|  |
| --- |
| **模式识别与机器学习课程报告** |
| **基于AlexNet卷积神经网络的图像分类辨识** |
| |  |  | | --- | --- | | 姓　 名： |  | | 学 科 专 业： |  | | 指 导 教 师： |  | | 班 级： |  | | 学 号： |  | |
| 二○二二年十二月 |

**摘要**

本文是《模式识别与机器学习》的课程设计报告，主要用于巩固和实践课程中所学的模式识别与机器学习基础知识，并进行一定的扩展。图像分类识别一直以来都是深度学习的入门学习项目之一，又由于最近在学习Python和Opencv，于是利用Python语言实现图像分类识别。

图像分类识别技术的应用场景很多，如邮局的手写邮政编码、物体分类、行人检测等，特别是在金融、交通等方面应用广泛。当前环境保护问题日渐被重视，故联想到了使用神经网络进行图像识别，进而实现垃圾分类。本文在和鲸社区、百度飞浆社区的开源数据集中寻找各类垃圾的图片作为训练集和测试集，主要利用了AlexNet卷积神经网络探讨图像分类识别。并针对数据集进行了神经网络的搭建。其中搭建卷积神经网络使用了tensorflow库，实现了对垃圾的分类，训练集准确率准确率最高达96.3%，测试集准确率最高达82.4%。此外通过OpenCV库调用外接摄像头实现了基于卷积神经网络的实时垃圾分类，效果良好。

**关键词**：模式识别与机器学习、AlexNet、卷积神经网络、垃圾分类、图像辨识

**目录**

[摘要 I](#_Toc19498)

[第一章 问题描述 1](#_Toc20394)

[1.1 本文的研究背景 1](#_Toc18126)

[1.2 图像辨识技术现状 1](#_Toc30748)

[1.3 本文所做的工作 2](#_Toc20143)

[1.4 本章小结 3](#_Toc16240)

[第二章 数据集的制作和预处理 3](#_Toc13291)

[2.1 数据集的来源简介 3](#_Toc21936)

[2.2 垃圾分类数据集预处理 4](#_Toc14450)

[2.2.1 图像重命名 4](#_Toc9967)

[2.2.2 图像预处理 4](#_Toc2654)

[2.3 本章小结 5](#_Toc14798)

[第三章 建模分析 6](#_Toc11004)

[3.1 人工神经网络 6](#_Toc29416)

[3.1.1 神经元 6](#_Toc15085)

[3.1.2 神经网络模型 7](#_Toc27884)

[3.2 卷积神经网络 9](#_Toc15481)

[3.2.1 CNN理论 9](#_Toc21032)

[3.2.2 深度卷积神经网络（AlexNet） 10](#_Toc9543)

[3.3 本章小结 11](#_Toc19412)

[第四章 基于AlexNet卷积神经网络图像分类识别的设计 11](#_Toc8988)

[4.1 AlexNet卷积神经网络的构建 11](#_Toc31847)

[4.2 卷积神经网络参数的选择 13](#_Toc21328)

[4.3 本章小结 15](#_Toc26024)

[第五章 仿真实现与模型评估 15](#_Toc27410)

[5.1 基于AlexNet卷积神经网络的图像分类识别的仿真实现 15](#_Toc24687)

[5.2 模型评估 17](#_Toc32442)

[5.3 本章小结 18](#_Toc15140)

[第六章 利用外接红外摄像头实现垃圾实时分类 18](#_Toc18613)

[6.1 OpenCV简介 19](#_Toc32129)

[6.2 基于AlexNet卷积神经网络的垃圾图像识别 19](#_Toc5333)

[6.3 AlexNet神经网络实时识别的识别效果 22](#_Toc804)

[6.4 本章小结 23](#_Toc23299)

[第七章 结论 23](#_Toc27636)

[致谢 25](#_Toc11432)

[参考文献 25](#_Toc6246)

# 第一章 问题描述

## 1.1 本文的研究背景

随着我国垃圾分类政策的加速推进，推行垃圾分类落地已经成为各个城市的必然选择，但是在推行垃圾分类的过程中，存在垃圾分类识别难度大、成本高、效率低的难题[1]，如何借助现代技术来实现垃圾分类自动识别，减少人工投入，提高分类效率，将成为影响垃圾分类政策推行进度的重要因素。图像识别技术起源于20世纪40年代，当时由于技术不足以及硬件设施不完善，图像识别技术并未得到快速发展。直到20世纪90年代，人工神经网络与支持向量机相结合，促进了图像识别技术的发展，图像识别技术得到了广泛应用[2]。目前我国正在推行全民垃圾分类，但大众普遍缺乏垃圾分类知识，无法正确进行垃圾分类，所以借助图像辨识技术负责大众进行垃圾种类的辨识非常重要[3]。本文是《模式识别与机器学习》的课程设计报告，属于是在图像辨识领域的初步探索，因此仅在和鲸社区、百度飞浆社区的开源数据集中选取了6类垃圾的图片，进行初步的实践。

## 1.2 图像辨识技术现状

近年来针对图像辨识主要有以下几种方法：

1. 统计决策法

在概率论和数理统计的基础上产生了模式识别的一个经典方法，该方法有着严格的数学基础，优点是抗干扰能力强，但难以抽取好的特征且难以反映模式的精细结构特征。

1. 模糊判别法

该方法的理论基础是模糊数学，用隶属度将模糊集合模式类划分为若干子集，子集总书与模式的类别总数相同，然后根据就近原则分类。但是合适隶属度函数比较难以建立。

1. 逻辑推理法

该方法的基础是人工智能，主要是从事实出发运用知识库中的一系列规则进行推理得到不同的结果，从而对应不同的模式类

1. 神经网络法

神经网络法基于神经网络理论，主要是利用神经网络的学习和记忆功能，让神经网络学习大量的样本，从而使其能够确定新样本的模式类别。该方法能够通过自身学习形成决策区域，而不需要给出经验知识和判别函数。在相继出现了各种分类方法之后，支持向量机（SVM）方法出现了，该方法从理论上来说是全局最优的，目前支持向量机的性能已经在数字识别领域得到了验证。

在GPU没有大量普及的20世纪90年代，计算机的硬件能力较弱，很难完成一个多通道、多层和有大量参数的卷积神经网络训练任务[4]。而且包含许多特征的深度模型需要大量的有标签数据才能表现得比其他经典的方法好，故在20世纪90年代，神经网络在图像识别领域的应用较少。随着通用GPU的出现改变了这一格局，自2021年涌现出诸如OpenCL和CUDA之类的编程框架，使得GPU在2010年至今都在被机器学习的社区使用。

## 1.3 本文所做的工作

本文主要是为了加深对《模式识别与机器学习》课程内容的理解，并学习更多模式识别与机器学习、神经网络等领域的知识，属于是在图像辨识领域的初步实践，主要所做的工作有以下几项：

1. 在和鲸社区、百度飞浆社区收集垃圾的图片，并制作数据集。将自制的数据集划分为训练集和测试集，因此要保证其图像格式的规范化。自行编写Python脚本对该数据集进行了规范的命名化工作和图像数字化的处理。
2. 分析和实现了AlexNet卷积神经网络的图像分类识别。对于大尺寸的输入图像，使用全连接层容易导致模型过大，这会带来过于复杂的模型和过高的开销。对于卷积神经网络。一方面，卷积层保留输入形状，使图像的像素在高和宽两个方向上的相关性均有可能被识别；另一方面，卷积层通过滑动窗口将同一个卷积核与不同位置的输入重复计算，从而避免参数尺寸过大。针对自制数据集实现了基于AlexNet卷积神经网络的分类识别，识别准确率约为82%左右。
3. 利用OpenCV调用外接摄像头完成了垃圾的实时分类，效果良好。OpenCV是一个功能强大的计算机视觉库。使用训练好的垃圾分类模型，借用OpenCV库调用外接摄像头，并对采集到的图片进行修整，实现了垃圾的实时分类。

## 1.4 本章小结

本章主要对当前国内垃圾分类现状进行梳理，提出使用图像辨识技术辅助垃圾分类的意义，引出了图像分类识别的研究问题。之后对本文所做的工作进行了梳理。

# 第二章 数据集的制作和预处理

## 2.1 数据集的来源简介

和鲸社区（原“科赛网”）是中国知名的第三方数据科学社区之一，较早一批专注于大数据算法比赛的平台，拥有近20万注册数据科学家用户，辐射超过30万数据人才群体。和鲸科技旗下的ModelWhale（原K-Lab），是具备国际领先性的数据科学SaaS平台，可满足数据科学家、人工智能工程师、商业分析师等数据工作者在线完成分类、建模、分析、可视化、结果输出等任务，并支持私有化部署和云端协同，帮助企业、高校、科研机构开展工业级数据科学应用与人工智能研发。

