

文章编号: 2095-2163(2021)03-0009-07

中图分类号: TP183

文献标志码: A

# 基于自我监督预处理的食物图像分类

姚伟盛<sup>1</sup>, 沈宇帆<sup>1</sup>, 彭玉波<sup>2</sup>, 沈 炜<sup>1</sup>

(1 浙江理工大学 信息学院, 杭州 310018; 2 浙江省电子商务促进中心, 杭州 310004)

**摘要:** 随着社交网络的快速发展,人们通常会上传、分享和记录食物图片,因此食物图像分类的应用价值也越来越大,对食品推荐、营养搭配、烹饪文化等方面都产生了积极的影响。尽管食物图像分类有着巨大的应用潜力,但从图像中识别食物仍然是一项具有挑战性的任务。为了解决食物的细粒度识别问题,本文提出了一种基于自我监督预处理的食物图像分类模型,通过自我监督的学习方式更高程度地学习食物图像特征。该模型在基于密集连接网络的食物图像分类模型 DenseFood 基础上搭建,采用上下文恢复的自我监督策略,将训练好的网络权重用于初始化 DenseFood 模型,训练微调完成分类任务。上下文恢复的自我监督策略和密集连接网络都是专注于图像特征的提取,同时结合两者,充分学习食物图像特征,来达到更好的食物图像分类精确度。为了进行性能比较,使用 VIREO-172 数据集对基于自我监督预处理的食物图像分类模型、未预处理的食物图像分类模型 DenseFood 以及基于 ImageNet 数据集训练预处理的 DenseNet、ResNet 这四个模型进行训练。实验结果表明,本文提出的食物图像分类模型优于其他策略。

**关键词:** 图像分类; 自监督学习; 卷积神经网络

## Food image classification based on self supervised preprocessing

YAO Weisheng<sup>1</sup>, SHEN Yufan<sup>1</sup>, PENG Yubo<sup>2</sup>, SHEN Wei<sup>1</sup>

(1 School of Informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China;

2 Zhejiang Province E-commerce Promotion Center, Hangzhou 310004, China)

**[Abstract]** With the rapid development of social networks, people usually upload, share and record food images, so the application value of food image classification is also increasing, which has a positive impact on food recommendation, nutrition collocation, cooking culture and so on. Although food image classification has great application potential, it is still a challenging task to recognize food from images. In order to solve the problem of fine-grained food recognition, this paper proposes a food image classification model based on self supervised preprocessing, which can learn food image features to a higher degree through self supervised learning. The model is based on DenseFood, a food image classification model based on dense connected network. The self-monitoring strategy of context recovery is adopted. The trained network weight is used to initialize DenseFood model, and fine-tuned trained to complete the classification task. The self-monitoring strategy of context recovery and dense connection convolution network are both focused on the extraction of image features. The research combines them to fully learn the food image features to achieve better classification accuracy of food image. In order to compare the performance, VIREO-172 data set is used to train four food image classification models; self supervised preprocessing based food image classification model, non preprocessed food image classification model densefood, and ImageNet data set based training preprocessing DenseNet and ResNet. The experimental results show that the proposed food image classification model is superior to other strategies.

**[Key words]** image classification; self supervised learning; convolution neural network

## 0 引言

食物是人类生活的必需品,关系到人民群众的身体健康和生命安全。随着社交网络的快速发展,人们通常会记录、上传和分享食物图片,因此食物图像分类的应用价值也越来越大,对营养搭配、食品推荐、餐饮、社交等方面都产生了积极的影响,受到了广泛的关注。在营养搭配方面,营养习惯被认为是导致糖尿病和肥胖症等健康问题的主要原因。因此,食物摄入评估是肥胖管理的一个重要方法<sup>[1]</sup>,可以帮助人们了解和保持良好的饮食习惯,在跟踪

