



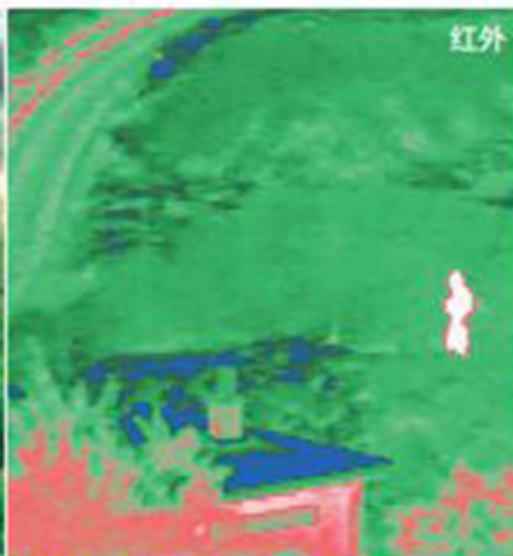
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2019
06
VOL.24

ISSN1006-8961
CN11-3758/TB



红外与可见光融合 P0987

中国图象图形学报

刊名题字：宋健 月刊（1996年创刊）



第24卷第6期（总第278期）

2019年6月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@radi.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

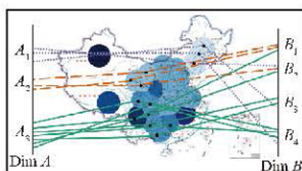
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



面向重复纹理及非刚性形变的像对高效稠密匹配方法
(第0924页)



平行坐标轴动态排列的地理空间多维数据可视分析
(第0956页)



细节感知的纹理滤波算法
(第0969页)

图像处理和编码

加强边缘感知的盲去模糊算法

邱枫, 侯飞, 袁野, 王文成 0847

图像分析和识别

3维灰度矩阵的钢板缺陷图像识别

兰红, 方治屿 0859

多级卷积特征金字塔的细粒度食物图片识别

梁华刚, 温晓倩, 梁丹丹, 李怀德, 茹锋 0870

综合边界和纹理信息的合成孔径雷达图像目标分割

谌华, 郭伟, 闫敬文 0882

融合分割先验的多图像目标语义分割

廖旋, 缪君, 储琚, 张桂梅 0890

级联优化CNN的手指静脉图像质量评估

曾军英, 谌瑶, 秦传波, 甘俊英, 翟懿奎, 冯武林 0902

使用人眼几何特征的视线追踪方法

苏海明, 侯振杰, 梁久祯, 许艳, 李兴 0914

图像理解和计算机视觉

面向重复纹理及非刚性形变的像对高效稠密匹配方法

贾迪, 赵明远, 杨宁华, 朱宁丹, 孟璟 0924

矩形件简单块占角排样方式的动态规划

潘卫平, 张瑞友 0934

Chinagraph 2018

结合细粒度特征与深度卷积网络的手绘图检索

李宗民, 刘秀秀, 刘玉杰, 李华 0946

平行坐标轴动态排列的地理空间多维数据可视分析

周志光, 余佳琚, 郭智勇, 刘玉华 0956

细节感知的纹理滤波算法

肖懿, 朱贤益, 何仰提, 郑燕 0969

用于快速服装搭配的FMatchNet算法

刘玉杰, 冯士贺, 李宗民, 李华 0979

BEMD分解和W变换相结合的红外与可见光图像融合

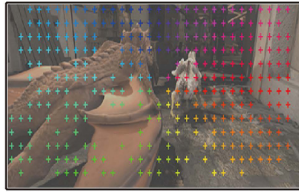
宫睿, 王小春 0987

柔性密封条截面误差自动视觉测量方法

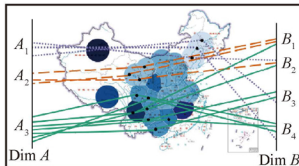
廖林, 李建华 1000

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Efficient dense matching method for repeated texture and non-rigid deformation(P0924)



Visual analysis of geospatial multi-dimensional data via a dynamic arrangement of parallel coordinates(P0956)



Detail-aware texture filtering algorithm(P0969)

Image Processing and Coding

Blind image deblurring with reinforced use of edges

Qiu Feng, Hou Fei, Yuan Ye, Wang Wencheng 0847

Image Analysis and Recognition

Image recognition of steel plate defects based on a 3D gray matrix

Lan Hong, Fang Zhiyu 0859

Fine-grained food image recognition of a multi-level convolution feature pyramid

Liang Huagang, Wen Xiaoqian, Liang Dandan, Li Huaide, Ru Feng 0870

Synthetic aperture radar image target segmentation method based on boundary and texture information

Chen Hua, Guo Wei, Yan Jingwen 0882

Multi-image object semantic segmentation by fusing segmentation priors

Liao Xuan, Miao Jun, Chu Jun, Zhang Guimei 0890

Finger vein image quality assessment based on cascaded fine-tuning convolutional neural network

Zeng Junying, Chen Yao, Qin Chuanbo, Gan Junying, Zhai Yikui, Feng Wulin 0902

Gaze tracking method for human eye geometric features

Su Haiming, Hou Zhenjie, Liang Jiuzhen, Xu Yan, Li Xing 0914

Image Understanding and Computer Vision

Efficient dense matching method for repeated texture and non-rigid deformation

Jia Di, Zhao Mingyuan, Yang Ninghua, Zhu Ningdan, Meng Lu 0924

Dynamic programming algorithm for simple block corner-occupying pattern of rectangular blanks

Pan Weiping, Zhang Ruiyou 0934

Chinagraph 2018

Sketch-based image retrieval based on fine-grained feature and deep convolutional neural network

Li Zongmin, Liu Xiuxiu, Liu Yujie, Li Hua 0946

Visual analysis of geospatial multi-dimensional data via a dynamic arrangement of parallel coordinates

Zhou Zhiguang, Yu Jiajun, Guo Zhiyong, Liu Yuhua 0956

Detail-aware texture filtering algorithm

Xiao Yi, Zhu Xianyi, He Yangti, Zheng Yan 0969

FMatchNet algorithm for fast clothing matching

Liu Yujie, Feng Shihe, Li Zongmin, Li Hua 0979

Infrared and visible image fusion based on BEMD and W-transform

Gong Rui, Wang Xiaochun 0987

Automatic vision-based deviation measurement method for cross sections of flexible sealing strips

