**基于神经网络的食品**

**图像分类与热量分析**

项目成员：

范博坤 王建军 高峰

何晓曦 龚茗荣 贾亚伟

时间：2021年9月27日—2021年12月20日

## 目 录

目录

[目 录 2](#_Toc91490355)

[一、项目背景 3](#_Toc91490356)

[1.1 国内外现状 3](#_Toc91490357)

[1.2 问题分析 4](#_Toc91490358)

[1.3 项目创新点 6](#_Toc91490359)

[二、项目研发方案 7](#_Toc91490360)

[2.1 系统设计 7](#_Toc91490361)

[2.2 软件设计 7](#_Toc91490362)

[三、测试与结果 10](#_Toc91490363)

[3.1 软件测试 10](#_Toc91490364)

[3.2 系统测试 10](#_Toc91490365)

[四、小结与未来工作 13](#_Toc91490366)

[4.1 项目小结 13](#_Toc91490367)

[4.2 未来工作 13](#_Toc91490368)

[五 组员分工 14](#_Toc91490369)

[5.1 范博坤 工作概述 15](#_Toc91490370)

[5.2 王建军工作概述 15](#_Toc91490371)

[5.3 高峰工作概述 16](#_Toc91490372)

[5.4 何晓曦工作概述 17](#_Toc91490373)

[5.5 龚茗荣工作概述 17](#_Toc91490374)

[5.6 贾亚伟工作概述 18](#_Toc91490375)

[6 参考文献 19](#_Toc91490376)

## 一、项目背景

### 1.1 国内外现状

食物是供给机体活动所需能的重要物质来源，与人类日常生活息息相关。由于食物种类繁多，所含营养成分也干差万别，在普通条件下无法快速测定食物的种类及营养成分，但随着深度学习的火热，计算机视觉技术可以完成对食物图像的识别任务，因此，如何搭建一个泛化性较强、识精度较高的食物图像识別分类算法是当前热门的研究领域之一。在当下的食物图像识别分类领域，在国外，学者 Yoshiyuki Kawano[1]等提出了ー种基于能移动设备端的食物识技Joutou等[2]提出采用多重核函数学习的方法集成颜色、纹理等图像特征用于食物图像的识别工作；Austin Myer[3]等通过卷积神经网络实现对食物的识别，并基于食物的常用烹方法对其养成分进行分析，keiji Anais等人提出使用Pre- training和Fine- training来提高深度卷积神经网络对食物图像的识别精度[4]， Simon Mezgecs等[5]融4种卷积神经网络形成深度神经网络进行食物识别。

在国内， Zhihui Fu等[6]提出使用残差网络模型来进行中餐品食物图像的识别郭礼华等[7]使用SIFT特征、颜色直方图特征、梯度直方图、SURF特征、LBP特征和 Gabor.特征等对食材图像数据库进行了分类，并与传统特征分类方法进行了分析与比较 Shulin Yangs等[8]对食物不同分成之间的特征关系进行两两统计形成食物识别分类器 Xin Chens[9]等通过分别对不同的深度神经网络进行训练，提出了一种简单集成训练好的深度神经网络模型用于中餐食物识别的研究方法。

### 1.2 问题分析

本项目旨在我们使用基于机器学习的神经网络和目前所推出的API，来实现常见中餐与常见水果的识别，并根据识别结果和相应的称重来得到该食物的总热量，最后用户可以个人选择认为更为准确的结果。

首先是寻找数据集的过程：

最开始我们找到的数据集是Food-101 Dataset。该数据集由斯坦福大学于 2014 年发布，相关论文有《Food-101 — Mining Discriminative Components with Random Forests》。

Food-101 Dataset 是包含 101 中食品类别的图像数据集，主要用于图像分类，共有 101,000 张图像，每个类别的测试图像和训练图像分别有 250 张和 750 张，其中训练图像未经数据清洗，但所有图像均以重新进行尺寸缩放，最大边长在 512 像素。

