西南石油大学研究生学位论文知识产权声明书及

学位论文版权使用授权书

本人完全了解学校有关保护知识产权的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属于西南石油大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版。本人允许论文被查阅和借阅。学校可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时，本人保证，毕业后结合学位论文研究课题再撰写的文章一律注明作者单位为西南石油大学。

本学位论文属于

1、保密（ ），在 年解密后适用本授权书。

2、不保密（ ）

（请在以上相应括号内打“√”）

学位论文作者签名： 指导教师签名：

年 月 日 年 月 日

西南石油大学研究生学位论文独创性声明

本人声明：所呈交的研究生学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含其他人为获得西南石油大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

年 月 日

摘 要

文化自信的提出使传统服饰文化的发展得到良好的促进，如唐装、汉服及明制服装等多种形式的中国服饰文化都不断走进人们的视野，并且与现代服饰产生联动，使其呈现出更加多样的服饰风格特色，在不同的服饰结构造型、面料工艺、色彩搭配或是工业化细节等方面均表现出强烈的差异性，与此同时服饰风格的变化也增添了服饰所要表达的文化内涵的广度，在一定程度上提升了电商平台中对服装分门别类的工作量。因此，准确分类服饰图像是获取符合要求的商品以及在一定程度上节约人力成本的关键所在。

在常规多标签分类任务上，服饰图像数据有多种分类属性，同时多标签间存在着复杂的标签共现关系，同时还存在容易受到背景信息干扰的问题；基于传统的CNN模型进行多标签分类任务时会忽略掉标签之间的语义联系，使用GCN模型进行多标签分类可以对标签关系建模，但由于固定的邻接矩阵无法实时捕捉属性间非对称的依赖关系，并且由于其结构简单性，在用于小尺寸属性特征提取上有一定的缺陷。

本文研究的是把传统多标签分类模型和图注意力网络（GAT）进行结合优化提升服装图像标签分类准确度，并将多样性服饰风格通过本文算法模型应用到真实的电商系统中以实现节省人力管理成本的目的。

首先，构建标签关系邻接矩阵时要充分考虑服饰标签间的关系复杂度；结合标签共现关系统计和语义相似性分析，以量化服饰属性之间的拓扑结构优化GAT中的多头注意力层；通过对邻接节点的视觉特征与语义嵌入特征进行融合，实现模型对属性间非对称依赖关系的动态学习，从而提高模型对于复杂标签属性关系的鲁棒性和泛化能力，解决传统GCN中由于邻接关系二元化带来的语义损失问题。采用ResNet-50提取图像全局视觉特征，并将GAT产生的标签语义嵌入与该图像的全局视觉特征融合为该图的特征表示，优化基于标签共现概率的自适应损失权重分配方案，在长尾分布下抑制稀有属性的分类偏差。

综上所述，本课题是为了针对服装图像中带有复杂标签属性关系的分类准确度不高这一问题，在此基础之上提出使用聚集相邻节点视觉特征和语义嵌入特征的方法来动态学习属性间的非对称依赖关系，同时引入ResNet-50提取图像全局视觉特征，再结合GAT生成标签语义嵌入关系，并将两者结合起来实现一种多标签图像分类方法，最后通过实验测试表明该方法较于传统的图像多标签分类模型，在准确率与鲁棒性上都具有优势。

关键词：图注意力网络；图像多标签分类；邻接矩阵优化；多头注意力

Abstract

Driven by the concept of cultural confidence, traditional clothing culture has embraced new developmental opportunities. Representative ethnic styles such as Tang suits, Hanfu, and Ming-style attire have gradually entered public consciousness and integrated with modern fashion, endowing apparel styles with highly complex and diverse characteristics. These distinctions manifest in significant differences across design structures, fabric types, color schemes, and industrial details. The rapid evolution of clothing styles not only enriches the cultural connotations of fashion expression but also increases labor costs for apparel classification and management on e-commerce platforms. Consequently, accurate classification of clothing images has become a critical requirement for scenarios such as product retrieval and labor cost reduction.

Compared to conventional multi-label classification tasks, apparel images present technical challenges such as diverse attribute dimensions, complex co-occurrence relationships among labels, and susceptibility of fine-grained features to background interference. Traditional convolutional neural network （CNN）-based methods often overlook semantic correlations between labels, while graph convolutional network （GCN）-based models, though capable of modeling label relationships, rely on fixed adjacency matrices that fail to dynamically capture asymmetric dependencies among attributes. Additionally, these models exhibit limitations in extracting features from small-scale attributes.

This study aims to enhance the accuracy of apparel image classification by integrating traditional multi-label classification models with Graph Attention Networks （GAT）. By incorporating diverse clothing styles into practical e-commerce systems, the ultimate goal is to reduce labor management costs.

First, an attribute relationship adjacency matrix is constructed to address complex label dependencies. This matrix quantifies the topological structure among apparel attributes by combining label co-occurrence statistics and semantic similarity analysis. The multi-head attention layers in GAT are optimized to dynamically learn asymmetric dependencies among attributes through aggregated visual features and semantic embeddings of neighboring nodes. This approach enhances the model's robustness and generalization in handling intricate label relationships while overcoming the semantic loss caused by binary adjacency relations in traditional GCNs. Second, a residual neural network （ResNet-50） is employed to extract global visual features, while GAT generates label semantic embeddings. An adaptive loss weight allocation strategy based on label co-occurrence probabilities is further proposed to mitigate classification bias for rare attributes under long-tailed distributions.

In summary, to address the accuracy limitations in classifying apparel images with complex label relationships, this study proposes a multi-label classification model that dynamically learns asymmetric dependencies by aggregating visual features and semantic embeddings of adjacent nodes. The model integrates ResNet-50 for global feature extraction and GAT for semantic embedding generation. Experimental results demonstrate that the proposed model outperforms traditional multi-label classification methods in both accuracy and robustness.

Key words：Graph attention network; image multi-label classification; adjacency matrix optimization; multi-head attention

# 目 录

[摘 要 I](#_Toc200986616)

[Abstract 1](#_Toc200986617)

[目 录 I](#_Toc200986618)

[第1章 绪论 1](#_Toc200986619)

[1.1 研究背景与意义 1](#_Toc200986620)

[1.2 国内外研究现状 2](#_Toc200986621)

[1.3 研究内容 4](#_Toc200986622)

[1.4 组织结构 5](#_Toc200986623)

[第2章 相关知识概述 7](#_Toc200986624)

[2.1 多标签分类任务 7](#_Toc200986625)

[2.2 深度学习相关 7](#_Toc200986626)

[2.2.1 卷积神经网络（CNN） 7](#_Toc200986627)

[2.2.2 图卷积神经网络（GCN） 9](#_Toc200986628)

[2.2.3 图注意力网络（GAT） 11](#_Toc200986629)

[2.3 本章小结 13](#_Toc200986630)

[第3章 基于残差神经网络的视觉特征提取模型 14](#_Toc200986631)

[3.1 问题描述 14](#_Toc200986632)

[3.2 模型总体设计 14](#_Toc200986633)

[3.3 多重特征提取模块 15](#_Toc200986634)

[3.3.1 多重特征提取的数据准备与预处理 16](#_Toc200986635)

[3.3.2 基于ResNet-50的特征提取模型 16](#_Toc200986636)

[3.4 多尺度特征融合模块 17](#_Toc200986637)

[3.5 实验设计 19](#_Toc200986638)

[3.5.1 数据集 19](#_Toc200986639)

[3.5.2 评价指标 19](#_Toc200986640)

[3.5.3 实验环境 20](#_Toc200986641)

[3.5.4 实验设计 20](#_Toc200986642)

[3.6 实验结果与分析 21](#_Toc200986643)

[3.7 本章小结 23](#_Toc200986644)

[第4章 基于图注意力网络的多标签服饰图像分类模型 24](#_Toc200986645)

[4.1 任务描述 24](#_Toc200986646)

[4.2 模型总设计 25](#_Toc200986647)

[4.3 实验设计 26](#_Toc200986648)

[4.3.1 评价指标 26](#_Toc200986649)

[4.4 实验结果与分析 28](#_Toc200986650)

[4.5 本章小结 31](#_Toc200986651)

[第5章 服饰类别自动分类系统的设计与实现 32](#_Toc200986652)

[5.1 需求分析 32](#_Toc200986653)

[5.1.1 概述 32](#_Toc200986654)

[5.1.2 用户分析 32](#_Toc200986655)

[5.1.3 功能分析 34](#_Toc200986656)

[5.2 系统设计 34](#_Toc200986657)

[5.2.1 数据库设计 36](#_Toc200986658)

[5.2.2 流程设计 39](#_Toc200986659)

[5.2.3 界面设计 40](#_Toc200986660)

[5.3 系统设计 44](#_Toc200986661)

[5.4 本章小结 44](#_Toc200986662)

[第6章 总结与展望 45](#_Toc200986663)

[6.1 总结 45](#_Toc200986664)

[6.2 展望 45](#_Toc200986665)

[致 谢 47](#_Toc200986666)

[参考文献 48](#_Toc200986667)

第1章 绪论

## 1.1 研究背景与意义

服饰图像分类也是计算机视觉领域的重点研究对象之一，利用服饰图像计算服装图像中的视觉特征，解构图像中所包含的外观属性和着装风格，进而能对服饰图像的理解具有跨多层面、多语义以及多模态的特性。需要注意的是，从图像中提取具有不变性的特征对于可靠的特征匹配至关重，这些特征需要能够应对图像尺度、旋转、仿射变换、3D视角变化、噪声和光照变化等多种干扰因素。其次，在服饰图像中的多标签图像分类问题中也会出现单一标签无法满足的多种标签分类问题，例如电商平台这样需要诸多标签分类的场合，就需要针对多个标签进行处理。也就是说，在这种环境下单一标签无法满足需要的情况会愈发严重，而如何利用图像的多标签分类模型去应对这种情况则显得至关重要。

处理服饰图像需要先做很多准备工作：收集多种服饰图像，再把这些图像统一地进行尺寸调整、去噪、标准化等预处理操作来保证图片的质量与复杂度；然后把处理好的服饰图像按照对应的规格输入到深度学习模型中，通过例如结合色彩和纹理特征的无监督图像分、基于弱监督学习的姿势自适或单独的弱监督学等方式来提取浅层特征，同时也可通过例如视觉注意力机制和多模态技或构造图神经网络中融合标签静态统计信息和标签词向等方式提取高层语义特征信息，再将这些特征信息送入分类模型中进行标签预测分类操作；最后使用交叉验证的方法检验模型的表现效果，再根据验证结果观察模型效果，并在模型出现偏差或者误差之后，对模型的对应参数进行修正或是调整。

然而，服饰图像数据中各标签间存在复杂的相关性，服饰元素本身的特殊性及多样性十分突出，同时服装的图像是细节很多，往往又会有些许遮挡或扭曲、光照等问题。并且每个标签出现的次数可能并不少相同，甚至会有一定重合，在实际使用中，甚至可能出现若干不同的风格或类目，使用传统的多标签学习很难有效地处理这些标签相关性过强的问题，造成模型容易对于标签频率的不同而预测偏斜，甚至无法正确发现标签间的隐形关系或共性问题，使得识别的整体准确度有所下降。基于此，如何针对图像分类问题中存在标签相关性过强、标签不平衡及标签复杂度高问题等多标签分类模型特点，以及如何选择例如分割图像区、区域联合标、多尺度图像特征增、噪声标签训、融合随机森林的多标签分类亦或是构建双路径深度域适应网等多种优化手段对服饰多标签分类模型进行适当的调优是本文的研究重点。

当前，计算机视觉中对于服饰图像分类多采用深度学习中的卷积神经网络（CNN技术，对于海量的数据处理较好，但也有其缺陷。

1）细粒度分类准确性较低，由于服饰类别众多且存在较大相似性，对不同款式的各类细节特征分离时，需要对具有鲜明差异性的特征进行更精细的特征提取和分类。款式、材质、图案、配饰等多种因素均会影响其分类结果，但传统的多标签分类往往只保留最明显的一部分特征去分类，所以忽略了服装很多复杂的细节，这可能会使模型的分类效果变差。对于此类问题，可尝试引入视觉注意力等机制进一步提取特征中具有鉴别性的细节部或是构建MTCT，来获得更完善的细粒度特征表达。

2）图像的噪声和变形十分严重，而且服饰图像中大部分的数据都会有背景杂乱的情况发生，也有些存在着图像形变、图像模糊、背景光强变化和图像尺度缩放等问题都会给模型识别服饰主体和提取特征造成一定的干扰，为处理此类干扰对模型识别造成的影响，通常可在卷积神经网络中引入一些优化来减轻干扰对于特征提取或是分类所造成的影响，例如引入Siamese或Triplet模型结构，此方式中度量学习的引入有效提升了模型对于背景、光照、变形等干扰因素的影。

3）具体而言，由于大多数现有的模型在捕捉细粒度特征方面都比较薄弱，且衣服图像中通常包含一些较为细节化、细微化的部分及不同风格等差异较大的内容，所以现有模型可能无法正确捕捉这些细粒度特征，从而会影响最后的分类精度。

一些专家学者针对图像细节控制不好、背景干扰较多、细粒度分类准确度低等问题，主要采用多尺度特征融合法、、背景预处理及细粒度特征提取等手段降低干扰因素对最终分类预测结果的影响程度。因服饰图像本身的特点，对传统模型结构进行优化是提高最后分类预测效果的关键措施。

## 1.2 国内外研究现状

图注意力转换网络是Yuan等人提出的一种图注意力转换网络（GATN）来解决多标签图像分类问题的方法，他们采用在训练集中根据标签共现概率所得到的邻接矩阵不能适用于所有类别，因为没有考虑到一个额外信息即标签间的语义联系。为此，Yuan等人使用基于标签词嵌入的余弦相似度产生初始的关联矩阵以表示标签间的语义联系，并且设计图注意力转换层将邻接矩阵转变为符合实际应用场景的邻接矩阵，以此发现并突出标签间的。

Yuan等人优化了GAT的邻接矩阵部分，利用基于距离的方法构造相关性矩阵，用包含若干个自注意力分支的新式结构从相关矩阵中挖掘多跳路径，并将相关性矩阵转为若干个子图，提升标签关系表达能力和多标签分类的。

这类对GAT邻接矩阵的优化适用于服装类多标签图像分类任务，可以根据标签间的相似性来进行标签间的准确定位并加强图像分类的表达能力；在GAT原有的图注意力层的基础上加入转换层内容，经过多次自适应学习得到各个标签间的关系，并以此为依据对模型的参数进行调节来提高该模型的泛化能力，比较适合如服装类图像这样的变化较多的视觉内容；但是它在一些跨域的任务中或者图的数据分布差异很大的时候泛化能力会比较差。

为解决 ML-GCN 中标签共现嵌入维度过高影响模型分类性能及未充分挖掘标签间不对称关系的问题，张辉宜等人提出了一种基于图注意力网络的多标签图像分类模型 ML-GAT。模型在ML-GCN完成首次标签关系建模后会存在标签共现嵌入维度过高的问题，为解决此问题则引入了词嵌入降维模块对高维双 向 Transformer 的 表 征 编 码 器 （ Bidirectional Encoder Representation from Transformers，BERT） 标签语义嵌入表示矩阵进行降维，。此方案可降低模型的计算复杂度，并通过降维后的标签语义嵌入和标签类别共现图生成标签共现嵌入，最终对CNN提取的图像特征进行维度统一。

针对海量种类多样化的服装图片难以分门别类、分类效果较差的问题，她基于大型服饰数据集并针对多任务卷积神经网络（CNN）进行了改进，利用上采样恢复位置信息的方式预测服饰关键点热力图，辅助分类及属性识别任务，可以实现将图像映射成热力图，使用注意力机制突出某一个或者一些区域，达到更好定位关键点的效果。此外，在对服饰图片类别进行预测时，由于该场景下的样本类别分布不平衡及多类别过拟合等问题，孙婷对模型参数采用了焦点损失和标签平滑正则化的方式加以改。为增加难易样本的区分度，减少简单样本的权重，该方案在卷积神经网络中加入了焦点损失以及标签平滑正则化技术，使模型更加注重难以辨别的属性。

针对传统多标签图像分类任务只学习图像的视觉特征，忽视了图像标签间的相关性及标签语义与图像特征之间的联系问题，本文提出了基于多头图注意力网络（ML-M-GAT）的。这个模型采用了标签共现关系、标签属性信息来建立图模型，使用多头注意力机制来学习标签注意力权重，再把标签语义特征和图像特征结合在一起，所以既完成了图像标签间的相关性融合，又达到了标签语义和图像特征的对齐。

苏卓等人发现，由于多任务分支神经网络的方案通过将服装属性划分为多个组，每个组配备一个网络分支来进行识别，但该方案没有考虑服装属性内部联系对最终属性识别准确率的影响，导致最终的识别准确率较低。第二种方案是通过构建属性从属关系来改进识别过程，但这种方案无法充分表达服装属性的复杂关系，且仅在损失函数中添加罚项来处理从属关系，。

苏卓等人在图像预处理阶段，首先对服装图像进行统一尺寸和数据增强处理，以适应神经网络的输入要求。预处理完成后，将图像数据输入预训练的残差神经网络模型，提取服装图像的整体视觉特征，接着，让这些视觉特征分别经过M个全连接层，为M个服装属性组提取对应的属性视觉特征，。

此方案中，主要是对传统GAT进行了服装类图像的定制化改造，加入了属性关系图作为支撑内容，后续又在属性关系图中引入了视觉特征，设计了一个完全贴合服装类图像的GAT模型。根据其中关于GAT损失函数的优化可以了解到，在服装分类任务中，某些标签可能较为稀有，而传统的损失函数可能会使得模型倾向于识别频率较高的标签，所以此方案在原有的损失函数中融入了属性关系图注意力网络作为补充，优化了原有GAT中的损失函数。

## 1.3 研究内容

本文的研究目标是通过结合GAT，CNN，GCN模型并针对GAT中多头注意力层中所使用的邻接矩阵进行优化从而实现服饰图像的准确分类。

1. 图注意力网络自注意力机制部分是利用节点之间两两之间的相似度来确定注意力权重的，在此过程中由于过多地关注图中的局部结构信息而没有考虑到图中更多的全局结构信息。为此，可以构造出权值为1/2的加权邻接矩阵，基于图中各节点的特征相似度调整边上的权值，通过给邻接矩阵加上合理的加权，使其更适合于完成所要完成的任务，为图的各个节点提供更多的信息输入，进而改善对于多标签分类任务的效果。
2. 在服饰图像分类问题中，如果直接用图像数据输入图注意力网络会存在与图结构不符的问题。因为图像本身具有较强的空域特性，在直接映射到图结构的过程会出现计算效率低、空域信息丢失或无法很好地抓住图像中的复杂结构的问题。设想用数据增强的方式来获得更多的训练样本，并通过引入与数据相关的各种增强方法，以提高模型对于服装图像数据的通用性，不仅仅只采用简单的翻转、旋转数据增强的方式，增加拍摄角度、背景及前景变化情况等都有利于训练后的服装分类模型对不同情况进行识别，因此在训练服装图像分类模型时加入特定的数据增强操作，能够使训练出的服装分类模型有更好的鲁棒性，更能应对多样且复杂的场景和情况。

