CNN 食物图像分类的优化

杨思涵

软件 2204 - 2226114139

摘 要 在这篇论文中，我们专注于优化 CNN 模型在食物图像分类上的表现。通过一个系列的实验，我们首先使用最基本的 CNN 模型，然而准确率仅为 0.255，显示出过拟合的问题。我们随后依次增加举措，尝试了图像增广与学习率调度器，半监督学习以及 ResNet50 模型，举措的改变使得验证集上的准确率逐步提高，达到了 0.568，0.597 和 0.700。从这些结果中，我们得出结论：通过从数据、模型和训练策略三个方面进行优化，我们能有效地提高食物图像分类的准确性。这三个优化方向同样对于其他任务有所借鉴意义。

关键词 卷积神经网络，食物图像分类，过拟合，图像增广，学习率调度器，半监督学习，ResNet50

Optimization of Food Image Classification Based on CNN Model

SIHAN YANG[1)]

[1)](Department of Software, Xi'an Jiao Tong University, Xi'an, Shannxi)

Abstract In this study, our focus lies in optimizing the performance of CNN models for food image classification. Through a series of experiments, we initially employed a basic CNN model which exhibited overfitting with an accuracy of only 0.255. Subsequently, we incorporated various techniques including image augmentation, learning rate scheduler, semi-supervised learning, and ResNet50 model to address this issue. As a result, the accuracy on the test set gradually improved to 0.568, 0.597, and eventually reached 0.700 respectively. Based on these findings, we conclude that by optimizing data quality, refining the model architecture and employing effective training strategies; significant improvements can be achieved in food image classification accuracy. These three optimization directions also hold relevance for other related tasks.

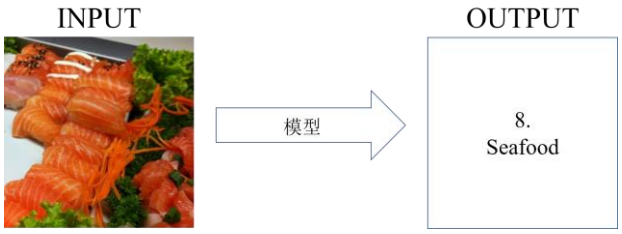
Key words Convolutional neural network, Food image classification, Overfitting, Image augmentation, Learning rate scheduler, Semi-supervised learning, ResNet50

1. 引言

在随着深度学习和计算机视觉的快速发展，图像分类正逐渐成为人工智能领域的一个关键问题。尤其是食物图像分类，由于其巨大的实用应用价值和研究挑战，引起了广大研究者的关注。然而，已有的分类模型往往会在处理这类任务时遇到各种挑战，如过拟合问题，以及训练数据和测试集之间的巨大差异等。

为了解决这些问题并优化食物图像分类的准确性，我进行了一系列关于数据，模型和训练策略三方面的优化实验。我使用来自 Kaggle 的数据集作为研究对象，数据集包括 11 个类别，分别是：Bread, Dairy product, Dessert, Egg, Fried food, Meat, Noodles-Pasta, Rice, Seafood, Soup, Vegetable-Fruit。

提交作业日期：2023年12月30日；



从简单的 CNN 模型开始，我逐步尝试了图像增广与学习率调度器，半监督学习，以及使用 ResNet50 模型，从而对模型进行改进和优化，使得测试集上的准确率得到显著提高。这不仅提供了一种有效的食物图像分类解决方案，也为其他图像分类任务提供了可借鉴的方法。

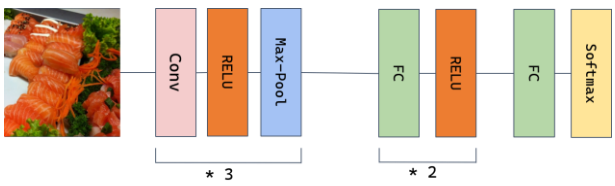
我希望这篇论文能够为食物图像分类的模型优化提供帮助和新的方向，同时，对于图像分类的广泛性，我们也寄望于这些方法能够对处理其他类别的图像分类问题产生借鉴和启示。

