



理学硕士学位论文

面向小样本食品图像识别方法研究

Research of Few-Shot Food Recognition

作 者 吕永强

导 师 段 华 副教授

山东科技大学

二〇一九年五月

中图分类号 TP391

学校代码 10424

UDC

密 级 公 开

山东科技大学

理学硕士学位论文

面向小样本食品图像识别方法研究
Research of Few-Shot Food Recognition

作 者 吕永强 入学时间 2016.9

导 师 段 华 职 称 副教授

申请学位 理学硕士 所在学院 数学与系统科学学院

学科(类别) 系统科学与集成 方向(领域) 系统建模与优化

答辩日期 提交日期

学位论文使用授权声明

本人完全了解山东科技大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理。

作为申请学位的条件之一，学校有权保留学位论文并向国家有关部门或其指定机构送交论文的电子版和纸质版；有权将学位论文的全部或部分内容编入有关数据库发表，并可以以电子、网络及其他数字媒体形式公开出版；允许学校档案馆和图书馆保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。

(保密的学位论文在解密后适用本授权)

作者签名：

日 期： 年 月 日

导师签名：

日 期： 年 月 日

学位论文原创性声明

本人呈交给山东科技大学的学位论文，除所列参考文献和世所公认的文献外，全部是本人攻读学位期间在导师指导下的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：

年 月 日

学位论文审查认定书

研究生 在规定的学习年限内，按照培养方案及个人培养计划，完成了课程学习，成绩合格，修满规定学分；在我的指导下完成本学位论文，论文中的观点、数据、表述和结构为我所认同，论文撰写格式符合学校的相关规定，同意将本论文作为申请学位论文。

导师签名：

日 期：

摘 要

在食品的自动检测和饮食管理、食品流行趋势分析、食品健康和智能家居等诸多领域的驱动下，食品图像识别成为计算机视觉、数据挖掘以及多媒体领域的重要课题。目前大部分的食品图像识别工作是基于大规模标记样本的深度神经网络，这些工作无法鲁棒地识别只有少量样本的类别，因此小样本食品识别是个亟待解决的问题。在现有工作成果基础上，本文对小样本食品图像识别问题进行详细的研究和探讨。论文主要工作如下：

第一，首先研究了食品原料信息对小样本食品图像识别的影响，食品图像与其他数据集不同，其丰富的原料信息可以作为食品视觉信息的补充，为此本文提出一个多视角小样本食品图像识别框架(MVFSL)。实验结果表明，MVFSL 相较基准方法可以取得更高的分类准确率，且验证了信息在小样本食品识别中的优越性。

第二，为研究更加细粒度的区分对小样本食品识别的影响，本文使三元卷积神经网络来学习类内与类间更具区分信息。然而对于食品图像而言，目前基于线性度量的三元神经网络的鉴别能力不足，为此通过引入可学习的关系网络作为三元卷积神经网络的非线性度量函数，进一步提出一个基于非线性度量的三元神经网络用于小样本食品识别方法。此外，本文提出一个新的三元组样本在线采样方案“limited batch hard”，可以使训练过程更加稳定。实验结果表明，本文提出的非线性度量的三元神经网络和新的三元组筛选方案的有效性。

关键词：食品图像识别；小样本识别；度量学习；三元神经网络；非线性度量

Abstract

Food recognition has received a significant amount of attention in various fields, such as computer vision, data mining and multimedia communities motivated by many applications in automated food monitoring and dietary management , food trend and popularity analysis smart home and food safety. However, most existing work focuses on food recognition with large amounts of labelled samples, thus fail to robustly recognize food categories with few samples, under this condition, few-shot food recognition is an urgent problem. The main works of the paper are as follows:

1. To explore additional ingredient information for few-shot food recognition, a Multi-View Few-Shot Learning (MVFSL) framework is proposed to exploit rich food ingredients for few-shot food recognition.. We conduct the comprehensive experimental evaluation on various food benchmarks and experimental results verify the effectiveness of MVFSL, also the experimental results again demonstrate the advantage in exploiting ingredient information.

2. To study the impact of more fine-grained differentiation on few-shot food recognition, we use the triplet network to learn the inter-class and intra-class information, however the liner metric function is not discriminative enough for measuring similarities of food images. To address this problem, we use the learnable relation network as non-linear metric and propose a triplet network with relation network to solve the above two disadvantages of the few-shot learning and triplet network. In addition, we proposed an on-line mining rule for triplet samples, which makes the model stable in the training stage. The experimental results verify the effectiveness of the proposed model and sample rule.

Keywords: Food recognition; few-shot learning; Fine-grained; Relation Network; Triplet Network;

目 录

图清单.....	I
表清单.....	I
变量注释表.....	II
1 绪 论.....	1
1.1 研究背景.....	1
1.2 研究意义.....	2
1.3 研究内容及贡献.....	2
1.4 章节安排.....	4
2 相关研究工作.....	5
2.1 食品图像识别.....	5
2.2 小样本识别.....	6
2.3 预备知识.....	8
3 基于多视角小样本学习框架(MVFSL)的食品图像识别.....	12
3.1 模型结构.....	12
3.2 实验及结果分析.....	14
3.3 本章小结.....	20
4 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.....	21
4.1 模型结构.....	21
4.2 实验及结果分析.....	24
4.3 本章小结.....	29
5 总结与展望.....	30
5.1 研究工作总结.....	30
5.2 研究工作展望.....	30
参考文献	
作者简历	
致谢	
学位论文数据集	

Contents

List of Figures	I
List of Tables	I
List of Variables	II
1 Introduction	1
1.1 Research background.....	1
1.2 Research significance	2
1.3 Research contents and contributions.....	2
1.4 Arrangements.....	4
2 Related Work	5
2.1 Food recognition.....	5
2.2 Few-shot learning.....	6
2.3 Prerequisite knowledge.....	8
3 Few-Shot Food Recognition via Multi-view Representation Learning	12
3.1 Structure of the model.....	12
3.2 Experiment and result analysis.....	15
3.4 Summary.....	22
4 Few-shot food recognition via triplet network with relation network	23
4.1 Structure of model.....	23
4.2 Experiment and result analysis.....	25
4.3 Summary.....	30
5 Summary and Expectation	31
5.1 Summary of the work.....	31
5.2 Expectation of the work.....	31

References

Author's Resume

Acknowledgements

Thesis/Dissertation Data Collection

图清单

图序号	图名称	页码
图 1.1	食品图像对比	3
Fig.1.1	Comparison of food images	3
图 2.1	Siamese Network 网络结构	9
Fig.2.1	The structure of the Siamese Network	9
图 2.2	Matching Network 结构	10
Fig.2.2	The structure of Matching Network	10
图 2.3	Relation Network 结构	10
Fig.2.3	The structure of Relation Network	10
图 2.4	三元神经网络的结构	11
Fig.2.4	The structure of Triplet Network	11
图 3.1	多视角小样本学习(MVFSL)框架	12
Fig.3.1	The architecture for Multi-View Few-Shot Learning	12
图 3.2	部分食品图像的鉴别定位图	13
Fig.3.2	Discriminative localization maps from some food images	13
图 3.3	关系网络结构	14
Fig.3.3	Structure of Relation Network	14
图 3.4	Food-101, VIREO Food-172, ChineseFoodNet 数据集示例图	15
Fig.3.4	Some examples from Food-101, VIREO Food-172 and ChineseFoodNet	15
图 4.1	基于非线性度量学习的三元卷积神经网络框架图	22
Fig.4.1	Architecture of triplet network with relation network	22
图 4.2	limited batch hard 流程图	25
Fig.4.2	Flow diagram of limit batch hard	25
图 4.3	不同阈值的实验结果	28
Fig.4.3	Experimental results with different margins	28
图 4.4	Food-101 与 VIREO Food-172 部分训练集和测试集对比	29
Fig.4.4	Samples comparison of training set and test set in Food-101 与 VIREO Food-172	29

表清单

表序号	表名称	页码
表 3.1	MVFSL 的实验结果	17
Table 3.1	Table 3.1 Performance comparison on MVFSL	17
表 3.2	不同关系网络设置下 MVFSL-LC 在 Food-101 上的性能	18
Table 3.2	The performance of MVFSL-LC with different relation network settings on Food-101	18
表 3.3	不同关系网络设置下 MVFSL-TC 在 Food-101 上的性能	18
Table 3.3	The performance of MVFSL-TC with different relation network settings on Food-101	18
表 3.4	不同网络结构在 Food-101 的实验性能	19
Table 3.4	The performance with different networks on Food-101	19
表 3.5	不同网络结构在 VIREO Food-172 的实验性能	19
Table 3.5	The performance with different networks on VIREO Food-172	19
表 3.6	不同网络结构在 ChineseFoodNet 的实验性能	19
Table 3.6	The performance with different networks on ChineseFoodNet	19
表 3.7	Siamese Network 的对比实验结果	20
Table 3.7	Performance comparison on SN	20
表 3.8	Matching Network 的对比实验结果	20
Table 3.8	Performance comparison on MN	20
表 3.9	不同方法的对比实验结果	21
Table 3.9	Performance comparison on different methods	21
表 4.1	不同小样本方法在不同数据集上的实验结果	27
Table 4.1	The experimental results of different datasets	27
表 4.2	三元组采样的对比实验结果	28
Table 4.2	The experimental results of triplet with different setting	28
表 4.3	不同参数初始方案的实验结果	28
Table 4.3	The experimental results of different parameter initialization	28

变量注释表

变量	注释	初现页
k	未知样本在所有已知样本中最近的 k 个邻居	6
C	代表数据集中随机采样类别数	8
N	代表从每个类别中采样的样本数	8
L	代表测试集类别数	8
M	代表测试集的样本数	8
D_{base}	表示训练集	8
D_{novel}	表示测试集	8
x_i^{base}	表示训练集中支持集的样本	8
y_i^{base}	表示训练集中支持集的样本标签	8
x_j^{base}	表示训练集中查询集的样本	8
y_j^{base}	表示训练集中查询集的样本标签	8
S	表示支持集	8
Q	表示查询集	8
K	表示支持集中每个类别的样本数	8
x	表示锚(anchor)样本	11
x^-	表示异类样本	11
x^+	表示同类样本	11
$Net(x)$	表示样本 x 的特征向量表示	11
$\tau(\cdot)$	融合算子	14
$\delta(\cdot)$	融合算子	14
$r_{i, j}$	图像对的关系值得分	15
α	采样算法的超参数，用于限制正样本图像筛选	24
β	采样算法的超参数，用于限制负样本图像筛选	24

1 绪论

1 Introduction

1.1 研究背景(Research background)

食品识别是计算机视觉^[1]、数据挖掘^[2]以及多媒体社交^[3]等领域的重要研究课题,在食品自动化检测^[4]、食品管理、食品的趋势和流行性分析以及智能家居和食品安全中有着广泛的应用,例如智能厨房和智能个人营养日志^[5]等。人们的饮食和饮食行为会影响健康问题^[6],随着移动设备的快速发展,越来越多的基于视觉方法的饮食管理系统^[7]出现,其中必要的一步就是自动识别图像中的食品。

食品图像识别面临许多挑战。据维基百科统计,已经有超过 8000 个种类的食品类别^[8],在如此多的食品类别下,从现实世界收集的食品数据集符合典型的长尾分布,即许多不常见食品类别只能收集到少量的标记样本。对于一个鲁棒的食品图像识别模型,不仅可以识别常见食品类别,还需要有效的识别不常见且只有少量标记样本的食品类别。现有基于深度神经网络的食物识别方法^{[8]-[11]},需要大规模的标注样本才能进行有效的模型训练,对于只有少量标注样本的食品类别,这些方法的识别性能很差。基于此,本文主要探究和解决面向小样本的食品识别问题。

近年来,学者们对小样本学习的兴趣越来越浓厚^{[12]-[17]},包括小样本的字符识别^[12],图像识别^{[13]-[14]}和图像分割^[17]等领域。小样本学习的目标是学习识别一个只有少量标记样本的类别。目前基于小样本的学习方法有很多,例如 Siamese Network^[12]、Matching Network^[13] 和 Relation Network^[14]等,这些工作被提出并用于多个领域,例如字符识别和常见的物体识别等,本文第一次提出将小样本学习应用到食品识别领域。目前小样本识别领域的大部分工作只使用类别信息作为监督信息,并没有探索像丰富的属性和语义等其他类型信息对于小样本学习的性能的影响,因此本文使用丰富的食品原料信息作为辅助信息探索对小样本食品识别性能的影响。

三元卷积神经网络^[18]主要应用于人体重识别^{[19]-[20]},车辆重识别^[21]等领域。相比于孪生网络(连体网络),三元卷积神经网络通过同时控制类内和类间图像对的相似度差异,可以学习到更加具有类别区分特征表示。然而大多数基于三元卷积神经网络的工作使用不可学习的线性度量,例如余弦距离或者欧氏距离等。这种利用固定距离作为度量函数的方法存在以下三个缺陷:第一,模型过度依赖特征学习网络的学习能力,即整个模型会受到特征学习网络产生的特征表示的区分程度的限制;第二,对于需要细粒度区分的食品图像而言,线性的度量方法的辨别能力不够强;第三,固定的度量算法需要根据不同的网络模型以及不同的数据

集进行人工选择,无法自适应的根据数据集和网络结构进行学习。最近, Sung 等^[14]提出一种基于卷积神经网络的关系学习网络,此网络可以根据不同的数据集和模型自适应的学习一个非线性度量函数,通过损失函数可以同时调节特征嵌入网络和关系网络,使模型能够取得更好性能。因此本文通过有机融合三元神经网络与关系网络,提出一个可端到端训练的小样本食品识别模型。

1.2 研究意义(Research significance)