和鲸社区中提供了大量的数据集供深度学习、图像识别领域的学者使用，本文的数据集大部分来自于和鲸社区，小部分来自百度飞浆社区。

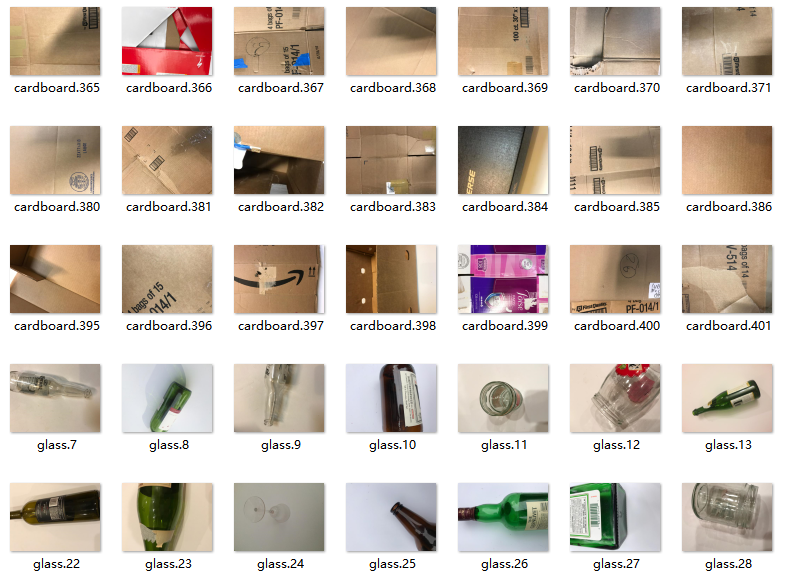


图 2.1 垃圾分类数据集样张

## 2.2 垃圾分类数据集预处理

### 2.2.1 图像重命名

由于在实现垃圾分类图像识别的过程中需要知道训练集图像的属性标签，因此需要给训练集图像有序命名，以便后续得知真实的分类。

由于训练集图像过多，手动命名过于费时，故使用PytQt编写了一个文件重命名软件，将文件按要求自动重命名。

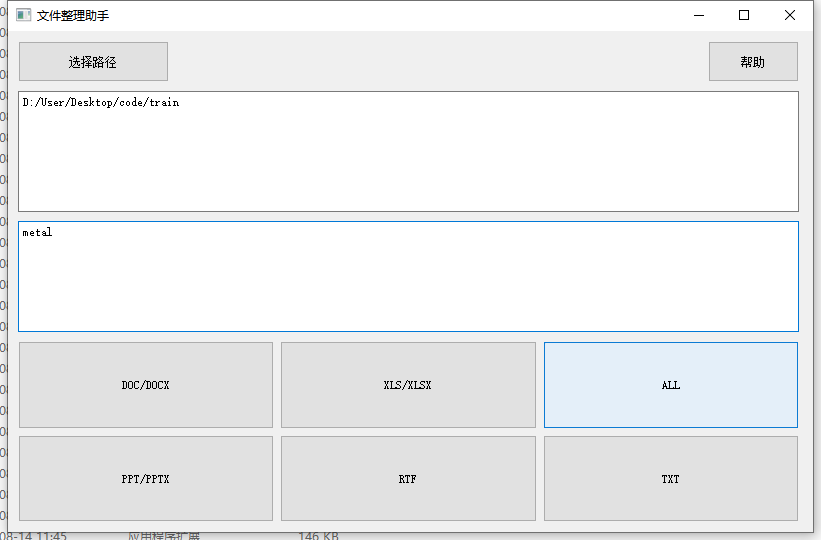


图 2.2 自制重命名软件界面图

文件命名遵循以下规律：

1. 第一个字符是图片所代表的真实类别
2. 第二个字符是“.”，以便后续分割文件名，从而得知真实数字标签
3. 后续字符为从0开始的序号，如上图2-1所示。

测试集的处理也类似。

### 2.2.2 图像预处理

由于在上一步的操作中，图像被命名为“cardboard.1”一类的格式。但计算机对字符串型的数据处理的效率不如对整型数据处理的效率高，故在此需要对六个类别做一个0-1的映射，以此来提高计算机的处理效率，降低程序编写的难度。类别映射表如表2.1所示。

表 2.1 类别映射表

|  |  |
| --- | --- |
| 类别 | 映射 |
| cardboard | 0 |
| glass | 1 |
| metal | 2 |
| paper | 3 |
| plastic | 4 |
| trash | 5 |

在设置映射表后，便可开始读取数据并进行数据预处理。为了方便后续的神经网络对图像的辨识，我们将图片同意处理为100\*100的大小，并使用img\_to\_array函数将图像转化为数组形式。最终我们获得两个列表，分别为标签列表和图像列表，他们之间是一一对应的关系。最后对图像进行归一化处理，可以增强模型的泛化能力。



图 2.3 数据集图像预处理

## 2.3 本章小结

本章主要讲述了数据集的具体来源和数据集预处理的方法。通过对图像标签的提取和对图像进行数组化、向量化，从而使得图像可以直接运用到神经网络算法中。

# 第三章 建模分析

## 3.1 人工神经网络

人工神经网络（Artificial Neural Network ANN）即神经网络（Neural Network NN)，它是对人脑组织结构和运行机制的某种抽象模型，具有单层和多层之分。每层包含若干人工神经元，通过对已知信息的反复学习训练，逐步调整神经元连接的权重值，从而达到信息处理的功能[5]。类比于生物神经元结构，神经元是构成神经网络的基本单元。

### 3.1.1 神经元

神经网络可以看成是以神经元为节点，用有向加权弧连接起来的有向图。神经元接受一组输入信号并产生输出，并使用激活函数输出一个活性值。

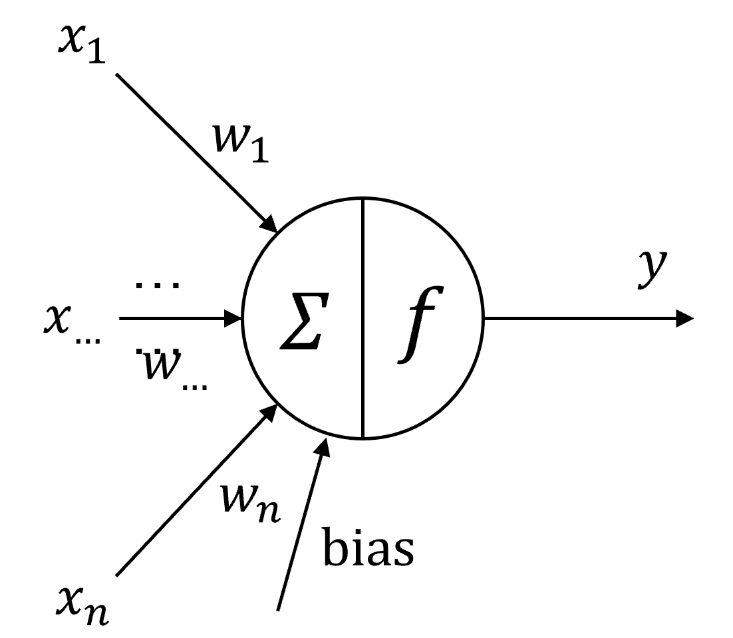


图 3.1 神经元模型

如图3.1所示，单个神经元接受*n*个输入，它表示一开始的数据来源或前一层网络的输出。则输出*y*的如公示3.1计算得到：

|  |  |
| --- | --- |
|  | （3.1） |

其中为*n*维权重向量，*b*表示偏置值，*f*表示激活函数，激活函数是一种非线性阈值过滤函数，由于每个神经单元的输出都是输入的线性组合，如果没有激活函数进行非线性过滤操作，那么无论网络层数有多深，本质上都是输入的线性组合，与单层神经网络的效果无异，失去了深度神经网络的意义。激活函数的选择随着研究的深入也在不断变化中，目前常用的激活函数共有四种:

1. *sigmoid*函数，其函数的表达式如公示3.2所示，

|  |  |
| --- | --- |
|  | （3.2） |

*sigmoid*函数，一般只用在二元分类的输出层，因为二元分类一般要求输出结果，而该函数刚好阈值就在区间内。其它层更加建议用其他的激活函数，一个神经网络可以使用多种激活函数，用*g[i]*表示第*i*层的激活函数。

1. *tanh*函数，其函数表达式如公示3.3所示，

|  |  |
| --- | --- |
|  | （3.3） |

函数可以看作放大且平移的*sigmoid*函数，上下界限分别是1和-1。它相比于表现更好，因为它的均值在0附近，有数据中心化的效果，神经网络在进行训练的时候更加方便和快速。

由于其特性，早期的神经网络模型广泛应用了*sigmoid*函数和函数。但是和 有一个共同的缺点，就是当z很大或很小的时候，它们的斜率就会趋向于0，这回使得梯度下降的学习速率降低。

1. *Relu* (The Rectified Linear Unit)，其表达式如公式3.4所示,

|  |  |
| --- | --- |
|  | （3.4） |

*ReLu*是目前应用最广泛的激活函数，使用它可以极大地加快网络训练速度。相较于传统的*sigmoid*和tanh函数，*ReLu*的优点在于它不会随着输入值的变化而趋向饱和，从而减轻了反向传播过程中梯度弥散的问题。*ReLu*函数的另一个优点是计算速度快。正向传播过程中, *sigmoid*和函数运算时需要计算指数，而*ReLu*函数仅需要判断输入值是否大于零，加快了正向传播的计算速度。

*ReLu*函数也存在一些不足，当非常大的梯度更新完参数之后，可能会将梯度永久设置为零，即不会对数据有激活现象了

### 3.1.2 神经网络模型

神经网络将许多个单一的人工神经元像链路一样联结在一起。神经网络根据神经元的连接方式，可以分为前馈神经网络(Feedforward Neural Networks)、反馈神经网络（Feedback Neural Networks）和自组织神经网络（ Self-Organizing Neural Networks）三大类。