代码仓库：
https://github.com/Hhankyangg/Classification_Food

2. 训练过程

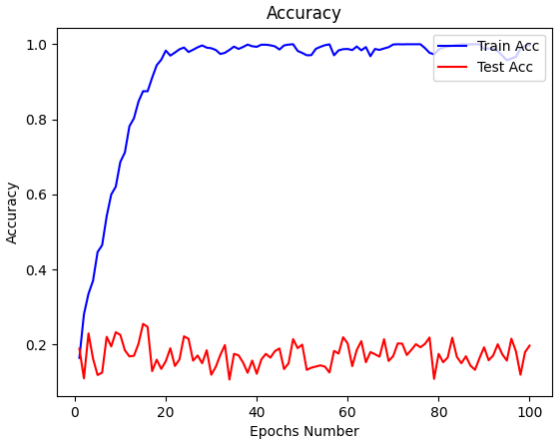
1.1 简单的 CNN 模型训练

在最初的训练架构中，我使用了简单的 CNN 模型训练。模型如下：



1. **输入层**：这是模型的开始，它接收一张食物图片作为输入。
2. **卷积块**：
   在卷积部分，我使用了三个相同的卷积块，每个卷积块有以下组成：
   1. **卷积层（Convolutional Layers）**：这是 CNN 的核心部分。卷积层会对输入图片进行卷积操作，提取图片中的局部特征。每一层卷积层都会从上一层中学习并提取特定的特征。
   2. **池化层（Pooling Layers）**：池化层位于卷积层之后，用于降低数据的维度，减少计算量并提高模型的泛化能力。我是用的池化操作是最大池化。
3. **全连接层（Fully Connected Layers）**：在提取了足够的特征之后，全连接层会对这些特征进行分类。全连接层的输出是每个类别的概率分布。我使用了三个相邻的全连接层。
4. **输出层**：输出层是一个 softmax 层，它将全连接层的输出转换为概率分布，表示图片属于每个类别的可能性。
5. **交叉熵**：交叉熵损失函数是多项逻辑回归中常用的一种损失函数，主要用于评估两个概率分布之间的差异。在深度学习中，一般使用交叉熵来度量模型预测的概率分布与实际情况的差距，进而通过优化器将这个差距最小化。交叉熵的最小值是 0，对应于模型预测的概率分布和实际概率分布完全相同的情况。交叉熵损失函数的公式为：$H(p,q) = -\Sigma\, p(x) \log q(x)$，其中 p(x)表示真实的概率分布，q(x)表示模型预测的概率分布。可以看到，如果模型在某一类别上的预测概率与真实概率越接近，则对应的损失函数会越小，说明模型的预测性能越好。
6. **Adam 优化器**：一种自适应学习率的优化算法，它通过计算梯度的一阶矩估计和二阶矩估计来调节学习率，相比于一般的随机梯度下降法（SGD），在处理大规模和复杂的机器学习问题时，Adam 可以提供更快的收敛速度并且更稳定。

然而，这种简单的训练方式在测试集的结果并不理想：



我们可以看到，随着 epochs 的增加，训练数据的损失曲线降低的很快。然而模型在测试数据的表现并不好，甚至差异过大，可以认为模型自始至终都没有学会怎么“分类”食物！这种现象被称为过拟合，即当我们的模型训练过度，以至于开始记忆训练数据，而非从中学习模式时，就会出现过拟合。具体来说，过拟合的模型在训练数据上表现非常好，但在未见过的验证和测试数据上表现却很差。这是因为模型已经过度适应了训练数据，捕捉到了训练数据中的噪音和异常值，而非更一般的数据分布。

1.2 图像增广与学习率调度器

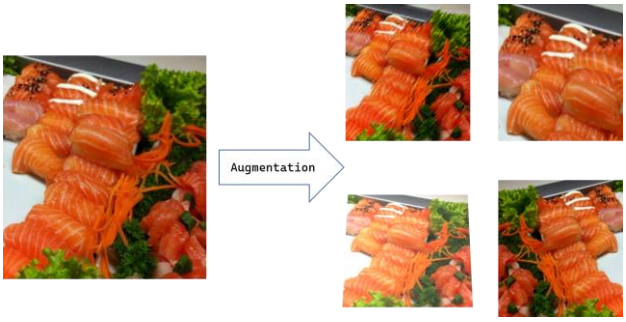
1.2.1 图像增广

为什么在我们的数据上会出现过拟合？一个原因是食物分类的特殊性。食物分类是一个较为广义的分类行为，比如西瓜，桃子，黄瓜，草莓都属于 Vegetable-Fruit 类。不像各式各样黄瓜，他们的相似性更强，食物分类中食物同类之间的相似度更低，由此，模型训练很容易过度适应了训练数据，捕捉到了训练数据中的噪音和异常值，而非更一般的数据分布。

