DOI:10.13878/j.cnki.jnuist.2019.06.013

李辉[1,2]　闵巍庆[2]　王致岭[2,3]　彭鑫[1]

基于图卷积网络的多标签食品原材料识别

摘要

当前,食品图像的营养成分识别主要还是集中在食品类别的识别以及作为多标签任务的识别.但是这两种方法并不具备很好的判别性,因为它们忽略了原材料之间的潜在关系.因此,本文在前期工作的基础上引入了原材料之间的关系.具体地说,我们的工作主要分为图像特征提取和原材料关系学习两部分.图像特征提取通过卷积神经网络提取到图像的低维特征向量.图卷积网络通过使用图数据(图的每个节点表示原材料的词嵌入,边表示节点之间的相关性),将图数据直接映射到一组相互依赖的分类器中,并与图像的低维特征向量融合,最后进行分类.通过在 Food-101 和 VireoFood-172 两个食品数据集上进行实验,并与当前最好的实验模型进行对比,发现基于图卷积的食品多标签分类方法可以有效地提高食品图像的分类性能.

关键词

多标签分类;食品原材料;食品图像;卷积神经网络;图卷积网络

中图分类号 TP391.41
文献标志码 A

收稿日期 2019-09-22
资助项目 国家自然科学基金(61772195,61602437);湖南省自然科学基金(2018JJ2156);湖南省学位与研究生教育教改研究课题(JG2018B119);湖南省"十三五"教育科学规划课题(XJK17BXX004)

作者简介

李辉,男,硕士生,研究方向为物联网技术.huili@vipl.ict.ac.cn

彭鑫(通信作者),男,博士,副教授,硕士生导师,研究方向为物联网技术.pengxin@hnist.edu.cn

1 湖南理工学院 信息科学与工程学院,岳阳,414006
2 中国科学院计算技术研究所 智能信息处理重点实验室,北京,100190
3 中国科学院大学,北京,100049

0 引言

随着人们生活水平不断提高,人们对食品的营养成分越来越重视,所以食品的营养成分分析成为食品和健康等领域最有前途的应用之一.通过识别食品图像中的原材料已了解食品所含营养成分已受到了越来越多研究者的关注.

随着深度学习理论的提出和计算设备的改进,卷积神经网络(Convolutional Neural Networks,CNN)得到了飞速的发展,并且被广泛应用于图像识别领域.例如文献[1]通过 AlexNet 网络来提取图像的视觉特征,从而实现图像的检测和分类;文献[2]通过 GoogLeNet 网络来提取图像视觉特征,从而实现食品图像和非食品图像的分类.文献[3]通过 Faster R-CNN 网络来提取图像的视觉特征.此外,还有近期一些工作[4]采用了多尺度多视图特征聚合方案,经过最后的二级融合,使得特征更加健壮、全面和具有判别力.但是上述研究对食品图像的分类都是基于单标签分类,需要根据得到的食品图像来推测其所含的原材料信息,这种方法并不能有效地判别食品的营养成分.所以本文提出了使用图卷积网络(Graph Convolutional Networks,GCN)的多标签模型[5]来直接识别食品图像中的原材料,以便可以更好地了解食品的营养成分,更好地规划我们的饮食.在现实生活中,虽然菜肴不计其数,但是它们大多由少量几种原材料组成,所以与传统食品图像识别不同的是,通过直接识别食品图像中的原材料,可以将其推广到系统从未见过的数据集,解决例如文献[6]数据集类别不多的问题.本文模型的整体框架主要分为图像特征提取和原材料关系学习.在图像特征提取中,首先通过 CNN 得到食品图像的特征图,然后通过降维,得到一组低维的特征向量.在原材料关系学习中,通过 GCN 学习原材料之间的关系,并将学到的原材料关系与得到低维图像特征向量进行融合,然后进行食品原材料的分类.

本文的主要贡献包括以下两个方面:1)本文首次将 GCN 的多标签模型应用于食品图像的多标签分类任务中;2)通过使用食品图像及其相关成分的数据集,验证了模型的有效性.

本文的组织结构如下:第 1 节介绍食品图像领域的最新技术;第 2 节介绍基于 GCN 的多标签模型在食品领域的应用;第 3 节对实验结果进行定性定量分析;第 4 节对本文工作进行总结与展望.