Liao Lin, Li Jianhua 1000

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2019)06-0870-12

论文引用格式: Liang H G, Wen X Q, Liang D D, Li H D, Ru F. Fine-grained food image recognition of a multi-level convolution feature pyramid[J]. Journal of Image and Graphics, 2019, 24(06): 0870-0881. [梁华刚, 温晓倩, 梁丹丹, 李怀德, 茹锋. 多级卷积特征金字塔的细粒度食物图片识别[J]. 中国图象图形学报, 2019, 24(06): 0870-0881.] [DOI:10.11834/jig.180495]

多级卷积特征金字塔的细粒度食物图片识别

梁华刚, 温晓倩, 梁丹丹, 李怀德, 茹锋

长安大学电子与控制工程学院, 西安 710064

摘要: **目的** 食物图片具有结构多变、背景干扰大、类间差异小、类内差异大等特点,比普通细粒度图片的识别难度更大。目前在食物图片识别领域,食物图片的识别与分类仍存在精度低、泛化性差等问题。为了提高食物图片的识别与分类精度,充分利用食物图片的全局与局部细节信息,本文提出了一个多级卷积特征金字塔的细粒度食物图片识别模型。**方法** 本文模型从整体到局部逐级提取特征,将干扰较大的背景信息丢弃,仅针对食物目标区域提取特征。模型主要由食物特征提取网络、注意力区域定位网络和特征融合网络3部分组成,并采用3级食物特征提取网络的级联结构来实现特征由全局到局部的转移。此外,针对食物图片尺度变化大的特点,本文模型在每级食物特征提取网络中加入了特征金字塔结构,提高了模型对目标大小的鲁棒性。**结果** 本文模型在目前主流公开的食物图片数据集 Food-101、ChineseFoodNet 和 Food-172 上进行实验,分别获得了 91.4%、82.8%、90.3% 的 Top-1 正确率,与现有方法相比提高了 1%~8%。**结论** 本文提出了一种多级卷积神经网络食物图片识别模型,可以自动定位食物图片区分度较大的区域,融合食物图片的全局与局部特征,实现了食物图片的细粒度识别,有效提高了食物图片的识别精度。实验结果表明,该模型在目前主流食物图片数据集上取得了最好的结果。

关键词: 食物图片识别; 卷积神经网络; 注意力网络; 细粒度识别; 特征金字塔

Fine-grained food image recognition of a multi-level convolution feature pyramid

Liang Huagang, Wen Xiaoqian, Liang Dandan, Li Huaide, Ru Feng

School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China

Abstract: **Objective** Food images have special characteristics, uncertainties in food appearances, complex backgrounds, inter-class similarities, and intra-class differences. Hence, these images are more difficult to identify than ordinary fine-grained pictures. Traditional food image recognition mainly uses manual design features, including color, histogram of oriented gradient (HOG), and local binary pattern (LBP), then utilizes a classifier (e.g., support vector machine (SVM)) to deal with features. However, manual design features cannot establish the connection between various features. Several integrated feature methods only superimpose numerous features; thus, the recognition accuracy on each food image data set is up to 70% only. Compared with the weak expression capability of manual design features, deep learning has a stronger feature representation capability. They both use large-scale, labeled food images to train multi-level convolutional neural network models for food image recognition to improve recognition accuracy. However, in the current method of using the sonorous convolutional neural network for food image classification, the food image is directly inputted into the convolutional

收稿日期: 2018-08-22; 修回日期: 2018-12-04; 预印本日期: 2018-12-11

基金项目: 陕西省重点研发计划项目(2018ZDCXL-GY-05-03-02)

Supported by: Shaanxi Provincial Key Research and Development Program (2018ZDCXL-GY-05-03-02)

neural network to extract features. The food image has a relatively complicated background information, which critically influences the recognition result. We developed a model called multi-level convolution feature pyramid for fine-grained food image recognition to improve the accuracy of food image recognition and take full advantage of the local details. **Method** We extracted features from the whole to local, which not only avoids the shortcomings of baseline methods but also retains the global information and local details. We extracted features only from the target areas of the food image and discarded the background information with large interference. The multi-level convolution feature pyramid model consists of three main parts, namely, food feature extraction, attention localization, and feature fusion networks. The single-level feature extraction network cannot obtain the global and local features of the food image simultaneously. We developed a three-level food feature extraction network by cascading, which can transfer features from global to local. Moreover, a feature pyramid network was constructed between the feature maps of each food feature extraction network to deal with the large variation of food image scale. To automatically locate the network to the fine-grained area, an attention area localization network was designed between the levels of the feature extraction network, and the feature extraction range was reduced from global to local. Then, the fine-grained area of the original picture was cropped, enlarged, and inputted to the next-level feature extraction network. The features extracted by each level of the feature extraction network were subsequently sent to the feature fusion network. The merged features included the global features of the food image and the detailed features of the food target. For our model, two loss functions were used to optimize the feature extraction, feature fusion, and attention localization networks. For the feature extraction and feature fusion networks, the SoftMax loss function, which is referred to as the classification loss function, was used. The inter-stage loss function was utilized for the attention area positioning network.

Result We adopted step-by-step and alternating training methods to train the feature extraction and attention localization networks and the cascade model separately. We conducted experiments on current mainstream datasets of food images. Our model obtained the top accuracy rates with 91.4%, 82.8%, and 90.3% on Food-101, ChineseFoodNet, and Food-172 datasets, respectively. The implemented framework showed the best performance compared with baselines for food picture recognition, with 1%~8% improvement in recognition accuracy. Moreover, we trained the model in the Food-202 dataset, which we constructed ourselves, to verify the performance of our model fully. Food-202 is a food image dataset of 202 classes, and the number of food images in each class is more than 1 000; it includes Chinese and Western food. Results show that the accuracy of the model with the feature pyramid network increased by 2.4%. **Conclusion** We built a fine-grained food image recognition model with a multi-level feature pyramid convolutional neural network. The model can automatically locate areas with large discrimination of food images and integrate the global and local features of food images to achieve fine-grained recognition. It can effectively enhance the accuracy of food recognition and the robustness of the target size. Experimental results show that the proposed model demonstrated better performance than the baseline models in current mainstream food image datasets.