卡路里消耗量的日常健身中进行使用<sup>[2]</sup>。在食品推荐方面,可以帮助社交媒体平台为餐馆和饮料公司针对其用户提供食品推荐方案。在餐饮方面,通过客户提供的图像样本帮助餐厅和食堂识别食物,并通过识别托盘中的食物帮助出纳自动计费。使用自动计费可以在保证减少错误的前提下提高工作效率<sup>[3]</sup>。在社交方面,可以通过社交媒体与朋友分享食物照片,根据其食物偏好对使用者进行聚类。此外,还可以帮助人们使用图像搜索食物。

尽管食物图像分类有着巨大的应用潜力,但从图像中识别食物仍然是一项具有挑战性的任务,挑

**作者简介:** 姚伟盛(1995-),男,硕士研究生,主要研究方向:机器学习;沈宇帆(2000-),男,本科生,主要研究方向:云计算、区块链。

**收稿日期:** 2021-03-24

战来自 3 个方面:

(1) 缺乏用于食物识别的大规模数据集。现有的工作主要集中在利用较小的数据集进行食物识别,如 ETH-food-101<sup>[4]</sup> 和 Vireo-food-172<sup>[5]</sup>。例如,Bossard 等人<sup>[4]</sup>发布了一个来自西餐的食物数据集 ETH food-101,包含 101 个食物类别和 101 000 张图片。Chen 等人<sup>[5]</sup>介绍了来自 172 个中国食品类别的 Vireo Food-172 数据集。这些数据集缺乏食品类别的多样性和覆盖面,没有包括广泛的食品图像。因此,可能并不足以构建更复杂的食物识别深度学习模型。

(2) 不同种类的食物可能从外观上看极其相似,但是类间相似性非常高,如图 1 所示。由图 1 可知,麻婆豆腐和红烧豆腐从外观上看基本上无法分辨。虽然已经有许多方法用于解决食物识别问题,但这些方法大多侧重于提取特定类型或某些类型的特征,而忽略了其他方面。例如,研究<sup>[6]</sup>的工作主要是提取颜色特征,而 Martinel 等人<sup>[7]</sup>设计了一个用于食物识别的网络来捕获特定的垂直结构。



图 1 不同种类的食物外观极其相似

Fig. 1 The appearances of different kinds of food are very similar

(3) 一种特定的食物可能有数千种不同的外观,但其本质上是相同的食物,如图 2 所示。由图 2 可知,烤鱼以不同形式的外观表现出来。因为相同食物的配方可以根据位置、食材成分以及最后但并非最不重要的个人口味而有所不同。食物图像中有细微的辨别细节,在很多情况下很难捕捉到。食物识别属于细粒度识别,其中有很多细微差别的细节,现有的图像分类技术还无法很好地获取并区分图像细节特征<sup>[8]</sup>。

由于有监督学习需要大量的手动数据注释,这一要求耗时耗力,因此无监督学习越来越受到了关注,尤其是在自我监督学习方面。自我监督学习是一种特殊的无监督学习,其目标是监督特征学习,其中监督任务是从数据本身生成的。模型必须充分学习图像特征,才能有效完成这类监督任务,所以基于自我监督的卷积神经网络预训练会产生有用的权

重,有助于后续的学习任务<sup>[9]</sup>。



图 2 同一种类的食物外观极其不同

Fig. 2 The appearances of the same kind of food are extremely different

本文主要研究食物图像的自我监督,提出了一种基于自我监督预处理的食物图像分类网络模型。文中的方法是将上下文恢复作为一项自我监督任务,上下文恢复策略训练的卷积神经网络专注于学习有用的语义特征,学习的卷积神经网络特征对后续分类任务有用<sup>[10]</sup>。同时由于基于密集连接网络的食物图像分类模型 DenseFood 在食物识别应用中表现出的优异性能<sup>[11]</sup>,研究中选择其作为后续的分类任务。建立自我监督预处理模型,训练好的网络权重初始化 DenseFood 网络,训练微调完成分类任务。通过使用 VIREO-172 数据集,对基于自我监督预处理的食物图像分类网络、无预处理的 DenseFood 网络模型以及基于 ImageNet 数据集训练预处理的 DenseNet、ResNet 这四个模型进行评估,实验结果表明,所提出的基于自我监督预处理的食物图像分类网络模型优于其他策略。