1 相关工作

本节将对近年来食品图像分类相关技术和方法进行介绍,主要包括两部分:第 1 部分主要介绍食品图像的分类及相关技术方法;第 2 部分主要介绍基于 GCN 的多标签分类.

1.1 食品图像分类

随着人们生活水平的提高,食品健康问题受到了越来越多的研究者关注.例如文献[7]采用了食谱的成分、食物图像和属性来执行第一个跨区域的食谱分析.文献[8]采用了谷歌的深度卷积网络 Inception 提取图像的视觉特征实现食品图像的分类.此外,更有研究者综述了计算机技术在食品领域的应用,并提供了一系列的研究方法,包括食品图像的检索、分类等[9].与上述文献对食品图像的单标签分类相比,食品图像原材料的多标签分类要困难得多,即每输入一个食品图像样本,就可以预测一个以上的原材料类别.原因是食品图像中的原材料尺寸较小,外观可能出现较大差异[10].

尽管食品图像的原材料识别相对于食品图像的识别要困难得多,但是随着研究者们在食品领域的深入探索,食品图像的原材料识别已经有了较大的进步.例如文献[6]提出一种 VGG 多任务学习框架,可用于同时识别食品类别和成分标签.文献[11]提出采用 CNN 模型,用 Sigmod 激活函数替换最后的 Softmax 激活函数进行食品图像的原材料分类.这两种方法在对食品图像的原材料分类中有比较好的效果,但是却忽略了原材料之间的关系.本文使用文献[6]的方法,将 GCN 的多标签模型应用到食品图像的多标签分类中.该方法考虑了食品原材料之间的相关关系,并采用得到的原材料关系来辅助完成多标签分类任务,最后将获得的视觉特征向量和原材料关系融合进行分类.

1.2 基于 GCN 的多标签分类

随着计算机的计算能力的提升以及大型数据集,如 ImageNet[12]、MSCOCO[13] 和 PASCALVOC[14] 的出现,图像的多标签识别得到了快速的发展.一种图像的多标签识别直接方法是为每个类或者标签训练独立的二进制分类器.但是,此方法没有考虑标签之间的关系,并且随着类别数量的增加,预测标签的数量也呈指数增长.例如文献[15]使用基于排名的学习策略训练深度 CNN 进行多标签图像识别,并发现加权近似排名的损失最有效.文献[16]利用循环神经络将标签转换为嵌入标签向量形式,从而利用标签之间的相关性.文献[17]引入空间变换层和 LSTM 单元来获取标签之间的相关性.文献[18]使用最大生成树算法获取标签中的相关性.文献[19]基于图形化的套索框架产生了图像相关的条件标签结构.文献[20]结合图卷积网络,描述了多个标签之间的关系.

虽然上述方法都引入了多标签之间的关系,但尚无人将此方法运用到食品图像的领域中.本文利用食品原材料之间的关系,然后利用图结构来捕获和探索原材料之间相关性.具体来说,基于该图结构,利用图卷积网络在多个原材料之间传播信息,为每个图像的原材料学习相互依赖的分类器.这些分类器可以从原材料图中吸收信息,然后将这些信息进一步应用于最终多标签预测的全局图像表示.实验结果验证了基于图卷积网络的多标签模型相对食品图像的多标签分类的有效性,整个过程为端到端的训练.

2 基于图卷积网络的食品图像原材料分类

本文利用文献[6]中基于图卷积网络的多标签模型框架进行食品多标签的识别,如图 1 所示,主要分为两部分:食品图像的特征提取以及原材料关系学习.在食品图像的特征提取中,首先利用卷积神经网络得到食品图片的局部特征,然后经过最大池化进行降维,最后得到一组 D 维的特征向量.在原材料关系学习模块中,首先通过原材料之间存在的相关性,构建出原材料的关系图,然后利用图卷积网络将原材料关系图直接映射到一组相互依赖的分类器中,并与图像的 D 维特征向量融合,最后进行分类.其中图的每个节点表示原材料的词嵌入,边表示节点之间的相关性.图 1 中,d 表示原材料特征的维度,C 表示原材料的类别数量.

