# 目 录

[目 录 1](#_Toc31935265)

[1 绪论 3](#_Toc31935266)

[1.1 研究背景 3](#_Toc31935267)

[1.2 国内外研究现状 5](#_Toc31935268)

[1.3 主要研究内容及内容安排 6](#_Toc31935269)

[2 相关技术概述 7](#_Toc31935270)

[2.1卷积神经网络的核心思想 7](#_Toc31935271)

[2.1.1深度学习介绍 7](#_Toc31935272)

[2.1.2 卷积神经网络 7](#_Toc31935273)

[2.1 经典卷积神经网络模型 8](#_Toc31935274)

[2.1.1 LeNet-5模型 9](#_Toc31935275)

[2.2.2 AlexNet模型 9](#_Toc31935276)

[2.1.2 Inception-v3模型 10](#_Toc31935277)

[2.2 基于卷积神经网络的迁移学习 11](#_Toc31935278)

[2.2.1 迁移学习 11](#_Toc31935279)

[2.2.2 卷积神经网络迁移学习 11](#_Toc31935280)

[2.3 图像识别介绍 12](#_Toc31935281)

[2.3.1 图像识别及其难点 12](#_Toc31935282)

[2.3.2 图像数据处理 12](#_Toc31935283)

[2.4 本章小结 12](#_Toc31935284)

[3 数据获取及预处理 13](#_Toc31935285)

[3.1 数据挖掘目标 13](#_Toc31935286)

[3.2 数据抽象及介绍 13](#_Toc31935287)

[3.3 数据增强 14](#_Toc31935288)

[3.3.1 图像随机裁剪 14](#_Toc31935289)

[3.3.2 图像随机旋转 15](#_Toc31935290)

[3.3.3 图像颜色调整 15](#_Toc31935291)

[3.4 本章小结 16](#_Toc31935292)

[4 基于卷积神经网络的图像分类 17](#_Toc31935293)

[4.1 问题的分析 17](#_Toc31935294)

[4.2 卷积网络模型 17](#_Toc31935295)

[4.3 实验环境及参数设置 17](#_Toc31935296)

[4.4 实验结果分析 17](#_Toc31935297)

[4.5 本章小结 17](#_Toc31935298)

[5 总结与展望 17](#_Toc31935299)

[5.1 全文总结 17](#_Toc31935300)

[5.2 研究展望 17](#_Toc31935301)

[致谢 17](#_Toc31935302)

[参考文献 17](#_Toc31935303)

# 1 绪论

## 1.1 研究背景

中国服饰产业的历史源远流长，作为我国传统优势产业之一，其在我国经济发展中占有举足轻重的地位。随着社会生产水平的提高，人们在服饰相关的消费支出在不断增长。根据国家统计局数据，2011-2017年全国服装类商品零售额逐年增长，但增速逐年放缓。2017年全国服装类商品零售额达到10365.4亿元，同比增长1.4%;2018年全国服装类商品零售额为9870.4亿元，同比下降4.8%，服装商品零售额首次出现负增长。截至2019年上半年，服装类商品零售额达到4749.7亿元。2011年起服饰类商品零售额及增长情况数据如图1-1所示:

随着移动互联网的崛起，传统的服饰销售不断地转移到线上，其规模和速度也前所未有当前，我国电子商务市场的容量巨大，目前比较有名的电商平台有京东商城、苏宁易购、拼多多、唯品会、网易严选、淘宝商城等。每年的双十一电商节，都是经销商去库存的好机会，也是消费者疯狂购物的盛会。2019年的双十一全天销售额达到了4101亿元,其中天猫占了65.50%的份额；京东、拼多多、苏宁易购占了28.20%；其他的一些电商占了6.30%的份额。具体数据点图1-2：

中国电商对服饰行业的整体发展有着举足轻重的作用，服饰电商市场的高速发展已是不争的事实，以服装为例，据国家统计局的数据显示2015年以来，网购的消费习惯推动服饰电商高速发展，2015年达到了约4306.4亿元，2019年更是达到了约10133亿元，并且这个数据还在逐年的增加中，具体数据见图1-3：