Key words: food picture recognition; convolutional neural network; attention network; fine-grained recognition; feature pyramid

0 引言

食物是保障人类生存与活动的重要物质来源,与人类日常生活关系密切。相比普通图片识别,食物图片具有形态复杂、颜色不固定、目标体积较小等特点。即使是同一种食物,由于地区等差异,也可能具有不同的特征。目前,利用计算机视觉技术进行食物图片识别已经成为一个热门研究领域。

早期的食物图片识别主要是先采用手工设计特征的方法,然后使用支持向量机(SVM)等分类器进

行分类^[1-5]。常用的手工设计特征包括颜色特征、LBP(local binary pattern)特征、HOG(histogram of oriented gradient)特征等简单的单一特征。由于手工设计特征无法建立各个特征之间的关联性,即使是一些集成特征方法,也只是将各种特征简单地叠加,所以在各个食物图片数据集上的识别准确率最高仅为70%左右。自2012年深度卷积神经网络在图片分类领域取得重大突破以来,将卷积神经网络应用于食物图片识别领域开始成为主流的方法,人们开始利用深度卷积神经网络提取食物图片特征并进行分类^[6-13]。当前使用深度卷积神经网络进行食物图

片分类的方法都是将食物图片直接输入卷积神经网络提取特征,由于食物图片往往具有较为复杂的背景,此时背景干扰信息就会对识别结果造成很大的影响。

除了以上的识别方法,也有一些关于细粒度识别的工作将食物图片作为一项实验结果进行对比^[14],但主要是针对 CUB-200-2011 数据集进行分类,没有针对食物图片的特点设计模型,也没有针对食物识别进行相关分析。

本文针对食物图片的特点,设计了一种多级卷积特征金字塔神经网络的细粒度食物图片识别模型,可以自动定位食物图片区分度较大的区域,融合食物图片的全局与局部特征,实现了食物图片的细粒度识别。该模型的核心是利用卷积神经网络自带注意力机制的特点,通过注意力区域定位网络定位食物图片细粒度区域。为了提高模型的识别效果,主要从两个方向进行了改进:

1) 通过特征金字塔网络提高细粒度区域的表示与定位能力。

对于相似度较高的食物图片,高区分度细粒度区域的定位尤为重要。本文采用特征金字塔结构,综合利用卷积神经网络浅层高分辨率结构信息与深层低分辨率语义信息,获得了更好的细粒度区域定位能力。

2) 设计特征融合网络,融合食物图片全局粗粒度信息与局部细粒度信息。

食物图片的细粒度信息虽然可以提高相似度较大的食物的区分能力,但是丢失了全局的结构信息,过度依赖细粒度信息反而会造成模型识别精度的下降。所以采用特征融合网络对各级食物特征提取网络提取到的特征进行融合,获得食物图片更加全面的特征信息。

1 模型设计

本文模型主要由食物特征提取网络、注意力区域定位网络和特征融合网络 3 部分组成,分别负责特征提取、细粒度局部区域定位和全局局部特征融合。由于单级食物特征提取网络无法同时获得食物图片全局与局部特征,因此采用 3 级食物特征提取网络级联的结构实现了特征由全局到局部的转移。针对食物图片尺度变化大的特点,在每级食物特征提取网络的特征图之间构建了特征金字塔网络,提

高了网络的特征描述能力,在加入特征金字塔后,模型获得了 2.4% 的性能提升。

图 1 是本文模型的总体结构图,图中有 3 级特征提取网络,分别是 Grade_1、Grade_2 和 Grade_3,每一级网络都加入了特征金字塔结构。AN1 和 AN2 分别代表注意力区域定位网络。每级特征提取网络之间是注意力区域定位网络。图片右边是特征融合网络,负责融合食物图片全局和局部特征。因为食物图片特征较为复杂,为了获取更加细节的特征,第 1 级网络的输入样本尺寸为 448×448 像素。为了获取更大的感受野,在 Grade_1 的第 1 层卷积层使用大小为 5×5 像素的卷积核,步长为 2 个像素,感受野达到 391×391 像素,相比使用大小为 3×3 像素卷积核感受野的 196×196 像素扩大了将近 4 倍。第 2 级和第 3 级特征提取网络的输入样本都是在原始图片上裁剪和放大得到。

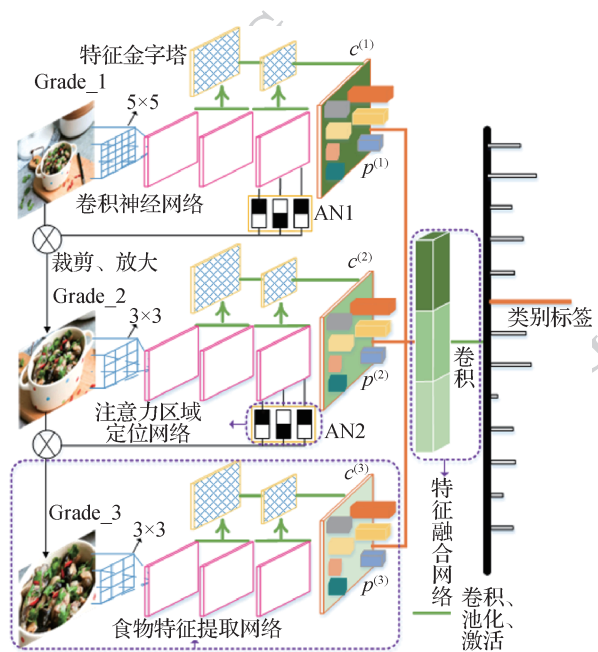


图 1 模型总体结构图

Fig. 1 The framework with three main parts

1.1 食物特征提取网络

图 1 中每一级食物特征提取网络(简称特征提取网络)都由全卷积结构组成,网络的结构类似于 VGG16,由 5 个卷积模块组成。同 VGG16 的命名方法一样,本文也采用 Conv1_2, Conv2_2, Conv3_3, Conv4_3, Conv5_3 分别命名每个卷积模块的最后一层特征。训练样本首先输入第 1 级特征提取网络,通过卷积神经网络提取图片特征,然后输出类别和

每一类的概率值。

$$p^{(i)} = \delta(\mathbf{X}_i \otimes \mathbf{W}_i) \quad (1)$$

$$C^{(i)} = \delta(\mathbf{X}_i \otimes \mathbf{W}_i) \quad (2)$$

式中,输入图片为 \mathbf{X} , \mathbf{X}_i 和 \mathbf{W}_i 表示第 i 级卷积神经网络的输入和权重参数, \otimes 表示在进行卷积过程中进行的一系列操作(卷积、池化、激活等), $\delta(\cdot)$ 表示特征提取网络。 $p^{(i)}$ ($i = 1, 2, 3$) 表示第 i 级卷积神经网络对于每一类别的输出概率, $C^{(i)}$ ($i = 1, 2, 3$) 表示第 i 级卷积神经网络的输出类别标签。

