

学校代码：10015

密级：公开

北京印刷学院  
硕士学位论文

基于深度学习的机械零件分类关键技术研究

Research on key technologies of machine parts classification based on  
deep learning

作者姓名：杨 北

学 号：2017065106

导师姓名：张勇斌

导师职称：副教授

学科名称：机械工程

学位级别：硕士

2020 年 5 月

## 致谢

论文从选题到最终完成经过了漫长的过程，当论文完成时，心里很高兴。三年的研究生学习生涯让我受益匪浅、收获甚多，这一切首先要感谢我的导师张勇斌教授在学业上给了我很大的帮助，不仅如此，每当研究遇到挫折时，总是能给我鼓励和指导，老师的耐心指导也使得我对自己的研究方向增加信心。每一篇论文，都是在老师严谨的科学态度和谆谆教导中逐步完善的。每天的积累让我对深度学习这个领域有了更多的认识和见解，同时也增加了兴趣。这次论文能够顺利完成，除了我的导师，还有刘玉琴老师对我专业上的指导。

其次我要感谢的是三年来遇到的同学和舍友，尤其是我的舍友陈远爱同学，他为人质朴，乐于助人，我们每天一起看书，讨论问题，一起运动，彼此互帮互助，让我的研究生阶段充满了意义。

最后要感谢我的家人，论文的很大一部分由于疫情是在家里完成的，他们在我写论文的过程中为我提供了精神上的支持，保证我学习的环境。

## 摘要

伴随着新一轮的科学技术革命，人类将进入一个新纪元。传统工业中很多需要人工来手动操作的工作，渐渐地被机器所替代，工业将朝着自动化和智能化的方向发展。对不同的机械零件进行分类识别以实现自动化分拣正是顺应了这一趋势的发展。针对工业领域的生产线零件分拣系统存在识别零件准确率较低的问题，本文提出利用卷积神经网络进行零件的识别分类，主要从两个部分来进行相关研究，构建数据集和设计网络模型。

在实际采集数据的过程中，由于受多种条件的限制，采集到的数据非常有限，没有足够的数据量就不能训练出较好的网络模型，针对这一问题，本文通过多种数据扩充方法来对采集到的原始数据集进行扩充，如镜像变换法、随机裁剪、高斯噪声、白化处理、改变对比度等，这在一定程度上扩充了数据的数量。要想满足数据集的多样性，并且能够训练出更好的网络模型，传统的方法过于单一。因此，本文还采用了生成式对抗神经网络来扩充数据，更加丰富了数据集的多样性，从原始的 400 张图片组成的小样本数据集扩充到 4400 张。其次是设计网络模型，本文选取了 VGG 网络模型为基础，分析了 VGG 网络模型的优缺点，然后针对 VGG 网络模型的不足在其基础上进行改进，调整参数，降低阶层，优化损失函数和梯度下降法等，设计出适合小样本数据集训练的网络模型。本文是基于 64 位 Ubuntu 18.04 版本的 Linux 系统来搭建 Tensorflow 学习框架和 Pycharm 编程环境，完成网络的设计和实验。通过将 VGG 模型和改进后的网络模型进行对比分析，结果表明改进后的模型在识别精度和速度上都略高于 VGG 模型，说明本文的网络模型是可行的。最后将本文提出的基于深度学习的零件分类方法和传统的一些方法进行比较，本文的方法更有优势，主要表现在特征提取方面，这也为工业进一步迈向自动化和智能化提供了一定的参考价值。

**关键词：**深度学习；TensorFlow；零件识别；小样本数据集

## ABSTRACT

With the new round of scientific and technological revolution, mankind will enter a new era. In the traditional industry, many jobs requiring manual operation are gradually replaced by machines, and the industry will develop towards automation and intelligence. The classification and identification of different mechanical parts to realize automatic sorting conforms to the development of this trend. Aiming at the problem of low accuracy in identifying parts in production line parts sorting system in industrial field, this paper proposes to use convolutional neural network to identify and classify parts. The author mainly conducts related research from two parts, constructs data set and designs network model.

In the process of actual data, due to various constraints, the collected data is very limited, there is not enough data quantity will not be able to better trained network model, in order to solve this problem, this paper by the method of multiple data expansion to the original data sets were collected, such as image transformation method, random cutting, gaussian noise, bleaching processing, such as change the contrast, it expanded the amount of data to a certain extent. In order to satisfy the diversity of data sets and train a better network model, the traditional method is too simple. Therefore, this paper also adopts the generative adversarial neural network to expand the data and enrich the diversity of the data set, from the original small sample data set consisting of 400 pictures to 4400 pieces. Followed by the design of network model, this paper selected the VGG network model as the foundation, analyzes the advantages and disadvantages of VGG network model, and then aiming at the shortcomings of the VGG network model based on its was improved, the adjustment parameters, lower class, optimize the loss function and gradient descent method, etc., design a training network model suitable for small sample data set. This paper builds the Tensorflow learning framework and Pycharm programming environment based on the Linux system of 64-bit Ubuntu 18.04 to complete the network design and experiment. By comparing and analyzing the VGG model and the improved network model, the results show that the improved model is slightly higher in recognition accuracy and speed than the VGG model, indicating that the network model in this paper is feasible. Finally, we compare the method of parts classification based on deep learning proposed in this paper with some traditional methods. The method in this paper has more advantages, mainly in the

aspect of feature extraction, which also provides certain reference value for the industry to further move towards automation and intelligence.

**KEYWORDS:** Deep learning; TensorFlow; Part identification; Small sample data set

# 目录

致谢.....	iii
摘要.....	iii
ABSTRACT.....	iv
1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 研究现状.....	2
1.2.1 深度学习的研究现状.....	2
1.2.2 深度学习在图像识别领域的应用现状.....	3
1.2.3 深度学习在工业领域的应用现状.....	4
1.2.4 现有技术存在问题.....	5
1.3 论文研究内容与结构.....	5
1.3.1 论文研究内容.....	5
1.3.2 论文结构安排.....	6
1.4 本章小结.....	7
2 深度学习理论算法.....	8
2.1 神经网络基础.....	8
2.1.1 生物神经元.....	8
2.1.2 神经元数学模型.....	9
2.1.3 人工神经网络.....	10
2.2 卷积神经网络的特点.....	11
2.2.1 局部连接.....	11
2.2.2 权值共享.....	12
2.2.3 最大池采样.....	13
2.3 卷积神经网络的组成结构.....	13
2.3.1 卷积层.....	13
2.3.2 池化层.....	15
2.3.3 激活层.....	15
2.3.4 全连接层.....	17
2.3.5 其它层.....	18

2.3.6 损失函数.....	18
2.4 卷积神经网络训练 .....	19
2.4.1 前向传播算法.....	19
2.4.2 反向传播算法.....	19
2.4.3 梯度下降法.....	22
2.5 TensorFlow 框架.....	22
2.6 本章小结.....	24
3 构建数据集.....	20
3.1 构建小样本数据集.....	25
3.2 数据扩充.....	25
3.2.1 镜像变换法.....	26
3.2.2 随机裁剪.....	27
3.2.3 高斯噪声.....	27
3.2.4 白化处理.....	28
3.2.5 改变对比度.....	29
3.3 使用对抗生成网络进行数据扩充.....	30
3.3.1 DCGAN 网络原理及特点.....	30
3.3.2 DCGAN 网络模型训练过程.....	31
3.3.3 扩充后的数据集.....	32
3.4 本章小结.....	33
4 设计网络模型.....	35
4.1 构建模型.....	35
4.1.1 VGG 网络模型.....	35
4.1.2 改进的网络模型.....	38
4.1.3 损失函数设计.....	40
4.2 实验平台搭建.....	42
4.2.1 系统硬件实现.....	42
4.2.2 系统软件实现.....	42
4.3 数据集准备.....	43
4.3.1 数据标注.....	43
4.3.2 数据集分类.....	43
4.4 网络训练与实验结果分析.....	44

4.4.1 数据读取.....	44
4.4.2 网络训练方法.....	45
4.4.3 训练结果与分析.....	46
4.5 传统的零件分类方法.....	46
4.5.1 基于有向无环图的 SVM 的机械零件分类识别.....	47
4.5.2 基于形状的图像分类算法.....	48
4.5.3 基于纹理的图像分类算法.....	48
4.5.4 基于深度学习的图像分类算法.....	48
4.6 本章小结.....	49
5 总结与展望.....	50
5.1 总结.....	50
5.2 展望.....	50
参考文献.....	52
作者攻读学位期间取得的研究成果.....	55
独创性声明.....	56
学位论文使用授权说明.....	56



## 图表录

图 1-1 各章关系示意图.....	7
图 2-1 生物神经元结构.....	8
图 2-2 人工神经元模型.....	9
图 2-3 单隐含层神经网络示意图.....	10
图 2-4 局部连接示意图.....	12
图 2-5 权值共享示意图.....	13
图 2-6 最大池化操作.....	15
图 2-7 平均池化操作.....	15
图 2-8 Sigmoid 函数图像.....	16
图 2-9 Tanh 函数图像.....	16
图 2-10 ReLU 函数图像.....	17
图 2-11 全连接示意图.....	18
图 2-12 损失函数波动示意图.....	22
表 2-1 各学习框架对比表.....	23
图 3-1 选取的 8 种零件部分样本.....	25
图 3-2 镜像变换示意图.....	27
图 3-3 随机裁剪效果图.....	27
图 3-4 添加噪声效果图.....	28
图 3-5 白化处理效果图.....	29
图 3-6 白化前后像素值对比图.....	29
图 3-7 改变对比度前后效果图.....	29
图 3-8 DCGAN 网络判别结构示意图.....	31
图 3-9 DCGAN 训练过程中损失值变化曲线.....	32
图 3-10 测试效果示意图.....	33
表 3-1 扩充后的数据集.....	33
图 4-1 VGG 网络结构示意图.....	35
图 4-2 随机失活示意图.....	36
表 4-1 VGG 网络模型具体参数配置.....	37
表 4-2 VGG 网络模型优缺点对比表.....	37
表 4-3 VGG 网络模型参数统计表.....	38
表 4-4 改进后的网络模型具体参数.....	39
表 4-5 VGG 与改进后的网络模型参数对比表.....	40

图 4-3 损失函数结构.....	42
图 4-4 数据集标注.....	43
表 4-6 数据集分类.....	44
图 4-5 部分零件数据样本.....	44
图 4-6 动量梯度下降法示意图.....	45
图 4-7 损失值变化曲线.....	46
表 4-7 VGG 与改进后的模型测试结果对比表.....	46
图 4-8 DAGSVM 分类结构图.....	47
图 4-9 基于不变矩和 SVM 的机械零件分类流程图.....	47
图 4-10 基于纹理的图像分类示意图.....	48
图 4-11 基于深度学习的图像分类示意图.....	49

# 1 绪论

## 1.1 研究背景与意义

随着科学技术的迅速发展，工业自动化与智能化的重要性不断提升，现代化工厂对工业机器人的要求也越来越高<sup>[1][2]</sup>。零件的识别和定位作为工业机器人的核心模块，对推进生产自动化有着重要意义<sup>[3]</sup>。传统的零件识别算法有基于图像分割、形状匹配、特征分析等方法<sup>[4][5]</sup>。近年来在提取特征的识别算法方面，提出了很多融合算法，如局部二值模式与向量机融合识别法<sup>[6]</sup>，基于 SVM 与组合矩识别法<sup>[7]</sup>，基于角点检测和霍夫变换识别法<sup>[8]</sup>。这些较传统的识别方法与零件本身的特点相关性较强，对不同的零件识别就要重新设计算法，不具有通用性。另一方面由于受到很多外界环境因素的影响，导致很难提取出零件的相关特征。

人类在人工智能（Artificial Intelligence, AI）领域取得了巨大的进步，目前，人工智能技术已经广泛地应用于各个领域，并且发挥着举足轻重的作用，如医药、交通、汽车、安防等。人工神经网络（Artificial Neural Networks, ANN）是人工智能领域中十分重要的运算模型<sup>[9]</sup>，通过模拟人的大脑来处理复杂的问题，如分类、预测、回归、智能控制、系统识别等。ANN 的发展起源于上个世纪 50 年代<sup>[10]</sup>，到现在为止，已经衍生了很多的网络模型，卷积神经网络（Convolutional Neural Network, CNN）是比较典型的代表。

目前，传统的制造业工厂对于零件的分类大都是依靠人工分拣，这种方法不仅增加劳动成本，而且费时费力，效率低，不利于库存管理。本文就是针对不同的机械零件通过卷积神经网络来实现分类<sup>[11]</sup>。在工业 4.0 的大背景下<sup>[12]</sup>，越来越多的工业生产线使用机器人分拣零件，其中识别零件是分拣系统的关键环节。零件分类的核心是其网络模型的设计<sup>[13]</sup>，卷积神经网络是近年来比较热门的一种识别方法，它避免了人工提取特征的复杂步骤，能够靠自身去学习相关特征，从而提高图像分类的精度和效率。机器零件辨别类型渐渐朝着数字化、信息化、以及高速智能化的方向发展，传统的检测手段主要还是以人工为主，但是随着高新技术的不断发展，这显然已经不再适应现代化工业的需求。在机械制造业领域，要想实现产品自动化、智能化生产，就要寻找一种新技术对制造的机械零件进行检测识别，并完成自动化分拣。深度神经网络便是对零件类型识别和分类技术研究的重点方向，它具有识别精度高、效率快、不与机械零件接触的优势，因此，是当前制造业重点关注的技术之一。

本文将采集到的零件图像信息，通过卷积神经网络进行特征提取，从而进行准确的分类，完成相应的自动化分拣工作。卷积神经网络的应用非常广泛，各行各业都能看到深度学习的身影，工业领域亦是如此。如今社会的工业大方向就是由传统工厂向智能化工厂的转变，进一步执行中国制造 2025 战略<sup>[14][15]</sup>。要想实现工厂的智能化，其自动化程度就得达到一定的水平<sup>[16]</sup>，通过信息化和自动化系统的集成，采用智能的装备与生产系统，实现更加高效、优化、柔性的生产。零件识别是深度学习在机械工业领域中的重要应用之一，它作为机械加工自动化的基础，不仅将人类从繁重枯燥的体力劳动中解放出来，还提高了生产率，降低了成本，更是完善了工厂的自动化体系。在如今的时代背景下，各国争相成为制造业的主导，因此智能工厂、智能制造、智能系统、智能装备成为时代工业领域的新兴代名词。机械零件识别也是迈向智能化的一个关键技术，已经在国内外引起了广泛关注，现有的方法主要集中在两个方面，模板匹配和神经网络，前者相比后者过于传统，本文就是利用卷积神经网络来对零件进行识别分类。

## 1.2 研究现状

### 1.2.1 深度学习的研究现状

卷积神经网络能够发现最早得益于 Rumelhart 于 1986 年提出的 BP 算法<sup>[17]</sup>，随后由 LeCun<sup>[18]</sup>将 BP 算法应用至多层神经网络中并最后于 1998 年提出 LeNet-5 模型<sup>[19]</sup>，其主要对手写数字进行分类，此模型成了之后神经网络的雏形。在此之后的十年里，神经网络停滞不前，神经网络的层数越多，越复杂，需要的计算量就越大，在当时没有能够支撑庞大的计算量的硬件设备。直到 2006 年 Hinton 在《Science》上发表的一篇文章让学者们重新把目光转向多层神经网络，硬件设备相比较以前有了很大的更新，GPU 加速的投入使用使得 DeepLearning 和 CNN 的研究进度有了质的飞跃<sup>[20]</sup>。2012 年的 ImageNet 大赛<sup>[21]</sup>上使用 CNN 的第一名正确率超出第二名接近 10%，人们才重视起来，并且不再局限于手写数字分类，而是广泛地应用于各个领域，如人脸识别，质量检测，图像识别等等。卷积神经网络是图像识别的核心算法之一，并在大量学习数据时有稳定的表现<sup>[22]</sup>。图像特征的提取与分类一直是计算机领域的一个重要的研究方向<sup>[23]</sup>，卷积神经网络提供了一种端到端的学习模型，模型中的参数可以通过梯度下降算法进行训练，本文就是基于传统的梯度下降算法进行改进，采用动量梯度法来训练网络模型，使其能够更好地学习并提取到图像中的特征信息，最终完成对输入图片的分类预测。对于一般的大规模数据集分类问题，卷积神经网络可用于构建阶层分类器<sup>[24]</sup>，也可以在

精细分类识别中用于提取图像的判别特征以供其它分类器进行学习。

在工业化生产过程中自动化、智能化、高精度化的实现也离不开该技术的应用。国内外学者在该技术的理论研究以及实际应用中取得较大的成功与进步，随着计算机技术、信息技术的飞速发展，基于卷积神经网络的各种仪器、系统得到应用和推广，从而推进传统的工厂向智能化工厂的转变。想要将深度学习更恰如其分的应用到工业领域中，就需要对深度学习在图像处理领域以及工业领域的研究现状进行分析。

### 1.2.2 深度学习在图像识别领域的应用现状

深度学习模型从根本上来说属于神经网络，换句话说，神经网络是深度学习的前身。神经网络试图通过模拟人的大脑机理来解决各种机器学习问题，所以深度学习是机器学习的一个新的研究方向。在深度学习出现之前，机器学习在解决实际问题时通常主要精力用于提取特征工程，特征的优劣往往在很大程度上决定了模型的识别精度。由于挑选合适的特征难度大，费时费力，人们打算采用机器代替人工自动学习特征，而深度学习的出现正好解决了这一大难题。深度学习可以从输入的原始数据中自动抽象出目标所需要的特征信息，整个过程不需要人工的干预<sup>[25]</sup>。近年来由于深度学习的广泛使用，越来越多的领域都可以看到深度学习的影子，因此在诸多图像处理问题上深度学习相比很多传统方法甚至其他的深度网络模型具有更优的预测结果<sup>[26]</sup>。不仅如此，深度学习还被应用到图像分类，目标检测和语义分割等多个领域。