2.1 食品图像的特征提取

图 1 中采用卷积神经网络(如 ResNet50、ResNet101)为基本模型对食品图像进行特征提取.当输入食品图像时,首先获取经过卷积操作后的特征图,然后通过池化得到一组低维的食品图像特征向量.其表达式如下:

$$\boldsymbol{x} = V_{\mathrm{GMP}}(F_{\mathrm{cnn}}(I;\theta_{\mathrm{cnn}})) \in \mathbf{R}^{D} \tag{1}$$

式(1)中,I 表示输入的食品图片,θ_{cnn} 表示卷积神经网络模型参数,$F_{\mathrm{cnn}}(\cdot)$ 表示在卷积神经网络中提取到的食品图像的特征图函数,$V_{\mathrm{GMP}}(\cdot)$ 表示的是

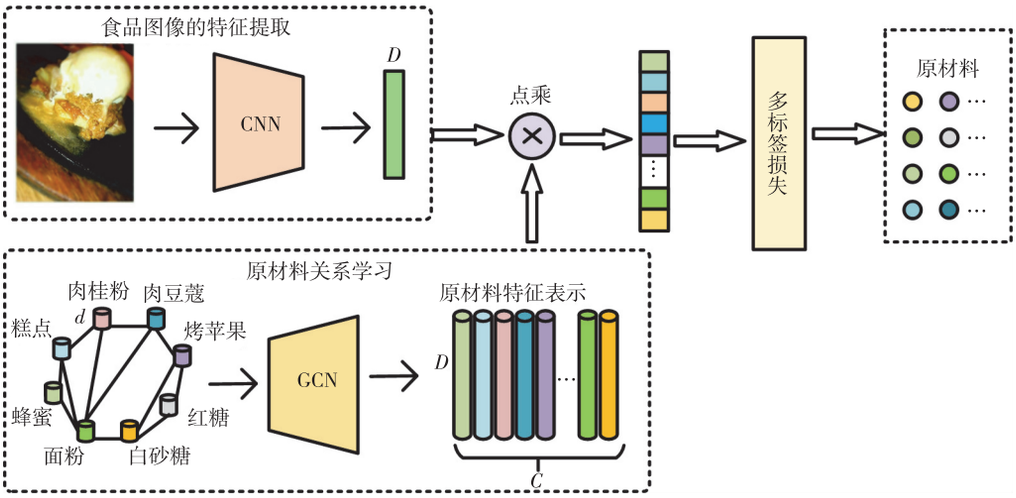

图 1　图卷积网络的多标签模型框架

Fig. 1　Multi-label model framework for graph convolutional networks

经过最大池化得到的食品图像的特征向量函数, $\boldsymbol{x}$ 表示得到的食品图像的低维特征向量, D 表示特征向量的维度.

2.2　原材料关系学习

图 1 原材料关系学习中利用图卷积网络学习原材料之间的关系.在图卷积网络中,每个节点表示为原材料的词嵌入,边表示为原材料之间的共现关系, C 表示为原材料的类别数量.由于图卷积网络是基于相关矩阵在节点之间传播信息工作,因此找到节点之间的相关性是建立相关矩阵 $\boldsymbol{A}$ 的关键性问题.在大多数应用中,相关矩阵是预先定义的.但是,在任何标准的多标签图像识别数据集中都没有提供相关矩阵.

本文通过在数据集中挖掘原材料的共现关系来定义原材料之间的相关性,并以条件概率的形式对原材料共现关系进行建模.其中 $P(L_i \mid L_j)$ 表示原材料 L_j 出现时原材料 L_i 出现的概率.为了构造该共现矩阵,首先计算训练集中出现的原材料对,得到矩阵 $\boldsymbol{M} \in \mathbf{R}^{C\times C}$.其中 C 是原材料的类别数, M_{ij} 是原材料 L_i 和原材料 L_j 同时出现的次数.然后通过共现矩阵,得到 $P_i = M_i / N_i$,其中 N_i 表示训练集中 L_i 的出现总次数.但是如果简单使用 P_{ij} 表示原材料 L_i 和原材料 L_j 的相关性,可能会存在两个缺陷:1)当 P_i 非常小的时候,表示原材料 L_i 和 L_j 的关系较弱或者没有关系,此时 P_i 不但增加了原材料之间的多余关系,而且增加模型的复杂度;2)来自训练和测试共现的绝对数量可能不完全一致,拟合到训练集的相关矩阵会降低泛化能力.因此可以将相关性 P_{ij} 进行二值化处理,并引入了阈值 τ 来缓解上面两种缺陷.相关矩阵的表达式如下:

$$\boldsymbol{A}_{ij} = \begin{cases} 0, & P_{ij} < \tau, \\ 1, & P_{ij} \geqslant \tau. \end{cases} \tag{2}$$