每一级网络都会生成关于输入图片的特征图, 用 F_j^i 表示每一级网络生成特征图的集合, i 表示每一级网络最后一层输出的特征图个数, j 表示模型包含的网络级数, 本文中 $j = (1, 2, 3)$, 因此输入图像 \mathbf{X} 可用以下特征图表示来描述

$$\{F_1^i, F_2^i, F_3^i\} \quad (3)$$

图2为特征金字塔详细结构图。卷积神经网络通过每一层的卷积核逐层提取特征, 不同卷积层的特征图对应图片的不同区域与目标。为了解决食物形状大小不一的问题, 同时为了更好地利用浅层卷积层的物体边缘信息, 本文在卷积神经网络各个卷积层之间构建了特征金字塔。由于卷积神经网络中浅层卷积层的尺度较大, 而深层卷积层尺度较小, 为了统一尺度, 首先将后一层的卷积层进行反卷积操作, 然后将两张特征图相加, 生成的特征图就是整张图片的最终表示。

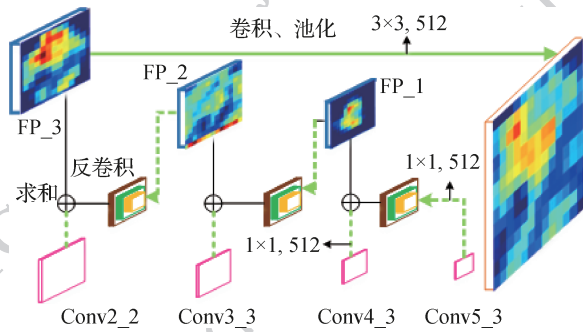


图2 特征金字塔结构图

Fig. 2 The detail of feature pyramid network

1.2 注意力区域定位网络

食物图片的全局信息可以提供食物的外观形状, 局部细粒度特征可以提供精细化的细节信息。为了实现食物从全局特征到局部细粒度特征的过渡与转移, 本文设计了注意力区域定位网络, 如图1所示, 注意力区域定位网络在特征提取网络之间, 负责

定位原始图片中的细粒度食物区域, 并将对应的区域裁剪送入下一级网络。目前大量的工作表明, 深度卷积神经网络如同人类观察图片的方式一样, 训练完成的网络可以自动将显著性区域(对识别效果有重要影响力的区域)激活^[15]。文献[16]证明了在深度卷积神经网络中, 特征图中响应最大的区域对应原始图片中对识别影响最大的区域。例如在行人识别时, 网络会自动将激活区域集中在人的脸部区域。所以精准地定位特征图中响应最大的区域并将其映射到原始图片, 即可得到原始食物图片的细粒度区域, 对细粒度区域进行特征提取实现图片的细粒度识别。本文采用逐级定位的注意力网络自动定位对识别具有重要影响力的区域。

注意力区域定位网络将输入的特征图映射为以 (t_x, t_y) 为中心、 t_a 为边长一半的正方形注意力区域, 并将注意力区域映射到原始图片, 进行裁剪、放大送入下一级网络, 实现食物图片由粗粒度图片到细粒度区域的精细化识别^[17]。

将特征图中高响应区域映射到原始图片的关键是获得高响应区域的坐标。由于神经网络具有很强拟合能力, 本文设计了一个2层全连接网络的注意力区域定位网络, 网络输入为特征图, 输出为高响应区域边界框的坐标, 如图3所示。通过该全连接网络可以实现高响应区域的自动定位。在高响应区域边界框生成的过程中, 为了避免边界框太大超出图片边界或太小无法有效定位物体, 本文对正方形的尺度进行了限制, 正方形边长不能大于图片尺寸边长的 $3/4$, 不得小于边长的 $1/6$ 。

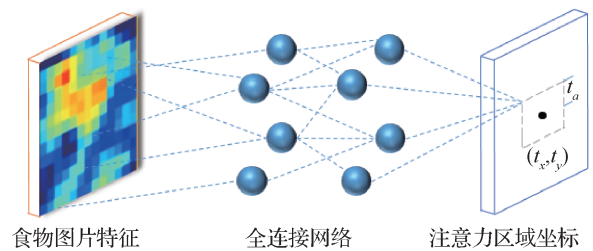


图3 注意力区域定位网络结构图

Fig. 3 The structure of attention localization neural network

$$[t_x, t_y, t_a] = \varphi(\mathbf{X} \otimes F_j^i) \quad (4)$$

式中, (t_x, t_y) 代表注意力区域正方形边界框的中心坐标, t_a 表示注意力区域正方形边界框边长的一半, $\varphi(\cdot)$ 表示注意力网络。

神经网络权重参数的初始化对网络的收敛性和

上的坐标映射回原始图片进行裁剪。

卷积神经网络在卷积过程中没有改变图片目标的空间位置信息,所以可以将特征图看做是原始图像的缩小。目标在特征图中的空间位置比例与目标在原始图片中的空间位置比例是一致的,因此依靠比例一致性的特点可以将特征图注意力区域映射到原始区域,即

$$\alpha = \frac{w_x}{w_f}, \quad \beta = \frac{h_x}{h_f} \quad (12)$$

式中, w_x 和 h_x 分别代表原始输入图片的宽和高; w_f 和 h_f 分别代表每一级特征提取网络最后一层提取到的图片特征图的宽和高; α 和 β 分别表示 x 轴方向和 y 轴方向的缩放因子。则注意力区域对应的原始图片区域坐标为

$$t_x^* = \alpha \times t_x, \quad t_y^* = \beta \times t_y \quad (13)$$

式中, (t_x, t_y) 为目标在特征图中的坐标信息, (t_x^*, t_y^*) 为目标在原始图中的坐标信息。

经过以上操作,可以在原始输入图片上裁剪注意力区域,为了获得更加细粒度的图片,提取更加精细的特征,本文将裁剪后的图片放大输入下一级网络。为了兼顾速度与精度,本文采用双线性插值算法。根据图像的几何性质,原始图片注意力区域与放大后的图片的比例关系为

$$\frac{p_{\text{Att}}^{(n,m)}}{p_{\text{Mag}}^{(i,j)}} = \frac{2t_a}{B_{\text{wide}}} \quad (14)$$

