整个网络。Mejia 等人<sup>[124]</sup>简化了 Mask R-CNN 网络<sup>[125]</sup>, 使得它仅专注于菜品图像识别任务。相比于原始网络, 改进的 Mask R-CNN 不仅具有更高的识别性能, 而且更易于训练, 计算量更少。Martinel 等人<sup>[54]</sup>提出了一种 WISER 网络, 通过融合来自 WRN 网络 (Wide Residual Networks)<sup>[126]</sup> 和文中所提出的切片网络 (Slice Network) 的视觉特征来进行识别。切片网络用于捕获特定的菜品垂直结构。通过融合这两个分支的视觉特征, 该方法在当时多个食品图像数据集上达到了当时最好性能。Zhao 等人<sup>[127]</sup>提出一种联合学习蒸馏网络方法用于移动端的菜品识别。该网络能够利用两个网络中不同的中间层来同时训练大型教师网络和小型学生网络。它能够从大型教师网络中学习, 同时保持小型学生网络规模来实现小型学生网络的高食品识别准确率。该模型还引入一种新的实例激活学习, 能够在每个训练样本的实例级激活图上共同训练学生和教师模型。Phiphiphatphaisit 等人<sup>[128]</sup>改进 MobileNets<sup>[129]</sup>, 包括增加了全局池化层、批量归一化层和 Relu 层。结合数据增强技术, 在菜品识别性能上有了明显提高。

还有一些研究工作将食品图像识别看做细粒度视觉识别, 设计面向细粒度图像识别的深度学习网络, 主要思路是从食品图像中找到判别性的区

域, 然后融合这些区域特征进行识别。例如, 陈<sup>[55]</sup>首先对菜品图像进行餐具分割, 排除背景图像干扰, 得到菜品图像最显著性区域, 然后提取区域的深度特征进行识别。基于此项研究, 陈<sup>[55]</sup>又提出了一个基于注意力模型的菜品图像识别算法。通过该注意力模型, 可以得到菜品图像的注意力特征图, 然后利用注意力特征图对原特征图进行改进, 以找到显著性区域, 最后融合显著性区域特征作为最终的特征表示进行识别。梁等人<sup>[56]</sup>提出一种方法自动定位菜品图像显著性区域并从中提取局部特征, 然后通过融合菜品图像的全局与局部特征来实现菜品图像识别。针对菜品图像尺度变化大的特点, 它还引入了特征金字塔结构, 提高了模型对目标尺度的鲁棒性。Qiu 等人<sup>[10]</sup>提出了一种新颖的菜品图像识别网络 PAR-Net, 该网络不仅能以弱监督的方式充分挖掘食品图像显著性区域, 同时可以通过“对抗擦除”的方式提供互补信息。Wu 等人<sup>[130]</sup>为了解决全连接层对输入尺寸的限制, 提出了多尺度采样模块, 对图像进行多个尺度的采样。并进一步提出了基于注意力机制的双线性网络, 从通道和空间两个方向构建注意力网络。此外为了降低过拟合, 还在带权重的交叉熵损失函数中加入正则化机制。而 Liao 等人<sup>[131]</sup>在多分类损失函数的基础上, 针对菜品图像类间相似性大的问题, 提出了最大类间距损失函数, 以增大相似类之间的距离, 实现相似类的区分。上述研究工作主要通过多尺度和注意力两个角度去寻找食品图像最显著性区域, 通过融合显著性区域特征来提高识别性能。

此外, 有些研究人员还尝试融合手工特征和深度特征进行食品图像识别。例如 Kawano 等人<sup>[132]</sup>同时提取 HoG 和颜色手工特征及 AlexNet 网络的深度视觉特征, 最后融合这几种不同类型的特征作为最终的视觉表示。

### 2.2.3 融入上下文和外部知识的食品图像识别

随着菜谱分享网站和其他各种社交网络的发展, 用户上传和分享了大量的食品多模态数据, 除了食品图片, 还包含丰富的上下文 (Context) 信息 (如位置和时间信息等) 及食品相关的知识 (如食谱信息、食材、风味和营养信息等), 这些上下文信息和外部知识提供了额外的信息源, 充分利用这些信息有助于提高食品图像识别方法的性能<sup>[133]</sup>。当前食品图像识别常用的上下文信息为位置信息, 而最常用的外部知识为食材信息。

Bettadapura 等人<sup>[42]</sup>提出了一种利用位置信息

和各种手工特征来自动识别餐厅菜品的方法。Xu 等人<sup>[40]</sup>提出了在地理区域中实现特定餐厅的菜品图像识别的框架, 并引入了地理区域模型的概念, 其中利用了 DeCAF 的深度特征和餐厅位置信息。Herranz 等人<sup>[16]</sup>提出了一个概率图模型来融合菜品、餐厅和位置信息进行菜品图像识别。Wang 等人<sup>[134]</sup>认为菜品识别和餐厅识别任务是密切关联的, 提出了一种多任务 CNN 从菜品图像中同时预测菜品和餐厅类别。Bolaños 等人<sup>[96]</sup>将 CNN 与循环神经网络 (Recurrent Neural Network, RNN) 结合起来通过菜品和关联的菜单信息识别对应餐厅的菜品。Wei 等人<sup>[135]</sup>提出了一个菜品识别系统, 该系统在在线阶段, 使用检索算法, 将查询的菜品图像及其 GPS 与预定距离内的候选菜品图像库进行匹配。

另外, 还有一些工作融合了食品的食材和菜谱等外部知识。例如 Zhou 等人<sup>[12]</sup>利用二分图建模菜品类别和食材的关系, 并结合深度学习模型通过多任务学习的方式实现菜品识别。Jiang 等人<sup>[136]</sup>认为食材信息可以从另外一个视角来描述菜品图像。因此, 他们提出一个多尺度多视角模型 MSMVFA 进行菜品图像识别。该模型分别训练了一个分类网络和食材属性网络, 提取这两个网络的视觉特征, 最后通过多尺度多视角融合的方式得到菜品最终的特征表示。Min 等人<sup>[18]</sup>则提出一个食材信息指导的多注意力网络 IG-CMAN 实现菜品图像识别。该模型能够从粗粒度到细粒度、基于类别信息和食材信息从菜品图像中找到多个区域, 最后融合这些注意力区域特征, 经过 Softmax 分类器进行识别。

### 2.2.4 小结

早期的食品图像识别通常包括手工特征提取和分类器训练两个阶段。而考虑到不同手工特征之间的互补性, 通常提取多种类型的手工特征, 然后采用不同的集成学习方法融合不同类型的特征以改进食品图像识别的性能。在所有的手工特征中, SIFT 描述子由于其尺度和旋转不变性、对形变不敏感等特点, 和词袋模型结合在一起被广泛应用到食品图像识别中, 尤其在菜品图像识别任务中。自从 2012 年 Alex Krizhevsk 等人在 ImageNet 上训练一个 8 层的深度模型 AlexNet 并在 ImageNet 竞赛上取得非常好的效果后, CNN 在图像识别领域受到了广泛关注。深度网络学习的特征由于其强大的表达能力随后在 2014 年被引入到食品图像识别领域。基于深度学习的食品图像识别方法从最初简单的使用预训练或者微调的模型提取视觉特征逐渐转向

针对食品图像识别任务的神经网络设计。由于 CNN 可以自动从食品图像中学习特征因而基于深度特征的食品图像识别可以将特征学习和分类器训练同时放到一个框架里面进行端到端的学习。而食品相关的上下文信息和外部知识的引入可进一步改进食品图像识别的性能。未来随着大规模食品图像数据库资源的发展,深入挖掘食品图像的特性设计针对食品图像识别任务的深度学习网络将是食品图像识别未来发展的重要方向。此外,将各种丰富的上下文信息和外部知识有效组织成食品知识图谱,将食品知识图谱融入到食品图像特征学习的过程中将对改进食品图像识别性能有重要作用<sup>[137]</sup>。

### 2.3 学习方法

在常规食品图像分类任务中,为了保证训练得到的分类模型具有高准确性和高可靠性,需要满足

(1) 训练样本和测试样本服从独立同分布和 (2) 训练样本足够多。然而在现实世界中,总有一些食品类别的样本很难收集<sup>[44]</sup>,因而这些类的样本数量有限;此外现实世界的数据是开放的,新类别和原有类别的新样本是动态增加的<sup>[138]</sup>,食品图像也不例外。因此为了解决以上问题,需要放松常规食品图像识别方法的假设,于是就产生了基于迁移学习、小样本学习及增量学习的食品图像识别研究。因此,从机器学习方法维度出发,食品图像识别可以分为 (1) 常规食品图像识别; (2) 基于迁移学习的食品图像识别; (3) 基于小样本学习的食品图像识别和 (4) 基于增量学习的食品图像识别。

#### 2.3.1 常规食品图像识别

大部分食品图像识别的研究工作都是面向封闭数据集,且满足 (1) 训练样本和测试样本服从独立同分布和 (2) 训练样本足够多。在这种设置下,食品图像识别通常的流程包括特征提取和分类器训练,早期基于手工特征的食品图像识别方法<sup>[48-49]</sup>基本上都属于这类方法。而基于深度学习的食品图像识别由于大规模食品图像数据集的缺失,很多情况下采用了迁移学习方法以充分利用额外的大规模数据集。

#### 2.3.2 基于迁移学习的食品图像识别

深度学习已成为食品图像识别的主流方法。由于当前的食品图像识别缺乏公开的大规模食品图像基准数据集,为了有效训练 CNN,需要采用迁移学习方法以充分利用额外的大规模数据集。在 ImageNet 上训练的 CNN 已被广泛应用于迁移学习,主要包括如下两种方法:直接使用预训练的网络作

为特征提取器和采用目标数据集对网络进行微调。近期研究工作<sup>[38,50-53]</sup>直接使用不同的 CNN 进行食品图像识别。相比于简单将预训练的模型作为特征提取器,微调是在目标食品图像数据集上重新调整网络参数,从而使 CNN 能够捕获更具有区分性的特征<sup>[121]</sup>。基于网络微调的方法由于其识别性能的改进更常用于食品图像识别任务中<sup>[38,50-51]</sup>。近来一些工作<sup>[139-140]</sup>深入探讨了迁移学习从大规模非食品数据集向小规模食品图像数据集迁移所产生的性能影响。

#### 2.3.3 基于小样本学习的食品图像识别

常规食品图像识别方法尤其是基于深度学习的方法往往需要足够多的训练样本才能得到好的训练模型,但是对某些食品类别,从现实中收集大量的训练数据十分困难,因而对于只有少量样本的食品类别,模型的识别准确率较低。为了解决该问题,近来一些研究者将小样本学习应用到食品图像识别任务中<sup>[44,141-143]</sup>。吕等人<sup>[141]</sup>第一次提出面向小样本学习的食品图像识别技术,通过融合三元卷积神经网络与关系网络<sup>[144]</sup>进行小样本食品图像学习。Zhang 等人<sup>[142]</sup>则基于二阶池化的特性提出了一种多尺度关系网络学习小样本设置下的图像关系。Song 等人<sup>[143]</sup>提出一个注意力模型用于食品图像的小样本学习。该模型首先通过 CNN 捕获图像特征。然后基于注意力机制产生图像注意力权重,并且作用于食品图像特征图获得重要信息。Jiang 等人<sup>[44]</sup>通过引入了原材料信息提出了多视角食品小样本学习模型。该方法提取类别特征和食材特征后,经过多视角特征融合和关系网络实现小样本食品图像识别。

