



# 小样本食品图像识别@宜宾

闵巍庆  
中国科学院计算技术研究所  
智能信息处理重点实验室  
2018-11-19



# 研究动机

- 人类在历史上面临着识别食物的艰巨任务以满足人类的生存与营养需求



Article | [OPEN](#) | Published: 15 December 2011

## Flavor network and the principles of food pairing

Yong-Yeol Ahn, Sebastian E. Ahnert✉, James P. Bagrow & Albert-László Barabási✉

*Scientific Reports* **1**, Article number: 196 (2011) | [Download Citation](#) ⬇



## 食品图像识别



食品图像组织和管理



移动饮食记录  
mobile food Journaling



能量摄入估计  
food calorie estimation



POI 推荐



饮食偏好和健康分析

- 人类能够通过泛化一个例子就可以学习新的概念，拥有单样本学习的能力
- 对食品图像新类别的学习同样如此



- 食品数据与其他的物体类别数据相似，其分布是长尾的，以至于许多食品类别只有少量的样本



Food name: Wagafi bread  
Baidu search : 10  
Bing Search: 20  
Google Search: 20



Food name: Babute  
Baidu search : 0  
Bing Search: 16  
Google Search: 20

# 我们需要研究 小样本的食品图像识别





# 相关工作-食品图像识别

- 食品图像识别的代表性工作

基于手工特征的食品图像识别

Yang et al. CVPR'10

结合深度视觉特征和餐馆信息

Xu et al. TMM'15

设计针对食品特点的计深度学习结构

Martinel et al. WACV'18

Yanai et al. ICPR'12  
Yanai et al. IECC'14  
Bossard et al. ECCV'14

使用深度学习特征

Chen et al. ACM MM'16  
Zhou et al. CVPR'16

结合深度视觉特征和原料信息



# 相关工作-食品图像识别

## • 相关食品图像数据集

数据集: ETHZ Food-101  
类型: 图片和类别  
数量: 101,000/101

**Bossard et al.  
ECCV'14**

**Yanai et al. ICPR'12  
Yanai et al. IECC'14**

UEC Food100, 14,361  
UEC Food256, 25,088

数据集: Dishes  
类型: 图片、类别  
和GPS  
数量: 117,504

**Xu et al. TMM'15**

数据集: Vireo Food-172  
类型: 含图片和原料的食谱  
数量: 110,241

**Chen et al.  
ACM MM'16**

**Min et al. TMM'17  
Min et al. TMM'18  
Min et al. ACM MM'17**

Yummly-66K  
类型: 图片和食谱  
数量: 66,000张

数据集: Recipe1M  
类型: 图片和食谱  
数量: 1,000, 000

**Salvador et al.  
CVPR'17**



# 相关工作-食品图像识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

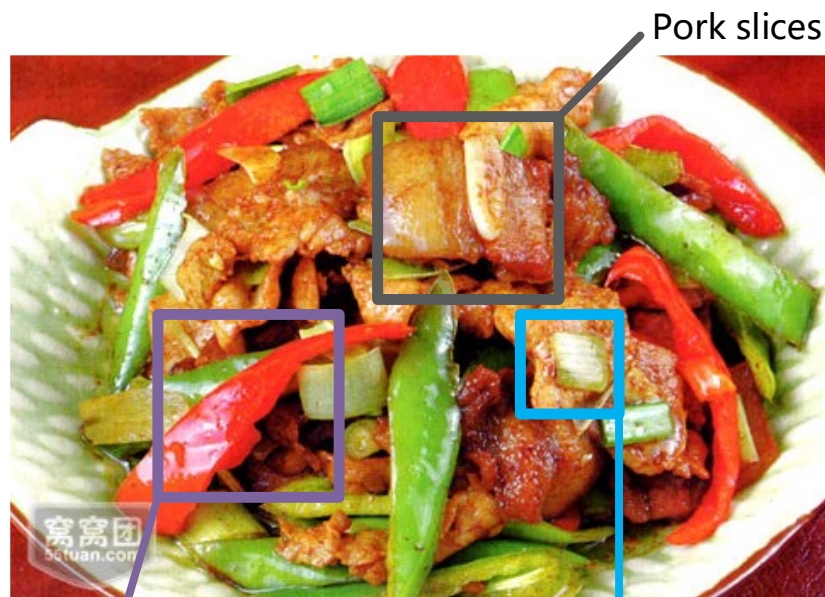
现有的工作**并未对**小样本食品图像识别问题  
进行研究!



# 小样本食品图像识别

## 小样本食品图像识别有其自身的特点

食品图像带有丰富的原料  
属性信息



Pepper slices

Minced green onion

**Category:** Shredded Pork with pepper  
**Ingredients:** Minced green onion, Pepper slices, Pork slices

小样本食品图像识别是  
细粒度图像识别





## 小样本食品图像识别自身的特点

食品图像带有丰富的原料属性信息



工作1. 基于多角度融合表示的小样本食品图像识别

小样本食品图像识别是细粒度图像识别



工作2. 基于三联体非线性关系网络的小样本食品图像识别



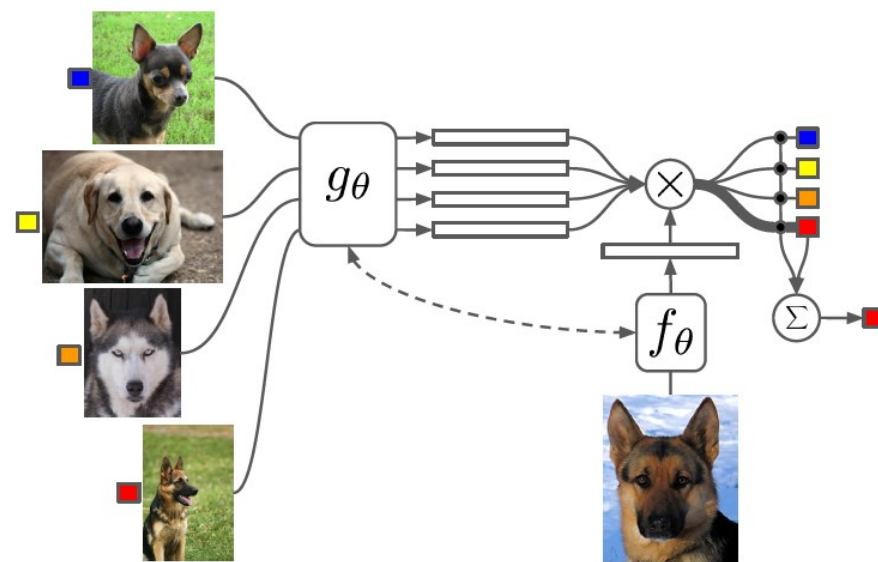
# 预备知识-小样本学习(Few-Shot Learning)