食品是个人和社会生活的重要组成部分,食品的多样性导致不同的食品偏好,不同的饮食习惯与饮食健康对身体健康有直接的影响。随着社会的发展进步,人们越来越关心食品安全,并且催生出许多检测饮食信息的应用,同样也推动了智能家居、智能厨房以及智能个人营养日志的发展。目前实现这些方法和应用的重要且必要的步骤是自动识别输入的图像并能够显示食品的相关信息,然而主流的基于深度学习方法的食品识别往往需要大量的标记样本,对于只有少量样本的类别的识别性能很差。据维基百科统计,已经有超过 8000 个种类的食品类别,从现实世界中收集如此多种类的数据是十分困难的,且收集到的数据集是符合典型的长尾分布。对于许多不常见的食品类别只能收集到少量的样本,对于基于深度学习的食品识别方法无法得到鲁棒模型。因此如何识别一个只有一个或者几个标记样本的类别是亟待解决的问题,也是本文的主要研究方向。

目前基于小样本学习方法主要应用于手写字体识别和物体识别领域,这些图像信息相对更加简单,例如手写字体,每张图片只是简单的笔划,对于图像内容更加复杂的食品图像而言,小样本食品识别面临更多的困难和挑战。首先,食品识别属于细粒度图像识别范畴,因此食品图像的类内与类间细微差异对食品识别是至关重要的。食品数据集与其他类型的数据集不同,除食品图像外,且有丰富的原料信息等额外信息,不同的食品类别是由不同的原料组合构成,因此原料信息有助于获取食品类内与类间更加细粒度的区分信息,从而帮助提升食品识别的准确性。其次,对于小样本识别问题,由于训练集与测试集类别空间是完全不相关的,利用训练集获得更加具有区分性的信息以提升测试集食品类别的准确率显得更加困难。因此除了考虑像丰富的原料信息等额外信息之外,如何从模型层面提升小样本方法的学习能力,以获得更具有区分度的特征同样需要关注。基于以上的问题,本文针对这两方面分别进行大量的研究工作。

1.3 研究内容及贡献(Research contents and contributions)

本文主要针对深度学习无法识别只有少量样本的食品识别,通过查阅大量的文献和大量的实验分析,实现了小样本食品图像识别,并且提出两个小样本食品识别方法的改进模型。

1.3.1 多视角小样本食品识别

在许多分享菜谱的网站上，除食品图片外，还有相关食品的原料信息。原料信息可以作为食品的一种属性信息，在食品识别中起着重要的作用。原料信息为食品识别提供了辨别性信息，特别对于小样本食品识别，许多食品类别只有少量的标记样本，原料信息作为一种常见信息，可以作为一种重要的补充信息可以用于提高小样本食品识别的性能。本文提出一个同时考虑食品类别信息和原料信息的小样本学习框架(MVFSL)。该框架主要有三部分组成：类别信息与原料信息特征学习子网络、多角度特征融合和多角度信息关系学习子网络，本框架第一次实现小样本食品识别问题。与传统的小样本学习研究方法不同，本框架不仅考虑到食品图像类别信息，同时引入丰富的食品原料信息，构建了一个端到端学习的多视角小样本食品识别模型。由于食品原料信息在不同的食品类别中的组合是不同的，食品原料信息可以作为类别信息的互补信息，通过融合两种信息可以获得更加丰富且具有区分性的图像信息，因此本框架相较于只考虑类别信息的小样本学习模型可以获得更好的性能。

1.3.2 融合三元神经网络与关系网络的小样本食品识别



图 1.1 食品图像对比

Fig.1.1 Comparison of food images

目前基于小样本学习的工作是利用图像对之间的相似度对图像进行分类，即相似度大的为同类，相似度小为异类。这些方法忽视图像对在类内与类间更加细粒度的对比信息。如图 1.1 所示，图片样本来自 VIREO Food-172^[22]，其中上半部分(a)是“糖醋鱼”类别的两个样本，下半部分(b)分别是“水煮鱼”和“水煮肉”类别的样本。数据集中存在很多视觉上相差甚远同类图像对(a)或者视觉上很相似的异类图像对(b)，由于忽略了更加细粒度区分信息，这些模型对此类的图像对的识别能力较弱。

本文提出一个基于非线性度量学习三元神经网络用于小样本食品识别。模型主要由两部分组成：(1)用于图像特征学习的特征嵌入子网络；(2)用于非线性度量函数学习的关系学习子网络。此外，本文提出一个新的样本在线采样方案。相较基于孪生网络的小样本识别模型，三元卷积神经网络代替孪生网络以学习食品图像在类内与类间更加细粒度的区分信息，从而提升小样本食品图像识别的分类性能。针对基于线性度量的三元神经网络对于复杂的食品图像的鉴别能力不够强，提出使用可学习的关系网络作为非线性度量方案，同时融合三元神经网络，进一步提出基于非线性度量学习的三元神经网络，提升小样本食品分类的性能。

1.4 章节安排(Arrangements)

全文总共分为五章，全文围绕实现小样本食品识别问题展开，每章的概要如下所示：

第一章是绪论部分，主要介绍了目前两种算法的研究背景、研究现状以及研究意义和本论文要研究的方向及创新点。

第二章为相关工作，首先介绍了传统的食品识别方法以及相关工作，然后介绍小样本学习方法的问题定义和常见的基于度量学习的小样本学习方法的定义，并详细解释小样本学习方法如 Siamese Network、Matching Network 等的具体实现，最后介绍三元神经网络的定义和三元神经网络的应用领域和实现方法等。

第三章主要探究食品原料信息对小样本识别问题的影响，提出了一个同时考虑食品类别信息和原料信息的小样本学习框架。本章首先介绍多视角小样本食品识别框架，详细描述每个模块的功能，并对每个模块的实现的具体步骤进行总结。最后通过大量的实验验证，结果显示多视角食品识别方法要优于使用单一信息的性能。本文还针对框架的每个模块进行详细的实验分析。

在第四章中，首先对基于非线性度量的三元神经网络模型结构进行详细描述，分别介绍图像特征嵌入模块的实现细节，关系学习模块的实现细节，以及对实现端到端的联合优化的细节进行详细描述。同时介绍一个新的三元组筛选的优化算法，称为“limited batch hard”。最后进行丰富的实验分析，验证不同数据集上所提方法与不同小样本方法的效果对比，同时对模型中超参数，例如三元损失函数中的阈值和“limited batch hard”中超参数设置进行分析。

第五章是总结和展望部分，首先对论文研究内容做出全面总结，然后提出研究算法的不足之处，从而为日后的研究提供可行的方向。

2 相关研究工作

2 Related works

食品是人类生活的基础。随着社交网络、移动网络和物联网的快速发展,对食品的相关研究可以支持多种应用和服务,如指导人类行为,改善人类健康和了解烹饪文化等。人们通常上传、分享和记录食品图片、食谱、烹饪视频和食品日记,产生了大规模的食品数据。食品数据提供了丰富的食品知识,可以帮助解决人类社会的许多核心问题。其中食品图像识别作为众多以食品为导向的任务和应用的基础,受到越来越多的关注。然而,在现实生活中收集食品数据集,需要耗费大量的成本且许多食品类别只能收集到少量的数据样本,因此小样本食品识别是一个亟待解决的问题。

本文的工作主要面向解决小样本食品图像识别问题,与本文工作最相关的是食品图像识别和小样本学习两个方向。本节将分别详细介绍相关工作及研究进展。

2.1 食品图像识别(Food image recognition)

与饮食有关的慢性疾病,如肥胖、糖尿病和心血管疾病,已成为一个主要的健康问题,饮食管理是预防和治疗这些疾病的关键因素。传统的食品日志方法是通过手工方式记录食品的种类和食用的份量,因此很难保证准确性。随着智能手机的广泛使用和计算机视觉的进步,新型的智能食品识别系统成为可能。一旦确定食品的种类或原料,就可以进一步进行各种与健康相关的分析,例如卡路里摄入量估计、营养分析和饮食习惯分析等。此外,直接从食品图像中识别食品对于食品相关的应用是十分重要的,以自助餐厅为例,食品识别不仅可以监控食品的消费,还可以自动为顾客抓取的食品买单。最后,对于那些想要更好地了解他们不熟悉的食品,或者他们从来没有见过的食品的人,他们可以简单地拍张照片,就可以了解更多的细节。因此食品图像识别成为实现上述功能的关键。

最近,越来越多的学者关注食品识别方法的研究,由于食品图像有较大的外观差异,这种差异不仅来自光照和焦点的变化,还有许多食品由于不同烹饪方法或者不同的食材搭配和切割方式造成的形状、纹理和其他视觉属性的变化,因此食品识别成为计算机识别领域的难题之一。目前已存在的食品识别的工作可被分为两个方向:传统方法和深度学习方法。对于传统的识别方法, Yang 等^[23]利用不同原料之间的空间关系,食品由不同食品原料的局部特征表示,但是这种方法只适用于标准化的食品。Bossard 等^[1]采用随机森林方法挖掘食品图像中的辨别性区域,以提高食品识别的准确率,除此之外,许多工作使用人工设计的特征用于食品识别。Joutou 等^[24]采用多核学习方法探索多种图像特征,将颜色直方图的 Bag-of-SIFT 与伽柏滤波器联合的方法用于区分图像,与传统的人工设计的特征相比, Kawano 等^{[25][26]}发现深度特征明显优于人工设计的特征。Kagaya 等^{[26][27]}进一步提取深度的视觉特征用于食品的检测和识别。然而以上工作只考虑到食品图像的视觉信息,许多工作关注面向餐馆上下文信息的食品识别,如 Xu 等^[27]探索餐馆的地理位置信息和菜单信息简化食品的分类问题。此外,一些工作关注使用食品的原料信息和食谱信息构建多模态的识

别或检索模型提高识别和检索性能: Min 等^[28]提出一个多模态多任务模型同时学习食品图像的视觉特征表示和食品的原料信息的特征表示, 并利用这两种信息提升模型的识别能力。Min 等^[29]利用丰富的食谱信息和食品图像的视觉信息, 提出一个多模态食品检索模型。文献[30]进一步考虑食品的原料信息, 利用多任务的方法建模原料信息视觉信息之间的关联, 提出一个多任务识别模型。此外, 一些工作关注到图像背景噪音的干扰, Mei 等^{[31][43]}使用 Faster R-CNN 提取图像的判别性区域特征, 避免噪音的干扰且使提取的特征更具判别性。

然而以上工作都是在大规模数据基础上建立模型, 在样本数量较小时并不适用, 基于此, 本文研究面向小样本的食品识别问题。

2.2 小样本识别(Few-shot learning)

深度学习(Deep Learning)已经在各个领域得到广泛的应用, 许多学者利用深度学习方法解决各类问题, 如图像分类问题、视频分类、物体跟踪等, 且在其领域内取得了较好的性能。然而, 深度学习是一种数据饥渴方法, 需要大量的有标记样本才能训练得到鲁棒模型, 这与人类的认知模式有很大不同, 人类可以从很少的监督信息中学习一个新概念, 例如一个小孩可以从书本的一张图片中概括出“狮子”的概念, 然而深度学习却需要成千上万的标记样本。这激发了越来越多的学者将目光投向小样本学习问题, 即如何从少量样本中学习一个类别。

近年来, 学者们采用不同的策略处理小样本学习问题。对于小样本学习, 其主要目标是识别只有少量标记的样本。2006 年 Li 等^[32]提出一个变分贝叶斯模型用于小样本, 通过利用已学习类别预测只有一个或者几个样本的新类别。随后 Lake 等^[33]提出了多层贝叶斯模型, 并且在字符识别任务上与人类水平相匹配。目前主流的小样本学习方法有两种: 主要有基于度量学习(Metric Learning)的小样本学习方法^{[12]-[14]}和基于元学习(Meta Learning)^{[15]-[16]}的小样本学习方法。

2.2.1 基于度量学习的小样本方法

在机器学习中, K 近邻分类算法和支持向量机(SVM)算法是处理文本或者图像相似度的较好的方法, 而 K 近邻算法由于无法处理不同维度之间对分类的影响以及对高维数据, 往往不能取得良好的分类效果, 而且 K 近邻算法对类别密度高度敏感, 当类间密度分布高度不均匀时, 对最终分类效果会有较大影响。为解决上述问题, 首要解决的是距离度量问题。度量学习的基本原理为根据不同的任务定义来自主学习出针对某个特定任务的度量距离函数。度量学习在人脸识别问题中是常用的机器学习方法, 后来度量学习又被迁移至其他领域中得到广泛使用, 尤其是针对高维数据的文本处理, 度量学习有很好的分类效果。与传统的度量学习相似, 基于度量学习的小样本方法旨在学习一组函数, 这样在学习特征嵌入模块时, 使用简单的最近邻或线性分类器就可以很容易地识别图像之间的相似度。

目前已经有许多基于度量学习的小样本工作, Koch 等^[12]提出使用孪生网络作为图像特征嵌入网络, 通过学习特征嵌入网络获得图像对特征表示, 使其能够通过固定的度量算

法实现图像的分类。在孪生网络的基础上, Vinyals 等^[13]提出 Matching Network, 通过构建支持集和查询集, 并将其嵌入到一个共享的特征空间中, 利用余弦距离计算损失函数学习整个网络。Siamese Network 和 Matching Network 都是采用线性度量方法, Sung 等^[14]提出一种基于非线性度量方法的 Relation Network, 利用可学习的卷积神经网络作为非线性度量函数, 相比于线性度量方法, Relation Network 取得更好的性能。由于基于度量学习的小样本学习方法取得较好的性能, 因此本文主要关注基于度量学习的小样本识别方法。