#### 2.3.4 基于增量学习的食品图像识别

在现实开放场景下,每个类别中的样本数量持续增加,并且也会出现新的类别。为此一些研究者开始研究将增量学习应用到食品图像识别任务中。Tahir 等人<sup>[145]</sup>提出了一种新的开放式增量学习框架,该框架使用基于深度模型的迁移学习技术和 Relief F 技术分别进行特征提取和特征选择,并使用新型自适应精简类增量式核极限学习机以进行分类。Horiguchi 等人<sup>[146]</sup>则针对每个用户提出了个性化增量学习框架,该框架是基于深度特征的最近类别均值分类器和 1-最近邻分类器的组合,以解决个性化的食品分类问题。He 等人<sup>[147]</sup>提出了一个增量学习框架,该框架使用改进的交叉蒸馏损失函数进行增量学习以实现食品图像识别。

### 2.3.5 小结

当前大部分食品图像识别仍然是在传统分类范式下开展研究的。然而由于大规模食品图像基准数据库的缺失,一些食品类别样本的稀少、开放世界中食品图像类别和实例动态增加等问题,一些研究者开始探索将迁移学习、小样本学习和增量学习等应用到食品图像识别任务中。但目前基于小样本学习和增量学习的食品图像识别处于初步探索阶

段,性能也不够高。随着机器学习社区小样本学习和增量学习方法的不断发展和食品图像本身特点的深入挖掘,基于小样本学习和增量学习的食品图像识别将会得到更多的关注。

表 1 进一步从特征类型、学习方法和识别对象三个维度对代表性的食品图像识别方法给予了总结。

表 1 食品识别方法总结

算法	识别对象	特征类型	学习方法	发表年份
Bolle 等人 <sup>[31]</sup>	果蔬	图像特征(Histogram、Color、Texture)	常规食品识别	1996
Wu 等人 <sup>[48]</sup>	菜品	图像特征(SIFT)	常规食品识别	2009
Joutou 等人 <sup>[104]</sup>	菜品	图像特征(BoF、Color Histogram、Gabor Texture Features)	常规食品识别	2009
Yang 等人 <sup>[35]</sup>	菜品	图像特征(PFD、多维直方图)	常规食品识别	2010
Rocha 等人 <sup>[27]</sup>	果蔬	图像特征(BIC、CCV、UNSER)	常规食品识别	2010
Zhu 等人 <sup>[100]</sup>	菜品	图像特征(Gabor、Color Moment)	常规食品识别	2011
Faria 等人 <sup>[75]</sup>	果蔬	图像特征(BIC、CCV、LAS、EOAC)	常规食品识别	2012
Kawano 等人 <sup>[132]</sup>	菜品	图像特征(CNN)	常规食品识别	2014
Zhang 等人 <sup>[74]</sup>	果蔬	图像特征(Color Histogram、Texture、FNN)	常规食品识别	2014
Su 等人 <sup>[91]</sup>	食品风味	食材特征	常规食品识别	2014
Martinel 等人 <sup>[108]</sup>	菜品	图像特征(Color Histogram、PHOG、LBP、GIST 等)	常规食品识别	2015
Yanai 等人 <sup>[38]</sup>	菜品	图像特征(CNN)	常规食品识别	2015
Xu 等人 <sup>[40]</sup>	菜品	图像特征(CNN)、地理位置	常规食品识别	2015
Wang 等人 <sup>[76]</sup>	果蔬	图像特征(小波熵、PCA、FNN)	常规食品识别	2015
Ravi 等人 <sup>[110]</sup>	菜品	图像特征(HoG、LBP、color features)	常规食品识别	2015
Kawano 等人 <sup>[111]</sup>	菜品	图像特征(color histogram、BoF、HoG、Color Fisher Vector)	常规食品识别	2015
Martinel 等人 <sup>[54]</sup>	菜品	图像特征(CNN)	常规食品识别	2016
Herranz 等人 <sup>[16]</sup>	菜品、餐厅	图像特征(CNN)、地理位置	常规食品识别	2016
Bolaños 等人 <sup>[59]</sup>	菜品	图像特征(CNN)	常规食品识别	2016
He 等人 <sup>[147]</sup>	菜品	图像特征(CNN)	增量学习的食品识别	2016
Pandey 等人 <sup>[122]</sup>	菜品	图像特征(CNN)	常规食品识别	2017
Aguilar 等人 <sup>[60]</sup>	菜品	图像特征(CNN)	常规食品识别	2017
Pan 等人 <sup>[66]</sup>	食材	图像特征(CNN)	常规食品识别	2017
Bolaños 等人 <sup>[67]</sup>	食材	图像特征(CNN)	常规食品识别	2017
Zeng 等人 <sup>[83]</sup>	果蔬	图像特征(CNN)	常规食品识别	2017
Hou 等人 <sup>[13]</sup>	果蔬	图像特征(CNN)	常规食品识别	2017
Pouladzadeh 等人 <sup>[119]</sup>	菜品	图像特征(CNN)	常规食品识别	2017
Liu 等人 <sup>[175]</sup>	菜品	图像特征(texture features)	常规食品识别	2017
Chen 等人 <sup>[65]</sup>	食材	图像特征(CNN)	常规食品识别	2018
Wan 等人 <sup>[88]</sup>	果蔬成熟度	图像特征(特征色值、BPNN)	常规食品识别	2018

Qiu 等人 <sup>[10]</sup>	菜品	图像特征(CNN)	常规食品识别	2019
Liang 等人 <sup>[56]</sup>	菜品	图像特征(CNN)	常规食品识别	2019
Wei 等人 <sup>[135]</sup>	菜品	图像特征(CNN)、地理位置	常规食品识别	2019
Jiang 等人 <sup>[136]</sup>	菜品	图像特征(CNN)、食材特征	常规食品识别	2019
Min 等人 <sup>[18]</sup>	菜品	图像特征(CNN)、食材特征	常规食品识别	2019
Steinbrener 等人 <sup>[26]</sup>	果蔬	高光谱图像特征(CNN)	常规食品识别	2019
Lv 等人 <sup>[141]</sup>	菜品	图像特征(CNN)	小样本食品识别	2019
Pan 等人 <sup>[69]</sup>	食材	图像特征(CNN)	常规食品识别	2020
Zhao 等人 <sup>[127]</sup>	菜品	图像特征(CNN)	常规食品识别	2020
Jiang 等人 <sup>[44]</sup>	菜品	图像特征(CNN)、食材信息	小样本食品识别	2020
Tahir 等人 <sup>[145]</sup>	菜品	图像特征(CNN)	增量学习的食品识别	2020

### 3 数据库与性能评估

#### 3.1 数据库

在食品图像识别算法快速发展的同时, 用于评估识别算法的数据库也在日益完善。我们将在本节中对食品图像识别的主流数据库进行概述, 从中也可以一窥食品图像识别的发展。在介绍不同数据库时, 将主要从数据集规模、所属菜系及包含的上下文信息和多模态信息等不同方面进行阐述, 表 2 全面总结了食品图像识别领域的数据库。

2009 年 Chen 等人<sup>[34]</sup>在计算机视觉领域发布了第一个快餐食品数据库 PFID 用来评估菜品识别方法。该数据集由来自 11 个快餐厅的 101 类西方菜品组成, 共有 4545 张菜品图片。此外该数据集还提供了 606 对立体图像、303 个 360 度运动结构视频和 27 个志愿者提供的饮食活动 (eating events) 视频。2012 年来自日本电气通信大学的研究团队<sup>[57,111]</sup>发布了食品数据集 UEC Food100。该数据集包括 100 类日本菜品和 12905 菜品图片。随后该研究团队对该数据集进行扩充, 发布了 UEC Food256<sup>[148]</sup>, 该数据集仍然全部由日本菜品构成, 菜品种类从原来的 100 类扩充为 256 类, 其相应的菜品图片数量增加到 25088 张, Kawano 等人<sup>[149]</sup>提供了其数据库扩充的详细过程。相比于 PFID, UEC, Food256 无论从菜品类别还是每类菜品的图片数量都有了较大幅度的增加。由于数据规模相对较大, 经常被用作深度学习领域菜品图像识别的评测数据集。

2014 年 Bossard 等人<sup>[15]</sup>发布了第一个大规模的西方菜品图像数据库 ETHZ Food-101, 包含 101 类西餐和 101000 张菜品图像。每类包含 1000 张图像,

其中包括 750 张训练图像和 250 张测试图像。Bolaños 等人<sup>[67]</sup>进一步为每类菜品提供了相应的食材信息。由于 ETHZ Food-101 每一类的图片数量都较大, 非常适合复杂模型特别是深度学习模型的训练, 因而该数据集已被广泛用于检验菜品图像识别算法性能尤其是基于深度学习的识别方法的基准。截止 2014 年, 发布的数据集大部分为西方菜品数据库, 很少有东方菜品数据库, 尤其是中餐数据库。2016 年发布了第一个较大规模的中餐菜品数据库 Vireo Food-172<sup>[41]</sup>。该数据集有 172 个类, 共 110,241 张食品图像。随后 Chen 等人<sup>[37]</sup>发布了更大规模的中餐菜品数据库 ChineseFoodNet, 该数据库包括 208 类中餐菜品和 192000 张图像。近来, 中科院计算所团队于 2019 年发布了 ISIA Food-200 数据集<sup>[18]</sup>, 该数据集从 Wikipedia 中获取食品的类别列表, 然后使用该类别名作为查询从多个搜索引擎(例如 Google 和 Bing)中检索食品图像。经过人工筛选和预处理, 最终该数据集包括 200 类菜品种类, 319 种食材, 约 200000 张菜品图片, 同时涵盖中西方菜谱。随后该团队进一步将其从 200 类扩展到 500 类, 即 ISIA Food-500<sup>[43]</sup>, 该数据集包含 500 个类, 每个类包含的图像数目都大于 500, 共 399726 张食品图像。该数据库的提出对食品图像识别的算法提出了更大的挑战, 有望催生出优秀的理论与算法, 将食品图像识别的研究推向一个新高度。此外, 由于果蔬识别的研究主要集中在农业等相关领域, 大部分数据集均未公开。公开的几个数据库包括 Fruits360<sup>[85]</sup>、VegFru<sup>[13]</sup>和 FruitVeg-81<sup>[150]</sup>等。