## □ 基本概念

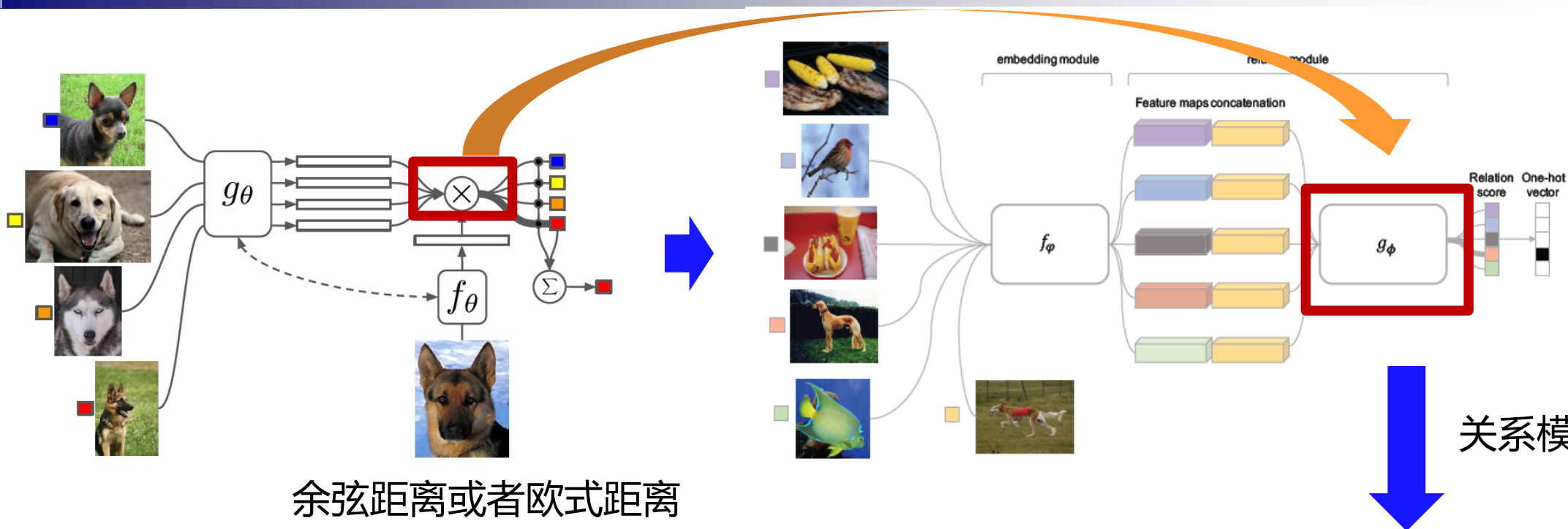
- **支撑集(Support Set)**: 以训练集为例, 随机选取C个类别, 从每个随机选取K个样本, 这C x K个样本称为支撑集。其相应任务称为 “**C-way K-shot**”。
- **查询集(Query Set)**: 从支撑集的C个类别中随机选取一个类, 并且并从中采样N个样本, 这N个样本与Support Set 中该类的样本不同。

## □ 训练和测试

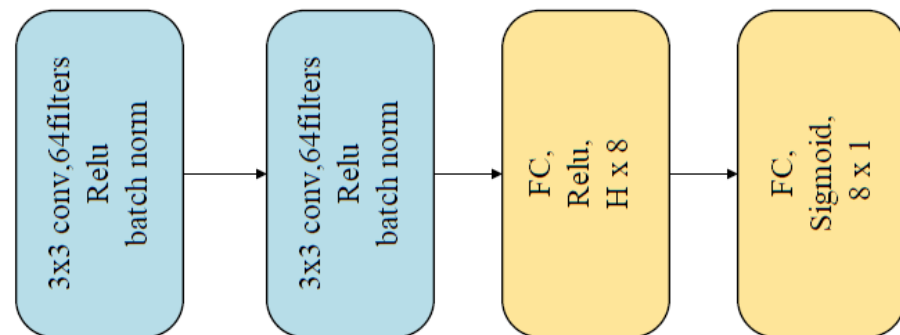
1. 数据集划分: 数据集的训练集与测试集的类别不相交, 即训练集的类别不会出现在测试集中。
2. 训练: 通过度量学习分别对比支撑集与查询集图片, 通过不断地学习使同类的图像相似度更大, 异类图像相似度降低。
3. 测试: 对于训练的模型, 当输入一张查询图像时, 可以基于支撑集的信息选择相似度最高的那个样本的标签作为预测的类的结果。



Matching Network



- **两个卷积块：** 每个卷积块=卷积核为3x3的卷积层+batch norm 层+ReLU激活函数+2x2最大池化层
- **两个全连接层：** ReLu+Sigmoid激活函数