式中, $p_{\text{Mag}}^{(i,j)}$ 和 $p_{\text{Att}}^{(n,m)}$ 分别代表放大后的图片与原始图片注意力区域对应坐标的像素值; B_{wide} 代表放大后图片的宽度,因为是正方形,所以长和宽是相同的。根据双线性插值算法可知,放大图片任意一个坐标可通过式(14)计算对应原始图片的坐标。 $2t_a/B_{\text{wide}}$ 是原始图片注意力区域与放大后图片的放大比率,用 λ 表示,一般为小数,故原始图片的坐标会出现小数坐标,这与图片坐标实际情况不符合。将小数坐标分为整数部分与小数部分, $[\cdot]$ 和 $\{\cdot\}$ 分别表示坐标的整数部分和小数部分。放大后图片的像素横纵坐标对应的原始图片的横纵坐标为

$$\begin{aligned} p_{\text{Att}}^{(n)} &= p_{\text{Mag}}^{(i)} \frac{2t_a}{B_{\text{wide}}} = p_{\text{Mag}}^{(i)} \times \lambda \\ [p_{\text{Mag}}^{(i)} \times \lambda] &= z, \quad (p_{\text{Mag}}^{(i)} \times \lambda) = u \\ p_{\text{Att}}^{(m)} &= p_{\text{Mag}}^{(j)} \frac{2t_a}{B_{\text{wide}}} = p_{\text{Mag}}^{(j)} \times \lambda \end{aligned} \quad (15)$$

$$(p_{\text{Mag}}^{(j)} \times \lambda] = c, \quad (p_{\text{Mag}}^{(j)} \times \lambda) = v \quad (16)$$

放大后图片的像素值为

$$\begin{aligned} p_{\text{Mag}}^{(i,j)} &= (1-u)(1-v)p_{\text{Att}}^{(z,c)} + uv p_{\text{Att}}^{(z+1,c+1)} + \\ &u(1-v)p_{\text{Att}}^{(z+1,c)} + v(1-u)p_{\text{Att}}^{(z,c+1)} \end{aligned} \quad (17)$$

将式(17)变形,可得

$$\begin{aligned} p_{\text{Mag}}^{(i,j)} &= \sum \left| 1 - \alpha - \left(\frac{i}{\lambda} \right) \right| \times \\ &\left| 1 - \beta - \left(\frac{j}{\lambda} \right) \right| p_{\text{Att}}^{([\frac{i}{\lambda}] + \alpha, [\frac{j}{\lambda}] + \beta)} \end{aligned} \quad (18)$$

根据以上所述就可利用双线性插值获得注意力区域的放大图片。

1.3 特征融合网络

模型每一级特征提取网络都会提取食物图片的特征,第1级特征提取网络提取食物图片的全局特征,第2级网络提取食物图片的局部特征,第3级网络提取食物图片中食物目标更加细节的特征。将每一级网络提取的特征按通道对特征图中的每个元素相加,最终得到256个 28×28 像素的特征图。然后使用 1×1 像素大小卷积核进行特征融合和池化,接着使用两个 3×3 像素大小的卷积核进行卷积操作和激活函数激活,最后进行池化,得到512个 7×7 像素的特征图,该特征既包含食物的全局特征,也包含食物目标的局部特征。特征融合网络可以融合食物图片的特征,也可以通过池化将之前 28×28 像素的高分辨率特征图降低到 7×7 像素,减少了模型的参数。

1.4 损失函数设计

按照上述方法设计好网络后,为了同时优化特征提取网络和注意力区域定位网络,本文采用了两种损失函数优化网络。使用分类损失函数优化每一级特征提取网络,级间损失函数用来优化特征提取网络之间的注意力区域定位网络。

对于每一级特征提取网络,本文采用平方损失函数,具体为

$$L_{\text{cls}}^i = \sum_k \frac{1}{2} |C^k - C^*|^2 + \lambda \|w\|^2 \quad (19)$$

式中, i 表示第 i 级特征提取网络, $C^* = 1$ 表示真实类别标签, $C^k \in (0, 1]$ 表示预测类别标签, λ 表示正则化因子。分类损失函数可以使每一级分类网络获得更加优异的特征提取能力。

分类损失函数只能优化每一级特征提取网络,无法优化每一级特征提取网络之间的注意力网络。

为了更加精确地定位高分辨率区域,本文采用级间损失函数优化每一级之间的注意力网络。级间损失函数^[19]为

$$L_{\text{grade}}^i = \max\{0, p^i - p^{i+1} + \Delta\} \quad (20)$$

式中, Δ 表示 p^i 与 p^{i+1} 之间的间隔, $i = 1, 2$ 。当 $p^i < p^{i+1}$, 且二者的差值超过间隔值时, 损失为零, 当 $p^i > p^{i+1}$, 且差值超过间隔值时, 产生损失。由此可知, 级间损失函数可以促使下一级特征提取网络每一类输出概率值大于上一级特征网络输出概率值, 即级间损失能够促进特征提取网络的特征学习能力, 而特征提取网络在增强特征提取能力后又可提供更加精准的注意力区域定位, 二者互相促进, 不断提高网络的性能。

当计算出类别损失与级间损失后, 就可得到模型的总损失函数, 具体为

$$L = \gamma \sum_{i=1}^3 L_{\text{cls}}^i + (1 - \gamma) \sum_{i=1}^2 L_{\text{grade}}^i \quad (21)$$

式中, 权重因子 $\gamma \in (0, 1)$, 用来平衡两种类型的损失。模型中包括裁剪在内的操作全部是可导的, 可以实现端到端的训练。

2 实验结果与分析

本文选取的数据集包括 Food-172^[11]、Chinese-FoodNet^[12]、UEC-100^[20] 和 Food-101^[21]。此外, 为了使模型得到更加充分的训练, 本文构建了一个类别为 202 类的食物图片数据集, 每类食物图片数量大于 1 000 张, 包括中国食物和西方食物, 称为 Food-202。数据集的部分样本如图 5 所示。



图 5 Food-202 数据集部分样本

Fig. 5 Partial samples of Food-202 dataset

2.1 模型训练

本文采用分步训练和交替训练的方法, 分别训练特征提取网络、注意力区域定位网络和级联模型。

1) 分步训练。首先在食物数据集上训练特征提取网络。用特征提取网络提取训练样本的特征作

为注意力区域定位网络的输入, 计算每一个训练样本特征注意力区域的坐标信息, 并将此作为注意力区域定位网络的监督信息, 然后训练注意力区域定位网络。网络的具体训练细节如表 1 所示。表 1 中的每一行代表模型中的 1 个子网络, 第 2 列和第 3 列分别表示网络的输入和输出, 参数的初始化采用 xavier 方式。本文使用了常用的数据增强方法, 包括随机裁剪、添加噪声、平移、旋转等。在训练模型时, 在保持长宽比不变的情况下, 本文将训练图片的短边调整为 512 像素。

表 1 模型训练细节

Table 1 Training details for our model

模型	输入	输出
Grade_1	448 × 448 像素 训练样本	类别概率值, 图片的特征向量
Grade_2	224 × 224 像素 训练样本	类别概率值, 图片的特征向量
Grade_3	224 × 224 像素 训练样本	类别概率值
AN1	高响应区域坐标值, Grade_1 输出的图片 的特征向量	高响应区域坐标值
AN2	高响应区域坐标值, Grade_2 输出的图片 的特征向量	高响应区域坐标值