但是我们下载了该数据集而后分析发现，该数据集主要包含的为西方的常见食品，且仅包含101种常见的食品。对于我们这次的需求来说并不太相符。而后又找到了一些关于印度的常见食物的数据集，同样不符合我们的要求。

最后我们找到了Chinese-FoodNet。该数据集由美的全球人工智能团队提供，其中包括了共计208类中国各大常见菜系的实物和菜谱图片，合计180,000张，共18GB，是目前世界上最大的中国成品菜数据集。在下载评估过后，我们发现该数据集非常适合我们目前的情况。同时，通过对该数据集的研究我们发现关于食品热量的调研变得更加复杂，因为中国的成品菜成分并不单一，其单位质量的热量含量需要进行更细致的调研，这部分我们将在之后的工作中加强。

综合考虑我们选择了Chinese-FoodNet数据集，1）：数据量更多，食物的种类更多；2）：更符合我们日常饮食情况，我们的饮食大部分都是中餐。但是选择了该数据集，意味着之后我们的对应食物的单位热量含量情况变得更加复杂。

在水果数据集的收集中，我们采用了爬虫来完成。爬虫的过程为首先通过用户输入要爬取的信息的关键字，随后再根据关键字建立URL，通过“request”机制来获取URL对应的信息，再进行正则匹配“objURL”，最终再进行图片的下载工作，随后将该类图片保存在本地的相应文件夹中以便后续的使用。

在神经网络的选取过程中：

我们开始选定了最近在图像识别领域十分热门的Tranformer神经网络，再分别针对菜品数据集和水果数据集，我们各自选定了不同的神经网络结构来对数据集进行划分，并且为了验证模型的效果，我们决定在用API与菜品识别模型进行对比，在水果识别时，用不同的神经网络模型进行对比。

最终菜品数据集的分类我们采用了华为的Transformer-in-Transformer神经网络模型，水果数据集的分类我们采用了CNN模型和Vision-Transformer模型。在训练完成之后我们用食堂拍摄的菜品图像和日常拍摄的水果图像对模型的功能进行测试。

在选取API的过程中：

在各大互联网公司的API库中，我们进行了广泛的调研，目前已有三家公司提供了菜品识别的API。包括百度云，京东云，腾讯云。但是目前京东云和腾讯云的菜品识别API仍在测试阶段，而百度云的菜品识别API已经有了较为成熟的实例，已经有多款调用了百度云的菜品识别API并且已经投入了市场开始使用。

在系统集成中：

为了将模型的功能和API的功能集成起来，做一个集中的展示。考虑到不同组员训练神经网络模型的环境都不一样，我们综合考虑之后，决定将模型上传到云服务器上，在通过远程与服务器通讯上传图片，并得到返回结果。

### 1.3 项目创新点

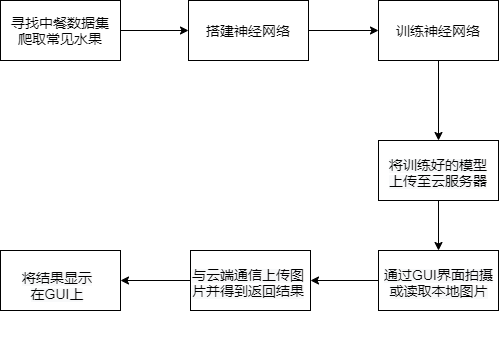
1）基于华为Transformer-in Transformer神经网络，实现了高准确度的食物识别，在划分的数据集上准确可以达到87.54%。