## 1 相关研究

自我监督学习的关键挑战是确定一个合适的自我监督任务,即通过数据生成模型输入输出对。Chen 等人<sup>[10]</sup>(2019)提出了一种医学图像自监督学习策略。具体来说,给定一幅图像,随机选择并交换 2 个补丁。多次重复此操作会产生一个新的图像,该图像的强度分布被保留,但其空间信息被改变,然后通过训练卷积神经网络将改变后的图像恢复到原来的版本。所提出的上下文恢复策略有 3 个优点:在该任务中训练的卷积神经网络专注于学习有用的语义特征;在该任务中学习的卷积神经网络特征对后续不同类型的任务(包括分类、定位和分割)有用;实现简单明了。

针对医学图像分析中的 3 个常见问题,即:分

类、定位和分割,对该自监督学习策略进行了评估。评估使用了不同类型的医学图像:对二维胎儿超声(US)图像进行图像分类;对腹部计算机断层扫描(CT)图像进行器官定位;对脑磁共振(MR)图像进行分割。在这三个任务中,基于上下文恢复策略的预训练都优于其他的自监督学习策略,也优于没有自监督训练的学习策略。

食物图像分类方面,卷积神经网络取得了广泛的成功,其性能优于其他方法。卷积神经网络有许多不同的架构,如 AlexNet<sup>[12]</sup>、VGG<sup>[13]</sup>、GoogleNet<sup>[14]</sup>、ResNet<sup>[15]</sup>、DenseNet<sup>[16]</sup> 等,其中 DenseNet 的性能表现优异。密集连接网络 DenseNet 在 2017 年的 ImageNet 大规模视觉识别大赛 (ILSVRC) 上表现出出众的效果。DenseNet 专注于图像特征的提取与复用,加强了图像特征的传递,一定程度上减轻了梯度消失的问题,通过对图像特征的极致利用达到了更少的参数和更好的效果。Metwalli 等人<sup>[11]</sup> (2020) 提出了基于 DenseNet 的食物图像识别模型 DenseFood,使用了 softmax 损失函数和中心损失函数相结合的方法,该模型从头开始训练就达到了 81.23% 的准确率,仅次于基于大规模 ImageNet 数据集训练的 ResNet 和 DenseNet。

基于此,本次研究中针对食物图像分类,提出了一种基于自我监督预处理的网络模型,使用基于上下文恢复的自我监督预处理方法,训练好的权重用于初始化食物分类网络模型 DenseFood,再进行训练微调完成分类任务,来达到更好的分类效果。

## 2 本文方法

研究中基于自我监督预处理的食物图像分类网络模型由 2 部分组成,分别是:基于上下文恢复的自我监督预处理模型和食物图像分类模型。对此拟展开研究论述如下。

### 2.1 基于上下文恢复的自我监督预处理模型

基于上下文恢复的自我监督预处理是将原始的图片进行打乱,再利用卷积网络将其上下文训练恢复为原始图片<sup>[10]</sup>。打乱的方式是选取图像中随机的 2 个小块进行交换,迭代多次,保证所有小块不会出现重叠,防止图片打乱过于复杂,最终将会生成一个乱序的图片。乱序算法的伪代码如图 3 所示,其中,  $x$  为原图,取小块大小为  $10 \times 10$ ,迭代次数  $N$  为 10。乱序图像生成过程如图 4 所示。通过上下文恢复的自我监督学习,卷积神经网络可以更加专注地学习食物图像的语义特征。如图 5 所示,基于上下文恢复的自我监督预处理模型由 2 部分组成,即:预处理部分和恢复部分。

```

while iter < N
  randomly select a patch p1 ∈ x
  randomly select a patch p2 ∈ x
  if set ∩ p1 == 0 and set ∩ p2 == 0 and p1 ∩ p2 == 0
    then
      swap p1 and p2
      add p1 to set
      add p2 to set
      iter ++
end
  