2.2.2 基于元学习的小样本学习

元学习在机器学习方法中已经受到广泛的关注, 相关工作在 80 年代和 90 年代早期就已经出现^[34], 目前接受最广泛的元学习概念可以概括为“学会学习”, 通过一些训练过程可以提高算法的学习能力(通过逐步和动态修改其归纳偏见)。元学习算法一般包括两个主要部分^[35]: 基础学习者(Base-Learner)和元学习者(Meta-Learner)。基础学习者主要用于处理独立事件(或任务), 对应于小样本学习独立事件可定义为小组标记的训练样本。元学习者依次从几个这样的事件中学习, 目的是提高基础学习者在不同情景中的学习能力。总而言之, 元学习的目标是通过通过对一组训练集进行泛化, 使基础学习算法能够有效地适应新的训练集。

近年来, 元学习方法是研究小样本学习方法的研究热点之一。Finn 等^[15]提出了 MAML 模型, 该模型并不是显式地设计一个元学习器模块来学习更新规则, 而是通过梯度下降的操作, 通过反向传播来优化超参数或元学习者的初始参数。为了减轻深度神经网络中对学习知识灾难性遗忘的缺点。文献[36]-[37]提出利用记忆增强模型, 它可以首先保留并访问与先前训练相关的重要信息和先前学习到而不可见的信息。然而这些记忆增强模型需要在长时间范围内其存储和访问模型。Vinyals 等^{[13][13]}提出了一种基于注意力机制来确定在一个事件中哪些知识片段更加重要。Mishra 等^[38]提出一个基于空间卷积网络的软注意力(soft attention)元学习模型, 该方法允许注意力机制利用与过去学习知识相关的信息。考虑到目前基于度量学习的小样本学习方法取得更好性能, 且对于食品数据集而言, 基于度量学习的方法可以学习和利用更加细粒度的区分信息, 因此本文主要关注基于度量学习的小样本学习方法

2.3 预备知识(Prerequisite knowledge)

2.3.1 小样本学习问题定义

小样本学习问题中, 训练与测试样本通常由一系列的训练集与测试集构成。假设有 C 个训练类别, 共有 N 个有标注的训练样本, 定义训练集 $D_{base} = \{(x_i^{base}, y_i^{base})\}_{i=1}^N, y_i^{base} \in \{B_1, B_2, \dots, B_C\}$, 其中 x_i^{base} 是指采样的图像, y_i^{base} 是指 x_i^{base} 的标签。对于测试集, 假设有 L 个新的类别, 共有 M 个测试样本, 定义测试样本集 $D_{novel} = \{(x_i^{novel}, y_i^{novel})\}_{i=1}^M$, 其标签集为 $y_i^{novel} \in \{N_1, N_2, \dots, N_L\}$ 。值得注意的是训练集与测试集的样本空间是完全不相关的。

对于小样本学习方法，需要定义支持集和查询集。以训练集为例，随机从 D_{base} 采样 C 个类别，并从每个类别中随机采样 K 个样本构成支持集 (Support Set) $S = \{(x_i^{base}, y_i^{base})\}_{i=1}^m$, $m = (C \times K)$ 。定义查询集 (Query Set) $S = \{(x_i^{base}, y_i^{base})\}_{i=1}^n$ ，从支持集的 C 个类别中随机选择一个类别，并从选中类别中随机采样 n 个样本。如果支持集中包含 C 个不同的类别，且每个类别包含 K 个样本，这个任务称为“C-way K-shot”。一般而言，在小样本学习的设置中 K 往往是很小的，例如 $K=1$ 或者 $K=5$ 。基于“C-way K-shot”任务的目标是提供一张查询图像 \hat{x} ，利用支持集学习一个分类映射 $c_S(\hat{x})$ 得到查询类别的概率分布 $P(\hat{y}|\hat{x}, S)$ ，其中 \hat{y} 为预测的标签。

2.3.2 Siamese Network

Siamese Network 最早用于签字认证识别的网络模型，其算法原理是利用神经网络提取目标特征，然后利用两个特征的相似性度量来确定两个目标是否相似。如图 2.1 所示，以图像识别为例，对于两个图像样本 X_1 和 X_2 ，分别输入 Siamese Network 的两个分支，即两个卷积神经网络，通过卷积操作、池化操作等相关运算后，分别得到 X_1 和 X_2 的特征向量 $G_w(X_1)$ 、 $G_w(X_2)$ ，其中 w 为网络参数。通过构造基于特征向量距离度量函数(如欧式距离等)计算两张图像之间的相似度。

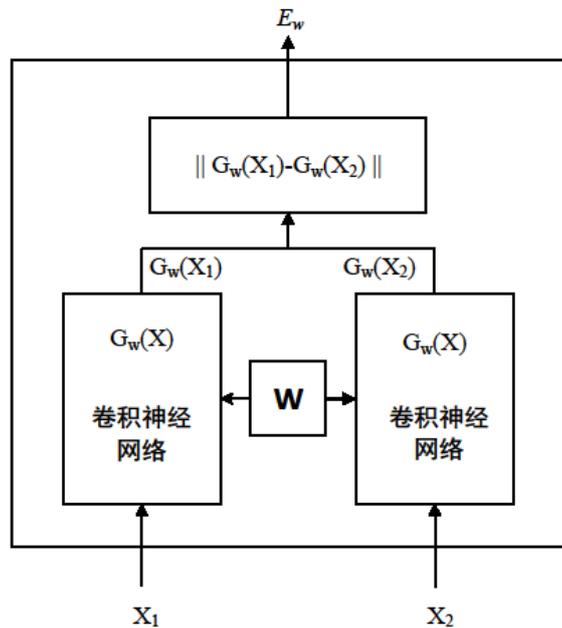


图 2.1 Siamese Network 网络结构
Fig 2.1 The structure of the Siamese Network

Siamese Network 作为一种基于相似性度量方法的分类方法，与传统的分类方法不同，其输入是一对样本(同类样本或异类样本)，且对每个样本不给定确定的样本标签，而是将同类的标签设置为 1，异类标签设置为 0，通过淡化样本标签，并通过构建样本对可以进一步扩张训练数据，因此对于处理样本类别多而类别样本数少的情况有明显的优势。对于两个分支网络，其网络结构相同且网络权值共享，对网络的输出进行度量计算，度量函数可以采取如 L1、L2 欧氏距离或者余弦距离等，其损失函数需要根据不同的度量函数进行针

对性的设计，其中最常见的是交叉熵损失函数，通过控制图像对之间的相似度进行学习。

Koch 等^[12]将 Siamese Network 应用于小样本识别领域，首先从训练集中随机生成同类和异类图像对进行模型训练。在测试阶段，与“C-way K-shot”的设置相似，训练完成的模型用于从C个类别中识别出一个未标记图像X。具体而言，给定一个无标签的图像X和另外一个图像集 $\{X_c\}_{c=1}^C$ ，其中 X_c 表示第c个类别中的一个样本，利用训练好的网络提取将图像对 $(X, X_c)_{c=1}^C$ 特征，根据图像对之间的余弦距离判断图像X的类别。

2.3.3 Matching Network

基于 Siamese Network 的基础思想，Vinyals 等^[13]提出 Matching Network 用于解决小样本学习问题，训练程序基于一个简单的机器学习原理：测试和训练条件必须匹配(match)。该工作第一次提出使用事件集(task)的思想去训练模型，其具体做法是将每次训练迭代定义为一个事件，其中每个事件由支持集和查询集构成。同样在测试阶段，将每个测试迭代也视为一个事件，同样每个事件由支持集和查询集组成。此外，该模型还引入记忆增强模型和注意力机制应用到网络结构中，其中记忆增强模型采用的是长短期记忆网络(LSTM)，并经注意力机制应用在 LSTM 的模型中。

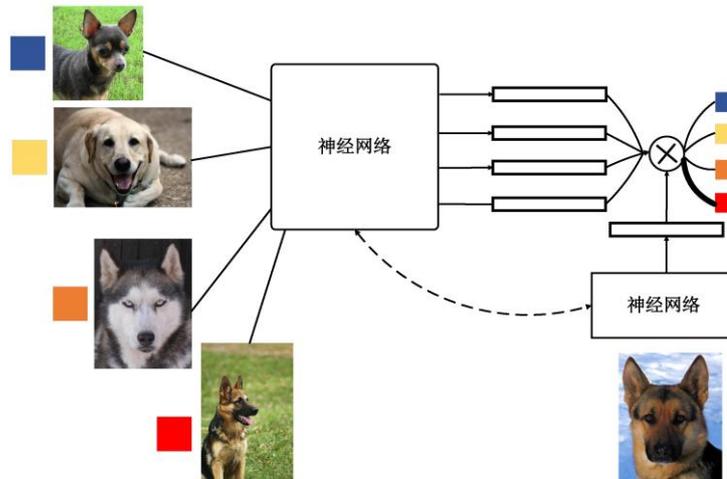


图 2.2 Matching Network 结构

Fig 2.2 The structure of Matching Network

Matching Network 的主要工作原理如图 2.2 所示：将查询图片与支持集图片输入神经网络，获得特征嵌入表示后，利用距离度量算法计算查询图像与支持集中每张图片的距离，距离近的认为是同类的图片，从而实现对查询图像的分类。在训练阶段，从数据集的训练集中构建支持集 S_{train} 和查询集 Q_{train} 组成训练事件(task)，并利用训练事件优化网络模型的分能力 $c_S(\hat{x})$ ，其中 \hat{x} 为训练样本。模型分类器可以定义为一个映射函数 $P_\theta(\hat{x}|\hat{y}, S_{train})$ ，其中 θ 为模型在训练过程中学习的参数。在测试阶段，给定一个未标记的样本 x_t 和支持集 S_{test} ，并从中预测 x_t 的样本标签 y_t 。

2.3.4 Relation Network

在基于度量学习的小样本学习方法的启发下，Sung 等^[14]提出了 Relation Network 用于小样本识别。其主要思想与 Siamese Network 和 Matching Network 相似，与先前工作最大的不同是 Relation Network 采用可学习的神经网络作为非线性度量函数代替 Siamese Network 和 Matching Network 中的线性度量。

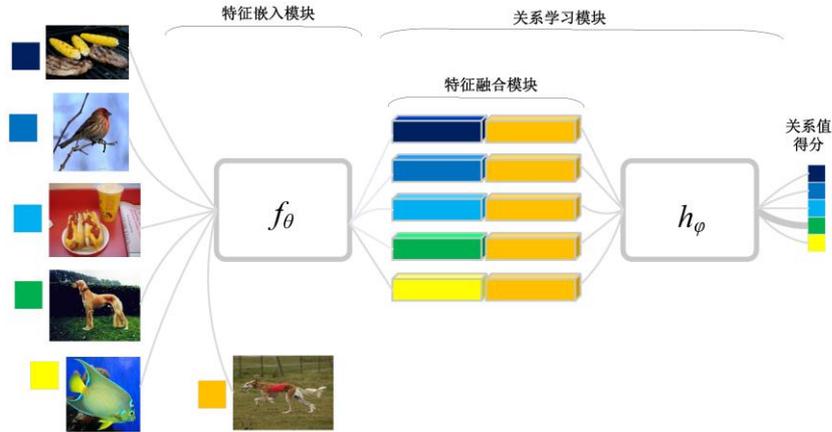


图 2.3 Relation Network 结构
Fig 2.3 The structure of Relation Network

其结构如图 2.3 所示。其结构主要由两大部分组成：特征嵌入模块 f_θ 和关系学习模块 h_ϕ ， θ 和 ϕ 分别为特征嵌入网络和关系学习网络的参数。关系学习模块包含特征融合模块，实现特征的融合过程。特征嵌入模块主要用于提取支持集和查询集的特征图，利用特征融合模块将提取的查询集的特征图与支持集每张图像的特征图进行融合，得到融合后特征图后，利用关系学习网络学习融合后两张图像之间的关系值得分。通过使用神经网络作为非线性度量函数，Relation Network 相较 Siamese Network 和 Matching Network 取得更好的性能。

2.3.5 三元神经网络

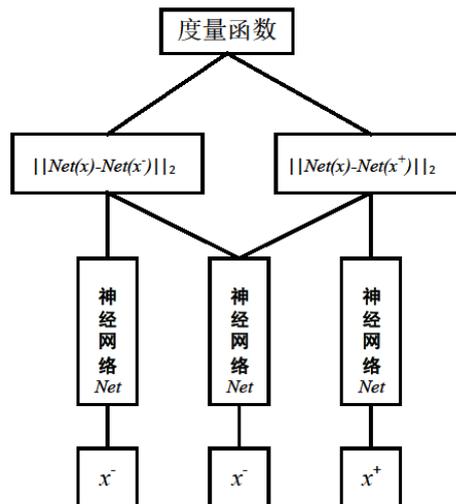


图 2.4 三元神经网络的结构
Fig 2.4 The structure of Triplet Networks

三元神经网络^[18]是基于连体网络的改进模型，其处理的问题与连体网络相似，都是面向样本种类多而可训练样本数量少的情况下，在人体重识别领域和车辆重识别领域应用十分广泛。三元神经网络结构如图 2.4 所示，与连体网络相同的是，都是将图像对映射到特征空间中，分别计算图像对之间的距离。但是与连体网络不同的是，三元神经网络以三元组作为一组输入，而不再是一对图像，其中三元组由锚图像 x (anchor) 和正样本图像 (positive) x^+ 及一张负样本图像 (negative) x^- 组成。将三张图像分别输入神经网络得到各自的特征嵌入表示分别为 $Net(x)$ 、 $Net(x^+)$ 和 $Net(x^-)$ ，分别计算正样本图像对和负样本图像对之间的距离，以平方欧式距离为例，其距离度量分别为 $\|Net(x) - Net(x^+)\|_2$ 和 $\|Net(x) - Net(x^-)\|_2$ ，最终利用三元损失函数实现网络模型的训练。