2）首次将中餐识别与热量分析相结合，填补了这一部分的空缺。

3）实现了GUI界面与后端的结合，使整个食物识别过程更加的人性化，便于用户进行操作使用。

## 二、项目研发方案

### 2.1 系统设计

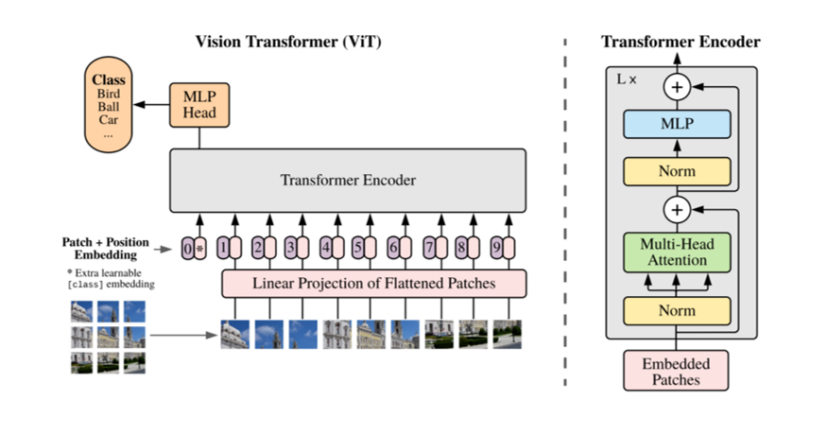
整个项目的设计如下流程图所示：

**图2.1 项目设计流程图**

### 2.2 软件设计

本项目使用了百度API和两个神经网络来实现食物的识别，神经网络分别是Transformer-in Transformer和Vision-Transformer。在具体功能实现上，先训练网络，接着利用训练好的网络进行食物种类的识别，识别成功之后在后台有一个热量表和每种食物一一对应。

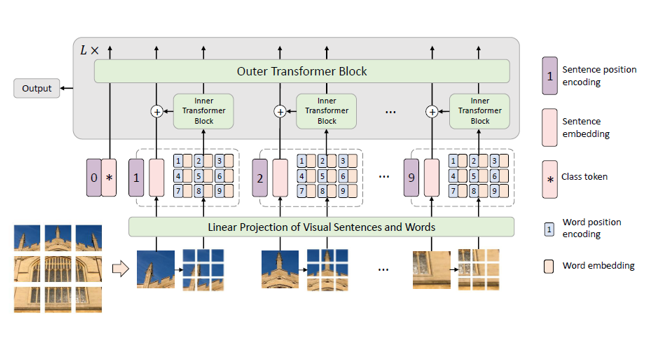
在软件设计部分我们会主要讲解下使用的两个用于分类的的神经网络：Vision-Transformer和Transformer-in Transformer。



**图2.2 Transformer-ViT神经网络的结构图**

Vision Transformer将CV和NLP领域知识结合起来，对原始图片进行分块，展平成序列，输入进原始Transformer模型的编码器Encoder部分，最后接入一个全连接层对图片进行分类。在大型数据集上表现超过了当前SOTA模型。

其主要原理为：经典的Transformer分为Encoder和Decoder两个模块，其中最主要的部分就是Multi-Head Attention。通过将分块、位置编码和分类与Transformer相结合来形成最新的Vision Transformer网络。网络结构图如上图所示。



**图2.3 Transformer in Transformer神经网络结构图**

Transformer是一种最初用于NLP任务的基于自注意力的神经网络。最近，提出了基于纯Transformer的模型来解决计算机视觉问题。这些视觉Transformer通常将图像视为一系列patches，而忽略每个patch内部的固有结构信息。

使用的Transformer in Transformer神经网络有以下特点：

从层次上，相比以前的vision transformer更加的完整，考虑了patch和pixel两个level的feature和relation。

从内容上，添加了position encoding对attention处理后的spatial structure损伤进行了修补，效果显著。

整体上对vision transformer进行了应用缺陷弥补，正如文中所写，如今的transformer已经基本具有CNN的深度挖掘能力。虽然性能上还不如Ef-CNN的深度卷积，但是已经在相同参数和计算量下精确度已经优于大部分CNN model了。