```

图 3 图片乱序伪代码

Fig. 3 Image disorder pseudo code

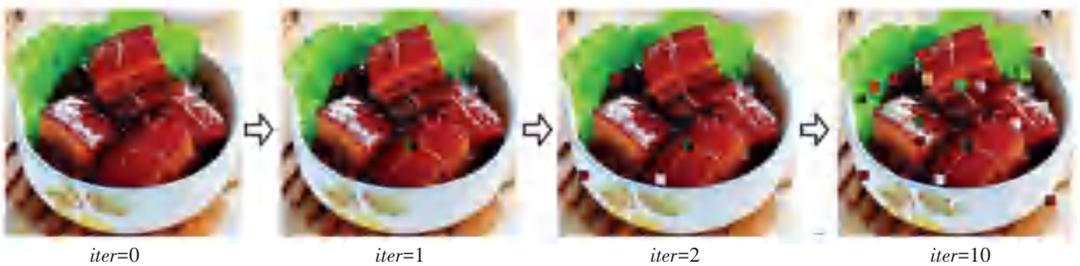


图 4 图片乱序过程

Fig. 4 Picture disorder process

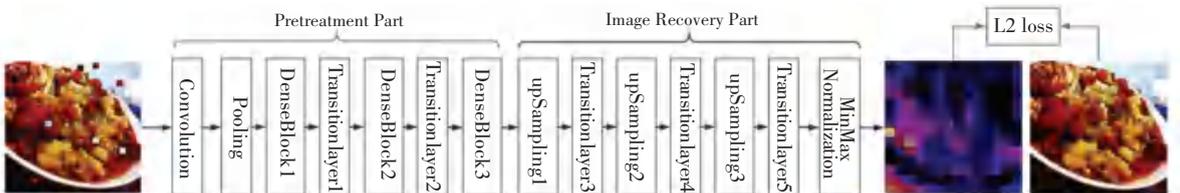


图 5 基于上下文恢复的自我监督预处理

Fig. 5 Self monitoring preprocessing based on context recovery

预处理部分主要由初始化层、密集连接块和过渡层组成,训练出的权重用于初始化后续的图片分类网络。初始化层由卷积层和汇聚层组成,在将信息输入密集块层之前从图像中提取出大量的信息,特征映射被下采样以减少参数的数量。密集连接块由批标准化、激活函数和卷积层组成。如图6所示,每一层的输出都作为输入提供给后续层,因此,第 $k$ 层接受来自先前所有层的特征映射。将 $X_0, X_1, \dots, X_{k-1}$ 视为输入,即:

$$X_k = H_k([X_0, X_1, \dots, X_{k-1}]), \quad (1)$$

其中,  $X_k$  表示第 $k$ 层的输出;  $[X_0, X_1, \dots, X_{k-1}]$  表示前面的 $k$ 层;  $H_k([\ ])$  表示批量归一化(BN)、指数线性单元(ELU)和 $3 \times 3$ 卷积(conv)操作的组合函数。批量归一化将输入值进行归一化加快训练速度,ELU解决梯度消失的问题,提高分类精度。

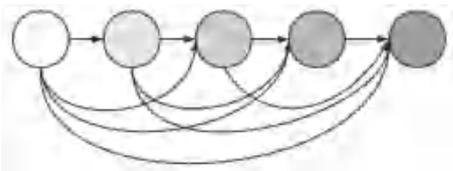


图6 密集连接块

Fig. 6 Dense connection block

为了降低特征映射的维数,将网络结构划分为4个密集连接块,密集连接块之间通过过渡层衔接,过渡层由批处理归一化、ELU、 $1 \times 1$ 卷积层和 $2 \times 2$ 最大池化层组成。池化层使用最大池、而不是平均池来减少特征映射的数量,避免过拟合,以使模型更具通用性。值得一提的是,简单模型可能无法很好地拟合数据,因此可能会出现欠拟合,而层数较多的模型可能会产生较高的计算成本,通常需要更大的数据集来避免过度拟合,提高精度,并实现可推广的性能<sup>[17]</sup>。考虑到本次研究的数据集不够大,无法训练一个复杂的模型,因此分别构造了4个6层、12层、24层和16层的密集连接块,以避免过度拟合和欠拟合。由于密集连接网络较深,导致图像恢复效果