## 小样本食品图像识别自身的特点

食品图像还带有丰富的  
原料属性信息



工作1. 基于多角度融合  
表示的小样本食品图像  
识别

小样本食品图像识别是  
细粒度图像识别

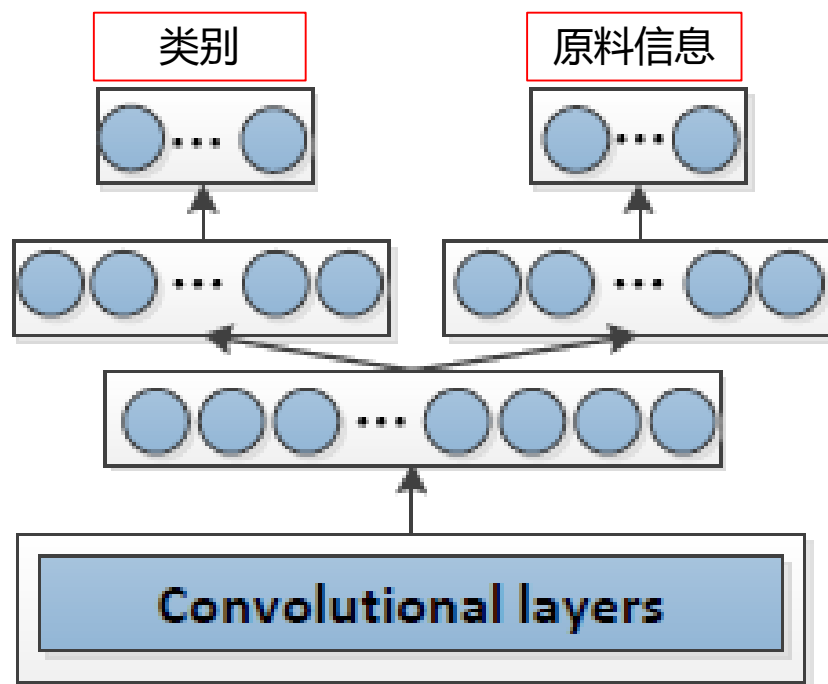


工作2. 基于三联体非线性  
关系网络的小样本食  
品图像识别



# 基于多角度融合表示的小样本食品图像识别

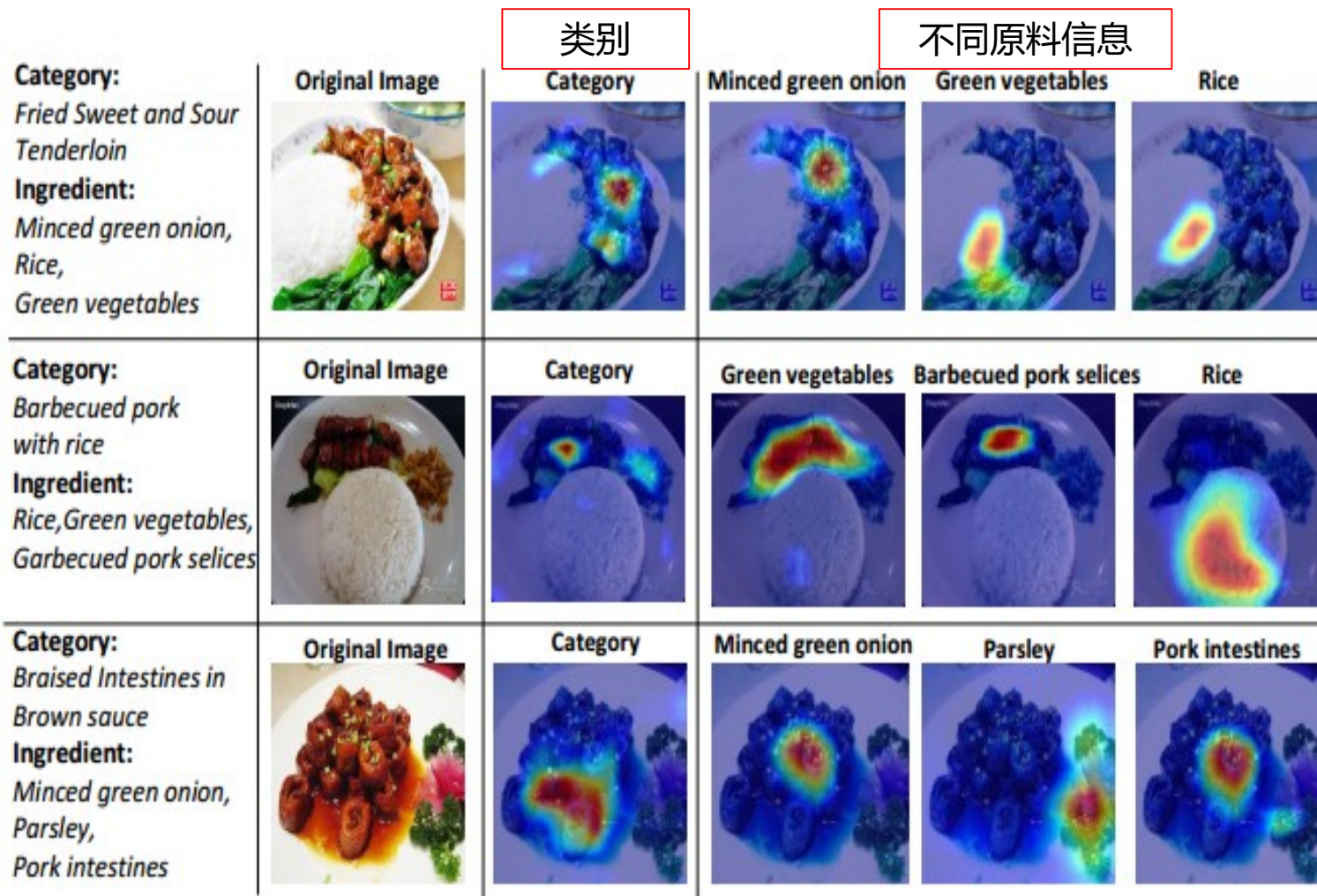
- 融入原料属性监督的多任务学习已被证明可以改进食品图像识别的性能



|             | Method  | Top-1 (%) | Top-5 (%) |
|-------------|---------|-----------|-----------|
| Baseline    | FC7     | 48.02     | 72.01     |
|             | Gist    | 15.39     | 31.85     |
|             | CM      | 16.54     | 39.76     |
| Single-task | AlexNet | 64.91     | 85.32     |
|             | VGG     | 80.41     | 94.59     |
| Multi-task  | Arch-A1 | 78.58     | 94.24     |
|             | Arch-A2 | 78.63     | 94.10     |
|             | Arch-B  | 79.05     | 94.70     |
|             | Arch-C  | 80.66     | 95.05     |
|             | Arch-D  | 82.06 ↑   | 95.88 ↑   |

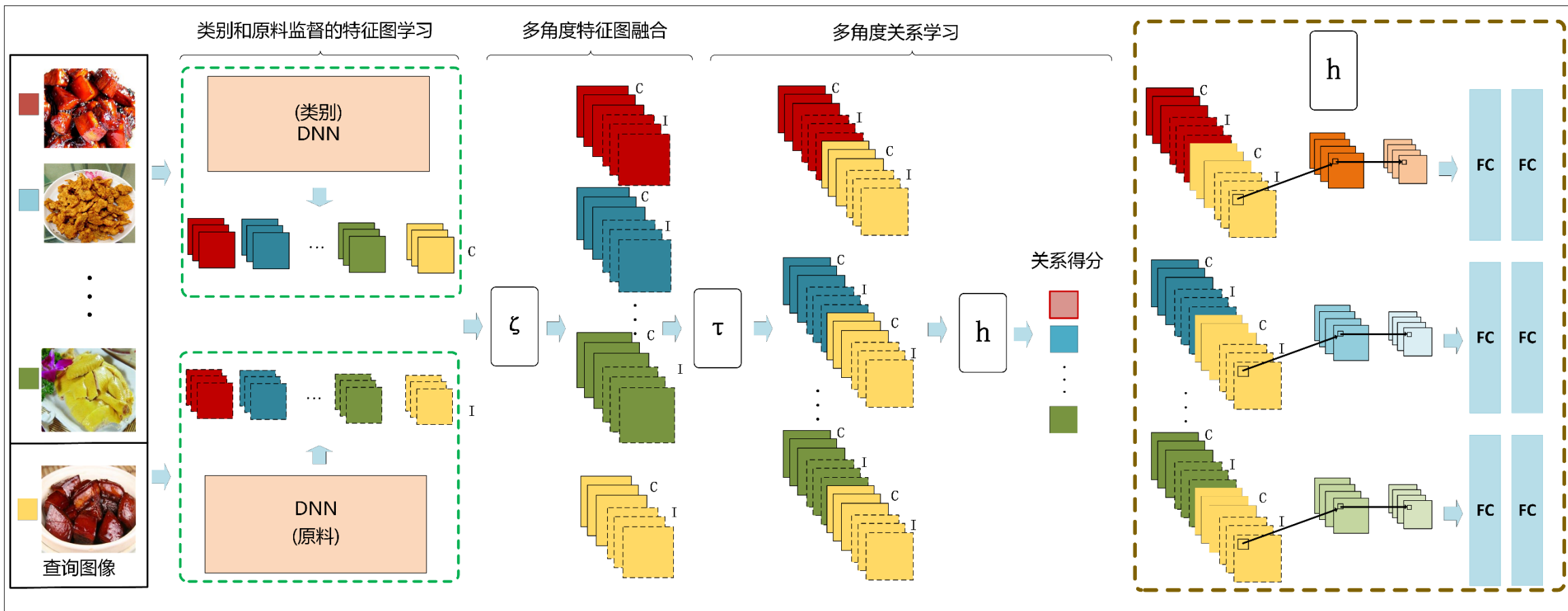
Jingjing Chen and Chong-Wah Ngo. 2016. Deep-based ingredient recognition for cooking recipe retrieval. In Proceedings of the ACM on Multimedia Conference. 32–41.