## 三、测试与结果

### 3.1 软件测试

在水果识别方面，我们将其他数据集的图片用我们训练好的模型进行了测试，将近1000张水果的图片，准确率达到了95%。因此我们认为该模型已经基本可用，将其上传到了云端。

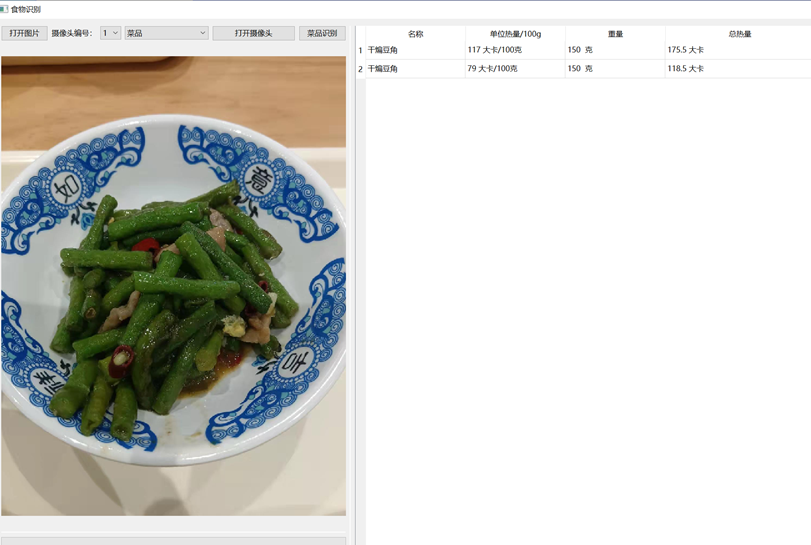
在菜品识别方面，我们通过日常拍摄食堂的菜品图片，对其进行识别，得到了较为准确的结果，且该模型在测试集上达到了93%的准确率，因此该模型也部署到了云端。