2) 交替训练。当训练好独立的特征提取网络与注意力区域定位网络后, 开始交替训练。设置 $\text{Margin} = 0.005$, 注意力模板生成函数的 $k = 20$ 。本文首先固定特征提取网络参数, 训练注意力区域定位网络; 接着固定注意力区域定位网络参数, 训练特征提取网络, 如此交替训练直到收敛。学习率为 0.001, 迭代次数为 50 000 次, 正则化系数为 0.1, 使用随机梯度下降 (SGD) 优化器, 特征提取网络采用 xavier 初始化的方式。

2.2 实验分析

目前关于食物图片识别的文献很多, 各个文献之间没有统一的检验标准。为了验证本文设计模型的性能, 且使对比结果可靠科学、公平合理, 本文分别在文献[11]和[12]使用的食物图片数据集上进行训练与测试。训练集和测试集的分割与原始数据集提供的完全相同, 具体结果如表 2 所示。

表2 本文模型与其他模型在 Food-172 和 ChineseFoodNet 数据集的比较

Table 2 Comparisons between our model and other models on Food-172 and ChineseFoodNet datasets

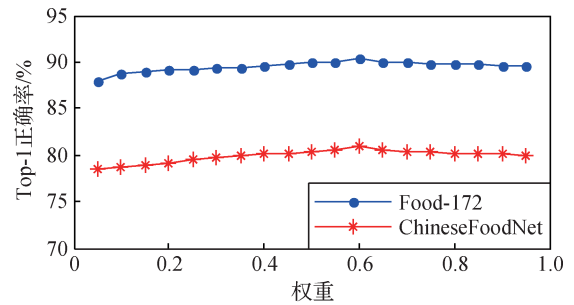
模型	数据集	Top-1 正确率 /%
Arch-D ^[11]	Food-172	82.12
本文模型(2级)	Food-172	90.30
本文模型(3级)	Food-172	88.20
TastyNet ^[12]	ChineseFoodNet	81.55
本文模型(2级)	ChineseFoodNet	80.70
本文模型(3级)	ChineseFoodNet	82.80

注:加粗字体为最优结果。

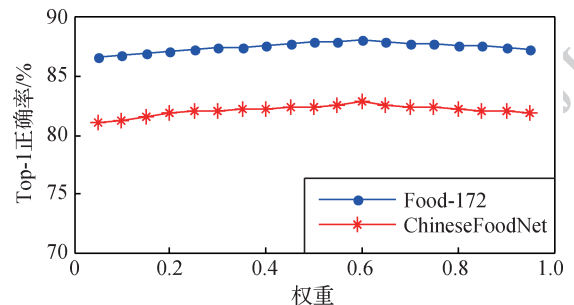
表2中前3行对比了本文模型与文献[11]中的模型在 Food-172 数据集上的 Top-1 正确率。由表2可知,在使用2级模型时达到最高的 Top-1 正确率 90.3%,提高了 8.18%,即使是3级模型也达到了 88.2% 的识别精度。后3行与文献[12]提出的模型在 ChineseFoodNet 数据集上进行了比较,精度提高了 1.25%。

图6分析了全局损失函数取不同权重因子时 Top-1 正确率的变化。实验结果表明,当 $\gamma = 0.6$ 时,本文模型在 Food-172 和 ChineseFoodNet 数据集上的 Top-1 正确率最高。因此,全局损失函数的权重因子 $\gamma = 0.6$ 。

此外,本文同时在食物图片识别领域与经常使用的食物图片数据集 Food-101 进行了比较,并取得了最好的结果,具体如表3所示。目前在 Food-101 数据上的识别率最高达到 90.27%,该模型使用了基于残差算法的宽残差网络。本文采用以注意力网络为基础的多级卷积神经网络识别模型,在只有两级网络的情况下达到 89.5% 的正确率,在3级网络的时候达到 91.4% 的正确率,超过了其他所有模型。文献[14]采用基于注意力机制在同一张图片上生成多个注意力区域,然后对每个区域提取特征进行分类,达到了 86.5% 的正确率,低于本文的 91.4%,主要原因是食物图片与鸟类图片的结构完全不同,没有清晰、明显的物理结构。同一张食物图片上多个区域的特征相关性很高,造成特征利用效率降低。同时也证明了本文设计的模型可以较好地定位食物目标细粒度区域。



(a) 本文模型(2级)



(b) 本文模型(3级)

图6 权重因子 γ 对 Top-1 正确率的影响

Fig. 6 The effect of weights on the Top-1 accuracy

- ((a) model with two feature extraction networks;
(b) model with three feature extraction networks)

表3 本文模型与其他模型在 Food-101 数据集的比较

Table 3 Comparisons between our model and other models on Food-101 dataset

模型	主干网络/层数	Top-1 正确率 /%
文献[7]	GoogLeNet/22	74.75
文献[8]	集成模型	72.12
WISeR ^[10]	—	90.27
文献[14]	—	86.50
CNNs Fusion(S2) ^[9]	—	86.07
本文模型(2级)	VGG/13	89.50
本文模型(3级)	VGG/13	91.40

注:加粗字体为最优结果。

为了更加全面地验证模型的性能,在本文构建的 Food-202 数据集对模型进行了训练。首先将数据集按照 2:8 的比例划分为测试集和训练集,然后从训练集中划分 20% 作为验证集。与之前相同,模型中的特征提取网络也在 ILSVRC 数据集上进行了预训练,具体结果如表4所示。

表4 模型在 Food-202 数据集的识别率
Table 4 The accuracy of our model on Food-202

模型	Top-1 正确率 /%
本文模型(无特征金字塔2级)	81.0
本文模型(2级)	83.4
本文模型(3级)	81.6

注:加粗字体为最优结果。

表4的实验结果表明,在没有特征金字塔的情况下,模型识别正确率下降了2.4%。特征金字塔对模型的主要影响体现在注意力区域的定位精度方面。从表4可以看出,本文模型在3级网络的情况下获得了81.6%的正确率,低于2级网络的83.4%。这是因为本文模型是通过不断定位高区分度区域,然后进行裁剪、放大注意力区域对应的目标送入下一级网络进行识别,当食物在图片中所占的比例较小时,过多级数会使注意力网络不断地裁剪图片,反而会造成食物信息的丢失,从而造成识别正确率下降。

为了验证平方损失函数的有效性,本文模型在不同的数据集上对特征提取网络采用平方损失函数与交叉熵损失函数时相应的Top-1正确率进行了对比,如表5所示。实验结果表明,本文模型的特征提取网络采用平方损失函数进行优化时,Top-1正确率的整体表现比交叉熵损失函数高0.6%~2.5%。