## 原料属性监督的激活信息与类监督的激活信息的互补性

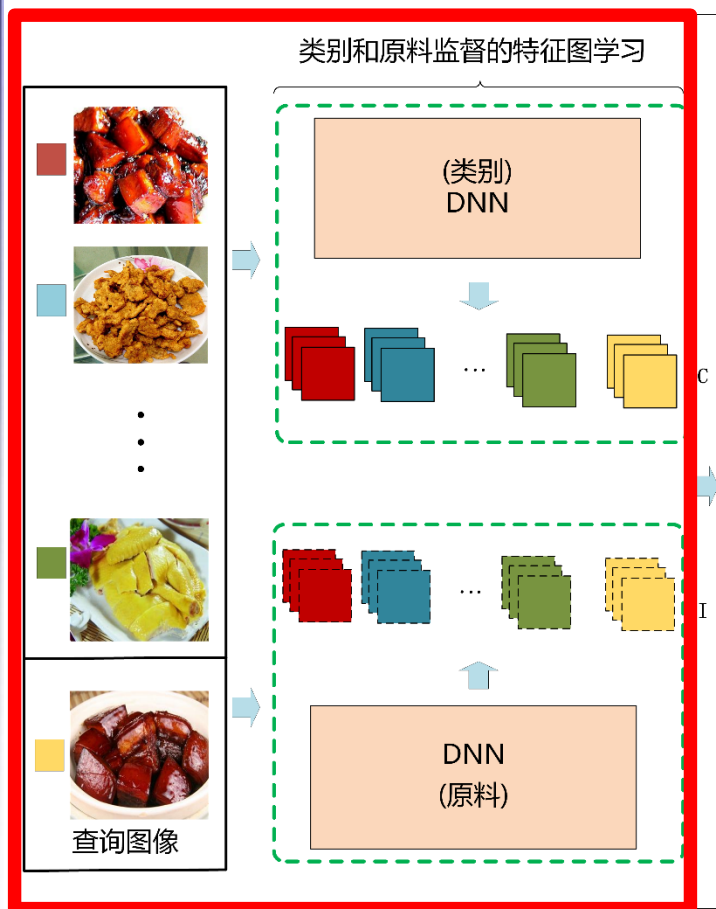


# 基于多角度融合表示的小样本食品图像识别

## □ 框架图

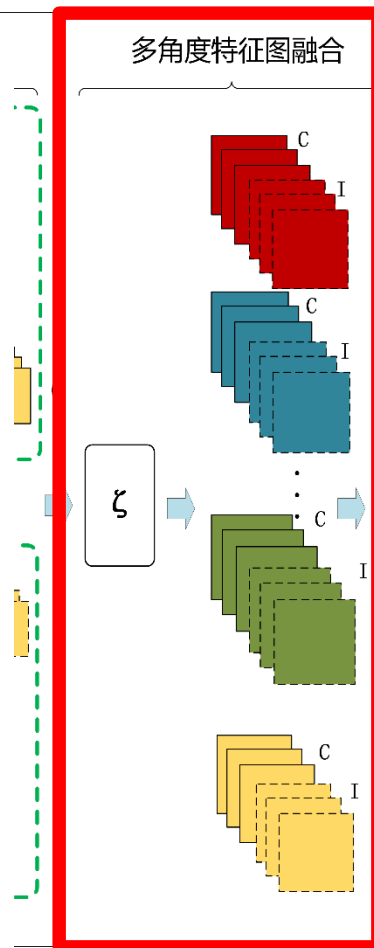


# 基于多角度融合表示的小样本食品图像识别



## 类别和原料监督的特征图(feature map)学习

1. 使用VGG16深度神经网络分别从类别信息监督和原料信息监督的深度网络提取图像特征
2. 采用Conv5层而不是全连接层作为特征，主要考虑食品图像识别属于细粒度图像识别，卷积层含有更多细节信息用于区分不同类