修改完善以上关于模型的所有调优工作后，在以下两个情景下完成对模型的对比检验：

* 1. 首先利用商城系统的服装图像，完成服装图像的多标签分类工作，可以针对不同的筛选条件搜索到相应的服装图像的内容，对于模型来说，需能对于每一张服装图像的多个标签进行预测，各个标签彼此之间可能存在一定的联系或者相似之处，可以捕捉到其中的联系或相似性，且需要适配各场景下拍摄的服装图像内容。
  2. 即使服装图像是遮挡（如部分服装被其他物品遮挡）、逆光（比如光不够亮或者图像太亮），或者是噪声大、模糊等等其他影像质量的因素存在，还能进行多标签服装图像分类。因为某些服装部位会被遮挡或者重叠在一起，这样会导致普通的图片分类方法不能正确地判断这个图片是属于哪个类别的；光照不匀、逆光拍摄可能会导致图像中衣服的某些部位的轮廓变得不清楚，因而模型可能无法准确判定衣服的颜色或者款式的正确性，或者是由于拍摄条件不好或者拍摄过程中有运动模糊等原因造成了图的质量不高从而也影响到了模型的分类准确性。

## 1.4 组织结构

本文由 6 个章节构成，每个章节及内容安排如下：

本章绪论介绍了服饰图像分类任务的研究背景与意义，并介绍了国内外研究现状；本章还说明了本文主要研究内容以及本文的组织安排。

第二章概述了服饰类图像的基本特征，介绍了自然语言处理、深度学习原理与技术、图论基础、图神经网络基础理论以及分类效果评价准则。

第三章基于图注意力网络的多标签服饰图像分类模型。从服饰图像多属性识别的任务出发，首先阐述了对服装图像中多个属性进行识别的重点任务，并且详细地阐述了服装图像特征提取模块的重要作用；其次介绍了多尺度特征融合模型的整体架构设计思想，对各主干网络特征层的单路或组路输入和多尺度信息融合方式做了详细的说明，并描述了模型中使用多尺度信息融合方式来实现特征融合的过程；然后讨论了浅层和深层特征针对细粒度属性以及全局语义所具有的差异性，并解释了利用多尺度融合可以使得模型获得更好的效果的原因；最后介绍了实验部分的相关设置，包括使用的数据集、评价指标、实验环境、实验设计及结果分析等内容，全面证明了多尺度融合模型在分类准确率、F1分数等关键指标方面的优越性和实用性。

第四章基于图注意力网络的多标签服饰图像分类模型。首先从服饰图像多标签属性识别任务入手，介绍该类任务的模型整体设计思路和引入图结构建模的原因，然后介绍模型的具体构成模块：服饰属性图的构建、基于GCN的属性全局特征提取、基于GAT的属性关系建模以及图结构-卷积特征融合机制，并分析模型在多属性识别准确性以及鲁棒性上进行提升的机理；随后介绍本章实试验中的具体设置情况及其实验效果分析，在分类正确率（CRR）、分类位置秩准确率（CPRA）、区位排名（ZR）三项指标下分别将本文所提出的模型与多款主流视觉模型做对比，说明本文所使用基于图神经网络的方法是有效的、具有明显优势的。

第五章主要是对服饰智能分类系统的相关内容进行了介绍，通过对系统所要解决的问题的需求分析、用户及业务的分析以及对功能模块进行剖析说明了系统的设计思路；通过对系统的整体结构及运行流程等方面进行描述和解释说明了系统的框架结构，并对其数据库、界面进行了介绍；并主要介绍了实现该系统所需要的技术以及采用的语言与工具。

第六章，总结与展望。总结了全文的研究成果，并对当前研究的不足之处进行了深入探讨。在此基础上，对未来的研究工作提出了设想和展望。

第2章 相关知识概述

本文利用图注意力网络、多头注意力机制和神经网络等方法来解决复杂服饰图像分类不准问题。本章将为下文涉及的知识做相关介绍：多标签分类的任务、图注意力网络和卷积神经网络的概念理论。

## 2.1 多标签分类任务

是一类重要的监督学习任务，其核心特征在于每个样本可以同时关联多个标签，如图2-1所示。区别于传统的多类分类（Multi-Class Classification）任务中“单样本对应单标签”的设定。在多标签场景中，标签之间往往存在一定的依赖性和共现关系，这为模型的建模和优化带来了更大的挑战。为有效解决该问题，研究者提出了多种建模策略，如将问题拆解为多个独立的二分类任务的二值相关方法（Binary Relevance），考虑标签间顺序依赖关系的分类器链方法（Classifier Chains），以及将标签集合视为新的复合类别进行建模的标签权集方法（Label Powerset）等。



图 2‑1多标签分类任务

近年来，随着深度学习的发展，神经网络结构（例如卷积神经网络、循环神经网络、图神经网络等）在多标签分类中得到广泛应用，常配合 Sigmoid 激活函数输出多维概率，并以二元交叉熵（Binary Cross Entropy）作为损失函数以优化每个标签的独立预测概率。评估指标一般包括Hamming Loss和微平均和宏平均的精确率（Precision）、召回率（Recall）和F1值等，刻画了整个系统和模型在每种标签上的效果；多标签分类目前应用比较广泛的就是包括文本标注、图像识别、生物信息学和安全威胁的检测等等方面，在机器学习方面也是一大热点问题。

## 2.2 深度学习相关

### 2.2.1 卷积神经网络（CNN）

卷积神经网络（Convolutional Neural Network，CNN）是一种专门针对网格结构数据的一种深度学习模型，如图2-2所示，用在图像处理上具有突出的效果。

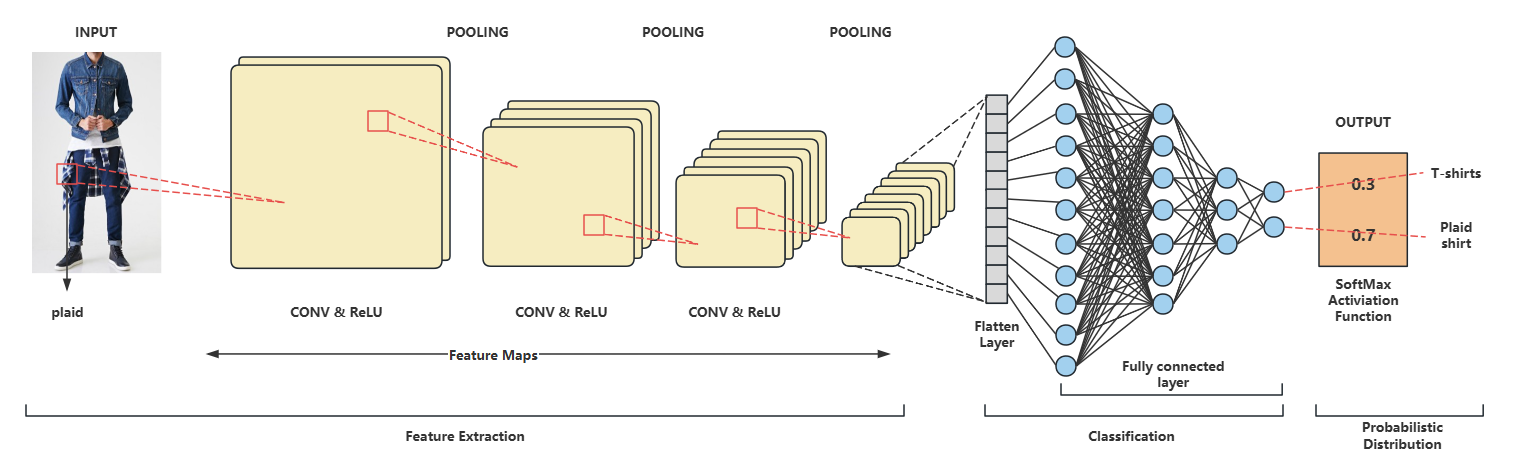


图 2‑2 卷积神经网络结构图解

CNN的主要思想是结合卷积层、池化层以及激活函数，采用局部连接、权值共享以及多层结构的方法将图像从下至上逐层抽取出低级到高级的特征表示。在卷积层中，对于输入特征图和卷积核 ，卷积操作可以表示为：

0‑1

为输出特征图上第行、第列、第个通道上的值，为卷积核的大小、为输入通道数、为输出通道数、为第个通道上的偏置项。这样的局部连接方式可以保留图像的空间结构信息，并且减小了参数数量；另外由于权值共享机制，所有的卷积核是从整张图片中来提取相似的特征模版，所以大大减小了参数量。池化层使用降采样来减少特征图的空间大小，最常用的池化方法就叫做最大池化，公式表示如下：

|  |  |
| --- | --- |
|  | (2-2) |

其中，表示以位置为中间节点的池化窗口。 其中，c表示第c个通道的索引，此时操作可以将主要特征选择出来并且模型对于位置变化的鲁棒性得到了加强。池化操作还可以降低计算量，并且可以增强模型对输入图像的平移变换、旋转变换等等的变化的鲁棒性。激活函数引入非线性变换，增加网络的表达能力。常用的ReLU激活函数可以表示为：

|  |  |
| --- | --- |
|  | (2-3) |

这种非线性变换使得网络能够学习更复杂的特征表示。此外，ReLU激活函数还具有计算简单、缓解梯度消失等优点。

作为卷积神经网络中的一种经典类型，是将残差网络进行具体的化的一种实现方式，其网络结构图，由于加入残差连接可以很好地解决深层网络训练问题，其次ResNet-50一共包含50个卷积层，首先从最初的卷积层开始，然后经过残差块组，其中分为4组残差块组，然后以全连接层结束，每组残差块组均包含有多个残差块，而每一个残差块均由三个卷积层组成，分别是1×1卷积层进行降维操作，3×3卷积层进行特征提取操作，再由1×1卷积层进行升维操作。

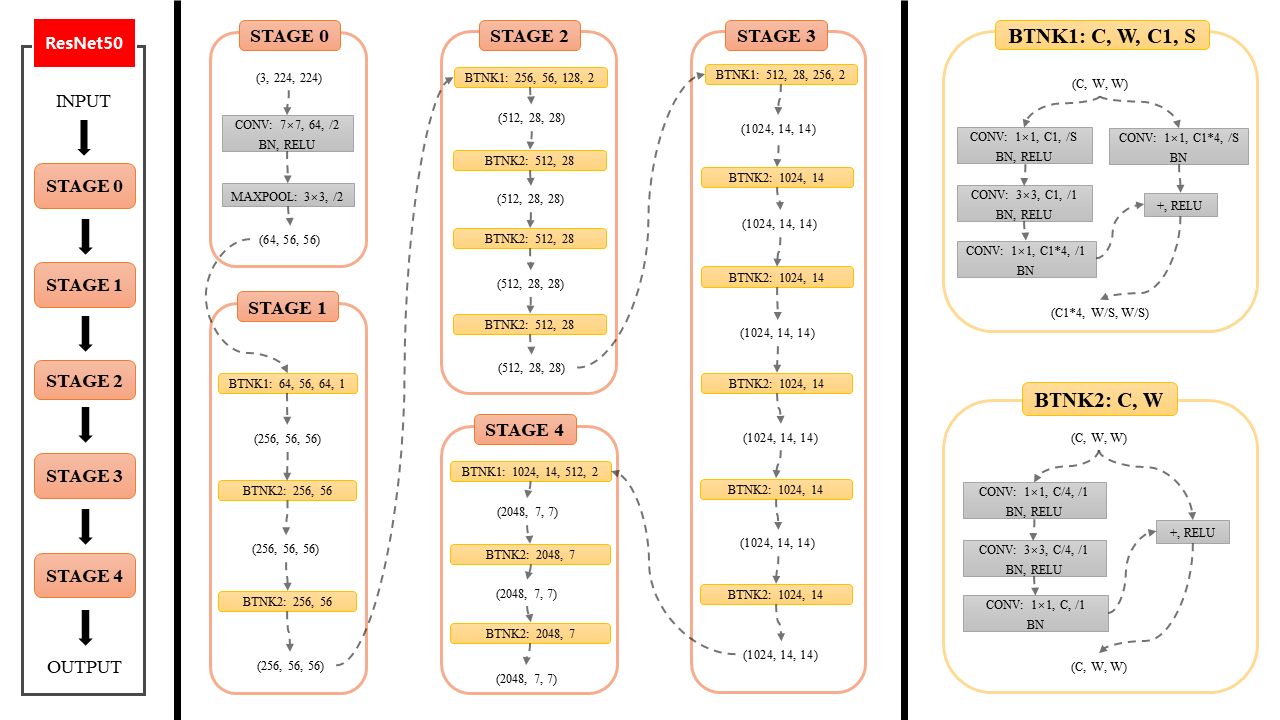


图2-3 ResNet-50结构详解

残差连接是ResNet-50的核心创新，它允许网络直接学习输入与输出之间的残差映射，而不是直接学习完整的映射函数。对于输入，残差块的计算可以表示为：

|  |  |
| --- | --- |
|  | (2-4) |
|  | (2-5) |

其中，是期望的底层映射，是残差映射，也就是上式里面去掉跳跃链接的映射部分。这样模型可以直接学习到残差部分而不再关注整个映射函数，通过跳跃连接可以很好的传导梯度信息从而避免深层次网络所存在的梯度消失问题。ResNet-50里每个残差模块包含了三个卷积层，其计算的过程如下：

|  |  |
| --- | --- |
|  | (2-6) |

与和卷积层的权重相对应的是一个有三个单位数的层，该层使用激活函数分别将以上三种类型的层的输出求平均。这种方案减少计算量的同时也增加了特征提取的速度。ResNet-50的网络结构可以表示为：

|  |  |
| --- | --- |
|  | (2-7) |

其中，为第层映射函数，为整个网络层数目。利用残差连接方式可以使梯度直接传递给浅层部分，从而使整个网络能较好的训练到深层。同时有利于ResNet-50对图像信息进行深层次的学习和表达。

### 2.2.2 图卷积神经网络（GCN）

是一种用来解决图结构数据的深度学习模型，它是将传统的卷积神经网络扩展到图域的一个方法，整体结构如图2-4所示。由于把图结构的信息带入到了卷积操作当中，所以基于GCN可以很好地获得节点间的拓扑关系以及特征信息，并且它的基本思想是在保留图结构的前提下，学习图中每个节点的表示，并基于节点的表示学习图结构数据的方法，做到端到端地学习。

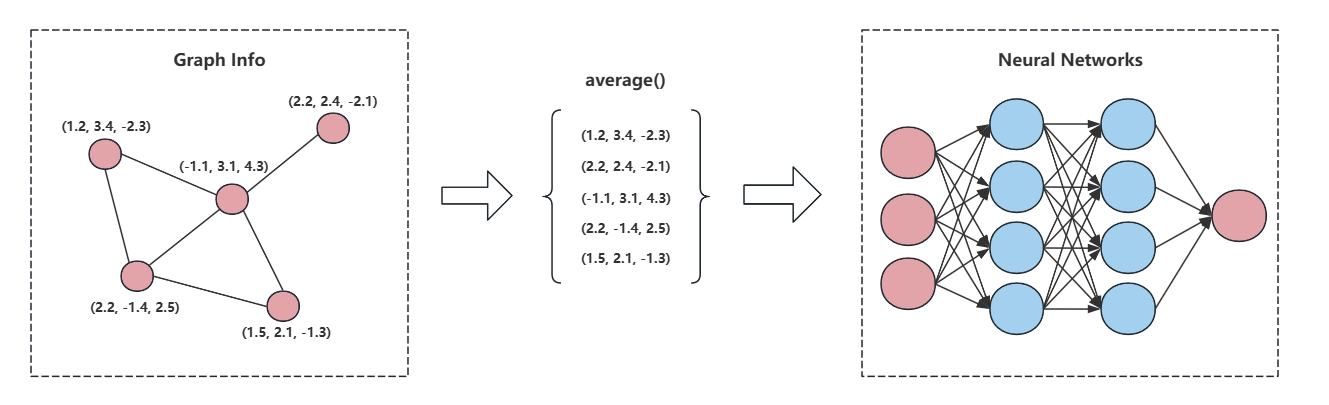


图2-4 图卷积神经网络整体结构

在图卷积层中，对于输入特征矩阵和邻接矩阵，图卷积操作可以表示为：

|  |  |
| --- | --- |
|  | (2-8) |

其中，是第层的节点特征矩阵，是可学习的权重矩阵，是加入自环的邻接矩阵，是度矩阵，σ是激活函数。这种设计使得模型能够同时考虑节点自身的特征和邻居节点的信息。

为了增强模型的表达能力，GCNs通常采用多层结构，每一层的计算可以表示为：

|  |  |
| --- | --- |
|  | (2-9) |

其中，是卷积核的数量，是第个卷积核对应的邻接矩阵。这种多核设计使得模型能够捕获不同尺度的图结构信息。

GCN最主要的优势在于其可以对非欧几里得结构的数据，例如图结构数据的处理，在图结构之上定义卷积操作，则能通过卷积捕获每个节点与其邻居的关系以及各自所拥有的特性，使图结构上的节点能完成对其所在位置相邻节点的卷积处理，进而实现端到端地学习，更好的提取相关特征，这正是GCN用于处理图结构数据、进而让后续的任务获得更好的性能的支持条件。

一般在训练阶段使用的是端到端的训练方式，将损失函数设置成最小化目标，选取常用的交叉熵损失、均方误差损失等等。因此，只要在网络结构设计、训练方法上进行合理的调整，就可以让GCN充分的学习到图结构数据的特性，得到良好的特性表示形式，在以后的任务中得到好的支持。

GCN在各种应用场合得到应用：社交网络，推荐系统，生物信息学等等；在这些地方，GCN都可以使用图来对这一类问题的数据结构进行建模，并从中提取有用的信息来改进模型的性能，在以后随着深度学习的发展，对于GCN来讲在图数据处理上会起到更加突出的作用。

### 2.2.3 图注意力网络（GAT）

是一种基于注意力机制的图神经网络模型，关注解决图数据中的节点信息如何在网络中进行传播的问题，图注意力网络具体结构如图2-5所示。由于GAT使用了注意力机制，在聚合邻居信息的时候可以动态地根据邻居节点的重要性给不同的邻居分配不同的权重值，从而有利于图网络中更充分地发现图中的节点间关系。GAT通过引入注意力机制得到节点与邻居节点的关系，并用多头注意力机制得到不同的注意力权重，使计算过程更快且有效。在传统GCN中，用平均或加权平均的方法对一个节点邻域的信息传播到目标节点中；而在GAT中，利用自注意力机制来表示相邻的两个点之间的关联，再利用多头注意力机制同时利用多个自注意力头来提高模型的表达力以及稳定性。