1. 前馈神经网络

前馈神经网络又称为前向神经网络。神经元分层排列，中间层神经元只与前一层的神经元相连，接收前一层的输出，并输出给下一层。这种网络只有在反向传播训练过程中信息会反馈传播，而在检测过程中的数据只能向前传输，从输入层经过中间隐藏层直接到达输出层，前层与后层之间没有反馈机制。这是目前研究、应用最广泛的网络。

常见的前馈神经网络包括感知器网络、BP网络、RBF网络。一般多层结构的前馈网络由输入层、隐藏层和输出层三部分组成，网络结构如图3.2所示:

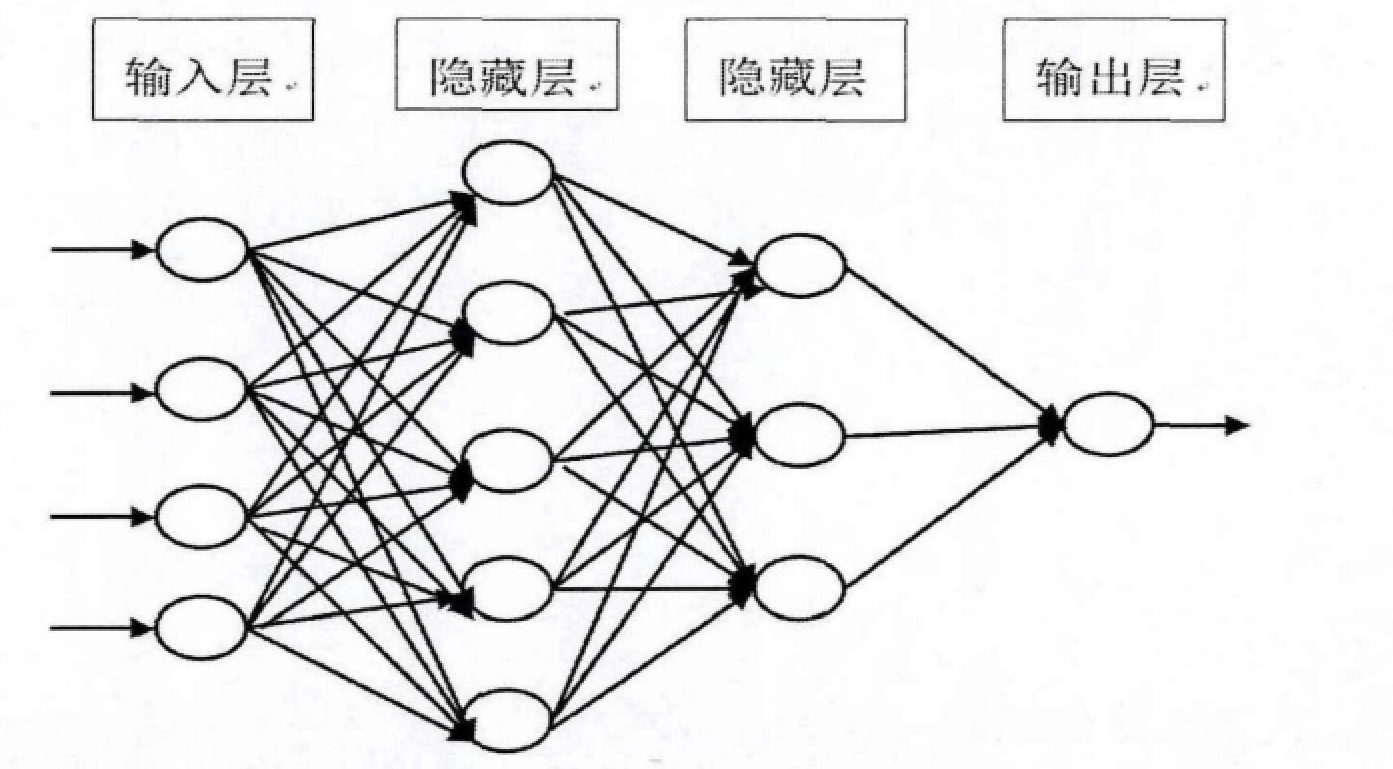


图 3.2 多层前馈神经网络模型

输入层:网络的输入主要为固定大小的非线性输入向量。

隐藏层:是输入和输出层之间的众多神经单元组成的各个层面，一个神经网络可以有很多的隐藏层。由于训练算法和运算能力的提高，隐藏层可以达到几十乃至几百层，一般来说，隐藏层越多，神经网络的非线性特性越显著，特征提取能力越强。于此同时，权重值参数量越大，训练难度提升，容易出现过拟合的问题。

输出层:是输入层向量经过众多隐藏层运算得出的结果。一般是神经网络经过训练后得出的结果，用于提供给后面进行实际应用。

1. 反馈神经网络

反馈型神经网络是一种具有输出反馈给输入机制的神经网络，其“反馈”体现在当前的结果会作为一个输入，影响到下一次的结果，即当前的结果是受到先前所有的结果的影响的。

只从系统设计的观点看，前馈神经网络模型由于不考虑输出与输入在时间上的滞后效应,只表达输出与输入在网络运算中的映射关系,缺少网络全局稳定性。而反馈神经网络采用了反馈动力学的思想,具有比前馈神经网络更强的计算能力,但是训练过程复杂，要求的计算能力更加苛刻。循环神经网络（Recurrent Neural Networks，RNNs)是当前最常见的反馈神经网络，RNN在自然语言处理(Nature Language Processing，NLP）领域取得了巨大的成功，也是当前深度学习的主流研究方向之一。

## 3.2 卷积神经网络

卷积神经网络(Convolutional Neural Networks，CNN)是一类具有深度结构且包含卷积层运算的前馈神经网络。它在上文介绍的普通前馈神经网络的基础上增加了卷积和池化操作的层级结构。上世纪八十年代由 Yann Lecun发明并运用于数字识别，取得了不错的效果。近年来，CNN广泛应用于计算机视觉、推荐系统以及自然语言处理等领域[6]。

### 3.2.1 CNN理论

卷积是数学分析领域中的一种重要运算，假设有二维离散函数*f(x,y)*和*g(x,y)*，那么它们的卷积运算定义如公式3.5所示,

|  |  |
| --- | --- |
|  | （3.5） |

卷积应用于图像处理中，是用一个模板（卷积核）和一幅图像中的像素值进行卷积运算。对于图像上的每一个点，让模板的原点和该点重合，模板上的点和图像上对应的点相乘，然后将各点的积相加，就得到该点的卷积值。给定图像和卷积核，其中*m<<M*，*n<<N*。其卷积的输出公式如公式3.5所示，

|  |  |
| --- | --- |
|  | （3.6） |

卷积是一种积分运算，用来求两个曲线重叠区域面积。可以看作加权求和, 可以用来消除噪声、特征增强。其本质把一个点的像素值用它周围的点的像素值的加权平均代替。

卷积神经网络具有三个特性:局部感知、权值共享和下采样。这三个特性能够保证卷积神经网络既能有效地提取出图像的特征,又能大大减少网络参数的数量。

1. 局部感知

社会科学的一种普遍观点是，人类对外界的认知是由局部到整体的。人脑神经元在处理图像信息时，首先感知的是局部的图像，然后大脑高级神经组织才将各种局部图像信息综合整理起来，得到全面的图片感知。生物学中人类的视觉系统结构，位于视觉皮层的神经元用来感受局部信息，只响应来自某些特定局部位置的刺激。CNN的设计思想，也是基于图像信息的空间联系。即局部图像空间的信息联系紧密，而相对距离远的图片信息相关性变弱。

局部感知模型的特点在于每个隐含单元相当于视觉皮层的神经元,只连接输入单元的部分区域，这部分区域称为神经元的感受野，每个隐含单元只负责处理其感受野的数据。在卷积神经网络中，感受野的大小即卷积核的大小。如图3.3所示，全连接层的参数数量为1000\* 1000\*1M, 而采用局部感知模型的参数量下降为10\*10\*1M，降低4个数量级。

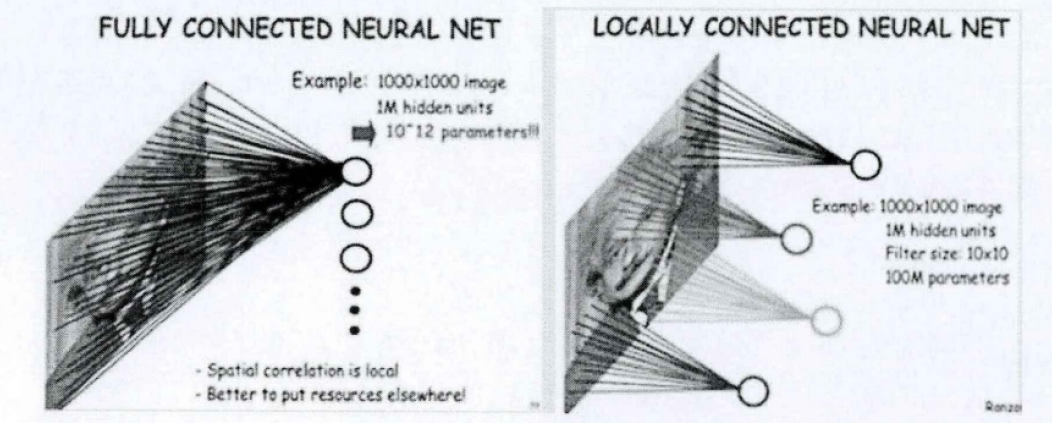


图 3.3 全连接模型和局部感知模型

1. 权值共享

图像信息的底层特征与特征在图片中的位置无关。比如说图像的边缘信息，无论是在图片中间的边缘特征，还是在图片边角处的边缘特征,都可以用类似的特征提取器获取。这就给卷积神经网络中的权值共享特性提供了理论基础。