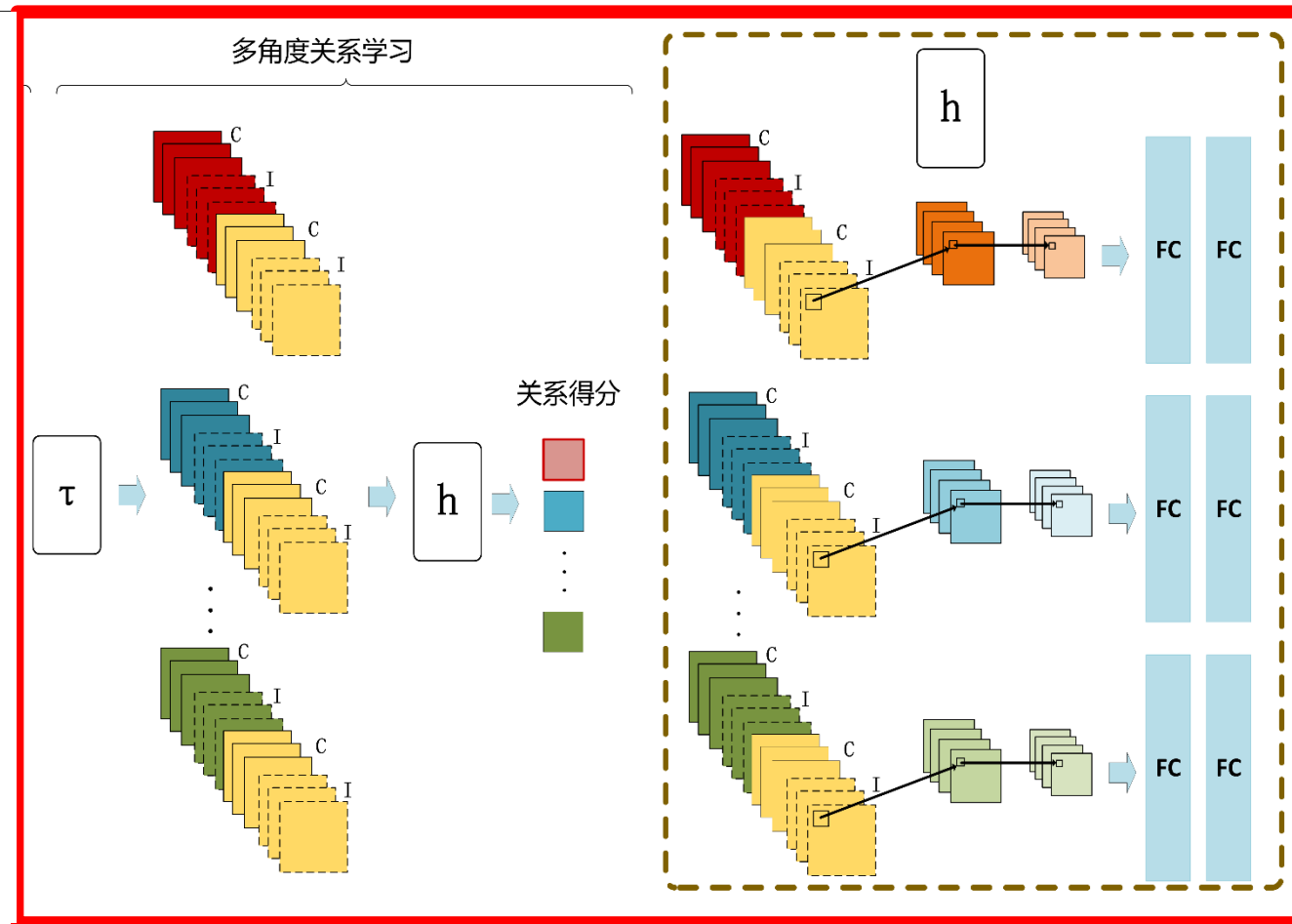


## 多角度特征映射融合

对来自同一张图片不同角度的特征进行融合

## 多角度关系学习

1. 进一步融合support set图像与query set 图像的特征图
2. 通过关系网络进行关系得分计算



# Experiment

| Dataset        | #Num/classes | #training classes | #test classes | #ingredients |
|----------------|--------------|-------------------|---------------|--------------|
| Food101        | 101,000/101  | 71                | 30            | 101          |
| VIREO Food-172 | 110,241/172  | 132               | 40            | 353          |
| ChineseFoodNet | 185,628/208  | 158               | 50            | 353          |

Dataset :

Food-101



**Category:**  
Fish and Chips  
**Ingredient:**  
Old bay , Flour,  
Pepper, Vinegar,  
Beer, Chili,  
Oil, Cod, Potato



**Category:**  
Cheesecake  
**Ingredient:**  
Ham , Cheese  
Sugar , Butter  
Vanilla, Egg



**Category:**  
Hot Dog  
**Ingredient:**  
Cheddar, Onion,  
Relish , Dogs ,  
Chili , Dog bun,  
Mayonnaise, Mustard



**Category:**  
Pizza  
**Ingredient:**  
Cherry , Cheese,  
Oregano , Basil ,  
Olive, Pizza,  
Oil , Salt

Dataset :

VIREO  
Food-172



**Category:**  
Fried Egg  
**Ingredient:**  
Minced green  
onion,  
Crushed pepper,  
Boild egg slices



**Category:**  
Braised Beef with  
Potatoes  
**Ingredient:**  
Minced green  
onion,  
Hob blocks of  
potato,Beef chunks



**Category:**  
Beef Seasoned  
with Soy Sauce  
**Ingredient:**  
Minced green  
onion,  
Beef slices



**Category:**  
Pork Lungs in Chili  
Sauce  
**Ingredient:**  
Black sesame,  
Crushed garlic,  
Beef slices, Chili oil

Dataset :

ChineseFoodNet



**Category:**  
Sirloin tomatoes



**Category:**  
Poached Egg



**Category:**  
Cold Rice Noodles



**Category:**  
Cheese Shrimp  
Meat



# 实验对比

| Model                       | Food-101     | VIREO Food-172 | ChineseFoodNet |
|-----------------------------|--------------|----------------|----------------|
| Relation network-Category   | 53.9%        | 74.0%          | 63.8%          |
| Relation network-Ingredient | 53.5%        | 70.5%          | 64.0%          |
| <b>MVFSL</b>                | <b>56.0%</b> | <b>74.9%</b>   | <b>66.2%</b>   |

| Model                       | Food-101     | VIREO Food-172 | ChineseFoodNet |
|-----------------------------|--------------|----------------|----------------|
| Matching network-Category   | 45.6%        | 73.6%          | 48.9%          |
| Matching network-Ingredient | 46.8%        | 65.9%          | 52.0%          |
| <b>MVFSL</b>                | <b>47.5%</b> | <b>74.1%</b>   | <b>53.0%</b>   |

| Model                      | Food-101     | VIREO Food-172 | ChineseFoodNet |
|----------------------------|--------------|----------------|----------------|
| Siamese network-Category   | 49.1%        | 60.3%          | 50.5%          |
| Siamese network-Ingredient | 54.5%        | 65.5%          | 62.5%          |
| <b>MVFSL</b>               | <b>55.0%</b> | <b>65.8%</b>   | <b>64.4%</b>   |



# 实验对比

## 不同网络结构的对比实验结果

| Model         | Food-101     |              |              | VIREO Food-172 |              |              | ChineseFoodNet |              |              |
|---------------|--------------|--------------|--------------|----------------|--------------|--------------|----------------|--------------|--------------|
|               | AlexNet      | VGG16        | VGG19        | AlexNet        | VGG16        | VGG19        | AlexNet        | VGG16        | VGG19        |
| RN-Category   | 48.6%        | 53.9%        | 54.7%        | 68.8%          | 74.0%        | 74.6%        | 59.5%          | 63.8%        | 62.9%        |
| RN-Ingredient | 51.4%        | 53.5%        | 55.1%        | 70.3%          | 70.5%        | 73.6%        | 59.4%          | 64.0%        | 65.6%        |
| <b>MVFSL</b>  | <b>51.8%</b> | <b>56.0%</b> | <b>56.4%</b> | <b>70.8%</b>   | <b>74.9%</b> | <b>75.7%</b> | <b>60.4%</b>   | <b>66.2%</b> | <b>66.3%</b> |

## 不同Relation network设置在Food-101对比实验结果

| Model        | 64 Filters | 128 Filters  |
|--------------|------------|--------------|
| 1 Conv layer | 55.1%      | 56.7%        |
| 2 Conv layer | 56.0%      | 56.9%        |
| 3 Conv layer | 57.3%      | 57.7%        |
| 4 Conv layer | 57.1%      | <b>57.8%</b> |