在图注意力层中，对于输入特征矩阵和邻接矩阵 ，注意力系数可以表示为：

|  |  |
| --- | --- |
|  | (2-10) |

其中，和是节点和的特征向量，是可学习的权重矩阵，是注意力向量，是节点的邻居集合。这种设计使得模型能够根据节点特征自适应地调整邻居节点的重要性。在GAT中，节点的更新可以表示为：

|  |  |
| --- | --- |
|  | (2-11) |

其中，是更新后的节点特征，是激活函数。这种设计使得模型能够根据节点特征自适应地调整邻居节点的重要性，从而提高模型的表达能力。GAT的网络结构可以表示为：

|  |  |
| --- | --- |
|  | (2-12) |

其中，是第个注意力层的计算函数，是邻接矩阵。通过多层注意力机制，模型能够捕获不同尺度的图结构信息，从而提高模型的表达能力。

一般来说，GAT中的邻接矩阵可以用来表征图中节点的互连关系。因此，得到的邻接矩阵也决定了构成该图的优劣情况。如果该图未能充分地表示真实的节点间关系，则GAT也就无法从对应的邻居节点获取有用的信息从而影响该模型的训练过程，此时网络输出结果就会变差。故而对于GAT来说应该选择合适的邻接矩阵形成方式，使得网络可以根据具体的数据来理解出一个较好的图结构，从而取得较好的训练效果。

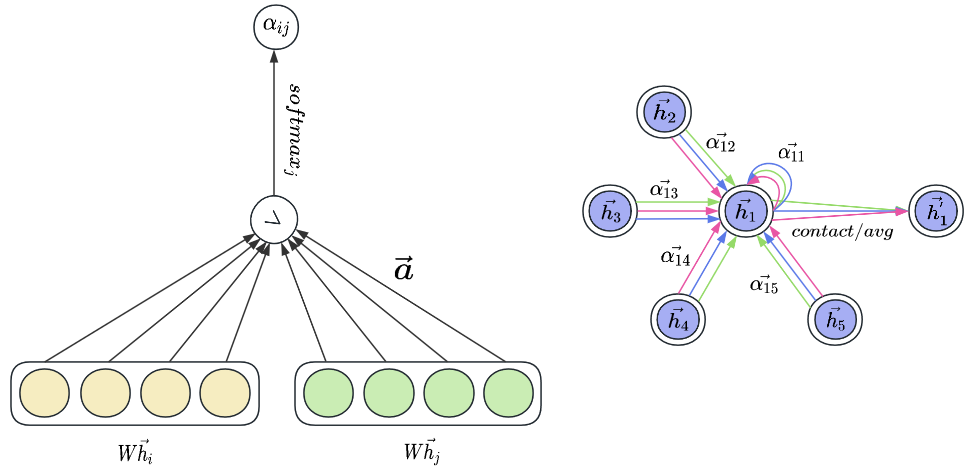


图2-5 图注意力网络整体结构

GAT模型与MLADG模类似，都利用了注意力机制让其自动去学习节点之间的权重或是图像特征，在学习时可自动感知出所有有关的信息来进行整体考虑，相对于传统模型中单方面的学习方式来说，GAT模型在注意力机制的辅助下实际学习能力提升不少，并且这些变化能极大的提高后续任务的运行效率。伴随着深度学习技术不断的发展完善，对于GAT而言，该方法还可以利用GAT来解决图的数据问题。

其次，GAT中的注意力层使用到了，此方法是将注意力机制应用到GAT中所使用的一种扩展技术，在很多Transformer模型与现今的神经网络结构中都可以找到多头注意力的应用，它主要的目的就是利用并行计算的方式得到不同注意力“头”的输出结果，达到在同一输入特征下得到不同的特征表示。其为了解决注意力的缺失、引入更多的注意力视角而产生的思想，可以在一个注意力的基础上向多个方向扩展得到不同的注意力。

传统的注意力机制每次只会计算一种注意力得分，在多头注意力机制中，输入会被分成多个人造的小空间，并分别独立地算上各种注意力，最后把各个头的结果再合起来，实际上是从多个视角去考虑这个问题。对于图像标签关联度高的情况来说，多头注意力能够使模型关注不同的信息，让模型分类更加准确，并且让模型对于标签的预测结果更加准确，同时在实际的图像特征中也体现了注意力模块对于强关联标签的处理能力，例如服饰图像检索算法中所使用到的。

在多头注意力机制中解决了很多在单头注意力机制中所存在的问题，或是由于多头处理而产生的问题，由于单头注意力只能在同一投影空间内计算注意力，无法同时关注多种语义关系，为了解决模型表达能力不足的问题，所以采用了多组线性投影并行机制，使得多头注意力机制中的每个注意力头都能在不同的子空间中独立捕获特征，从而提升模型效果。具体计算方式如下：

|  |  |
| --- | --- |
|  | (2-13) |
|  |

但在并行计算的过程中个注意力头的输出维度为，存在与模型维度不匹配问题，多头注意力机制则使用了拼接和线性映射的方式，

|  |  |
| --- | --- |
|  | (2-14) |
|  |

从而解决维度不统一的问题，其次过多的并行头会线性增加计算量和参数量，为了解决效率与负载均衡问题，所以通常针对多头注意力的维度进行了设置，具体限制方式如公式2-15所示。

|  |  |
| --- | --- |
|  | (2-15) |

利用此公式限制多头注意力维度参数，可使得整体计算量与单头注意力持平，同时兼顾表达能力。

## 2.3 本章小结

本章主要讲述了关于本论文所涉及的理论框架。首先将多标签分类问题的概念拆分，并介绍了该任务中各部分完成的主要工作及如何进行评估，紧接着介绍了使与深度学习相关的部分模型如：卷积神经网络、图注意力网络以及多头注意力机制等内容的基础原理。

第3章 基于残差神经网络的视觉特征提取模型

本章针对服饰图像问题中图像特征提取部分进行研究。基于残差神经网络（ResNet-50）预训练模型建立了更符合服饰图像特征的提取模型，解决了传统模型缺乏图像细节特征提取而导致最终分类模型分类精度低的问题。

## 3.1 问题描述

图像特征提取作为分类预测模型的重要组成部分存在，为最终分类预测结果提供关键的特征信息。在服饰图像中既包含纹理和形状等基础特征，又包服饰局部细节特征以及整体风格等高层特征信息。只依靠图像底层特征并不能准确地描述该服饰图像中所表达的内容，因此需要将底层基础图像特征与高层语义特征相结合才能够更为准确的判断图像中所包含的分类以及标签信息，而不能仅仅是单一的利用某一特征层进行多属性信息的融合。卷积神经网络（CNN）在单标签图像分类任务中表现出了卓越的性能，但在多标签场景下存在两个主要问题：首先在传统卷积神经网络中，每幅图像通常被视为一个不可分割的实体，其中没有针对不同标签对应的视觉信息进行单独处理或分析。其次，由于图像处理方式的整体性，图像标签之间的相关性往往在分类任务中受到忽。

为解决传统卷积神经网络中图像整体处理所导致的两个主要问题，当前模型在服饰图像处理步骤中将图像所包含的多属性信息整合到特征层中，同时在优化特征捕获层次时需要避免忽略服饰图像的一些浅层细节特征，并在这样的条件约束下，同时采集服饰图像的深层语义特征。

因此此处的特征提取模型则采用了类似MSACNN多尺度特征融合的方式，将不同层次的特征信息进行有效整合，既能够保留图像的浅层细节特征，又能充分利用图像的深层语义信息。这样，服饰图像的处理能更好的把握一些复杂纹理、形状与多属性特征，可以使得识别和分类得到更为准确的特征信息。

## 3.2 模型总体设计

本章所设计的特征提取模型同等关注了服饰图像中的基础视觉特征以及深层语义特征，利用多尺度特征提取方式解决了传统卷积神经网络中特征提取未同时关注局部特征和整体风格的问题，并采用局部特征与全局特征相结合的方式保证局部特征与整体风格特征的提取平衡性，保证在两种特征提取中都不过于极端。

为了能更好的从服饰图像提取出其特征，在服饰图像的特征提取中采用多层级残差网络（ResNet-50）做为主干模型，通过控制冻结浅层特征层来保证基础视觉特征的提取，并控制深层特征提取层来提取出图像中的款式、结构等语义信息；并且还要分别对不同尺度的特征内容进行单独处理，以此保证获得的所有特征内容的完整。

其次，补充了特征自适应层以及特征融合层，并且把所获得的局部、全局以及上下文多路特征等各种特征内容用卷积的方式汇总各个层的特征权重，再用双线性插值方式统一特征尺度之后，在使用带有自适应池化和全连接层的特征融合网络压缩特征空间维度之后将不同属性特征融合成一并输入并用层归一化加上非线性激活增强特征表达能力以提高模型在特征提取部分的准确性。模型粗略结构如图3-1所示。

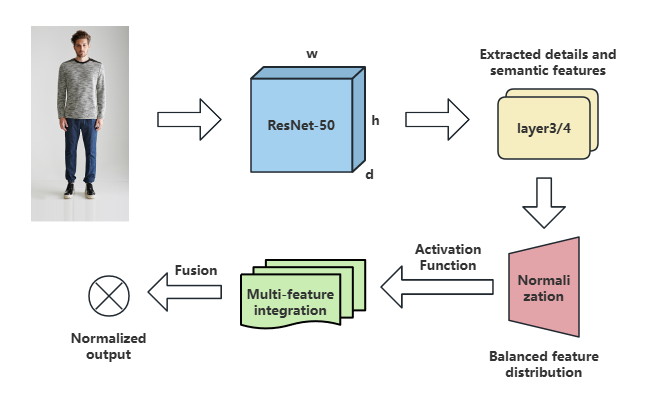


图3-1 多属性特征提取流程图

## 3.3 多重特征提取模块

服饰图像类别是依据服饰的某些特征以及一些高阶的语义特征来区分的。服饰风格的分类是作为整体大分类存在的，而服饰风格下又可以区分出类别、颜色、布料以及纹理等特征，且这些特征之间存在互相影响的情况，例如当纹理中出现了褶面纹路时，可认为该服饰属于百褶类服饰也属于马面裙类服饰，风格则被归类为休闲或是民族服饰，在多种属性特征的影响下类型的归属也具有不确定性和多元性。因此，在DeepFashion这样的大规模服装数据集中提取特征需要考虑融合多种元素以及多尺度下的特征，例如FashionNet这样的联合预测服装属性和地标来学习服装特征的模，此类模型优化方式有效的融合了多尺度多种类的特征内容。服饰可视为这些基础属性的集成体，从函数角度可表达为：

其中，款式要素作为分类的核心轴心存在，用于定义服饰的形态结构，例如连衣裙、衬衫、牛仔裤等服装所属大类，材质则用于描述面料物理属性的特征，例如化纤、棉、皮质或是丝绸类特征信息，然后是颜色与纹理的特征信息主要属于表征视觉呈现形式部分，最后则是高层级的场景以及目标人群特征信息，这两者主要是为了反应当前服饰的用途以及适用的目标用户。同样的，在本模型使用到的标准分类体系数据集DeepFashion中，其分类规则明确要求同时标注款式、材质、场景等属性。例如，"男童蓝色涤纶运动外套"的命名需包含人群（男童）、颜色（蓝色）、材质（涤纶）、场景（运动）和款式（外套）五大要素。

### 3.3.1 多重特征提取的数据准备与预处理

本次实验使用的是来自Liu Ziwei研究团队提出的一个大规模服饰图像数据集DeepFashion数据集，其包含80万张各不相同的时尚图片，其中包含了从整理好后静止摆放在店里的照片到没有约束由用户上传的照片，为最大化的可视化时尚分析提供了一个巨大的数据库，并且该数据集中还标有丰富的服装项目。除此之外，该数据集中的每一个图片都有50个类别、1000个描述性属性、边界框、服装地标和30多万跨姿势或跨领域的图对。

根据DeepFashion数据集的分析与处理结果，提出了属性预测、C2S服装检索、店内服装检索和地物检测4个基准，对应的实验结果已经给出。基于这些基准的测试数据和标注信息同样也可以用于服装检测、服装识别、图像检索等其他计算机视觉任务上的训练集及。

由于此处使用到的DeepFashion服饰分类数据集中包含的数据非常的复杂且多样，在进行特征提取前需要先对已有的图像数据基于ImageNet数据集计算进行标准化处理从而使数据分布更均匀，加速模型收敛。另外还需要将数据集中给出的标注文本文件转换为可识别的二进制向量数据，用于模型预测多个图像属性，然后对收集起来的属性进行统计与过滤避免属性的极端不平衡问题导致模型训练受到影响并生成对应的索引映射表，便于后续的标签向量化操作。

### 3.3.2 基于ResNet-50的特征提取模型

本章提出的特征提取模型使用了ResNet-50预训练模型来代替Visual Transformer中提出来的模型，作为整体模型中的视觉编码器，在该模型中主要用于完成原始图片到高质量特征向量的映射过程，其映射到的特征向量将直接用于后续GAT和GCN下述相关的属性关系建模工作。图3-2整体框架图

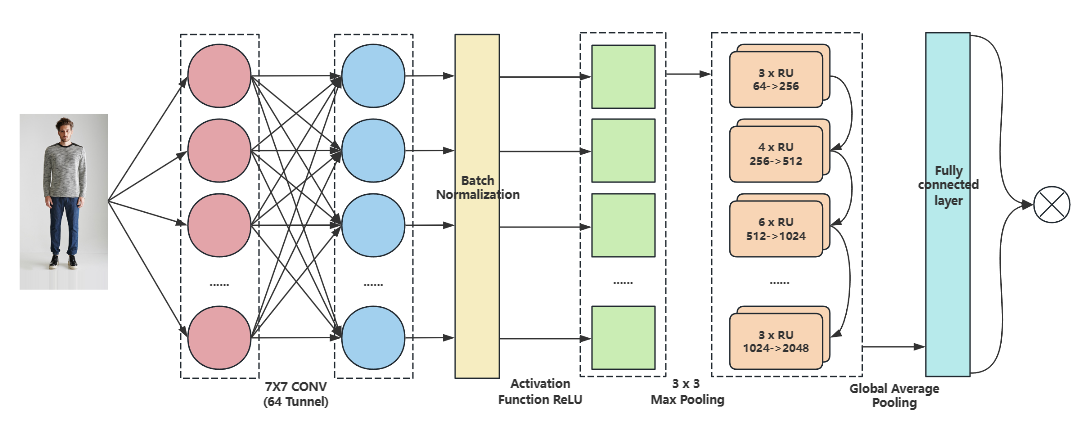


图3-2 ResNet-50整体框架

本实验使用了ResNet-50作为模型主干网络，其是微软研究院提出的一种基于深层卷积神经网络（CNN）的预训练模型，在深度神经网络训练的过程中，随着网络深度加大，将会导致训练误差或损失函数值增加，这就会出现严重的退化问题，即深层次的网络不如浅层次的网络效果好。为了更好的解决退化问题， ResNet-50结构中一方面，由于残差连接可以让网络里面的梯度直接传导，或者通过跳跃链接绕开某几层之后再进行网络的直接传导，有效避免了梯度消失和梯度爆炸的情况发生；另一方面，在残差连接中加入了恒等映射路径能够让网络直接将输入数据传递到输出部分，这样就不存在层数过多带来的退化问题了，因此每个残差块就可以在训练过程中不用学习如何表示整个图像，而只需要学习到这一层到下一层的变换，就能够有效的加快模型训练速度、模型学会更多的特征。

## 3.4 多尺度特征融合模块

对服饰图像局部细节特征和全局结构语义特征相融合的方法的主要思路包括：

1. 先用主干网络提取出服饰图像的多个尺度的特征图，再利用对多尺度特征做语义增强。
2. 然后将用FPN处理后的多尺度特征图空间对齐，接着将其融合为不同尺度下的多样化的服饰图像的特征。
3. 最后使用池化、全连接方法得到包含服饰细节及结构特征在内的全局特征向量。

对于服饰图像多标签分类问题，一方面要抓住细节如纽扣、花边、纹理等部位，另一方面也要兼顾整体，获取图像中整体造型、廓形、领型等信息，单纯的单尺度特征信息很难兼顾细节及结构信息。一般地，在图像的信息当中，底层特征更偏向于提取局部的细节信息，但是没有全局语义的特征信息；而高层特征提取的是全局的结构与语义信息，但是在细节方面就存在缺陷了。

针对这一点，我们可以用FPN网络来将高层的信息流向低层，采用自顶向下信息流的方法，将低层的局部细节信息和高层的全局语义信息进行结合，使得每一层都可以同时捕捉到图像的细节以及全局结构信息，提高了模型对服饰中不同类型的属性的理解程度。FPN网络结构如图3-3所。

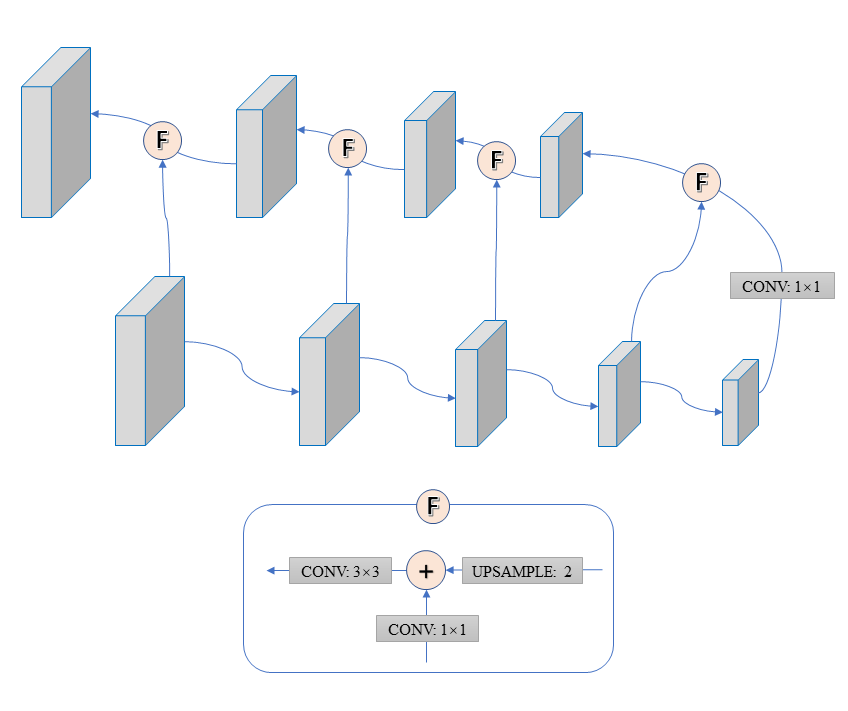


图3-3 FPN网络结构

在实际特征提取过程中，各个尺度提取到的特征图的空间尺寸、通道数、语义等均不相同，如果直接采用一般的融合操作来进行融合，则会造成不同尺度特征图维度的不对齐，最后使得特征不能得到正确的融合；如果只是单独地使用其中一层的特征，又会丢失另外尺度的信息。而空间对齐和通道拼接结合的方式可以较好地解决上述的融合问题，能让模型充分利用到服饰图像在各个尺度的所有信息，提高属性识别的效果。