为了解决这一问题，我首先采用了图像增广的方法。

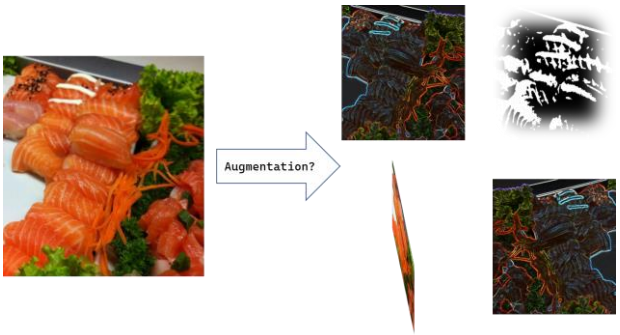
图像增广是一种在原始图像数据基础上引入随机性变化，生成新的训练样本，以增加训练集的多样性和数据量的常用技术。这可以帮助模型更好地泛化，提高对新数据的预测准确度。

以三文鱼的图片为例：



1. 不同角度：将图片旋转一定的角度。例如，你可能会拍摄到三文鱼的侧面或是从顶部看三文鱼。这种角度变化可以帮助模型理解不同视角下的三文鱼形态，并学习到这些特征与“三文鱼”这个类别的关联。
2. 不同部位：随机裁剪图片的一部分。以三文鱼为例，你可能会有一张只显示鱼尾部位的图片，又或者是只显示鱼头的图片。这种方法的好处在于，模型将会学习到一部分物体也可以表示为整个物体的特性，这有助于在实际应用中进行更准确的物体检测和分类。
3. 其他变化：还可以添加翻转（水平或垂直）、缩放、对比度、亮度调整等操作。例如，随机调整图片亮度可以模拟在不同光照条件下的三文鱼图片，让模型优化对这些条件下的识别能力。

但也要注意，图像增广不能失去图像本有的属于这一类别的信息，例如对图像做如下奇怪的变换：

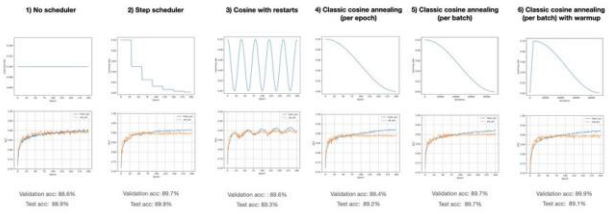


2.2.2 学习率调度器

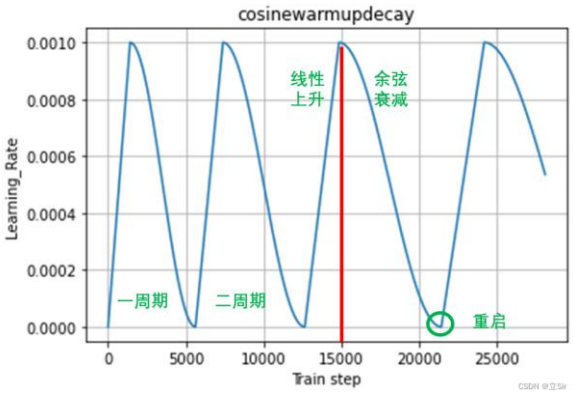
学习率调度器是调整模型训练过程中的学习率的工具。在模型的训练过程中，初始阶段通常使用较高的学习率以快速达到一个较好的优化区域，随后逐渐降低学习率，以使模型在找到的区域中进行更精细的优化。

虽然学习率调度器主要用于控制优化过程和加速训练，但其间接地也可以帮助抑制过拟合。具体来说，降低学习率可以使优化过程更稳定，避免模型在参数空间中跳过可能的更好的解。这种细致的搜索可以使模型不那么依赖一些特定的、对训练数据过度拟合的特征，从而提高模型的泛化性能。