较差,研究中就选取了前三个密集连接块作为预处理部分,既加强了图像特征学习,又保证了图像恢复效果。恢复部分主要由 upSampling 上采样层、过渡层和 Min-Max 标准化层组成,输出为图片。由于恢复部分不会应用于后续的分类工作,所以相对预处理部分较为简单,upSampling 上采样层采用重采样和插值方法,过渡层由批处理归一化、ELU、 $1 \times 1$ 卷积层组成,Min-Max 标准化层将输出进行标准化,Min-Max 标准化公式见如下:

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}} \times N. \quad (2)$$

其中,  $x_1, x_2, \dots, x_n$  为输入序列,对其进行变换得到标准化序列  $y_1, y_2, \dots, y_n$ 。由于该模型训练的数据标签是图片,会占用大量内存,导致可训练的数据集容量较小,因而将网络输出的图片大小以及做比对的原始图片都缩小为原来的 $1/4$ (宽和高各缩小为原来的 $1/2$ ),来保证可以训练更大的数据集,提高训练效率及效果。同时由于恢复部分的网络权重并不会应用到后续的分类工作中,因此缩小输出图片所带来的像素损失对分类结果的影响较小。

## 2.2 食物图像分类模型

将上下文恢复预处理模型中的预处理部分的网络权重对食物图像分类网络进行初始化,再进一步训练微调。如图7所示,主要由4部分组成,即初始层、密集连接块层、过渡层和完全连接层。前面介绍过,为了避免过度拟合和欠拟合,研究中构造了4个6层、12层、24层和16层的密集连接块,初始层、前三个密集连接块层、过渡层组成了预处理部分,由预处理模型权重进行初始化。最后一部分包括一个密集连接块层、过渡层和2个完全连接层,第一个完全连接层使用全局平均池将特征映射展平成一个包含1024个节点的数组,然后将其作为分类器输入到第二个完全连接层,该层包含172个神经元,每个神经元代表一个食物类。

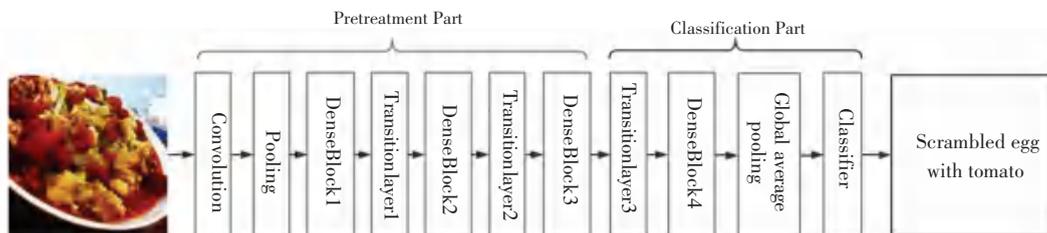


图7 基于密集连接卷积网络的食物图像分类网络

Fig. 7 Food image classification network based on dense connected convolutional network

### 2.3 损失函数

对于自我监督预处理模型,文中采用  $L_2$  损失训练网络来完成上下文恢复任务,  $L_2$  损失可以预估出图片的恢复程度:

$$L_2 = \sum_{i=1}^n (y_i - x_i)^2 \quad (3)$$

其中,  $x_i$  和  $y_i$  分别表示恢复图和原图的像素值;  $L_2$  损失函数又称为最小平方误差,把目标值和估计值的平方和最小化。尽管上下文恢复的输出可能是模糊的,但是  $L_2$  损失对于特征学习来说是足够的了。

对于分类网络模型,考虑到食物图像具有类间相似性和类内变化,采用 softmax 损失和中心损失相结合,并使用  $\lambda$  来平衡 2 个损失函数<sup>[18-19]</sup>,如下所示:

$$\text{loss} = l_{\text{softmax}} + \lambda \cdot l_{\text{center}}, \quad (4)$$

softmax 损失可以最大化类间差异<sup>[19]</sup>:

$$l_{\text{softmax}} = -\frac{1}{M} \sum_{i=1}^M \log \frac{e^{w_j f_i}}{\sum_{j=1}^c e^{w_j f_i}}. \quad (5)$$