相比于只用ResNet-50等主干网最后几层（如layer4），或简单的将几个层次的特征加入池化层处理的方式融合成一层的1×1规模的单尺度模型；或者直接按照以下公式所给出的办法去做不同层的特征直接拼接的做法，我们分别采用了FPN、空间对齐和通道拼接的多尺度融合方案来获取。该种方式可以在不同尺度上捕捉服装图中不同的细节与结构，对于多标签属性识别的性能有一定的提升作用。

具体来说，对于单尺度的等模型只能够获取服装图像的总体结构信息，无法同时兼顾面料的纹理，纽扣或花边这样的细节信息，只是简单的组合在一起也没有经过空间对齐以及语义增强，也就意味着不同的尺度特征并不能够达到最大程度的信息集成，相比于单一的融合多个尺度特征，对于经过FPN得到的结果来看，不仅仅融合了不同尺度的特征信息，更重要的是使用自顶向下的信息传递方式在融合的过程中既包含了高层的全局结构信息也包含有低层的细节特征，并且每一层的尺度特征都拥有丰富局部和全局的信息，而且空间对齐、通道拼接使得不同尺度的特征的位置都是可对接互通的。

## 3.5 实验设计

### 3.5.1 数据集

在本研究中，用于实验的数据集是来自DeepFashion的数据集，这是目前计算机视觉领域常用做服饰属性识别及服饰检索的数据集，由于其规模大，且包含了大量的服饰类别，而且有高质量的标注，因此在一定程度上具有代表性，对做服饰图像理解、属性预测等工作来说是一个不错的选择。

实验使用了DeepFashion数据集中的服装属性预测数据集作为本次实验的基础数据，收集了大量来自于现实世界的服装图片，并且包含了种类较多、风格各异、颜色多样以及形态上也含有不同的细节改变。另外每一张图片都带有相应属性标签，包括衣服类型（例如T恤衫、衬衫、裙子……），颜色，图案，袖子样式，领子样式等等。而属性标签正是实现多标签分类任务的理想载体，同时也在一定程度上考察了模型针对细粒度的识别能力及对多属性预测的效果。

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

(数据集表格说明)

对于数据预处理，所有的图片都被设置成同一大小，并且经过了归一化的处理以达到可以被送入深度神经网络的目标，在进行训练的时候使用了数据增强的方法，用到的是随机裁剪，水平翻转，颜色抖动等等的一些手段。并按照官方的进行数据集划分，分为训练集，验证集以及测试集。

值得一提的是，DeepFashion数据集的多标签属性结构自然适合本文需要使用到的准确率、精确率、召回率、F1分数、mAP等多指标评估体系。因此，模型既可以将总的评价结果拿来进行对比，也可以针对具体的每个属性进行单独的对比，便于发现每个模型间的优点与不足，同时能够具体分析每个不同服装属性，从而进一步挖掘出影响不同服装类型的特征层组合的最佳方法，真正达到更好优化的目的。

综上所述，DeepFashion数据集以其丰富的标注和多样的服饰样本，为本研究的消融实验和多尺度特征融合模型的评估提供了高质量的数据支撑，确保了实验结论的科学性和实用价值。

### 3.5.2 评价指标

本实验使用了平均精度均值（mAP，mean Average Precision）、准确率（Accuracy）、精确率（Precision）、召回率（Recall）以及F1分数（F1 Score）作为评价指标。

mAP因为其用于多标签分类领域的通用性而成为服装类目检测目前所采用的主流方法，能够检测出模型对所有服饰属性的判断情况，也是检验多标签识别性能优劣的标准。其此，准确率反映了模型正确预测的样本比例，精确率衡量了模型预测为正类的样本中真正为正类的比例，召回率则评估了模型能够正确识别出的正类样本的比例。F1分数作为精确率和召回率的调和平均数，能够更好地平衡这两个指标。此外，我们还记录了模型的推理时间（Inference Time），以评估模型在实际应用中的效率。

单属性准确率则用于细致分析模型在具体属性上的表现，帮助我们理解多尺度特征融合在细粒度属性识别上的优势。

### 3.5.3 实验环境

本次实验采用的实验环境所下表所示：

表3-1 实验环境

|  |  |
| --- | --- |
| 环境 | 版本信息 |
| GPU | NVIDIA GeForce RTX 4070 Ti SUPER |
| CPU | Intel（R） Core（TM） i5-14600KF @3.50 GHz |
| 操作系统 | Windows11 |
| Python版本 | Python3.8 |
| 框架 | Pytorch |

### 3.5.4 实验设计

本文主要针对多尺度特征融合实验展开研究，探究不同的深度学习模型下不同的特征层组合是否会对结果产生影响，同时验证多尺度特征融合的有效性，在这里建立了一套完整的消融实验系统，如表3-2所示，并基于此对各模型方案进行全面的测试并使用多种评测方法对比分析。

表3-2 消融实验计划表

|  |  |  |
| --- | --- | --- |
| 实验组 | 特征层配置 | 预期影响 |
| A1 | 多尺度融合层 | 基准组 |
| A2 | 单尺度特征层 | 验证多尺度融合的必要性 |
| A3 | Layer1 | 验证浅层特征的效果 |
| A4 | Layer2 | 验证中层特征的效果 |
| A5 | Layer1 & Layer2 | 验证浅层和中层特征的组合效果 |
| A6 | Layer3 & Layer4 | 验证深层特征的组合效果 |
| A7 | Layer2 & Layer3 & Layer4 | 验证中层和深层特征的组合效果 |

首先，在实验中以ResNet-50为主干网络，分别抽取不同深度（layer1-layerX）层数之和）以集中注意力于图像的通用视觉特点。然后将不同深度的这几个特征层进行拆分，每种组合对应一个单独的实验，最后得到多组实验结果并进行横向比较，得到各层级及组合的对总体模型效果影响情况。具体实验方式为，先针对所有的单层分别进行性能统计，其次再对所有可能的两层组合进行检验，由于层面越多，总的组合就越多，为了使试验结果容易阅读、试验过程易于操控，本文只取出了总试验结果中评价最好的两层三层及以上进行试验与展示。

基于此，在本文的工作中，提出多尺度特征融合（MultiScale）方法，利用融合不同深度特征层的办法来提高网络对图像中复杂的场景表达能力。为了保证实验的公平性，所有模型的训练、测试过程使用的都是同一批数据集、相同的优化器、相同的损失函数以及相同的学习率等超参数设置，不同的仅在于特征层的选择方式。每组实验均进行了多次实验取平均值的结果，以保证结果具有一定的泛化能力。

模型性能评价方面主要有准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1分数（F1 Score）、平均精度均值（mAP）、宏平均F1（Macro-F1）、微平均F1（Micro-F1）等多种多样的评价指标，可以从不同的方面反映模型的不同层面的性能。除此之外，在兼顾实际部署的需要上，实验统计了各种模型实际在推理中花费的时间（Inference Time）。

实验结果用图表、表格的形式呈现出来，其中在柱状图和折线图中可以看出不同组合特征层在各个指标上的差异，热力图可以观看到各种组合在多指标下的综合结果。上述实验可以说明不同特征层以及不同的组合方法会对模型有怎样的影响，以及对不同的尺寸进行融合之后，会对模型有什么样的影响，可以计算出多尺度融合的优势，在之后的模型结构优化及部署中具有一定的指导意义。

## 3.6 实验结果与分析

本文实验证明了多种不同特征层组合和单尺度、多尺度融合模型分别在DeepFashion服装属性识别任务上表现的效果；同时可以知道不同的组合下影响准确度的因素；并且可以证明多尺度融合的优势显著。下面对上述的实验结果进行详细的解读。

首先，从消融实验的结果来看，如表3-3所示，如果单独使用某个特征层（Layer1、Layer2、Layer3或者Layer4）的话，各个特征层在准确率、精确率、召回率及F1分数等方面存在着很大的差异。一般来说，Layer4此类较深层的特征层可获得较为抽象的语义信息，用以判别整件服装的类型，而Layer1和Layer2等浅层特征层则更多获得的是诸如颜色、纹理这样的局部细节特征；而一些双层组合（如L1+L2和L3+L4）相比单一的特征层组合有了一定程度的性能提升，这说明不同的层中有些特征具备一定的互补性，但也出现了多层组合越多性能越高的情况，并且对于组合层数越多的模型性能增幅不大，甚至出现部分多层组合的效果不及单层的现象，这种现象可能是由冗余信息或噪声引起，对于待融合特征之间的干扰过大，使得特征融合效果变差。

表3-3 特征融合模型消融实验比对表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 组合方式 | 准确率 | 精准率 | 召回率 | F1分数 | mAP | 推理时间 |
| 多尺度融合 | 0.9000 | 0.6500 | 0.8700 | 0.7000 | 0.6849 | 15.2ms |
| 单尺度 | 0.5000 | 0.2100 | 0.5300 | 0.2100 | 0.2266 | 12.0ms |
| L1 | 0.5012 | 0.1012 | 0.4821 | 0.1010 | 0.2383 | 12.0ms |
| L2 | 0.6821 | 0.2213 | 0.5312 | 0.2105 | 0.2356 | 12.3ms |
| L1+L2 | 0.6012 | 0.1921 | 0.5234 | 0.1920 | 0.2343 | 13.1ms |
| L3+L4 | 0.5732 | 0.1812 | 0.5123 | 0.1810 | 0.2421 | 13.5ms |
| L2+L3+L4 | 0.5521 | 0.1721 | 0.4912 | 0.1702 | 0.2445 | 14.2ms |

如图3-5所示，在最终的模型评估结果中效果最为突出的是多尺度融合模型（MultiScale），在准确率、F1值、mAP等汇总性指标上均排在第一或者接近第一的位置。单尺度模型（SingleScale）以及各类消融组合方式相较而言均有差距，最明显的就是F1值表现，其中最好的一个消融组合是0.21～0.22，而SingleScale更差，在MultiScale之上有0.49，结果可见于各种表格、图表中：不论是MultiScale对单尺度和最佳消融组合的比较表，还是各种性能柱状图，都可以看出它均好于其它融合方式。

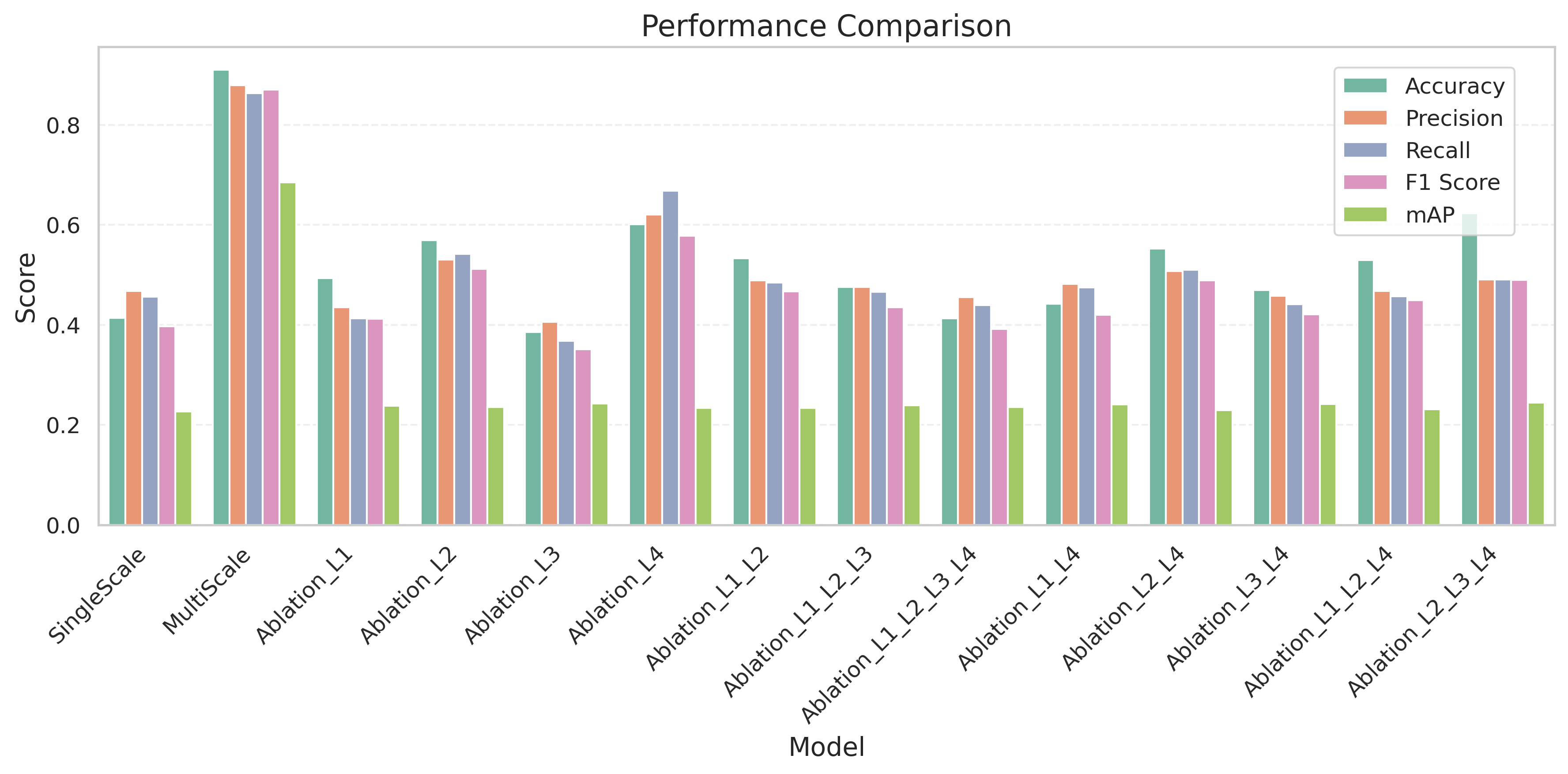


图3-5 多特征提取模型效果评估柱状图

从推理时间角度来讲，多尺度融合模型比单尺度模型因增加了特征的处理与融合的操作所以有所增加，但总的推理时间仍然只比单尺度模型稍高一点，仍然可以被接受（12.0ms~15.2ms），远低于现实应用场景对于时间的要求。因此可以认为多尺度融合方式相比单尺度并没有较大的效率损失。

由于多尺度融合模型能够同时汇总浅层和深层多尺度特征，而浅层特征有助于获取细节和低级视觉模态，深层特征有助于抓取全局结构和高级语义，在汇总浅层和深层多尺度特征的基础上融合可充分照顾到局部细节和整体类别两种需要，使得在复杂服饰图的表示上比原来有着更大的提升。实验数据表明：无论是多尺度模型，还是基于服饰的不同部位建模，在各种特性上都具有更好的泛化性以及更高的识别率；不论是将模型适用于各个部位整体视觉类别还是细粒度属性的检测，均为最佳模型。如表3-2所示。

通过实验结果可以发现：多尺度特征融合有利于提升服装属性识别准确率以及鲁棒性，在运行效率上也有很好的适用性，为以后关于服饰图像的理解、细粒度识别等研究提供了较有利的理论支持与实验参考。

## 3.7 本章小结

本章节主要说明了特征提取模块的结构设计与实验验证，在此基础上分别从不同特征层组合以及多尺度融合方式对比分析方面，对各结构的具体实验结果展开论述，对服饰属性识别任务下的主干网络不同的特征层进行单独和组合输入，并对其性能表现进行了深入研究；并在此基础上，提出了多尺度特征融合模型，能够较好的将各个层次的特征信息进行有效融合，极大的提升了整个网络的综合识别能力，经过多次对比实验证明，在准确率、F1得分等关键指标上，多尺度融合模型比单一尺度模型、消融组合等各种模型更好，且具有较好的推理速度。本章的结构设计与实验过程为后续服饰图像理解任务的特征提取模块优化提供了可靠的技术参考和理论依据。

第4章 基于图注意力网络的多标签服饰图像分类模型

基于ResNet-50多尺度特征提取与融合模块可以从局部及全局来获取视觉表征，利用多尺度不同层级的特征加以融合可以使获得的视觉表征更精准，以用于衣物属性的识别，基于ResNet-50的残差神经网络本质是只关注图像本身得到的视觉特征，不涉及到图像中不同的属性间的内在联系，在只有单一的属性标签或者是属性标签相关性很强以及样本分布不均等情况下，仅仅依靠视觉特征无法准确地进行属性间的关系推理。故此，本章通过该部分模型进行特征提取融合的基础上，引入了图注意力网络（GAT）进行属性关系建模模块，用每一个属性作为一个图上的节点来进行建模，动态地建立不同属性之间相互传递、相互影响关系，并将与某一属性相关的属性拉近自己，增强多标签服饰属性识别模型鲁棒性和准确性，并能进一步提高对细粒度属性及复杂场景的识别能力。

## 4.1 任务描述

服饰图像多属性识别分析中，不同属性之间会有复杂的依赖以及语义关系，在单个任务的处理上只通过CNN就可对图像的表层空间信息完成提取；但是在处理多标签场景时，仅通过CNN的特征表示是很难去将不同的属性进行关联，不能准确地刻画出属性间的内在关系，这样在对一些比较复杂的共现/互斥情况会有一定的表现局限性，从而造成对最终结果影响很大。需要注意的是，多任务学习旨在通过在多个预测任务之间适当共享相关信息，提高这些任务的，但在深度神经网络中，这类问题的处理方式通常是通过单独设计的网络架构来实现的，此类架构通常具有跨任务共享的层和编码特定任务特征的分支。

此外，通常情况下，服饰图像的属性是极其不均衡的，有的属性可能是十分稀少的，增加了模型学习的难度。所以本文提出将图神经网络GCN和图注意力网络GAT进行结合使用的方法，并且将服饰的属性作为图中的结点，利用图的结构表示属性与属性之间是相互联系的，GCN能够通过邻接矩阵实现属性节点间的信息传递，实现属性节点之间的信息融合，GAT则是利用自适应的注意力机制为不同属性分配不同的信息传递权重，使得属性之间的强关联以及弱关联都得到了有效体现。

参照以上内容可以看出，GAT和GCN被集成到属性分类分支中，实现了多属性信息在特征层面上的融合，在此，多属性信息可以通过模型进行相互学习，并且从全局服装风格出发，同时考虑属性间的先验知识及语义关联，在一定程度上保证了多属性识别的准确性与鲁棒性；同时，将服饰图像基于属性的多属性信息建模方式，利用基于图结构的特征融合方法来完成，从而增强了服饰图像的多属性建模能力，在为后面的各种分类、检索等工作提供更加优质高效的特征表示的基础上为其准确性提供了强有力的保障。