服饰电商行业的发展为我们生活带来方便的同时，也对传统的搜索提出了新的要求:对图片中物体的识别进行搜索，其中应用领域较为广阔的一类问题就是服装种类识别，即通过输入图片，对其中包含的服装的种类做出识别分类。传统的识别分类技术主要借助数字图像处理、模式识别的方法，通过对图像检测分割、特征提取分类识别等操作。然而上面提及的特征多数是人为规定的特征，并不能在所有的场景下都很好地刻画待识别物体的属性，而且往往需要较大的计算量，不适合在便携式设备上进行计算。

随着深度学习概念的提出，开始通过神经网络来模拟人类大脑多层抽象的机制来对现实的对象进行抽象表达，主要包含文本、语音和图像等形式的数据，而且对数据特征的抽象过程尽可能少地减少人为干预。实践已经证明深度学习方法在识别以及分配问题上取得了很高的准确率。常见的深度学习模型有卷积神经网络、受限波尔兹曼机和堆栈式自编码器，本文正是利用卷积神经网络，结合Softmax回归模型，提出了对服装图片进行分类的方法。

## 1.2 国内外研究现状

在使用深度学习算法之前，词袋模型是在图像识别时使用较多的方法。词袋模型是自然语言处理和信息检索下的简化表达模型。在此模型中，可以用包含这些单词的包来表示一段文本，而不管语法和单词顺序如何。对于图像，单词袋模型需要构建字典。

基于深度学习的图像分类方法可以使用无监督或受监督的方法来学习分层特征描述，而不是手动设计或选择图像特征。近年来，卷积神经网络在深度学习模型中的应用在图像领域取得了令人瞩目的成就。CNN直接使用图像像素信息作为输入，并最大程度地保留输入图像的所有信息。通过卷积运算提取特征，并且模型的输出可以直接用作图像识别的结果。

为了检验图像识别技术的水平，ILSVRC每年都会举办，ILSVRC作为全球最大的图像识别大赛，它一直是衡量图像分类任务的基准，基本上代表了计算机视觉领域的最高水平。自2012年以来，几乎每年基于深度学习模型的图像分类准确性都取得了重大突破，如图1-4所示。从2012年到2015年，随着深度学习技术的发展，Imagenet图像分类的错误率每年下降4％。随着模型结构的深化，深度学习识别模型前5名的错误率逐渐降低。目前，它已经超过了人眼的识别能力，已经降低到3.5％左右。在相同的数据集上，人眼识别的误识率约为5.1％。

1998年，Yann-Lecun完成的LeNet-5模型有7层，是最早的神经网络模型。它包含了深度学习的基本模块：卷积层、池层、全连接层。同时，它被认为是一种非常有效的用于手写体识别的卷积神经网络，然而，由于当时计算机水平的限制，相关研究者并不使用更复杂的CNN模型，而是更倾向于使用支持向量机。

2012年，Alex以15.3％的错误率在ILSVRC比赛中获胜，使CNN取得了历史性的进步，标志着深度学习的兴起。2014年，Google团队采用GoogLeNet模型，以6.67％的错误率夺得冠军。2015年，微软团队使用Resnet模型以3.57％的错误率夺得冠军，该模型共有152层。从以上数据可以看出，神经网络模型的层次和精准度在不断提高，在现实生活中的应用也越来越广泛。卷积神经网络甚至已经成为图像识别的主流技术。随着图像识别研究的不断深入，许多卷积神经网络算法已经被应用到实际业务中。

## 1.3 主要研究内容及内容安排

本文主要研究在应用卷积神经网络基础下所实现的图像分类方法，在现有的图像处理技术中，通过学习应用卷积神经网络等方法深层学习理论，再结合迁移学习技术，利用改进的模型进行研究学习。本文各章的主要内容如下：