在卷积神经网络中，权值共享本质上是滤波器（即卷积核）共享。一种特征可以出现在图像的任何地方，用同一种滤波器去滑动，就可以提取出特定的特征。在低层的卷积层操作中，每一种过滤器遍历整个图像时，过滤器的参数是固定不变的，即相当于整个图像的所有元素共享了相同的权值。每个卷积层具有多个特征图，每个特征图通过一种卷积滤波器提取局部感受野的特征信息，其次是每个特征图中由多个神经元,即卷积层中每个滤波器都有其对应关注的图像特征，比如垂直与水平边缘特征、颜色特征、凹凸特征等等。最后，将这些神经元整合起来，就能够得到整张图像的特征集合。这样既能够保证提取到有效的图像特征，又能进一步降低网络权重参数的数量。

### 3.2.2 深度卷积神经网络（AlexNet）

AlexNet网络是深度卷积神经网络的起源。在2012年的ImageNet竞赛图像分类任务中,参赛人员采用AlexNet 网络结构取得了最好的成绩，这也使得卷积神经网络在图像分类中取得了阶段性成功[7]。本文使用AlexNet神经网络对垃圾图像进行分类。

AlexNet包含8层变换，其中5层卷积和2层全连击隐藏层，以及1个全连接输出层。其中第一层中的卷积窗口形状是11×11。因为ImageNet中绝大多数图像的高度和宽度比MNIST手写数字数据集的图像大10倍以上，大的图像会占用更多的像素，所以需要更大的卷积窗口来捕获物体。第二层中的卷积窗口形状为3×3。此外，第一、第二和第五个卷积层之后都是用了窗口形状为3×3、步幅为2的最大池化层，紧接着最后一个卷积层的是两个输出个数为4096的全连接层。另外，AlexNet使用较为简单的ReLu作为激活函数，使得模型的计算更加简单、更为灵活。

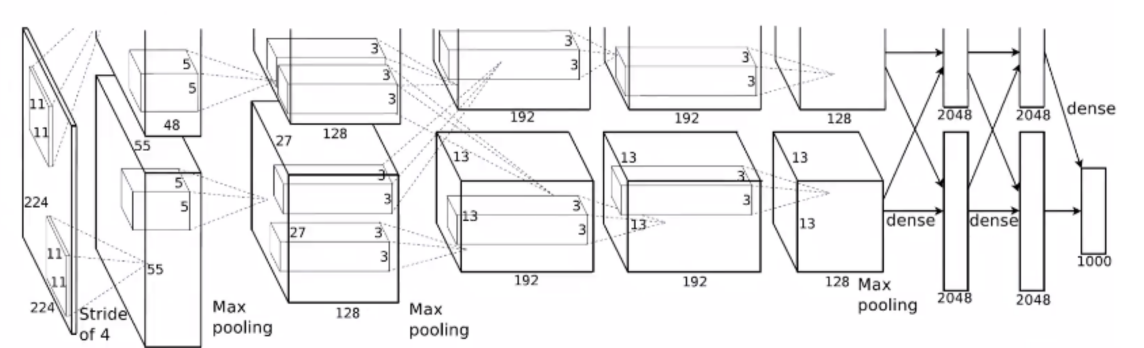


图 3.4 AlexNet网络结构图

## 3.3 本章小结

本章主要内容为建模分析，从单一的神经元到深度卷积神经网络进行了初步的阐述，是搭建卷积神经网络的基础知识。下一章将会重点介绍AlexNet卷积神经网络图像分类识别的设计。

# 第四章 基于AlexNet卷积神经网络图像分类识别的设计

## 4.1 AlexNet卷积神经网络的构建

根据3.2.2节中提到的深度卷积神经网络模型，构建神经网络。通过tensorflow和keras的学习库，我们可以快速搭建卷积神经网络。本文的神经网络结构如下：

卷积层>标准化+激活层(Prelu)>卷积层>标准化+激活层>最大值池化层>>

卷积层>标准化+激活层(Prelu)>卷积层>标准化+激活层>最大值池化层>>

卷积层>标准化+激活层(Prelu)>卷积层>标准化+激活层>最大值池化层>>

卷积层>标准化+激活层(Prelu)>卷积层>标准化+激活层>均值池化层>>

Dropout层>线性层>激活层

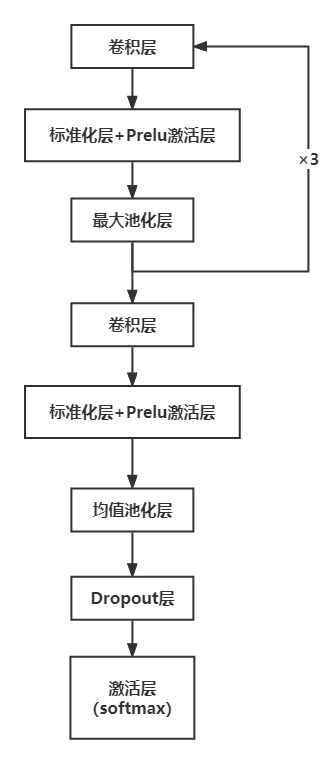


图 4.1 AlexNet卷积神经网络结构图

主要网络层为卷积层，池化层，全连接层

**卷积层:**（包含有卷积核、卷积层参数、激励函数）：使用卷积核进行特征提取和特征映射。

**池化层：**

在卷积层进行特征提取后，输出的特征图会被传递至池化层进行特征选择和信息过滤。池化层包含预设定的池化函数，其功能是将特征图中单个点的结果替换为其相邻区域的特征图统计量。池化层选取池化区域与卷积核扫描特征图步骤相同，由池化大小、步长和填充控制。

**全连接层：**

全连接层位于卷积神经网络隐含层的最后部分，并只向其它全连接层传递信号。特征图在全连接层中会失去空间拓扑结构，被展开为向量并通过激励函数。全连接层的作用则是对提取的特征进行非线性组合以得到输出，即全连接层本身不被期望具有特征提取能力，而是试图利用现有的高阶特征完成学习目标。

## 4.2 卷积神经网络参数的选择

1) 卷积核大小的选择

从理论上来说，卷积核的大小会影响最终模型的训练效果，但是实际上没有一种科学的方法来确定哪种大小的卷积核是最合适的[8]。但是随着卷积核的不断增大，带来的参数量和计算量是巨大的。由于本文选取的数据集与ImageNet的数据集有一点的差距，本文数据集图像的大小相对较小，同时在课程设计中不需要非常高的识别精度，为了加快训练速度，在本文搭建的卷积网络中，所选择卷积核的大小均为3×3。

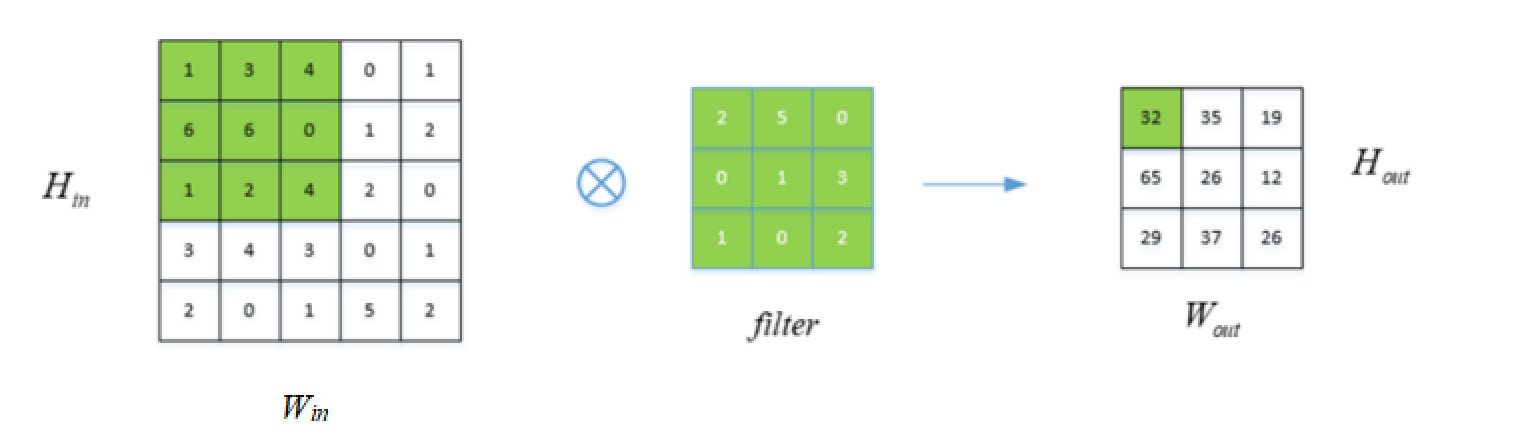


图 4.2 卷积示意图

2) 池化层的选择

池化本质上是下采样，通常使用的由均值池化和最大池化。在本文的卷积神经网络中，采用了4个池化层，前三个池化层为最大池化，最后一个为均值池化。池化层的大小均为2×2。