## 小样本食品图像识别自身的特点

食品图像还带有丰富的原料属性信息



工作1. 基于多角度融合表示的小样本食品图像识别

小样本食品图像识别是细粒度图像识别

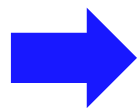


工作2. 基于三联体关系网络的小样本食品图像识别



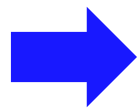
# 动机

连体网络



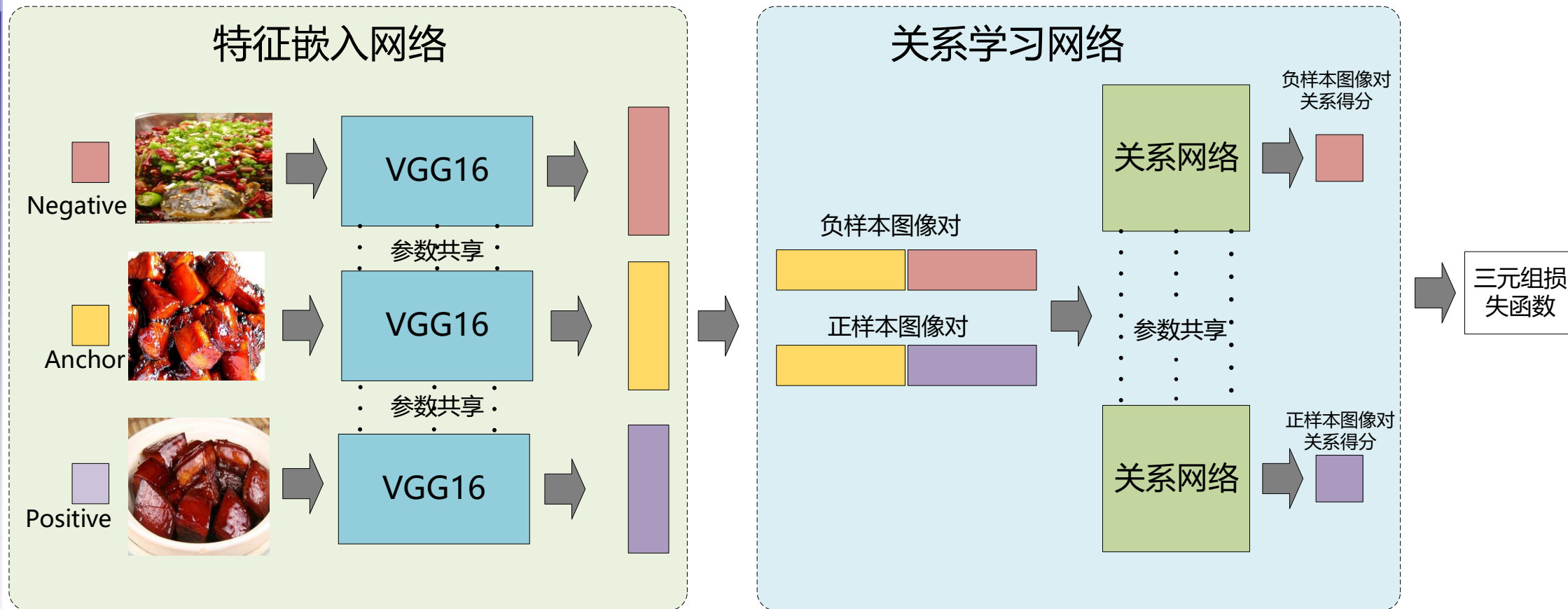