由于节点的特征为其自身特征与相邻节点的特征的加权总和,所以二进制相关矩阵可能导致过度平滑,使得节点变得难以区分[21].为了缓解这个问题,提出以下重新加权方案,具体方程式为:

$$\boldsymbol{A}'_{ij} = \begin{cases} \dfrac{p}{\sum\limits_{j=1, i\neq j}^{c} \boldsymbol{A}_{ij}}, & i \neq j, \\ 1 - p, & i = j, \end{cases} \tag{3}$$

其中 $\boldsymbol{A}'$ 是加权相关矩阵, p 为分配给节点本身和其他相关节点的权重.当更新节点特征时,获得节点本身的固定权重,相关节点的权重由邻域分布确定.当 $p \to 1$ 时,不考虑节点本身的特征;当 $p \to 0$ 时,邻近信息倾向于被忽略.

在本节中,基于图卷积网络的映射函数从标签表示中学习相互依赖的对象分类器,即 $\boldsymbol{W} = \{w_i\}_{i=1}^{C}$,其中 C 表示类别数.在图卷积网络中 $H^{l+1} = f(H^l, \boldsymbol{A}')$,其中 H^{l+1} 为 $l+1$ 层的节点表示,是将 l 层的节点表示 H^l 作为输入而产生的新的节点表示.对于第一层,输入是 $\boldsymbol{Z} \in \mathbf{R}^{C\times d}$ 矩阵,其中 d 是原材料嵌入的维数.对于最后一层,输出为 $\boldsymbol{W} \in \mathbf{R}^{C\times D}$,其中 D 表示图像表示的维数.将学习到的分类器应用于图像表示,可以得到预测分数:

$$\hat{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{x}. \tag{4}$$

假设图像的标签为 $\boldsymbol{y} \in \mathbf{R}^C$,其中 $\boldsymbol{y}^i = \{0,1\}$ 表

示原材料 i 是否出现在图像中. $\boldsymbol{y}^c$ 和 $\hat{\boldsymbol{y}}^c$ 分别表示类别 C 的标签和预测值.整个网络使用以下传统的多标签分类损失进行训练,损失函数 L 如下:

$$L = \sum_{c=1}^{c} \boldsymbol{y}^c \log(\sigma(\hat{\boldsymbol{y}}^c)) + (1 - y^c)\log(1 - \sigma(\hat{\boldsymbol{y}}^c)), \tag{5}$$

其中 $\sigma(\cdot)$ 是 Sigmod 激活函数.

3 实验评估

本节首先描述数据集、实现细节等.然后在 Food-101 和 VireoFood-172 两个食品数据集上验证本文所提方法在多标签分类任务中的有效性.

3.1 数据集

1) Food-101[22] 数据集:为了验证本文模型对食品多标签分类的可靠性,首先在 Food-101 数据集上进行了实验.该数据集一共包含 101 000 张图像,对于每个类别,将其中的 750 张图像用于训练以及剩余的 250 张图像用于测试.在训练集中没有对图像进行清理,因此仍然包含一些噪音,主要是以强烈的色彩和一些错误的标签形式出现的.在图数据的构建中,我们利用了数据集中全部的 446 个原材料信息作为节点,共现关系为边构建了原材料的关系图.图 2 中展示了 Food-101 数据集上一些食品图片类别和原材料的例子.