在图像识别中，人脸识别是深度学习应用最多的领域。能够引发图像识别变化的因素很多，而这些因素产生的变化往往存在着极为复杂且多变的特征，对于这些非线性的特征，传统的线性模型很难将其区分开来<sup>[27]</sup>。而深度学习可以通过非线性变换获取新的特征，从而避免由诸多因素带来的干扰而导致没有办法将识别对象有效的区分开。人脸识别技术已经越来越成熟，在机场，火车站，企业等场所都可以看到非常成熟的人脸识别系统，人脸识别和刷脸认证已成为时代的主流。随着科技的不断进步，基于深度学习的图像识别技术在交通领域的应用也更加广泛，如车牌识别<sup>[28]</sup>，车道偏离预警<sup>[29]</sup>以及交通标识识别等。同时交通识别技术还被广泛应用于交通控制，收费管理，智能停车等方面。当然深度学习在交通领域的应用远不止于此，近年来，很多研究人员尝试应用到更深层次的领域中，如交通标志检测领域。深度学习在农业领域的应用也非常广泛，如质量检测，种类划分，对水果和蔬菜按照品质进行等级划分。许多专家和学者提出了水果分类识别算法，吕秋霞等人<sup>[30]</sup>设计了基于 BP 神经网络的水果分类系统，该系统主要是

通过对水果的颜色、形状和纹理等表面特征进行分析提取，最终完成对不同种水果的分类，识别率达到了 95% 以上。王水平等人<sup>[31]</sup>采用 SVM 分类器设计了水果种类识别系统，首先对其图像进行分类，然后提取相关特征，利用分类器对水果图像进行分类预测。相关方面的研究还有很多，但这些传统水果图像识别系统的大量工作主要被用来进行图像预处理和特征提取，就会耗费很大的人力和物力，过程极为复杂且效率低。

随着标记数据和 GPU 技术的发展，卷积神经网络的研究取得了很大进展<sup>[32-34]</sup>，相比较传统的神经网络，卷积神经网络避免了人工特征提取和训练分类器等复杂操作，因此引起了大家对这种高效深度学习识别算法的广泛关注。深度学习涉及的领域还有很多，当然在工业领域中也是如此，下面主要分析深度学习在工业领域的应用现状。

### 1.2.3 深度学习在工业领域的应用现状

在机器学习与深度学习还没有发展起来之前，国外有一些非常好的视觉软件，如 Halcon，Visionpro，Cognix，MIL 等，具有定位，测量，检测等功能。但是这些传统的方法来做缺陷检测大多都是依靠人工来提取相关特征，从颜色，形状，长度，宽度来确定被检测的目标是否符合标准，最终定义出一系列的规则来进行缺陷检测。在一些简单的案例中应用的很好，唯一的缺点就是随着被检测物体的变动，所有规则和算法都要重新设计和开发，即使是同样的产品，不同批次的变化同样都会造成不能重复的现实。

凭借自身在图像处理上的优势，深度学习也越来越多的被应用到工业检测领域，在一定程度上代替了传统的机器视觉处理技术，目前主要集中在缺陷检测方面。缺陷检测通常是指对物品表面是否存在瑕疵和不足的检测，从而保证物品的质量。在对工件进行缺陷检测时，往往会存在一系列问题，如工件表面缺陷定位不准和检测率低等，因此乔丽提出“两次识别两次检测”的缺陷检测方法<sup>[35]</sup>，但这样显然增加了一定的工作量，给检测带来不便。陈康提出了一种基于 Faster-RCNN 模型的汽车检测算法<sup>[36]</sup>。南阳等将卷积神经网络应用到工业易拉罐喷码字符检测中<sup>[37]</sup>。张玉燕等人针对 CT 扫描技术得到的金属点阵结构内部灰度图中的缺陷识别和定位问题，提出了一种基于 Faster R-CNN 的缺陷识别算法<sup>[38]</sup>。吴静兰提出了一种基于一类分类的卷积神经网络对电子元件的缺陷进行检测<sup>[39]</sup>，实验表明，该模型在此类数据集上表现出了较好的性能。除了缺陷检测，卷积神经网络还为工业上的其他方向提供了极大的参考价值，如对不同的零件进行分类。徐一丁等选取 10 种工件构建数据集，利用卷积神经网络对其进行识别，有效解决了传统算法中

图像特征难以提取，通用性差的问题<sup>[40]</sup>。周玉强等人使用图像预处理和卷积神经网络相结合搭建了流水线零件识别与分拣系统，通过对不同的零件识别将其搬运到指定的容器，此方法有较高的识别率，分拣系统稳定，可以满足工业流水线零件分拣任务<sup>[41]</sup>。

虽然深度学习在工业领域相较于传统的缺陷检测和分类方法取得了很大的提升，但是技术还没有更加成熟，而且在实际应用中还是很少，因此，将深度学习的优势与工件缺陷自动相结合仍然有很大的发展空间，本文将利用深度学习对不同的零件进行分类识别。

#### 1.2.4 现有技术存在问题

虽然学术界于深度学习技术的研究取得了很大进展，并且在工业领域的应用越来越广泛，但仍然存在很多亟待解决的问题，如：

（1）深度学习模型需要依赖海量的数据做支撑，在实际生产中有些数据难以获取足够的样本，因此对深度学习模型进行前期训练时，容易造成数据集不平衡的问题，导致模型最终的识别率。

（2）需要对不同种类的样本做标签，会耗费大量的时间和精力。

（3）利用大量的数据在训练模型时，对硬件设备的要求非常高。

（4）模型参数较多，训练时间长，调参需要花费较长的时间，相对复杂。

### 1.3 论文研究内容与结构

#### 1.3.1 论文研究内容

在总结前人工作的基础上，本文利用卷积神经网络来对不同种类的零部件进行识别分类，由于条件有限难以获得大量的图片数据，因此，本文主要从扩充数据和构建符合数据集的卷积神经网络两个方面入手，在原始数据不足的情况下实现分类。本文研究内容具体如下：

（1）本文总共选取了 8 种零部件，传动齿轮、弹簧、垫片、滑轮、螺钉、螺母、丝杠和轴承，每种零部件收集 50 张不同样式的图片，构建一个 400 张图片的小数据集。为了使数据量更加具有多样性，并且达到供模型训练的数量，本文将采用不同的扩充技术对原始数据集进行扩充。单一的扩充技术会使数据显得过于单一，影响模型的识别率，因此本文除了利用翻转、裁剪等一些传统的方法外，

还采用了生成对抗神经网络来扩充数据，更加丰富了数据的多样性。

(2) 构建卷积神经网络模型对不同种类的零部件进行分类。卷积神经网络发展至今诞生了很多模型，本文首先选取了 VGG 网络模型，并对其进行了分析，然后在其基础上进行了改进，设计出适合小数据集的网络模型。最后将改进前和改进后的网络模型通过小样本数据集进行训练，并将两者的实验结果进行对比分析，得出最佳网络模型。通过实验发现，此方法可以对各种零件进行有效的分类，相比原有的方法有很多优势，从而更好的实现零件的自动化分拣。

### 1.3.2 论文结构安排

本文一共分为 5 个章节，图 1-1 为各章节的关系示意图。

第一章是绪论部分，首先介绍了在传统的工厂下对不同的零部件进行识别分类的研究意义，然后分析了使用卷积神经网络对不同零部件进行分类在各个领域的研究和应用现状，并总结了深度学习技术在工业领域应用上的一些问题和不足，最后对本文的主要研究内容和结构安排做了简要论述。

第二章介绍了相关的理论知识，首先是关于卷积神经网络的基本概念，主要包括数学模型、工作原理、相关算法、特点以及结构。

第三章是构建数据集，首先对不同的数据扩充方法进行了简单介绍，对原始的数据样本通过不同的方法进行扩充，达到网络模型所需要的数据量，保证数据的多样性以及测试结果的识别准确率。

第四章就是本文的核心部分，设计卷积神经网络模型，主要是在 VGG 网络模型的基础上不断地进行参数调整，降低层数，一方面设计出适合小样本数据的模型，另一方面使其识别率达到最佳，从而确定最终的神经网络模型，并且将本文的方法和之前的零件分类方法做了比较。

第五章对本文的工作进行了概括性的总结，并对未来的研究工作和方向进一步完善展望。



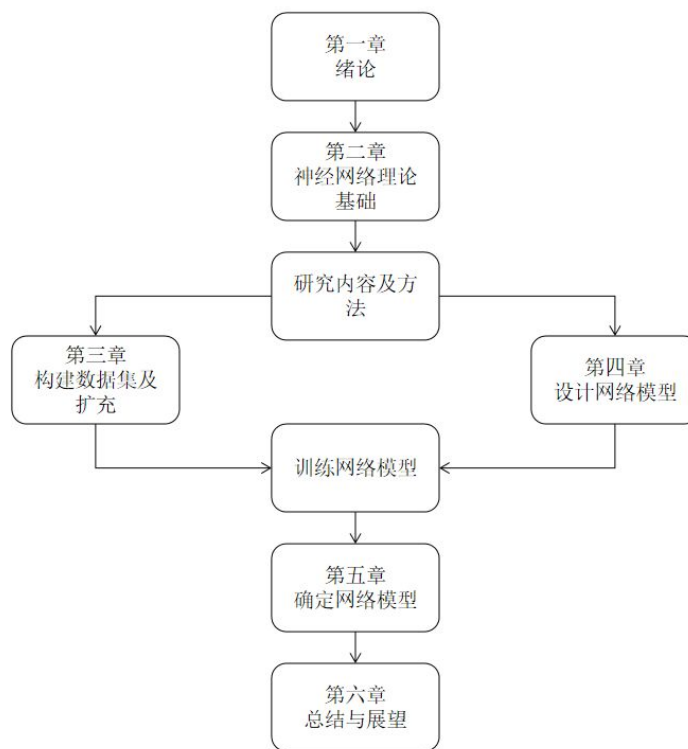


图 1-1 各章关系示意图

## 1.4 本章小结

本章主要是对论文的研究意义，国内外研究现状以及本文的主要研究内容和整体结构安排做了简要的概述。第二章主要介绍深度学习的相关理论知识和算法，为之后的实验部分做铺垫。

## 2 深度学习理论算法

### 2.1 神经网络基础

人类在认识世界和改造世界的过程中，也从自然界和生物特征上汲取了大量的灵感和经验，基于脑与认知的通用人工智能研究已经成为近年来学术界和业界的重要研究热点<sup>[42][43]</sup>。人工神经网络（Artificial Neural Network，即 ANN），是 20 世纪 80 年代以来人工智能领域兴起的一个重要研究方向。它的构筑理念是受到生物神经网络功能的运作启发而产生，以计算机建模为手段，通过软硬件协同实现的机器智能。

#### 2.1.1 生物神经元

神经元使大脑处理信息的基本单元，如图 2-1 所示，可以清楚地看到，整个神经元以细胞体为主，向周围无规则的延伸，形状看似一个树干，这就是所谓的神经元。除了细胞体本身外，还有轴突、树突、突触和神经末梢等部分。

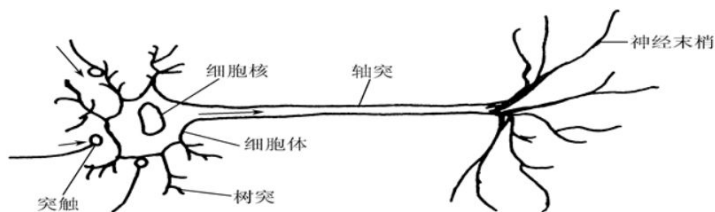


图 2-1 生物神经元结构

所谓的生物神经元存在于人类的大脑中，在自然界，任何生物都有神经元。比如大家熟悉的老鼠大脑拥有 13 个神经元，而人类的大脑有数千亿个神经元，这也说明为什么人类能成为高智慧生物。经相关研究发现，一个典型的生物神经元其实就是一个单细胞，每个细胞都是由细胞主体和很多向四周散发延伸的无规则的卷须组成的。这些卷须整体分为两种：树突和轴突。树突是接受其他神经元发来的信息，而轴突则是向别的神经元发送信息。这期间信息的传递大致分为三个阶段：首先发射神经元产生的电脉冲沿着分支传到轴突的末端，然后通过突触连接到接收神经元的树突，发射神经元的突触释放化学物质在两者间扩散传播，因此神经元之间信息的传递实际上就是一场化学反应。当一个神经元接收到来自多个发射神经元发来的电脉冲时，它会不断的积累，到达一定程度后，它就会向其它神经元发射电信号，从接收神经元转化为发射神经元，以此类推。因此可以看

出，整个大脑里的神经元都是相互连接的，而且结构相当复杂，人工神经网络就是模仿这样的运作原理并基于数学和算法来设计的。

### 2.1.2 神经元数学模型

人工神经元的研究源于脑神经元学说，在生物、生理学领域，Waldeger 等人于 19 世纪末创建了神经元学说。1943 年，Warren McCulloch 和 Walter Pitts 首次创建了一种基于数学和算法的人工神经网络模型，被称为 M-P 模型，该模型是所有神经网络模型的雏形，如图 2-2 所示。该模型是通过模拟生物学上神经细胞的结构和原理来构造出的神经网络模型<sup>[44]</sup>，可以把它看作是一个处理单元，也称作人工神经元。人工神经网络就是以这些神经元为节点，用有向加权弧连接起来的有向图。

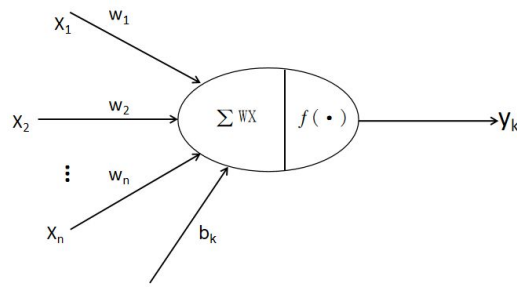


图 2-2 人工神经元模型

在  $n$  个输入中，针对每个输入  $x_i$ ，都有其相应权重项  $w_i$ ，在接收到这  $n$  个输入后，对各条输入进行加权求和。记为式 (2-1)：

$$y_k = \sum_{i=1}^n w_i x_i \quad (2-1)$$

而在求和点处，通常需要加一个偏置项  $b_k$ ，也称预激励，记为式 (2-2)：

$$u_k = y_k + b_k \quad (2-2)$$

而如果将偏移项也视为一个输入  $x_0$ ，将  $x_0$  设置为 1，其权重  $w_{k_0}$  设置为  $b_k$ ，就可以得到式 (2-3)：

$$u_k = \sum_{i=0}^n w_i x_i \quad (2-3)$$

大家都知道，一个简单完整的神经网络模型由多个神经元组成，如图 2-3 所示，这是一个仅含有一个隐含层的神经网络模型，每层含有 4 个神经元，它们之间没

有进行连接，只是在不同的层之间进行全连接或者局部连接，局部连接相比较全连接降低了模型的参数量，因此在训练过程中大大提高了网络模型的计算速度，这也是卷积神经网络的一个重要特点，也正是由于卷积神经网络的这种连接方式，因此训练速度上相比较其他网络具有很大的优势。

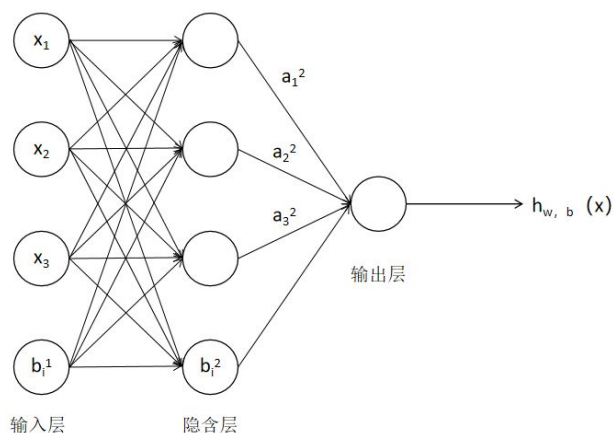


图 2-3 单隐含层神经网络示意图

对于上述的网络，其实还可以继续添加更多的隐含层，构成多隐含层神经网络。整个神经网络会随着隐含层数的增多越能够提取更多更抽象的特征，越抽象的特征越有区分性，越能精准的识别不同的物体图像，从而解决实际问题。

### 2.1.3 人工神经网络

人类自有史以来就是地球上智慧最高的生物，这与人的大脑结构息息相关。本世纪初，很多科学家就开始探索人的大脑，研究发现人类的大脑就像是一个复杂的非线性的并行信息处理系统，它含有数千亿个神经元。本文在前面小节里对生物神经元的运作原理进行了简单介绍，无数个神经元相互连接在一起会形成一个相当复杂的网络，称为神经网络。整个网络就是通过来自外界的刺激从而产生电脉冲不断在多个神经元之间相互传播，发生化学反应，以此来促进人的思考。人工神经网络正是基于这一研究得到的启发，从而对人脑神经元结构进行抽象和模仿，建立某种简单的数学网络模型，在学术界被简称为人工神经网络。从数学角度来说，神经网络是一种运算模型<sup>[45]</sup>。这些大量的神经元可以看成是一个一个的节点，每个节点代表一种特定的输出函数，称为激励函数。每两个节点间的连接都代表一个加权值，称之为权重，这相当于人工神经网络的记忆。神经网络的提出为人工智能迎来了又一春，历史上人工智能的发展经历了 3 次浪潮<sup>[46]</sup>，前两次都是由于缺乏好的运算模型和强大的计算机设备才陷入了低谷，神经网络为人

工智能提供了良好的运算模型。

神经网络是一种经典的机器学习算法，随着对神经网络研究的不断深入，目前，它在模式识别、物体检测、视频分析和图像识别等领域发挥着越来越重要的作用，它的理论已经发展成为一个国际前沿研究领域。本文主要是利用神经网络来识别不同的零件图片，并进行分类，而卷积神经网络则是图片识别领域中最佳的网络模型，接下来作者主要从卷积神经网络特点、结构以及算法等方面介绍卷积神经网络。