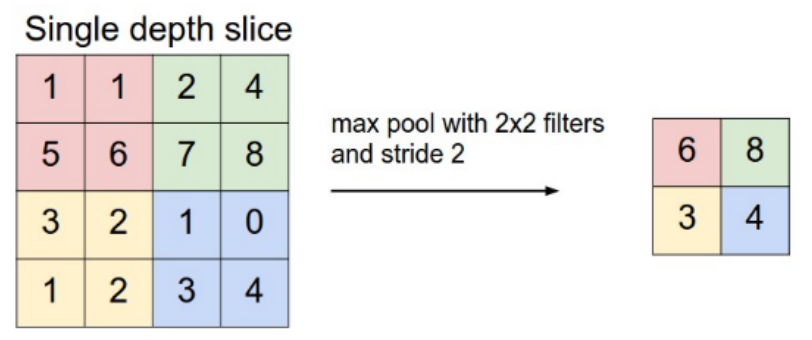


图 4.3 最大池化示意图

1. 激活层函数选择

使用激活函数主要是为了对卷积操作之后的数据进行滤波，形成非线性的映射，滤除一定的噪声。在第三章中曾经提到过，常用的激活函数有*Sigmoid*函数、*Relu*函数、*maxout*函数等。由于AlexNet网络相对于LeNet的改进之一就是使用了更为简单的激活函数*Relu*，而本文中的卷积神经网络是基于AlexNet网络简化得来，故在本文中依旧选取*Relu*作为激活函数。

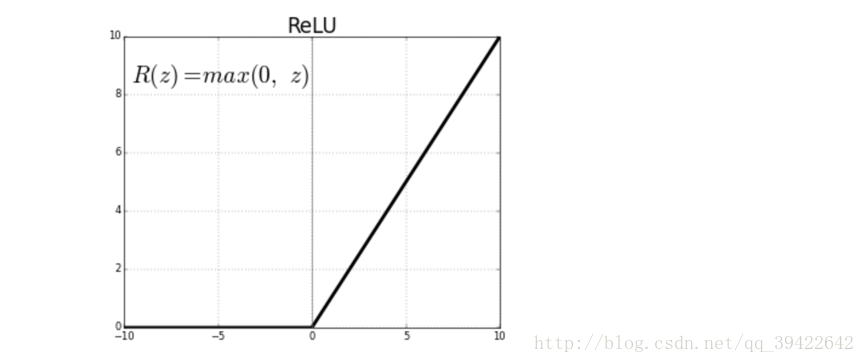


图 4.4 激活函数*Relu*示意图

1. Dropout层参数选择

Dropout的作用是，在模型训练时，随机冻结部分隐含层，使其权重不工作，不工作的节点可以暂时不被看做网络的一部分，但是它们的权值是被保留的，在下一次输入样本时，这些被冻结的节点可能会被重新启用。在训练卷积神经网络是，如果训练样本相对较少，为了防止模型过拟合，可以加入Dropout层。

我们可以通过设置Dropout层的参数dropout\_ratio，来控制冻结节点的多少。dropout\_ratio越大舍弃的信息越多，loss下降越慢，准确率增加越慢。经查阅资料和实践，将dropout\_ratio设置为0.5。

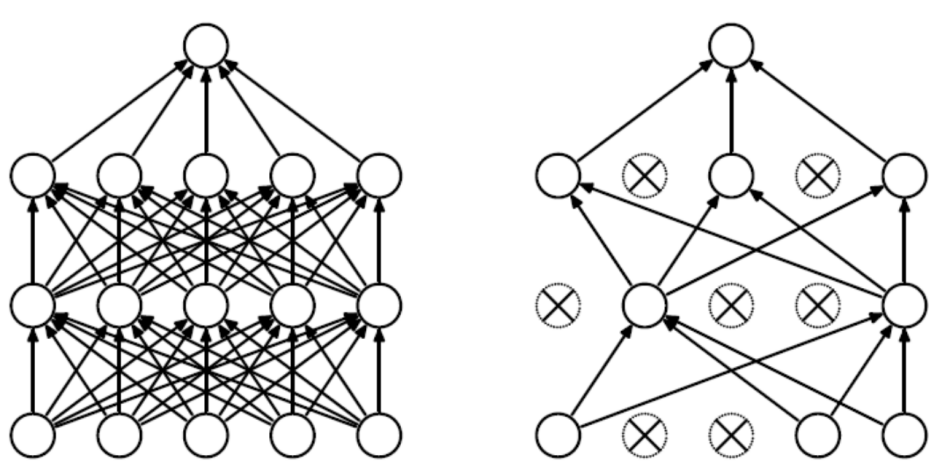


图 4.5 Dropout层作用示意图

## 4.3 本章小结

本章主要介绍了AlexNet卷积神经网络的构建以及卷积神经网络参数的选择。具体参数了卷积神经网络各层的作用，说明了每个参数选择的依据。下一章将重点介绍AlexNet卷积神经网络的仿真实现与模型评估。

# 第五章 仿真实现与模型评估

## 5.1 基于AlexNet卷积神经网络的图像分类识别的仿真实现

根据第四章卷积神经网络的设计，将神经网络搭建完成后，可以按照一下步骤进行训练。

1.初始化。包括以下的几部分。

1) 将数据集导入，并进行训练集和测试集的划分，在此选择0.3，即训练集占70%，测试集占30%；

2) 对训练集和测试集的数据进行预处理，将图片尺寸统一为100×100；

3) 对训练集的图片进行图片增强，随机加入旋转、宽度偏移、高度便宜、缩放等操作，从而提升模型的泛化能力；

4) 设置神经网络参数，包括迭代次数，卷积核大小，池化层大小等。

2.开始训练模型，并记录每一次训练模型的准确率，方便用于后续的模型评估；

3.保存准确率最高的模型。理想状态下，随着迭代次数的增加，模型的准确率会不断提高，但当迭代次数增加到一定程度，模型的准确率可能不会一直保持最优值，而是在一定范围内震荡，故在训练过程中需要实时保存当前准确率最高的模型。

4.使用准确率、Loss等指标评估AlexNet卷积神经网络模型。

AlexNet卷积神经网络的算法流程图及具体的网络训练过程如下图所示：

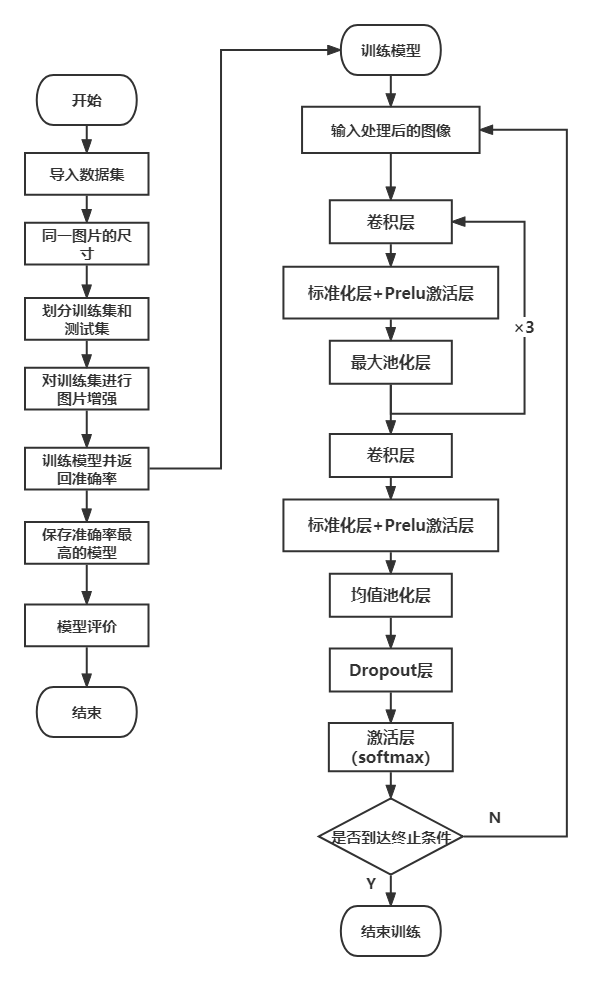


图 5.1 AlexNet卷积神经网络训练流程图

## 5.2 模型评估

根据每一次迭代返回的模型准确率和Loss值，使用numpy库可以绘制出卷积神经网络准确率及误差。由于硬件资源有限，仅进行了100次迭代。测试集和训练集的acc曲线和loss曲线如下图所示：

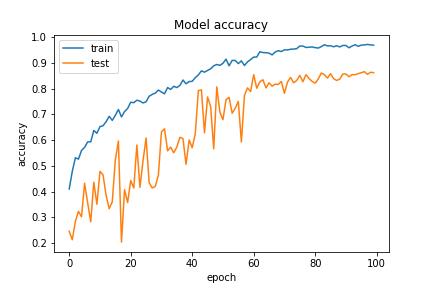


图 5.2 AlexNet卷积神经网络网络的准确率

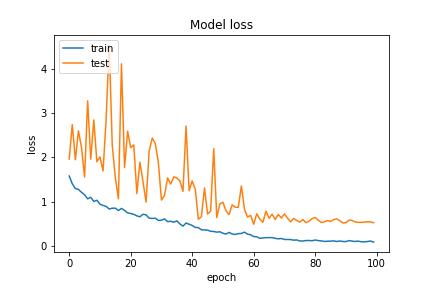


图 5.3 AlexNet卷积神经网络网络的误差

由图5.2、图5.3可以看出，随着迭代次数的增加，训练集的准确率平稳升高，误差平稳下降，效果较为理想，最终的准确率能达到95%左右。而测试集的准确率和误差在前期震荡比较大，但随着迭代次数超过60次，测试集的误差和准确率趋于平稳，最终准确率能达到80%左右。由此可以认为，该模型取得了较好的识别效果。