在菜品识别API方面，我们成功调用了百度的API，用于与我们的模型进行对比。

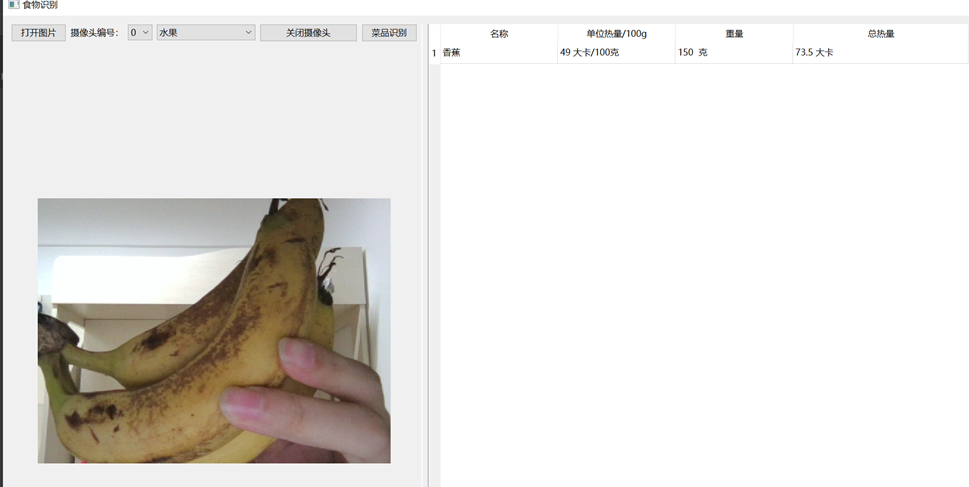
### 3.2 系统测试

 在最终的实验过程中，得到了以下的效果呈现：

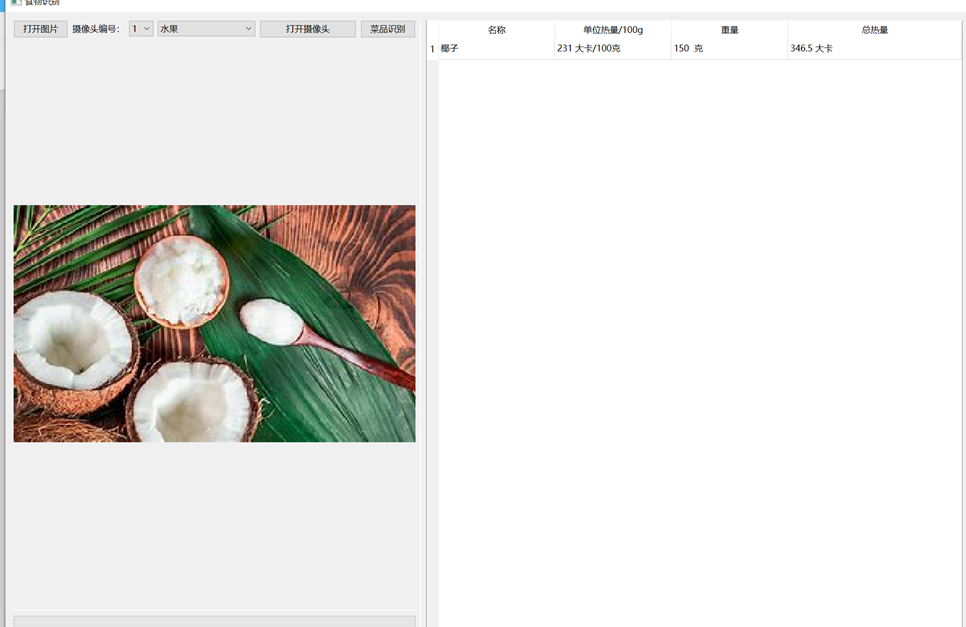
**图3.1 清蒸鲈鱼的识别**



**图3.2基于GUI界面的干煸豆角识别**



**图3.3 使用Vision-transformer神经网络实现对于香蕉的识别**



**图3.3 使用Vision-transformer神经网络实现对于椰子的识别**

## 四、小结与未来工作

### 4.1 项目小结

本项目针对基于神经网络的中餐菜品名称、水果识别进行了实现，在整个的识别过程中，我们按照7:2:1划分训练集、验证集和测试集，我们在测试集上的识别准确率高达87.54%，即使在imgnet任务中也是相当可观的，进一步验证了网络的性能。在识别速度方面，我们可以在RTX2080Ti的显卡条件下实现68img/s的高效识别效率，能够有效满足现实场景中的实时要求。取得了较好的识别性能; 与传统的CNN 以及SVM－HOG 方法相比，本文的方法可以更快速地对中餐菜品的进行识别，且具有较好的可移植性，因此将训练好的模型与移动端架构相结合，实现智能移动端设备上的APP 模块化操作也将成为后续工作的重点研究方向。