总的来说, (1) 从食品的类型来看, 这些数据库涵盖了来自世界上不同地区的菜品。PFID<sup>[34]</sup>和 ETHZ Food-101<sup>[15]</sup>主要由西方菜系组成, ChineseFoodNet<sup>[37]</sup>和 Vireo Food-172<sup>[41]</sup>主要由中国

菜肴组成, UECFood100<sup>[57]</sup>和 UECFood256<sup>[148]</sup>主要由日本菜组成, ISIA Food-200<sup>[18]</sup>和 ISIA Food-500<sup>[43]</sup>包括中西方菜系。除了菜品, 还有一些果蔬数据集, 如 Fruits360<sup>[85]</sup>主要由各种水果组成, 而 VegFru<sup>[13]</sup>和 FruitVeg-81<sup>[150]</sup>则同时包括水果和蔬菜。最近一些研究学者发布了一些其他类型的食品数据集, 如零售食品数据库 Grocery<sup>[29]</sup>。多样化的食品数据集能够支持各种场景下食品识别方法的研究, 使其能够满足各种现实的需求。图 1 展示了一些代表性食品数据集的样例图像。(2) 从数据库的规模来看, 由于当前主流发布的数据库多集中在菜品数据库, 我们以此为例进行介绍。现有的菜品数据集可以分为小规模、中等规模和大规模菜品数据集。小规模数据集主要指数据量在 10000 以下的数据集, 代表数据集有 PFID<sup>[34]</sup>、和 UNIMIB2015<sup>[151]</sup>等。这些数据集规模较小, 不足以支持复杂模型尤其是深度学习方法的食品识别方法研究。中等规模的数据集主要指数据量在 10,000 到 100,000 的数据集, 代表性的有 UEC Food256<sup>[148]</sup>、UEC Food100<sup>[57]</sup>、Food-975<sup>[12]</sup>、UNICT-FD1200<sup>[32]</sup>等, 这些数据集已初具规模, 能够适应特定场合下菜品识别方法的研究。大规模数据集特指数据量在 100000 以上的数据集, 代表性数据集包括 ETHZ Food-101<sup>[15]</sup>、Vireo Food-172<sup>[41]</sup>和 ISIA Food-500<sup>[43]</sup>等, 其中 ETHZ Food-101<sup>[15]</sup>是第一个用于菜品识别研究的大规模数据集。这些大规模的菜品数据集对菜品识别具有里程碑式的意义, 能够显著地推动菜品识别方法的发展。(3) 从数据的模态来看, 大部分的菜品数据集都由含有类别标签的图像组成, 也有一些数据集包含食材信息, 例如 Vireo Food-172<sup>[41]</sup>和 ISIA Food-200<sup>[18]</sup>。还有一些数据集包含其他模态信息, 如 RGBD-FG<sup>[19]</sup>包含深度信息, 能够支持基于 RGB-D 的食品识别方法研究。充分利用每类菜品所包含的其他上下文信息和属性能够改进菜品识别性能。虽然当前发布了一些多模态菜谱数据集如 Yummly66K<sup>[190]</sup>和 Recipe1M<sup>[191]</sup>, 但是这些数据集主要是面向菜谱总结或检索的任务。构建面向菜品图像识别的大规模多模态菜品数据集将进一步推进该领域的继续发展。在表 2 中, 我们列出了一些食品识别任务中常用的数据集, 包括提出的年份、数据集规模、所属菜系、包含的上下文信息和是否开源等信息。每个数据集的详细信息可参见表格。

### 3.2 性能评估

和其他识别方法类似, 食品图像识别主要采用 Top-1 和 Top-5 分类准确率作为评价指标。Top-1 分类准确率表示测试图像中概率最大的预测标签正确的百分比。Top-5 分类准确率表示真实标签在排名前 5 位的预测标签中的测试图像所占的百分比。另外性能评估包括 1-crop 和 10-crop 两种设置。此外, 近年来也有一些工作对人类在食品识别性能的评估<sup>[46]</sup>进行了探讨。

随着食品数据库的日益丰富, 食品图像识别算法也得到了快速发展, 本文将展示常用的食品识别方法在三个数据库 Vireo Food-172、UEC Food256 和 ETHZ Food-101 上的性能。表 3、表 4 和表 5 分别展示了食品图像识别方法在这三个数据库上的性能。可以看出, 基于卷积神经网络的深度学习方法是食品图像识别方法的主流方法, 在所有数据库上均取得了最佳识别性能: IG-CMAN 在 Vireo Food-172 上取得了最佳识别性能。WISER 在 UEC Food256 取得了最佳的识别性能。EfficientNet 和 GPipe 在 Food-101 上取得了最佳性能。以上几种模型均是基于卷积神经网络的改进模型。以上最优方法中, 在基于深度学习方法的基础上, EfficientNet 通过平衡网络深度、宽度和分辨率等引入了一种新的尺度化(Scaling)方法改进了识别性能。IG-CMAN 充分挖掘食材信息以得到更为判别性的区域, WISER 引入了额外的切片网络用于捕获针对性的视觉特征。

## 4 应用

食品图像识别在现实中具有广阔的应用前景。例如食品图像识别通过自动识别食品类别, 可进一步实现后续的营养分析和卡路里估算。在自助餐厅中, 食品图像识别可以实现自动计费。在生活中, 人们也可以通过简单拍照更好的了解他们不熟悉或以前从未见过的食物, 并了解其细节, 例如烹饪方法、食材、口味和营养成分等。下面介绍几种代表性的食品图像识别应用。

### 4.1 饮食质量评估

饮食质量评估 (Dietary Assessment) 是预防和治疗各种慢性疾病 (如糖尿病) 的重要手段。传统的饮食管理方法主要依赖用户报告数据<sup>[167]</sup>。一方面, 由于主观性和估计偏差, 传统方法的准确性无

表 2 现有食品图像识别数据集

数据集	年份	类别数	图片数	菜系	上下文信息	是否公开
PFID <sup>[34]</sup>	2009	101	4545	西方菜系	视频数据	否
Food50 <sup>[104]</sup>	2010	50	5000	混合	-	否
Food85 <sup>[105]</sup>	2010	85	8500	混合	-	否
UEC Food100 <sup>[57]</sup>	2012	100	14361	日本菜	边界框	是
UEC Food256 <sup>[148]</sup>	2014	256	25088	日本菜	边界框	是
ETHZ Food-101 <sup>[15]</sup>	2014	101	101000	西方菜系	食材信息	是
Diabetes <sup>[49]</sup>	2014	11	4868	混合	-	否
UPMC Food-101 <sup>[152]</sup>	2015	101	90840	西方菜系	文本描述	是
UNIMIB2015 <sup>[151]</sup>	2015	15	2000	混合	-	是
UNICT-FD889 <sup>[153]</sup>	2015	889	3583	混合	-	是
UNIMIB2016 <sup>[22]</sup>	2016	73	1027	混合	-	是
Vireo Food-172 <sup>[41]</sup>	2016	172	110241	中国菜	食材信息	是
Food-975 <sup>[12]</sup>	2016	975	37785	混合	-	否
Food500 <sup>[154]</sup>	2016	508	148408	混合	-	否
Food11 <sup>[155]</sup>	2016	11	16643	混合	-	否
UNICT-FD1200 <sup>[32]</sup>	2016	1200	4754	混合	-	是
Food524DB <sup>[156]</sup>	2017	524	247636	混合	-	是
ChineseFoodNet <sup>[37]</sup>	2017	208	192000	中国菜	-	是
VegFru <sup>[13]</sup>	2017	292	160000	果蔬	超类标签	是
FruitVeg-81 <sup>[150]</sup>	2017	81	15737	果蔬	-	是
ISIA Food-200 <sup>[18]</sup>	2019	200	197323	混合	食材信息	是
Fruits 360 <sup>[85]</sup>	2019	131	90483	果蔬	-	是
Sush-50 <sup>[10]</sup>	2019	50	3963	日本菜	-	是
Grocery <sup>[29]</sup>	2019	33	5125	包装食品	-	是
Refrigerator-Food dataset <sup>[157]</sup>	2019	80	49557	混合	-	否
ISIA Food-500 <sup>[43]</sup>	2020	500	399726	混合	-	是
RGBD-FG <sup>[19]</sup>	2020	50	93051	果蔬	深度信息	否

表 3 不同模型在 Vireo Food-172 上性能比较(%)

方法	设置	Top-1 准确率	Top-5 准确率
AlexNet <sup>[46]</sup>	1-crop	64.91	85.32
VGG-16 <sup>[115]</sup>	1-crop	80.41	94.59
MTDCNN(VGG-16) <sup>[41]</sup>	1-crop	82.06	95.88
DenseNet-161 <sup>[116]</sup>	1-crop	86.93	97.17
MTDCNN(DenseNet-16) <sup>[41]</sup>	1-crop	87.21	97.29
SENet-154 <sup>[117]</sup>	1-crop	88.71	97.74
PAR-Net <sup>[10]</sup>	1-crop	89.6	-
Two-Scale CNN <sup>[166]</sup>	1-crop	89.72	98.40
PAR-Net <sup>[10]</sup>	10-crop	90.2	-
MSMVFA <sup>[136]</sup>	1-crop	90.61	98.31
IG-CMAN <sup>[18]</sup>	1-crop	<b>90.63</b>	<b>98.40</b>

表 4 不同模型在 UEC Food256 上性能比较(%)

方法	设置	Top-1 准确率	Top-5 准确率
DeepFoodCam <sup>[118]</sup>	1-crop	63.77	85.82
DeepFood <sup>[160]</sup>	1-crop	63.8	87.2
DCNN-FOOD <sup>[38]</sup>	1-crop	67.57	88.97
Two-Scale CNN <sup>[166]</sup>	1-crop	71.75	91.49
Inception V3 <sup>[51]</sup>	1-crop	76.17	92.58
ResNet-200 <sup>[114]</sup>	1-crop	79.12	93.00
WRN <sup>[54]</sup>	1-crop	79.76	93.90
WISeR <sup>[54]</sup>	1-crop	<b>83.15</b>	<b>95.45</b>

表 5 不同模型在 ETHZ Food-101 上的性能比较 (%)