2) VireoFood-172[5] 数据集:为了进一步验证 GCN 的多标签模型在食品多标签分类中的可靠性,接着将该模型应用于 VireoFood-172 数据集中.其中 VireoFood-172 数据集涵盖了 172 种最常见的中国菜,共包含了 110 241 张食品图片以及 353 个原材料信息.本实验对 353 个原材料信息进行了简化处理,并删除了 5 张不合格的图片,得到了 276 个的原材料信息,包括类似"辣椒"和"猪肉"等热门食品,然后用得到的 276 个原材料信息作为分类的节点.在图 3 中展示了 VireoFood-172 数据集上一些食品图片类别和原材料的例子.

3.2 实现细节

3.2.1 超参数设置

本实验对模型参数的设置如下:训练中批次大小为 16,阈值 τ 为 0.4, p 为 0.2,迭代次数为 100 个周期,衰减 momentum 参数为 0.9,权重衰减参数为 0.000 5,初始的学习率为 0.01,并且在每 20 个周期之后,学习率将调整为当前学习率的 0.1.

3.2.2 评价指标

在多标签分类任务中,通常使用查准率(precision)、查全率(recall)和 F 度量 (F1)对模型的性能进行评估.查准率为预测为正例的样本中预测正确的概率,查全率为正例样本被预测正确的概率,F 度量为查准率和查全率的调和平均.其中,F 度量综合

a. 食品图片类别　　b. 原材料

图 2　Food-101 数据集上食品图片类别及原材料信息

Fig. 2　Food image categories (a) and ingredients (b) on Food-101 dataset

a. 食品图片类别　　b. 原材料

图 3　VireoFood-172 数据集上食品图片类别及原材料信息

Fig. 3　Food image categories (a) and ingredients (b) on VireoFood-172 dataset

考虑了查全率和查准率,对性能评价的影响相对较大.在本文的实验中,我们分别评估了总体的查准率(OP)、查全率(OR)、F1(OF1)和每个类别查准率(CP)、查全率(CR)、F1(CF1).按照文献[16,23]的设置,对于每个图像,如果它们的置信度大于0.5,则正确预测了该标签.

3.3　分类结果及分析

为了验证本文方法的有效性,我们在 Food-101 数据集上将本文方法与 ResNet50[24]、ResNet101 进行比较,并同时做了相同的数据增强处理,包括缩放、截取、翻转等.定量的结果如表 1、表 2 所示.由表 1 可以看出:数据增强后的图卷积网络的多标签模型总体性能都要优于 ResNet50.为了公平比较,我们将 ResNet101 做了相同的数据增强处理,其结果依然不如数据增强后的图卷积网络的多标签模型(本文模型).这是由于本文模型引入了原材料的共现关系,使得分类性能更好.由表 2 可以看出:在对食品图像的数据做增强后,模型的性能有显著提高.

表 1　本文模型在 Food-101 数据集上与当前最好方法的比较

Table 1　Comparison of the proposed model with current methods on Food-101 dataset　%

方法	OP	OR	OF1	CP	CR	CF1
ResNet50[24]	88.11	73.45	80.11			
ResNet101	**91.23**	82.81	86.82	**91.66**	81.20	86.12
本文模型	91.07	**85.73**	**88.32**	90.51	**83.03**	**86.61**

表 2　数据增强和未做数据增强时在 Food-101 数据集上的实验比较

Table 2　Experimental results on Food-101 dataset with/without data enhancement　%

方法	OP	OR	OF1	CP	CR	CF1
本文模型(未做数据增强)	87.99	79.70	83.64	88.28	78.41	83.06
本文模型(数据增强)	91.07	85.73	88.32	90.51	83.03	86.61

阈值选取:如图 4 所示,随着 τ 逐渐增加,多标签识别平均精度的均值(mean Average Precision, mAP)将上升,这是由于忽略了原材料的不相关节点.但是,当 τ 增加到一定值时,由于相关的原材料节点也会被忽略,因此精度会下降.对于 Food-101 数据集,τ 的最佳值为 0.4.