因为我们无法确保模型已经找到了一个合理的优化区域，如果过早地降低学习率，可能会使模型陷入不佳的局部极小值中。因此，现代的学习率调度不局限于“先快后慢”，而有以下众多学习率变化方案：



在实验中，我最终使用了学习率余弦退火算法：



最终，这一次优化的训练经过 400 个 epochs 的训练，在验证集的最大精度达到了 0.568。可认为基本克服了过拟合的问题，但是精确度仍然不理想。

表 1 目前的实验结果

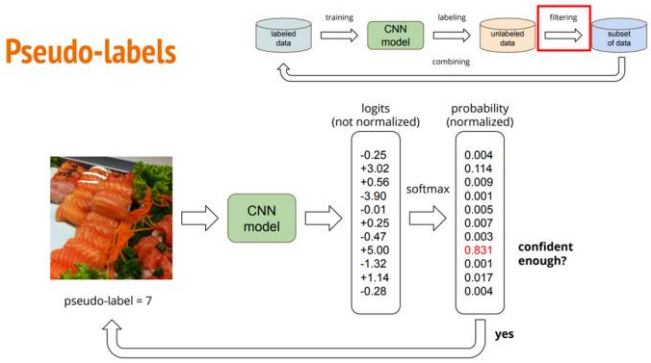
方法	测试集最大准确率
初版方案 100 epochs	0.255
数据增强 + 学习率调度器 400 epochs	0.568

1.3 半监督学习

在不改变模型架构，策略的前提下，要使模型准确率提高，我们能做的只有引入更多数据了。然而。在食物图像分类中，标记每一张图片需要专门的人工工作，这是非常耗费人力物力的。例如，你可能需要专门的标注员去识别每张图片并标明它是什么。另一方面，获取未标记的图像（例如从网页，摄像头，社交媒体等）是容易且成本低的。

半监督学习也就应运而生，而且可以很完美的应用到我们的任务之中：

半监督学习利用这大量的未标记数据来帮助模型理解数据的底层结构和分布，从而可以通过这些信息去“预测”对未标记数据的正确标签，然后利用这些“预测”标签来进行进一步的训练。这样做的结果是，即使我们只有相对较少的标记数据，也可以训练出进行准确分类的强大模型。



这是我用到的半监督学习示意图，模型会在指定的轮次时事进行预测，若“足够自信”分类正确，则会将这个图片纳入训练集中进行训练。

最终，这一次优化的训练经过 400 个 epochs 的训练，在验证集的最大精度达到了 0.597。相比于上一个版本有所进步，但精确度仍然不理想。

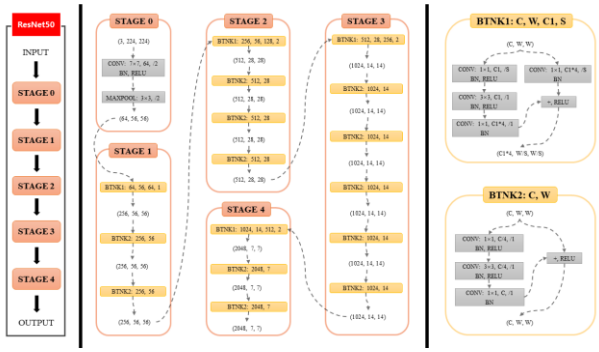
表 2 目前的实验结果

方法	测试集最大准确率
初版方案 100 epochs	0.255
数据增强 + 学习率调度器 400 epochs	0.568
方案优化 + 半监督学习 400 epochs	0.597

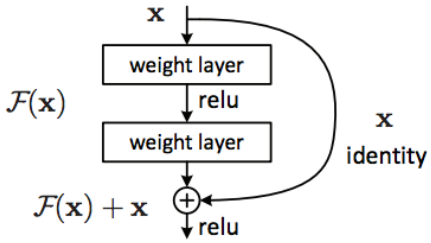
1.4 ResNet50

在数据和策略都进行了一定的优化后，我将目光转向了模型的优化。模型的简陋也会导致图像识别性能差，于是我弃用了自己手写的简陋的网络架构，使用了 ResNet50。

ResNet50 是一种深度卷积神经网络模型，它是微软推出的深度残差网络（ResNet）的一种配置。ResNet50 这个名称中的“50”代表这个网络中含有大约 50 层的卷积层，用于提取图像或其他数据中的特征。