第一章介绍了本文的主要内容、内容安排、国内外研究现状、研究背景。

第二章介绍了卷积神经网络、迁移学习、图像识别等方法的背景和知识。

第三章介绍了利用网络爬虫技术，从京东网站上获取相关的图像。

第四章介绍了使用迁移学习的方法调整神经网络模型，提取和识别图像的特征。

第五章总结和展望，对本文研究的主要内容进行总计，同时提出本文存在的问题与不足，并对未来的研究方向进行了展望。

# 相关技术概述

## 2.1卷积神经网络的核心思想

### 2.1.1深度学习介绍

深度学习通过学习样本数据的表示层次和内在规律，建立多层神经网络，用于解释语音、文本、图片、视频的算法集合。深度学习的核心是自动进行特征学习，进而学习和抽取特征信息，加快特征工程的进行，从而改善人工特征提取的重大难题。

HOG、SURF等特征提取算法能够对一定程度内的图像缩放、视觉改变、平移、旋转、亮度调整等操作具有不变性，然而依旧不能达到较低的错误率。但是卷积神经网络可以做到很好的自动识别特征，并且可以和分类训练同时处理。由于深度学习算法往往具有多层结构，因此它具有较强的特征表达能力和学习能力，尤其是在处理复杂的上下文信息和全局特征提取的时候，浅层模型的效果远远不如深度学习算法的效果好。深度学习算法能够取代传统算法的重要原因之一是它能够省去人工设计特征阶段的工作，因为它可以使用多层次的非线性映射方法，从海量数据中自动学习对象的特征。深度学习算法还可以识别出图片中的隐含因素，而人工提取的方法却很难辨别，因为这些因素往往以很复杂的非线性方式相互关联，而深度学习算法可以通过大量的训练，将这些相互关联的因素分开，从而使分类变得精准和简单。

由此我们可以得知，深度学习有很多优势：首先，由于精巧的模型设计，深度学习可以通过并行计算来处理大量数据；其次，深度学习不要用通过人工来提取特征，一个网络结构可以同时进行特征选择与分类；最后，深度学习在图像分类上的准确率要远远高于传统分类方法。

### 2.1.2 卷积神经网络

卷积神经网络经常被用于图像处理，它可以省去很多人工预处理的过程，直接处理原始图像，在许多计算机视觉任务中被用作为核心方案。卷积神经网络包括输入层、卷积层、激活层、池化层、全连接层以及输出层，具有权值共享、有效的图像分类、局部连接、极强的特征提取等优点。从根本上来说，卷积神经网络通过卷积层来进行特征提取，通过池化层来减少参数，最后通过传统神经网络来完成分类等任务。

（1）卷积层是卷积神经网络的核心层，由一组滤波器组成。通过从图像中

随机选出训练样本中学习特征，并且将这些特征与原图像做卷积运算，可以得到原图像中任意位置上的不同特征的激活值。通过卷积层的计算，可以将当前层的一个字矩阵转化为下一层的一个单位节点矩阵。卷积运算过程如图2-1所示：

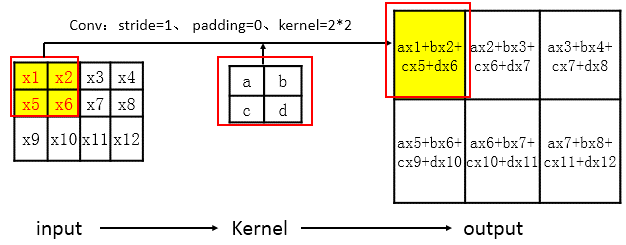


图2-1 卷积运算操作

（2）池化层同时也叫做下采样层，在连续的的卷积层中间加入池化层，可以减少数据和参数的数量，使得模型加快收敛，起到减少过拟合的作用。池化层的具体操作和卷积层的操作基本相同，一般最常见的是平均池化层与最大池化层，即取对应位置的平均值或者最大值。最大池化层的操作如图2-2所示：