3 基于多视角小样本学习框架的食品图像识别

3 Few-Shot Food Recognition via Multi-view Representation Learning

本章主要研究原料信息对于小样本食品识别问题的影响，提出了一个多视角小样本学习框架(MVFSL)。在本框架中除使用面向类别的深度视觉特征外，还探索并使用面向原料信息监督深度特征。面向原料特征作为食品重要的语义属性信息，是对面向类别特征重要补充信息，对提高食品识别能力具有重要作用。特别在小样本食品识别中，原料信息可以进一步缩小不相交训练类别空间和测试类别空间的差距。为了充分利用原料信息，将这两种特征融合在一起，首先将两种特征在各自微调深度神经网络的最后一个卷积层的特征图结合起来，利用算子将两种特征信息深度融合，然后在结合后的特征图上进行卷积，这种卷积被进一步合并到一个多视角关系网络中。模型通过采用端到端联合优化的方法进行训练，在不同的食品数据集上进行了大量的实验，证明了 MVFSL 在多视角特征融合方面的优势。在此基础上，进一步扩展到另外两种常见的小样本学习网络，即 Siamese Network 和 Matching Network。实验结果表明，在这两种网络中引入原料信息同样可以提高小样本食品图像识别的性能。下面将介绍 MVFSL 的框架及实现方法描述，并详细分析实验结果。

3.1 模型结构(Structure of the model)

如图 3.1 所示，多视角小样本学习(MVFSL)框架主要由三个部分组成：面向类别和面向原料的特征学习、多视角特征图融合和多视角特征关系学习。下面分别对框架中每部分实现过程和功能作用进行详细阐述。

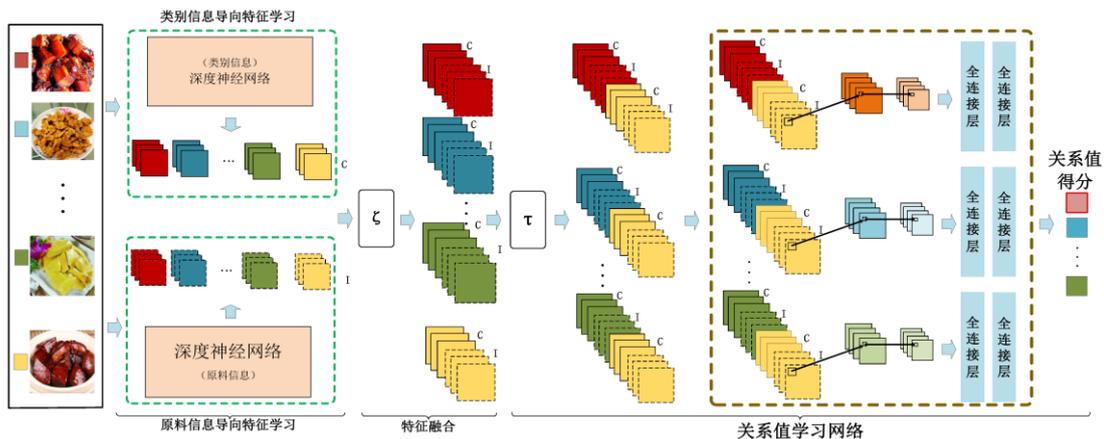


图 3.1 多视角小样本学习(MVFSL)框架

Fig.3.1 The architecture for Multi-View Few-Shot Learning

3.1.1 类别和面向原料特征学习

对于食品类别信息导向的特征学习模块，该模块主要由一个深度神经网络组成。在训练过程中，使用训练集中的图像及其与类别信息微调分类深度神经网络模型，然后提取最后一个卷积层的特征图谱 $f_{\mu}(x_i)$ ，其中 μ 为深层网络参数。对于食品原料信息导向的特征学

习模块，将原料信息做为每个图像的多标签向量，使用训练集的图像及其原料多标签来微调另一个用于多标签原料分类深度神经网络，然后提取面向原料特征图谱 $f_v(x_i)$ ，其中 v 表示面向原料的深层网络参数。值得提出的是，任意的深度神经网络都可以应用到此结构中，本文主要以 VGG16 作为基础模型来介绍本文工作。

3.1.2 多视角特征图融合

根据文献[39]的研究，深度卷积层通常是稀疏的，并且倾向于选择更高层次的视觉概念。如图 3.2 所示，展示了来自 VIREO Food-172 数据集的三幅食品图像样本及其深度卷积层特征得到可视化效果，本文采用 Grad-Cam^[40]实现类别和原料判别定位区域可视化(叠加图像颜色越暖，像素提供判别能力越强)。从左至右依次为：类别及其原料信息；原始图像；分类判别性定位图；原料判别性定位图。对于原料信息的可视化，利用训练的多标签分类模型^[41]，根据最后卷积层的特征图得到每个原料类别的判别定位区域。从图中可以发现许多特征图的判别性区域(暖色高亮显示)是存在语义意义的，例如，“米饭”的判别性区域倾向于定位在“糖醋里脊炒饭”和“叉烧饭”中的米饭区域，“欧芹”的判别性区域倾向于定位在“红烧肠”的欧芹区域。因此，面向原料的特征能够捕捉食品的更加丰富的细节，可以作为对面向类别信息特征的补充。例如，对于“红烧肠”，原料信息区域覆盖图像的不同部分，如“欧芹”和“葱花”的激活区域，与类别的激活区域具有互补性。

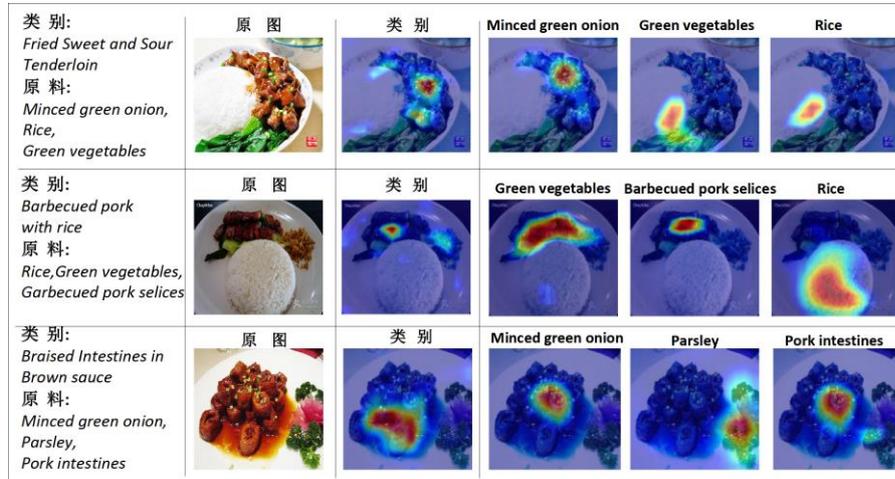


图 3.2 部分食品图像的鉴别定位图

Fig. 3.2 Discriminative localization maps from some food images

通过以上的观察和分析，将类别和原料判别性区域结合起来，可以增强特征表示能力。在本文中，通过一个算子 δ 计算结合特性映射表示 $\delta(f_\mu(x_i), f_\theta(x_i))$ ，其中 $\delta(\cdot)$ 采用的是特征图的深度拼接。具体而言，提取 VGG16 最后卷积层的特征图维度为 $14 \times 14 \times 512$ ，通过算子 δ 融合后，特征维度将变为 $14 \times 14 \times 1024$ 。这样的组合将被送入卷积层进行融合。

将一张图像的两种不同类型的特征图进行融合后，并对融合后的特征图进行卷积。同时，将卷积与下文将要提及的多视角关系学习结合起来进行参数学习，下面的小节中给出更多细节。融合的特征包含丰富的食品属性信息，弥补了不相交训练类别与测试类别之间的差距，提高了小样本食品识别的性能。

3.1.3 多视角关系学习

如图 3.3 所示，与文献[14]相似，关系学习网络包含 2 个卷积块和 2 个全连接层。每个卷积块由 64 个 3×3 的卷积核构成的卷积层和一个 2×2 的最大池化层，在卷积层中使用批正则化处理，ReLU 非线性激活。对于两个全连接层分别是 8 维和 1 维的输出。第一个全连接层使用 ReLU 函数作为激活函数，第二个全连接层使用 Sigmoid 函数作为激活函数。

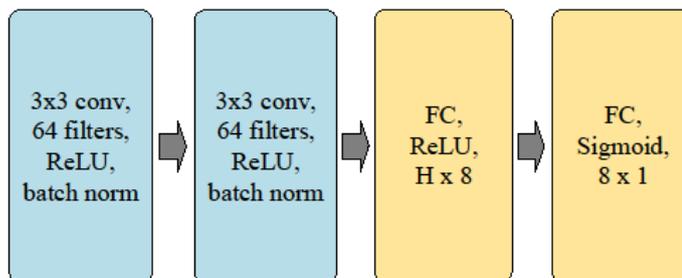


图 3.3 关系网络结构

Fig.3.3 Structure of relation network

多视角关系学习用于将查询图像与已标记的样本图像进行比较，以根据图像对之间的关系得分确定这些图像对是否来自相互匹配的类别。对于多视角关系学习，从支持集 S 中采样融合的多视角表示 $\delta(f_\mu(x_i), f_\vartheta(x_i))$ 和以及对应查询集中的样本 $\delta(f_\mu(x_j), f_\vartheta(x_j))$ 。这些融合的特征图谱通过另一个融合算子 $\tau(\cdot)$ ，其中算子 $\tau(\cdot)$ 同样采用的是特征的深度融合。具体而言，对于分别来自支持集和查询集的两个样本 x_i 和 x_j ，融合类别信息和原料信息后其特征表示为 $\delta(f_\mu(x_i), f_\vartheta(x_i))$ 和 $\delta(f_\mu(x_j), f_\vartheta(x_j))$ ，其特征维度为 $14 \times 14 \times 1024$ 。经过算子 $\tau(\cdot)$ ，通过融合 x_i 和 x_j ，其最终的特征维度变为 $14 \times 14 \times 2048$ 。通过关系子网络 h_θ 进一步处理查询和支持集中样本的组合特征映射，其中 θ 为关系网络的参数。关系网络由两个卷积层和两个全连接层组成，最后一个全连接用于产生关系值得分。关系得分 $r_{i,j}$ 的定义如下式所示：

$$r_{i,j} = h_\theta \left(\tau \left(\delta(f_\mu(x_i), f_\vartheta(x_i)), \delta(f_\mu(x_j), f_\vartheta(x_j)) \right) \right). \quad (3.1)$$

得到关系得分后，本文采用均方误差损失函数训练整个模型，通过回归关系值得分 $r_{i,j}$ 至 Ground-Truth 标签，即同类图像对的相似度为 1，异类图像对相似度为 0。最终的损失函数如下式所示：

$$\operatorname{argmin}_\theta \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - 1(y_i == y_j))^2. \quad (3.2)$$

其中 m 为来自支持集的图像数量， n 为来自查询集的图像数量。

3.2 实验及结果分析(Experiment and result analysis)

本章提出的多视角小样本食品识别模型可以应用到其他小样本的学习方法中。在本节中，将介绍多视角小样本食品识别模型另外两种常用的小样本学习方法，即 Siamese Network 和 Matching Network，这两个工作采用固定余弦距离作为度量函数。首先描述实验设置，包括数据集和实现细节；然后在不同的食品数据集上对该框架进行了详细的定性

和定量的评价；接下来，评估了另外两种小样本学习方法 Siamese Network 和 Matching Network 在食品识别上的性能；最后，给出分析和讨论。

3.2.1 数据集



图 3.4 Food-101, VIREO Food-172, ChineseFoodNet 数据集示例图

Fig.3.4 Some examples from Food-101, VIREO Food-172 and ChineseFoodNet

由于目前并没有食品数据集应用到小样本食品识别，本文使用以下三个食品数据集，分别是 Food-101^[1]、VIREO Food-172^[23]和 ChineseFoodNet^[42]来模拟小样本的食品识别。

Food-101:数据集包含 101 类的食品类别，每个类别包含 1000 张食品图像，共有 101000 张食品图像，其中大多数食品为西方菜品。由于小样本学习的训练集与测试集是完全不交叉的，本文随机分割数据集，其中 71 类为训练集，30 类为测试集。

VIREO Food-172:数据集包含 172 个食品类别，共有 110241 张食品图像。所有的食品图像都是中国菜，并且所有的食品图像是从百度和谷歌搜索中获取的。同样本文随机分割数据集，其中 132 类作为训练集，40 类作为测试集。

ChineseFoodNet:数据集主要由许多不同烹饪风格的中国菜品构成，包含 208 个食品类别，共有 185628 张食品图像。本文随机分割数据集，其中 158 类作为训练集，50 类作为测试集。

在图 3.4 中分别展示了三个数据集中的样本以及相关信息。

3.2.2 实验细节

在现有的小样本学习设置的基础上，采用常见的“5-way 1-shot”策略，即 $C=5$ 且 $K=1$ 。在训练阶段，对于每个事件 MVFSL 除采样 $C \times K = 5$ 张图像外，同时采样 15 张查询图像。对于 Siamese Network 和 Matching Network，本文遵循其在原文中的设置，每个事件只包含一个查询图像。此外，在实验中采用 VGG16 网络作为基本的深度网络结构进行特征提取。

对于 MVFSL，采用基于事件的训练策略，每个事件中 共有 $15 \times 5 + 1 \times 5 = 80$ 张图像。在训练阶段，从训练集中构建 10 万个事件以保证事件的多样性，使用 Adam^[43]对小样本学习进行随机优化，初始学习率为 10^{-4} ，每训练 2 万个事件学习率减半。对于最终小样本食