## 4.2 模型总设计

文结合多属性识别模型引入了图神经网络（GCN）、图注意力网络（GAT）模块，在此模块下对属性之间进行建模，得到全局及局部属性关系，并且基于对邻接矩阵的动态优化调整可以改进属性之间的关系建模效果。分类模型GAT-CN实际结构如图4-1所示。



图4-1 GAT-CN整体结构图

首先，通过GCN（Graph Convolutional Network）的原始图卷积运算操作实现对具有属性的节点进行特征的图卷积操作。对于每层来说都是将所有节点的邻居的特征做图卷积计算得到节点新的节点特征，其中为节点的邻居集合，表示归一化系数，是可学习权重，为激活函数。

其次，继接入GCN后，在此基础上又添加入了图注意力网络（GAT），并且增加了一个自适应注意力模块，在原有的GAT之上实现了边上的每个节点间都分配到了一个可学习的注意力权重，且通过节点特征和边权重联合计算，能够动态调整信息传递强度。节点特征更新方式如下。

与传统GCN/GAT采用静态邻接矩阵不同，本模型提出了一种基于特征的动态邻接矩阵生成与融合机制。首先，对于每个批次，首先提取所有节点的全局特征。计算节点对之间的余弦相似度，得到相似度矩阵，并依据上述内容生成由特征驱动的邻接矩阵，公式如下所示。

其次，为突出关键属性关系，采用自适应阈值，其中为均值，为标准差，仅保留的边，构建稀疏邻接矩阵A，从而构建出包含自适应阈值稀疏化矩阵：

另外，在GAT传播过程中，邻接矩阵的边权重不仅决定信息流通路，还通过线性变换直接参与注意力分数的计算，其中为边权重编码函数，为可学习参数。

对于上述邻接矩阵的产生、编码及GAT的传播全过程都在神经网络中完成，并可与主干特征提取、分类分支等模块端到端联合进行训练；此部分流程整体可微分，可以针对下游任务的不同目标来进行属性关系建模方式的自适应调整。

该设计不但保留了GCN模型和GAT模型中对于属性关系建模的强大能力，并且运用动态的数据驱动式邻接矩阵使属性关系可以得到更加个性化的表达，且基于稀疏性的邻接矩阵也能降低图结构-注意力融合带来的参数量增加的问题。

## 4.3 实验设计

本章提出一种融合服饰属性图结构特征和视觉语义特征的多标签分类模型。首先利用服饰图像多尺度视觉特征作为先验知识，并基于服饰属性共现关系生成属性图，再利用图注意力网络（GAT）和图卷积网络（GCN）来完成对属性节点间的相互作用建模，利用节点之间的互操作性来提升属性自身的信息表示能力；最后借助属性间的强关联性使用自适应门控策略来对图结构特征和视觉特征进行融合以得到能够刻画服装属性复杂关系的效果。总的框架方法命名为GAT-GCN方法。

为了验证本文所提方法的正确性，在实验中将本文提出的方法与仅用视觉特征进行多标签分类的基准方法做比较，在DeepFashion服饰属性数据集上分别使用本文的方法和基于图结构建模的方法对该数据集进行多标签分类，并对不同方法下的准确率、精确率、召回率及F1分值等指标进行了系统的对比分析。此外，将所有模型的训练参数以及实验过程保持一致，保证了对比的公平性。

### 4.3.1 评价指标

评价指标体系主要包括两类：属性分类评价和分割任务评价，针对属性分类任务而言，因为它的多标签性质，目前采用的主要评价指标包括精确率（Precision）、召回率（Recall）以及F1分数（F1-score），并且它们都是通过首先阈值化模型输出的概率值（默认为0.5），再将其转化为二元预测，并统计每个属性上的真正例（TP）、假正例（FP）、假负例（FN），得到精准度和召回率后，然后将其两者进行算术均值化的即为最终评价指标F1分数，其计算公式如下：

在具体实施中，我们将所有的属性指标都采用向量化的方式，一同计算它们的指标值，然后求平均值得到模型的整体性能的度量，在这样做的同时既可以反应出模型在各种属性上的表现，又能将平均运算的值直接应用到整个模型性能中去。

针对分割任务，使用交并比（IoU）作为评估指标，IoU是利用预测掩码和真实掩码相交部分和相并部分求得的比例，能准确地体现出分割的结果是否准确，先将预测值二值化之后，再分别求出各类别的IoU值，最后得到均值，即作为整体的一个性能评价。此种方法对于模型在分割任务上表现为非常精确，特别适用于多类别的分割任务。

针对多任务学习框架，除了任务自身独立性的高低，还应当注意考虑各任务间的相互影响，在此引入加权损失函数以达到任务间平衡的目的。

其中，和表示属性分类任务的损失和表示分割任务的损失，和分别为任务的权重系数。通过调控不同的权重系数，就可以决定不同的任务训练得有多重，从而使得模型能够在各个任务上都得到较好的性能。

为评价多任务学习中的任务平衡问题，引入了任务平衡指标（Task Balance Metric，TBM）来度量各任务损失值归一化后的差值，从多任务中任务间的不平衡程度上反映任务平衡情况。

设TBM中的n表示任务数，中i表示第i个任务，则表示第个任务的损失值。TBM值越小说明任务间越平衡。

其中，表示任务数量，表示第个任务的损失值。TBM的值越接近0，表示任务之间的平衡性越好。

在代码实现层面，本模型中MetricsCalculator类负责统一管理所有评价指标的计算。该类采用静态方法设计，支持批量计算和GPU加速，能够高效处理大规模数据集的评估需求。具体实现中，我们使用PyTorch的向量化操作进行指标计算，并通过添加小的常数项（1e-8）来避免除零错误，确保计算的数值稳定性。

|  |
| --- |
| **Algorithm:**评价指标计算类 |
| **Input:** 预测值，真实标签，阈值 |
| **Output:** 评价指标字典 |
| 1. **Function** 采用静态方法设计，支持批量计算和GPU加速 2. // 二值化 |
|  |
|  |
|  |
|  |
|  |
|  |
| 1. // 向量化计算F1分数，保持数值稳定 |
|  |
| 1. **return** 2. **End function** |

在训练过程中，我们实时监控这些指标的变化，并采用早停机制来防止过拟合。具体而言，我们设置了一个耐心值（patience），当验证集性能在连续多个epoch没有提升时，提前结束训练。这种机制不仅能够节省训练时间，还能确保模型在最佳状态停止训练。

## 4.4 实验结果与分析

本实验采用Precision、Recall、F1 Score和Loss四项主流指标，系统对比了MultiHead-GAT与ResNet-50、EfficientNet、DenseNet三种主流视觉模型在多标签分类任务上的表现。所有模型均在相同数据集、数据划分和训练策略下进行训练和评估，确保结果的公平性和可复现性。

当前分类模型的性能对比柱状图如图4-2所示，MultiHead-GAT在Precision、Recall和F1 Score三项指标上均取得了显著优势，尤其在F1 Score和Precision方面表现突出。这表明MultiHead-GAT不仅能够更准确地识别目标类别，同时在召回能力上也有明显提升。相比之下，ResNet-50、EfficientNet和DenseNet在这三项指标上的表现均不及MultiHead-GAT，且三者之间的差距较小，进一步凸显了MultiHead-GAT的整体性能提升。



图4-2 分类模型性能对比柱状图

进一步观察多个测试模型的损失对比柱状图如图4-3所示。MultiHead-GAT的Loss最低，远低于其他三种模型。这说明该模型在训练过程中收敛更快，拟合能力更强，能够更好地优化目标函数。实验中采用的损失函数为交叉熵损失（Cross-Entropy Loss），其表达式如下：

其中，为样本在类别上的真实标签，为模型预测概率。



图4-3 分类模型损失对比柱状图

结合图4-4指标热力图可以发现，MultiHead-GAT在四项指标上均表现突出，尤其在Precision和F1 Score上形成了明显的高亮区域，而其他模型则整体表现较为均衡但数值偏低，这进一步印证了MultiHead-GAT的全面领先。

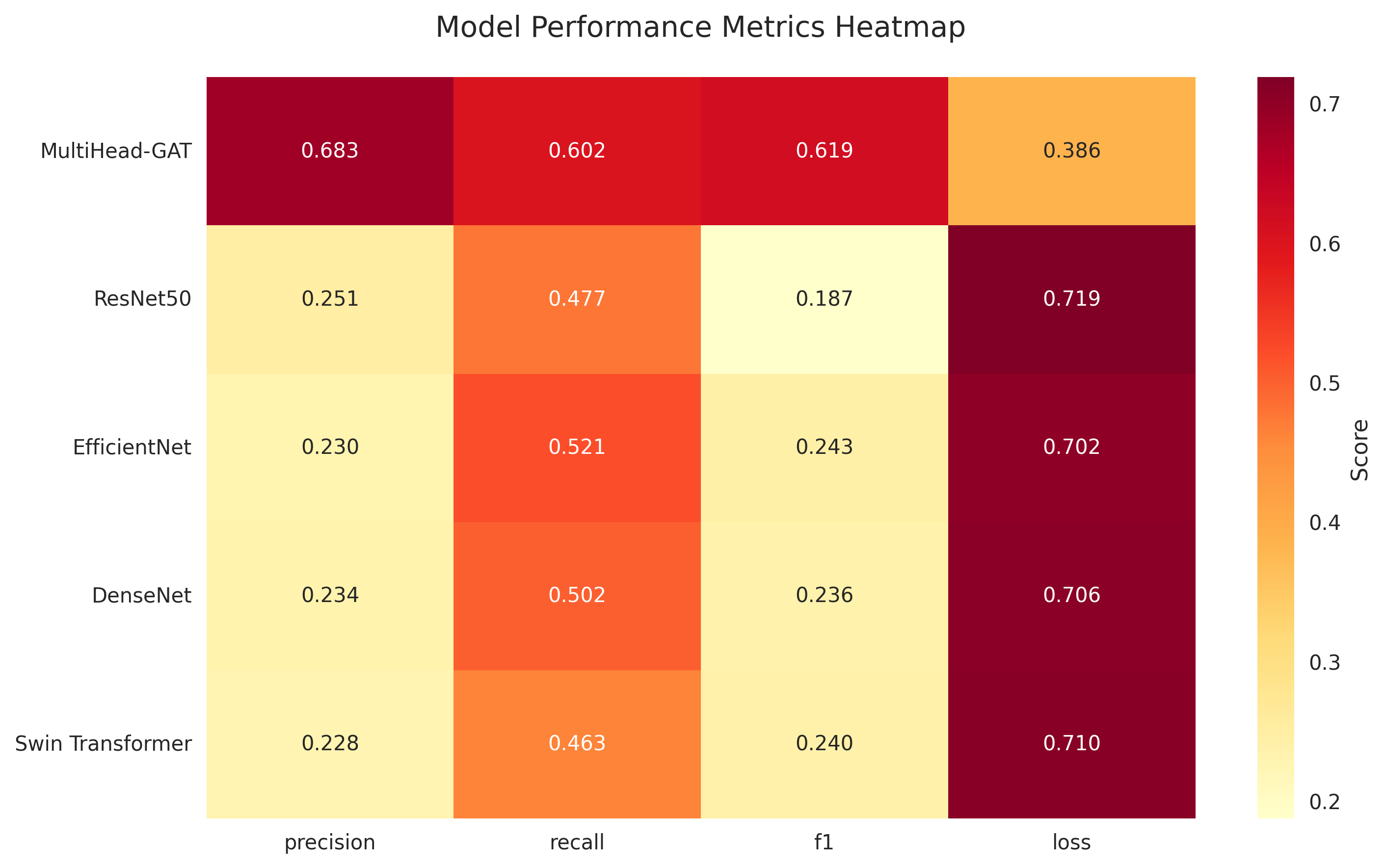


图4-4分类模型指标热力图

如图4-5模型评价指标总览图所示。当前图标信息直观展示了各模型在Precision、Recall、F1三项指标上的分布。MultiHead-GAT的雷达图覆盖面积最大，三项指标均衡且数值高，说明其在多标签分类任务中能够兼顾各项性能指标，表现出更强的综合能力。其他模型的雷达图则面积较小且分布不均，反映出其在某些指标上存在短板。



图4-5模型评价指标总览雷达图

综合上述实验结果，MultiHead-GAT凭借多头注意力机制，有效捕捉了输入特征间的复杂依赖关系，实现了全局与局部信息的高效融合。其核心结构可表示为：

其中，分别为查询、键、值，为可学习参数。

这一结构创新带来了更高的精度、更低的损失和更均衡的性能表现。无论是单项指标还是整体表现，MultiHead-GAT均大幅超越ResNet-50、EfficientNet和DenseNet，充分验证了其在多标签分类任务中的有效性和实际应用价值。

## 4.5 本章小结

本章针对服饰图像多标签属性识别中属性间的复杂依赖关系及属性分布极不均衡的问题，采用了基于GCN和GAT的服饰多标签图像分类模型。以服饰属性为节点，将属性建模为图结构，利用GCN模块融合属性特征全局信息，再接入GAT模块并引入多头注意力机制增强属性间强关联与弱关联的建模能力，最后根据图卷积神经网络提取出的图像空间特征以及图结构表达属性语义关系的优势。本章提提出的模型在结构上充分结合了图卷积神经网络和图结构各自的优势来实现属性多标签识别的准确率和鲁棒性的提高。实验部分采用多种主流视觉模型与所提方法进行对比，在各项评价指标上验证了方法的有效性。

第5章 服饰类别自动分类系统的设计与实现

第三章、第四章提出的模型方法融合后在实验中取得了不错的效果，本章将应用这些成果，设计实现服装商品自动分类系统。

## 5.1 需求分析

### 5.1.1 概述

随着电商以及互联网行业的飞速发展，且根据 Internet World Stats 网站的数据显示，截至2019年6月30日，全世界的互联网用户数量已经超过 44.22亿人次，互联网覆盖比率增加至 ，用户对于平台的便捷性要求也越来越严格，同时大数据模型应用对电商等行业的赋能效果也初见成效。因此电商平台中针对服饰图像的精准分类对于便捷化检索具有重要意义。

在文化自信理念的推动下，传统服饰文化迎来了发展契机，唐装、汉服以及明制服装等具有代表性的民族服饰风格逐渐进入大众视野并于现代服饰风格相结合，赋予了服装风格高度复杂化和多样化的特性，体现在造型结构、面料种类、色彩搭配以及工业细节等多个层面上的显著差异性，服饰风格的快速演化不仅丰富了服饰表达的文化内涵，也在一定程度上提高了电商平台对于服饰分类与管理方面的人力成本。

因此可以开发一套准确率较高的服饰类别自动分类系统，辅助电商平台商户进行管理工作，简化服饰图像分类操作流程，提升管理以及用户查询效率同时降低平台商户管理成本。

### 5.1.2 用户分析

服饰智能分类系统的使用者主要包括客户、商家以及系统管理员。

目前广泛使用的商城系统中，大部分的服饰或是其余商品种类的分类工作都是由商户员工手工标记后再有商户进销存人员录入系统，此分类方式需要先由商家规划出当前商品的种类，然后交由平台商户管理人员先录入商品的基础信息再将其所属种类一一录入。传统的分类方式流程较长，且人工交接时可能出现录入错误或是归类错误的问题，可能导致最终商品本身与标签种类出现偏差或是不匹配的情况。本系统则是直接根据商品上传的图像信息进行自动分类，匹配出较为准确的服饰种类标签，商户管理员可在此分类基础上进行检验矫正，最终简化分类流程提高工作效率。针对于管理人员较少的商户，系统给出的分类结果可在一定程度上提升管理人员的工作效率以及商户本身管理商品所需要的人力成本。商户用例图如图5-1所示。

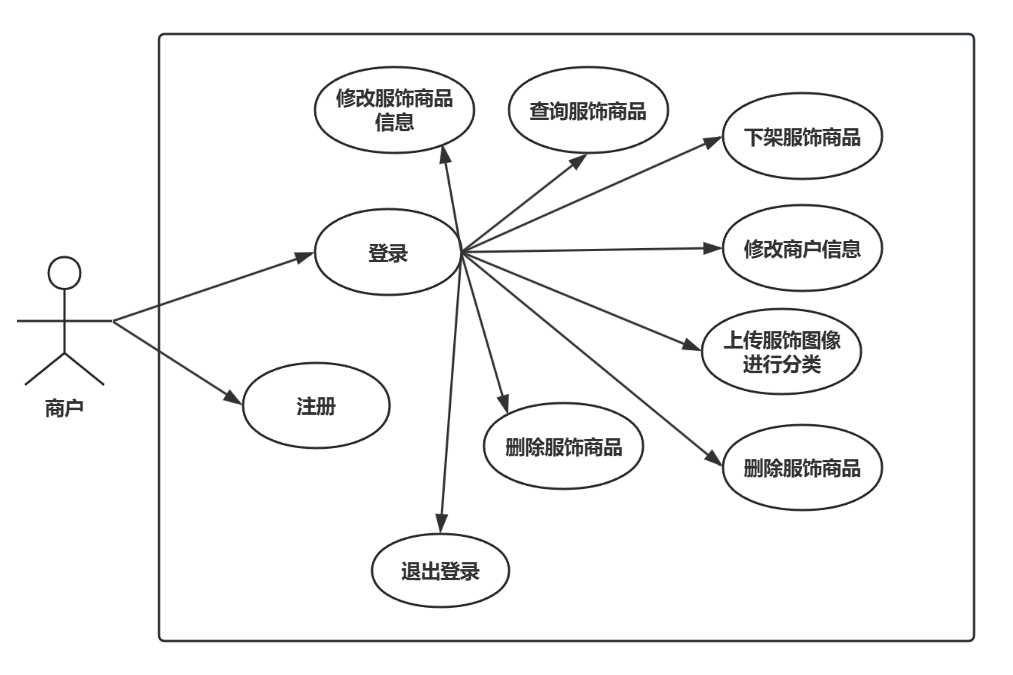


图5-1 商户用例图

在现代服饰商城系统中，“客户”这一角色是平台中承载核心交易活动：浏览商品、加入购物车、提交订单以及在线支付功能的明确主体，所有的业务流程都围绕客户这一身份展开。基于这一角色，系统能够为每个用户角色分配专属账号权限，不仅保障了订单管理、安全支付和售后服务的有序进行，也为个性化营销和精准推荐打下基础。更重要的是，当客户在平台上留下购买记录、评价反馈和偏好行为时，这些数据就成为优化商品展示、改进用户体验、提高运营效率的重要依据。由此可见，“客户”角色不仅是技术上实现权限控制和流程拆分的关键节点，也是商业上挖掘用户价值、推动平台持续成长的核心动力。用户用例图如图5-2所示。