接下来导入自行拍摄的图片进行辨识。在自行拍摄的30张照片中，大约有80%能够被正确识别，基本符合预期效果。

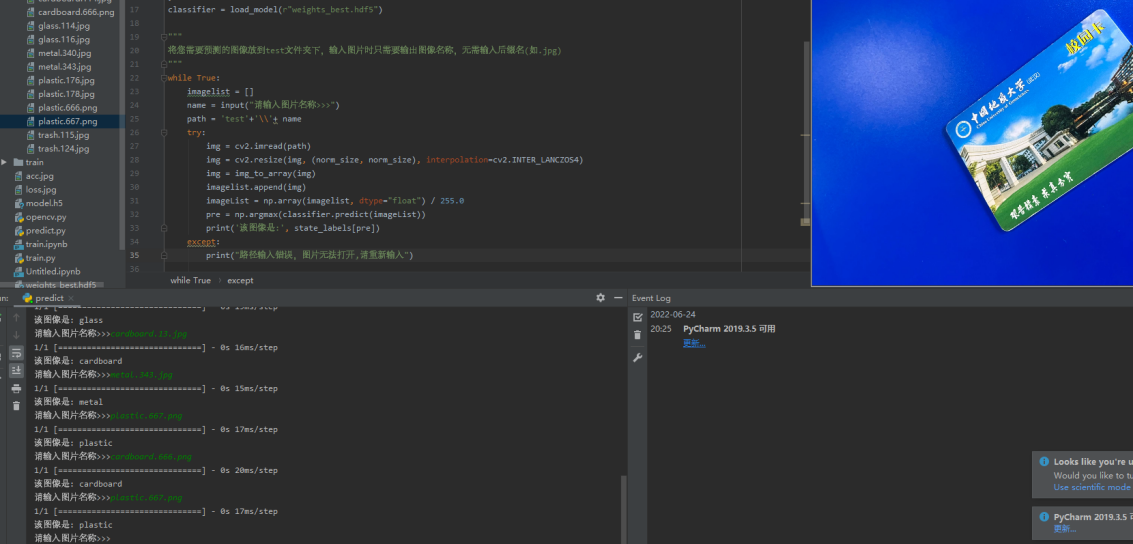


图 5.4 对自行拍摄的图片进行分类

## 5.3 本章小结

本章主要对AlexNet卷积神经网络的仿真实现和模型评估进行了重点介绍，同时利用在和鲸社区、百度飞浆社区收集的数据集对卷积神经网络进行了训练。最终识别的准确率能达到95%左右。下一章主要介绍利用外接摄像头实现实时的垃圾分类，为实际应用做初步的探索。

# 第六章 利用外接红外摄像头实现垃圾实时分类

显然利用AlexNet卷积神经网络只完成对测试集和上传图片的识别无法应用到实际生活中去，因为不可能将生活中的每个图像都上传到系统文件中再进行识别，如银行的支票、信件上的邮政编码以及生活中各种来不及上传的图像等。因此需要可以对图像分类识别进行实时识别的研究，使其能够进一步应用到生产生活中。本章利用自制垃圾图像数据集训练出来的神经网络作为实时识别的模型，通过OpenCV调用外接USB摄像头实现手写数字的实时识别。

## 6.1 OpenCV简介

OpenCV是一个基于Apache2.0许可（开源）发行的跨平台[计算机视觉](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%A7%86%E8%A7%89/2803351" \t "_blank)和[机器学习](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/217599" \t "_blank)软件库，可以运行在[Linux](https://baike.baidu.com/item/Linux/27050" \t "_blank)、[Windows](https://baike.baidu.com/item/Windows/165458" \t "_blank)、[Android](https://baike.baidu.com/item/Android/60243" \t "_blank)和[Mac OS](https://baike.baidu.com/item/Mac%20OS/2840867" \t "_blank)操作系统上。它轻量级而且高效——由一系列 C 函数和少量 C++ 类构成，同时提供了Python、Ruby、MATLAB等语言的接口，实现了[图像处理](https://baike.baidu.com/item/%E5%9B%BE%E5%83%8F%E5%A4%84%E7%90%86/294902" \t "_blank)和计算机视觉方面的很多通用算法。OpenCV用C++语言编写，它具有C ++，[Python](https://baike.baidu.com/item/Python/407313" \t "_blank)，[Java](https://baike.baidu.com/item/Java/85979" \t "_blank)和[MATLAB](https://baike.baidu.com/item/MATLAB/263035" \t "_blank)接口，并支持Windows，Linux，Android和Mac OS，OpenCV主要倾向于实时视觉应用，并在可用时利用MMX和SSE指令， 如今也提供对于C#、Ch、Ruby，GO的支持。近年来OpenCV越来越广泛地被应用到各个领域，如人机互动、物体识别、图像分割、人脸识别、机器人等等[9]。本章将利用OpenCV库调用外接摄像头结合前面实现的算法进一步探究垃圾图像分类的实时识别。