品识别的准确率计算，首先从测试集中随机生成 1000 个事件，将这 1000 个事件的平均准确率作为最终的性能。

在本文中，MVFSL 的训练过程分为两个版本，分别为 MVFSL-LC 和 MVFSL-TC。下面详细介绍两个版本的训练细节。在 MVFSL-LC 设置中，首先分别对食品分类监督信息的深度网络和原料监督信息的深度网络进行微调，随后分别提取整个训练集和测试集相应的特征图，基于提取的特征，通过融合算子两种类型的特征映射融合，最后将支持集与查询集中的样本特征图融合后的特征图输入到关系网络进行多视图关系学习。由于这种训练过程属于松散结合(Loosely Combined)，因此将此版本称为 MVFSL-LC。对于第二个版本，在对食品类别信息监督的深度网络和原料信息监督深度网络进行微调后，对 MVFSL 中这两类特征提取子网络和关系子网络的整体训练才用了端到端联合优化训练方式。基于该设置整个模型紧密地组合(Tightly Combined)在一起，因此命名为 MVFSL-TC。

3.2.3 实验分析

(1)定量评价：首先本文对 MVFSL-LC 和 MVFSL-TC 的性能进行定量评价，为了验证 MVFSL-LC 和 MVFSL-TC 的有效性，本文采用以下方法作为基准方法：类别信息导向的关系网络^[14](RN-类别)，该基准方法使用图像及类别信息训练特征提取网络和关系网络；面向原料关系网络(RN-原料)，该基准方法使用图像及原料信息来训练特征提取网络和关系网络。分别对 Food-101、VIREO Food-172 和 ChineseFoodNet 进行了综合评价。实验结果如表 3.1 所示。

表 3.1 MVFSL 的实验结果 (单位：%)
Table 3.1 Performance comparison on MVFSL

模 型	Food-101	VIREO Food-172	ChineseFoodNet
RN-类别	53.9	74.0	63.8
RN-原料	53.5	70.5	64.0
MVFSL-LC	55.1	74.8	65.8
MVFSL-TC	55.3	75.1	66.1

从表中可以看出：(a) 三个数据集上的 MVFSL-LC 和 MVFSL-TC 的性能比它们对应的基准方法要好。特别是在 Food-101 中，MVFSL-LC 较 RN-类别和 RN-原料有更好的表现，分别比这两个基准方法高 1.2%和 1.6%；与这两个基准方法相比，MVFSL-TC 具有更好的性能，其性能分别高出 1.4%和 1.8%。在 VIREO Food-172 中，MVFSL-LC 的性能也比 RN-类别和 RN-原料有所提高，分别比这两个基准方法提高 0.8%和 4.3%；MVFSL-TC 的性能优于这两个基准方法，分别优于 1.1%和 1.6%。同样，对于 ChineseFoodNet，MVFSL-LC 的性能比 RN-类别和 RN-原料有明显的提高，也都比这两个基准方法高出 2.0%和 1.8%；MVFSL-TC 的性能优于这两个基准方法，分别优于 2.3%和 2.1%。实验结果验证了 MVFSL 在融合面向类别和面向原料的特征表示方面的有效性。此外也证明了原料信息可以弥补不相交训练类别与测试类别之间的差距，从而提高了性能。(b)MVFSL-TC 通过端到端训练，在面向类别的模型和面向原料的模型中可以学习到更加合适的类别和原料表示。从表 3.1

中还可以看出，MVFSL-TC 的性能优于 MVFSL-LC。具体来说，对于 Food-101，MVFSL-TC 的性能优于 MVFSL-LC，且优于 MVFSL-LC 0.2%。对于 VIREO Food-172，MVFSL-TC 的性能优于 MVFSL-LC，且优于 MVFSL-LC 0.3%。同样，对于 ChineseFoodNet，MVFSL-TC 与 MVFSL-LC 相比，性能同样提高 0.3%。

(2)定性分析 MVFSL-LC。本文通过列举实例进一步验证了 MVFSL-LC 的有效性。如图 3.4 所示，给出了来自 MVFSL-LC 与其他两个基准方法的一些例子对比结果。

在图 3.4 中从左到右依次为：类别及其原料列表，查询图像，支持集图像。在每个例子中，显示了 MVFSL-LC 和其他两个基准方法的关系得分。关系得分越高，图像的相关性越强。其中红色框标中的图片为与查询集同类的图像。从图中可以得到：(a)与 RN-类别相比，RN-原料可以做出更合理的预测。对于“鸭脖”、“菠菜猪肝汤”等菜品，RN-类别没有做出正确的预测，而 RN-原料做出了正确的预测。(b) MVFSL-LC 能够对一些从支持集上很难识别的菜系做出准确的预测，如“鸭脖”、“菠菜猪肝汤”等。RN-类别和 RN-原料都没有做出正确的预测，

类别和原料	查询集图像	支持集图片															
数据集: Food-101 类别: Clam chowder 原料: Milk, Celery, Flour, Fat, Clam, Onion, Butter, Potato, Salt	 RN-Category RN-Ingredient MVFSL	 <table border="1"> <tr> <td>0.374371856</td> <td>0.339298308</td> <td>0.716337919</td> <td>0.234566808</td> <td>0.392835468</td> </tr> <tr> <td>0.126617000</td> <td>0.097398110</td> <td>0.139268380</td> <td>0.073396480</td> <td>0.104291120</td> </tr> <tr> <td>0.924340070</td> <td>0.651421490</td> <td>0.913648720</td> <td>0.860222160</td> <td>0.863520620</td> </tr> </table>	0.374371856	0.339298308	0.716337919	0.234566808	0.392835468	0.126617000	0.097398110	0.139268380	0.073396480	0.104291120	0.924340070	0.651421490	0.913648720	0.860222160	0.863520620
0.374371856	0.339298308	0.716337919	0.234566808	0.392835468													
0.126617000	0.097398110	0.139268380	0.073396480	0.104291120													
0.924340070	0.651421490	0.913648720	0.860222160	0.863520620													
数据集: VIREO Food 172 类别: Duck neck 原料: Duck Neck, Lettuce	 RN-Category RN-Ingredient MVFSL	 <table border="1"> <tr> <td>0.294173568</td> <td>0.0292214863</td> <td>0.1571647080</td> <td>0.456348667</td> <td>0.067083038</td> </tr> <tr> <td>0.609602990</td> <td>0.0276846200</td> <td>0.0313023800</td> <td>0.051866260</td> <td>0.070580680</td> </tr> <tr> <td>0.928961245</td> <td>0.0093683833</td> <td>0.0439097323</td> <td>0.488185138</td> <td>0.197947964</td> </tr> </table>	0.294173568	0.0292214863	0.1571647080	0.456348667	0.067083038	0.609602990	0.0276846200	0.0313023800	0.051866260	0.070580680	0.928961245	0.0093683833	0.0439097323	0.488185138	0.197947964
0.294173568	0.0292214863	0.1571647080	0.456348667	0.067083038													
0.609602990	0.0276846200	0.0313023800	0.051866260	0.070580680													
0.928961245	0.0093683833	0.0439097323	0.488185138	0.197947964													
数据集: ChinaFoodNet 类别: Spinach and pork liver soup	 RN-Category RN-Ingredient MVFSL	 <table border="1"> <tr> <td>0.219086170</td> <td>0.111917370</td> <td>0.336520430</td> <td>0.332341550</td> <td>0.335742620</td> </tr> <tr> <td>0.691743970</td> <td>0.097642060</td> <td>0.044169190</td> <td>0.097139980</td> <td>0.691015010</td> </tr> <tr> <td>0.791396260</td> <td>0.437382820</td> <td>0.344772220</td> <td>0.496775870</td> <td>0.903917430</td> </tr> </table>	0.219086170	0.111917370	0.336520430	0.332341550	0.335742620	0.691743970	0.097642060	0.044169190	0.097139980	0.691015010	0.791396260	0.437382820	0.344772220	0.496775870	0.903917430
0.219086170	0.111917370	0.336520430	0.332341550	0.335742620													
0.691743970	0.097642060	0.044169190	0.097139980	0.691015010													
0.791396260	0.437382820	0.344772220	0.496775870	0.903917430													

图 3.4 MVFSL-LC 与另外两个基准方法的实验结果

Fig.3.4 Some experimental results from MVFSL and other two baselines

而 MVFSL-LC 融合了面向类别信息和面向原料信息特征信息后，做出了正确的预测。这进一步验证了面向类别的特征和面向原料的特征是互补的，MVFSL-LC 能够融合这两种类型的信息，提高小样本食品识别的性能。

表 3.2 不同关系网络设置下 MVFSL-LC 在 Food-101 上的性能 (单位: %)

Table 3.2 The performance of MVFSL-LC with different relation network settings on Food-101.

模 型	64 个滤波器	128 个滤波器
1 个卷积层	54.7	55.9
2 个卷积层	55.1	56.3
3 个卷积层	56.1	56.9
4 个卷积层	54.6	55.6

表 3.3 不同关系网络设置下 MVFSL-TC 在 Food-101 上的性能 (单位: %)
Table 3.3 The performance of MVFSL-TC with different relation network settings on Food-101.

模 型	64 个滤波器	128 个滤波器
1 个卷积层	55.1	56.0
2 个卷积层	55.3	56.4
3 个卷积层	56.6	57.6
4 个卷积层	55.7	56.3

从表中可以看到:(a)当固定滤波器的个数时,随着卷积层数的增加,性能会有一致性的提高。(b) 128 个滤波器的卷积层性能要优于 64 个滤波器的卷积层。例如,在 MVFLS-LC 中,随着卷积层数的增加,128 个滤波器的性能优于 64 个滤波器,在不同卷积层的性能分别提高了 0.2%、1.2%、0.8%和 1.0%。(c)在每组参数设置中,MVFLS-TC 的性能都优于传统的 MVFLS-LC。(d)关系网络的设置为 3 个卷积层和 128 个滤波器时,MVFSL 的性能达到最优,然而随着层数和滤波器数的增加,MVFSL 的呈现性能下降趋势。这主要原因可能为模型复杂度的增加导致过度拟合。

(4)MVFSL-LC 和 MVFSL-TC 特征嵌入网络深度的分析。如表 3.4 所示,在 Food-101、VIREO Food-172 和 ChineseFoodNet 三个数据集上分别以 AlexNet、VGG16 和 VGG19 进行详细实验对比。

表 3.4 不同网络结构在 Food-101 的实验性能 (单位: %)
Table 3.4 The performance with different networks on Food-101

模型	AlexNet	VGG16	VGG19
RN-类别	48.6	53.9	54.7
RN-原料	51.4	53.5	55.1
MVFSL-LC	51.8	55.1	55.9
MVFSL-TC	52.1	55.3	56.5

表 3.5 不同网络结构在 VIREO Food-172 的实验性能 (单位: %)
Table 3.5 The performance with different networks on VIREO Food-172

模型	AlexNet	VGG16	VGG19
RN-类别	68.8	74.0	74.6
RN-原料	70.3	70.5	73.6
MVFSL-LC	70.8	74.8	75.2
MVFSL-TC	71.0	75.1	75.3

从表 3.4、表 3.5 和表 3.6 中我们可以看出:(a)随着层数的增加,性能也在不断的提高。随着层数的增加,MVFSL 与其他的基准方法基线获得了一致的性能提升。例如使用 VGG16 网络的 MVFSL 和其他基线的性能比使用 AlexNet 的模型高出 0.2%到 5.2%。同样,在大多数情况下,使用 VGG19 网络的 MVFSL 和其他基线的性能要优于使用 VGG16 网络的模型。只有一个例外,基于 VGG9 的 RN-类别的性能比基于 VGG16 性能要低。一个可能的原因是该数据集上的网络训练存在过拟合。(b) 与具有相同层的每个体系结构的基线相比,

MVFSL具有一致性增长。(c)对于基于 AlexNet、VGG16 和 VGG19 的 MVFSL-TC 与 MVFSL-LC 相比，性能得到一致提高，这再次验证端到端训练方式的有效性。

表 3.6 不同网络结构在 ChineseFoodNet 的实验性能 (单位: %)
Table 3.6 The performance with different networks on ChineseFoodNet

模型	AlexNet	VGG16	VGG19
RN-类别	59.5	63.8	62.9
RN-原料	59.4	64.0	65.6
MVFSL-LC	59.9	65.8	66.5
MVFSL-TC	60.2	66.1	66.7

(5)Siamese Network(SN)和 Matching Network(MN)的实验评价。对于 SN 和 MN，根据“5 way-1 shot”实验设置构造正样本图像对和负样本图像对。具体来说，随机抽取 5 个类别并从每个类中采样 1 幅图像作为支持集，随机选取其中一个类别额外采样 1 张图像作为查询集。将查询图像与支持集中的每个图像相结合，构造出正图像对和负图像对，其中正图像对表示这些图像来自同一个类，负图像对表示它们来自不同的类。因此每个事件(task)中共有 6 张图像。在训练阶段，从训练集中构建 100000 个事件样本，SN 使用 Nesterow 动量随机优化方法，初始学习率为 5×10^{-5} ，动量值为 0.9。MN 采用 Adam 随机优化算法，初始学习率设置为 5×10^{-4} 。SN 和 MN 每 20000 个事件学习率减半。在测试集中随机产生的 1000 个事件的平均准确率作为最终的分类精度。为了进一步验证引入成分信息的有效性，本文构建基准方法来评估 SN-多视角和 MN-多视角：

Siamese Network-类别(SN-类别): 使用图像及类别信息来训练 Siamese Network;
Siamese Network-原料(SN-原料): 使用图像及原料信息来训练 Siamese Network;
Matching Network-类别(MN-类别): 使用图像及类别信息来训练 Matching Network;
Matching Network-原料(MN-原料): 使用图像及原料信息来训练 Matching Network。