## 2.2 卷积神经网络的特点

卷积神经网络是受生物自然视觉认知机制启发而来的，属于前馈型神经网络。由于该网络在很多方面表现突出，已经成为众多科学领域的重要研究方向之一，尤其是在图像分类领域，该网络在前期不需要对数据集进行预处理，可以直接输入原始图像<sup>[47]</sup>。卷积神经网络在前馈神经网络的基础上做了很多改变，在前馈神经网络中，每一层都采用全连接层，也就是每层的神经元之间都要彼此相连接，这样会形成巨大的参数量，模型的计算量也就越大，对于软硬件的要求也不断提高。基于这方面的限制，卷积神经网络将前馈神经网络中神经元的全连接方式变为局部连接，在参数方面采用权值共享的方式，这样在保证不减少网络的深度的同时还大大减少了网络的参数，提高了计算速度和准确率。因此，局部连接和权值共享是卷积神经网络的两个重要特征。

### 2.2.1 局部连接

如果从神经网络的整个发展历史来看，在卷积神经网络出现之前，已经诞生了很多网络模型。最早的 M-P 模型于 1943 年就提出来了，具有简单的逻辑运算能力，人工神经网络就此拉开了帷幕。随后 50 年代末又提出了感知器模型，随后出现了衰退期，直至 80 年代出现了 hopfield，还有 BP 神经网络，90 年代的 SVM 等等。这些都算是卷积神经网络的前辈，但与这些网络相比，卷积神经网络也突出了自己的优势和特点。局部连接就是其中之一，从结构上可以发现，卷积神经网络并没有采用全链接的方式，而是局部连接。所谓局部连接，就是每个卷积层的节点仅仅与其前一层的部分节点相连接，只用来学习局部特征，并不需要与所有的节点相连接，这就是它与其他网络的区别所在。这种局部感知理念主要从动物视觉的皮层结构中受到启发，上个世纪五六十年代 Hubel 和 Wiesel 对此进行了相关研究，发现动物视觉表层中的神经元细胞只对某些刺激进行反应，而对于其他

刺激没有任何反应，说明动物视觉的神经元在感知外界物体的过程中只有一部分神经元起作用，研究者将这种只对某些刺激进行反应的区域称为感受野<sup>[48][49]</sup>。从计算机的角度来看，图像中任何一块区域的像素之间的相关性都与像素之间的距离相关，并且它们之间的关系成正比，也就是说距离较近的像素之间相关性较强，反之较弱。在图 2-4 中可以看到，第  $n+1$  层的每个神经元只与第  $n$  层的 3 个相连接，而并不是与所有的神经元都连接，这样就从原来的  $5*3=15$  个权值参数减少到  $3*3=9$  个权值参数，第  $n+2$  层与第  $n+1$  层之间采用了相同的连接方式。这种局部连接的方式大幅减少了参数数量，提高了计算速度和学习速率，同时也在一定程度上避免了过拟合反应。

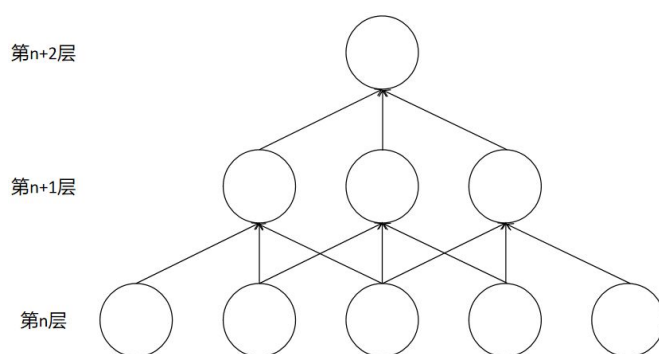


图 2-4 局部连接示意图

## 2.2.2 权值共享

卷积神经网络训练过程中，需要给模型输入图片，每输入一张图片，就会用一个卷积核去对整张图片从左到右依次进行扫描，这个卷积核里的数值就是权重，由于整张图片的每个区域都是同一个卷积核扫描，所以权重一样，这就是所谓的权值共享。经过卷积提取出图像的局部特征，构成了原始图像的特征图，这种方式在一定程度上也降低了模型的参数。比如一个  $5 \times 5$  大小的卷积核，共 25 个参数，它会和输入图片的所有区域进行卷积，来检测相同的特征。在对一张图片进行卷积时不止一个卷积核，而是通过多个不同的卷积核来进行卷积，不同的卷积核用来检测图片上不同的特征，最后将这些所有的局部特征图进行叠加就是整个图片的特征。如图 2-5 所示，通过权值共享的方法，这里一共只有 3 组不同的权值，如果只用了局部连接的方法，共需要  $3 \times 4=12$  个权值参数，如果再加上权值共享，仅仅只需要 3 个权值参数，很明显，又进一步地减少了参数量。

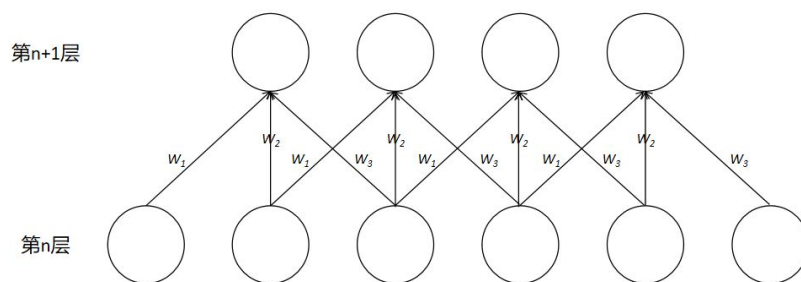


图 2-5 权值共享示意图

### 2.2.3 最大池采样

在通过卷积获取图像特征之后，特征信息还是很多，在对模型训练时仍然会产生很大的计算量。这里提出了卷积神经网络的又一个特点，最大池采样，也叫做最大池化操作，在接下来卷积神经网络的结构中作者作出了详细说明。通过最大池化采样，也可以对特征图进行降维，这些降维后的特征更容易进行分类。因此，除了局部连接和权值共享外，这也是卷积神经网络的一个重要特征。

## 2.3 卷积神经网络的组成结构

一个典型的卷积神经网络通常由若干个卷积层、激活层、池化层及全连接层组成，这些都只是卷积神经网络的最基本的层，除此之外，根据实际情况还会增加一些其它层。下面是对各层的简要概述。

### 2.3.1 卷积层

卷积层是卷积神经网络中较为核心的网络层，由一些大小统一的卷积核组成，每个卷积核的参数都是经过网络训练得到的。在卷积层，通过实现“局部感知”和“参数共享”这两个设计理念，可以达到降维和特征提取的目的。基于图像的空间局部相关性分别抽取图像局部特征，通过将这些局部特征进行叠加，可以形成图片的整体特征。这个过程被称为卷积操作。卷积的概念来源于现实的抽象，本文可以通过列举一个例子来加以说明。

假设有一个任务是实时监控一艘宇宙飞船，这艘飞船带有激光发射器。激光发射器在任意时刻 $t$ 都实时输出信号 $f(t)$ ，这里的 $f(t)$ 表示飞船在任意时刻 $t$ 所处的位置。通常来说，激光信号中都夹杂有噪声信号 $g(t)$ 。为了能更加准确地测量飞船的位置，需要减少噪声的影响，因此需要对获取的距离信号 $x(t)$ 进行平滑

处理。很显然，对于相邻时间的不同输出结果，距离当前时间较近的输出，它们对结果的输出影响也较大（分配较大的权值）。因此，加权平均后的飞船位置  $s(t)$  可以用如下公式(2-4)表示：

$$s(t) = \int_{-\infty}^{\infty} f(a) * g(t-a) da \quad (2-4)$$

其中，函数  $f$  和函数  $g$  是卷积对象， $a$  为积分变量，星号“\*”表示卷积。公式所示的操作，被称为连续域上的卷积操作。这种操作通常也被简记为如下公式(2-5)：

$$s(t) = f(t) * g(t) \quad (2-5)$$

在公式中，通常把函数  $f$  称为输入函数， $g$  称为滤波器或者卷积核，这两个函数的叠加结果称为特征图或者特征图谱。在理论上，输入函数可以是连续的，因此通过积分可以得到一个连续的卷积。但实际上，在计算机处理场景下，它是不能处理连续信号的，因此需要把连续函数离散化。事实上，一般情况下并不需要记录任意时刻的数据，而是以一定的时间间隔进行采样即可。对于离散信号，卷积操作可用公式(2-6)表示：

$$s(t) = f(t) \times g(t) = \sum_{-\infty}^{\infty} f(a)g(t-a) \quad (2-6)$$

当然，对于离散卷积的定义可推广到更高维度的空间上。例如，二维的公式可表示为公式(2-7)：

$$s(i, j) = f(i, j) \times g(i, j) = \sum_m \sum_n f(m, n)g(i-m, j-n) \quad (2-7)$$

卷积神经网络目前被广泛地应用于图像处理领域，通过卷积操作可以对图片中的特征进行提取，从而对图像进行识别。在人眼中，图片看上去就是一个色彩斑斓的整体画面，但是从计算机角度去看的话，它其实是由一个个的小像素点构成，因此可以用二维矩阵来表示一张图片，对图片进行卷积操作的具体表现为公式(2-8)，其中  $(M, N)$  表示输入图像的尺寸大小， $w$  表示卷积核的大小， $b$  为卷积层中的偏置大小。整个网络结构一般会在卷积层的后面添加激活函数  $f$ ，这样就能够保证网络模型能够学习到图像中的非线性特征信息，强化网络提取特征的能力。

$$y_{nm} = f(\sum_{j=0}^{J-1} \sum_{i=0}^{I-1} x_{m+i, n+j} w_{ij} + b), (0 \leq m < M, 0 \leq n < N) \quad (2-8)$$

为了提取更加丰富的特征信息，卷积神经网络中的每一个卷积核都会与图像进行卷积运算，然后对应输出一个特征图，随着卷积层中卷积核数量的不断增加，输出的特征图也越来越多，但需要注意的是不是因为特征图越多，特征信息就越多。因为卷积核增加，无用的特征信息量也会增多，对有用的特征信息造成干扰，从而影响卷积神经网络对图像的识别。因此卷积核的数量也要根据很多因素综合



来定，比如输入图像的大小以及数据集的规模等。

### 2.3.2 池化层

经过卷积后的特征信息依然很庞大，参数很多，影响计算性能，所以要对提取到的特征信息进行池化操作，也可以称之为采样。所谓采样，就意味着降低数据的规模，减少训练参数，提高计算性能，同时又能提取到具有代表性的特征信息，一定程度上避免过拟合。其实对图片进行采样就是对图像进行压缩，将某个局部区域的特征信息作进一步抽象化，虽然这样会减少信息量，但也只是舍去了一些不太重要的信息，留下的就是能够表达图像特征的重要信息，这也是池化的另一个作用：特征不变性。

如果将池化操作细分的话，可分为最大池化、平均池化、求和池化和随机池化，但一般来说，最大池化和平均池化是比较熟悉的，最大池化是将图片中每一个小区域选最大值作为池化结果，平均池化则是选取平均值作为池化结果。如图 2-6 和图 2-7 所示为池化步长为 2 的两种池化效果示意图。

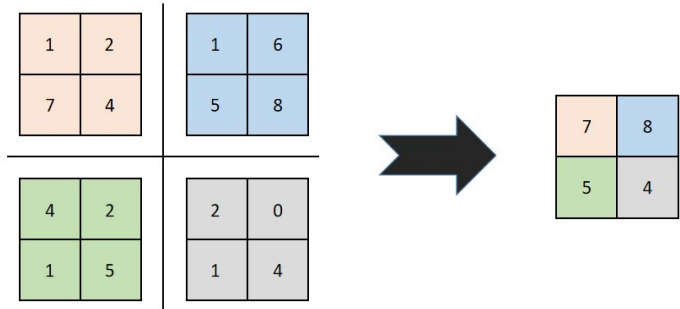


图 2-6 最大池化操作

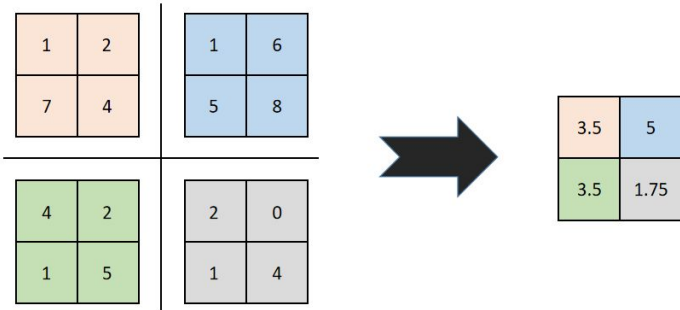


图 2-7 平均池化操作

### 2.3.3 激活层

在卷积神经网络中，神经元与神经元的连接都是基于权值的线性组合，线性之间组合在一起还是线性的，线性方程相对容易解决，但是它们的复杂性有限，并且从数据中学习复杂函数映射的能力更小，所以要对神经网络进行改进，不仅可以学习和计算线性函数，而且还可以学习更加复杂的非线性函数。如果没有激活函数，神经网络将无法学习和模拟其他复杂类型的数据。因此，需要在卷积层后面添加激活函数，深度神经网络才具备了分层的非线性映射学习能力。常见的激活函数有 Sigmoid 函数、Tanh 激活函数和 ReLU 激活函数等。

Sigmoid 激活函数的图像是一个 S 形的曲线，它有一个很大的缺点，就是导数值很小，取值范围仅为[0,1/4]。当输入的数据很大或者很小的时候，导数几乎趋近于 0，很容易产生所谓的梯度消失现象，而且函数在（0,1）之间是连续单调的。因此，Sigmoid 函数常被用于二分类问题。Sigmoid 激活函数的图像如图 2-8 所示。

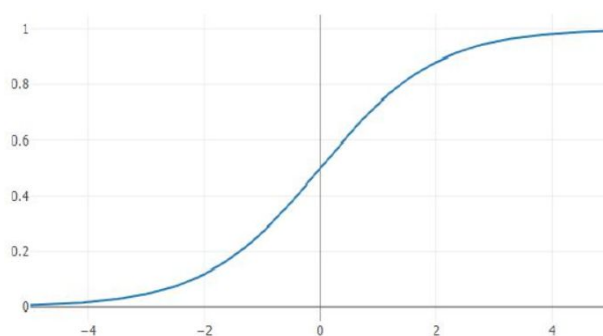


图 2-8 Sigmoid 函数图像

Sigmoid 激活函数的表达式如公式 (2-9) 所示：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2-9)$$

Tanh 激活函数的图像是一个双曲线，所以 Tanh 激活函数也被称为双切正切函数。实际上，Tanh 函数是 Sigmoid 函数的线性变换，二者具有一定的相似性。Tanh 激活函数的图像如图 2-9 所示。

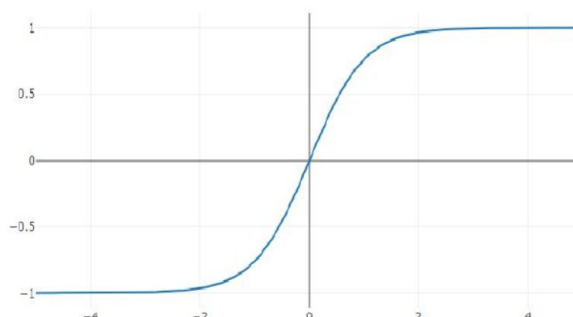


图 2-9 Tanh 函数图像

Tanh 激活函数的表达式如公式 (2-10) 所示：

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-10)$$

目前，在卷积神经网络中，最常用的激活函数就是修正单元 **ReLU**，它是一种非常饱和的激活函数。相比其他几种激活函数而言，**ReLU** 函数具有更强的表达能力。**ReLU** 激活函数的图像如图 2-10 所示。

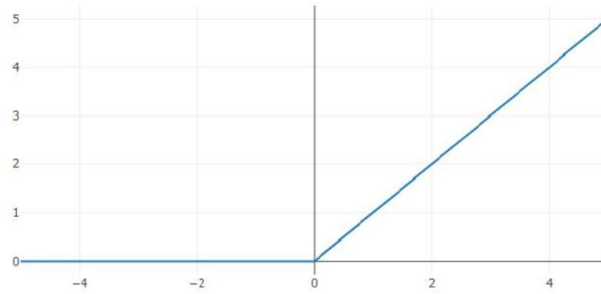


图 2-10 ReLU 函数图像

**ReLU** 激活函数的表达式如公式 (2-11) 所示：

$$f(x) = \max(0, x) \quad (2-11)$$

### 2.3.4 全连接层

卷积神经网络之前，其他的前馈神经网络一般都采用全链接的方式。所谓全连接，就意味着前一层网络中的所有节点都与下一层的所有节点彼此相连接，所以相对于局部连接，全连接的参数量要多得多，一般会占到整个网络参数的 80% 左右，全连接层通常位于网络的末端。如果说前面提到的卷积层、池化层和激活层是将原始数据映射到隐含层特征空间的话，那么全连接层的目的是将前面各层学习到的分布式特征映射到样本标记空间，然后利用损失函数对整个训练过程进行调控，直至损失函数曲线趋于稳定，此时网络的识别率也就达到了最佳，然后就可以对识别的对象进行更加准确的分类预测。所以在整个网络中，全连接层就相当于一个“分类器”。

全连接层的拓扑结构就是一个简单的  $n \times 1$  模式，而前面的卷积层和池化层中的数据通常都是多维的，所以在接入全连接层之前，必须将多维的张量拉成一维的数组，也就是  $n \times 1$  的形态，以便于和后边的全连接层进行连接。如图 2-11 所示，为全连接层的拓扑结构示意图。