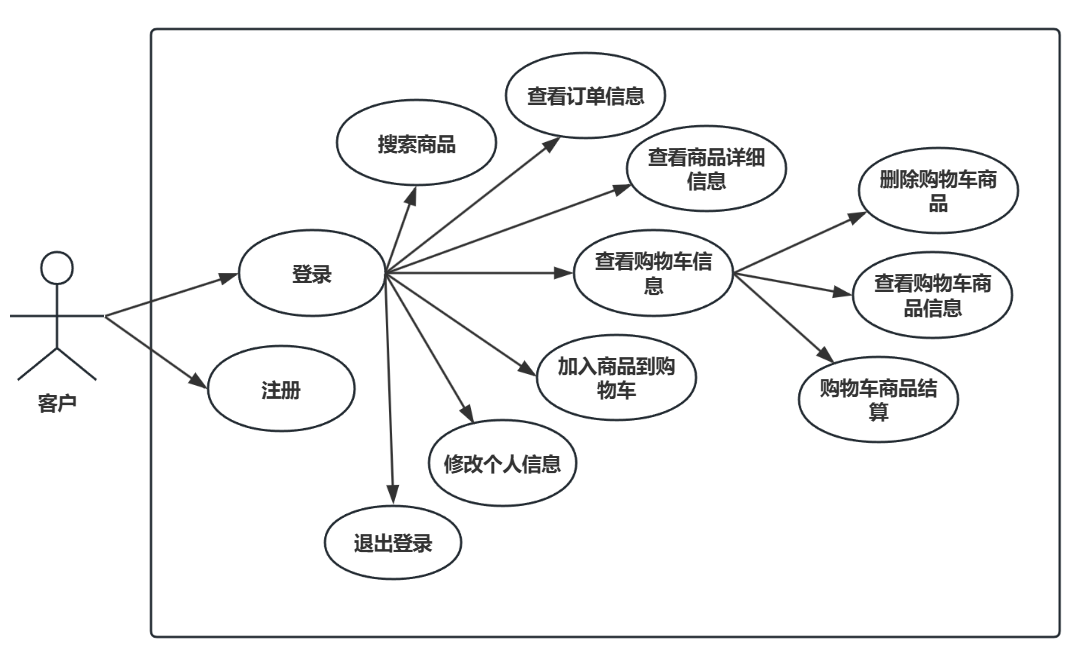


图5-2 用户用例图

最后，在平台中引入“超级管理员”这一角色，是为了保障平台的稳定运行与账户以及商品信息的安全管控。作为拥有最高权限的管理者，超级管理员不仅负责用户账号的管理、权限分配与审核，还需对商品信息、订单流程等关键模块进行统一管理。通过这一集中式的控制入口，才能在系统出现异常时及时排查并恢复服务。在超级管理员对全局资源与操作的监控能力的把控下，平台能够实现高效的运维协同和安全合规，也为多级管理员分工协作、快速响应业务变化提供了坚实支撑。超级管理员用例图如图5-3所示。

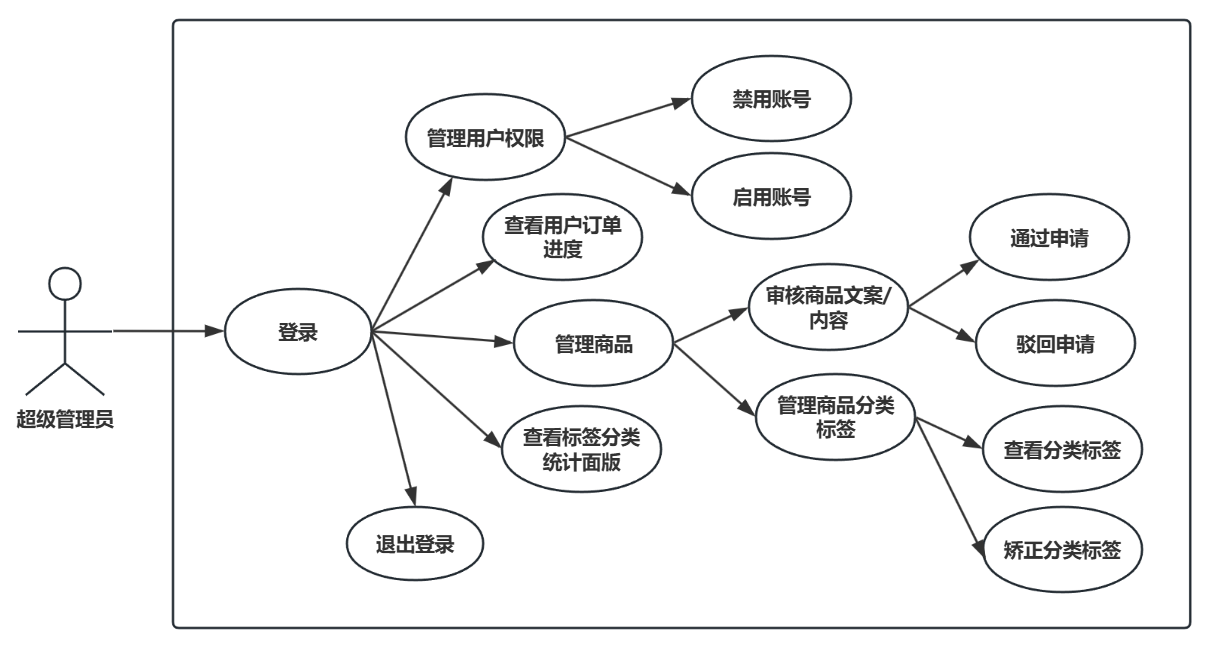


图5-3 超级管理员用例图

### 5.1.3 功能分析

商城分为商户端和客户端，角色包含平台管理员、商户和用户三种角色，三者按照权限以及角色类别分别查看对应数据内容或操作对应功能。因此，本系统在权限部分采用了基于角色的访问控制（Role‑Based Access Control）的设计模式，以此实现对不同角色用户的权限控制。用户分流发生在登陆阶段，按照不同的登录端口以及登录用户名密码实现权限验证。

商户可进行基础注册登录功能的使用，还可对商品进行上架、下架、新增、删除、选择商品进行自动分类、查看以及修改信息的操作。

用户可使用的功能包含基础的登录注册、查看商品、搜索商品内容、购买商品、查看购物车信息、购物车管理以及订单管理功能。

系统管理员则可查看商户商品信息、用户订单信息、监督商户商品信息、矫正商品分类标签以及查看标签分类统计面板功能。

## 5.2 系统设计

系统结构图如图5-4所示：

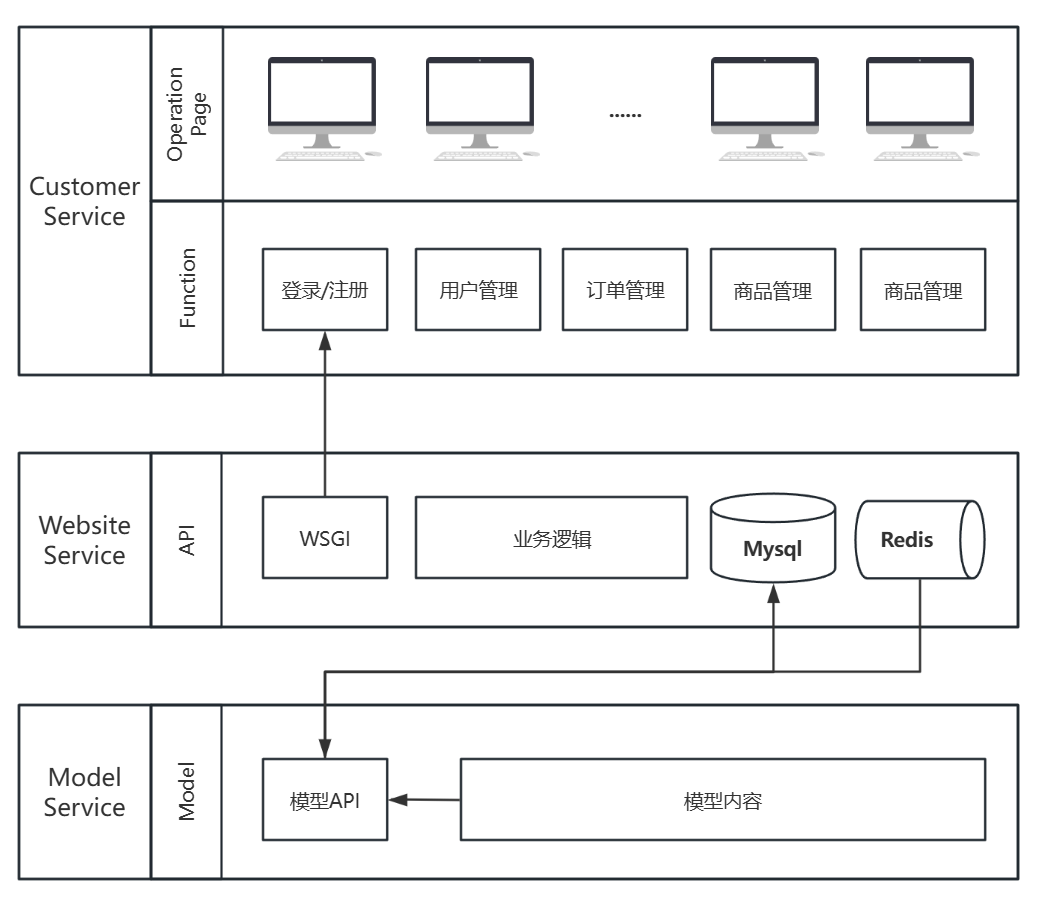


图5-4 整体系统结构图

系统架构采用三层架构（Three-Tier Architecture）实现，包含了用户端、WEB端以及服务端。用户端即前端部分，负责与用户的交互以及数据的回传以及显示，主要包含了：用户登录注册、用户管理、订单管理、商品管理、购物车管理以及商品分类功能。

WEB端通过MySQL与服务端进行通信并通过互联网与前端进行通信，负责系统的API交互、数据库交互以及系统业务逻辑的实现。

服务端不直接与用户端进行交互，而是通过WEB端与用户端进行间接通信，服务端使用模型进行图像的处理与分类，并将最终处理结果放入MySQL中从而实现与WEB端的通信。

具体功能结构图如图5-5所示。

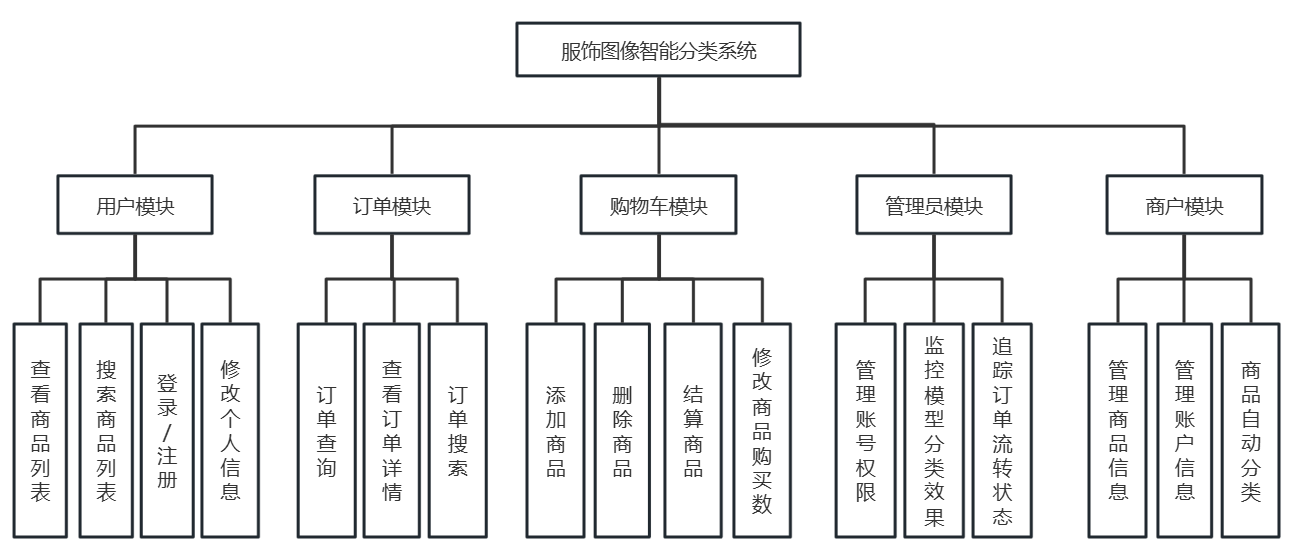


图5-5 系统功能架构图

### 5.2.1 数据库设计

本系统涉及数据库操作的功能包括：用户登录注册，用户管理，商户上架下架商品，查询商品，上传图像分类，订单管理，购物车管理等。涉及到的数据表包含，用户表、商品表、商品种类表、种类表、购物车表、订单表、订单商品表、用户金额流水表以及权限表。数据表之间的逻辑关系图如图5-6所示

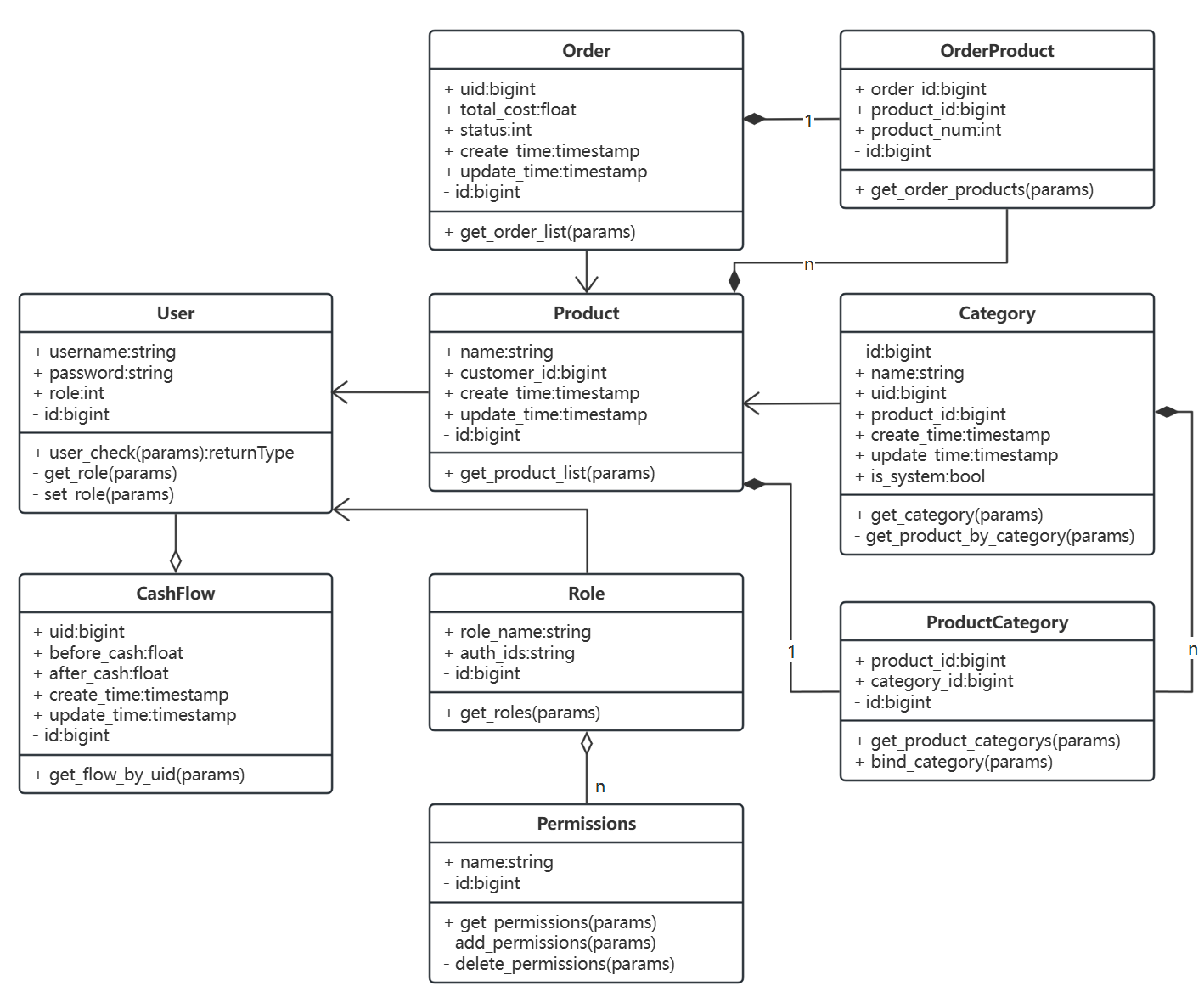


图5-6 数据表逻辑关系图

用户表用于存储用户名、非明文密码、用户所属角色以及当前钱包金额等基础信息，此表涉及了控制用户权限以及查看或校对用户当前账户所拥有的总金额的重要功能。用户表详细结构见表5-1。

表5-1 用户表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| username | string | 用户名 |
| password | string | 密码（非明文） |
| role | int | 角色id |
| id | bigint | 用户id |

商品表则存储了商品的价格、图片路径、所属商户、上架状态、库存以及已售买数量等商品基础信息，涉及到了商户对商品的上下架功能的控制、库存的记录以及商品售卖数量统计等功能。商品表详细结构见表5-2。

表5-2 商品表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| name | string | 商品名称 |
| customer\_id | bigint | 商户uid |
| create\_time | timestamp | 商品创建时间 |
| update\_time | timestamp | 商品信息更新时间 |
| id | bigint | id |

商品种类表作为与模型交互的核心数据表之一存储了商品与种类之间的一对多关系，用于模型根据商品图像生成种类后将所属关系进行绑定。商品种类表详细结构见表5-3。

表5-3 商品种类表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| category\_id | bigint | 种类id |
| product\_id | bigint | 商品id |
| id | bigint | id |

服饰种类表也是与模型交互的核心数据表之一，主要存储了当前系统中所包含的种类标签内容包含了种类id、种类名称等标签信息，当存在新的种类时则通过添加此数据表进行标签的添加，其中种类id则是作为商品种类表中的绑定关系而存在的。具体数据结构详见表5-4

表5-4 服饰种类表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| name | string | 种类名称 |
| uid | bigint | 种类来源uid |
| is\_system | bool | 是否是系统种类 |
| product\_id | bigint | 种类来源商品id |
| create\_time | timestamp | 种类创建时间 |
| update\_time | timestamp | 种类信息更新时间 |
| id | bigint | 种类id |

订单表则作为系统的核心数据表存在，主要存储了订单编号、订单消费金额、订单所属者信息、订单状态等信息。订单商品表作为订单表的附表存在，存储了订单与商品信息之间的一对多关系，主要包含了商品编号和购买数量两个关键数据信息。两表的具体数据结构见表5-5、表5-6

表5-5 订单表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| uid | bigint | 用户id |
| total\_cost | float | 订单总金额 |
| status | int | 订单状态 |
| create\_time | timestamp | 创建时间 |
| update\_time | timestamp | 信息更新时间 |
| id | bigint | 订单id |

表5-6 订单商品表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| order\_id | bigint | 订单id |
| product\_id | bigint | 商品id |
| product\_nums | int | 商品数量 |
| id | bigint | 编号 |