## 6.2 基于AlexNet卷积神经网络的垃圾图像识别

将训练过程中保存的最优模型导入，按照一下步骤具体实现垃圾的实时分类。

1.使用轮询的方式查看外接摄像头在拿一个通道，一般在通道1；

2.建立如表2.1的类别映射关系；

3.将训练好的神经网络导入程序，利用OpenCV调用外接摄像头，并实时读取图像；

4.将外接摄像头的图像resize为100×100的大小，放入训练好的神经网络模型中进行识别；

5.不断重复该过程，实现实时识别。

为了方便调试，该部分内容使用Jupyter Notebook编写，图6.1、图6.2为AlexNet神经网络实时识别的实例。



图 6.1 手持摄像头实时识别废旧金属

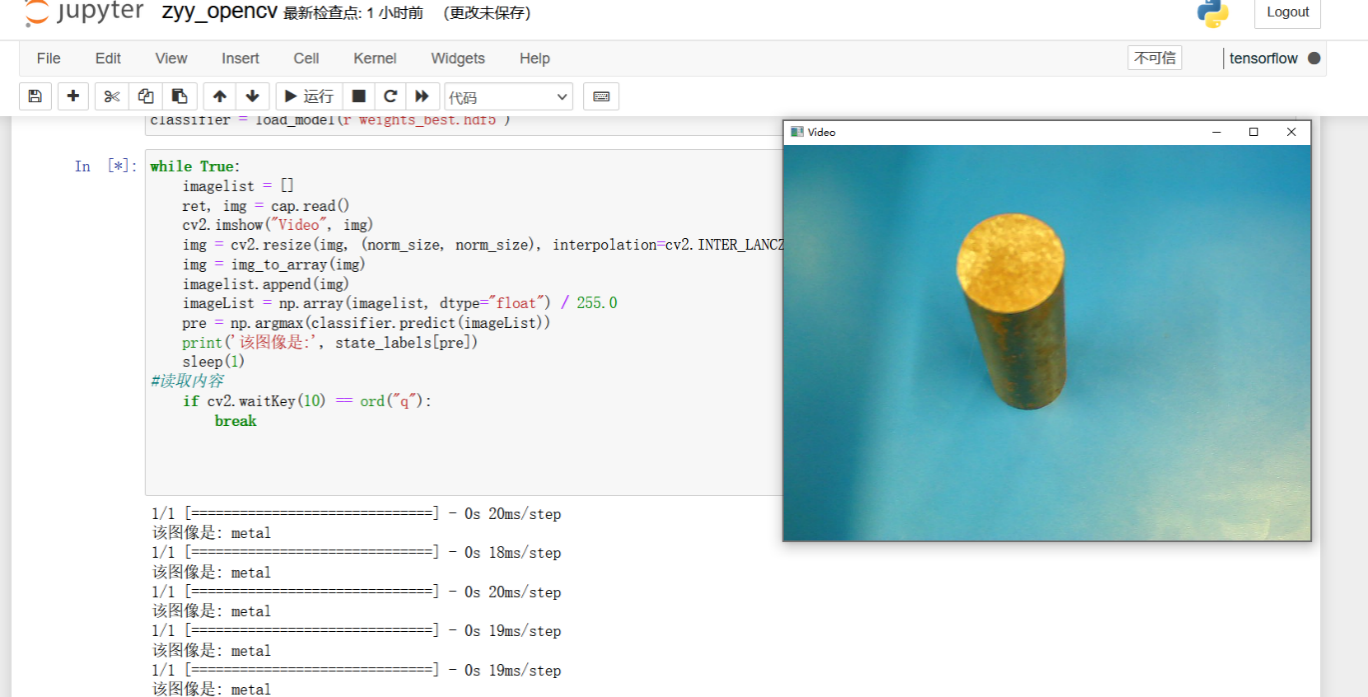


图 6.2 实时识别废旧金属界面图

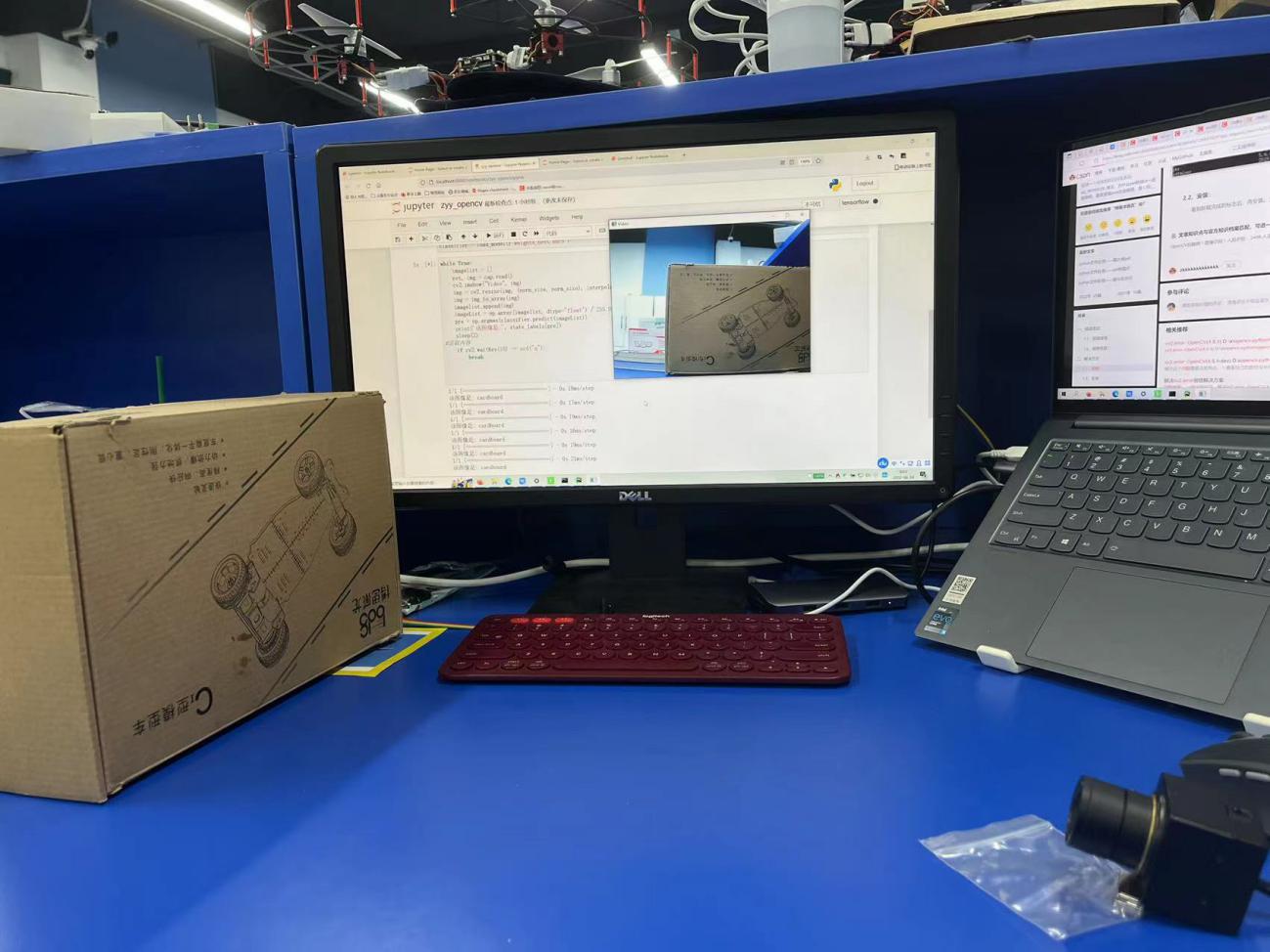


图 6.3 简易的识别系统

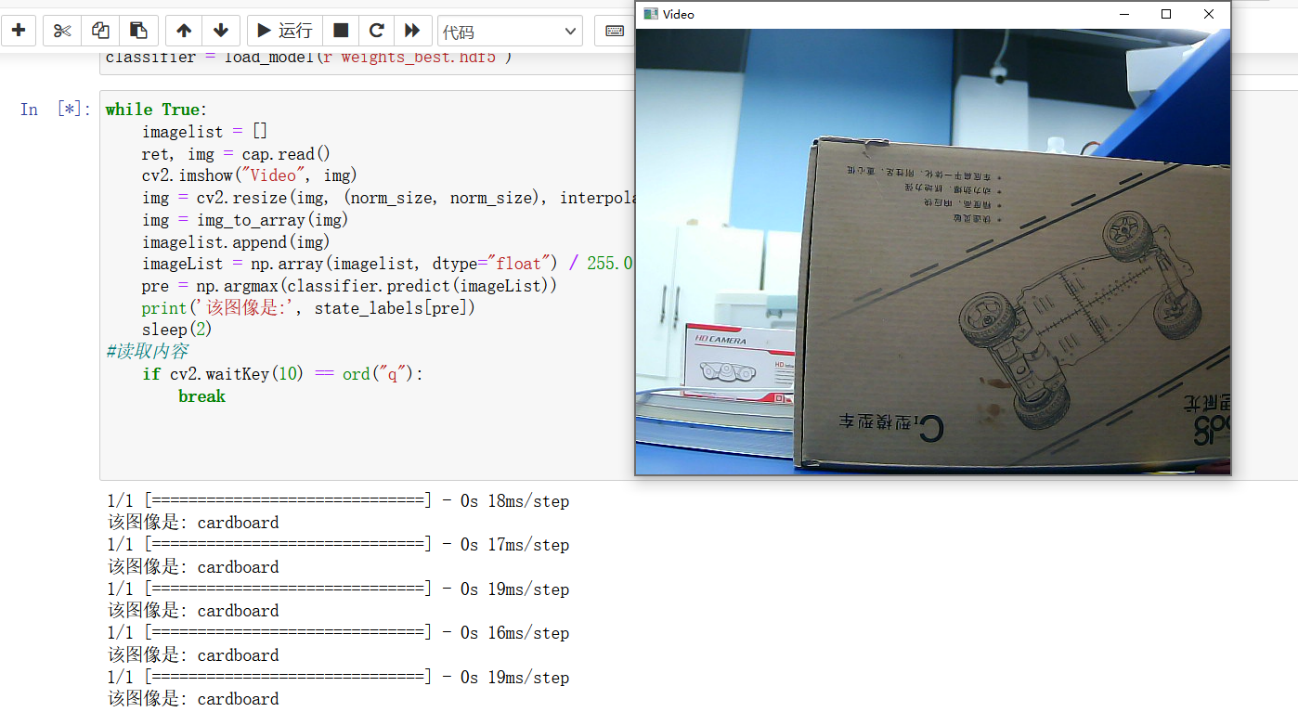


图 6.4 实时识别废旧纸箱界面图

由于AlexNet神经网络在进行垃圾分类实时识别之前已经完成了神经网络的训练，只需要将图像信息放入到神经网络中得到结果即可，因此其识别速度非常快，由图6.2和图6.4可以看出，在20毫秒内可以输出分类结果，基本上可以摄像头的图像只要变化，便可以识别出来。同时实时识别的准确率也有保证。

## 6.3 AlexNet神经网络实时识别的识别效果

从算法的实时识别情况来看，训练过后的神经网络能对垃圾类别进行实时识别，并且由于前期训练的识别率很高，能够达到96.3%，因此实时识别的准确率也很高。粗略统计下，实时识别的准确率在80%以上。

但目前搭建的简易系统还有致命的缺陷，例如不能在图像中标明所识别物体的位置，如图6.5所示，图中有黄色、白色两个纸箱，神经识别的结果也为纸箱。但是神经网络识别的纸箱具体是哪一个呢？我们不得而知。



图 6.5 摄像头实时传回的图像

在今后可以对实时识别的算法进行改进，例如在将图像送入神经网络进行判别之前，先通过图像二值化、大津阈值、边缘检测等方法，将需要识别的部分先框选出来，再送入神经网络进行识别。如图6.6所示，是去年学习OpenCV时做的一个边缘检测，但使用的语言是C++，由于期末复习时间紧迫未在本文中实现。



图 6.5 通过边缘检测提取数字并框选

## 6.4 本章小结

本章通过利用OpenCV库来调用外接摄像头。并实现基于AlexNet卷积神经网络的垃圾实时分类，效果良好。分析了识别过程中算法的缺陷，并提出改进方法，以便后续进行进一步的研究。

# 第七章 结论

本文主要通过图像分类识别这一主题来对《模式识别与机器学习》这一课程的知识进行巩固和延伸。在选择课程设计选题时，曾经有考虑过使用BP神经网络做数字识别，但是由于BP神经网络是全连接的方式，不适合做更多其他图像的识别。为了能够实现更多种类图像的识别，我继续了解卷积神经网络的相关知识。在对比LeNet和AlexNet的区别后，选择表现更加优秀的AlexNet卷积神经网络进行设计，最终训练集的辨识准确率达到96.3%，测试集的准确率达到82%。最后通过训练好的模型，调用OpenCV库实现了垃圾实时分类，效果良好。

但实现过程中仍有许多不足：

1. 没有进行多种网络效果的对比，比如使用LeNet网络和AlexNet网络同时进行训练，并对比两个模型训练效果。也可以使用BP神经网络进行识别，用来对比全连接和局部连接神经网络的区别，体现出局部连接的优势。由于时间原因，这些都没有实现，我将会在后续继续进行研究。

2. 本次设计中，网络的搭建采用了tensorflow库和keras库，虽然说学习这两个库的基本使用方法也花费了大量的时间，但是相对于不使用深度学习库而是自己手动搭建神经网络，难度还是大有区别。在后续时间充裕后，将尝试手动搭建卷积神经网络。