表5 平方损失函数与交叉熵损失函数的 Top-1 正确率对比
Table 5 Comparison of Top-1 accuracy between squared loss function and cross entropy loss function

数据集	本文模型(2级)		本文模型(3级)	
	平方损失函数	交叉熵损失函数	平方损失函数	交叉熵损失函数
Food-172	90.3	89.7	88.2	86.4
ChineseFoodNet	80.7	80.9	82.8	80.3
Food-101	89.5	88.3	91.4	90.7
Food-202	83.4	82.6	81.6	80.1

2.3 区域定位与特征金字塔的有效性

在本文模型中,每一级特征提取网络提取食物图片特征,注意力网络负责定位细粒度区域。为进一步证明注意力网络的有效性,本文选取 UEC-100

食物图片数据集作为训练样本。UEC-100 数据集的每张图片中的食物目标都有人工标记的边界框,便于进行评估,本文模型在训练时仅使用了图片的类别标签信息,没有使用任何的边界框标注信息。但是 UEC-100 数据集存在一个严重问题,该数据集样本量极少。由于本节主要研究注意力区域网络的定位能力,所以选取了 UEC-100 数据集中数量最高的4类(green salad, miso soup, rice, ramen noodle)作为训练样本。按照数据集中的分割方法将数据集分为训练集和测试集,在 UEC-100 数据集上,加入特征金字塔结构的模型在3级网络时,达到了81.4%的Top-1正确率。

为了获取注意力区域定位网络输出的注意力区域坐标,本文将图片输入模型前向传播,提取第1个注意力区域定位网络输出的坐标信息。注意力区域信息的可视化结果如图7所示。

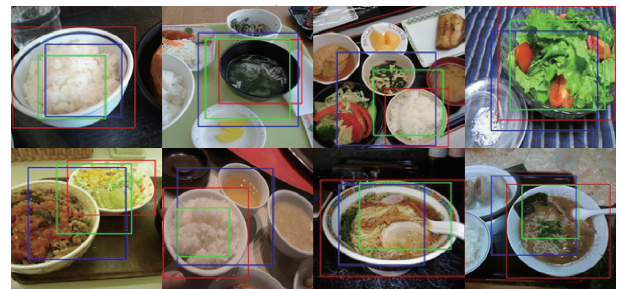


图7 注意力区域定位网络定位结果

Fig. 7 The result of the attention localization network

图7中红色框为数据集标注的标准框,蓝色框是没有特征金字塔结构的特征提取网络定位的细粒度区域,绿色框是使用特征金字塔结构的特征提取网络定位的细粒度区域。

由图7可以发现,使用特征金字塔的模型定位的食物图片细粒度区域包含更少的背景信息,对食物目标的覆盖能力较为精准。这也解释了表4中的结果,说明使用特征金字塔的模型识别结果优于不使用特征金字塔的模型,主要原因是特征金字塔使食物细粒度区域的定位能力更加准确。观察图片还可以发现另外一个问题,当同一张食物图片中有多个食物目标时,注意力网络的定位能力会下降,如图7第1行中的第3幅图片。因为注意力网络仅输出一个区域的中心坐标与边长,当图片中有多个目标时,需要多个注意力区域坐标,所以出现了该种情况。这种现象也引出了另外一个问题,即注意力网

络的多目标定位能力。

为了对结果进行统一评价,选取目标检测领域的常用指标 IOU(intersection over union)作为目标定位能力的评价标准,IOU 的计算方法为

$$IOU = \frac{G \cap At}{G \cup At} \quad (22)$$

式中, G 表示数据集中给定的边界框区域, At 表示注意力网络输出的区域。

IOU 的取值范围为 $[0,1]$,当 IOU 接近 0 时,说明两个框几乎没有重叠。当 IOU 接近 1 时,说明两个框几乎重叠。但是对于本文,该指标存在一个问题。由于在 UEC-100 数据集中边界框内往往包含图片背景或其他非食物目标(例如食物容器),而注意力网络的目标是定位出食物目标的细粒度区域,该区域并不是全部与数据集的标准框完全吻合。一般而言,数据集中标准框区域要大于注意力网络输出的区域。此时又会出现另外的情况,当注意力网络定位区域大于数据集的标准框区域,此时 IOU 越接近 0,定位能力反而是越差。基于以上分析,对于本文 IOU 的衡量标准是,尽量使 IOU 值小。但是注意力区域 IOU 不可以太小,太小会造成食物信息的丢失或定位区域噪声太大。同时 IOU 也不能太大,太大也会加入噪声。同上一部分一样,本文选取第 1 个注意力网络的输出参数,具体结果如表 6 所示。表 6 中第 4 列和第 5 列是注意力网络输出定位区域的面积。通过实验表明,IOU_FP 和 IOU_NFP 之间的关系与图 7 有着密切关系,当图片中只包含 1 个目标时,IOU_FP < IOU_NFP,如表 6 中的第 3 行和第 4 行所示。

表 6 注意力网络 IOU 指标评价

Table 6 IOU indicator evaluation
for attention neural network

类别	IOU_FP	IOU_NFP	注意力区域 面积(FP)	注意力区域 面积(NFP)
Green salad	0.488	0.254	204 ²	293 ²
Miso soup	0.609	0.526	249 ²	296 ²
Rice	0.336	0.448	197 ²	227 ²
Ramen noodle	0.343	0.497	211 ²	254 ²

表 6 中,IOU_FP 表示每一级特征提取网络加入特征金字塔结构,IOU_NFP 表示没有加入特征金字塔结构。

当图片中只有一个目标时,注意力区域小于数据集标准框,此时,加入特征金字塔结构的模型具有更好的定位能力,如图 7 中第 1 行第 1 张图片和第 2 行第 3 张图片。当图片包含多个目标时,注意力区域与数据集标准框部分重叠,此时 IOU 值越大,说明重叠值越大,越接近数据集标准框,定位性能越好。如表 6 中第 1 行和第 2 行所示,IOU_FP > IOU_NFP,对应的图片为图 7 中的第 1 行第 3 张、第 2 行第 1、第 2 张。通过以上分析,证明了特征金字塔对细粒度区域定位的有效性,这也是加入特征金字塔结构的模型识别正确率较高的原因。从表 6 中的第 4 列和第 5 列可知,有特征金字塔结构的模型的注意力区域面积都小于没有特征金字塔结构模型注意力区域的面积。

图 8 为注意力区域裁剪放大效果图。图 8(c)是注意力区域过度裁剪的情况,此时全局信息丢失较为严重,这也是 3 级模型正确率低于 2 级模型的原因。



图 8 注意力区域裁剪放大效果图

Fig. 8 Attention area cropping enlargement effect diagram

((a) original pictures; (b) attention areas;
(c) attention areas enlargement)