方法	设置	Top-1 准确率	Top-5 准确率
SELC <sup>[109]</sup>	1-crop	55.89	-
AlexNet-CNN <sup>[15]</sup>	1-crop	56.40	-
ResNet-152+SVM-R	1-crop	64.98	-
BF <sup>[120]</sup>			
DCNN-FOOD <sup>[38]</sup>	1-crop	70.41	-
LMBM <sup>[158]</sup>	1-crop	72.11	-
Ensemble Net <sup>[122]</sup>	1-crop	72.12	91.61
DeepFOOD <sup>[160]</sup>	1-crop	77.40	93.70
GoogLeNet <sup>[159]</sup>	1-crop	78.11	-
ILSVRC <sup>[161]</sup>	1-crop	79.20	94.11
WARN <sup>[54]</sup>	1-crop	85.50	-
Two-Scale CNN <sup>[166]</sup>	1-crop	86.21	97.19
CNNs Fusion <sup>[162]</sup>	1-crop	86.71	-
Inception V3 <sup>[51]</sup>	1-crop	88.28	96.88
SENet-154 <sup>[117]</sup>	1-crop	88.62	97.57
WRN <sup>[54]</sup>	10-crop	88.72	97.92
PAR-Net <sup>[10]</sup>	1-crop	89.3	-
SOTA <sup>[139]</sup>	1-crop	90.00	-
DLA <sup>[163]</sup>	1-crop	90.00	-
WISER <sup>[54]</sup>	10-crop	90.27	<b>98.71</b>
IG-CMAN <sup>[18]</sup>	1-crop	90.37	98.42
PAR-Net <sup>[10]</sup>	10-crop	90.4	-
Inception-Resnet-v2	1-crop	90.4	-
SE <sup>[140]</sup>			
MSMVFA <sup>[136]</sup>	1-crop	90.59	98.25
多级卷积特征金 字塔 <sup>[56]</sup>	1-crop	91.04	-
GPipe <sup>[164]</sup>	1-crop	<b>93.0</b>	-
EfficientNet-B7 <sup>[165]</sup>	1-crop	<b>93.0</b>	-

法保证, 另一方面, 频繁地报告数据会影响用户体验。随着计算机视觉的发展, 学者们提出被动饮食监测的概念<sup>[168]</sup>, 即依靠摄像机等传感器来记录饮食情况并自动执行食品识别及后续任务。在整个过程中, 食品图像识别是第一步, 也是至关重要的一步, 在饮食管理系统中起着举足轻重的作用。

当前各种便携式拍摄设备(如手机和可穿戴相机等)的普及使得自动高效的多媒体饮食日志记录和饮食质量评估成为可能。Mezgec 等人<sup>[169]</sup>提出了一种饮食管理系统, 该系统可以从单个图像中识别出用户一餐中含有的菜品条目, 然后预测其营养成分及热量等。在整个系统中, 食品识别模块采用基

于 CNN 的多标签分类器来预测一餐中含有的菜品类别, 预测得到的结果通过营养表查找和体积估算等后续操作以预测相应的营养成分含量。近来一些工作<sup>[23,62,170]</sup>采用可穿戴的第一视角相机拍摄饮食场景, 通过菜品图像识别<sup>[62,170]</sup>或者饮食场景识别<sup>[23]</sup>自动识别日常饮食或者饮食环境, 对人的饮食习惯研究及饮食管理具有重要的现实意义。

## 4.2 食品自动结算

食品图像识别在餐厅、超市、农贸市场等各种应用场景中可以实现食品的自动结算, 提高结算效率, 降低运营成本, 改善用户就餐或者购买体验。例如作为无人餐厅的核心技术之一, 食品图像识别在监控食品的消费量、对客户的点餐进行自动计费结账等方面起着重要作用。例如 Aguilar 等人<sup>[60]</sup>提出了一种在食堂和饭店环境下对食物托盘进行自动分析的方法, 能够检测和预测托盘中的菜品类别, 为自助餐厅环境下的消费量监控和自动结账提供了重要的技术支持。Xiao 等人<sup>[21]</sup>通过对食材的识别实现其自动结算服务。

## 4.3 食品组织、检索和推荐

在大数据时代, 各种社交媒体尤其是饮食或食谱相关的网站通过食品图像识别方法可以进一步有效的组织和检索食品图像。大规模的餐饮网站(例如美团和 Yelp)为我们的日常饮食提供了便利。引入食品识别可满足人们了解不熟悉或未见过的食物, 并可进一步实现个性化菜品推荐。Min 等人<sup>[171]</sup>提出了一个系统的食品推荐框架。在整个系统中, 食品图像识别技术对理解用户需求及改进个性化推荐结果有重要影响。以美团为例, 通过菜品图像识别算法识别用户在评论区上传的食品图像的具体类别, 可根据识别结果对用户进行个性化食品推荐。

## 4.4 智能厨具

食品图像识别在智能厨具中也有广泛应用, 最为代表性的是智能冰箱。食品识别是其核心功能之一, 在食品种类识别或者食品新鲜度识别方面发挥重要作用。例如 Zhang 等人<sup>[172]</sup>提出了一种智能冰箱中的水果识别方法, 该方法使用神经网络与多模型融合技术进行水果识别。Zhu 等人<sup>[157]</sup>提出了一种基于有监督 Transformer 网络的食品识别系统, 能够在含有多种食物的冰箱环境下检测和识别食品的类别。Mohammad 等人<sup>[173]</sup>提出使用较少的传感器并结合软件云平台进行冰箱食品监测。整个计算框架中, 食品识别被用于感知后的食品分析过程。该方法基于卷积神经网络和迁移学习技术进行

食品识别。除了智能冰箱外,食品识别还被应用于其他的智能厨具中。例如 Hashimoto 等人<sup>[174]</sup>基于食物追踪和食材识别等设计了一种以人为中心的智慧厨房系统。一些研究工作<sup>[175-176]</sup>则是面向边缘设备的食品图像识别。例如 Liu 等人<sup>[175]</sup>设计了一种基于边缘计算设备的食品识别系统。Liu 等人<sup>[176]</sup>提出了一种新颖的边缘计算系统 EdgeVegfru, 用于果蔬图像识别。该系统能在有限时间和计算资源内表现出出色的识别性能, 可应用于自助超市和零售业等应用场景。

#### 4.5 智能穿戴设备

不健康的饮食习惯是罹患各种慢性疾病的重要危险因素之一。近年来, 能够追踪个人日常食品摄入量的饮食监测应用在技术界引起了广泛的关注。随着各种便携式的智能穿戴设备的兴起, 越来越多的学者开始将这些饮食监测应用部署在智能穿戴设备上。食品图像识别作为饮食监测中一项基础和核心的技术, 其识别准确率对整个系统的性能有显著的影响。Rachakonda 等人<sup>[177]</sup>提出了一个部署在智能眼镜上的饮食监测系统。该系统通过智能眼镜自动地从现实环境中获取食品图像, 然后使用基于卷积神经网络的食品识别方法进行识别, 并进一步进行检测和摄入量估计, 最后对用户进行压力分析以确定用户的饮食状态。最终实现对用户的饮食行为的自动分析以发现生活中的过度饮食和压力饮食等问题。随着智能设备的进一步普及, 饮食检测系统的应用将越来越广泛, 对嵌入式的高性能食品识别方法的研究也将成为学者们的关注热点。

#### 4.6 餐饮机器人

随着技术的发展, 智能机器人逐渐被应用于人们的日常生活中, 其中一个重要的应用场景就是日常饮食。例如在无人餐厅中, 传菜机器人和送菜机器人被用于自动向顾客运送餐饮。在智能餐厅中, 炒菜机器人被用于自动烹饪食品。在物联网医疗中, 喂饭机器人能够向肢体残疾的人喂饭。Feng 等人<sup>[178]</sup>为实现机器人辅助喂食, 开发了一个咬合获取框架, 该框架将整个盘子图像作为输入, 分割和识别出菜品, 并应用所提出的 SPANet 模型来选择目标菜品和相应的动作提高辅助喂食成功率。在这些机器人的内部计算框架中, 食品识别模块用于识别经传感器感知后的食品类别, 并将识别结果送入分析模块以获取食品的状态信息。食品识别的性能关系到智能机器人对食品的感知情况和后续决策, 对于智能机器人的发展有重要意义。

#### 4.7 食品安全

食品安全已成为亟待解决的社会问题。当下, 高质量食品供给能力的短缺、监管资源和能力的不足使得我国的食品安全形势日益严峻。传统的基于化学的检测方法周期冗长, 流程复杂, 很难大规模地投入使用中。得益于较快的评估速度和无损坏的评估过程, 基于计算机视觉的食品质量检测成为近年来热门的研究方向。在整个食品质量检测系统中, 稳健而准确的食品识别算法至关重要, 能否精准地识别自然环境下的食品类别将直接影响到系统的评估性能。同时也可以帮助食品监管部门追溯食品信息、自动分拣过期食品, 从而加大食品安全的监察力度<sup>[179-180]</sup>。

### 5 展望

在过去 10 年里, 食品图像识别无论从算法设计还是基准数据库构建都已经取得了一系列进展。识别方法建立了以深度学习模型为基础的体系框架, 基准数据库出现了菜品、果蔬和零售食品等不同类型的数据库。在基于之前食品图像识别方法、数据库及应用的基础上, 本文接下来对食品图像识别存在的问题进行了讨论, 并重点探讨了食品图像识别发展的方向。

#### 5.1 面向食品特点的食品图像识别方法

食品图像识别尽管从概念上来说属于细粒度图像识别, 但不同于传统的细粒度图像识别。以主流的菜品图像识别为例, 一方面很多类的菜品图像具有一定的细粒度特性, 需要挖掘和放大局部的细节区域才可以有效的区分不同菜品类别。另一方面, 很多菜品类别的差异和普通的图像识别也比较相似, 这就需要考虑食品图像的整体特征, 使得类间差比较大的类通过全局特征更容易区分。因此设计的菜品图像识别模型要兼顾这两个方面。此外, 常规的细粒度识别任务通常挖掘固定的语义部件(例如鸟类的头、胸和翅膀)以捕获图像判别性的视觉特征。然而很多类的菜品图像并不具有固定的语义模式, 很难像定义鸟类的翅膀和头部一样定义菜品图像的局部判别性区域, 因而也不能直接使用现有的细粒度视觉分类方法来进行食品图像识别。考虑卷积神经网络强大的表达能力, 如何充分挖掘食品图像的特点设计针对性的神经网络用于食品图像识别是未来重要的发展方向。此外, 随着智能便携式设备、智能家居等的迅速普及, 将食品图像识别

应用于移动设备和边缘设备上的需求日益增加。因此,基于轻量化卷积神经网络的食品图像识别也将受到越来越多关注。

## 5.2 构建大规模食品图像识别基准数据库

在当前深度学习时代,大规模 ImageNet 数据库<sup>[35]</sup>促进了物体识别算法的发展。同样,发展食品图像识别技术,也需要大规模的食品图像数据库的支撑。以主流的菜品数据库为例,目前确实存在一些基准菜品数据库,例如 ETHZ Food-101<sup>[15]</sup>和 Vireo Food-172<sup>[41]</sup>及近来发布的 ISIA Food-500<sup>[43]</sup>。但是,与 ImageNet 相比,这些数据集的类别和图像数量在规模上还不够大。构建食品识别数据库有其特殊的挑战。例如,以菜品为例,同一道菜因区域差异可能有几种不同的名称。同样,有些菜品虽被标记为相同的菜品名称,但实际上属于包含不同食材的不同菜品。另外重要的一点是 ImageNet 依据 WordNet 的标准概念体系构建,而对于菜品数据库,东西方的菜品分类体系具有明显的差异。再比如包装食品,种类更为庞杂,很难有统一标准的分类体系。这也给构建大规模食品图像识别数据库带来了困难。因此针对不同的食品类型,如何构建标准统一的菜品拓扑体系是构建大规模食品数据库首要解决的问题。此外相比于 ImageNet 数据库,食品图像识别数据库的标注可能需要专家标注,如何设计一种有效方案保证高质量和低成本的标注也是未来构建大规模食品图像识别数据库需要解决的一个关键问题。