构建了完整的从拍照或读取本地图片到调用模型识别并返回结果的体系，并将其展示到GUI界面上。

### 4.2 未来工作

目前我们通过识别常见中餐和水果并给出相应的热量，虽然达到了较高的识别准确率但仍有以下方面可以改善：

1） 改进算法模型，模型参数较大，不便于在移动端部署，后续需要进一步压缩网络模型。

2）扩充数据集。目前只做了常见中餐与常见水果数据集，下一步计划新增常见糕点的种类识别与热量分析。

## 五 组员分工

|  |  |  |
| --- | --- | --- |
| 模块 | 功能 | 备注 |
| 个人（范博坤） | 项目整合 | 将各个组员的工作集成到一个GUI上。 |
| GUI设计 | GUI调用摄像头对水果拍照，读取本地图片，再与云端通信，得到返回结果。 |
| 个人（何晓曦） | 水果数据集收集 | 收集数据集，同时通过网络爬虫构建新的数据集。 |
| 个人（贾亚伟） | 菜品数据集收集 | 收集菜品数据集。 |
| 文档工作 | 负责海报，ppt，及文档的撰写 |
| 水果数据集分类 | 范博坤，龚茗荣 | |
| （两人合作） | |
| 寻找不同的网络模型，对水果数据集进行分类。 | 构建Vision-Transformer模型对水果数据集进行分类。 |
| 构建cnn神经网络对水果数据集进行分类。 |  |
| API识别菜品（高峰） | API | 调用百度的API对菜品的图片进行识别，返回菜品种类和热量。 |
| 个人（王健军） | 菜品数据集分类 | 构建Transformer-in-Transformer模型对菜品数据集进行分类。 |

**表5.1项目分工表**

### 5.1 范博坤 工作概述

主要工作

1）采用Vision-transformer神经网络对水果数据集进行分类

2）GUI设计

3）小组成员工作整合

收获与体会

提升了运用神经网络进行图像分类的能力，学习了与云服务器相关的一些知识，对项目整体的结构和过程有了更清晰的认识，同时也提高了领导能力。

### 5.2 王建军工作概述

主要工作

搭建多种网络模型进行中餐识别，并对多种网络进行任务对比。具体化为以下部分：

1）数据预处理

下载使用中餐数据集，并将数据处理成imgnet格式

2）搭建传统卷积算法以及transformer算法作为两种技术路线

传统的卷积算法以resnet50作为主干网络，并采用FPN特征金字塔结构融合底层与顶层信息用于中餐数据的分类。transformer算法我们复现了华为诺亚方舟实验室的Transformer in Transformer（TNT）模型，该模型对 patch级和像素级表示进行建模。在每个TNT块中，外部transformer块用于处理patch嵌入，而内部transformer块则从像素嵌入中提取局部特征。通过线性变换层将像素级特征投影到patch嵌入的空间，然后将其添加到patch中。通过堆叠TNT块，建立了用于图像识别的TNT模型。

3）对比两者的速度以及准确度。

速度方面传统的卷积算法胜出，以RTX2080ti显卡为例，传统的卷积算法每秒可以处理124张图片，而采用transformer的TNT结构，每秒只能处理10张图片，但是在精度方面TNT胜出，比卷积结构准确度提高了近4个点

收获与体会

物联网有广阔的前景，需要学习的科目也是很多的，通过了这段时间对物联网技术课程的了解和学习，更加向往物联网的独特魅力。物联网是新一代信息技术的重要组成部分。其英文名称是“The Internet of things”。由此，顾名思义，“物联网就是物物相连的互联网”。这有两层意思：第一，物联网的核心和基础仍然是互联网，是在互联网基础上的延伸和扩展的网络；第二，其用户端延伸和扩展到了任何物品与物品之间，进行信息交换和通信。 因此，物联网的定义是通过射频识别（RFID）、红外感应器、全球定位系统、激光扫描器等信息传感设备，按约定的协议，把任何物品与互联网相连接，进行信息交换和通信，以实现对物品的智能化识别、定位、跟踪、监控和管理的一种网络。物联网被视为互联网的应用扩展，应用创新是物联网的发展的核心，以用户体验为核心的创新是物联网发展的灵魂。其目的是实现物与物、物与人，所有的物品与网络的连接，方便识别、管理和控制。通过该课程对神经网络的搭建有了更清晰的认识，动手能力得到了进一步的提高。

### 5.3 高峰工作概述

主要工作

调用各大互联网公司（例如百度云、京东云、阿里云等）的云API来识别中餐菜品，经过调试和对比，由于百度云的菜品识别的精确度较高，决定使用百度云的菜品识别API。

收获与体会

通过这次大的项目，我学到了如何调用公开的API去进行我们的实验，以及如何与同组成员进行项目分工与讨论，学到了如何与其他小组进行对接、讨论并完成相应的任务。

### 5.4 何晓曦工作概述

主要工作

1）利用爬虫爬取了23种水果，每种1000张的图片。爬虫采集图片数据的流程为用户输入关键字、根据关键字构建url进行搜索、requests获取对应的html内容、正则匹配要采集的内容、下载图片。