3 结论

本文采用以注意力机制为基础的多级卷积神经网络模型对食物图片进行细粒度的识别。为了适应食物图片形状大小变化较大的特点,在每级网络中

加入了特征金字塔结构,利用各个卷积层中不同尺度的特征图,使模型对目标大小的鲁棒性更高。卷积神经网络中各层特征图分别表示不同的特征,由简单到复杂逐步递进。为了避免所有特征图的同一化,本文采用反卷积的方法将深层低分辨率的特征图与前一层较高分辨率的特征图归一化到同一尺度,然后进行融合。在每级特征提取网络提取食物图片特征后,通过特征融合网络将食物图片的全局特征与局部特征融合。实验结果表明,本文模型在主流食物图片数据集上获得了最好的结果。然而,本文模型也存在一些问题。本文模型对多目标食物图片的定位效果差,而且,特征提取网络的技术需要手动设计,有时会出现过度裁剪的问题造成食物目标结构信息的丢失,导致模型识别正确率下降,解决这些问题是下一步工作的重点。

参考文献 (References)

- [1] Liu C, Cao Y, Luo Y, et al. Deepfood: deep learning-based food image recognition for computer-aided dietary assessment [C] // Proceedings of the 14th International Conference on Smart Homes and Health Telematics. Wuhan: Springer, 2016: 37-48. [DOI: 10.1007/978-3-319-39601-9_4]
- [2] He H S, Kong F Y, Tan J D. DietCam: multiview food recognition using a multikernel SVM [J]. IEEE Journal of Biomedical and Health Informatics, 2016, 20(3): 848-855. [DOI: 10.1109/JBHI.2015.2419251]
- [3] Kawano Y, Yanai K. FoodCam: a real-time food recognition system on a smartphone [J]. Multimedia Tools and Applications, 2015, 74(14): 5263-5287. [DOI: 10.1007/s11042-014-2000-8]
- [4] Martinel N, Piciarelli C, Micheloni C. A supervised extreme learning committee for food recognition [J]. Computer Vision and Image Understanding, 2016, 148: 67-86. [DOI: 10.1016/j.cviu.2016.01.012]
- [5] Chen M Y, Yang Y H, Ho C J, et al. Automatic Chinese food identification and quantity estimation [C] // Proceedings SA'12 SIGGRAPH Asia 2012 Technical Briefs. Singapore: ACM, 2012. [DOI: 10.1145/2407746.2407775]
- [6] Zhao W T, Wang Y H, Chen X X, et al. Learning deep feature fusion for group images classification [C] // Proceedings of the 2nd Chinese Conference on Computer Vision. Tianjin: Springer, 2017: 566-576. [DOI: 10.1007/978-981-10-7302-1_47]
- [7] Bolaños M, Radeva P. Simultaneous food localization and recognition [C] // Proceedings of 2016 International Conference on Pattern Recognition. Cancun, Mexico: IEEE, 2016: 3140-3145. [DOI: 10.1109/ICPR.2016.7900117]
- [8] Pandey P, Deepthi A, Mandal B, et al. FoodNet: recognizing foods using ensemble of deep networks [J]. IEEE Signal Processing Letters, 2017, 24(12): 1758-1762. [DOI: 10.1109/LSP.2017.2758862]
- [9] Aguilar E, Bolaños M, Radeva P. Food recognition using fusion of classifiers based on CNNs [C] // Proceedings of the 19th International Conference on Image Analysis and Processing. Catania, Italy: Springer, 2017: 213-224. [DOI: 10.1007/978-3-319-68548-9_20]
- [10] Martinel N, Foresti G L, Micheloni C. Wide-slice residual networks for food recognition [J]. arXiv: 1612.06543, 2016.
- [11] Chen J J, Ngo C W. Deep-based ingredient recognition for cooking recipe retrieval [C] // Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands: ACM, 2016: 32-41. [DOI: 10.1145/2964284.2964315]
- [12] Chen X, Zhu Y, Zhou H, et al. ChineseFoodNet: a large-scale image dataset for chinese food recognition [J]. arXiv: 1705.02743, 2017.
- [13] Hassannejad H, Matrella G, Ciampolini P, et al. Food image recognition using very deep convolutional networks [C] // Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. Amsterdam, The Netherlands: ACM, 2016: 41-49. [DOI: 10.1145/2986035.2986042]
- [14] Liu X, Xia T, Wang J, et al. Fully convolutional attention networks for fine-grained recognition [J]. arXiv: 1603.06765, 2016.
- [15] Zhou B L, Khosla A, Lapedriza A, et al. Object detectors emerge in deep scene CNNs [C] // Proceedings of 2015 International Conference on Learning Representations. arXiv: 1412.6856, 2015.
- [16] Bau D, Zhou B L, Khosla A, et al. Network dissection: quantifying interpretability of deep visual representations [C] // Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: IEEE, 2017: 3319-3327. [DOI: 10.1109/CVPR.2017.354]
- [17] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916. [DOI: 10.1109/TPAMI.2015.2389824]
- [18] Wei X S, Luo J H, Wu J X, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881. [DOI: 10.1109/TIP.2017.2688133]
- [19] Zhang X F, Zhou F, Lin Y Q, et al. Embedding label structures for fine-grained feature representation [C] // Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1114-1123. [DOI: 10.1109/

CVPR. 2016. 126]

- [20] 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. IEEE, 2015: 1-6. [DOI: 10. 1109/ICMEW. 2015. 7169816]
- [21] Bossard L, Guillaumin M, Gool L V. Food-101: Mining discriminative components with random forests [C]. Computer Vision- ECCV2014. Zurich, Switzerland: Springer, 2014. [DOI: 10. 1007/978-3-319-10599-4_29]

作者简介



梁华刚, 1980 年生, 男, 副教授, 主要研究方向为机器视觉和智能交通。
E-mail: hgliang@chd.edu.cn



温晓倩, 通信作者, 女, 硕士研究生, 主要研究方向为模式识别与智能系统。
E-mail: wenxiaoqian@chd.edu.cn

梁丹丹, 女, 硕士研究生, 主要研究方向为模式识别与智能系统。E-mail: 1261523826@qq.com

李怀德, 男, 硕士研究生, 主要研究方向为模式识别与智能系统。E-mail: lihuaide@chd.edu.cn

茹锋, 男, 教授, 主要研究方向为图像处理与机器视觉、车联网技术及嵌入式系统、多源信息融合技术。

E-mail: fengru@chd.edu.cn