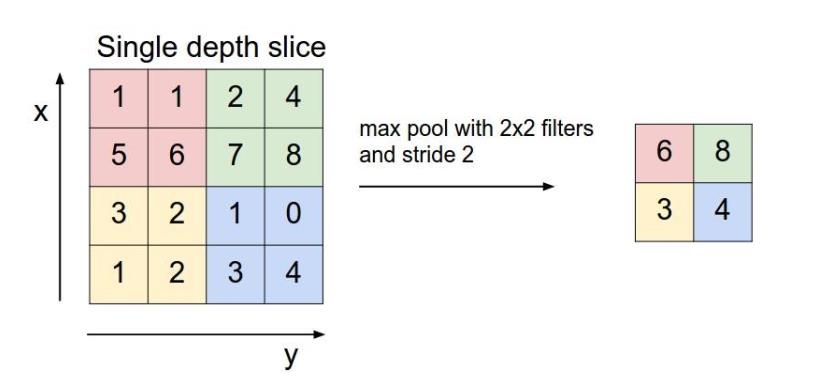


图2-2 最大池化操作

（3）全连接层主要起到“分类器”的作用，在卷积神经网络中，全连接层出现在最后几层，连接前面处理过的特征信息，将结果输出给分类器。

（4）Softmax层用来实现最后的分类问题，最终得出样例属于不同种类的概率。

## 经典卷积神经网络模型

经典的神经网络模型有LeNet-5、AlexNet、VGG-16、GoogLeNet、ResNet等模型，经过大量研究人员对深度学习不断探索，模型的效果越来越好，甚至已经超过人眼的识别能力，模型的具体信息如图标2-3所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | AlexNet | GoogLexNet | VGG | ResNet |
| 年代 | 2012 | 2014 | 2014 | 2015 |
| Top-5错误率 | 16.40% | 6.70% | 7.30% | 3.57% |
| Data Augmentation | + | + | + | + |
| Inception | - | + | - | - |
| 模型层数 | 8 | 22 | 19 | 152 |
| 全连接层数 | 3 | 1 | 3 | 1 |
| 全连接结构 | 4096,4096,1000 | 1000 | 4096,4096,1000 | 1000 |
| 卷积层数 | 5 | 211 | 16 | 151 |
| 卷积核大小 | 11,5,3 | 7,1,3,5 | 3 | 7,1,3,5 |
| Dropout | + | + | + | + |
| Batch Normalization | - | - | - | + |
| LRN | + | + | - | - |

通过表2-2可以看出，在准确率提高的同时，模型的层数越来越多，复杂度也越来越搞。

### LeNet-5模型

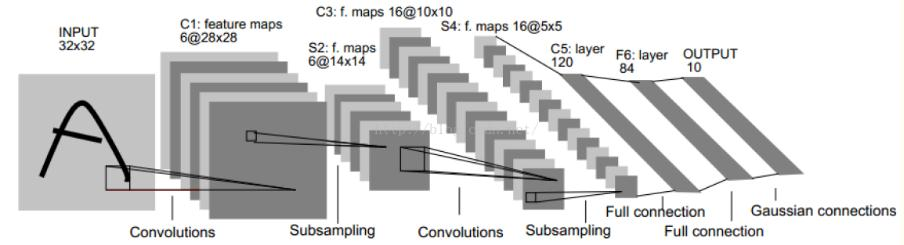
LeNet-5模型是由Yann-Lecun所提出的一个经典的神经网络模型，主要用于手写体的识别。LeNet-5模型共有7层，包括2个卷积层，2个池化层，3个全连接层，LeNet-5模型可以很好的学习训练样本的特征信息，但是当训练样本大规模缺乏标签的时候，不适合使用该模型。其具体结构如图2-3所示：

图2-3 LeNet-5模型

### 2.2.2 AlexNet模型

AlexNet模型的层数要多于LeNet-5模型，包括5个卷积层，3个全连接层和1个softmax层，运用了LRN、Dropout、Data Augumentation等方法解决过拟合的问题，使用2块GPU加快模型的运算速度，使用收敛速度更快的的Relu激活函数取代Sigmoid函数。因此在2012年ILSVRC大赛上取得冠军，并远超第二名，AlexNet模型具体结构如图2-4所示：