2）初步采集的图片有一些无关和格式损坏的图片，人工对爬取的数据集进行了剔除筛选。

3）配合其他小组成员完成最后项目的结项汇报工作。

收获与体会

非常开心在本学期选修了《物联网技术》这门课，在这门课的学习过程中，我了解了物联网技术在生活中的应用，掌握了部分原理的实现原理。在完成小组项目的过程中，为了配合其他成员完成分类网络的训练，我用爬虫技术采集了图片数据集，提升了爬虫技术。跟其他小组成员学习了图片分类相关的算法，提升了合作沟通的技巧，为以后科研与工作打下了一定的基础。

### 5.5 龚茗荣工作概述

主要工作

1）寻找合适的分类算法

2）构建模型对水果数据集进行分类

收获与体会

### 这学期选修了《物联网技术》这门课，让我收获颇多，我觉得这门课十分注重动手操作的能力，本来我认为开发一个项目最重要的就是写代码，似乎整个软件都是编代码，因为自己动手能力不强所以就很排斥做项目。可是经过我们学习软工课程到团队做项目再到学习软件工程实践课程之后，我才真正意识到实施一个软件工程项目并不是说简单的会编码就能够解决问题的，因为一个软件的生命周期分为三个时期：软件定义时期、开发时期、维护时期，而这三个时期整体又分为七个阶段，他们分别是：问题定义、可行性研究、需求分析、总体设计、详细设计、编码和单元测试、综合测试，由此可看出，当我们开发一个项目时，更多的精力不是放在编码上，编码只是一个很小的模块，而是项目的整体结构上。让我学会了团队合作，总之这门课是一门十分有意义的课程。

### 5.6 贾亚伟工作概述

主要工作

负责菜品数据集的收集以及辅助何晓曦完成常见的水果数据集的爬取和筛选工作，到最终的项目验收阶段项目海报的设计与实现、答辩PPT的制作和项目报告的完成。

收获与体会

学会了在整个项目中要多与队员沟通，互相通报各自的进展以便项目更好的进行。在整个项目的进行过程中遇到问题要勇于提出并给出解决方案，以免到最后堆积成大的致命的问题。在做文字工作的过程中，更加熟练使用文字编辑软件和绘图软件。

## 6 参考文献

［1］KAWANO Y，YANAI K． A real－time food recognition system on a smartphone［J］． Multimed Tools Appl，2014，74( 14) : 1－25．

［2］ JOUTOU T，YANAI K． A food image recognition system with multiple kernel learning ［C］/ /2009 16th IEEE International Conference on Image Processing ( ICIP) ． IEEE，2009: 285－288．

［3］ MEYERS A，JOHNSTON N，RATHOD V，et al． Im2Calories: towards an automated mobile vision food diary［C］/ /Proceedings of the IEEE International Conference on Computer Vision．2015: 1233－1241．

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

［5］MEZGEC S，KOROUIC' SELJAK B． NutriNet: a deep learning food and drink image recognition system for dietary assessment ［J］． Nutrients，2017，9( 7) : 657．

［6］FU Z，CHEN D，LI H． Chinfood1000: A large benchmark dataset for chinese food recognition［C］/ /International Conference on Intelligent Computing． Springer，Cham，2017: 273－281．

［7］ 郭礼华，罗材． 食材数据库统计与对比实验性能分析［J］． 中国图象图形学报，2017，22( 8) : 1079．

［8］ YANG S，CHEN M，POMERLEAU D，et al． Food recognition using statistics of pairwise local features ［C］/ /2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition． IEEE，2010: 2249－2256．

［9］ CHEN X，ZHU Y，ZHOU H，et al． Chinesefoodnet: A largescale image dataset for chinese food recognition［J］． arXiv preprint arXiv: 1705．02743，2017．