图8 图像预处理

Fig. 8 Image preprocessing

## 3 实验与结果

### 3.1 数据集

VIREO-FOOD 数据集是一个大型的公共中餐食品数据库,包含 172 类的 110 241 张图片,如图 9 所示,用于训练卷积神经网络有不错的效果。文中



图9 VIREO-172 数据图片

Fig. 9 VIREO-172 data picture

### 3.2 实验平台

文中使用 TensorFlow 实现了研究中的分类网络模型, TensorFlow 是由 Google 开发的一个端到端开放源代码机器学习平台,具有灵活和全面的工具、库

其中,  $M$  表示训练样本数;  $c$  表示类数;  $f_i$  表示第  $i$  个样本的特征;  $y_i$  表示实类别标签;  $\mathbf{W}$  表示倒数第二层 (softmax 损失之前的层) 的权重矩阵;  $w_j$  表示  $\mathbf{W}$  的第  $j$  列, 对应于第  $j$  个类。中心丢失可以通过减少特征与其对应中心之间的距离来最小化类内变化<sup>[19]</sup>:

$$l_{\text{center}} = \frac{1}{2} \sum_{i=1}^M \|x_i - c_{y_i}\|_2^2. \quad (6)$$

其中,  $M$  表示训练样本的数目;  $C_{y_i}$  表示  $y_i$  类的特征中心;  $x_i$  表示第  $i$  类的归一化特征。

### 2.4 图像预处理

研究中使用的数据集的大小有限,平均每个类 641 张图片的 172 个类只包含 110 241 个图像。为了解决这一问题,防止训练中的过度拟合,通过使用水平翻转、旋转、错切变换、缩放和平移等方法来增加数据,如图 8 所示。图像的大小总是调整为  $224 \times 224$ , 以适应模型。

将数据集分为 2 个子集,即: 80% 用于分类网络训练, 20% 用于分类网络测试。同时用于训练的子集部分中, 80% 作为训练集, 20% 作为校验集。分类网络的训练集的 40% 用于自我监督预处理的训练, 校验集的 30% 用于自我监督预处理的校验。

和资源生态系统<sup>[20]</sup>。训练过程是在 Intel Core™ i7 2.8 GHz CPU、32 GB RAM 和一个 6 GB 的 Nvidia GeForce GTX 1060 GPU。训练以平均 52.48 张图片/s 的速度进行。

### 3.3 卷积神经网络训练

自我监督预处理模型的训练初始学习率设置为 0.01, 分类网络模型是在预训练基础上训练的, 为了避免权重失真, 初始学习率改为 0.005。此外, 在学习速率表中使用余弦衰减来降低学习速率。同时使用中心损失函数和 *softmax* 分类交叉熵函数相结合进行分类, 其中  $\lambda$  的值设为 0.5, 以平衡损失函数, 因为大多数食品类别的类内变化较小。为了避免过度拟合和提高精度, 在训练期间使用了数据扩充, 使用随机的水平翻转、旋转、错切变换、缩放和平移。由于计算资源有限, 批量大小设为 16, 训练时长设为 40 个周期。

### 3.4 实验结果

研究中, 在 VIREO-172 数据集上评估了上下文

恢复模型以及分类网络模型, 上下文恢复训练结果如图 10 所示, 还原图较为模糊, 但对于预训练分类网络已经足够了。紧接着, 训练了无预处理的食物图像分类网络 DenseFood、基于 ImageNet 数据集预处理的 DenseNet121、ResNet50 和基于自我监督预处理的食物图像分类网络四个模型, 提供 Top-1 和 Top-5 精度, 实验结果见表 1。由表 1 可以看出, 基于自我监督预处理的食物图像分类网络的 Top-1 和 Top-5 精度分别为 84.25% 和 96.97%, 准确率高于其他模型, 验证了本文的网络模型具有更有效的食物图像特征学习。在此基础上, 还绘制了这 4 个模型的损失曲线和精确度曲线, 如图 11、图 12 所示, 文中的模型对损失值做到了更好的最小化, 其损失值为 0.69。