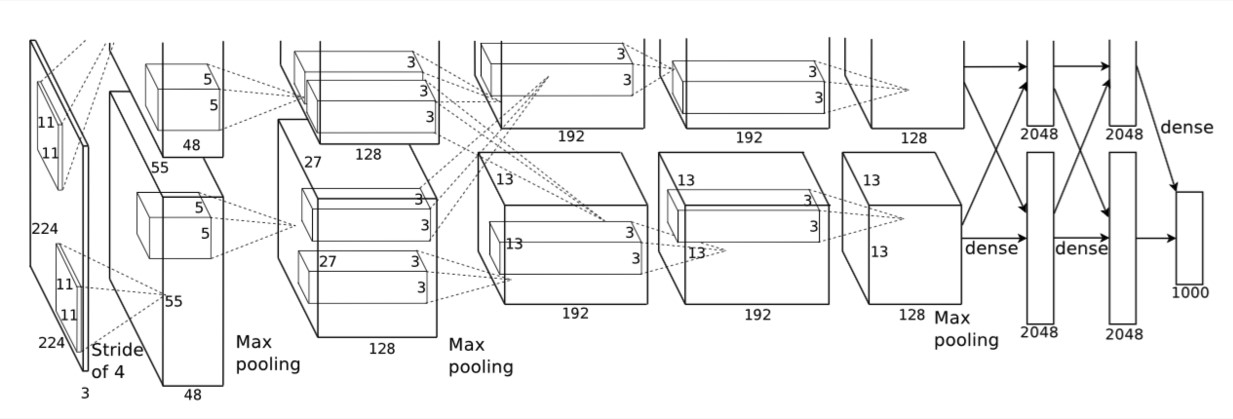


图2-4 AlexNet模型结构

### Inception-v3模型

在GoogLeNet被提出之前，网络层数的加深是网络性能突破的主流选择，但是伴随着网络层数的增加，神经元的数目也大大增加，导致网络的参数过多，复杂性变高，容易过拟合的问题的出现，网络加深，梯度越可能消失，模型也越来难以优化，为解决这些问题，Inception-v3模型采用了更小的核，采用Asymmetric方式，并且将全连接转化为系数连接。Inception模型结构如图2-4所示：

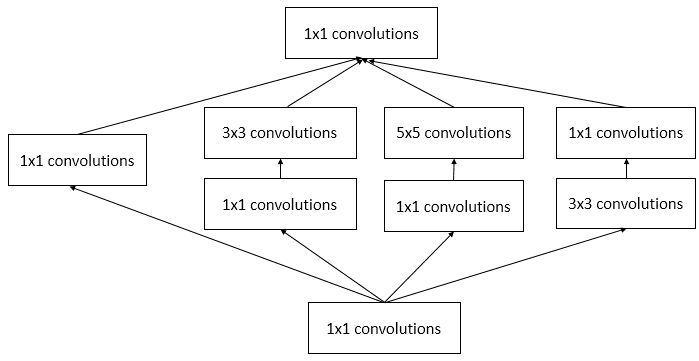


图2-4 Inception 结构

从图2-4中可以看出，Inception-v3模型与AlexNet模型的结构完全不同，GoogLeNet中设计了Inception模块，该模块可以关联不同大小的卷积层。Inception-v3模型在增加特征表达能力的同时减少运算，通过增加BN层稳定训练。为了减少参数，Inception-v3使用了更小的核，即将一个5x5的滤波器换成两个3x3的滤波器，将一个3x3的卷积核换成2个1x3的卷积核。GoogLeNet模型由11个inception模块构成，共96个卷积层，Inception-v3模型结构如图2-5所示：