## 5.3 基于多传感器信息融合的食品识别

除了视觉信息,食品数据中还蕴含着丰富的其他多模态信息,如嗅觉信息、近红外信息、触觉信息和光谱信息等。近年来,各种传感器由于能够捕获食品不同类型的多模态信息而被广泛应用于食品识别研究,例如,一些工作<sup>[28,45,181]</sup>提出了基于 RGB-D 的菜品或者水果识别方法。Tapomayukh 等人<sup>[182]</sup>提出了一种使用触觉进行食品识别的方法,以实现康复机器人的自动喂食。Shirin 等人<sup>[183]</sup>提出了一种使用高光谱图像检测一餐中不同食物的方法,可以快速自动识别食物。Amit 等人<sup>[184]</sup>提出了一种使用近红外信息进行冰箱食品识别的方法。Ghasemi 等人<sup>[185]</sup>将电子鼻、电子耳和生物传感器用于食品变质检测系统中。这些传感器能够提取食品的嗅觉、触觉等信息。随着传感器技术的快速发展,研究选取哪些多模态信息集合用于食品识别以及采用何种多模态特征融合方法也将成为食品识别

的一个重要发展方向,值得深入探讨。

## 5.4 基于食品图像识别的新应用

目前食品图像识别已经应用到饮食质量评估、食品自动结算、食品组织、检索和推荐等应用中,但未来还有更多食品图像识别的应用场景,如在智能厨具中,当前的智能烤箱依然需要根据所烤食物手工选择对应的烘烤类型,如果在烤箱中加入摄像头和食材识别算法,那么就可以实现烤箱的自动化烘烤。李等人<sup>[186]</sup>提出了一种智能烤箱及烘烤方法,其中食品识别模块用于识别烤箱中的食材信息。随着食品图像识别算法的日益成熟,开拓新的应用场景对推动未来食品图像识别的发展具有重要价值。

## 5.5 从食品图像识别走向检测和分割

当前食品图像识别的对象主要针对含有一个食品对象的图像。但是现实很多场景下,一张食品图像通常包含多个食品识别对象,这时候就需要食品图像的检测和分割<sup>[63,187-188]</sup>。例如 Aguilar 等人<sup>[60]</sup>提出了一种在食堂和饭店环境下对食物托盘进行自动分析的方法,能够自动地预测食物托盘中的多种食品类别。该方法将食品的定位、识别和细分集成到一个框架中以进行食物托盘识别。董<sup>[189]</sup>提出了一种基于深度可分离卷积的轻量级食品识别方法,该方法由食物检测网络和食物分类网络组成。食物检测网络使用深度可分离卷积和多层融合检测网络生成食物区域的候选框并判断是否存在食物,食物分类网络则根据候选框预测食物的类别标签。深度可分离卷积的广泛应用极大程度上降低了模型的计算成本和额外开销,使得模型能够成功地部署在各种移动设备上。苏<sup>[63]</sup>和 Lu<sup>[187]</sup>考虑到了菜品检测数据集的匮乏,采用了多目标分割算法识别图像中的菜品区域,随后对候选区域通过提取 CNN 特征进行识别。为了更有效解决现实世界中多个识别对象的问题,食品图像的检测和分割是未来亟需探索的重要研究方向。

## 6 结论

食品图像识别在计算机视觉和多媒体等研究领域中具有重要的理论意义和实际应用价值,但同时目前仍存在诸多问题与挑战。本文首先从识别对象、特征类型和学习方法三个维度总结了食品图像识别的研究进展,然后介绍并分析了食品图像识别领域主流的数据库及性能评价。本文还从饮食质量评估、食品自动结算、食品组织检索和推荐、智能

厨具、智能穿戴设备、餐饮机器人和食品安全等七个方面介绍食品图像识别的典型应用。最后本文还从面向食品特点的图像识别方法、构建大规模食品图像识别基准数据库、基于多传感器信息融合的食品识别、基于食品图像识别的新应用、食品图像检测和分割五个方面对食品图像识别的发展方向进行了分析与展望。随着更大规模的食品识别数据库的出现,深度学习在食品图像识别中将有望取得进一步的突破性进展,与此同时还将催生出更多的新应用。

## 参 考 文 献

- [1] Khanna S K. Food and Culture: A Reader, by Carole Counihan and Penny Van Esterik. *Ecology of Food and Nutrition*, 2009, 48(2): 157-159
- [2] Sajadmanesh S, Jafarzadeh S, Ossia S A, et al. Kissing cuisines: Exploring worldwide culinary habits on the web//Proceedings of the 26th International Conference on World Wide Web Companion. Perth, Australia, 2017: 1013-1021
- [3] Chung J, Chung J, Oh W, Yoo Y, et al. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Scientific Reports*, 2017, 7: 41690
- [4] Nestle M, Wing R, Birch L, et al. Behavioral and Social Influences on Food Choice. *Nutrition Reviews*, 2009, 56(5): 50-64
- [5] Sørensen L B, Møller P, Flint A, et al. Effect of sensory perception of foods on appetite and food intake: a review of studies on humans. *International Journal of Obesity*, 2003, 27(10): 1152-1166
- [6] Pauly D. A simple method for estimating the food consumption of fish populations from growth data and food conversion experiments. *National Oceanic and Atmospheric Administration*, 1986, 84(4): 827-840
- [7] Chen Z, Tao Y. Food safety inspection using “from presence to classification” object-detection model. *Pattern Recognition*, 2001, 34(12): 2331-2338
- [8] Harris, Marvin. Comment on Vayda's review of good to eat: Riddles of food and culture. *Human Ecology*, 1987, 15(4): 511-517
- [9] Min Weiqing, Jiang Shuqiang, Liu Linhu, Rui Yong, Jain Ramesh C. Jain. A Survey on Food Computing. *ACM Computing Surveys*, 2019, 52(5): 1-36
- [10] Qiu Jianing, Frank Po Wen Lo, Yingnan Sun, Siyao Wang, Benny Lo. Mining Discriminative Food Regions for Accurate Food Recognition//Proceedings of British Machine Vision Conference, Cardiff, UK, 2019: 158
- [11] Zhang X, Zhou F, Lin Y, et al. Embedding label structures for fine-grained feature representation//Proceedings of the Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1114-1123
- [12] Zhou F, Lin Y. Fine-grained image classification by exploring bipartite-graph labels//Proceedings of the Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1124-1133
- [13] Hou Saihui, Feng Yushan, Wang Zilei. VegFru: A Domain-Specific Dataset for Fine-Grained Visual Categorization//Proceedings of the International Conference on Computer Vision. Venice, Italy, 2017: 541-549
- [14] Austin Meyers, Nick Johnston, Vivek Rathod, et al. Im2Calories: Towards an automated mobile vision food diary//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1233-1241
- [15] Lukas Bossard, Matthieu Guillaumin, Luc Van Gool. Food-101 - Mining Discriminative Components with Random Forests//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 446-461
- [16] Herranz L, Jiang Shuqiang, Xu Ruihan. Modeling restaurant context for food recognition. *IEEE Transactions on Multimedia*, 2016, 19(2): 430-440
- [17] Deng L., Chen J., Sun Q., He, X., et al. Mixed-dish recognition with contextual relation networks//Proceedings of the 27th ACM International Conference on Multimedia, 2019, 112-120
- [18] Min Weiqing, Liu Linhu, Luo Zhengdong, Jiang Shuqiang. Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition//Proceedings of the 27th ACM International Conference on Multimedia. Nice France, 2019: 1331-1339
- [19] Y. Tan, K. Lu, M. M. Rahman, J. Xue. RgbD-Fg: A Large-Scale Rgb-D Dataset For Fine-Grained Categorization//Proceedings of 2020 IEEE International Conference on Multimedia and Expo. London, UK, 2020: 1-6
- [20] M. S. Hossain, M. Al-Hammedi, G. Muhammad. Automatic fruit classification using deep learning for industrial applications. *IEEE Transactions on Industrial Informatics*, 2018, 15(2): 1027-1034
- [21] Xiao G, Wu Q, Chen H, et al. A Deep Transfer Learning Solution for Automating Food Material Procurement using Electronic Scales. *IEEE Transactions on Industrial Informatics*, 2019, 16(4): 2290-2300
- [22] Gianluigi Ciocca, Paolo Napoletano, Raimondo Schettini. Food Recognition: A New Dataset, Experiments, and Results. *IEEE J. Biomed. Health Informatics*, 2017, 21(3): 588-598
- [23] E. T. Martinez, M. Leyva-Vallina, M. M. K. Sarker, D. Puig, N. Petkov and P. Radeva, Hierarchical Approach to Classify Food Scenes in Egocentric Photo-Streams. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(3): 866-877