三联体网络

欧式度量或者余弦

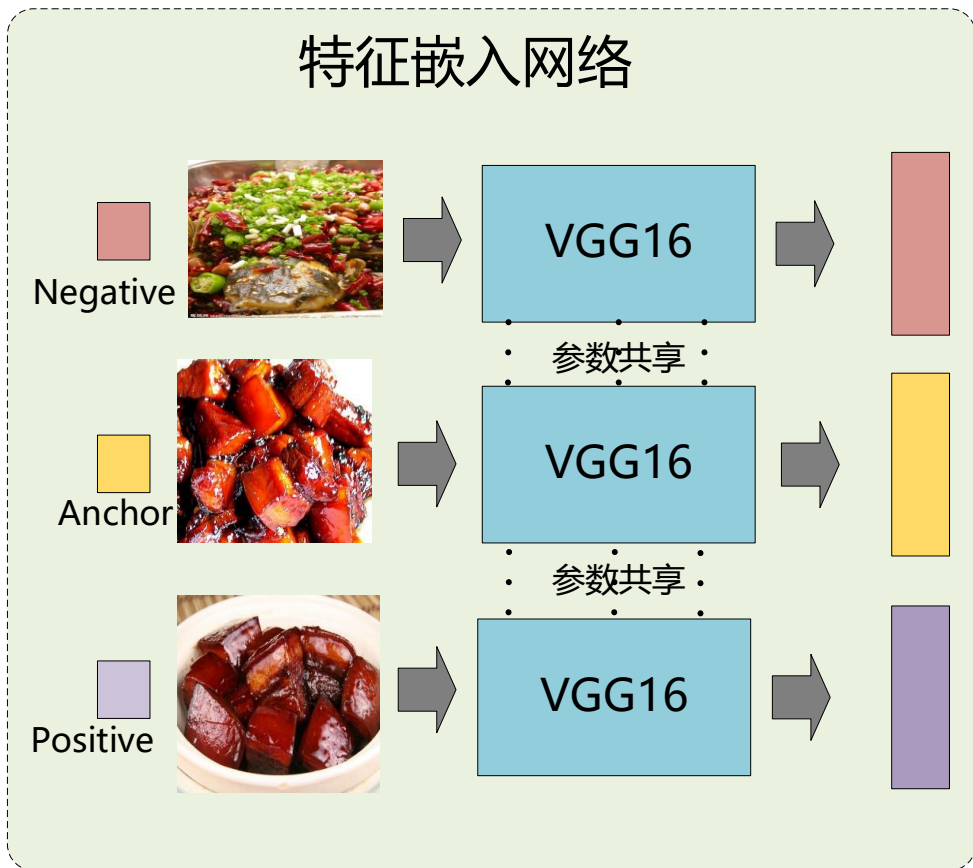


关系网络

# 三连体关系网络



□ 整个过程是一个端到端的训练

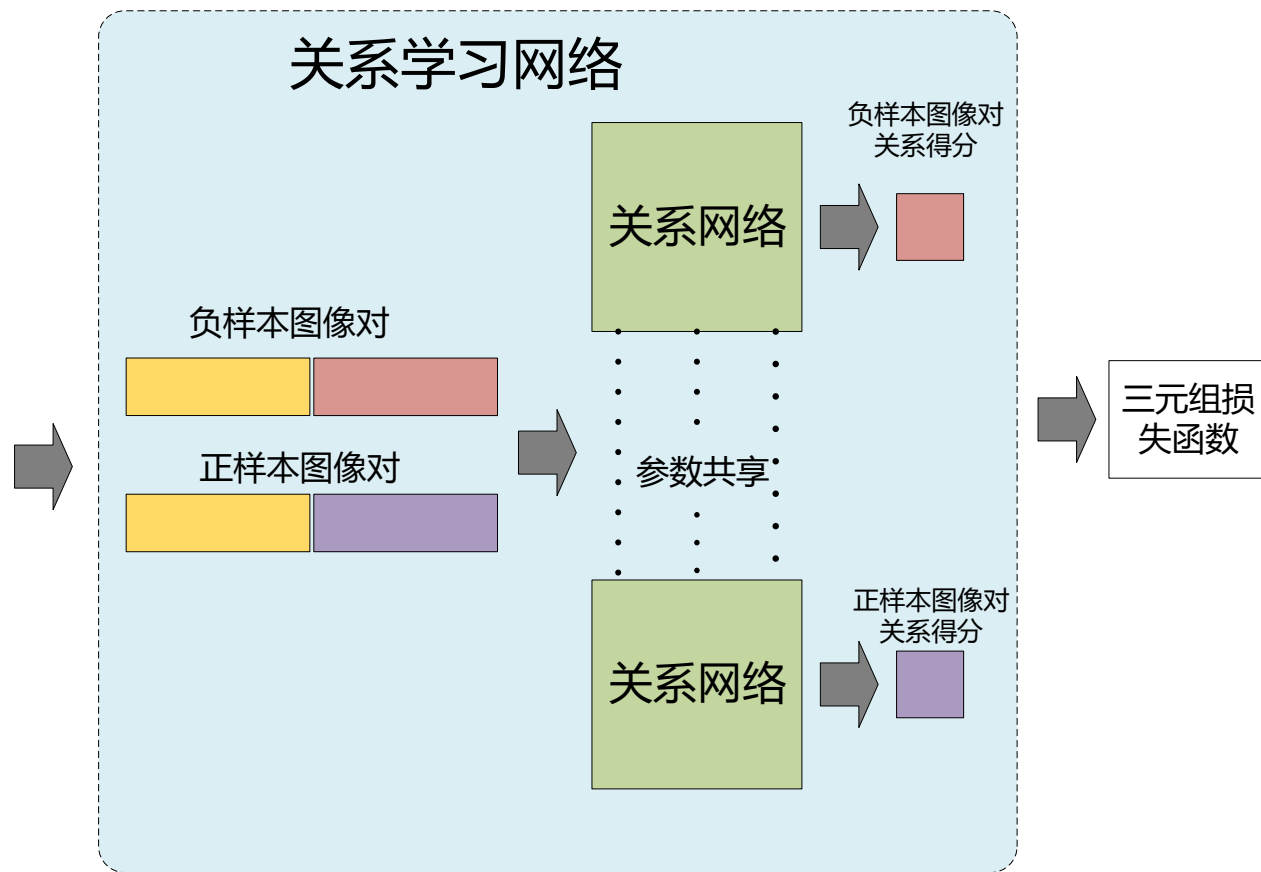


## 特征嵌入网络

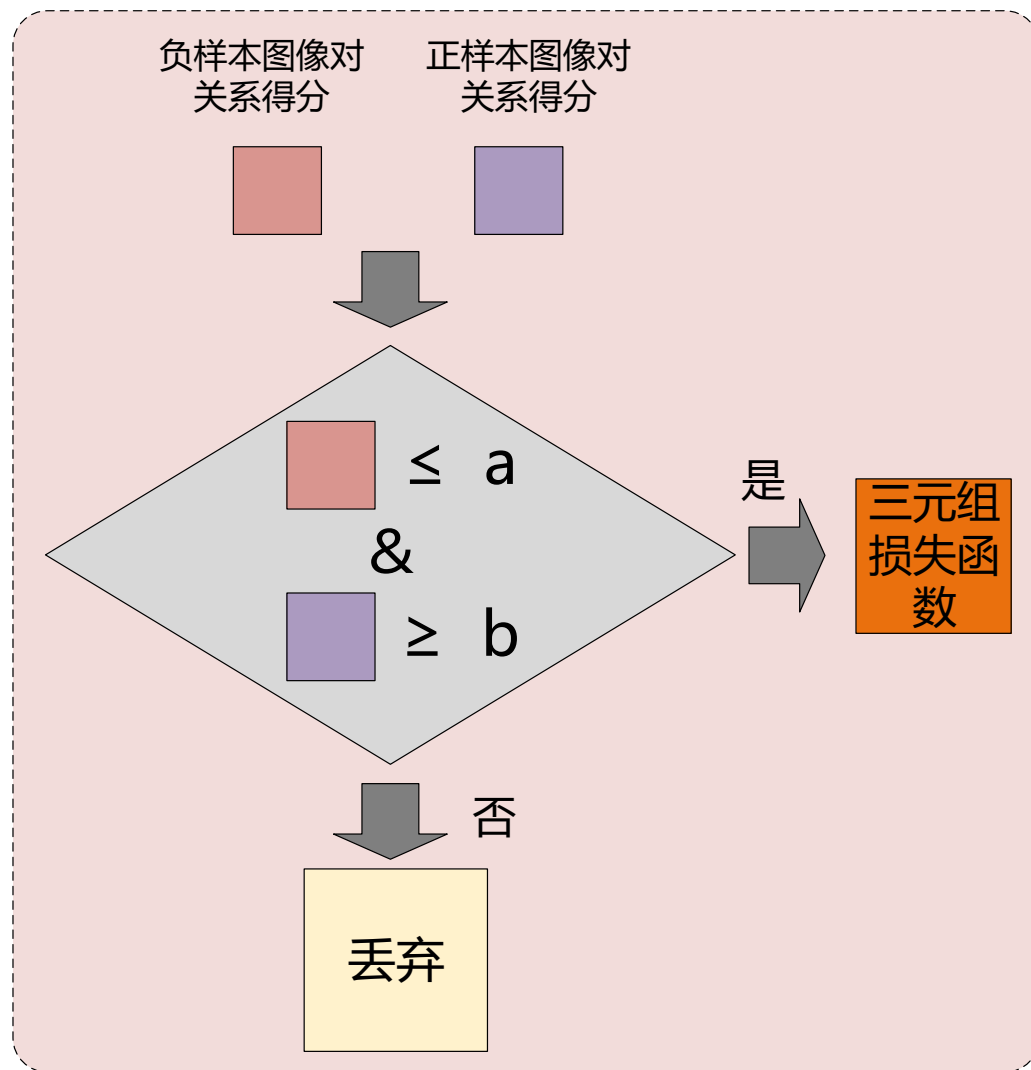
1. 使用VGG16深度神经网络提取图像特征
2. 采用Conv5层而不是全连接层作为特征