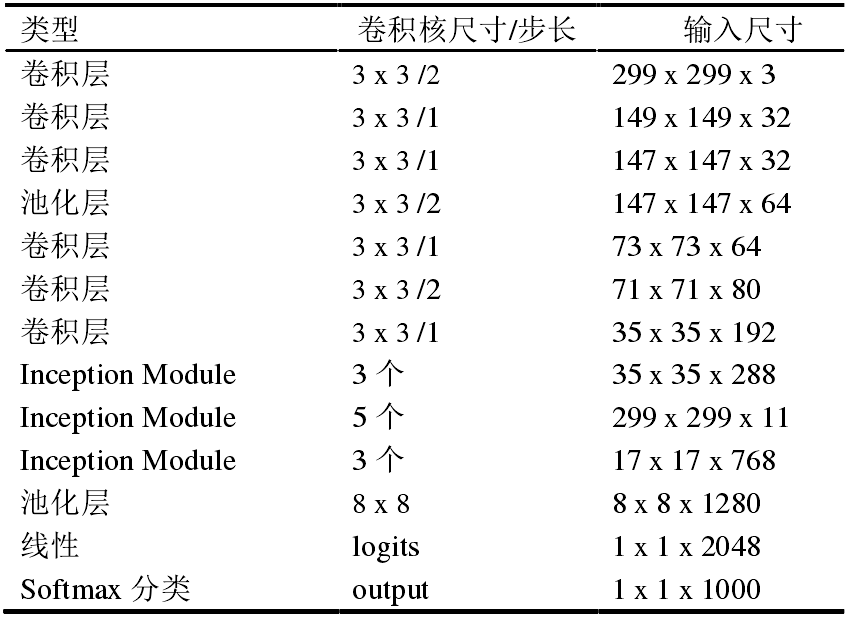


图2-5 Inception-v3结构

## 基于卷积神经网络的迁移学习

### 迁移学习

大部分的数据或者任务存在相关性，我们可以基于参数、样本、问题相关性、特征等数据进行研究学习。迁移学习就是利用这一特性调整和修改训练后的模型参数，使其能够适应新的问题，帮助新的模型进行训练，加快并优化新的模型的学习，而不用从零开始重新学习。但是迁移学习在应用领域并不成熟，它所需要的条件要求也未形成一套正统的体系。

机器学习模型的更新与训练过于依赖数据的标注，尽管我们可以获取到大量的数据，但是很少有数据能够被正确的标注，为了解决这类问题，查找与目标数据接近的标记数据，使用该数据构建模型，并为目标数据添加标签。当数据集过小，模型没有更多的数据来提高它的鲁棒性和泛化性，我们可以使用迁移学习的方法。

### 卷积神经网络迁移学习

卷积神经网络在图像识别、目标检测、语音识别等领域得到了成功的应用，并且发挥这良好的性能。但是，从头开始训练复杂的卷积神经网络模型需要大量标记的数据集以及大量的计算资源，有时可能需要几天或几周的时间。

由于数据标注最耗费时间和精力，复杂的卷积神经网络模型只有在大量的标注数据的基础下才能达到比较理想的效果。在实际生活中，研究院通常采用微调方法和迁移学习方法，将已经在大型数据集上进行修改、调整，应用与新的问题上。而且实验表明，效果很好。当我们的数据集规模不大的时候，可以使用在类似领域受过训练的模型。本文介绍了基于Inception-v3的迁移学习方法在服装图像分类中的应用。

在本文中，使用Tensorflow将训练好的Inception-v3模型迁移到服装图像数据集。通过替换模型的最后一层，使用瓶颈层的输出重新训练新的全连接层，以解决商品服装的分类问题。因为ImageNet有1000个分类，因此需要将原来模型的Softmax分类器改变给本文所需要的输出类型数量，然后再训练。

## 图像识别介绍

### 图像识别及其难点

图像识别是指对图像进行特征提取，将输入图像与存储图像对比，已识别不同类型的对象的技术。要做到图像识别，需要做图像分割、物体检测、物体识别和定位等工作，一般情况下需要对图像进行预处理操作，例如标准化、预处理、翻转、调整色彩等，然后训练识别图像所设计的模型，最终就可以识别图像。

图像的不确定性是实现图像识别的难处之一，很难直接找到图片的特征信息，例如空间关系、边界、色彩、形状、纹理，同时概念层的信息更加难以提取，精准度也难以控制。并且由于受到背景掺杂、视角变化、光照影响、扭曲变形、尺度变化、遮挡干扰等影响，图片之前的差异性较大，导致计算机更加难以理解图像。

### 图像数据处理

为了使数据集能够满足各种情况，需要对原始图像进行标准化处理，例如：边缘切割、图片清洗去噪、大小变换，进而为模型的训练提供保障。同时为了防止在训练过程中出现过拟合的情况，需要使用数据增强技术扩充数据集，比如对图像进行色彩变换、填充、随即裁剪、旋转、标注等操作。