表 3.7 Siamese Network 的对比实验结果 (单位: %)
Table 3.7 Performance comparison on SN

模 型	Food-101	VIREO Food-172	ChineseFoodNet
SN-类别	49.1	60.3	50.5
SN -原料	54.5	65.5	62.5
SN-多视角	55.0	65.8	64.4

表 3.8 Matching Network 的对比实验结果 (单位: %)
Table 3.8 Performance comparison on MN

模 型	Food-101	VIREO Food-172	ChineseFoodNet
MN-类别	45.6	73.6	48.9
MN -原料	46.8	65.9	52.0
MN -多视角	47.5	74.1	53.0

表 3.7 和表 3.8 分别展示 SN-多视角和 MN-多视角与基准方法的性能比较。具体而言，对于 SN-多视角，比基于类别信息的基准方法和基于原料信息的基准方法性能提高 4%到

5%左右。类似地，MN-多视角同样比与基准方法的性能有所提升，并比基准方法高出 0.4% 到 8.2%左右。这进一步验证了本章提出的方法在使用原料信息方面的有效性。

表 3.9 不同方法的对比实验结果 (单位: %)

Table 3.9 Performance comparison on different methods

模 型	Food-101	VIREO Food-172	ChineseFoodNet
SN-多视角	55.0	65.8	64.4
MN -多视角	47.5	74.1	53.0
MVFSL-LC	55.1	74.8	66.5
MVFSL-TC	55.3	75.1	66.7

表 3.9 进一步总结了 MVFSL-TC、MVFSL-LC、SN-多视角和 MN-多视角之间的实验结果。与其他三种方法相比，MVFSL-TC 的性能最好。对于 Food-101，MVFSL-TC 的性能优于 MVFSL-LC、SN-多视角和 MN-多视角，分别为 0.2%、0.3%和 7.8%。与 VIREO Food-172 和 ChineseFoodNet 类似，MVFSL-TC 的性能最好。

3.3 本章小结(Summary)

本章提出的 MVFSL 是受到文献^[14]提出的 Relation Network 的启发，但是本章工作与之有着本质的不同：(1)Relation Network 只学习利用了面向类别信息用于小样本学习，而 MVFSL 同时采用类别信息与原料信息用于小样本学习。(2)Relation Network 只在类别信息上进行卷积操作，而 MVFSL 通过在融合的分类信息和原料信息上卷积达到更好的融合效果。

本章提出了一种多视角小样本学习(MVFSL)框架，以通过使用食品丰富的原料信息探索多视角小样本食品识别。为了充分利用原料信息，这两种特征首先将各自微调深网的最后一层卷积中提取的特征图并进行有机融合，然后对融合后的特征图进行卷积，从而有效地使用这两种特征信息。该卷积操作被合并到一个多视角关系网络中，该网络经过训练，用于将查询图像与已标记的样本进行比较，以获得图像对之间关系得分。通过对包括西方和东方食品在内的三种不同食品数据集进行综合实验评价，验证了本文提出的 MVFSL-LC 和 MVFSL-TC 的有效性。此外，还将框架扩展到另外两种小样本学习方法，实验结果验证了利用食品原料信息进行小样本食品识别的优越性。

4 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

4 Few-shot food recognition via triplet network with relation network

本章主要介绍三元卷积神经网络用于小样本食品识别的模型。相比于孪生网络(连体网络),三元卷积神经网络通过同时控制类内和类间图像对的相似度差异,可以学习到更加具有类别区分特征表示。本文将三元神经网络应用于小样本食品识别领域,下面将详细介绍模型结构及具体实现细节,同时通过综合实验验证模型的有效性。

4.1 模型结构(Structure of model)

如图 4.1 所示,基于非线性度量学习的三元神经网络模型主要由两个子网络构成:图像嵌入网络和关系学习网络。在下面小节中将详细介绍每个模块的实现细节。

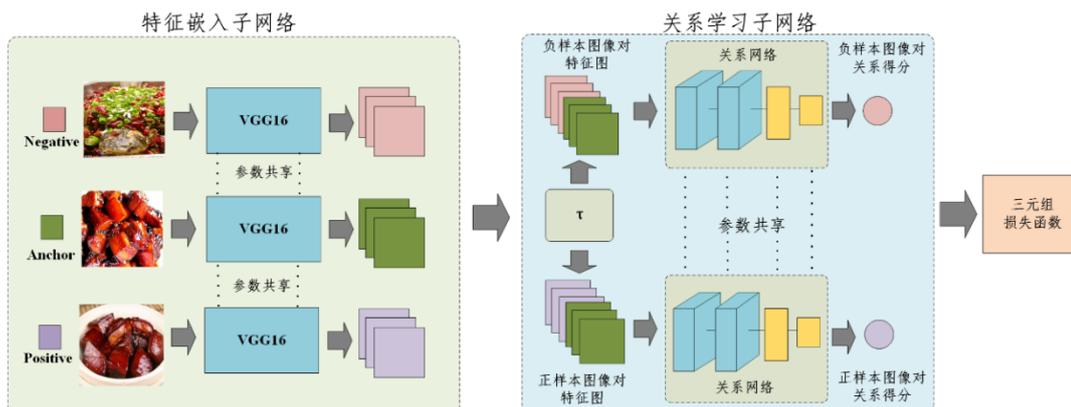


图 4.1 基于非线性度量学习的三元卷积神经网络框架图

Fig.4.1 Architecture of triplet network with relation network

4.1.1 图像特征嵌入网络

基于三元卷积神经网络的图像特征嵌入网络 f_{θ} 是受到孪生网络(连体网络)的启发,其结构由三个参数共享的前馈深度神经网络构成,本文采用三个 VGG16 深度神经网络作为特征嵌入网络,分别用于三元组 $X = \{x^-, x, x^+\}$ 的特征提取。具体而言,提供三个图像输入 x^- , x 和 x^+ , 分别表示 negative, anchor 和 positive 图像,其中 x 与 x^+ 是属于同一类别样本, x^- 与 x 属于不同类别的样本。假设样本的特征嵌入表示为 $f_{\theta}(x)$, 其中 θ 为特征嵌入网络的参数。先前的工作^{[20]-[21]}是采用分类层之前的全连接层作为图像的嵌入表示,然后利用固定距离算法(如 L2 距离)得到三元组的距离表示:

$$Triplet(x^-, x, x^+) = \left[\begin{array}{c} \left\| \left(f_{\theta}(x), f_{\theta}(x^-) \right) \right\|_2 \\ \left\| \left(f_{\theta}(x), f_{\theta}(x^+) \right) \right\|_2 \end{array} \right]. \quad (4.1)$$

本文特征嵌入表示为卷积层提取的特征图，不仅可以契合关系学习网络的输入，而且相较全连接层，卷积层的特征包含更加丰富的图像信息。

4.1.2 关系学习网络

与 MVFSL 中关系网络相同，关系学习网络 g_{φ} 主要包含 2 个卷积块和 2 个全连接层，其中每个卷积块由 64 个 3×3 的卷积核构成的卷积层和一个 2×2 的最大池化层，在卷积层中同样使用批正则化处理，ReLU 非线性激活。对于两个全连接层分别是 8 维和 1 维的输出。第一个全连接层使用 ReLU 函数作为激活函数，第二个全连接层使用 Sigmoid 函数作为激活函数。

为契合关系网络的输入，本文使用一个算子 τ 用于特征图的融合，该算子采用的是特征图的深度拼接，例如从最后一个卷积层提取的特征图维度为 $14 \times 14 \times 512$ ，通过正样本或者负样本图像对特征融合，融合特征图维度为 $14 \times 14 \times 1024$ 。则最终的三元组的特征中间表示为：

$$Triplet(x^-, x, x^+) = \left[\begin{array}{c} \tau \left(f_{\theta}(x), f_{\theta}(x^-) \right) \\ \tau \left(f_{\theta}(x), f_{\theta}(x^+) \right) \end{array} \right]. \quad (4.2)$$

将深度融合后的正负样本图像对的特征图分别输入到两个参数共享的关系学习网络，得到两个范围在 0 到 1 的关系值表示正负样本图像对之间的相似度，称之为非线性关系得分。在本文中，本文认为非线性关系得分越高，代表两张图像越相似，相反关系值越低表示两张图像越不相似。最终的关系得分表示为：

$$r(x^-, x, x^+) = \left[\begin{array}{c} g_{\varphi} \left(\tau \left(f_{\theta}(x), f_{\theta}(x^-) \right) \right) \\ g_{\varphi} \left(\tau \left(f_{\theta}(x), f_{\theta}(x^+) \right) \right) \end{array} \right], \quad (4.3)$$

其中 r 为关系值得分， φ 为图像特征嵌入网络的参数， $\tau \left(f_{\theta}(x), f_{\theta}(x^-) \right)$ 为负样本图像对的特征深度融合， $\tau \left(f_{\theta}(x), f_{\theta}(x^+) \right)$ 为正样本图像对的特征深度融合。

4.1.3 端到端的联合优化

与先前工作不同，本文没有选择用于学习两张图像相似度的损失函数，例如交叉熵损失函数或者均方误差损失函数等，而是使用能够学习三张图像区分信息的铰链损失函数。铰链损失函数通过同时限制正负样本关系值可以使模型学习到更加具有区分度的信息。具体而言，当输入一个三元组 (x^-, x, x^+) ，如前文中提到的， x 与 x^+ 属于同类， x 与 x^- 属于异类，将三元组输入到图像特征嵌入网络得到特征图分别为 $f_{\theta}(x^-)$ ， $f_{\theta}(x)$ ， $f_{\theta}(x^+)$ ，经过算子 τ 本文得到融合后的正样本图像对和负样本图像对的特征图分别为 $\tau \left(f_{\theta}(x), f_{\theta}(x^+) \right)$ 和 $\tau \left(f_{\theta}(x), f_{\theta}(x^-) \right)$ 。训

练阶段融合的特征图 $\tau(f_\theta(x), f_\theta(x^-))$ 和 $\tau(f_\theta(x), f_\theta(x^+))$ 分别输入到关系学习网络 g_φ ，最终的铰链损失函数定义为：

$$L_{triplet}(x, x^-, x^+) = \max \left\{ m + g_\varphi \left(\tau \left(f_\theta(x), f_\theta(x^-) \right) \right) - g_\varphi \left(\tau \left(f_\theta(x), f_\theta(x^+) \right) \right) \right\} \quad (4.4)$$

其中 m 为阈值， $g_\varphi(\cdot)$ 为关系网络得到的关系值得分， φ 为关系网络的参数。铰链损失函数使得正样本图像对之间的关系值得分大于负样本图像对之间关系值得分。铰链损失函数不仅指导特征嵌入模型产生图像的嵌入，而且同时指导关系网络学习。

4.1.4 Limited batch hard 三元组挑选方案

三元卷积神经网络的模型训练条件十分苛刻，首先随着数据集的增大，三元组的采样空间要远远大于样本空间，训练的时间将变得冗长；其次在构建的三元组之中有很多对训练不利，例如对于容易训练的三元组，即正样本图像对的关系值很大，负样本图像对关系值很小，铰链损失函数的值接近 0 或者等于 0，这些三元组对模型的调节能力很小或者不调节模型；更糟糕的是，如果选取过于难训练的三元组，网络将很难收敛，以至于无法学习到具有区分性的信息。因此，三元组的选择直接关系到三元卷积神经网络训练的稳定性。Hermans^[20]等提出一种三元组挑选方案，称为“batch hard”。具体而言，对于每个批量数据，随机选择 P 个类别，然后从每个类别中随机选择 K 个样本，这样每个批量样本中就含有 $P \times K$ 个图像。对于每个样本 x 在构建三元组时可以选择出最难训练的正样本图像对和最难训练的负样本图像对，最终筛选出 $P \times K$ 个三元组。因此基于“batch hard”的关系网络的损失函数表示为：

$$L_{BH}((\theta, \varphi); x) = \sum_{i=1}^P \sum_{a=1}^K \left[\left\{ m - \max_{p=1 \dots K} g_\varphi \left(\tau \left(f_\theta(x_a^i), f_\theta(x_p^i) \right) \right) \right\} + \min_{\substack{j=1 \dots P \\ n=1 \dots K \\ j \neq i}} g_\varphi \left(\tau \left(f_\theta(x_a^i), f_\theta(x_n^j) \right) \right) \right] \quad (4.5)$$

但是利用“batch hard”的方法筛选基于关系网络的三元组，仍然存在很多对训练不利的三元组。具体而言，关系网络最后一层全连接的输出经过 Sigmoid 函数后被规范到 0 到 1 之间，具体来讲就是越接近 1 图像对越相似，越接近 0 越不相似。通过本章研究发现，在一个批量的样本中存在很多不利于训练的三元组，例如有许多正样本图像对的关系值在 0.01 左右，且负样本的关系值得分在 0.9 左右。如果直接用这些三元组训练网络，会造成网络训练不稳定或者不收敛。