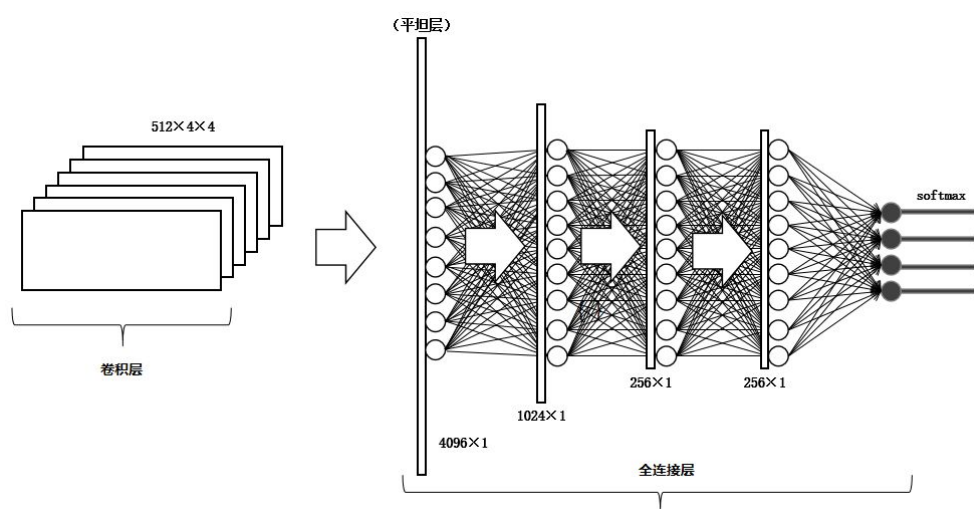


图 2-11 全连接层示意图

由于全连接层的参数冗余，导致该层的参数总数占据整个网络参数的大部分比例，稍有不慎，就会出现过拟合的情况，导致网络的泛化能力下降。为了解决这一问题，通常会采用正则化技术，或者随机舍去一些神经元。

与卷积层不同，全连接层的计算相对简单，直接将输入的数据与向量做矩阵乘法。假设全连接层的输入为  $x$ ，输入维度为  $D$ ，输出维度为  $V$ ， $W$  为  $V \times D$  的权值矩阵，则全连接层的表达式为公式 (2-12)：

$$y = Wx \quad (2-12)$$

### 2.3.5 其它层

以上介绍的网络层都属于卷积神经网络最基本的层结构，除此之外还有一些其它层，如循环层（常用于循环神经网络）、噪声层、局部连接层等等，都是根据特定的应用场景来构造合适的神经网络模型，不同的构造应用效果也不同。

### 2.3.6 损失函数

神经网络模型训练得以实现是通过前向传播来计算损失函数，根据损失函数的值进行反向传播，以此来调整网络的相关参数，使网络达到最优化。因此，损失函数是指导参数进行调整的方向性的指导，如果损失函数随意指示下降的方向，那么可能无论经过多少次迭代，都不可能到达损失函数的最低点。所以，在神经网络的训练中对于损失函数的选取以及梯度下降的各个参数的调整尤为重要。

卷积神经网络的损失函数一般由正则化项和损失项两部分组成，通过后向传播算法进行计算。对于分类问题，通常使用较为普遍的损失函数 Softmax 损失函数，数学表达式为：

$$Ls = - \sum_{i=1}^m \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (2-13)$$

其中， $x_i$  是第  $i$  个输入特征， $y_i$  是第  $i$  个输入对应的类别，也就是样本标签， $m$  是分批时的样本容量， $W$  是全连接层的权重， $n$  是输出的类别数， $b$  是全连接层的偏置。

## 2.4 卷积神经网络训练

### 2.4.1 前向传播算法

在训练卷积神经网络的过程中，一共分为两个部分：前向传播和反向传播。前向传播是从最开始的输入层逐层计算每层的输出，直到最后的输出层。这里以最简单的三层网络模型为例来简单地介绍一下前向传播算法，包括输入层的三个单元  $x_1, x_2, x_3$ ，隐含层和输出层三层， $l$  表示神经网络的层数， $W_{ij}^{(l)}$  表示第  $l+1$  层第  $i$  个节点和第  $l$  层第  $j$  个节点相连的权值， $b_i^{(l)}$  表示第  $l+1$  层上第  $i$  个节点的偏置， $a_i^{(l)}$  表示第  $l$  层第  $i$  个节点的输出值，则每层的输出结果为：

$$a_1^{(2)} = f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)}) \quad (2-14)$$

$$a_2^{(2)} = f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)}) \quad (2-15)$$

$$a_3^{(2)} = f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)}) \quad (2-16)$$

根据上式计算得到隐含层各个节点的输出可以计算出输出层的最终输出  $h_{w, b}(x)$ ，表达式为：

$$h_{w, b}(x) = a_1^{(3)} = f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)}) \quad (2-17)$$

上述介绍的就是神经网络训练时需要使用的前向传播算法，相对反向传播算法来说较为简单，计算量也少。

### 2.4.2 反向传播算法

卷积神经网络的训练主要是利用大量的输入和输出，不断调节网络节点的权值和参数，使之成为一种“端到端”的映射关系。前向传播主要是为了计算神经网络的预测值，反向传播则是利用梯度下降算法对构建的参数模型进行调整和优化，使得预测的输出结果与样本标签之间的误差最小。

本文以单隐含层的神经网络为例，首先要计算样本标签和预测值的误差，然后利用反向传播对参数进行修改。假设神经网络的训练样本数量是  $m$  个，样本集用  $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$  表示，其中  $x$  表示输入，经过前向传播可以计算出预测值用  $h_{w, b}(x)$  表示， $y$  是样本的真实标签，那么由此可以计算出样本标签和预测值的误差，也就是损失代价函数。

对于单个样本，损失代价函数为：

$$E(W, b, x, y) = \frac{1}{2} \|h_{w, b}(x) - y\|^2 \quad (2-18)$$

对于整个训练集，损失代价函数为：

$$\begin{aligned} E(W, b) &= \left[ \frac{1}{m} \sum_{i=1}^m E(W, b, x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[ \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{2} \|h_{w, b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (2-19)$$

式中， $l$  表示神经网络的层数， $s_l$  表示在该层中神经元的个数， $\lambda$  是一个超参数，主要是用来调节此公式中的前后两项的比重。其中，第一部分是均方差项，第二部分是权值衰减项，也就是正则化项，它可以降低权值调节的幅度，减少过度拟合现象的发生。

在神经网络中，训练的目的就是损失函数的值达到最小，因为这个差值越小，说明该神经网络的预测能力越好。深度学习就是根据损失函数采用优化算法来计算参数  $W$  和  $b$ ，通常  $W_{ij}^{(l)}$  和  $b_i^{(l)}$  都会设置一个初始值，一般是一个接近 0 的随机值，也可以初始化成一个符合正态分布的值，之后利用梯度下降法不断对算法进行迭代和优化，使得输出的预测值接近样本标签值，此时的  $W_{ij}^{(l)}$  和  $b_i^{(l)}$  就是神经网络参数的最优解。梯度下降法对权值和偏置项的求解可用公式 (2-20)、(2-21) 进行计算。

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} E(W, b) \quad (2-20)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} E(W, b) \quad (2-21)$$

式中， $\alpha$  表示学习速率，要计算  $W$  和  $b$ ，就得先计算出单个样本损失函数  $E(W, b, x, y)$  对权值和偏置项的偏导数，之后就可以推导出整体损失函数  $E(W, b)$  的

偏导数，具体计算方法可用公式(2-22)和(2-23)表示。

$$\frac{\partial}{\partial W_{ij}^{(l)}} E(W, b) = \left[ \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} E(W, b, x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)} \quad (2-22)$$

$$\frac{\partial}{\partial b_i^{(l)}} E(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} E(W, b, x^{(i)}, y^{(i)}) \quad (2-23)$$

对于单个样本，可以计算出每一层网络输出节点的残差，也就是该神经网络节点的输出值与真实值的差距。以第 $l$ 层为例，残差可以用 $\delta_i^{(l)}$ 表示，假设激活函数用 $f(z)$ 表示，那么最后一层的输出残差 $\delta_i^{(n_l)}$ 为：

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \|y - h_{W, b}(x)\|^2 = -(y_i - a_i^{n_l}) f'(z_i^{n_l}) \quad (2-24)$$

对于第 $n_l-1, n_l-2, \dots, 2$ 层，每个神经网络节点的残差可以由公式(2-25)进行计算。

$$\delta_i^{(l)} = \left( \sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}) \quad (2-25)$$

式中， $\delta_i^{(l)}$ 是根据前一层计算出的残差求均值得到的，然后，通过上面两个公式就可以计算出单个样本损失函数对权值和偏置项的偏导数，具体可以通过公式(2-26)和(2-27)进行计算。

$$\frac{\partial}{\partial W_{ij}^{(l)}} E(W, b, x, y) = a_j^{(l)} \delta_i^{(l+1)} \quad (2-26)$$

$$\frac{\partial}{\partial b_i^{(l)}} E(W, b, x, y) = \delta_i^{(l+1)} \quad (2-27)$$

根据此公式求出单个样本损失函数的偏导数，就可以计算出整体对每层网络中每一个节点的权值和偏置项的偏导数，计算方法如公式(2-28)和(2-29)所示。

$$\frac{\partial}{\partial W_{ij}^{(l)}} E(W, b) = \left[ \frac{1}{m} \sum_{i=1}^m a_j^{(l)} \delta_i^{(l+1)} \right] + \lambda W_{ij}^{(l)} \quad (2-28)$$

$$\frac{\partial}{\partial b_i^{(l)}} E(W, b) = \frac{1}{m} \sum_{i=1}^m \delta_i^{(l+1)} \quad (2-29)$$

### 2.4.3 梯度下降法

在训练卷积神经网络是权值参数的更新都是通过反向传播来确定的，其中最常用的就是梯度下降法，可以很快找到误差函数下降最快的方向，使损失函数值最小。常见的梯度下降算法有三种：

（1）批量梯度下降算法：它是在每次迭代时先求出所有样本的梯度值，再计算出平均值来确定梯度方向，这样确定的方向可以很好的代表所有的样本，准确的朝极值所在的方向对参数进行优化，并且可以并行实现，但是样本数目较大时，计算量会比较复杂，导致训练速度变慢。如果对训练模型的准确度要求高，就可以选用这种方法。

（2）随机梯度下降法：它和第一种方法不同，每次迭代只是随机的选择一个样本的梯度，使得每轮训练的参数更新速度加快，训练时间也在很大程度上缩短了，但是由于单个样本不能代表全体样本的趋势，有可能会使训练结果收敛于局部最优。这种训练方法适用于对训练速度要求高的情况。但是有一个缺点就是，如果出现损失函数波动，就很难判断是否收敛，如图 2-12 所示。

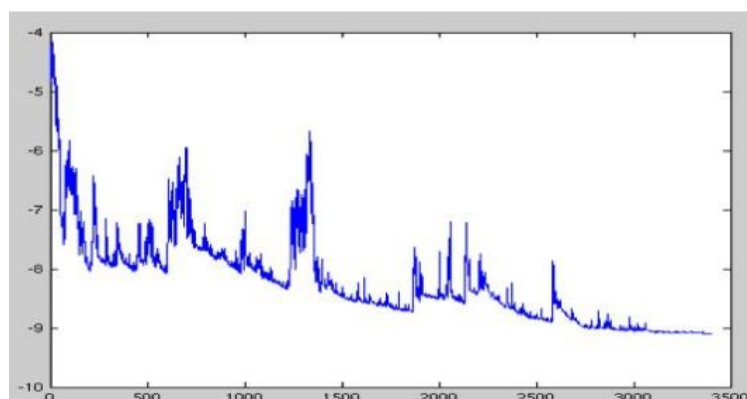


图 2-12 损失函数波动示意图

（3）小批量梯度下降法：它是前两种方法的结合，每次迭代选择其中的一部分样本求导作为梯度方向。

梯度下降算法比较简单，但是如果损失函数不是凸函数时，就会陷入局部最优或者梯度突然消失的情况，因此在训练的时候，可以适当的加入正则化，它可以避免发生过拟合，同时也可以提高神经网络的性能。

## 2.5 TensorFlow 框架

深度学习框架的出现能够更好的帮助研究人员去研究深度神经网络，近年来，



一些团队研发出了很多的学习框架，如 Caffe<sup>[50]</sup>、TensorFlow<sup>[51]</sup>、Keras<sup>[52]</sup>等，本文基于 TensorFlow 框架来完成相关实验。TensorFlow 和其他的框架不同，是采用数据流图的方法对数据进行计算的，它具有以下几大优势：

（1）高度灵活性：TensorFlow 不是严格的神经网络库，但是只要可以构件图，把计算表示成数据流图的形式，描写驱动计算的内部循环，它就可以通过工具帮助搭建组装“子图”。它可以在各种类型的机器上运行，从超级计算机到嵌入式系统。它的分布式架构使大量数据集的模型训练不需要太多的时间。

（2）多语言支持：TensorFlow 的底层设计都是用 C++完成的，同时也提供了多种语言接口，如 Python、C++、Java 等，使得用户可以根据自己的实际需求构建和执行自己的程序。

（3）可移植性：在没有特殊的硬件环境下，TensorFlow 可以在任何具有 CPU 和 GPU 的设备上运行，也可以在自己的笔记本上进行机器学习，还可以把训练好的模型移植到手机 APP 上。

（4）可用性：TensorFlow 工作流程相对容易，API 相对稳定，兼容性好，并且 TensorFlow 能与 Numpy 完美结合，这使大多数精通 Python 数据科学家很容易上手。与其他一些库不同，TensorFlow 不需要任何编译时间，这允许你可以更快地迭代想法。

（5）性能最优化：TensorFlow 框架可以将数据流图在多台设备上并发运行，有效的将计算机的潜能发挥出来，提高了计算速度。

（6）科研和产品相结合：TensorFlow 提供的新算法，可以直接进行训练和计算模型，并将其提供给用户。同时也可以让应用型研究者能将他们自己的想法应用到产品中，提高科研产品率。

TensorFlow 因其巨大的优势，目前被广泛应用于各个领域，GitHub 上的大多数深度学习算法都使用它。作为深度学习最流行的库之一，是谷歌在深刻总结了其前身 DistBelief 的经验教训上形成的。它不仅便携、高效、可扩展，还能再不同计算机上运行，小到智能手机，大到计算机集群。它是一款轻量级的软件，可以立刻生成你的训练模型，也能重新实现它。TensorFlow 拥抱创新，有强大的社区、企业支持，因此它广泛用于从个人到企业、从初创公司到大公司等不同群体。表 2-1 是 TensorFlow 与其他学习框架在 GitHub 上的数据统计表。

表 2-1 各学习框架对比表

框架	框架	支持语言	Stars	Forks	Contributors
TensorFlow	Google	Python/C++/Go...	41628	19339	568
Caffe	BVLC	C++/Python	14956	9282	221
Keras	fchollet	Python	10727	3575	322

表 2-1（续表） 各学习框架对比表

CNTK	Microsoft	C++	9063	2144	100
MXNet	DMLC	Python/C++/R...	7393	2745	241
Torch7	Facebook	Lua	6111	1784	113
Theano	U. Montreal	Python	5352	1868	271
Deeplearning4J	Deeplearning4J	Java/Scala	5053	1927	101
Leaf	AutumnAI	Rust	4562	216	14
Lasagne	Lasagne	Python	2749	761	55
Neon	NervanaSystems	Python	2633	573	52

从表中可以看出，TensorFlow 在 star 数量、fork 数量、contributor 数量上都完胜其他框架，有一统江湖之势。

## 2.6 本章小结

本章中，首先对神经网络的背景、数学模型等方面做了简要的阐述。紧接着在神经网络的基础上，介绍了卷积神经网络的几个特点，然后对卷积神经网络的各个阶层以及各层之间的运作原理做了详细的论述，最后介绍了卷积神经网络的训练方法以及训练卷积神经网络的框架 Tensorflow。

## 3 构建数据集

### 3.1 构建小样本数据集

在使用卷积神经网络训练模型时，为了保证网络模型的学习能力和泛化能力，就需要尽可能地准备大量的数据集来训练网络模型。本文在实际应用过程中所能收集到的数据量非常有限，而且费时费力。甚至在某些特殊领域，根本得不到任何相关数据，更何况小规模数据，比如大家熟悉的医学领域。所以在小样本下进行物体识别越来越得到广泛的关注，这也是数据科学家的一个重要技能。基于卷积神经网络的物体识别技术是一种通过不断学习物体特征来训练模型的方法，只要给神经网络输入合适的数据集，网络模型就可以不断地进行前后向传播调整参数，以此来提高模型提取特征信息的能力，从而更好的对识别对象进行分类预测。所以从中可以看出在使用卷积神经网络对物体进行分类的过程中，主要着重于两个方面，构建小样本数据集和设计神经网络模型，所以本文就是从这两个方面入手来进行研究，设计出对不同种类的零部件进行识别分类的卷积神经网络。

本文主要选取了八种比较典型的零件来作为训练对象，如图 3-1 所示弹簧、垫片、螺母、螺钉、轴承、丝杠、齿轮和轮滑，每种零件采集 50 个，构成一个由 400 张图片组成的小样本数据集。由于受到了多方面因素的影响，本文的数据量较少，而且大部分都来自网络，因此质量不是很好。



图 3-1 选取的 8 种零件部分样本

## 3.2 数据扩充

基于卷积神经网络识别图像，数据集的作用在于训练网络模型，从而让模型提取图片的相关特征，使模型具有更好的学习能力。当然数据越丰富多样，模型的泛化能力越强，识别图片的准确率就越高。在没有足够的情况下，利用小样本数据集对网络模型进行训练会导致网络还没有对图片进行有效地特征提取就不够用了，使得训练的模型不具有很强的识别能力，甚至还会出现过拟合反应。这就相当于一个小学生做题，要掌握一个知识点，就要大量的刷题，刷的越多，对于知识点的理解就越深刻，否则，同样的问题还是会有可能出错。为了提高模型的学习能力就需要采用各种图像处理方法对原始数据集进行扩充，使数据量和模型的规模要匹配，这样模型就可以提取到更多更详细的特征信息，准确率也就随之提高。在训练模型时，要保证这些扩充的图片不会改变原始图片的重要特征，避免带来一定的困扰，否则就会适得其反。随机生成的图片属于原始图片的某一类，但又有所区别，这样在训练模型时既能保证数据的多样性，又不会使计算机认为这些属于完全相同的图片，从整体上提高了模型的泛化能力。下面介绍几种扩充数据的方法。