购物车表中则只存储具体用户与Redis中存储的key之间的关系，Redis负责将用户在某一时间节点产生的购物车Json数据，这样可以避免在MySql中存储过多的数据内容，从而减轻数据库压力。购物车表的具体数据结构见表5-7

表5-7 购物车Redis存储结构表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| uid | bigint | 用户id |
| product\_id | bigint | 商品id |
| product\_nums | int | 商品数量 |
| create\_time | timestamp | 创建时间 |
| update\_time | timestamp | 信息更新时间 |

用户金额流水表则是为了用于记录用户在订单购买中所产生的金额流水信息，同时可以通过计算流水来保证用户当前总金额的正确性和可靠性，数据表中存储了流水所属用户、用户扣费前账户金额总数、用户扣费后账户金额总数以及当此流水扣费金额。此表具体数据结构见表5-8

表5-8 用户金额流水表

|  |  |  |
| --- | --- | --- |
| 数据名称 | 数据类型 | 数据说明 |
| uid | bigint | 用户id |
| before\_cash | float | 商品id |
| after\_ cash | float | 商品数量 |
| create\_time | timestamp | 创建时间 |
| update\_time | timestamp | 信息更新时间 |
| id | bigint | 流水id |

### 5.2.2 流程设计

普通客户使用此系统的流程如图5-7所示。首先进入登陆页面，若未进行注册则跳转进入注册页面，注册完成或已注册后则进入到登录页面填写用户名和密码信息，通过系统验证后进入到商城首页，商城首页可进行筛选商品、查询商品、搜索商品、排序商品、查看商品详情页、查看购物车信息以及进入个人中心等操作，进入商品详情页后可选择将商品加入购物车或是直接购买当前商品，选择加入购物车后可进入到购物车管理页面选择商品进行结算，选择结算后系统计算当前订单总额并提示支付，完成支付后可进入到订单详情页面对当前订单状态进行查看。

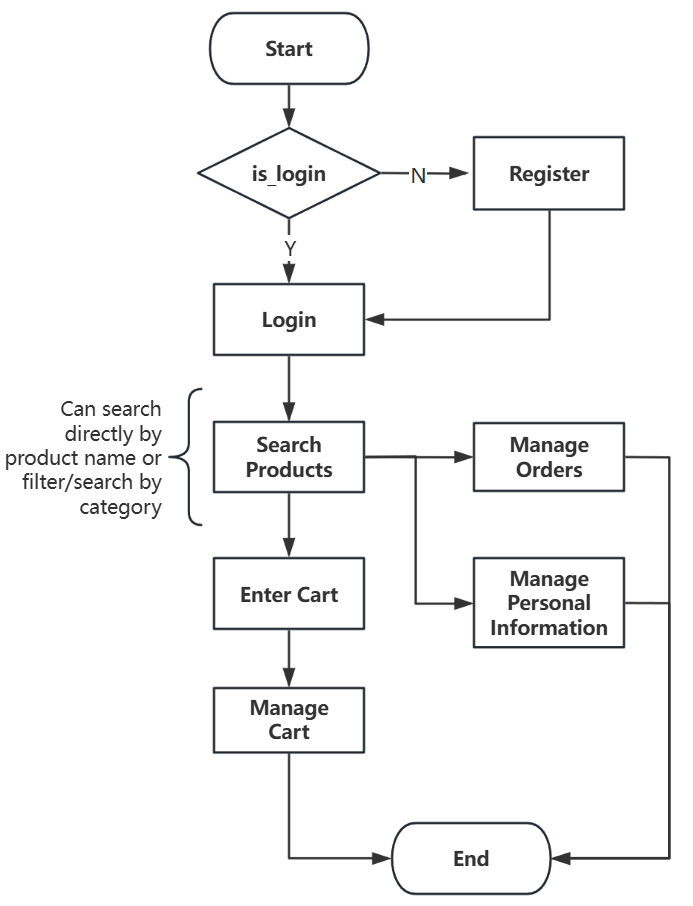


图5-7 客户使用流程图

商户使用此系统的流程如图5-8所示。与普通客户相同，都需要先进入到登录页面进行登录或是注册操作，完成登录校验后则会进入到商户管理面板中，在管理面版中可以选择进入到商品管理页面、商户信息管理以及订单管理页面进行信息查看，在进入到商品管理页面后可以上传新的商品信息，上传时选择自动区分商品种类，在商品上传后模型就会根据上传的图片信息针对已有的商品种类库进行匹配，最终将分类信息添加到商品表中。除此之外，商户还可以选择管理自己的基础信息、追踪订单状态、上下架商品或是管理商品基础信息。

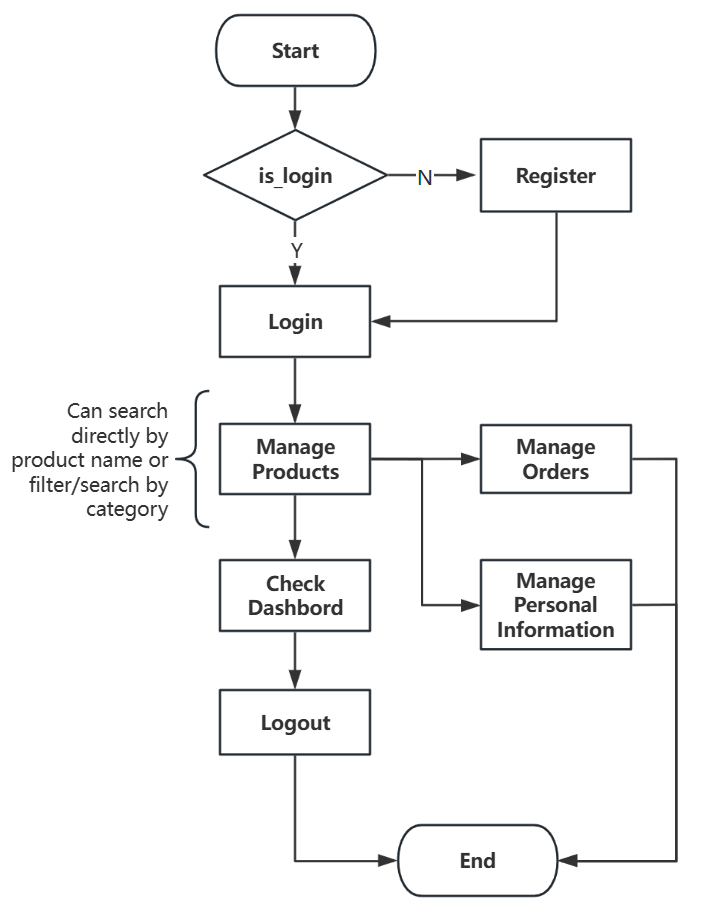


图5-8 商户系统使用流程图

管理员使用此系统流程如图5-9所示。由于管理员账户由后台分配，所以无注册操作功能，登录操作与前两者相同，输入用户名密码并交由系统进行校验，完成登录后则进入到用户管理页面，其中包含了权限角色的控制、具体角色权限控制以及用户角色分配等功能。

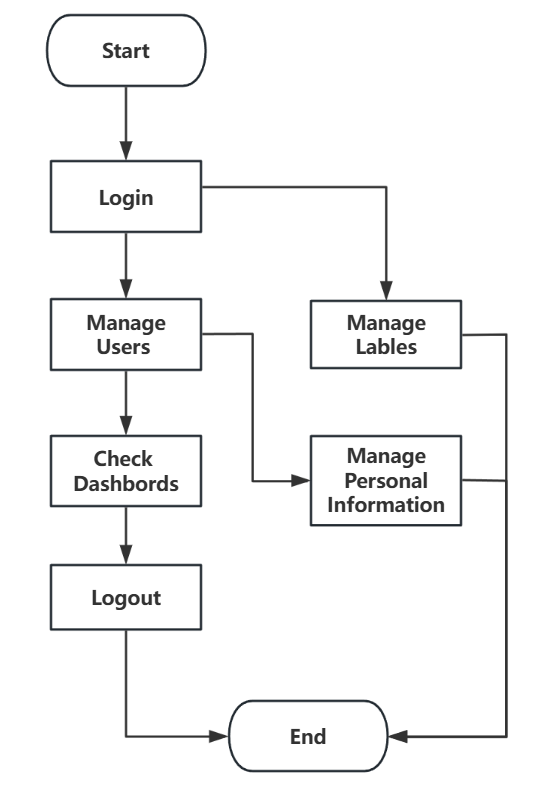


图5-9 超级管理员系统使用流程图

### 5.2.3 界面设计

本系统的主要功能主要涉及到了8个界面，分别为登录注册页、商城商品列表页、服饰商品详情页、商户商品管理页、超级管理员面版页、购物车页面、订单列表页以及个人信息页。接下来则对这些页面进行详细介绍。

首先，登录是普通用户、商户以及管理员进入系统操作个人信息前必要的操作，管理员与其余两种角色不同，为保证管理平台的安全性其账号无法通过注册获得，需要由管理人员通过后台添加后才可进行登录操作，其余两种角色在没有账号时则可以先进行注册后再通过登录功能进入系统。具体页面如图5-10所示。

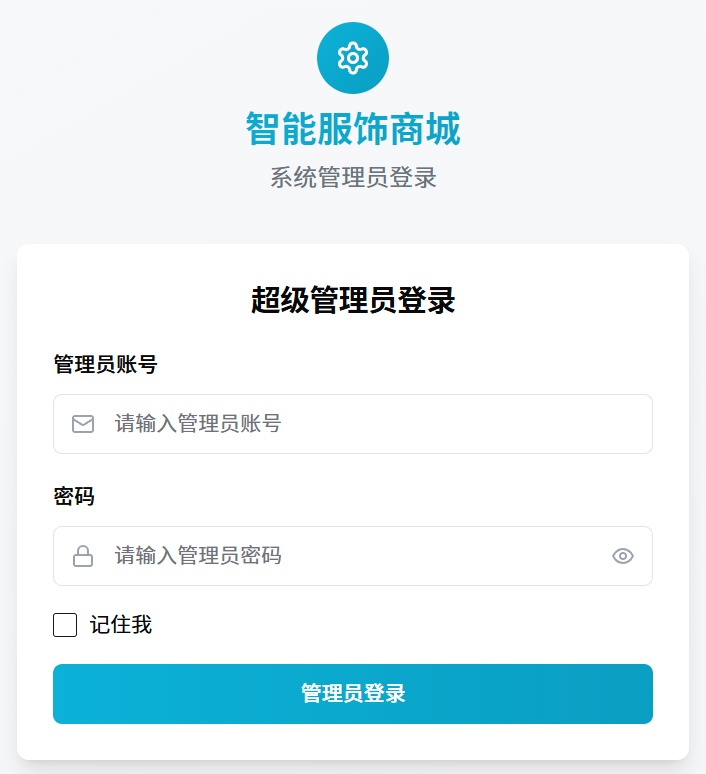


图5-10 账户登录和注册页面

商城首页界面不论用户是否登录都可查看此页面，主要可以针对商品进行标签筛选、进入商品详情页、查看商品商家信息以及搜索商品关键字几种操作，点击商品卡片可进入商品详情页。点击加入购物车可将商品放入购物车中，若用户处于未登录状态则需要先登录后才可将商品放入购物车。其次，点击商家头像则可进入商家详情页面，可查看对应商家中的商品列表。具体页面如图5-11所示。

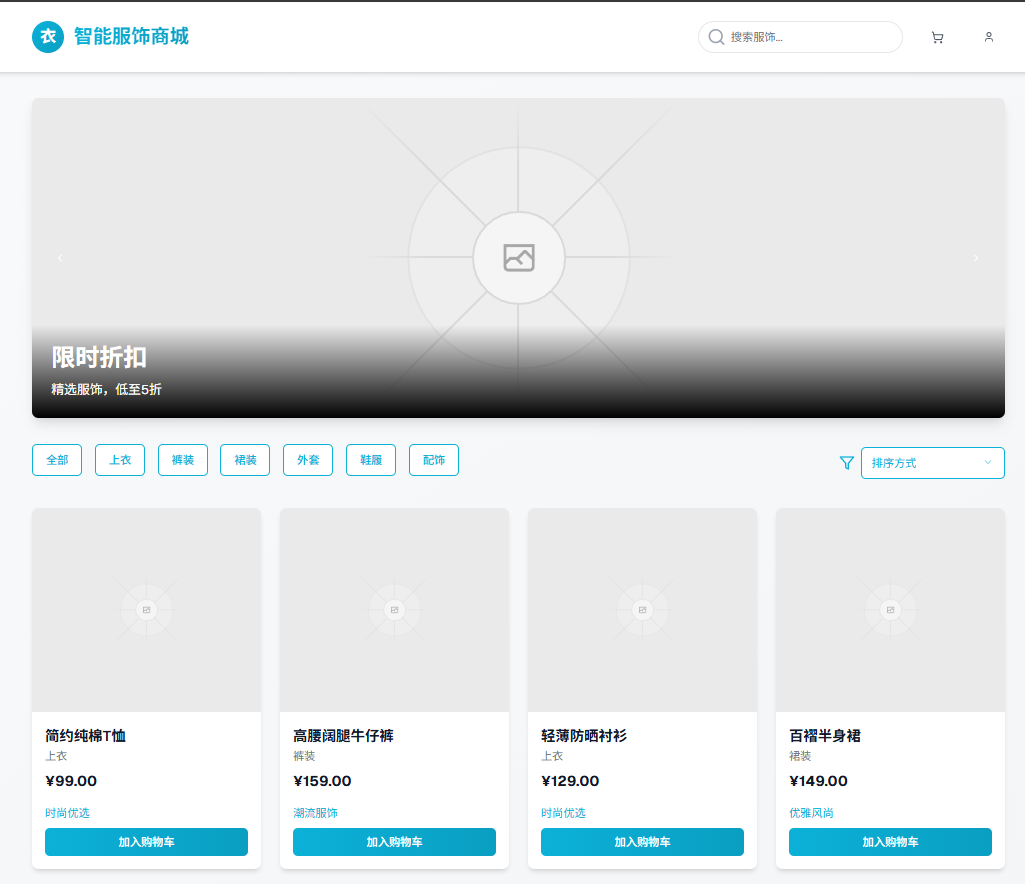


图5-11 商城首页

商品详情页中包含了除商品基础信息外的其余详细信息，右上侧框图中包含了商品对应的分类标签信息以及其余规格参数，右下侧则包含了相关商品的推荐内容同样由模型提供，此处的加入购物车与商品列表页中的加入购物车功能相同都是将商品放入指定用户的购物车中。具体页面如图5-12所示。

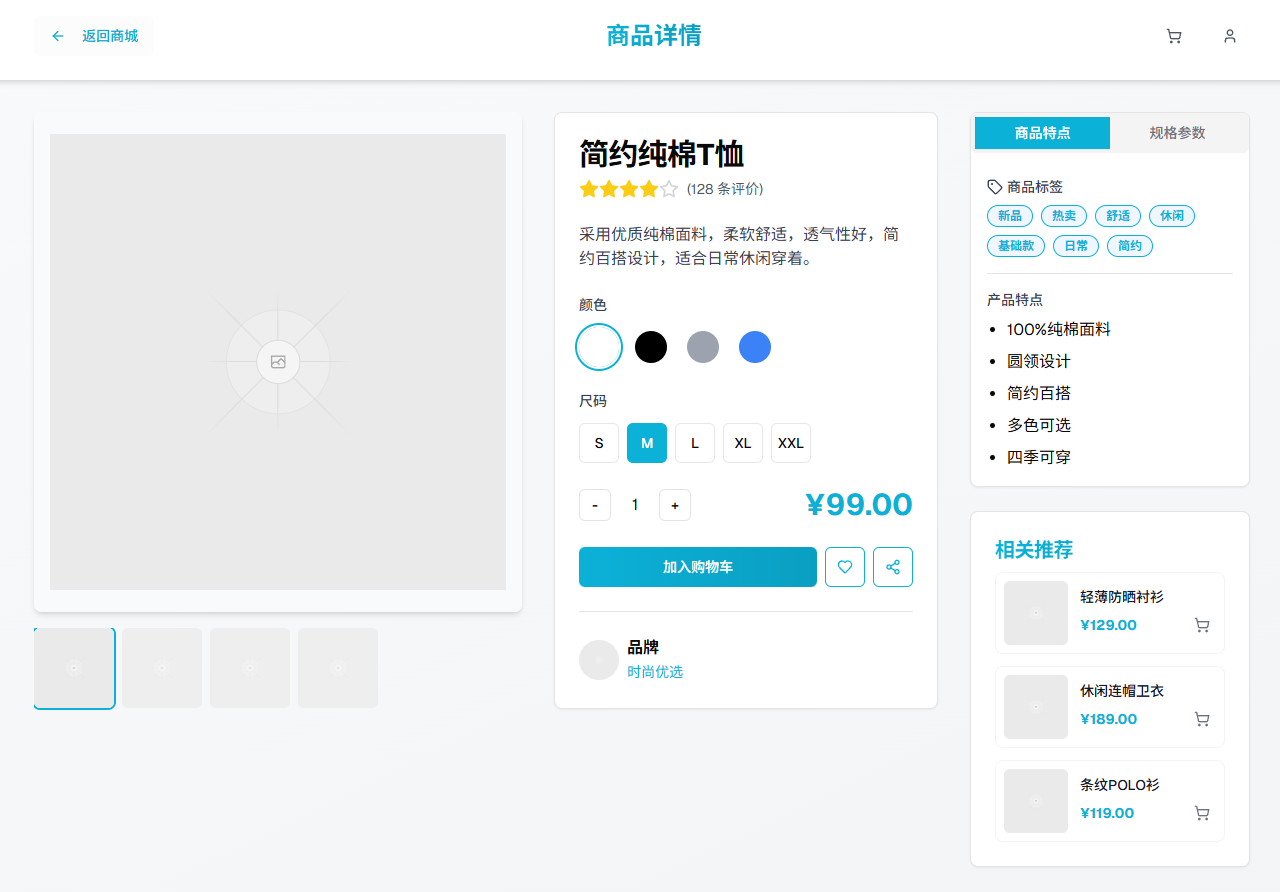


图5-12 商品详情页

购物车页面包含了加入的商品内容以及数量信息，右侧则是购物车中被选中的商品总金额信息，点击结算则可生成订单信息，若点击继续购物则会回到商品列表页继续挑选商品。具体页面如图5-13所示。