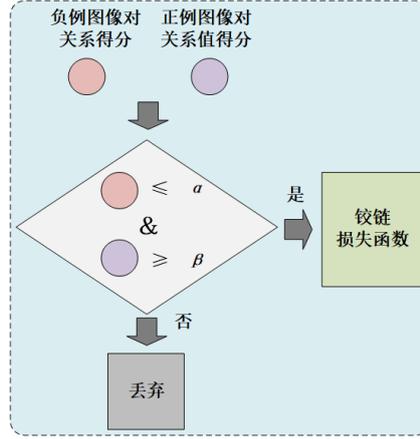


图 4.2 limited batch hard 流程图

Fig.4.2 Flow diagram of limit batch hard

如图 4.2 所示，本章提出一个新的三元组采样方案。与“batch hard”方法相似，对于一个样本 x ，从其构成的三元组中采样，但是在采样最难训练的三元组之前，对所有的正样本图像对和负样本图像对进行限制，例如只选取正样本图像对关系得分大于等于 α ，同时负样本图像对的关系得分小于等于 β 的三元组用于训练。本文称这种方法为“limited batch hard”。基于新的三元组采样规则，可以采样出对训练有益的三元组，使模型可以稳定的训练。最终基于新的三元组采样方案的损失函数为：

$$L_{BH}((\theta, \varphi); x) = \sum_{i=1}^P \sum_{a=1}^K \left[\left\{ m - \max_{p=1 \dots K} g_{\varphi} \left(\tau \left(f_{\theta}(x_a^i), f_{\theta}(x_p^i) \right) \right) + \min_{\substack{j=1 \dots P \\ n=1 \dots K \\ j \neq i}} g_{\varphi} \left(\tau \left(f_{\theta}(x_a^i), f_{\theta}(x_n^j) \right) \right) \right\} \right]$$

$$s.t. \quad g_{\varphi} \left(\tau \left(f_{\theta}(x_a^i), f_{\theta}(x_n^j) \right) \right) \geq \alpha, \quad (4.6)$$

$$g_{\varphi} \left(\tau \left(f_{\theta}(x_a^i), f_{\theta}(x_p^i) \right) \right) \leq \beta.$$

4.2 实验及结果分析(Experiment and result analysis)

首先介绍数据集以及实验的实现细节，之后验证本文所提模型在不同数据集上的实验性能，最后分析模型初始化对性能影响以及铰链函数阈值和“limited batch hard”参数的敏感度。

4.2.1 数据集及划分

为验证方法的有效性，本章在三个重要的食品数据集：Food-101^[1]、VIREO Food-172^[23]和 ChineseFoodNet^[42]上进行全面对比试验分析。对于数据集的划分与 MVFSL 中数据集的设置保持一致：

Food-101: 数据集包含 101 类，其中 71 类为训练集，30 类为测试集。

VIREO Food-172: 包含 172 类，其中 132 类为训练集，40 类为测试集。

ChineseFoodNet: 包含 208 类, 其中 158 类作训练集, 50 类为测试集。

4.2.2 实现细节

本节主要介绍模型的初始化方式以及模型参数设置。

(1)模型的初始化: 在实验过程中, 模型初始化对模型训练至关重要。首先模型初始化对三元组采样有决定性作用, 关系网络作为采样三元组的度量方法, 如果使用随机初始化, 则无法挑选出适合训练的三元组。同样, 图像的嵌入网络作为关系网络提供输入的关键, 图像嵌入网络的初始化也需要符合三元组初始筛选的条件。为此本章分别对图像嵌入网络和关系网络采用合适的预训练模型初始化。

对于图像嵌入网络, 初始化是采用基于训练集得到的分类模型参数, 目的是可以获得具有区分性的特征, 具体方法如下, 首先使用基于 ImageNet 数据集预训练的模型^[44]初始化 VGG16 深度神经网络, 然后利用训练集数据训练一个分类模型。最终得到一个具有训练集类别区分度的分类模型。

对于关系网络的初始化, 首先利用经过训练集微调过的 VGG16 分类模型, 分别提取出训练集与测试集最后一个卷积层的特征, 在基于“5-way 1-shot”的训练机制下使用均方误差损失训练关系网络, 最终利用测试集上准确率最高的一组参数作为非线性关系学习网络的初始化。

采用此初始化方案的优点如下: (a)模型获得一个具有训练集类别区分度的非线性分类器, 有利于三元组的在线采样; (b)测试集信息在微调网络过程中没有流入整个模型中; (c)相比于随机初始化, 使用预训练的初始化方案, 模型的收敛效果更好。

(2)参数设置: 对于模型的优化算法, 采用的是 Adam^[43]优化算法, Adam 优化算法相比于随机梯度下降, 模型的收敛速度更快且更稳定。在实验中使用 Adam 优化算法的默认超参数设置($\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-8}$)。

只考虑“5-way 1-shot”的训练机制。由于网络模型的复杂度较高, 批量样本大小设置为 49, 包含 $P=7$ 个类别以及每个类别随机选取 $K=7$ 个样本。由于模型初始参数是采用预训练的参数, 因此训练过程使用较低的初始学习率为 10^{-5} , 训练过程中微调图像特征嵌入网络最后一个卷积层以及整个非线性关系学习网络。对于其他实验参数设置, 铰链损失函数的阈值为 0.9, “limited batch hard”参数设置为(0.4, 0.6)。为防止过拟合现象, 在每个卷积层中本文加入批正则化处理和 ReLU 非线性层, 全连接层的 dropout 率设置为 0.5, 模型在训练大约 20000 次左右就可以收敛。

4.2.3 实验结果与分析

本节的实验结果主要验证不同数据集上所提方法与不同小样本方法的效果对比，以及实验参数敏感度的实验结果。

(1)方法对比实验：为验证模型的性能，分别在 Food-101、VIREO Food-172 与 ChineseFoodNet 三个数据集上，选取流行的小样本方法 Siamese Network^[12]、Matching Network^[13]、Relation Network^[14]和基于平方欧式距离的三元卷积神经网络^[44]作为基准方法，进行全面对比试验。

本章算法与其他文献算法的对比如表 4.1 所示，从表 4.1 中可以看出：(1)基于三元卷积神经网络的方法要比基于孪生网络的小样本方法取得了更好的性能。主要原因是三元卷积神经网络相较于孪生网络可以学习到更加细粒度的类别差异信息，可以提取更加具有辨识度的特征，因此模型的识别性能更强，具体而言，对比三个数据集的实验结果，相较 Siamese Network 性能分别提高 7.8%、16.6% 和 15.6%，相较 Matching Network 的性能分别提高 11.3%、1.6%和 17.2%，同样相较 Relation Network 提高 3.0%、2.9%和 2.3%。(2)关系网络作为可学习的非线性度量函数，可以根据数据集和特征嵌入的网络结构自适应学习，其识别能力更强。与基于平方欧氏距离的三元卷积神经网络相比，模型性能分别提升 0.8%、1.7%和 3.7%。最终所有实验结果表明，基于非线性度量学习的三元神经网络相较于基于孪生网络的方法和基于线性距离的三元卷积神经网络可以取得更好的性能。

(2)铰链函数阈值 m 的敏感度分析：为了探究阈值对于实验的影响，设计了基于不同阈值的对比实验。如图 4.4 所示，随着阈值的逐渐提升，模型的识别能力呈现出先提升后下降的趋势，当阈值为 0.9 时，模型的性能达到最高。阈值作为铰链损失函数的一部分，控制着正样本图像对与负样本对之间的距离学习；如果阈值设置过小，则正样本图像对与负样本图像对之间的区分度不够，很难通过训练得到具有区分性的模型。

表 4.1 不同小样本方法在不同数据集上的实验结果 (单位：%)

Table 4.1 The experimental results of different datasets

模 型	度量方法	Food-101	VIREO Food-172	ChineseFoodNet
Siamese Network	余弦距离	49.1	60.3	50.5
Matching Network	余弦距离	45.6	73.6	48.9
Relation Network	非线性距离	53.9	74.0	63.8
Triplet Network	平方欧式距离	56.1	75.2	62.4
本文所提方法	非线性距离	56.9	76.9	66.1

(3)三元组采样参数 α 和 β 的敏感度分析:本文在 VIREO Food-172 数据集上分析三元组采样方案的参数敏感度,通过设置“limited batch hard”中 α 和 β 的取值,不断的缩小三元组采样的范围。通过对比“batch hard”的采样方案,验证了提出的采样规则的有效性,实验结果如表 4.2 所示。

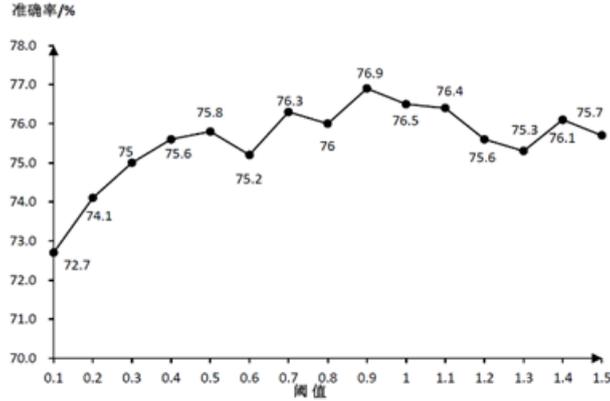


图 4.4 不同阈值的实验结果

Fig.4.4 Experimental results with different margins

表 4.2 三元组采样的对比实验结果

Table 4.2 The experimental results of triplet with different setting

模 型	准确率/%
batch hard	56.2
(0.1, 0.9)	72.7
(0.2, 0.8)	73.9
(0.3, 0.7)	75.5
(0.4, 0.6)	76.9
(0.5, 0.5)	75.5

从表 4.2 中可以看出: (1)“batch hard”三元组采样方案并不适用于基于非线性度量的三元卷积神经网络的模型,模型性能只能达到 56.2%。主要原因是基于非线性度量的“batch hard”方法会采样出许多对训练不利的三元组,导致模型收敛效果差。(2)基于“limited batch hard”三元组采样规则,模型可以更好的收敛,随着“limited batch hard”的加入以及限制逐渐加强,模型的识别能力从 72.7%开始提升,当限制正样本图像对关系值大于等于 0.4,且负样本图像对关系值小于等于 0.6 时性能达到 76.9%,相比于“batch hard”性能提升 20.7%。“limited batch hard”利用正负样本图像对的关系值,限制三元组的选择,可以采样出更加适合融合关系网络和三元卷积神经网络的三元组,最终相比于“batch hard”挑选规则,模型训练更加稳定,准确率更高。

(4)模型初始化分析:初始化参数对于本章所提出的模型十分重要,在 VIREO Food-172 数据集上探究了不同的初始化方案对模型性能的影响。为了分析特征嵌入网络与关系网络不同初始方案,使用“微调”表示初始化方案特定的设置,具体含义如下:对于特征嵌入网络,“否”代表使用 ImageNet 预训练模型初始化,“是”代表由数据集的训练集预训练模型初始化;对于关系网络,“否”代表随

机初始化，“是”代表由训练集预训练的模型初始化。实验结果验证了本章提出的模型参数初始方案的重要性。从表 4.3 可以看出，第一，关系网络的初始化对于模型的训练是至关重要的。如果采用随机初始化，无论特征嵌入网络采用何种初始化方案，模型在三个数据集上无法收敛。关系网络是采样三元组的度量方法，如果非线性网络使用随机初始化，则挑选出的三元组对于三元神经网络的训练是不具有参考性的，因此需要对关系网络一个较好得到初始化。第二，特征嵌入网络的初始化对实验性能有影响，相比于 ImageNet 初始化，基于训练集的初始化可以获得更好的性能，在三个数据集上性能提升大约 3.0%。主要原因是特征嵌入网络为三元组选择提供特征表示，特征表示区分程度也决定采样的三元组是否适合训练。因此最终采用基于训练集获得的分类模型参数用于模型的初始化，并取得性能的提升。

表 4.3 不同参数初始方案的实验结果(单位:%)

Table 4.3 The experimental results of different parameter initialization

微调(特征嵌入网络)	微调(关系网络)	Food-101	VIREO Food-172	ChineseFoodNet
否	否	—	—	—
是	否	—	—	—
否	是	53.2	73.3	60.9
是	是	56.3	76.9	66.1

4.2.4 数据集性能差异探讨



图 4.5 Food-101 与 VIREO Food-172 部分训练集和测试集对比

Fig 4.5 Samples comparison of training set and test set in Food-101 与 VIREO Food-172

从实验那结果中可以发现，本章提出的模型在 VIREO Food-172 和 ChineseFoodNet 数据集上的实验性能要明显优于 Food-101 数据集，可能的原因在于训练集和测试集的差异程度。基于度量学习的小样本识别算法属于迁移学习

的范畴,在小样本学习的设置下,训练集与测试集完全不相关,从训练集学习的可迁移知识对于测试的准确性至关重要。如果训练集与测试集的样本分布差距很大,小样本方法在训练阶段无法获得对测试集有用的迁移信息。反之,如果训练集和测试集之间有相似的分布,如外观和形状等,则性能可能会提高。如图 4.5 所示,一些来自 Food-101 和 VIREO Food-172 的图片样本,对于 Food-101,训练集与测试集在光线和色彩的差异较大,训练集许多类别的样本色彩比较明亮且主要以红色和黄色为主,然而在测试集中许多类别样本色彩偏暗且主要以灰色和白色为主。训练集与测试集的较大差异,会导致从训练集获得的迁移信息对测试集的指导能力有限。对于 VIREO Food-172 数据集,训练集与测试集在外观和颜色等方面有许多相似之处,如训练集中的“鲫鱼汤”和测试集中的“清蒸鱼”,在这两个类别中大部分图像内容的形状相似且颜色均以白色和灰色为主。训练集与测试集之间的这种相似的分布使得基于训练集的迁移信息更好的指导测试集以获得更好的性能。

4.3 本章小结(Summary)

本章提出一个融合非三元卷积神经网络和关系网络的小样本食品识别模型,该模型可以学习到图像类内和类间的区分信息,同时关系网络作为可学习的非线性度量函数可以更好地指导图像特征嵌入网络的学习和更好的契合数据集的特性。为训练该模型,进一步提出一个新的基于关系网络的三元组的采样规则“limited batch hard”,新的样本采样方法可以有效的筛选去除对训练不利的三元组。在不同的食品数据集上进行全面实验,所有的实验结果证明方法的有效性,同时通过相比于现有流行的样本学习方法,本章模型取得更好的性能。

5 总结与展望(Summary and Expectation)