### 3.2.1 镜像变换法

镜像变换也就是对图片进行翻转，它分为水平镜像变换和垂直镜像变换。两种方法都是以图片的中线为对称轴进行翻转，前者是以水平中线为对称轴翻转，后者则是以垂直线进行翻转。假设图片的宽度为 $W$ ，长度为 $H$ ， $(x_0, y_0)$ 为原图片中像素的坐标， $(x, y)$ 为变换过的图片坐标。这里可以用如下公式(3-1)和(3-2)分别表示水平翻转和垂直翻转后的图片坐标。

$$\begin{cases} x=W-x_0-1 \\ y=y_0 \end{cases} \quad (3-1)$$

$$\begin{cases} x=x_0 \\ y=H-y_0-1 \end{cases} \quad (3-2)$$

如图 3-2 所示为经过镜像变换的图片示意图。从中可以看出翻转后的图像和原始图片没有太大的区别，但这只是人眼看到的变化，在计算机眼中这是完全不同的三幅图片，因为计算机识别图片是以像素的方式进行的，所以变换后的图片的像素也会随之改变，因此可以通过此方式来对原始数据样本进行扩充。

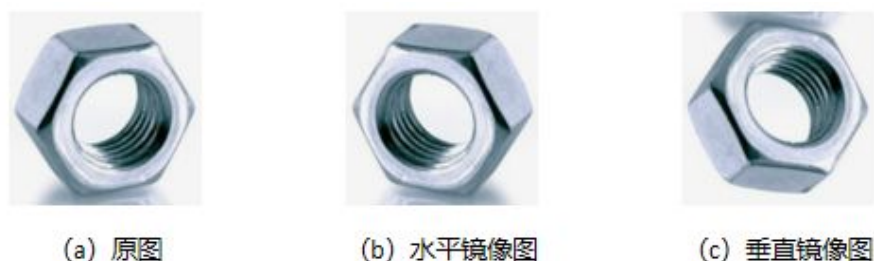


图 3-2 镜像变换示意图

### 3.2.2 随机裁剪

随机裁剪也可以称为多区域裁剪，主要是对零件的不同区域进行随机的裁剪。网络模型主要是根据不同零件的特征来进行提取，来达到识别分类的目的。在现实生活中由于受很多外界因素的影响，被识别的目标会被损坏或者遮挡，通过对零件进行随机裁剪，不仅可以扩充数据样本，还可以提高网络的泛化能力，从而得到更好的识别效率。图 3-3 是对某个零件进行随机裁剪的效果图。

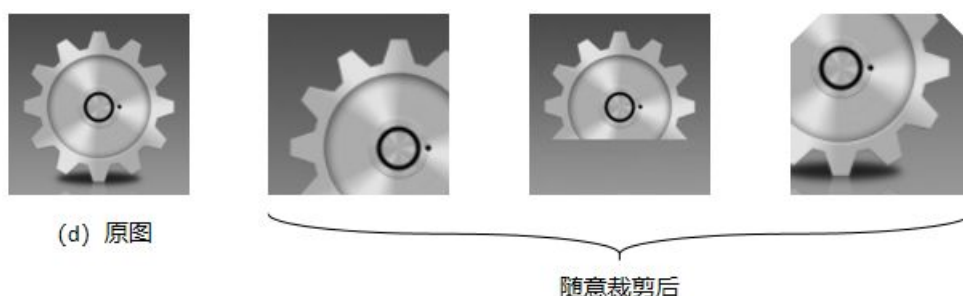


图 3-3 随机裁剪效果图

### 3.2.3 高斯噪声

噪声通常是指图片中的一些干扰的像素信息，常常表现得较为孤立，可有可无，但是这些像素点的存在往往会影响整个图片的质量，因此图片看起来比较模糊。在捕获图片的过程中，往往会受到一些外界或人为添加的噪声。因此，本文将通过给图像中添加适合的高斯噪声来扩充数据样本，丰富数据的多样性，提高模型的泛化能力，这里给出高斯的概率密度函数公式(3-3)。

$$p(x)=\frac{1}{\sqrt{2\pi}\delta}e^{\frac{-(\lambda-\mu)}{2\delta}} \quad (3-3)$$

在使用卷积神经网络解决图片的分类问题时，经常会出现过拟合反应。在对图片进行特征提取时，往往会出现一些次数较多的高频特征信息，这些高频特征信

息很多都是没有实质性帮助的信息，就会对模型学习带来一定的干扰，导致最终的识别效果降低。为了解决这一问题，就需要适当的添加一些噪声使高频特征失真，因此利用高斯噪声可以达到扩充数据的目的。图 3-4 为添加高斯噪声效果图。



图 3-4 添加噪声效果图

### 3.2.4 白化处理

现实生活中，图片成像总会受到环境照明强度、物体反射、拍摄相机等诸多因素的影响，为了能获得图片中包含的那些不受外界环境影响的肯定信息，就需要对采集到的图片进行白化处理。图片的白化处理方式就是改变图片的平均像素值为 0，改变图片的方差为单位方差 1。这里需要先计算原图片的均值和方差，然后对原图片的每个像素值做变换。假设图像  $p$  有  $I$  行  $J$  列，每个像素的值为  $P_{ij}$ ，均值和方差的计算公式分别为公式(3-4)和(3-5)所示：

$$\mu = \frac{\sum_{i=1}^I \sum_{j=1}^J P_{ij}}{IJ} \quad (3-4)$$

$$\sigma^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J (P_{ij} - \mu)^2}{IJ} \quad (3-5)$$

变换后新图片的每个像素值  $x_{ij}$  为公式(3-6)所示：

$$x_{ij} = \frac{P_{ij} - \mu}{\sigma} \quad (3-6)$$

如图 3-5 所示，可以看到经过白化处理过的图片曝光程度减弱了。图 3-6 是图像直方图，白化变换似乎是对原来直方图做了一个横向的拉伸，使得像素值的分布更加的均匀，而不是集中在一个有限的范围内。





图 3-5 白化处理效果图

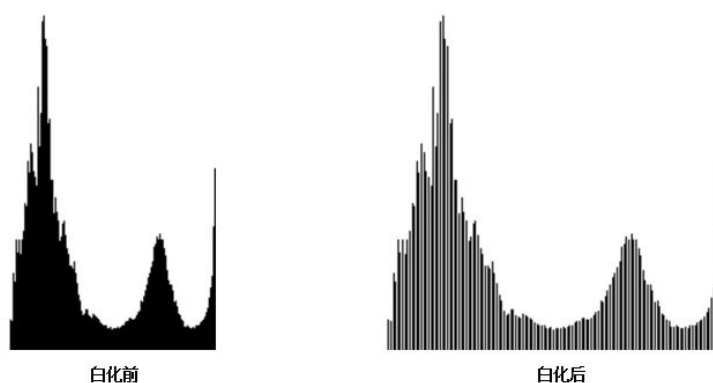


图 3-6 白化前后像素值对比图

### 3.2.5 改变对比度

大家都知道评价一张图片的好坏有很多指标，对比度就是其中关键之一。对比度是指图片中黑与白的比值，也就是从黑到白或者从白到黑的渐变层次，黑主要是指图片中色彩较暗的部分，白则是比较亮的部分。两者的比值越大，渐变层次就越多越明显，图片的色彩表现就越丰富。对比度对视觉效果的影响非常关键，一般来说对比度越大，给人的反差就越大，图像看起来就越清晰醒目，色彩也越鲜明艳丽，反之，对比度越小，整个图片给人感觉不是那么清晰可见。因此对比度越高对于图像的清晰度、细节表现、灰度层次等各方面都有很大帮助，本文可以通过改变图片对比度这一方法来扩充数据集。图 3-7 为改变对比度后的效果图。



图 3-7 改变对比度前后效果图

### 3.3 使用对抗生成网络进行数据扩充

以上都是基于传统的方法对小样本数据集进行扩充,数据类型过于单一,仅靠这些方法,不足以训练出良好的网络模型。针对这一问题,本文又使用了生成对抗神经网络来进行数据扩充,与传统图像增强方法扩充图像集不同,使用生成式对抗网络生成的图像明显区别于原始图像,不仅可以得到数量更多,内容更丰富的高质量图像,数据集扩充效率也得以提升。

生成式对抗网络(GAN)<sup>[53]</sup>自2014年被 Goodfellow 提出后,被誉为近十年来最富想象力的深度学习网络模型,并且被不断改进,形成了 CGAN、DCGAN、WGAN、BEGAN、infoGAN<sup>[54]</sup>等用于图像、自然语言处理甚至语音处理等领域的新模型。由于传统方法在数据增强方面的效率较低,且生成的图像数据带有很多的冗余信息,给网络训练带来更多不确定性。为了得到更丰富的图像数据,降低网络训练难度,一定程度上缓解因数据集太小造成网络模型泛化能力不强的问题,文中受生成式对抗网络启发,将带深度卷积的 GAN 网络用于生成图片,从本质上分离传统图像增强时带来的生成样本相关性,实现数据增强。目前生成式对抗网络主要使用了一些常见的激活函数如 Tanh、Relu<sup>[55]</sup>和 Batch Normalization 等方法简化网络模型参数,使训练变得快速高效,但这些方法实际使用中还是会由于参数调整不当而造成求解损失函数最小值时不收敛或震荡,生成图像质量较差等问题,为此引入一种新的激活函数 Selu 使训练更加稳定,改善训练生成式对抗网络不够稳定的缺点。

#### 3.3.1 DCGAN 网络原理及特点

DCGAN 是一种带深度卷积的 GAN 模型。GAN 模型如图 3-8 所示,主要包括了一个生成模型  $G$  (generator) 和一个判别模型  $D$  (discriminator)。  $G$  负责生成图像,它接受一个随机的噪声  $z$ ,通过该噪声生成图像,将生成的图像记为  $G(z)$ ,  $D$  负责判别一张图像是否为真实的,它的输入是  $x$ ,代表一张图像,输出  $D(x)$  表示  $x$  为真实图像的概率,实际上  $D$  是对数据的来源进行一个判别:究竟这个数据是来自真实的数据分布  $P_{data}$ ,还是来自于一个生成模型  $G$  所产生的一个数据分布  $p(z)$ 。所以在整个训练过程中,生成网络  $G$  的目标是生成可以以假乱真的图像  $G(z)$  让  $D$  无法区分,即  $D(G(z))=0.5$ ,此时便得到了一个生成式模型  $G$  用来生成图像扩充数据集。模型训练的过程构成了  $G$  和  $D$  的一个动态博弈,可以构造式 (3-7) 所示的交叉熵损失函数:



$$V(D,G)=E_{x\sim P_{data}(X)}[InD(x)]+E_{z\sim P_z(Z)}[In(1-D(G(z)))] \quad (3-7)$$

其中,  $x$  为用于训练的真实图像数据,  $P_{data}(X)$  为图片数据的分布,  $z$  为输入  $G$  网络的噪声且已知噪声  $z$  的分布为  $P_z(Z)$ , 而  $G(z)$  为网络生成的图片。在理想状态下, 经过训练,  $G$  尽可能学习到真实的数据分布  $P_{data}(X)$ ,  $G$  将已知分布的  $z$  变量映射到了未知分布的  $x$  变量上。判别模型  $D$  的训练目的就是要尽量最大化自己的判别准确率。当这个数据被判别为来自于真实数据时, 标注 1, 来自于生成数据时, 标注 0。而与这个目的和过程相反的是, 生成模型  $G$  的训练目标是要最小化判别模型  $D$  的判别准确率。在训练过程中, GAN 采用了一种交替优化方式, 它分为两个阶段, 第一个阶段是固定判别模型  $D$ , 然后优化生成模型  $G$ , 使得判别模型的准确率尽量降低。而另一个阶段是固定生成模型  $G$ , 来提高判别模型的准确率。

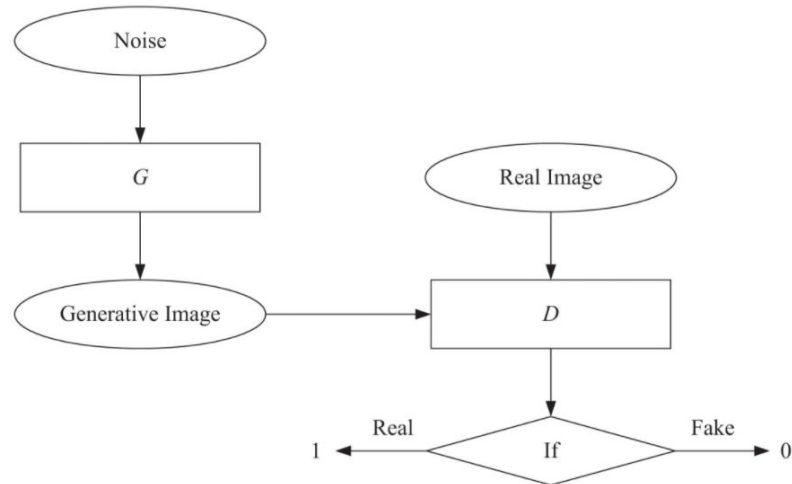


图 3-8 DCGAN 网络判别结构示意图

DCGAN 的特点是在  $D$  和  $D$  网络中使用了深度卷积网络。  $D$  网络与一般卷积神经网络类似, 使用了带步长的二维卷积, 提取深层特征的同时不断减小特征图的大小, 实现下采样。而  $G$  网络则使用了一种转置卷积, 从一维向量逐渐扩充成二维图像, 实现上采样过程。

### 3.3.2 DCGAN 网络模型训练过程

#### (1) 下载 DCGAN 代码

使用 git 工具下载 DCGAN 代码, DCGAN 目录结构介绍如下:

**Checkpoints:** 保存训练后的模型

**Data:** 数据处理和读取的 python 脚本

Launch: 训练模型的接口

Models: DCGAN 模型实现

Networks: DCGAN 实现依赖的网络和基类

Options: 参数设置

Requirements.txt: 运行 DCGAN 需要依赖的包

Sample\_dataset: 包含训练所需的数据集

Utils: 工具类

## (2) 使用 Openface 生成动作单元

在训练 DCGAN 网络模型前,需要使用 Openface 来生成数据集中每张图像的动作单元。生成后会在 Openface 目录下生成一个 processed 目录,该目录保存了对齐后的图片、提取的方向梯度直方图特征、生成的 CSV 文件,其中 CSV 文件中包含了动作单元强度值。

## (3) DCGAN 训练结果

如图 3-9 所示为 DCGAN 训练过程中损失值变化曲线,从图中可以看出在训练过程中随着迭代次数的增加损失值不断减小并趋于平稳。本文通过收集小样本数据集来训练 DCGAN 网络模型,在显存为 4G 的 GTX970 的图像处理器上进行训练。

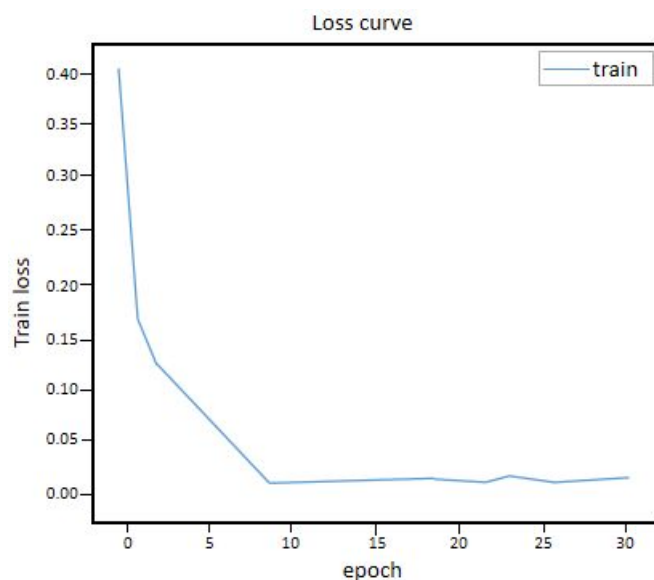


图 3-9 DCGAN 训练过程中损失值变化曲线

如图 3-10 所示是对训练好的 DCGAN 模型测试的效果示意图,第一张为原图,后边是经过 DCGAN 模型生成的三张图片。

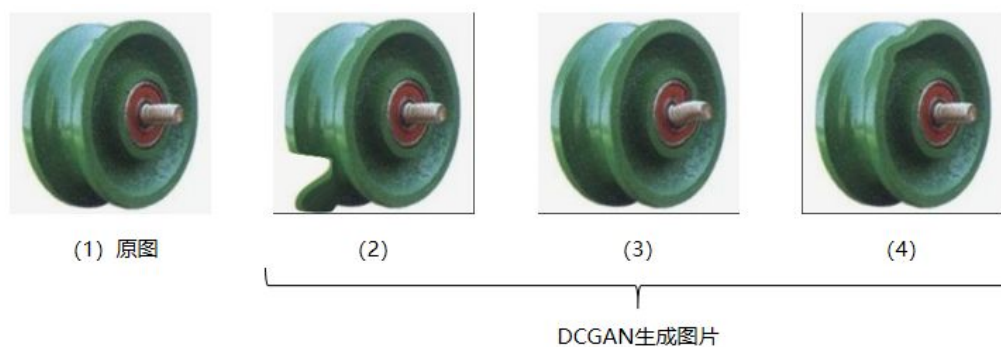


图 3-10 测试效果示意图

### 3.3.3 扩充后的数据集

本节主要对扩充后的数据做一个简单的统计，首先原始的小样本数据集总共有 400 张，本文先是采用了传统的扩充方法，如裁剪翻转等 5 种方法进行了扩充，扩充后的数据为 2800 张。然后又采用了 DCGAN 网络模型对每个原始数据样本进行训练并生成多个样本，每个原始样本另外生成 3 个，一共是 1200 张，前后加起来总共有 4400 张数据，如表 3-1 所示。

表 3-1 扩充后的数据集

	弹簧	垫片	螺母	螺钉	轴承	丝杠	齿轮	轮滑
原始数据	50	50	50	50	50	50	50	50
传统方法 扩充	350	350	350	350	350	350	350	350
DCGAN扩充	150	150	150	150	150	150	150	150
总计	550	550	550	550	550	550	550	550

4400

## 3.4 本章小结

为了实现对不同的零件识别分类任务，本章主要从数据样本方面进行了研究，并构建了小样本数据集。为了使训练的模型达到更好的识别效果，作者对最初的原始样本通过各种扩充方法进行了扩充，其中包括传统的扩充方法和生成对抗神