- [24] Lo F P W, Sun Y, Qiu J, et al. Image-based Food Classification and Volume Estimation for Dietary Assessment: A Review. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(7): 1926-1939
- [25] Zhu L, Li Z, Li C, Wu J, et al. High performance vegetable classification from images based on Alexnet deep learning model. *International Journal of Agricultural and Biological Engineering*, 2018, 11: 217-223
- [26] Steinbrener J, Posch K, Leitner R. Hyperspectral fruit and vegetable classification using convolutional neural networks. *Computers and Electronics in Agriculture*. 2019, 162: 364-37
- [27] Anderson Rocha, Daniel C, Hauagge, Jacques Wainer. Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture*, 2010, 70(1): 96-104
- [28] Rodrigo Méndez Perez, Fernando Alfredo Auat Cheein, Joan Ramón Rosell Polo. Flexible system of multiple RGB-D sensors for measuring and classifying fruits in agri-food Industry. *Computers and Electronics in Agriculture*, 2017, 139: 231-242
- [29] Marcus Klasson, Cheng Zhang, Hedvig Kjellström. A Hierarchical Grocery Store Image Dataset with Visual and Semantic Labels//Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision. Waikoloa Village, USA, 2019: 491-500
- [30] E.Parrish, A.K.Goksel, Pictorial pattern recognition applied to fruit harvesting. *Transactions of the ASA*, 1977, 20(5): 822-827
- [31] Ruud M. Bolle, Jonathan H. Connell, Norman Haas, Rakesh Mohan, Gabriel Taubin. VeggieVision: a produce recognition system//Proceedings of the Third IEEE Workshop on Applications of Computer Vision. Sarasota, USA, 1996: 244-251
- [32] Giovanni Maria Farinella, Dario Allegra, Marco Moltisanti, Filippo Stanco, Sebastiano Battiato. Retrieval and classification of food images. *Computers in Biology and Medicine*, 2016, 77: 23-39
- [33] Keigo Kitamura, Toshihiko Yamasaki, Kiyoharu Aizawa. Food log by analyzing food images//Proceedings of the ACM Multimedia, 2008: 999-1000
- [34] Chen Mei, Kapil Dhingra, Wu Wen, Yang Lei, et al. PFID: Pittsburgh fast-food image dataset//Proceedings of the International Conference on Image Processing. Cairo, Egypt, 2009: 289-292
- [35] Yang S, Chen M, Pomerleau D, et al. Food recognition using statistics of pairwise local features//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2249-2256
- [36] Chomtip Pornpanomchai, Khomkhwan Srikeaw, Voranun Harnprasert, Kamonwan Promnurakkit. Thai fruit recognition system (TFRS)//Proceedings of the First International Conference on Internet Multimedia Computing and Service. Yunnan, China, 2009: 108-112
- [37] Chen Xin, Zhou Hua, Diao Liang. ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition. *arXiv preprint arXiv: 1705.02743*, 2017
- [38] Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre-training and fine-tuning//Proceedings of the International Conference on Multimedia & Expo Workshops. Turin, Italy, 2015: 1-6
- [39] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. Menu-match: Restaurant-specific food logging from images//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Hawaii, USA, 2015: 844-851
- [40] Xu Ruihan, Herranz L, Jiang Shuqiang, et al. Geolocalized modeling for dish recognition. *IEEE Transactions on Multimedia*, 2015, 17(8): 1187-1199
- [41] Chen Jingjing, Chong-Wah Ngo. Deep-based Ingredient Recognition for Cooking Recipe Retrieval//Proceedings of the 24th ACM international conference on Multimedia. New York, USA, 2016: 32-41
- [42] Bettadapura V, Thomaz E, Parnami A, et al. Leveraging context to support automated food recognition in restaurants//Proceedings of the International Conference on Applications of Computer Vision. Waikoloa, USA, 2015: 580-587
- [43] Min Weiqing, Liu Linhu, Wang Zhiling, Luo Zhengdong, et al. ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA, 2020
- [44] Jiang Shuqiang, Min Weiqing, Lv Yongqiang, et al. Few-Shot Food Recognition via Multi-View Representation Learning. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2020, 16(3): 1-20
- [45] Hameed K, Chai D, Rassau A. A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing*, 2018, 80: 24-44
- [46] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks//Proceedings of the International Conference on Advances in Neural Information Processing Systems. Nevada, USA, 2012: 1097-1105
- [47] Yang S, Chen M, Pomerleau D, et al. Food recognition using statistics of pairwise local features//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2249-2256
- [48] Wu W, Yang J. Fast food recognition from videos of eating for calorie estimation//Proceedings of the International Conference on Multimedia and Expo. New York, USA, 2009: 1210-1213
- [49] Anthimopoulos M M, Gianola L, Scarnato L, et al. A food recognition system for diabetic patients based on an optimized bag-of-features model.

- IEEE Journal of Biomedical and Health Informatics, 2014, 18(4): 1261-1271
- [50] Kagaya H, Aizawa K, Ogawa M. Food detection and recognition using convolutional neural network//Proceedings of the International Conference on Multimedia. Orlando, Florida, USA, 2014: 1085-1088
- [51] Hassannejad H, Matrella G, Ciampolini P, et al. Food image recognition using very deep convolutional networks//Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. Amsterdam, The Netherlands, 2016: 41-49
- [52] Ming Z Y, Chen J, Cao Y, et al. Food photo recognition for dietary tracking: System and experiment//Proceedings of the 24th International Conference on Multimedia Modeling. Bangkok, Thailand, 2018: 129-141
- [53] Yang H, Kang S, Park C, et al. A Hierarchical deep model for food classification from photographs. KSII Transactions on Internet and Information Systems, 2020, 14(4): 1704-1720
- [54] Martinel N, Foresti G L, Micheloni C. Wide-slice residual networks for food recognition//Proceedings of the International Conference on Applications of Computer Vision. Lake Tahoe, USA, 2018: 567-576
- [55] Chen Ya-Li. Research on Food Image Classification Based on Deep Learning[M.D. dissertation]. University of Electronic Science and Technology of China, Chengdu, 2018(in Chinese)
- (陈雅丽. 基于深度学习的菜品图像分类方法研究[硕士学位论文]. 电子科技大学, 成都, 2018)
- [56] Liang Hua-Gang, Wen Xiao-Qian, Liang Dan-Dan, et al. Fine-grained food image recognition of a multi-level convolution feature pyramid. Journal of Image and Graphics, 2019, 24(6): 0870-0881(in Chinese)
- (梁华刚, 温晓倩, 梁丹丹, 等. 多级卷积特征金字塔的细粒度食物图片识别. 中国图象图形学报, 2019, 24(6): 0870-0881)
- [57] Matsuda Y, Yanai K. Multiple-food recognition considering co-occurrence employing manifold ranking//Proceedings of the International Conference on Pattern Recognition. Tsukuba, Japan, 2012: 2017-2020
- [58] Matsuda Y, Hoashi H, Yanai K. Recognition of multiple-food images by detecting candidate regions//Proceedings of the International Conference on Multimedia and Expo. Melbourne, Australia, 2012: 25-30
- [59] Bolaños M, Radeva P. Simultaneous food localization and recognition//Proceedings of the International Conference on Pattern Recognition. Cancun, Mexico, 2016: 3140-3145
- [60] Aguilar E, Remeseiro B, Bolaños M, et al. Grab, pay, and eat: semantic food detection for smart restaurants. IEEE Transactions on Multimedia, 2018, 20(12): 3266-3275
- [61] Anthimopoulos M, Dehais J, Diem P, et al. Segmentation and recognition of multi-food meal images for carbohydrate counting//Proceedings of the International Conference on BioInformatics and BioEngineering. Chania, Greece, 2013: 1-4
- [62] Than Le. Mask R-CNN with data augmentation for food detection and recognition. techrxiv preprint, 2020
- [63] Su Guo-Yang. Image-nased Chinese Dish Segmentation and Recognition[M.D. dissertation]. Zhejiang University. Hangzhou, 2019(in Chinese)
- (苏国炀. 基于图像的中餐菜品分割与识别[硕士学位论文]. 浙江大学, 杭州, 2019)
- [64] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection//Proceedings of the Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 779-788
- [65] Chen H, Xu J, Xiao G, et al. Fast auto-clean CNN model for online prediction of food materials. Journal of Parallel and Distributed Computing, 2018, 117: 218-227
- [66] Pan L, Pouyanfar S, Chen H, et al. Deepfood: Automatic multi-class classification of food ingredients using deep learning//Proceedings of the Conference on Collaboration and Internet Computing. San Jose, USA, 2017: 181-189
- [67] Bolaños M, Ferrà A, Radeva P. Food ingredients recognition through multi-label learning. Lecture Notes in Computer Science, 2017, 10590: 394-402
- [68] Zhang X J, Lu Y F, Zhang S H. Multi-task learning for food identification and analysis with deep convolutional neural networks. Journal of Computer Science and Technology, 2016, 31(3): 489-500
- [69] Pan L, Li C. A Novel Combinational Convolutional Neural Network for Automatic Food-Ingredient Classification. Computers, Materials & Continua, 2020, 62(2): 731-746
- [70] Shiv Ram Dubey, Anand Singh Jalal. Species and variety detection of fruits and vegetables from images. International Journal of Applied Pattern Recognition, 2013, 1(1): 108-126
- [71] Shiv Ram Dubey, Anand Singh Jalal. Fruit disease recognition using improved sum and difference histogram from images. International Journal of Applied Pattern Recognition(IJAPR), 2014, 1(2): 199-220
- [72] Susovan Jana, Saikat Basak, Ranjan Parekh. Automatic Fruit Recognition from Natural Images using Color and Texture Features//Proceedings of the Conference on Devices for Integrated Circuit. Kalyani, India, 2017
- [73] Li Dahua, Shen Mingming, Li Dong, Yu Xiao. Green apple recognition method based on the combination of texture and shape features//Proceedings of the IEEE International Conference on Mechatronics and Automation. Takamatsu, Japan, 2017: 264-269

- [74] Zhang Yudong, Wang Shuihua, Ji Genlin, Preetha Phillips. Fruit classification using computer vision and feedforward neural network. *Journal of Food Engineering*, 2014, 143: 167-177
- [75] Fábio Augusto Faria, Jefersson Alex dos Santos, Anderson Rocha, Ricardo da Silva Torres. Automatic Classifier Fusion for Produce Recognition//Proceedings of the 25th SIBGRAPI Conference on Graphics, Patterns and Images. Ouro Preto, Brazil, 2012: 252-259
- [76] Wang Shuihua, Zhang Yudong, Ji Genlin, Yang Jiquan, et al. Fruit classification by wavelet entropy and feedforward neural network trained by fitness-scaled chaotic abc and biogeography-based optimization. *Entropy*, 2015, 17(8): 5711-5728
- [77] Yu-Dong Zhang, Preetha Phillips, Shuihua Wang, Genlin Ji, et al. Fruit classification by biogeography-based optimization and feedforward neural network. *Expert Systems*, 2016, 33(3): 239-253
- [78] Zhang, Y., Wu, L. Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine. *Sensors*, 2012, 12(9): 12489-12505
- [79] Lu, S., Lu, Z., Aok, S., Graham, L. Fruit classification based on six layer convolutional neural network//Proceedings of the 2018 IEEE 23rd International Conference on Digital Signal Processing. Shanghai, China, 2018: 1-5
- [80] Wang, S.H., Chen, Y. Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. *Multimedia Tools and Applications*, 2018, 1-17
- [81] Katarzyna, R., Paweł, M. A Vision-Based Method Utilizing Deep Convolutional Neural Networks for Fruit Variety Classification in Uncertainty Conditions of Retail Sales. *Applied Sciences*, 2019, 9: 3971
- [82] Patino-Saucedo, A., Rostro-Gonzalez, H., Conradt, J. Tropical Fruits Classification Using an AlexNet-Type Convolutional Neural Network and Image Augmentation//Proceedings of the International Conference on Neural Information Processing. Berlin, Germany, 2018: 371-379
- [83] Zeng, G. Fruit and Vegetables Classification System Using Image Saliency and Convolutional Neural Network//Proceedings of the IEEE 3rd Information Technology and Mechatronics Engineering Conference. Chongqing, China, 2017: 613-617
- [84] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition//Proceedings of the International Conference on Learning Representations. San Diego, USA, 2015
- [85] Horea Muresan, Mihai Oltean. Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 2018, 10(1): 26-42
- [86] Ma Yue. Research and Implementation of Fruits and Vegetables detection and Classification Based on Deep Learning[M.D. dissertation]. Beijing University of Posts and Telecommunications, Beijing, 2018(in Chinese)
- (马越. 基于深度学习的果蔬检测与分类方法的研究与实现[硕士学位论文]. 北京邮电大学, 北京, 2018)
- [87] A. Wajid, N.K. Singh, P. Junjun, M.A. Mughal, Recognition of Ripe, Unripe and Scaled Condition of Orange Citrus Based on Decision Tree Classification//Proceedings of the International Conference on Computing, Mathematics and Engineering Technologies. Sukkur, Pakistan, 2018: 2-5
- [88] Wan Peng, Arash Toudeshki, Tan Hequn, Reza Ehsani. A methodology for fresh tomato maturity detection using computer vision. *Computers and Electronics in Agriculture*, 2018, 146: 43-50
- [89] L.F.S. Pereira, S. Barbon, N.A. Valous, D.F. Barbin, Predicting the ripening of papaya fruit with digital imaging and random forests, *Computers and Electronics in Agriculture*, 2018, 145: 76-82
- [90] Zhang M M. Identifying the Cuisine of a Plate of Food. University of California San Diego, La Jolla, USA, 2011: 190
- [91] Su H, Lin T W, Li C T, et al. Automatic recipe cuisine classification by ingredients//Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing. Seattle, USA, 2014: 565-570
- [92] Min Weiqing, Jiang Shuqiang, Sang J, et al. Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration. *IEEE Transactions on Multimedia*, 2016, 19(5): 1100-1113
- [93] Druck G. Recipe attribute prediction using review text as supervision. *Cooking with Computers//Proceedings of the International Joint Conference on Artificial Intelligence workshop*. Beijing, China, 2013
- [94] Ciocca, G., Micali, G., & Napoletano, P. . State Recognition of Food Images Using Deep Features. *IEEE Access*, 2020, 8: 32003-32017
- [95] Md. Mostafa Kamal Sarker, Hatem A. Rashwan, Estefanía Talavera, et al. MACNet: Multi-scale Atrous Convolution Networks for Food Places Classification in Egocentric Photo-Streams//Proceedings of the European Conference on Computer Vision Workshops. Munich, Germany, 2018: 423-433
- [96] Bolaños M, Valdivia M, Radeva P. Where and What Am I Eating? Image-Based Food Menu Recognition//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 590-605
- [97] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91-110
- [98] Navneet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005: 886-893
- [99] Antonio Ramón Jiménez, Anil K. Jain, Ramón Ceres Ruiz, José Luis Pons Rovira. Automatic fruit recognition: a survey and new results using Range/Attenuation images, *Pattern Recognition*, 1999, 32(10): 1719-1736