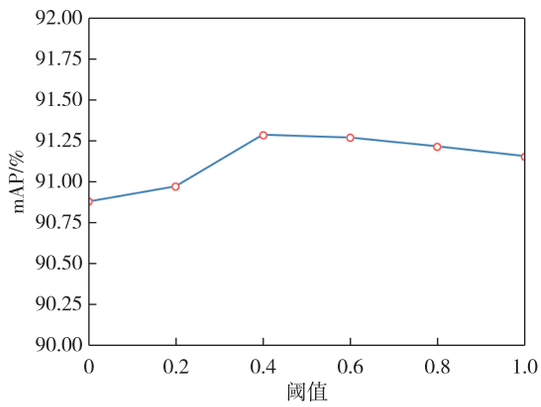

图 4　阈值对 mAP 值的影响

Fig. 4　Impact of threshold on mAP value

图 5 展示了基于 ResNet101 和基于图卷积网络的多标签模型的 top10 一些例子.实验结果按升序排列,黑色字体表示正确的原材料信息,红色表示错误的原材料信息.从中可以看到本文方法在所有方法中表现最好.

为了验证图卷积网络的多标签模型在其他食品数据集上的优越性,我们还将它应用于 VireoFood-172 数据集.由于 VireoFood-172 数据不均衡,不同的原材料,数量相差很大.实验结果表明,基于图卷积网络的多标签模型优于当前最好的方法.定量的结果见表 3.因为 OP、OR、OF1 是全部食品图像样本的累加统计,因此受少量大类别的影响较大.而 CP、CR、CF1 中考虑了每个食品图像类别然后做平均,所以它的值受大量小类别的影响较大.通过两种平均数据的比较,可以了解大多数食品图像样本的分类效果和大多数食品图像类别的分类效果.

图 6 展示了 ResNet172 和图卷积网络的多标签模型的多标签分类 top10 结果.实验结果按升序排列,黑色字体表示正确的原材料信息,红色表示错误的原材料信息.可以看到本文方法表现最好.

表 3　本文模型(增强)在 VireoFood-172 数据集上与当前最好的方法比较结果

Table 3　Comparison of the proposed model (enhanced) with current model on VireoFood-172 dataset　%

方法	OP	OR	OF1	CP	CR	CF1
ReaNet101	98.56	92.71	95.55	30.54	26.53	28.40
本文模型	98.17	95.12	96.62	48.58	41.43	44.72

4　总结与展望

本文提出了一种基于图卷积网络的多标签食品原料识别方法.该方法主要分为两部分:一是食品图像特征的提取,二是原材料之间关系学习,最后将得到的全局的视觉特征和局部的原材料特征相融合,实现多标签分类.食品图像的原材料关系的构建则是利用原材料之间在数据集中的共现关系.通过定量和定性的分析可以得出结论,对于食品成分识别的多标签问题,基于图卷积网络的食品多标签分类可以有效地提高食品图像的多标签分类性能.

在未来的研究中可以考虑以下研究方向:

1)研究原材料之间的其他相关性,然后基于不

图 5　Food-101 数据集上多标签分类的一些例子

Fig. 5　Examples of multi-label classification on Food-101 dataset

图 6　VireoFood-172 数据集上多标签分类的一些例子

Fig. 6　Examples of multi-label classification on VireoFood-172 dataset

同的相关性构图进行分类;

2)虽然食品种类繁多,但是食品大多由小部分的原材料构成,图卷积网络的多标签模型可以应用在更多的食谱中识别图像的原材料,进行营养成分的分析.

参考文献

References

[1] Kagaya H,Aizawa K,Ogawa M.Food detection and recognition using convolutional neural network[C]//Proceedings of the ACM International Conference on Multimedia,2014:1085-1088

[2] Singla A,Yuan L,Ebrahimi T.Food/non-food image classification and food categorization using pre-trained GoogLeNet model[C]//Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management,2016:3-11

[3] 梅舒欢,闵巍庆,刘林虎,等.基于 Faster R-CNN 的食品图像检索和分类[J].南京信息工程大学学报(自然科学版),2017,9(6):635-641
MEI Shuhuan,MIN Weiqing,LIU Linhu,et al.Faster R-CNN based food image retrieval and classification[J].Journal of Nanjing University of Information Science & Technology (Natural Science Edition),2017,9(6):635-641

[4] Jiang S Q,Min W Q,Liu L H,et al.Multi-scale multi-view deep feature aggregation for food recognition[J].IEEE Transactions on Image Processing,2020,29:265-276

[5] Chen Z M,Wei X S,Wang P,et al.Multi-label image recognition with graph convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:5177-5186