经网络扩充方法，并且对 DCGAN 的相关原理作了简要概述。通过数据扩充，在一定程度上丰富了数据的多样性，提高了模型的学习能力。

## 4 设计网络模型

### 4.1 构建模型

上一章主要是构建小样本数据集，并通过各种扩充技术对数据集进行了扩充。本章主要从网络结构方面入手，设计出适合小样本数据集训练的网络结构。本章首先对 VGG 网络模型结构进行了分析，然后根据模型的各项参数对 VGG 网络模型进行了改进。通过利用构建好的数据集分别对 VGG 网络和改进后的网络模型进行训练，最后对训练结果进行分析。

#### 4.1.1 VGG 网络模型

卷积神经网络发展至今，诞生了多种网络模型，如 LeNet 模型、AlexNet 模型、GooLeNet 模型、ResNet 模型等，这些模型本质上并没有太大的区别，主要是采用了不同的激励函数，还有就是网络的层数、卷积核的大小、数量等参数的不同。VGG 网络模型由牛津大学的视觉几何组（Visual Geometry Group）和 Google DeepMind 公司的研究员一起研发的深度卷积神经网络，在 2014 年的 ImageNet 比赛中获得了分类项目的亚军和定位项目的冠军，将 Top-5 错误率降到 7.3%。VGG

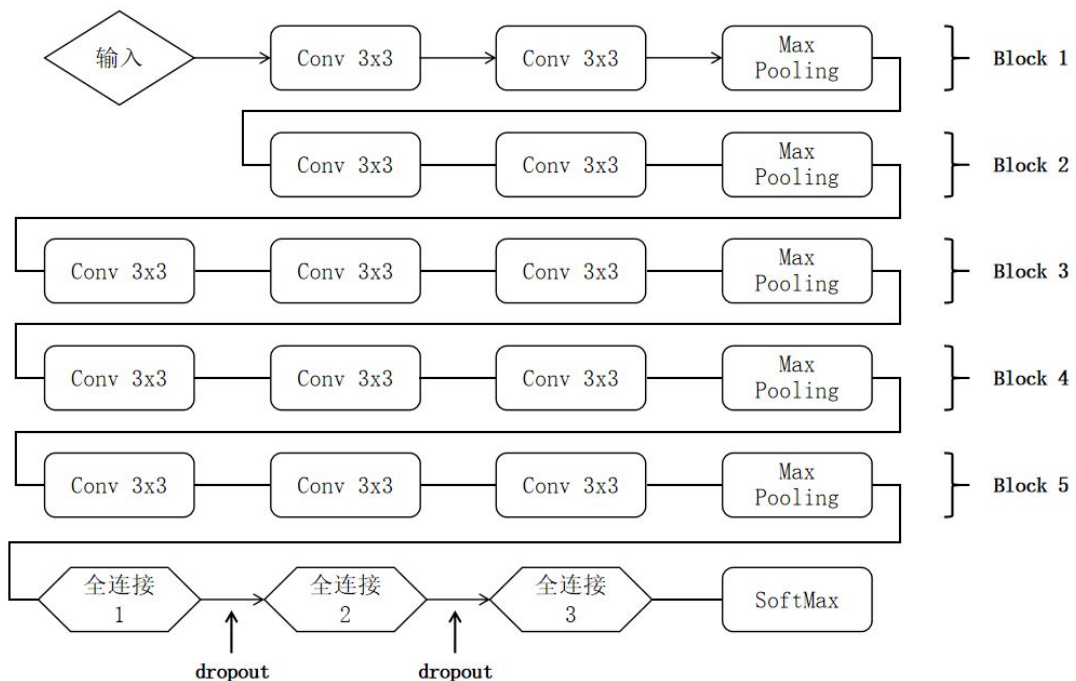


图 4-1 VGG 网络结构示意图

在图像特征提取方面优势十分明显，它主要的贡献是展示出网络的深度（depth）是算法优良性能的关键部分，到目前为止 VGG 依然被广泛的用于提取图像的特征信息。VGG 网络之所以在各领域备受青睐与它独特的结构有关，其中最重要的一点就是该网络采用了更深层次的网络结构，通过添加卷积层的数量加深了网络的深度，提升了卷积神经网络提取特征的能力。同时 VGG 用小卷积核和小池化核代替了大卷积核和大池化核，卷积核的大小可以影响模型的参数量，因此 VGG 模型的参数较少，这也就是该模型能够在控制参数的情况下还能保证其提取特征信息的能力不被减弱的原因所在。

如图 4-1 所示为 VGG 网络结构示意图。从图中可以看出 VGG 网络模型共有 5 个大的卷积模块，每个卷积模块由 2 至 3 个卷积层组成，一共有 13 个卷积层。但并不是每个卷积层后面都会有一个池化层，而是在每个模块后面接一个最大池化层，它的主要作用就是对输入图片的特征信息进行压缩，这样不仅可以简化网络计算的复杂程度，还可以提取到图片的主要特征信息。VGG 网络相对之前的网络模型扩大了图片的输入尺寸，图片尺寸为  $224 \times 224$ ，将图片输入网络模型后，经过一系列反复的卷积和池化操作，接着会经过三个全连接层，为了防止网络出现过拟合和梯度离散问题，每个全连接层的后面还会加上 dropout。以上是对图片进行特征的提取，最后经过一个 SoftMax 分类器，完成最后的图片分类。

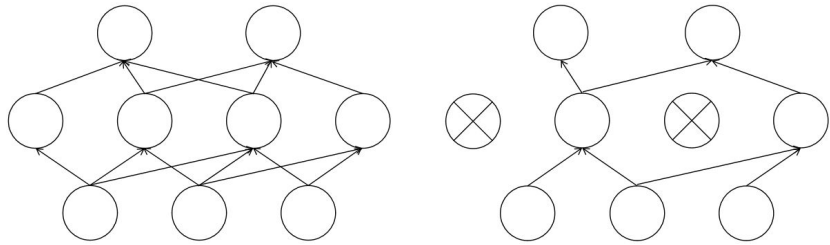


图 4-2 随机失活示意图

VGG 网络模型每层的参数如表 4-1 所示，表中对 VGG 网络中每层卷积核的尺寸、步长、数量、填充等参数进行了详细说明。VGG 网络模型之所以能在 2014 年的 ImageNet 比赛中取得惊人的成绩，主要是因为它将之前模型中的大卷积核换成了小卷积核。比如采用 2 个  $3 \times 3$  的小卷积核代替 1 个  $5 \times 5$  的卷积核，因为 2 个  $3 \times 3$  的卷积核串联在一起相当于 1 个  $5 \times 5$  的卷积核。换句话说，2 个  $3 \times 3$  卷积核相当于 1 个  $5 \times 5$  卷积核的感受野大小，虽然两者的作用相同，但是从参数量来看，前者却只有后者的一半左右，显然后者的计算量要比前者大很多。通过以上各方面参数的比较，可以发现小卷积核能够获得来自不同方向上的像素值变化，使得网络模型具有更强的学习能力，同时也大大减少了网络中参数数量和计算量。此外 VGG 网络中的每个卷积模块中的卷积层彼此之间没有连接池化层，而是在每一

个模块后面连接一个大的池化层，这种方式可以让卷积层参数更少一些，增强了模型的学习能力。

表 4-1 VGG 网络模型具体参数配置

	Conv 3×3	Conv 3×3	Conv 3×3	Pooling	输出大小
输入			3×3×224		
Block1	步长: 1 扩充: 1 核大小: 3 核数量: 64	步长: 1 扩充: 1 核大小: 3 核数量: 64		类型: Max 步长: 2 核大小: 2	112×112×64
Block2	步长: 1 扩充: 1 核大小: 3 核数量: 128	步长: 1 扩充: 1 核大小: 3 核数量: 128		类型: Max 步长: 2 核大小: 2	56×56×128
Block3	步长: 1 扩充: 1 核大小: 3 核数量: 256	步长: 1 扩充: 1 核大小: 3 核数量: 256	步长: 1 扩充: 1 核大小: 3 核数量: 256	类型: Max 步长: 2 核大小: 2	28×28×256
Block4	步长: 1 扩充: 1 核大小: 3 核数量: 512	步长: 1 扩充: 1 核大小: 3 核数量: 512	步长: 1 扩充: 1 核大小: 3 核数量: 512	类型: Max 步长: 2 核大小: 2	14×14×512
Block5	步长: 1 扩充: 1 核大小: 3 核数量: 512	步长: 1 扩充: 1 核大小: 3 核数量: 512	步长: 1 扩充: 1 核大小: 3 核数量: 512	类型: Max 步长: 2 核大小: 2	7×7×512
	全连接层-1		4096 输出节点		1×1×4096
	全连接层-2		4096 输出节点		1×1×4096
	全连接层-3		1000 输出节点		1×1×1000

VGG 网络模型在图像识别领域的应用已经更加成熟，但同时也暴露出一些问题和局限。VGG 模型参数数量多，计算量相当大，而且训练模型需要的数据量也很多，这些都影响了 VGG 网络模型的发展。针对本文的小数据集应用，本文需要对最初的 VGG 网络模型进行改进。如表 4-2 所示为 VGG 网络模型优缺点对比表。

表 4-2 VGG 网络模型优缺点对比表

VGG网络模型的特点	VGG网络模型的局限性
识别率高	参数数量多
小卷积核	计算量大
小池化核	需要大量的训练集进行训练
网络的特征提取能力好	模型太大

### 4.1.2 改进的网络模型

作者在此之前对 VGG 网络模型的结构等方面进行了分析，虽然 VGG 模型在某些方面相比其他模型有很大的优势，但是也存在一些不足。针对小样本数据集，本文在 VGG 模型的基础上使用小卷积核代替大卷积核成功地降低了卷积层的计算量，并且增强了网络模型提取图片信息特征的能力，也在一定程度上避免了过拟合反应，因此采用几个小卷积核的组合要比一个大的卷积核的效果要好很多。在实际应用中，VGG 模型容易受到很多因素的影响，比如训练的数据量不足，很难与现有的网络模型相匹配，导致模型不能够提取到更好的特征信息；其次就是在训练网络模型时还会受到硬件资源的影响，因为所有的实验都要通过计算机去实现。

整个模型的训练就是不断反复地自我调节的过程，不断地处理海量的数据，不断地提取相关特征，不断地计算反向传播的误差，直至误差曲线震荡越小，并趋于稳定。这些所有的计算过程都需要基于一个强大的计算机来实现，因此也会增加硬件资源的成本。为了克服这些问题，并且能让 VGG 网络模型达到实际的应用要求，本文需要在其基础上进行改进，降低模型的参数量和计算量，和现有的计算机硬件资源相匹配，保证模型提取特征的能力。

VGG 相较于之前的网络模型有一定的优势，比如 Alex 模型，其主要原因也是通过采用几个连续的  $3 \times 3$  的卷积核代替 Alex 模型中较大的卷积核。对于本文的小样本数据集，VGG 网络模型的卷积层显得较多，与数据集规模不匹配。通过分析可以发现模型的主要参数都集中在最后的全连接层，其参数量占整个网络模型的 90% 左右。通过对 VGG 模型每层的参数量进行分析，本文要减少参数量可以从全连接层入手。在对模型进行训练时，输入图片的尺寸大小也会对计算的复杂程度有一定的影响，因为图片越多，像素值越大，计算量也会随之增大，因此本文可以通过改变卷积核的数量和图片的大小来改变参数和计算量。如表 4-3 所示为 VGG 网络模型每层参数统计表。

表 4-3 VGG 网络模型参数统计表

卷积层部分						
	层数	核大小	核数量	步长	计算量	参数个数
Block1	Conv3-64	$3 \times 3$	64	1	86, 704, 128	1, 728
	Conv3-64	$3 \times 3$	64	1	1, 894, 668, 061	36, 864
	Max Pooling	$2 \times 2$		2		
Block2	Conv3-128	$3 \times 3$	128	1	924, 844, 032	73, 728
	Conv3-128	$3 \times 3$	128	1	1, 894, 668, 061	147, 456
	Max Pooling	$2 \times 2$		2		



表 4-3（续表） VGG 网络模型参数统计表

Block3	Conv3-256	3×3	256	1	924, 844, 032	294, 912
	Conv3-256	3×3	256	1	1, 894, 668, 061	589, 824
	Conv3-256	3×3	256	1	1, 894, 668, 061	589, 824
	Max Pooling	2×2		2		
Block4	Conv3-512	3×3	512	1	924, 844, 032	1, 179, 684
	Conv3-512	3×3	512	1	1, 894, 668, 061	2, 359, 296
	Conv3-512	3×3	512	1	1, 894, 668, 061	2, 359, 296
	Max Pooling	2×2		2		
Block5	Conv3-512	3×3	512	1	462, 422, 016	2, 359, 296
	Conv3-512	3×3	512	1	462, 422, 016	2, 359, 296
	Conv3-512	3×3	512	1	462, 422, 016	2, 359, 296
	Max Pooling	2×2		2		
全连接层部分						
	FC-4096		4096个输出节点		102, 760, 448	102, 760, 448
	FC-4096		4096个输出节点		16, 777, 216	16, 777, 216
	FC-1000		1000个输出节点，与样本类别有关		4, 096, 000	4, 096, 000
	总数				15, 473, 144, 320	138, 342, 976

早在 2014 年的 ImageNet 挑战赛上，VGG 网络模型就表现出了很不错的成绩，因此可以看出 VGG 网络底层对于图片的特征提取能力非常强。在对 VGG 网络进行改进时，为了保证网络提取图片特征的能力不被减弱，就需要尽可能的保留底层。表 4-4 是改进后的网络模型参数，明显比之前的要少很多。图像输入尺寸改为单通道变为灰度图像，每个像素点只有一个值表示颜色，比之前的 3 通道图像参数少很多。图像大小也由原来的 224×224 改为 96×96，图片越大，包含的特征信息就越多，训练需要的图片就越大，因此图片的大小要与数据规模相匹配。不仅如此，数据量还会影响卷积核的数量，数据量不足时，过多的卷积核会导致样本稀疏，容易出现过拟合反应。反之，过少的卷积核会发生欠拟合反应。经过不断地训练模型，最终确定了卷积核的数量为 8、12、16、16 和 12。

表 4-4 改进后的网络模型具体参数

层	核大小	核数量	步长	扩充	输出尺寸	计算量	参数个数
conv1	3×3	8	1	1	96×96×8	663, 552	72
maxpooling	2×2		2	0	48×48×8		
conv2	3×3	12	1	1	48×48×12	248, 832	108
maxpooling	2×2		2	0	24×24×12		
conv3	3×3	16	1	1	24×24×16	82, 944	144
maxpooling	2×2		2	0	12×12×16		
conv4	3×3	16	1	1	12×12×16	20, 736	144
maxpooling	2×2		2	0	6×6×16		
conv5	3×3	12	1	1	6×6×12	3, 888	108
maxpooling	2×2		2	0	3×3×12		

表 4-4（续表） 改进后的网络模型具体参数

FC1	24	1, 728	1, 728
FC2	6	144	144
总数		1, 021, 824	2, 448

表 4-5 为 VGG 与改进后的网络模型参数对比表，通过表格可以明显的看出改进后的网络参数和计算量都大大减少了，网络层数由原来的 16 层减少为 7 层。

表 4-5 VGG 与改进后的网络模型参数对比表

	VGG	改进后的VGG
卷积层数	13	5
全连接层数	3	2
输入尺寸	224×224	96×96
通道数	3	1
参数数量	138, 342, 976	2, 448
计算量	15, 473, 144, 320	1, 021, 824

### 4.1.3 损失函数设计

在卷积神经网络中，损失函数扮演着极其重要的作用，它的选取直接影响网络模型的训练结果。在图像分类任务中，最常见的就是 softmax 损失函数，它在很多领域表现出了良好的效果，它更多是侧重于对不同类别的分类，提取到的特征向量也不适合利用欧式距离或者余弦相似度进行度量。虽然本文对不同的零件分类属于不同类别的分类，但是所选取的对象有些相似之处，为了使提取到的特征更具有判别性和判断力，因此本文在 softmax 损失函数的基础上引进一种新的损失函数 center loss<sup>[53]</sup>，使得网络在训练的过程中，不仅可以加大内间的差距，也可以减少类内的差距。

Center loss 函数的定义为：

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|^2 \quad (4-1)$$

其中， $x_i$  表示零件样本的特征值， $c_{y_i}$  表示样本  $x_i$  所属类别  $y_i$  的类中心。 $M$  为样本的个数。由于对网络模型进行训练时需要的样本很多，如果每次迭代都要在整个训练样本中选择每个类的类中心，会在很大程度上加大工作量，因此选择基于小批次的方法进行更新。在训练的过程中，为每一个类都维护一个类中心，不断更新类中心，且减小同类样本到类中心的距离，拉大不同类类中心的距离，距离的计算采用欧氏距离。

本文将 center loss 函数和 softmax 函数通过一个参数  $\lambda$  进行结合，得到最终的损失函数，表达式如下：

$$L = L_s + \lambda L_c = -\frac{1}{M} \sum_{i=1}^M \log \frac{e^{W_j^T x_i + b_j}}{\sum_{j=1}^N e^{W_j^T x_i + b_j}} + \frac{\lambda}{2M} \sum_{i=1}^M \|x_i - c_{y_i}\|^2 \quad (4-2)$$

其中， $\lambda$  是两个损失函数之间的平衡系数，一般取值在 0 和 1 之间， $\lambda$  越大表示 center loss 占的比重越大，当  $\lambda=0$  时，该损失函数变为 softmax 损失函数，当  $\lambda=1$  时，变为 center loss 损失函数。

在反向传播过程中， $L_c$  的梯度根据  $x_i$  和  $c_{y_i}$  进行计算，计算方式如 (4-3) 所示。每个批次中类中心  $c_{y_i}$  的变化可以通过公式 (4-4) 进行计算，每个类中心特征的更新公示可由公式 (4-5) 进行计算。

$$\frac{\partial L_c}{\partial x_i} = x_i - c_{y_i} \quad (4-3)$$

$$\Delta c_j^t = \frac{\sum_{i=1}^M \delta \cdot (c_j^t - x_i)}{1 + \sum_{i=1}^M \delta(y_i = j)} \quad \text{其中, } \delta = \begin{cases} 1 & y_i = j \\ 0 & \text{其他} \end{cases} \quad (4-4)$$

$$c_j^{t+1} = c_j^t - \alpha \cdot \Delta c_j^t \quad (4-5)$$

其中，当  $y_i = j$  时， $\delta=1$ ，否则就等于 0， $c_j^t$  表示样本在  $t$  批次中类中心的特征，它会随着网络的训练不断进行更新。 $\alpha$  表示的是类中心的学习率，为了避免个别错误标记的样本对训练造成影响。