- [100] Zhu F, Bosch M, Schap T R, et al. Segmentation assisted food classification for dietary assessment//Proceedings of the International Society for Optical Engineering, San Francisco, USA, 2011: 78730B
- [101] Giovanni Maria Farinella, Marco Moltisanti, Sebastiano Battiato. Classifying food images represented as bag of textons// Proceedings of the International Conference on Image Processing. Paris, France, 2014:5212–5216
- [102] L.M. Oo, N.Z. Aung. A simple and efficient method for automatic strawberry shape and size estimation and classification. Biosystems Engineering, 2018, 170:96-107
- [103] Kong F, Tan J. Dietcam: Regular shape food recognition with a camera phone//Proceedings of the International Conference on Body Sensor Networks. Dallas, USA, 2011: 127-132
- [104] Joutou T, Yanai K. A food image recognition system with multiple kernel learning//Proceedings of the International Conference on Image Processing. Cairo, Egypt, 2009: 285-288
- [105] Hoashi H, Joutou T, Yanai K. Image recognition of 85 food categories by feature fusion//Proceedings of the International Symposium on Multimedia. Taichung, Taiwan, 2010: 296-301
- [106] Nguyen D T, Zong Z, Ogunbona P O, et al. Food image classification using local appearance and global structural information. Neurocomputing, 2014, 140: 242-251
- [107] Oliveira L, Costa V, Neves G, et al. A mobile, lightweight, poll-based food identification system. Pattern Recognition, 2014, 47(5): 1941-1952
- [108] Martinel N, Piciarelli C, Micheloni C, et al. A structured committee for food recognition//Proceedings of the International Conference on Computer Vision Workshops. Santiago, Chile, 2015: 92-100
- [109] Niki Martinel, Claudio Piciarelli, Christian Micheloni. A supervised extreme learning committee for food recognition. Computer Vision and Image Understanding, 2016, 148: 67-86
- [110] Ravi D, Lo B, Yang G Z. Real-time food intake classification and energy expenditure estimation on a mobile device//Proceedings of the International Conference on Wearable and Implantable Body Sensor Networks. Cambridge, USA, 2015: 1-6
- [111] Kawano Y, Yanai K. Foodcam: A real-time food recognition system on a smartphone. Multimedia Tools and Applications, 2015, 74(14): 5263-5287
- [112] He Hongsheng, Kong Fanyu, Tan Jindong. DietCam: Multiview food recognition using a multikernel SVM. IEEE Journal of Biomedical and Health Informatics, 2016, 20(3): 848–855
- [113] Zeiler M.D., Fergus, R. Visualizing and understanding convolutional networks//Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 818--833
- [114] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778
- [115] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, et al. Going Deeper with Convolutions//Proceedings of the Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1-9
- [116] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks//Proceedings of the Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 4700-4708
- [117] Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks//Proceedings of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7132-7141
- [118] Tanno R, Okamoto K, Yanai K. DeepFoodCam: A DCNN-based real-time mobile food recognition system//Proceedings of the International Workshop on Multimedia Assisted Dietary Management. Amsterdam, The Netherlands, 2016: 89-89
- [119] Pouladzadeh P, Shirmohammadi S. Mobile multi-food recognition using deep learning. ACM Transactions on Multimedia Computing, Communications, and Applications. 2017, 13(3s): 1-21
- [120] McAllister P, Zheng H, Bond R, et al. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. Computers in Biology and Medicine, 2018, 95: 217-233
- [121] Azizpour H, Razavian A.S., Sullivan J, et al.. From Generic to specific deep representation for visual recognition//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA, 2015: 36-45
- [122] Pandey P, Deepthi A, Mandal B, et al. FoodNet: Recognizing foods using ensemble of deep networks. IEEE Signal Processing Letters, 2017, 24(12): 1758-1762
- [123] Metwalli A S, Shen W, Wu C Q. Food Image Recognition Based on Densely Connected Convolutional Neural Networks//Proceedings of the Conference on Artificial Intelligence in Information and Communication. Fukuoka, Japan, 2020: 027-032
- [124] Mejia J, Ochoa-Zezzatti A, Contreras-Masse R, et al. Intelligent System for the Visual Support of Caloric Intake of Food in Inhabitants of a Smart City Using a Deep Learning Model. Applications of Hybrid Metaheuristic Algorithms for Image Processing. 2020: 441-455
- [125] He Kaiming, Georgia Gkioxari, Piotr Dollár, Ross B. Girshick. Mask R-CNN//Proceedings of the IEEE international conference on computer vision. Venice, Italy, 2017: 2980-2988
- [126] Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint, arXiv:1605.07146, 2016

- [127] Zhao H, Yap K H, Kot A C C, et al. JDNet: A Joint-learning Distilled Network for Mobile Visual Food Recognition. *IEEE Journal of Selected Topics in Signal Processing*, 2020: 1-1
- [128] Phiphiphatphaisit S, Surinta O. Food Image Classification with Improved MobileNet Architecture and Data Augmentation//Proceedings of the 2020 The 3rd International Conference on Information Science and System. 2020: 51-56
- [129] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications.. arXiv preprint, arXiv:1704.04861, 2017
- [130] Wu Zheng-Dong. Research on Chinese Food Dishes Image Classification Algorithm Based on Deep Learning[M.D. dissertation]. University of Electronic Science and Technology of China, Chengdu, 2020(in Chinese)
- (吴正东. 基于深度学习的中餐菜品图像分类算法研究[硕士学位论文]电子科技大学, 成都, 2020)
- [131] Liao En-Hong, Li Hui-Fang, Wang Hua, Pang Xiong-Wen. Food Image Recognition Based on Convolutional Neural Network. *Journal of South China Normal University(Natural Science Edition)*, 2019,51(4): 113-119(in Chinese)
- (廖恩红, 李会芳, 王华, 庞雄文. 基于卷积神经网络的食品图像识别. 华南师范大学学报(自然科学版), 2019,51(4): 113-119)
- [132] Kawano Y, Yanai K. Food image recognition with deep convolutional features//Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. Washington, USA, 2014: 589-593
- [133] Luis Herranz, Min Weiqing, Jiang Shuqiang. Food recognition and recipe analysis: integrating visual content, context and external knowledge. arXiv preprint arXiv:1801.07239, 2018
- [134] Wang Huayang, Min Weiqing, Li Xiangyang, et al. Where and what to eat: Simultaneous restaurant and dish recognition from food image. *Pacific Rim Conference on Multimedia*. 2016, 9916: 520-528
- [135] Wei Z, Chen J, Ming Z, et al. DietLens-Eout: Large Scale Restaurant Food Photo Recognition//Proceedings of the International Conference on Multimedia Retrieval. Ottawa, Canada, 2019: 399-403
- [136] Jiang Shuqiang, Min Weiqing, Liu Linhu, et al. Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing*, 2019, 29: 265-276
- [137] Kenneth Marino, Ruslan Salakhutdinov, Abhinav Gupta. The More You Know: Using Knowledge Graphs for Image Classification//Proceedings of the Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 20-28
- [138] Chen Chengpeng, Min Weiqing, Li Xue, Jiang Shuqiang. Hybrid incremental learning of new data and new classes for hand-held object recognition. *Journal of Visual Communication and Image Representation*, 2019, 58: 138-148
- [139] Simon Kornblith, Jonathon Shlens, Quoc V. Le. Do Better ImageNet Models Transfer Better//Proceedings of the Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019: 2661-2671
- [140] Cui Yin, Song Yang, Sun Chen, Andrew Howard, Serge J. Belongie. Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning//Proceedings of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4109-4118
- [141] Lv Yong-Qiang, Min Wei-Qing, Duan Hua, et al. Few-shot Food Recognition Combining Triplet Convolutional Neural Network with Relation Network. *Computer Science*, 2020, 47(1): 136-143(in Chinese)
- (吕永强, 闵巍庆, 段华, 等. 融合三元卷积神经网络与关系网络的小样本食品图像识别. *计算机科学*, 2020, 47(1): 136-143)
- [142] Zhang H, Torr P H S, Koniusz P. Few-shot Learning with Multi-scale Self-supervision. arXiv preprint, arXiv:2001.01600, 2020
- [143] Song G, Tao Z, Huang X, et al. Hybrid attention-based prototypical network for unfamiliar restaurant food image few-shot recognition. *IEEE Access*, 2020, 8: 14893-14900
- [144] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, et al. Learning to Compare: Relation Network for Few-Shot Learning//Proceedings of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 1199-1208
- [145] Tahir G A, Loo C K. An Open-Ended Continual Learning for Food Recognition Using Class Incremental Extreme Learning Machines. *IEEE Access*, 2020, 8: 82328-82346
- [146] Horiguchi S, Amano S, Ogawa M, et al. Personalized classifier for food image recognition. *IEEE Transactions on Multimedia*, 2018, 20(10): 2836-2848
- [147] He Jiangpeng, Mao Runyu, Shao Zeman, Zhu Fengqing. Incremental Learning In Online Scenario//Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020:13926-13935
- [148] Yoshiyuki Kawano, Keiji Yanai. FoodCam-256: A Large-scale Real-time Mobile Food Recognition System employing High-Dimensional Features and Compression of Classifier Weights//Proceedings of the 22nd ACM international conference on Multimedia. Orlando Florida, USA, 2014: 761-762
- [149] Yoshiyuki Kawano, Keiji Yanai. Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014:3-17
- [150] Georg Waltner, Michael Schwarz, Stefan Ladstätter, Anna Maria Weber, et al. Personalized Dietary Self-Management Using Mobile