在深度学习的背景下，传统的神经网络存在着所谓的梯度消失问题。这是由于在网络中通过多层传播梯度，会使梯度变得非常小，导致网络太深时，早期层的学习会变得非常困难。为了解决这个问题，ResNet 引入了残差块的概念。



每个残差块包含两条路径。一条是正常卷积路径，另一条是“跳过”一层或多层的跳跃连接。因此，每个残差块都会输出两部分的结果，一部分是正常卷积的结果，另一部分是短路连接的输入。然后再把这两部分结果加在一起。这么做的好处是既使网络很深，每层的输入信号也可以直接达到后面的层，从而缓解了梯度消失问题。

在 ResNet50 中使用的残差块的一种类型有三层卷积，分别是 1x1、3x3 和 1x1 的大小。这种设计有利于降低复杂性和计算成本，并且有助于提升模型的性能。

由于 ResNet50 具有从浅层到深层逐步提取特征的能力，因此它在诸如图像识别、物体检测等计算机视觉任务上，通常能够表现出较好的性能。

我使用了 ResNet50 作为新的网络架构进行训练，只训练了 200 轮，就得到了 0.700 的测试集最大精确度，效果立竿见影

表 3 目前的实验结果

方法	测试集最大准确率
初版方案 100 epochs	0.255
数据增强 + 学习率调度器 400 epochs	0.568
方案优化 + 半监督学习 400 epochs	0.597
方案优化 + ResNet50 200 epochs	0.700

3. 总结

在处理计算机视觉问题和执行图像分类任务

时，卷积神经网络作为一种强大且灵活的工具已经得到了广泛的应用。然而，优化图像分类的任务并不是一个简单的过程。特别是处理特定领域任务，如食物图像分类时，可能会遇到一些特殊挑战，例如数据不平衡、类别间边界模糊等。

在论文中，我详细介绍了如何通过在数据、模型和训练策略三个方面进行优化，改进 CNN 在食物图像分类任务上的表现。

在初始实验中，作者使用了一个基础的 CNN 模型进行分类，然而效果并不理想，出现了过拟合问题，测试集的准确率仅为 0.255。这主要是由于模型在训练集上的表现过于优秀，而在未知的测试数据上表现却相对较差。

为解决过拟合的问题，我首先尝试引入图像增广与学习率调度器。图像增广通过对原始图像进行各种变换，创造新的训练样本，增加了数据的多样性，提高了模型泛化能力。学习率调度器则通过在训练过程中动态调整学习率，保证了模型在学习过程中能够更加灵活地适应数据变化。这两种策略的使用显著提高了模型的性能，使测试集的准确率提升至 0.568。

然后，我使用了半监督学习技术。半监督学习能够有效利用标注和未标注的数据，通过自我训练和标签传播的方式，最大化地利用可用的数据，进一步提高了模型的准确率，达到了 0.597。

最后，我决定采用 ResNet50 作为基本结构进行训练。ResNet50 通过其深层结构和并行连接的方式，显著增加了模型的容量，可以更加深入地学习图像中的特征，从而提高分类精度。通过使用 ResNet50，测试集的准确率达到了 0.700。

论文提供了一种系统性的方法来优化食物图像分类任务。从**数据、模型和训练策略三个方面**展示了各种有效的优化策略，并通过实验验证了这些策略的有效性。

参 考 文 献

[1] Adam: A Method for Stochastic Optimization. *Diederik P. Kingma, Jimmy Ba*; Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015

[2] Deep Residual Learning for Image Recognition. *Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun*; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[3] ImageNet Classification with Deep Convolutional Neural Networks. *Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton;* Part of Advances in Neural Information Processing Systems 25 (NIPS 2012)

[4] Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. *Dong-Hyun Lee*

[5] SGDR: Stochastic Gradient Descent with Warm Restarts. *Ilya Loshchilov, Frank Hutter*; ICLR 2017 conference pape