将损失函数与改进后的 VGG 网络的全连接层相连接，对网络模型训练，提取深层次人脸特征，其中全连接层 Fc1 的输出就是本文算法最终提取到的零件特征。损失函数结构示意图如图 4-3 所示。

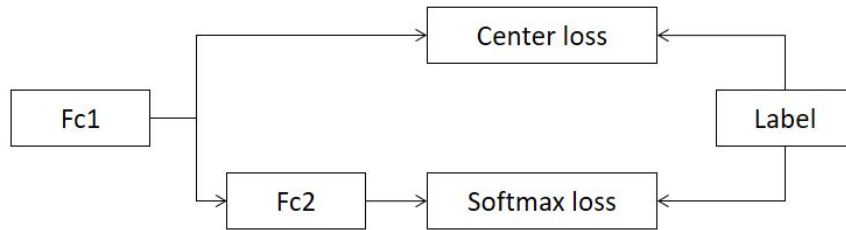


图 4-3 损失函数结构

## 4.2 实验平台搭建

### 4.2.1 系统硬件实现

本文实验使用的硬件环境如下：

- (1) 物理硬盘：英睿达 128GB SSD。
- (2) 内存：8GB，频率为 3200MHz；
- (3) CPU：intel 8400 主频为 2.8GHz；
- (4) GPU：型号 NVIDIA GeForce GTX 970，显存为 4GB，核心频率为 7010MHz；

GPU 起初作为图像处理器在图像处理方面表现出了过人的优势，除此之外，GPU 还擅于大规模的计算，因此也被更多地参与到计算任务中。在这方面，它和 CPU 有一定的区别，这里引用一个例子，对于 CPU 可以把它当做一个教授，GPU 则是由许多的小学生组成。因为 GPU 所参与的计算没有什么技术含量，在使用卷积神经网络识别图片的任务中会涉及到大量的参数计算，但是一些简单的加法和乘法，面对这样没有任何技术含量的大规模计算，最好的办法就是找很多个小学生来分工合作，这就是所谓的人海战术。而 GPU 就不一样了，它相当于一个教授，它可以执行很复杂的运算，比如高等数学中的微积分等，所以不同的计算任务可以选择不同的处理器来完成。而本文是基于卷积神经网络对不同的零件识别，属于前者，因此在训练过程中主要使用 GPU 进行计算。

#### 4.2.2 系统软件实现

软件编程环境如下：

- (1) 操作系统：64 位 Ubuntu 18.04 版本的 Linux 系统；
- (2) 学习框架：TensorFlow；
- (3) 开发环境：pycharm；
- (4) 编程语言：python。

TensorFlow 是谷歌公司研发的第二代人工智能学习系统，是近期比较火的深度学习框架，在 2016 年基于 TensorFlow 学习框架训练出来的人工智能机器人阿尔法狗大战世界围棋冠军李世石，并最终 4:1 的成绩赢得了比赛，至此人们不得不再次重视人工智能的崛起。除此之外，TensorFlow 框架在很多地方都有应用，尤其是在语音识别和图像识别等领域。它是一个采用数据流图，用于数值计算的开源软件库，它拥有灵活的架构，可以在多种平台上展开计算，因此多用于机器学习和神经网络研究。

## 4.3 数据集准备

### 4.3.1 数据标注

在读取数据前，要对数据进行简单的处理。首先要给数据打标签，做标注。本文采用 LabelImg 标注工具来对数据进行标注，它主要用于标注图像中物体名称和边界框，如图 4-4 所示，并生成相应的 xml 文件，物体名称常用英文标注。本文创建的数据集共有 8 种不同的零件，包括“nut”“bolt”“pulley”“gasket”“spring”“bearing”“screw”“gear”，每种零件 50 个，一共 400 个。

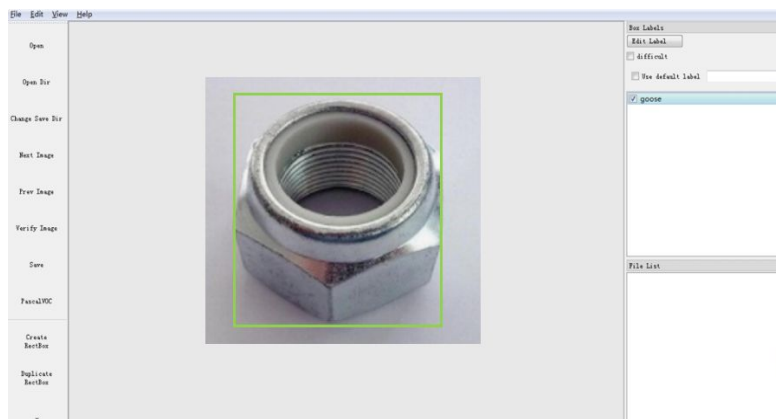


图 4-4 数据集标注

### 4.3.2 数据集分类

在训练模型之前先要对数据集进行相关处理，本文将采集并扩充后的数据按 8:1:1 的比例分为训练集、验证集和测试集三部分，如表 4-6 所示。

表 4-6 数据集分类

	数据样本	平均每个零件
训练集	3520	440
验证集	440	55
测试集	440	55

训练集主要是用来训练模型，让模型不断地提取图片特征并进行学习。验证集是对训练好的模型进行验证，不断地调整参数，使模型的识别率达到最佳。测试集用来测试训练好的模型。图 4-5 是部分数据集。



图 4-5 部分零件数据样本

## 4.4 网络训练与实验结果分析

### 4.4.1 数据读取

本文将原始数据集按照一定的比例分好后，开始进行读取，模型的数据读入由 CPU 完成。hw2.distorted\_inputs()负责生成训练所需的 batch，hw2.inputs()负责生成测试所需的 batch。基本过程如下：程序首先从给定的数据集文件夹中读取所有图片的文件名称及其标签，利用 tf.train.slice\_input\_producer()将其加入到 TF 的 QueueRunner 中，从该队列中读取图片，进行图片预处理后，将其加入到 tf.train.shuffle\_batch()（训练时）或者 tf.train.batch()（测试时）所形成的 QueueRunner 中，训练过程中 hw2.train()则可以按需从队列中读训练数据到 GPU 显存中。

### 4.4.2 网络训练方法

传统的梯度下降算法在每次进行迭代计算时都需要对整个数据集进行遍历，这会大大增加训练时间，降低训练效率。因此本文使用动量梯度下降法对训练过

程中的损失函数变化进行优化，该方法是在梯度下降算法上进行了改进，效果优于梯度下降算法。动量梯度下降算法是通过计算梯度的指数加权平均数来更新权重和偏置，公式如(4-6)、(4-7)和(4-8)所示。

$$v_{dW} = \beta v_{dW} + (1 - \beta) dW \quad (4-6)$$

$$v_{db} = \beta v_{db} + (1 - \beta) db \quad (4-7)$$

$$W = W - \alpha v_{dW}, \quad b = b - \alpha v_{db} \quad (4-8)$$

其中 $v_{dW}$ 为权重的指数加权平均数， $v_{db}$ 为偏置的加权平均数， $\alpha$ ， $\beta$ 是两个超参数， $\beta$ 控制着加权平均数， $\alpha$ 为学习效率影响着权重和偏置的更新。

通过以上的公式可以看出，它并不是直接减去 $\alpha dW$ 和 $\alpha db$ ，而是计算出了一个 $v_{dW}$ 和 $v_{db}$ 。这里需要引入一个新的概念，叫做指数加权平均，通过上面的公式，可以将之前的 $dW$ 和 $db$ 都联系起来，不再是每一次梯度都是独立的情况。其中 $\beta$ 是可以自行设置的超参数，一般情况下默认为0.9（也可以设置为其他数值）。 $\beta$ 代表了现在的 $v_{dW}$ 和 $v_{db}$ 与之前的 $1/(1-\beta)$ 个 $v_{dW}$ 和 $v_{db}$ 有关。0.9就是现在的 $v_{dW}$ 和 $v_{db}$ 是平均了之前10天的 $v_{dW}$ 和 $v_{db}$ 的结果。此时的梯度不再只是我现在的数据的梯度，而是有一定权重的之前的梯度，就像是把原本的梯度压缩一点，并且补上一个之前就已经存在的“动量”。

该方法每一次梯度下降都会有一个之前的速度的作用，如果我这次的方向与之前相同，则会因为之前的速度继续加速；如果这次的方向与之前相反，则会由于之前存在速度的作用不会产生一个急转弯，而是尽量把路线向一条直线拉过去。这就解决了普通梯度下降中存在的下降路线折来折去浪费时间的问题。如果要优化成本函数，函数形状如图4-6，红点代表最小值的位置，利用梯度下降：

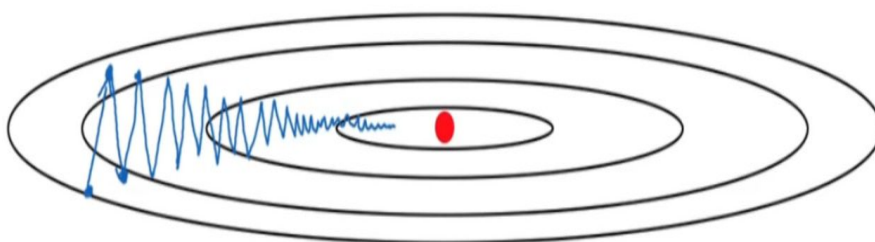


图 4-6 动量梯度下降法示意图

### 4.4.3 训练结果与分析

本文使用小数据集样本分别对 VGG 和改进后的网络模型进行了训练，并对其结果进行了对比分析。在使用卷积神经网络对图片进行分类时，本文主要从两个方面来分析，一个是训练时的损失函数的输出变化，一个是模型最后识别图片的准确率。损失函数越小则对应的模型就越收敛，迭代次数越多则模型的准确率就



会越高。如图 4-7（a）所示为改进后的网络模型损失值变化曲线图，4-7（b）为 VGG 网络损失值变化曲线图。

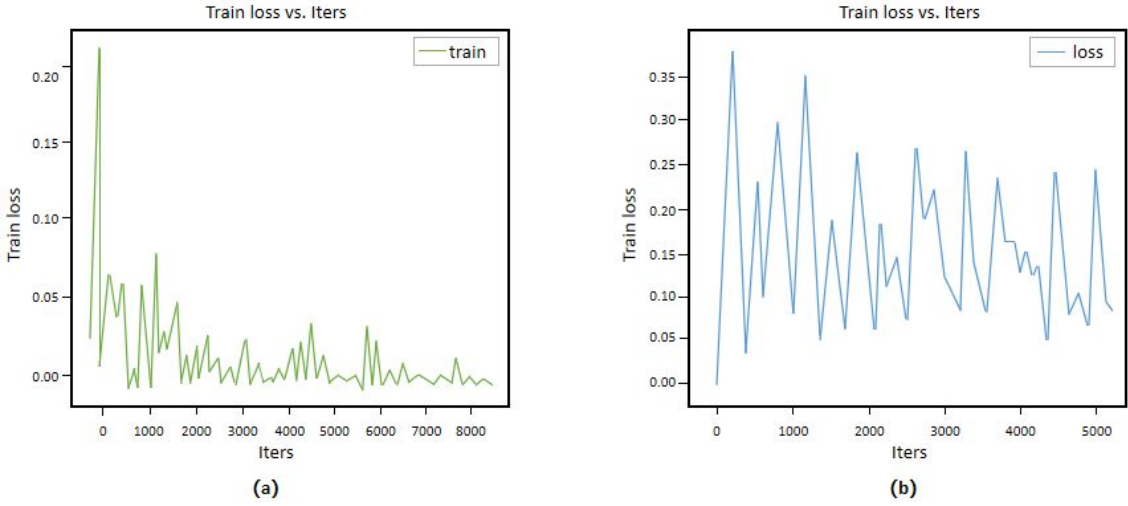


图 4-7 损失值变化曲线

从上图的损失值变化曲线可以看出改进后的网络模型的损失值随着迭代次数的增加逐渐降低，并趋于稳定，这说明改进后的模型在小样本数据集上的表现良好；而 VGG 网络模型的损失值变换很不稳定，说明该模型的训练效果不佳，与现有的小样本数据集规模不匹配。

表 4-7 为改进前的网络模型和 VGG 网络模型的对比表，从表中可以看出，改进后的网络模型不仅准确率比 VGG 模型略高一些，而且在训练时间上也减少了很多，并且提高了识别速度。

表 4-7 VGG 与改进后的模型测试结果对比表

网络模型	数据集规模	实现难易程度	训练时间	正确率
VGG	小样本	复杂	1周	69.6%
改进后的VGG	小样本	容易	6h	81.7%

## 4.5 传统的零件分类方法

### 4.5.1 基于有向无环图的 SVM 的机械零件分类识别

SVM 是 vapnik<sup>[57]</sup>等人在 1995 年基于统计学习理论上提出的一种新的机器学习方法。针对多类零件的分类，需要通过构造多个 SVM 分类器并将它们组合起来实现多类分类，有向无环图方法 DAG 可以将多个 SVM 二类分类器组合起来，



这种方法被称为 DVGSVM，其结构如图 4-8 所示。

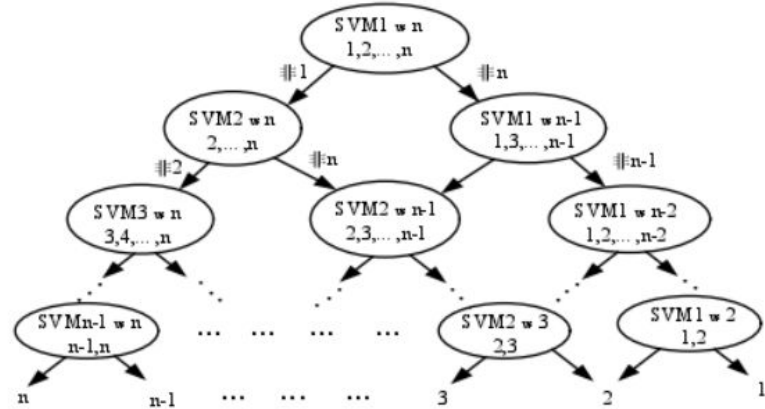


图 4-8 DAGSVM 分类结构图

利用 DAGSVM 实现机械零件的分类，先要训练分类器。训练的结果就是要在 DAGSVM 特征空间中形成由支持向量描述的分类超平面，将各类之间的特征空间划分开来。训练结束，就得到了 DAGSVM 分类器。然后就可以将待分类的机械零件样本的不变矩特征输入到 DAGSVM 分类器，分类过程中，某个零件样本的不变矩特征落入哪个类别训练样本所对应的特征空间的范围，该零件就属于该类别，从而实现了对不同零件的分类，具体流程图如图 4-9 所示。

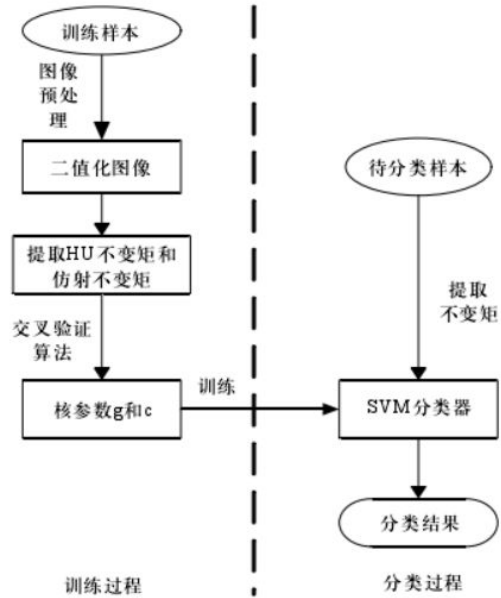


图 4-9 基于不变矩和 SVM 的机械零件分类流程图

#### 4.5.2 基于形状的图像分类算法

基于形状的图像分类算法，形状是由一条封闭的曲线所包围的区域，且该曲

线需要对轮廓边界和边界所包围的区域进行有效描述，因此可基于形状对图像进行分类。该方法以形状的轮廓边界和边界所包围的区域为特征，并基于此创建图像索引，进而到达分类的目的<sup>[58]</sup>。

### 4.5.3 基于纹理的图像分类算法

纹理作为一种重要的视觉线索，存在于任何物体的表面，也就是说任何一个物体都有它的纹理特征，因此可以通过纹理特征来实现对不同物体的分类。纹理特征以描述像素邻域中灰度的空间特征分布规律为本质，它广泛应用于视觉导航、场景分类、物体识别、工业检测和文本分类等领域。

纹理分类算法通过手工特征或者特征学习方法对整个图像进行全局描述，将原始数据变成一个合适的特征矢量，使得分类器能够基于该特征矢量进行分类任务。纹理特征的提取，需要经过特征编码、特征汇聚和特征分类三个基本步骤，具体流程如图 4-10 所示。然而，对于实际开放环境中多种多样的自然图像纹理分析和理解，仍然需要在理论和算法方面进行相关研究，以更好的学习和表达纹理特征。