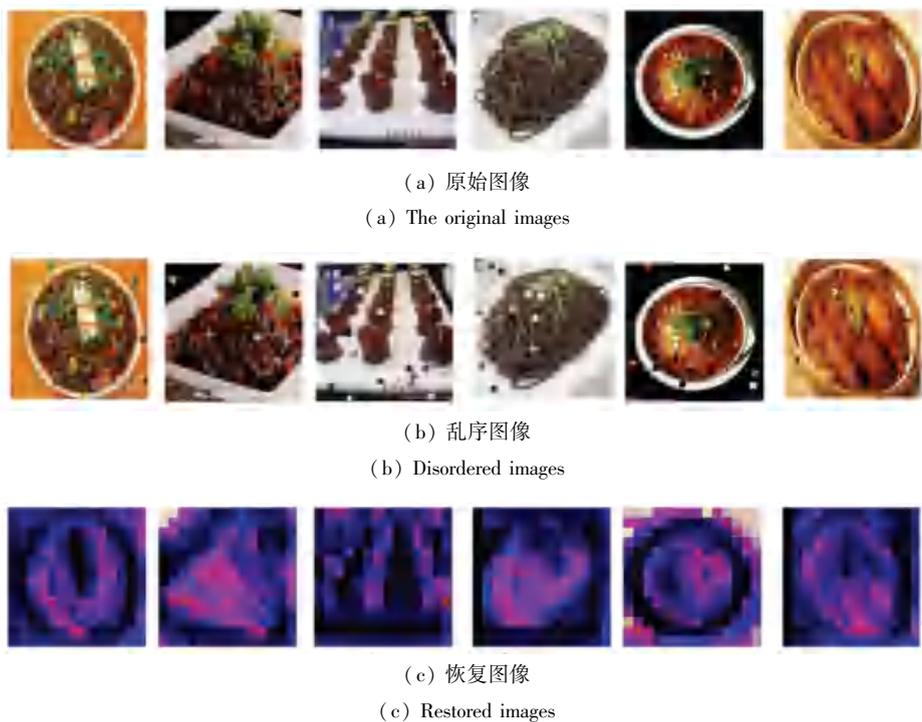


图 10 上下文恢复训练结果图

Fig. 10 The result of context recovery training

表 1 Top-1 和 Top-5 精确度  
Tab. 1 Top-1 and Top-5 accuracy %

模型	Top-1 Acc	Top-5 Acc
DenseFood	81.22	95.42
DenseNet121	83.89	96.38
ResNet50	82.52	95.83
基于自我监督预处理的食物图像分类网络	84.25	96.97

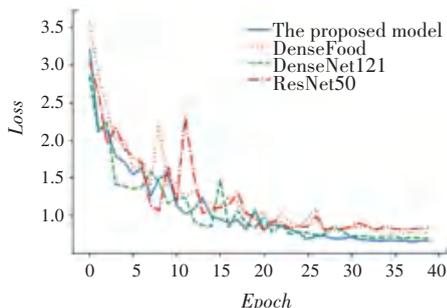


图 11 损失值曲线图

Fig. 11 Loss value curve

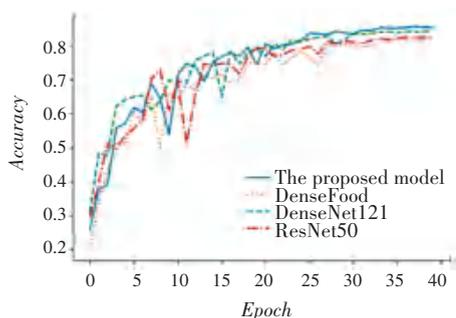


图 12 精确度曲线图

Fig. 12 Accuracy curve

## 4 结束语

本文提出了一种基于自我监督预处理的食物图像分类网络。研究中构建模型训练食物图片的上下文恢复能力来学习图像特征,将该模型训练出的网络权重在分类网络中进行初始化,再进一步训练微调,使用密集连接卷积网络进一步提取和复用图像特征,充分实现对食物图片细节上的图像特征的学习和提取,来提高食物图像分类的精确度。实验结果也验证了本文的论点,文中研发的模型 Top-1 和 Top-5 精确度高达 84.25% 和 96.97%, 优于其他模型。