- Vision-Based Assistance// Proceedings of the 3rd International Workshop on Multimedia Assisted Dietary Management. Catania, Italy, 2017: 385-393
- [151] Gianluigi Ciocca, Paolo Napoletano, Raimondo Schettini. Food Recognition and Leftover Estimation for Daily Diet Monitoring//Proceedings of the International Conference on Image Analysis and Processing Workshops. Genoa, Italy, 2015: 334-341
- [152] Wang Xin, Devinder Kumar, Nicolas Thome, Matthieu Cord, et al. Recipe recognition with large multimodal food dataset//Proceedings of the 3rd World Congress on Integrated Computational Materials Engineering Workshops. Colorado, USA, 2015: 1-6
- [153] Giovanni Maria Farinella, Dario Allegra, Filippo Stanco. A Benchmark Dataset to Study the Representation of Food Images//Proceedings of the European Conference on Computer Vision Workshops. Zurich, Switzerland, 2014: 584-599
- [154] Michele Merler, Hui Wu, Rosario Uceda-Sosa, Quoc-Bao Nguyen, John R. Smith. Snap, Eat, RepEat: A Food Recognition Engine for Dietary Logging//Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. Amsterdam, The Netherlands, 2016: 31-40
- [155] Singla A, Yuan L, Ebrahimi T. Food/non-food image classification and food categorization using pre-trained Googlenet model//Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. Amsterdam, The Netherlands, 2016: 3-11
- [156] Gianluigi Ciocca, Paolo Napoletano, Raimondo Schettini. Learning CNN-based Features for Retrieval of Food Images//Proceedings of the International Conference on Image Analysis and Processing Workshops. Catania, Italy, 2017: 426-434
- [157] Zhu Yousong, Zhao Xu, Zhao Chaoyang, Wang Jinqiao, Lu Hanqing. Food det: Detecting foods in refrigerator with supervised transformer network. Neurocomputing, 2020, 379: 162-171
- [158] Wu Hui, Michele Merler, Rosario Uceda-Sosa, John R. Smith. Learning to Make Better Mistakes: Semantics-aware Visual Food Recognition//Proceedings of the 24th ACM international conference on Multimedia. Amsterdam, The Netherlands, 2016: 172-176
- [159] Shuang Ao, Charles X. Ling. Adapting New Categories for Food Recognition with Deep Representation//Proceedings of The IEEE International Conference on Data Mining series Workshops. Atlantic City, USA, 2015: 1196-1203
- [160] Liu Chang, Cao Yu, Luo Yan, Chen Guanling, Vinod Vokkarane, Yunsheng Ma. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment//Proceedings of The International Conference on Smart homes and health Telematics. Wuhan, China, 2016: 37-48
- [161] Chen Long, Zhang Hanwang, Xiao Jun, Nie Liqiang, et al. SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning//Proceedings of the Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6298-6306
- [162] Eduardo Aguilar, Marc Bolaños, Petia Radeva. Food Recognition Using Fusion of Classifiers Based on CNNs//Proceedings of the International Conference on Image Analysis and Processing Workshops. Catania, Italy, 2017: 213-224
- [163] Fisher Yu, Wang Dequan, Evan Shelhamer, Trevor Darrell. Deep Layer Aggregation//Proceedings of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 2403-2412
- [164] Huang Y., Cheng Y., Chen D., Lee H., et al. Gpipe: Efficient training of giant neural networks using pipeline parallelism. arXiv preprint arXiv:1808.07233, 2018
- [165] Tan Mingxing, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, CA, USA, 2019: 6105-6114
- [166] C. S. Won. Multi-Scale CNN for Fine-Grained Image Recognition. IEEE Access, 2020, 8: 116663-116674
- [167] Frances E.Thompson, Amy F.Subar. Dietary Assessment Methodology. Nutrition in the Prevention and Treatment of Disease (Fourth Edition). 2017: 5-48
- [168] Benny Lo. An innovative passive dietary monitoring system. <http://detaryintake.org/index.php?id=1>, 2017
- [169] S Mezgec, T Eftimov, T Bucher, BK Seljak. Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. Public health nutrition, 2019, 22(7): 1193-1202
- [170] Jia W , Li Y , Qu R , et al. Automatic food detection in egocentric images using artificial intelligence technology. Public Health Nutrition, 2019, 22(7): 1168-1179
- [171] Min Weiqing, Jiang Shuqiang, Jain Ramesh C. Food Recommendation: Framework, Existing Solutions and Challenges. IEEE Transactions on Multimedia, 2019, PP: 1-1
- [172] Zhang Weishan, Zhang Yuanjie, Zhai Jia, Zhao Dehai, et al. Multi-source data fusion using deep learning for smart refrigerators. Computers in Industry, 2018, 95: 15-21
- [173] Mohammad, Iftekhar, et al. A Deep Learning Approach to Smart Refrigerator System with the assistance of IOT//Proceedings of the International Conference on Computing Advancements. Dhaka, Bangladesh, 2020: 1-7
- [174] Hashimoto A., Mori N., Funatomi T., Yamakata Y., et al. Smart Kitchen: A User Centric Cooking Support System//Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems. Málaga, Spain, 2008: 848--854

- [175] Liu C, Cao Y, Luo Y, et al. A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. *IEEE Transactions on Services Computing*, 2017, 11(2): 249-261
- [176] Liu Chang, Wang Xizhe, Ni Jing, Cao Yu, Liu Benyuan. An Edge Computing Visual System for Vegetable Categorization//Proceedings of the 18th IEEE International Conference On Machine Learning And Applications. Boca Raton, USA, 2019: 625-632
- [177] Laavanya Rachakonda, Saraju P. Mohanty, Elias Kougnanos. iLog: An Intelligent Device for Automatic Food Intake Monitoring and Stress Detection in the IoMT. *IEEE Transactions on Consumer Electronics*, 2020, 66(2): 115-124
- [178] Ryan Feng, Youngsun Kim, Gilwoo Lee, Ethan K. Gordon, et al. Robot-Assisted Feeding: Generalizing Skewering Strategies across Food Items on a Realistic Plate. *arXiv preprint arXiv:1906.02350*, 2019
- [179] Gu Li-Qin. Image Recognition Technology of the Food Species Detection Method. *Food Research and Development*, 2017, 38(2): 160-175 (in Chinese)
- (顾理琴. 基于图像识别技术的食品种类检测方法. *食品研究与开发*, 2017, 38(2): 160-175)
- [180] Wang Bo, Liu Jun-Kang, Lu Feng-Gui, Liu Deng-Yong, Cao Zhen-Xia. Application of convolutional neural network in image retrieval and classification of food image. *Journal of Food Safety and Quality*, 2019, 10(18): 6241-6247(in Chinese)
- (王博, 刘俊康, 陆逢贵, 刘登勇, 曹振霞. 基于卷积神经网络的食品图像识别. *食品安全质量检测学报*, 2019, 10(18): 6241-6247)
- [181] Ding Mei-Kun, Xu Yu-Lin, Fei Xin-Yi, Jiang Cai-Jun. Fruit Recognition and Location Based on RGB-D. *Industrial Control Computer*, 2016, 29(5): 46-50(in Chinese)
- (丁美昆, 徐昱琳, 费心怡, 蒋财军. 基于 RGB-D 信息的果蔬识别与定位. *工业控制计算机*, 2016, 29(5): 46-47)
- [182] Tapomayukh Bhattacharjee, Gilwoo Lee, Hanjun Song, Siddhartha S. Srinivasa. Towards Robotic Feeding: Role of Haptics in Fork-Based Food Manipulation. *IEEE Robotics and Automation Letters*, 2019, 4(2): 1485-1492
- [183] Shirin Nasr Esfahani, Venkatesan Muthukumar, Emma E. Regentova, Kazem Taghva, et al. Trabia. Complex Food Recognition using Hyper-Spectral Imagery//Proceedings of the 10th Annual Computing and Communication Workshop and Conference. Las Vegas, USA, 2020: 662-667
- [184] Amit Sharma, Archan Misra, Vengateswaran Subramaniam, Youngki Lee. SmrtFridge: IoT-based, user interaction-driven food item & quantity sensing//Proceedings of the 17th Conference on Embedded Networked Sensor Systems. New York, USA, 2019: 245-257
- [185] Ghasemi-Varnamkhasti, Mahdi, et al. Potential use of electronic noses, electronic tongues and biosensors as multisensor systems for spoilage examination in foods. *Trends in Food Science & Technology*, 2018, 80: 71-92
- [186] 李生明, 王月, 王侃, 白云飞. 一种智能烤箱及其烘烤方法, 中国. 2019-10-25
- [187] Y. Lu, T. Stathopoulou, M. F. Vasiloglou, S. Christodoulidis, et al. An Artificial Intelligence-Based System to Assess Nutrient Intake for Hospitalised Patients. *IEEE Transactions on Multimedia*, 2020: 1-1
- [188] Gordon Ethan K, Meng Xiang, Barnes Matt, Bhattacharjee, et al. Adaptive Robot-Assisted Feeding: An Online Learning Framework for Acquiring Previously-Unseen Food Items. *arXiv preprint arXiv:1908.07088*, 2019
- [189] Dong Tian-Jiao. Food Classification and Detection Based on Convolutional Neural Network[M.D. dissertation]. Hangzhou Dianzi University, Hangzhou, 2018(in Chinese)
- (董天骄. 基于卷积神经网络的饮食分类与识别[硕士学位论文]. 杭州电子科技大学, 杭州, 2018)
- [190] Min Weiqing, Bao Bingkun, Mei Shuhuan, Zhu Yaohui, Rui Yong, Jiang Shuqiang. You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis. *IEEE Transactions on Multimedia*, 2018, 20(4): 950-964
- [191] Javier Marín, Aritro Biswas, Ferda Ofli, Nicholas Hynes, et al. RecipeIM: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20