## 本章小结

本章介绍了卷积神经网络、迁移学习和图像识别的相关理论知识和应用现状，分析了图像识别中的技术难点和常用方法，包括，图像增强技术，图像预处理方法。

# 数据获取及预处理

## 数据挖掘目标

本文首先从京东网站获取有效的服饰图像，同时把异常图片以及不符合要求的图片进行清洗剔除。由于从网站获取的照片质量都比较好，大部分是使用单反相机拍摄的。但是在实际生活中，测试的图片质量良莠不齐，光线、角度、颜色等各有不同。这样就导致了从电商获取的图片和测试的图片差异较大，这些因素极大的影响了图片识别的精准度。为了解决这个问题，通过使用图片增强技术来扩充服饰数据集，提高图片识别的精准度。

图像增强可以给一个图片创建多个不同的副本，改变图片的角度、颜色、对比度等成分，提高图片的利用率，降低模型对图像结构学习过拟合的可能性。本文使用TensorFlow实现对图像的预处理，然后将处理后的图像进行学习。

## 数据抽象及介绍

通过使用爬虫技术，从京东商城获取各种服饰的图片作为本文的原始图像。分别获取了短靴、包、外套、套衫、凉鞋、短裤、连衣裙、运动鞋、长裤、短袖共十类服饰，样本数据如图3-1所示：

（a）短靴 （b）包 （c）外套

（d）凉鞋 （e）短裤 （f）连衣裙

（g）套衫 （h）运动鞋 （i）长裤



（j）短袖

图3-1 数据集示例

为了方便读取图像，本文将数据集中的图片按分类存放在文件夹中，文件夹的名称与类别名称相同。并将训练集、测试集、验证集，按照5.3：1：3.7的比例划分。

## 数据增强

由于从京东网站所获取的原始图像无法满足我们的需求，为了使模型的识别精准度更高，本文使用图片增强数据对数据集进行扩充。通过对比度调整、随机缩放、随机翻转、颜色变换、随即裁剪等方式，生产出更多的有效图像。

### 图像随机裁剪

图像随机裁剪的方式有两种：第一种方式是在原始图像的周围填充背景，然后再进行裁剪，第二种方法是将原始图像进行微调，最大程度地保留目标信息。效果如果3-2所示：

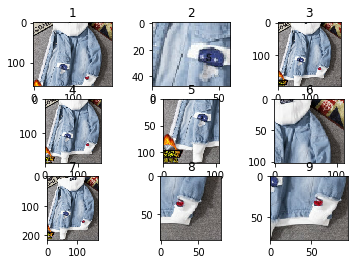


图3-2 随机裁剪效果

### 图像随机旋转

将图像上下旋转、左右旋转、对角线旋转，在有限的的样本基础上获取更多的数据，帮助模型识别不同角度的样本数据，提高模型的准确率。效果如图3-3所示：

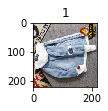
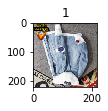
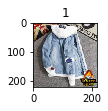


图3-3 随机旋转效果

### 图像颜色调整

通过调整图像的对比度、饱和度、亮度、色彩等因素，在不改变图片识别的结果上获取更多的图片，使模型尽量不受颜色变化的干扰。效果如图3-4，3-5，3-6，3-7所示：

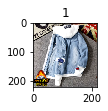
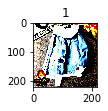
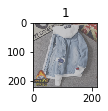


图3-4 对比度对比图

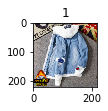
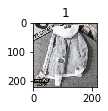


图3.5 饱和度对比图

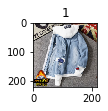
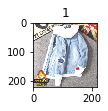
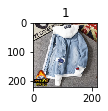


图3.6 亮度对比图

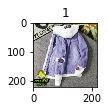
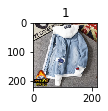


图3.7 色彩对比图

## 本章小结

本章介绍了模型所需要的数据的来源，并介绍了图像增强技术，利用色彩调整、裁剪、旋转的方式大大地扩充了数据集。

# 基于卷积神经网络的图像分类

## 问题的分析

## 卷积网络模型

## 实验环境及参数设置

## 实验结果分析

## 本章小结

# 总结与展望

## 全文总结

## 研究展望

# 致谢

# 参考文献