## 参考文献

- [1] HE Hongsheng, KONG Fanyu, TAN Jindong. DietCam: Multiview food recognition using a multikernel SVM [J]. IEEE Journal of Biomedical and Health Informatics, 2015, 20(3):848-855.
- [2] 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.
- [3] AGUILAR E, REMESEIRO B, BOLAÑOS M, et al. Grab, Pay and eat: Semantic food detection for smart restaurants [J]. IEEE Transactions on Multimedia, 2018, 20(12):3266-3275.
- [4] BOSSARD L, GUILLAUMIN M, GOOL L V. Food-101-Mining discriminative components with Random Forests [C]//European Conference on Computer Vision. Zurich: Springer, 2014:446-461.
- [5] CHEN Jingjing, NGO C W. Deep-based ingredient recognition for cooking recipe retrieval [C]//Proceedings of the 24<sup>th</sup> ACM on International Conference on Multimedia (MM'16). New York, NY, United States: ACM, 2016: 32-41.
- [6] BETTADAPURA V, THOMAZ E, PARNAMI A, et al. Leveraging context to support automated food recognition in restaurants [C]// IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE, 2015:580-587.
- [7] MARTINEL N, FORESTI G L, MICHELONI C. Wide-slice residual networks for food recognition [C]// IEEE Winter Conference on Applications of Computer Vision. Lake Placid, NY, USA: IEEE Computer Society, 2016: 567-576.

- [8] MIN Weiqing, LIU Linhu, WANG Zhiling, et al. ISIA Food-500: A dataset for large-scale food recognition via stacked global-local attention network [C]// Proceedings of the 28<sup>th</sup> ACM International Conference on Multimedia (MM '20). 2020: 393-401.
- [9] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations [J]. arXiv preprint arXiv:1803.07728, 2018.
- [10] CHEN L, BENTLEY P, MORI K, et al. Self-supervised learning for medical image analysis using image context restoration [J]. Medical Image Analysis, 2019, 58(11):101539.
- [11] METWALLI A S, SHEN W, WU C Q. Food image recognition based on densely Connected Convolutional Neural Networks [C]// 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC). Fukuoka, Japan: IEEE, 2020:27-32.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with Deep Convolutional Neural Networks [C]// Neural Information Processing Systems (NIPS). USA: Morgan Kaufmann Publishers, Inc., United States of America, 2012, 141: 1097-1105.
- [13] SIMONYAN K, ZIEEERMAN A. Very Deep Convolutional Networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [14] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inceptionresnet and the impact of residual connections on learning [C]// AAAI Conference on Artificial Intelligence (AAAI). San Francisco, California, USA: AAAI, 2017:4278-4284.
- [15] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions [C]//Proceedings of The IEEE Conference On Computer Vision and Pattern Recognition, Boston, MA: IEEE, 2015:1-9.
- [16] HUANG G, LIU Z, LAURENS V D M, et al. Densely Connected Convolutional Networks [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). Washington, DC: IEEE Computer Society, 2017: 2261-2269.
- [17] KABKAB M, HANDS E, CHELLAPPA R. On the size of Convolutional Neural Networks and generalization performance [C]// 2016 23<sup>rd</sup> International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE, 2016:3572-3577.
- [18] WANG F, XIANG X, CHENG J, et al. NormFace: L2 hypersphere embedding for face verification [C]//Proceedings of the 2017 ACM on Multimedia Conference. Mountain View, CA, USA: ACM, 2017: 1041-1049.
- [19] ZHANG Tong, WANG Rong, DING Jianwei, et al. Face recognition based on densely Connected Convolutional Networks [C]// IEEE Fourth International Conference on Multimedia Big Data (BigMM). Xi'an, China: IEEE Computer Society, 2018: 1-6.
- [20] Google. Tensorflow [EB/OL]. [2019]. <http://tensorflow.google.cn>.