5.1 工作总结(Summary)

本文主要针对深度学习无法识别只有少量样本的食品识别,通过查阅大量的文献,实现了食品的小样本学习,并且提出两个小样本食品识别的改进模型。第一,本文提出一个同时考虑食品类别信息和原料信息的小样本学习框架(MVFSL)。该框架主要由三部分组成:类别信息与原料信息特征学习子网络、多角度特征融合和多角度信息关系学习子网络。本框架首次实现解决小样本食品识别问题。与传统的小样本学习研究方法不同,本框架不仅考虑到食品图像的类别信息,同时引入丰富的食品原料信息,构建了一个端到端学习的多视角小样本食品识别模型。由于食品原料信息在不同的食品类别中的组合是不同的,食品原料信息可以作为类别信息的互补信息,通过融合两种信息可以获得更加丰富且具有区分性的图像信息,因此本框架相较于只考虑类别信息的小样本学习模型可以获得更好的性能。第二,本文提出一个基于非线性度量学习三元神经网络用于小样本食品识别。模型主要由两部分组成:用于图像特征学习的特征嵌入子网络和用于非线性度量函数学习的关系学习子网络。此外为了模型训练的稳定性,本文提出一个新的样本在线采样方案。较于基于孪生网络的小样本识别模型,三元卷积神经网络代替孪生网络以学习食品图像在类内与类间更加细粒度的区分信息,从而提升小样本食品图像识别的分类性能。基于线性度量的三元神经网络对于复杂的食品图像的鉴别能力不够强,提出使用可学习的关系网络作为非线性度量方案,同时融合三元神经网络,进一步提出基于非线性度量学习的三元神经网络,相较于孪生网络和基于线性度量的三元神经网络,本文提出的模型可以获得更具判别性的信息,从而提升小样本食品分类的性能。

5.2 研究工作展望(Expectation)

本文还存在一些未研究的问题,需要进一步研究:(1)可以采用不同的特征融合策略来实现小样本食品的识别。在当前的框架中,直接将两种类型的特性连接到一个统一的表示中。未来的工作还可以将融合层引入到的方法中来学习共享表示。(2)从食品数据集中挖掘更多的信息,提高小样本食品识别的性能。例如,除了原料信息外,还可以使用切削和烹饪属性或烹饪说明。(3)在研究过程中,发现训练集和测试集分布的差异会影响小样本学习的效果,如何处理这种差异是一个值得研究的问题。

有许多问题值得进一步探究:第一,可以探索食品数据集中更加丰富的信息提升性能。例如食品属性信息、地理位置信息和食谱信息等。第二,由于本文所提出的方法具有普适性,因此可以将此方法应用到更多的领域中,例如 Omniglot^[12]和 MiniImageNet^[13]等。第三,可以采用更多优化算法降低计算复杂度^{[45]-[49]}。

参考文献

- [1] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101—mining discriminative components with random forests[J]. In European Conference on Computer Vision, 2014, pages 446–461.
- [2] S. Ao and C. X. Ling. Adapting new categories for food recognition with deep representation[J]. In IEEE International Conference on Data Mining Workshop, 2015, pages 1196-1203.
- [3] L. Herranz, S. Jiang, and R. Xu. Modeling restaurant context for food recognition[J]. IEEE Transactions on Multimedia, 2017, pages 430 – 440.
- [4] K. Aizawa, Y. Maruyama, H. Li, and C. Morikawa. Food balance estimation by using personal dietary tendencies in a multimedia food log[J]. Multimedia, IEEE Transactions on, 2013, 15(8):2176–2185.
- [5] J. Zheng, Z. J. Wang, and C. Zhu. Food image recognition via superpixel based low-level and mid-level distance coding for smart home applications[J]. Sustainability, 2017, 9(5):856.
- [6] A. E. Mesas, M. Muozpareja, E. Lpezgarca, and F. Rodriguezartalejo. Selected eating behaviours and excess body weight: a systematic review[J]. Obesity Reviews An official Journal of the International Association for the Study of Obesity, 13(2):106, 2012.
- [7] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, G. Papandreou, J. Huang, and K. P. Murphy. Im2calories: towards an automated mobile vision food diary[J]. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1233 – 1241, 2015.
- [8] M. Bolanos, A. Ferra, and P. Radeva. Food ingredients recognition through multi-label learning[J]. In International Conference on Image Analysis and Processing, 2017, pages 394–402.
- [9] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Part-based r-cnns for fine-grained category detection[J]. In European Conference on Computer Vision, 2014, pages 834–849.
- [10] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou. Food recognition for dietary assessment using deep convolutional neural Networks[J]. In International Conference on Image Analysis and Processing, 2015, pages 58–465.
- [11] Martinel, Niki, G. L. Foresti, and C. Micheloni. Wide-Slice Residual Networks for Food Recognition[J]. IEEE Winter Conference on Applications of Computer Vision IEEE Computer Society, 2018, pages 567-576.
- [12] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural Networks for one-shot image recognition[J]. In International Conference on Machine learning, 2015, volume2.
- [13] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra. Matching Networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, pages 3630–3638.
- [14] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales. Learning to compare: Relation Network for few-shot learning[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2017.

- [15] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep Networks[J]. arXiv preprint arXiv:1703.03400, 2017.
- [16] M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas. Learning to learn by gradient descent by gradient descent[J]. In *Advances in Neural Information Processing Systems*, 2016, pages 3981–3989.
- [17] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taix’e, D. Cremers, and L. Van Gool. One-shot video object segmentation[J], 2017.
- [18] E. Hoffer and N. Ailon. Deep metric learning using triplet Network[J]. In *International Workshop on Similarity-Based Pattern Recognition*, 2015.
- [19] M. Geng, Y. Wang, T. Xiang, and Y. Tian. Deep transfer learning for person re-identification[J]. arXiv preprint arXiv:1611.05244, 2016.
- [20] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification[J]. arXiv preprint arXiv:1703.07737, 2017.
- [21] Y.Li, Yan H, et al. Deep joint discriminative learning for vehicle re-identification and retrieval[J]. *IEEE International Conference on Image Processing. IEEE*, pages 395-399, 2017.
- [22] J. Chen and C.-W. Ngo. Deep-based ingredient recognition for cooking recipe retrieval[J]. In *Proceedings of the ACM International Conference on Multimedia*, pages 32–41, 2016.
- [23] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar. Food recognition using statistics of pairwise local features[J]. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2249–2256, 2010.
- [24] T. Joutou and K. Yanai. A food image recognition system with multiple kernel learning[J]. In *IEEE International Conference on Image Processing*, pages 285–288, 2010.
- [25] Y. Kawano and K. Yanai. Food image recognition with deep convolutional features[J]. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pages 589–593.
- [26] H. Kagaya, K. Aizawa, and M. Ogawa. Food detection and recognition using convolutional neural Network[J]. In *Proceedings of the ACM International Conference on Multimedia*, pages 1085–1088, 2014.
- [27] R. Xu, L. Herranz, S.Jiang, et al. Geolocalized Modeling for Dish Recognition[J]. *IEEE Transactions on Multimedia*, 2015, pages 1187-1199.
- [28] W. Min, S. Jiang, J. Sang, H. Wang, X. Liu, and L. Herranz. Being a super cook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration[J]. *IEEE Transactions on Multimedia*, 2017, (5):1100–1113.
- [29] W. Min, B. K. Bao, S. Mei, Y. Zhu, Y. Rui, and S. Jiang. You are what you eat: Exploring rich recipe

- information for cross-region food analysis[J]. *IEEE Transactions on Multimedia*, 2017, 20(4):950–964.
- [30] H. Wang, W. Min, X. Li, and S. Jiang. Where and what to eat: Simultaneous restaurant and dish recognition from food image[J]. In *Pacific Rim Conference on Multimedia*, 2016, pages 520–528.
- [31] Mei S H, Min W Q, Liu L H, et al. Faster R-CNN based food image retrieval and classification [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2017(6).7(1): 189-206 (in Chinese)
(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(6).7(1): 189-206.)
- [32] L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 28(4):594–611, 2006.
- [33] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum. One-shot learning by inverting a compositional causal process[J]. In *International Conference on Neural Information Processing Systems*, pages 2526–2534, 2013.
- [34] Paul E Utgoff. Shift of bias for inductive concept learning[J]. *Machine learning: An artificial intelligence approach*, 1986.
- [35] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning[J]. *Artificial Intelligence Review*, 2002.
- [36] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta learning with memory-augmented neural networks[J]. In *International Conference on Machine Learning*, 2016.
- [37] Pablo Sprechmann, Siddhant M Jayakumar, Jack W Rae, Alexander Pritzel, Adrià Puigdomènech Badia, Benigno Uria, Oriol Vinyals, Demis Hassabis, Razvan Pascanu, and Charles Blundell. Memory-based parameter adaptation[J]. 2018.
- [38] Mishra N, Rohaninejad M, Chen X, et al. A simple neural attentive meta-learner[J]. *arXiv preprint arXiv:1707.03141*, 2017.
- [39] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks[J]. *CoRR*, abs/1311.2901, 2013.
- [40] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization[J]. In *IEEE International Conference on Computer Vision Workshop*, pages 618–626, 2017.
- [41] M. Bolaños, A. Ferr`a, and P. Radeva. Food ingredients recognition through multi-label learning[J]. In *International Conference on Image Analysis and Processing*, pages 394–402. Springer, 2017.
- [42] X. Chen, H. Zhou, Y. Zhu, and L. Diao. Chinesefoodnet: A largescale image dataset for chinese food

- recognition[J]. arXiv preprint arXiv:1705.02743, 2017.
- [43] D. Kingma and J. Ba. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [44] K. Simonyan, A. Zisserman. Very deep convolutional Networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [45] 王薇, 袁琪, 唐加会. 求解箱式约束全局优化问题的降维算法[J]. 数学建模及其应用, 2019, 8(01):38-43.
- [46] 孙诚, 杨磊, 王志海. 一种针对彩色视频图像中目标检测的改进模型[J]. 数学建模及其应用, 2017, 6(01):23-27+71.
- [47] 林森林, 徐圣兵, 方桂标, 等. 基于超声同步观测辅助技术的 CT 成像系统设计及图像融合[J]. 数学建模及其应用, 2019, 8(01):14-22.
- [48] 刘钱源, 杨欢欢, 刘培鑫, 张承进. 基于监控视频的前景目标提取[J]. 数学建模及其应用, 2018, 7(01):63-71.
- [49] 杨雨龙, 韩丛英. 错位指纹图像自动检测的深度学习[J]. 数学建模及其应用, 2018, 7(02):23-29+2.

作者简介

一、基本情况

姓名：吕永强 性别：男 民族：汉 出生年月：1992-06-14 籍贯：山东省潍坊市

1. 2011.09-2015.07 山东科技大学学士

2. 2016.09-2019.07 山东科技大学硕士

二、攻读学位期间发表学术论文情况

1. 吕永强,闵巍庆,段华,蒋树强。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在计算机科学期刊(自然科学版)(中文核心)。

2. 蒋树强,闵巍庆,吕永强,刘林虎. Few-Shot Food Recognition via Multi-view Representation Learning. IEEE Transactions on Multimedia(TMM) (在投)

三、攻读学位期间获奖情况

1. 2016年 校二等奖学金

四、攻读学位期间研究项目情况

1. 国家自然科学基金，多模态跨域适应性食品图像识别与理解.项目编号 6191101109. 参与人员

致谢

论文能够顺利完成，首先得感谢我研究生学习期间的导师段华副教授。从论文开题到撰写完成期间，得到了段华老师的悉心指导，在论文的细节部分，给出了宝贵意见和独到的见解。同时还要感谢段华老师在学习和科研生活上给予的帮助。

在攻读山东科技大学硕士学位这三年期间，我感谢数学学院给本文提供良好的学习环境以及齐全的学习工具。感谢实验室里所有的同学对我学习和生活上的帮助。在中国科学院计算技术研究所客座期间，感谢蒋树强研究员和闵巍庆副研究员对我在科研工作和学习生活上的悉心指导，同时还要感谢实验室同学给予的无私的帮助。

其次，我要感谢室友孙志鹏、余阳、王宁，他们对我生活上和学习上的帮助最大，这三年能与他们朝夕相处，是我人生中莫大的幸运。

最后，我感谢我的父母，感谢他们对我生活上无私的帮助，感谢他们支持和理解，感谢他们为这个家无私地奉献。

学位论文数据集

关键词*	密级*	中图分类号*	UDC	论文资助
小样本食品识别	公开	TP301.6	510	
学位授予单位名称*	学位授予单位代码*	学位类别*	学位级别*	
山东科技大学	10424	理学	硕士研究生	
论文题名*		并列题名*	论文语种*	
面向小样本食品图像识别方法研究		Research of Few-Shot Food Recognition	中文	
作者姓名*	吕永强	学号*	161520048	
培养单位名称*	培养单位代码*	培养单位地址	邮编	
山东科技大学	10424	山东省青岛市	266590	
学科专业*	研究方向*	学制*	学位授予年*	
系统分析与集成	系统建模与优化	3	2019	
论文提交日期*				
导师姓名*	段华	职称*	副教授	
评阅人		答辩委员会主席*	答辩委员会成员	
电子版论文提交格式 文本() 图像() 视频() 音频() 多媒体() 其他()				
推荐格式: application/msword; application/pdf				
电子版论文出版(发布者)	电子版论文出版(发布地)		权限声明	
论文总页数*				
注: 共 33 项, 其中带*为必填数据, 共 22 项。				



理学硕士学位论文

小样本食品图像识别方法研究与应用
Research and Application of Few-Shot Food Recognition

作 者 吕永强

导 师 段 华

山东科技大学
二零一九年五月