3. 实时垃圾分类中，可能会出现图像内容混杂的情况，对于内容混杂的图像，神经网络模型也会给出一个预测的结果，这一结果有可能是误识别，也有可能是在内容混杂的图像中识别到了对应的物体。对此可以使用OpenCV图像处理的技术，对图像进行边缘检测等操作，将物体框选出来后再送入神经网络进行识别。

上述的诸多不足之处也是后续研究需要解决的问题，对于该课程的内容也需要不断的复习巩固。《模式识别与机器学习》这门课程的学习，激发了我对模式识别与机器学习、神经网络、模糊控制、专家系统等模式识别与机器学习方法的兴趣。在后续时间充裕之后，我将会继续在这个领域进行更深入的研究。

# 致谢

对于本文我首先要感谢的是陆承达老师，在他的指导下我学习到了许多模式识别与机器学习的方法，并将其用于实践；同时，我还要在完成设计任务时给予我指导的学长，他给我讲授了一些Python-OpenCV的知识，并将他的外接摄像头借我使用；此外还需要感谢学校，给了我一个良好的学习平台，可以自由使用一些文献平台和下载端口；最后还要感谢我的家人们，是他们一直在给予我支持和鼓励。

# 参考文献

1. 张秀芳,龚小梅,占小洋,夏康顺,梁晓伟,宾晓蓓.垃圾分类之自动识别技术探索[J].科技创新与应用,2021(01):178-180+184.
2. 徐锐,陈敏谊,刘美红,郑丽莎,杨雯.人工智能在垃圾分类中的应用[J].品牌与标准化,2022(01):107-109.
3. 李望晨. BP神经网络改进及其在手写数字识别中的应用[D].哈尔滨工业大学,2006.
4. 张家怡.图像识别的技术现状和发展趋势[J].电脑知识与技术,2010,6(21):6045-6046.
5. 张驰,郭媛,黎明.人工神经网络模型发展及应用综述[J].计算机工程与应用,2021,57(11):57-69.
6. 张珂,冯晓晗,郭玉荣,苏昱坤,赵凯,赵振兵,马占宇,丁巧林.图像分类的深度卷积神经网络模型综述[J].中国图象图形学报,2021,26(10):2305-2325.
7. 季长清,高志勇,秦静,汪祖民.基于卷积神经网络的图像分类算法综述[J].计算机应用,2022,42(04):1044-1049.
8. 吕响,张书玉,宋英楠,王长忠,李雯琦,牛嘉瑞.基于深度学习下的卷积神经网络参数学习[J].渤海大学学报(自然科学版), 2021,42(04):369-375
9. 杨思源. 基于OPENCV的车辆牌照识别系统研究[D].西安电子科技大学,2013.

# 小作业

### 简述模式识别系统的主要组成部分和设计步骤

1）数据采集

显然，数据的采集是进行生物识别验证的前提条件。一个性能良好的生物特征识别系统一定需要首先捕获到好的生物特征数据。利用这些数据，我们就可以进行后续的预处理、特征提取、特征选择等工作。一般来说，这里的数据采集肯定需要借助相应的硬件设备，诸如，声音传感器、图像传感器等等。如果传感器的灵敏度不高，或者传感器的精确度不高，那么势必会对所采集到的数据产生一定的噪声污染。这样一来，尽管可以通过后续的预处理来减弱甚至消除一部分噪声，但是，终究无法做到完全去除噪声的干扰。所以，数据采集部分应该尽量保证所得到的数据纯正、干净。通常我们可以采集相当数量的数据，并从中选择最优、最好、最具有代表性的数据来作为原始的输入。这样，就从源头上保证了数据取样对最终生物识别验证系统的干扰最小。

另外，需要注意的是，针对不同的生物特征，数据采集的方法和原理是不同的。比如，虹膜识别系统是通过分析人的虹膜表面不同区域不同的纹理分布来进行分类辨别的。而掌纹识别，则是基于人的手掌脉络的不同分叉、线条的粗细等特征为依据来进行最终的识别。笔迹识别则是利用了不同的人在签名时笔划的长度、角度、偏移，握笔的力度、书写时的速度，加速度等特征来进行区分的。

（2）预处理

在基于统计方法的生物识别技术领域，所谓的预处理一般是指去除噪声的干扰，加强有效信息的过程。前面已经提到，原始数据的采集不可避免的要引入一些噪声的干扰，对于一个实际的生物识别系统而言，预处理是一个必要的环节。但是，需要注意的是，虽说预处理的作用都是减弱甚至消除噪声的干扰，同时增强有用信息的强度，不过，针对不同的特征，预处理的方法也是千差万别。

（3）特征提取

一般来说，从传感器得到的数据属于原始测量空间的数据，而原始测量空间的数据是无法直接进行判别分类的，或者说，直接利用原始测量空间得到的数据进行判别分类往往达不到期望的效果。通常来说，我们需要将数据从原始的测量空间“变换”到二次空间，而这个二次空间，研究人员一般将它称为特征空间。将数据从原始空间变换到特征空间后，我们就得到了表征某模式的二次特征，一般我们所指的特征就是这里所谓的二次特征。

（4）特征选择

数据从原始空间变换到特征空间后，得到了二次特征，而这里的二次特征不一定就是我们需要的特征，或者，变换得到的二次特征对最终的分类判别不一定有很好的作用。而特征选择的作用就是从若干二次特征中挑选对模式的最终分类判别最有帮助或者分类效果最后的一部分二次特征。不难发现，这里的特征选择，实际上是一个寻优问题。

（5）分类判别

在得到了对分类最有帮助的二次特征之后，最终需要做的就是对二次特征进行分类判别。一般而言，简单点的分类判别方法是将特征表示为向量的形式，然后，利用诸如欧氏距离、马氏距离、卡方距离、范数等手段来计算任意两个向量之间的相似性，之后，根据计算得到的向量的相似性来进行分类判别。复杂点的分类判别方法是利用SVM、ANN、CART、SNOW等方法将特征进行某种线性或者非线性的变换，从而在变换后的空间进行二次分类。

### 对“模式识别与机器学习”应用的认识

机器学习主要解决以下四类问题：

1.监督学习：指的是训练的数据既包括特征（feature）又包括标签（label），通过训练，让机器可以自己找到特征和标签之间的联系，在面对只有特征没有标签的数据时，可以判断出标签。监督学习主要分为两类，分别为回归问题（Regression）与分类问题（Classification）。回归问题的目标是通过对已有数据的训练拟合出恰当的函数模型，分类问题的目标是通过分析数据的特征向量与对应类别标签的关系，对于一个新的特征向量得到其类别。两者的区别是回归针对连续数据，分类针对离散数据。

2.非监督学习：指的是在未加标签的数据中，找到隐藏的结构，由于提供给学习者的实例是未标记的，因此没有错误信号（损失）来评估潜在的解决方案。典型的非监督学习类型包括聚类（Cluster）、隐马尔可夫模型、使用特征提取的技术降维（主成分分析）。

3.半监督学习：所给的数据有的是有标签的，而有的是没有标签的，试图利用大量的未标记示例来辅助对少量有标记示例的学习，常见的两种半监督的学习方式是直推学习（Transductive learning）和归纳学习（Inductive learning）。

4.强化学习（Reinforcement learning）：指的是机器以“试错”的方式进行学习，通过与环境交互获得奖赏指导行为，目标是使机器获得最大的奖赏。强化学习中由环境提供的强化信号对产生动作的好坏作评价，而不是告诉机器如何去产生正确的动作。

模式识别从十九世纪五十年代兴起，在二十世纪七八十年代风靡一时，是信息科学和人工智能的重要组成部分，主要被应用于图像分析与处理、语音识别、声音分类、通信、计算机辅助诊断、数据挖掘等方面。模式识别系统过程：特征提取与选择;训练学习;分类识别。例如人类见到一个东西之后，通常就会下意识地给其归类：是动物还是植物，属于哪一门纲目属科，是否可以药用，有果实吗，花朵是否漂亮，是否有毒……这一大串归类构成了人们对于这种事物的整体认知。这就属于人类对于模式的识别，这种技能对于人们甚至是一些动物来说，是非常简单而且几乎是与生俱来的。

模式识别与机器学习的区别：

1、发展趋势不同

从发展史上来讲，这对孪生兄弟都是人工智能历史中辉煌一时的流派。其中模式识别可以归入人工智能领域的元老级别。尽管模式识别看起来很高大上，而且也有了较长时间的应用，但是其效果似乎总是差强人意。似乎已经有些过气的征兆，正在慢慢没落消亡。而机器学习则是人工智能领域最基础也是人气最高的天王级选手。

2、应用范围不同

机器学习目前是在狭义的人工智能领域走得比较快，但是广度还是模式识广。模式识别在很多经典领域，如信号处理，计算机图像与计算机视觉，自然语言分析等都不断有新发展。

3、判断重点不同

机器学习根据样本训练模型，如训练好的神经网络是一个针对特定分类问题的模型;重点在于“学习”，训练模型的过程就是学习;机器学习的落脚点是思考。机器学习侧重于在特征不明确的情况下，用某种具有普适性的算法给定分类规则。而机器学习的概念可以类比聚类分析(聚类本身就是一种典型的机器学习方法)，对“类”的严格定义尚不明确，更谈不上检验。

模式识别是根据已有的特征，通过参数或者非参数的方法给定模型中的参数，从而达到判别目的的。学过多元统计的可以这样理解：模式识别的概念可以类比判别分析，是确定的，可检验的，有统计背景的(或者更进一步说有机理性基础理论背景)。