## 关系学习网络

1. 进一步融合support set图像与query set 图像的负样本图像对和正样本图相对的特征图
2. 通过关系网络分别计算关系得分



1. 先前的三元组筛选方案 “batch hard” 是指在一个batch中，寻找对于每个样本最难训练的三元组
2. 在我们设计的网络中，基于关系网络的三元组经过 “batch hard” 筛选后仍然存在很多过于难训练的三元组。
3. 为此我们提出了基于关系网络的一个新的三元组筛选方案。在一个batch中，寻找每个样本中最难训练的三元组之前，我们只挑选positive图像对的关系值得分不低于某一阈值  $b$ ，negative图像对的关系值不大于某一阈值  $a$  的三元组。





# 损失函数

$$L_{triplet}(f_{\theta}(x^{-}), f_{\theta}(x), f_{\theta}(x^{+})) \\ = \max \left\{ \left( m + g_{\varphi} \left( \tau(f_{\theta}(x), f_{\theta}(x^{-})) \right) \right) - g_{\varphi} \left( \tau(f_{\theta}(x), f_{\theta}(x^{+})) \right) \right), 0 \right\}$$

$m$ : 间隔值(margin)

$\theta, \varphi$ : 模型参数



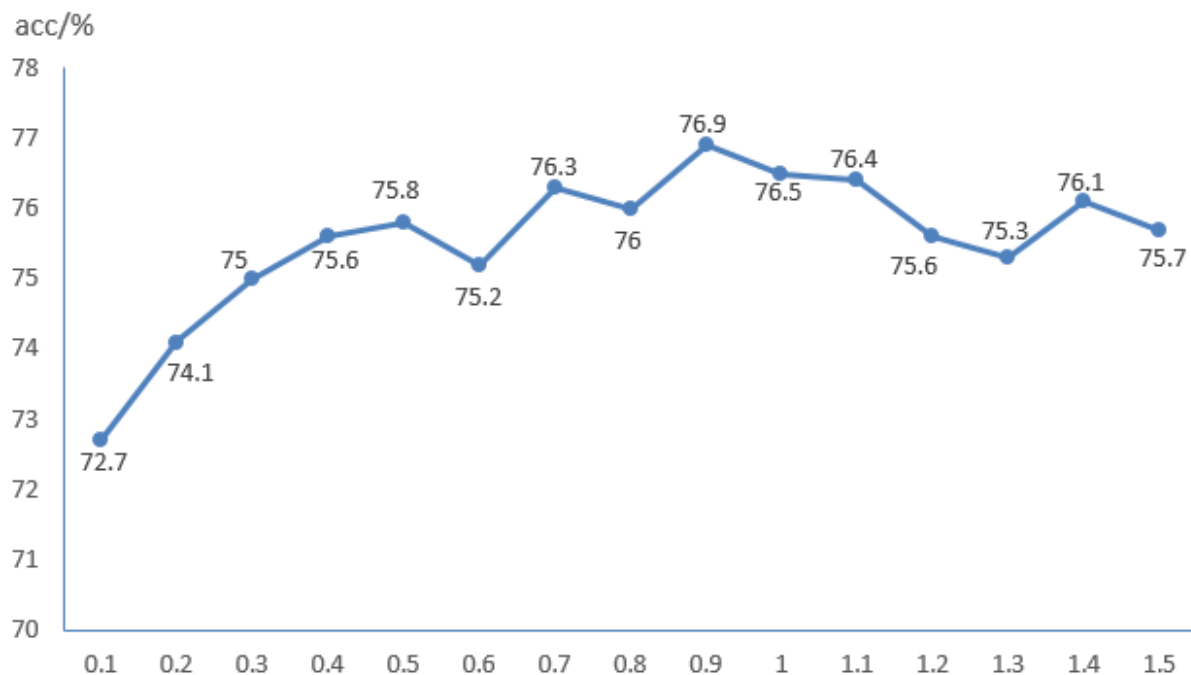
# 实验对比

| Model                     | Food-101     | VIREO Food-172 | ChineseFoodNet |
|---------------------------|--------------|----------------|----------------|
| Siamese network           | 49.1%        | 60.3%          | 50.0%          |
| Matching network          | 45.6%        | 73.6%          | 48.9%          |
| Relation network          | 53.9%        | 74.0%          | 63.8%          |
| Triplet network-Euclidean | 56.1%        | 75.2%          | 61.4%          |
| <b>Our method</b>         | <b>56.9%</b> | <b>76.9%</b>   | <b>66.1%</b>   |



# 实验对比

## 不同margin对比实验结果



## 不同三元组限制下对比实验结果

| Model      | 64 Filters |
|------------|------------|
| Batch hard | 56.1%      |
| (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%      |



# 总结

1

- 第一次提出了小样本食品图像识别的问题

2

- 提出了一种多角度表示学习的方法通过融入原料属性信息改进小样本食品图像识别

3

- 提出了一种三联体关系网络通过学习非线性度量改进小样本食品图像识别的性能

4

- 全面的实验结果验证了方法的有效性



# 展望

1

- 将离线的多角度表示学习改成端到端的训练

2

- 在三联体关系网络中引入原料属性监督的改进三联体非线性关系网络学习

3

- 在三联体关系网络中引入STN模块检测食品区域，改进基于食品图像区域的三联体非线性关系网络学习



# Thanks

- **Weiqing Min**, Shuqiang Jiang, Shuhui Wang, Jitao Sang, Shuhuan Mei: A Delicious Recipe Analysis Framework for Exploring Multi-Modal Recipes with Various Attributes. **ACM Multimedia 2017**: 402-410
- **Weiqing Min**, Shuqiang Jiang, Jitao Sang, Huayang Wang, Xinda Liu, Luis Herranz: Being a Supercook: Joint Food Attributes and Multimodal Content Modeling for Recipe Retrieval and Exploration. **IEEE Trans. Multimedia** 19(5): 1100-1113 (**2017**)
- **Weiqing Min**, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, Shuqiang Jiang: You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis. **IEEE Trans. Multimedia** 20(4): 950-964 (**2018**)
- Shuqiang Jiang ,**Weiqing Min**, Yongqiang Lv, Linhu Liu. Few-Shot Food Recognition via Multi-view Representation Learning. **submit to IEEE Trans. Multimedia 2018**
- **Weiqing Min**, Shuqiang Jiang , Linhu Liu, Rong Rui, Ramesh Jain A Survey on Food Computing. **submit to ACM Computing Survey (ACM CSUR) 2018**

<http://isia.ict.ac.cn/dataset>