[6] Chen J J,Ngo C W.Deep-based ingredient recognition for cooking recipe retrieval[C]//Proceedings of the ACM on Multimedia Conference,2016:32-41

[7] Min W Q,Bao B K,Mei S H,et al.You are what you eat:exploring rich recipe information for cross-region food analysis[J].IEEE Transactions on Multimedia,2018,20(4):950-964

[8] 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,2016:41-49

[9] Min W Q,Jiang S Q,Liu L H,et al.Asurvey on food computing[J].ACM Computing Surveys,2019,52(5):1-36

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

[11] Bolaños M,Ferrà A,Radeva P.Food ingredients recognition through multi-label learning[C]//International Conference on Image Analysis and Processing,2017:394-402

[12] Deng J,Dong W,Socher R,et al.ImageNet:a large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition,2009:248-255

[13] Lin T Y,Maire M,Belongie S,et al.MicrosoftCOCO:common objects in context[C]//European Conference on Computer Vision,2014:740-755

[14] Everingham M,van Gool L,Williams C K I,et al.The pascal visual object classes (VOC) challenge[J].International Journal of Computer Vision,2010,88(2):303-338

[15] Gong Y,Jia Y,Leung T,et al.Deep convolutional ranking for multilabel image annotation[J].arXiv preprint,2013,arXiv:1312.4894

[16] Wang J,Yang Y,Mao J H,et al.CNN-RNN:a unified framework for multi-label image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:2285-2294

[17] Wang Z X,Chen T S,Li G B,et al.Multi-label image recognition by recurrently discovering attentional regions[C]//IEEE International Conference on Computer Vision (ICCV),2017:464-472

[18] Li X,Zhao F P,Guo Y H.Multi-label image classification with a probabilistic label enhancement model[C]//Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence,2014:430-439

[19] Li Q,Qiao M Y,Bian W,et al.Conditional graphical lasso for multi-label image classification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition,2016:2977-2986

[20] Lee C W,Fang W,Yeh C K,et al.Multi-label zero-shot learning with structured knowledge graphs[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:1576-1585

[21] Li Q M,Han Z C,Wu X M.Deeper insights into graph convolutional networks for semi-supervised learning[C]//Thirty-Second AAAI Conference on Artificial Intelligence,2018

[22] Bossard L,Guillaumin M,van Gool L.Food-101-mining discriminative components with random forests[C]//European Conference on Computer Vision,2014:446-461

[23] Ge W F,Yang S B,Yu Y Z.Multi-evidence filtering and fusion for multi-label classification,object detection and semantic segmentation based on weakly supervised learning[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:1277-1286

[24] Bolaños M,Ferrà A,Radeva P.Food ingredients recognition through multi-label learning[C]//International Conference on Image Analysis and Processing,2017:394-402

Multi-label food ingredient recognition via graph convolution network

LI Hui[1,2] MIN Weiqing[2] WANG Zhiling[2,3] PENG Xin[1]

1 Hunan Institute of Science and Technology, Yueyang 414006

2 Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190

3 University of Chinese Academy of Sciences, Beijing 100049

Abstract People's awareness about their nutrition habits is increasing. Keeping track of what we eat will be helpful for us to follow a healthier diet. Currently, nutrient recognition of food images is mainly focused on food categories recognition, or is tackled as multi-label task recognition. These two approaches, however, are not very discriminative owing to their neglect of potential relationship between ingredients. In this paper, we introduce the relationship between ingredients to identify food nutrients based on previous work. The recognition approach includes two modules, namely the image feature extraction module and the ingredients relationship learning module. The low-dimensional image feature vectors are extracted by convolutional neural network (CNN), and the relationship between ingredients is learned through a graph convolutional network (GCN). Specifically, GCN uses graph data where nodes represent food ingredients as word embedding and edges represent the correlation between nodes. Then the GCN directly map the graph data into a set of interdependent classifiers. Finally, the low-dimensional image feature vectors are fused to make detailed classification. We conducted experiments on food data sets of Food-101 and VireoFood-172. Compared with state of the art food recognition methods, our GCN-based multi-label food image classification method offers very promising results and can effectively improve the recognition performance.

Key words multi-label classification; food ingredients; food images; convolutional neural network; graph convolutional network