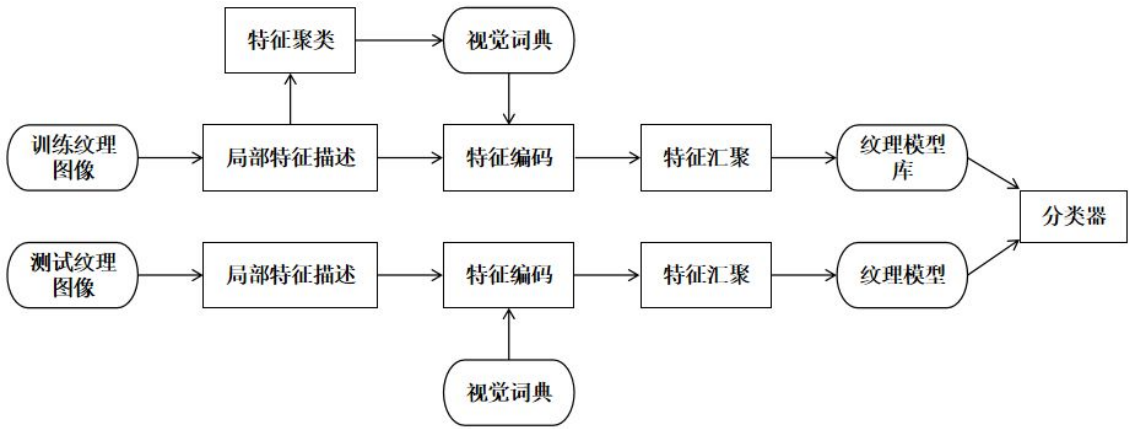


图 4-10 基于纹理的图像分类示意图

目前，该方法还存在一些亟需解决的问题，如构建大规模纹理数据库、纹理图像语义的理解问题、关于纹理的定义等。

### 4.5.4 基于深度学习的图像分类算法

深度学习作为端到端的网络模型，分类前无需手动生成图像特征表达，模型可通过训练过程自动生成图像的特征。本文采用卷积神经网络来实现对不同的零件进行分类并取得了良好的效果，它是深度学习中的典型模型之一，常用于图像

分类，如图 4-11 所示为基于深度学习的图像分类示意图。传统的图像分类法总是要花费更多的时间去进行特征信息的提取，而且提取到的特征往往带有局限性，不能更好的表达整体图像的特征。相比较那些传统图像分类法，卷积神经网络最大的优势就是不需要过多的时间去进行特征信息的提取，网络会自动提取并学习特征，而且识别率和效率更高。因此将卷积神经网络技术应用于对不同的机械零件进行分类能够进一步实现工业自动化和智能化。

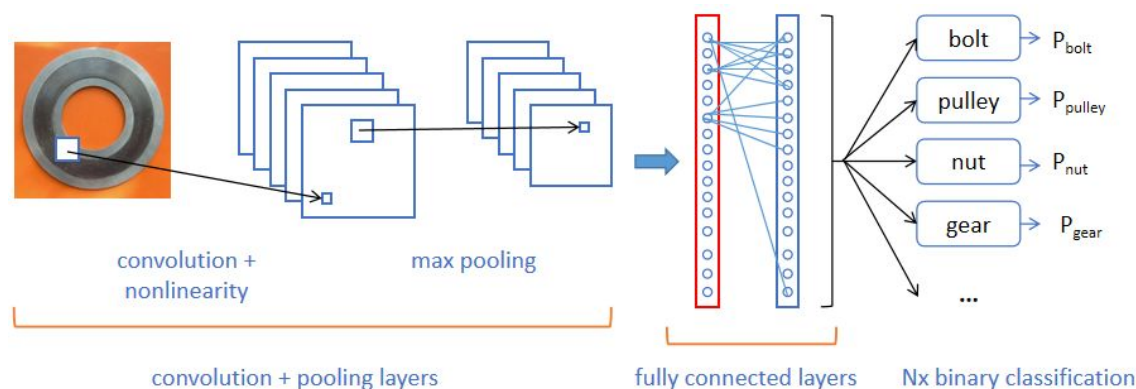


图 4-11 基于深度学习的图像分类示意图

## 4.6 本章小结

本章主要从网络模型这一方面来进行研究，首先对 VGG 网络模型的各方面进行了分析，然而并不适合小样本数据集的训练。VGG 网络的层数较多，为了避免过拟合现象，需要海量的数据来训练模型。因此会涉及到超强的计算量，对硬件设备的要求很高。对于实验室有限的计算机硬件资源来说，这无疑是无法满足的。作者在 VGG 网络模型的基础上进行了改进，不断进行参数调整，降低层数，优化损失函数，最终设计出一个轻量级的网络模型。通过对改进前和改进后的网络模型进行实验对比，确定了最佳的网络模型，实现对不同零件的分类。相比较传统的图像分类法，卷积神经网络能更好的实现图像分类。

## 5 总结与展望

### 5.1 总结

在工业领域对不同的零件进行分类有助于零件的自动分拣,减少更多的人工干预,从而使人类从枯燥单一的工作中解放出来。文中介绍了卷积神经网络在工业领域的应用现状,主要存在两个方面:质量检测和零件分类,在质量检测方面应用的较多,更加趋向于成熟,而在分类问题上显然类似的研究较少。本课题主要研究了使用卷积神经网络在特定的场景下对不同的零件进行识别分类。主要工作总结如下:

(1) 在使用卷积神经网络实现对不同的零件识别时,针对数据不足的情况本文介绍了多种数据扩充法,如镜像变换法、随意裁剪、改变对比度、添加噪声、白化处理等,并利用这些方法对每种零件进行了扩充,从而保证训练模型所需要的数据量。

(2) 作者从各方面分析了 VGG 网络模型的优势与不足,并在此模型的基础上加以改进。由于各种因素的影响采集到的数据量非常有限,为了设计出与之匹配的网络模型,需要降低该网络模型的层数,减少相应的参数量。另一方面受到了硬件资源的限制,减少参数可以保证现有的计算机设备能够支撑起相应的计算量。

(3) 构建好数据集,并设计好网络模型后,开始搭建实验环境,本文的实验是基于 TensorFlow 框架和 Ubuntu 系统环境下完成的。最后对改进前和改进后的网络模型进行了实验对比分析,确定最终的网络模型。

(4) 通过实验结果可以看出,卷积神经网络可以很好的对不同的机械零件实现分类,并取得了良好的效果。与传统的各种图像分类方法相比较,卷积神经网络表现出了更大的优势。

### 5.2 展望

本文针对零件小样本数据集不足的情况进行了扩充,并在 VGG 原本的网络模型基础上进行了改进,但是最终的模型识别率还是很低。主要是在构建数据集的时候由于受到一些外界因素的影响没有办法收集到高质量的图片,导致模型无法提取到更好的图像特征,因此识别率无法到达预期的效果,与此同时,虽然本文

使用了多种方法对数据集进行扩充，但是每个零件都只是一个角度，扩充后的图片虽然有一定的变化，但都是同一个角度，所以扩充后的数据集仍然显得很单一。因此今后此方面的研究需要对每个训练的对象要从多个角度去进行采集，并且使用更多能够改变原本图片角度的扩充方法，以此来丰富数据集的多样性。另一方面就是对网络模型还需要进一步的改进。卷积神经网络近年来在各行各业的应用越来越广泛，在工业领域亦是如此。但是应用方面比较单一，主要的研究还是集中在对零件的质量检测方面，对零件分类的研究很少，当然，要想实现真正意义上的智能化工厂，一方面是对其相关技术的不断研究与改进，另一方面就是将这些技术更多的应用到工厂的各个环节，实现两者的完美结合。

## 参考文献

- [1]王喜文.工业 4.0:智能工业[J].物联网技术,2013,3(12):3-4+6.
- [2]张曙.工业 4.0 和智能制造[J].机械设计与制造工程,2014,43(08):1-5.
- [3]C. Aldrich, C. Marais, B.J. Shean, 等. Online monitoring and control of froth flotation systems with machine vision: A review[J]. International Journal of Mineral Processing, 2010, 96(1-4):1-13.
- [4]王耀南,陈铁健,贺振东,吴成中.智能制造装备视觉检测控制方法综述[J].控制理论与应用,2015,32(03):273-286.
- [5]车录锋,周晓军,徐志农,程耀东.多传感器信息融合机器人工件识别[J].中国机械工程,2000(11):74-76+6.
- [6]吴益红,许钢,江娟娟,毕运锋.一种用于工件识别的 LBP-HOG 特征融合方法[J].南阳师范学院学报,2016,15(09):33-38.
- [7]黄山. 基于机器视觉的机械臂分拣系统的研究[D].江苏科技大学,2016.
- [8]Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems Man & Cybernetics, 2007, 9(1):62-66.
- [9]夏瑜潞.人工神经网络的发展综述[J].电脑知识与技术,2019,15(20):227-229.
- [10]McCulloch,W.S.&Pitts,W.Bulletin of mathematical biophysics ( 1943 ) 5:115.<http://doi-org-s.vpn1.bige.edu.cn/10.1007/BF02478259>
- [11]潘岳.指纹图像模式识别方法的研究[J].信息与电脑(理论版),2018(05):66-69.
- [12]周玉强,陈国栋,盛小明.基于卷积神经网络的零件识别与分拣系统[J].煤矿机械,2018,39(06):151-153.
- [13]周明浩,朱家明.基于卷积神经网络的机械零件表面缺陷分类[J].农业装备与车辆工程,2019,57(11):19-23.
- [14]L.GRoberts,J.T.Tippelt et al.Machine perception of three-dimensional solids.In Optional and Electro-optical Information Processing,MIT Press,Cambridge,1985:159-163.
- [15]Tsai D M , Chen J J , Chen J F . A vision system for surface roughness assessment using neural networks[J]. International Journal of Advanced Manufacturing Technology, 1998, 14(6):412-422.
- [16]林则衡.关于智能工厂的思考[J].科技风,2020(03):217-218.
- [17]Qing-bo Zeng,Hua Sun,Wei-hong Zhou.Monitoring configuration software and its application technology[M].Publication of Harbin Institute of Technology,2010.
- [18]ZHAO li, JIANG Chun-hui, ZOU Cai-rong, 等. A Study on Emotional Feature Analysis and Recognition in Speech[J]. Journal of China Institute of Communications, 2004, 1(4):418-420.
- [19]Logan B.Mel frequency cepstral coefficients for music modeling[J].Proc of Ismir,2000.
- [20]俞颂华.卷积神经网络的发展与应用综述[J].信息通信,2019(02):39-43.
- [21]Schr der J,Anemuller J,Goetze S.Performance comparison of GMM,HMM and DNN based approaches for a-coustic event detection within Task 3 of the DCASE 2016 challenge[C]// Proc.Workshop Detect.Classification Acoust.Scenes Events 2016:80-84.
- [22]Michael Egmont-Petersen, Dick de Ridder, Heinz Handels. Image processing with neural networks - a review. Pattern Recogn 35:2279C2301[J]. Pattern Recognition, 2002, 35(10):2279-2301.

- [23]李彦冬,郝宗波,雷航.卷积神经网络研究综述[J].计算机应用,2016,36(09):2508-2515+2565.
- [24]Srivastava,N.and Salakhutdinov,R.R.,2013.Discriminative transfer learning with tree-based priors.In Advances in Neural Information Processing Systems(pp.2094-2102).
- [25]尹宝才, 王文通, 王立春. 深度学习研究综述%Review of Deep Learning[J]. 北京工业大学学报, 000(1):48-59.
- [26]卢宏涛,张秦川.深度卷积神经网络在计算机视觉中的应用研究综述[J].数据采集与处理,2016,31(01):1-17.
- [27]孟令,朱正,林世忠,晏节晋.浅谈深度学习在图像识别领域的应用现状与优势[J].电工技术,2018(16):145-146+149.
- [28]李兵,易嘉闻,黄锋,晏鹏程,张一鸣,吴健辉,欧先锋.基于 TensorFlow 深度学习的车牌识别算法研究[J].成都工业学院学报,2018,21(04):28-33.
- [29]熊群芳,林军,刘悦,袁浩,游俊.深度学习研究现状及其在轨道交通领域的应用[J].控制与信息技术,2018(02):1-6.
- [30]吕秋霞,张景鸿.基于神经网络的水果自动分类系统设计[J].安徽农业科学,2009,37(35):17392-17394+17439.
- [31]王水平,唐振民,范春年,裴晓芳,谢永华.基于 SVM 的水果分类算法研究[J].武汉理工大学学报,2010,32(16):44-47.
- [32]刘方园,王水花,张煜东.卷积神经网络架构及其应用的研究[J].新型工业化,2017,7(11):40-51.
- [33]常亮,邓小明,周明全,武仲科,袁野,杨硕,王宏安.图像理解中的卷积神经网络[J].自动化学报,2016,42(09):1300-1312.
- [34]李丹,沈夏炯,张海香,朱永强.基于 Lenet-5 的卷积神经网络改进算法[J].计算机时代,2016(08):4-6+12.
- [35]乔丽. 基于 CNN 的工件缺陷检测方法研究及系统设计[D].华中师范大学,2016.
- [36]Xie H , Wu Q , Chen B , et al. Vehicle Detection in Open Parks Using a Convolutional Neural Network[C]// 2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA). IEEE Computer Society, 2015.
- [37]南阳,白瑞林,李新.卷积神经网络在喷码字符识别中的应用[J].光电工程,2015,42(04):38-43.
- [38]张玉燕,李永保,温银堂,张芝威.基于 Faster R-卷积神经网络的金属点阵结构缺陷识别方法[J].兵工学报,2019,40(11):2329-2335.
- [39]吴静兰. 基于深度学习的电子元件缺陷检测关键问题研究[D].华南理工大学,2019.
- [40]徐一丁,杜慧敏,毛智礼,张丽果,顾文宁.基于卷积神经网络的工件识别算法[J].组合机床与自动化加工技术,2018(04):37-40+45.
- [41]周玉强,陈国栋,盛小明.基于卷积神经网络的零件识别与分拣系统[J].煤矿机械,2018,39(06):151-153.
- [42]Terrence J. Sejnowski, Patricia Smith Churchland. Brain and cognition[M]// Foundations of cognitive science. MIT Press, 1989.
- [43]Hillman C H , Erickson K I , Kramer A F . Be smart, exercise your heart: exercise effects on brain and cognition[J]. Nature Reviews Neuroscience, 2008, 9(1):58-65.
- [44]王磊.人工神经网络原理、分类及应用[J].科技资讯,2014(03):240-241.
- [45]盛伯浩.人工神经网络的基本原理及其应用[J].世界制造技术与装备市场,1994(03):70-72.
- [46]全耀.浅谈人工智能的发展史[J].现代信息科技,2019,3(06):80-81+84.
- [47]周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.

- [48]Fukushima K . Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4):193-202.
- [49]Fukushima K . Neocognitron: A hierarchical neural network capable of visual pattern recognition[J]. Neural Networks, 1988, 1(2):119-130.
- [50]王茜,张海仙.深度学习框架 Caffe 在图像分类中的应用[J].现代计算机(专业版),2016(05):72-75+80.
- [51]章敏敏,徐和平,王晓洁,周梦昀,洪淑月.谷歌 TensorFlow 机器学习框架及应用[J].微型机与应用,2017,36(10):58-60.
- [52]范望,韩俊刚,苟凡,李帅.卷积神经网络识别汉字验证码[J].计算机工程与应用,2018,54(03):160-165.
- [53]GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al.Generative adversarial nets[C]//Advances in neural information processing systems.Montreal:NIPS Press,2014:2672-2680.
- [54]CHEN X,DUAN Y,HOUTHOOFT R,et al.Infogan:interpretable representation learning by information maximizing generative adversarial nets [C]//Advances in neural information processing systems.Barcelona,Spain:NIPS Press,2016:2172-2180.
- [55]NAIR V,HINTON G E.Rectified linear units improve restricted Boltzmann machines[C]//Proceedings of the 27<sup>th</sup> international conference on international conference on machine learning.[s.l.]:[s.n.],2010:807-814.
- [56]Yandong Wen, Kaipeng Zhang, Zhifeng Li. A Discriminative Feature Learning Approach for Deep Face Recognition[M]// Computer Vision – ECCV 2016. Springer International Publishing, 2016.
- [57]Gang C , Yang L , Wen-An Z , et al. Research on intelligent fault diagnosis based on time series analysis algorithm[J]. The Journal of China Universities of Posts and Telecommunications, 2008(01):72-78.
- [58]李向阳,庄越挺,潘云鹤.基于内容的图像检索技术与系统[J].计算机研究与发展,2001(03):344-354.



## 作者攻读学位期间取得的科研成果

姓名：杨北

### 一、发表论文/作品：

[1]杨北,杜艳平,朱磊,刘冠一,张媛.智慧物流系统中的新技术应用及案例分析[J].绿色包装,2019(06):35-44.

[2]刘冠一,杜艳平,朱磊,杨北,张媛.智慧物流背景下的供需匹配问题初探[J].绿色包装,2019(07):39-46.

### 二、参加项目：

[1]智慧物流管理与智能服务关键技术研究及应用示范.

### 三、获奖情况

[1]2017 学年获得北京印刷学院研究生学业奖学金和新生三等奖学金

[2]2018 学年获得北京印刷学院研究生二等学业奖学金

[3]2019 学年获得北京印刷学院研究生二等学业奖学金

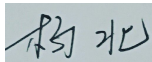
## 独创性声明

本人声明，所呈交的学位论文是本人在导师指导下，进行独立研究工作所取得的成果。本论文对本人和其他人的作品的使用的数量和质量均是在著作权法规定的合理范围之内。除了特别加以标注和致谢之处外，论文中不包含其他人已经发表或未发表的研究成果，也不包含为获得北京印刷学院或其他教育机构的学位或证书而使用过的论文或相关材料。对于本研究工作中给予帮助和所做的任何贡献的同事（同学）均已在论文中作了明确的说明并表示了诚挚的谢意。

如所声明的内容与事实不符，与北京印刷学院、导师及相关机构和人士无关。由此引起的法律纠纷或法律诉讼，由本人承担全部的法律风险。学院依据相关管理条例，视情节轻重，给予相应的处理。

特此声明。

学位论文作者签名：



签字日期：2020 年 6 月 13 日

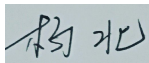
## 学位论文使用授权说明

本人完全了解北京印刷学院有关保留、使用学位论文的规定。特授权北京印刷学院可以将学位论文的全部或部分内容编入毕业论文集及有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意北京印刷学院向国家有关部门或中国学术期刊（光盘版）电子杂志社送交论文的复印件和电子文档。北京印刷学院应保证对本论文只做教学及其他合理目的的使用，并在使用中尊重作者的人身权利。

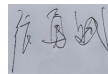
（保密期限： 年 月至 年 月）

（保密的学位论文在解密后使用本授权说明）

学位论文作者签名：



导师签名：



签字日期：2020 年 6 月 13 日

签字日期：2020 年 6 月 13 日