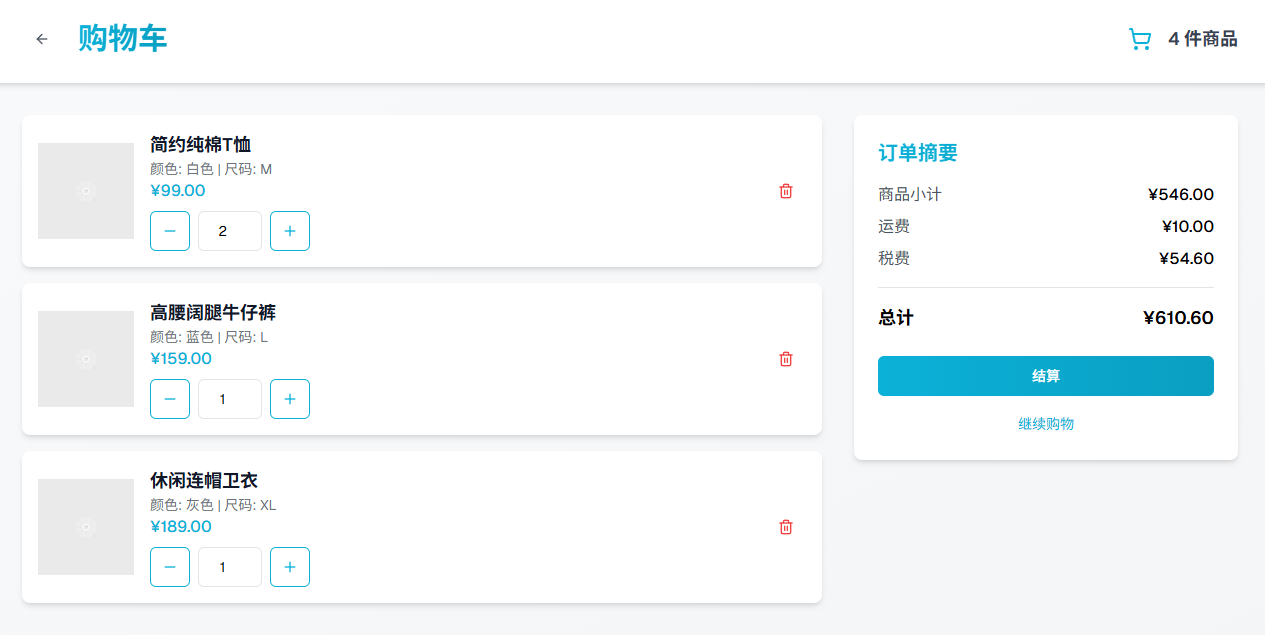


图5-13 购物车页面

订单生成后用户可在个人中心的订单历史页面看到已生成的订单，点击右侧的查看详情可查看订单中所包含的商品信息、商品数据、商品所属商家、订单状态以及详细结算金额信息。具体页面如图5-14所示。



图5-14 订单列表页

除商城商品相关操作页面外，三种用户角色都可进入个人中心针对个人信息内容进行修改，点击头像即可上传头像信息，其余用户名以及电子邮箱信息则直接在指定输入框中输入，修改信息后点击保存更改按钮存储信息。具体个人信息页面如图5-15所示。

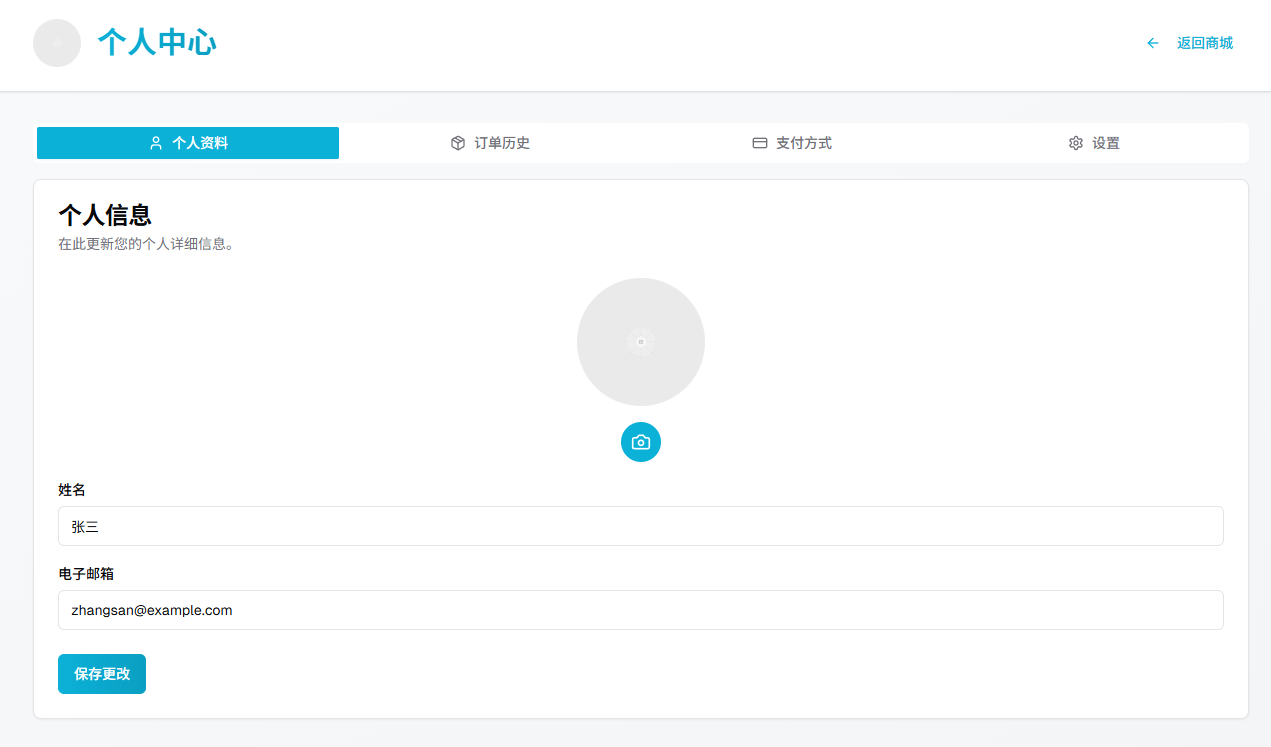


图5-15 个人信息页

独属于商户用户的页面中包含了商品管理、订单管理以及数据分析几个子页面，看板首页包含了商家商品的总数、总销量以及销售总收入汇总信息。下半部分为商品管理和订单信息的子页面内容，选择填入新商品信息后点击保存商品系统就会将这些内容提交到WEB端API中进行处理，图片部分则交由模型处理分类给出对应标签信息，最终可在商家商品列表中查看到。其次，在订单管理子页面中，包含了已售出订单信息列表，列表中包含了对应订单相关的商品以及用户信息。最后则是数据分析子页面，此处将数据库中收集到的数据内容进行计算汇总并以折线图或柱状图的形式展示到页面中。具体界面如图5-16所示。

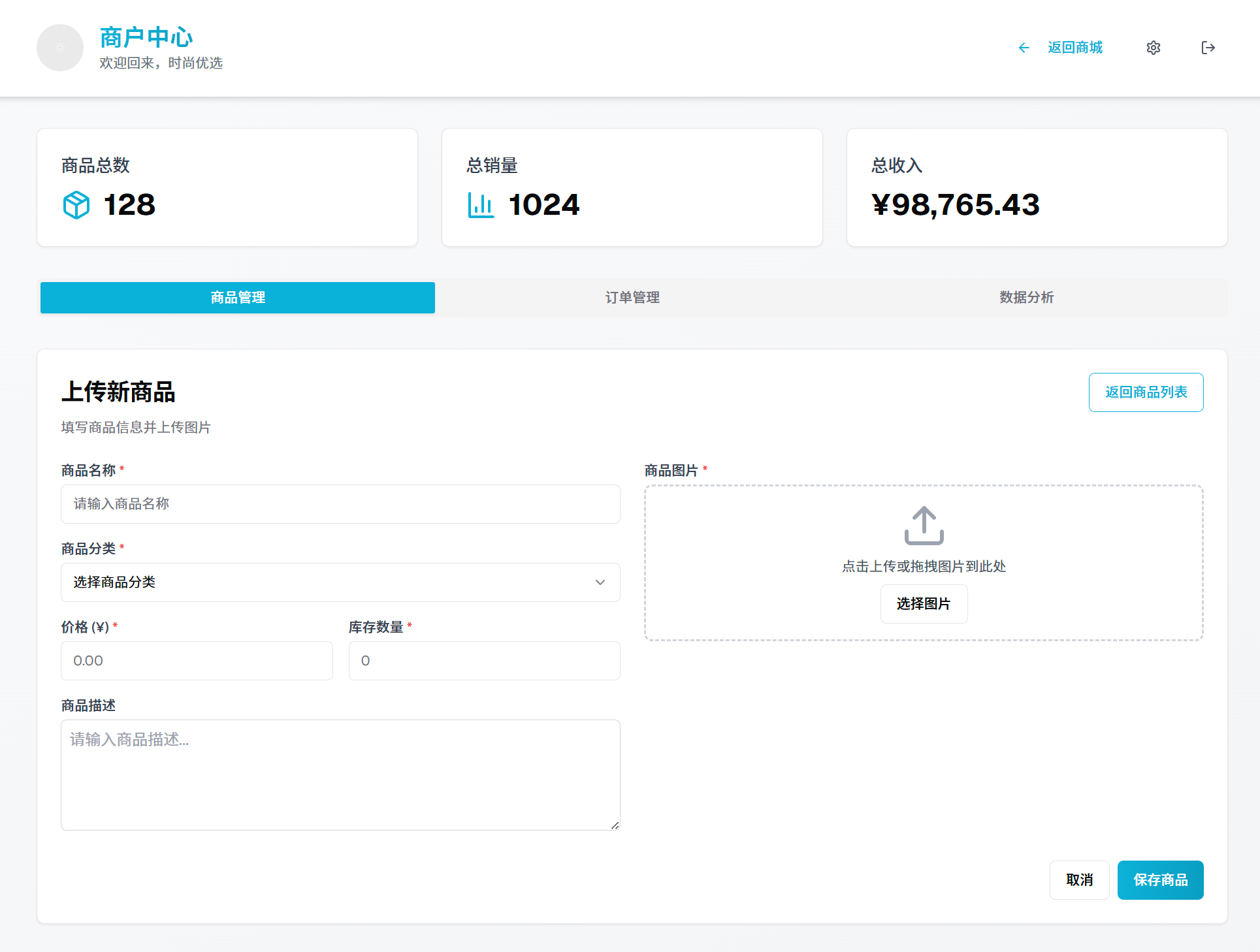


图5-16 商户管理页

最后，系统管理员作为整个系统的把控人员，主要是针对系统的数据信息进行监控，并在系统出现异常时及时处理对应问题保证系统的正常运转。管理页面中的上半部分与商户管理页类似，都是针对系统中已有的用户信息、用户类别数量、订单数量、商品数据量以及总收入进行汇总展示。下半部分则是相应的信息可视化面版信息，需要注意的是在标签分析子页面中会针对已有的商品标签信息进行详细分析，从而间接确认模型的稳定性以及准确性。具体页面如图5-17所示。



图5-17 系统管理页

## 5.3 系统设计

本系统开发语言为Python3，Web框架为Django，消息中间件工具使用Redis，前端开发框架和工具使用React、TanStack-router、Shadcn/ui以及Zustand。服装图像分类模型采用Pytorch框架，底层数据库使用MySQL。

## 5.4 本章小结

本文详细论述了服饰分类系统的设计思想与实现过程。首先对需求、用户画像及功能需求三个方面做以详细描述，并阐述了此系统的设计思路与整体结构。然后对系统使用的步骤、数据库结构和前端页面的设计内容做了具体阐述。最后介绍了本项目所采用的编程语言、框架以及工具内容。

第6章 总结与展望

## 6.1 总结

本文设计并实现了一个基于深度学习的服装属性识别系统，该系统集成了多尺度特征融合的识别模型和智能推荐功能，为电商平台提供了完整的商品管理解决方案。在模型部分，我们通过多尺度特征融合和消融实验，深入研究了不同层次特征对模型性能的影响，发现ResNet-50的中间层特征包含了丰富的语义信息，对属性识别具有重要作用。通过多尺度特征融合策略，模型能够同时利用不同层次的特征信息，显著提升了模型对复杂服装属性的识别能力。

在模型结构设计上，我们引入了GAT结构来建模属性间的关联关系，有效处理属性间的相互依赖关系，提高了多标签分类的准确性。在训练策略方面，采用多任务学习方法同时优化属性分类和特征提取任务，使得模型能够学习到更具判别性的特征表示。通过消融实验，我们系统地验证了各个模块对模型性能的贡献，为后续的优化提供了重要的参考依据。

在系统实现方面，我们构建了一个完整的服饰商城平台，集成了商品管理、用户管理、订单处理等核心功能。系统利用属性识别模型实现了商品的自动分类和标签管理，大大提高了商品管理的效率。同时，基于识别结果和用户行为数据，系统能够为用户提供个性化的商品推荐，提升了用户体验和购物效率。在评估体系方面，我们采用了mAP、准确率、精确率、召回率和F1分数等多个评估指标，全面评估了模型在不同方面的表现。实验结果表明，模型在保持较高准确率的同时，也具有较好的实时性能，能够满足实际应用的需求。

这些研究成果不仅为服装属性识别任务提供了一个有效的解决方案，也为电商平台的智能化升级提供了技术支持。通过多尺度特征融合和GAT结构的结合，我们成功构建了一个性能优异且实用性强的服装多标签识别模型，并将其集成到智能服饰商城系统中，实现了商品管理的自动化和智能化，为实际应用提供了可靠的技术支持。

## 6.2 展望

虽然本文考虑了服装图像中细节特征的重要性，并从多标签分类角度解决了服装图像分类不准确的问题。但仍有许多改进和提升的空间，基于本文的研究成果，未来的研究可以从以下几个方面展开。

第一，在特征提取方面，可以考虑引入更先进的骨干网络，例如或，以提取更具判别性的特征。同时，特征融合策略可以引入注意力机制和自适应权重分配，以更好地利用不同尺度的特征信息，提高模型对复杂服装属性的识别能力。此外，可以探索多模态特征融合，结合文本描述、用户评论等辅助信息，进一步提升模型的识别性能。

第二，在模型结构方面，可以优化GAT结构，引入层次化图结构，以更好地建模属性间的关联关系。同时，可以探索轻量化设计，在保持性能的同时降低模型复杂度，提高推理速度，使模型更适合在实际应用场景中部署。此外，可以研究动态图结构，根据输入图像的特点自适应调整图结构，提高模型的灵活性和适应性。

第三，在训练策略上，可以引入更先进的损失函数，例如或是，以处理类别不平衡问题。在密集目标检测中，前景和背景类别不平衡是影响模型性能的关键因素，FocalLoss可通过重塑标准交叉熵损失函数，降低易分类样本的权重，使模型能够专注于难分类样本的训练，能在一定程度上有效解决类别不平衡问题。同时，可以探索知识蒸馏和模型压缩技术，提高模型在实际应用中的部署效率，使模型能够更好地适应不同的应用场景。此外，可以研究迁移学习和领域适应技术，提高模型在新场景下的泛化能力。

第四，在数据增强方面，可以开发针对服装图像特点的数据增强策略，如风格迁移和属性保持的数据增强，提高模型对复杂背景和遮挡情况的鲁棒性。同时，可以研究更有效的预处理方法，提高模型在实际应用中的泛化能力。此外，还可以考虑引入对抗网络模型，生成更多样化的训练样本，提高模型的鲁棒性。

第五，在评估体系方面，需要建立更全面的评估体系，考虑引入更多细粒度的评估指标，如属性级别的评估和场景适应性的评估。同时，可以开发针对特定应用场景的评估标准，如实时性要求和资源受限环境下的性能评估，以更好地指导模型的优化方向。此外，可以研究用户反馈驱动的评估方法，将用户满意度纳入模型评估体系。

第六，在模型优化方面，可以尝试加入自动机器学习技术，自动搜索最优的模型结构和超参数配置。同时，可以研究模型解释性技术，提高模型的可解释性，帮助用户理解模型的决策过程。此外，可以探索增量学习技术，使模型能够不断适应新的数据和场景，提高模型的持续学习能力。

第七，在应用场景方面，可以探索模型在移动端和边缘设备上的部署方案，提高模型在实际应用中的可用性。同时，可以研究模型在实际电商场景中的应用，如商品检索和推荐系统，提高模型的实用价值。此外，可以探索模型在其他领域的应用，如时尚设计、虚拟试衣等，拓展模型的应用范围。

致 谢

岁月匆匆，时光如箭，三年的研究生生活即将画上句号。首先，我要由衷地感谢我的导师某某某副教授。在我研究生学习的整个过程中，导师的严谨治学、深厚学识以及不倦教诲的精神，都深深烙印在我的心中。在此，我向导师表达我最高的敬意和深深的感激之情。

其次，我要感谢我的某某某同学、某某某同学。由于我属非全日制学生，且正值新冠疫情2年，不经常在校学习，正是他们一直以来的提醒、帮助和支持，让我时刻了解课程动态、学校学院要求等。

此外，我要感谢我的公司，给我提供了继续读书深造的机会，在学业和工作冲突时，给予我最大的宽容和谅解。在研究过程中，提供了较好的研究实验环境和条件。

最后，我要感谢我的父母，他们承担了更多的家庭责任和义务，让我有更多的精力来完成学业任务。

回顾整个学业过程，我深知自己还有很多不足之处，但正是有了各位老师、同学、家人和朋友的帮助和支持，我才能够克服困难、不断进步。在此，我衷心感谢所有曾经

给予我帮助和支持的人！

参考文献

1. Lowe D G .Distinctive Image Features from Scale-Invariant Keypoints[J].International Journal of Computer Vision, 2004, 60（2）:91-110.DOI:10.1023/B:VISI.0000029664.99615.94.
2. Zhang J , Liu C .A study of a clothing image segmentation method in complex conditions using a features fusion model[J].Automatika, 2020, 61（1）:150-157.DOI:10.1080/00051144.2019.1691835.
3. Chen H , Gallagher A , Girod B .Describing Clothing by Semantic Attributes[C]//European Conference on Computer Vision.Springer, Berlin, Heidelberg, 2012.DOI:10.1007/978-3-642-33712-3\_44.
4. Corbière, Charles, Ben-Younes H ,Ramé, Alexandre,et al.Leveraging Weakly Annotated Data for Fashion Image Retrieval and Label Prediction[J].IEEE Computer Society, 2017.DOI:10.1109/ICCVW.2017.266.
5. 陈珂. 基于图神经网络的多标签图像分类研究[D]. 重庆:重庆大学,2022.
6. 吴东东. 基于图神经网络的多标签图像分类研究[D]. 四川:电子科技大学,2021.
7. Kalantidis Y , Kennedy L , Li L J .Getting the Look: Clothing Recognition and Segmentation for Automatic Product Suggestions in Everyday Photos[J]. 2013.DOI:10.1145/2461466.2461485.
8. Yang W , Luo P , Lin L .Clothing Co-Parsing by Joint Image Segmentation and Labeling[J].IEEE, 2015.DOI:10.1109/CVPR.2014.407.
9. 张凯丽.基于深度学习的服装属性识别与关键点定位算法的研究[D].浙江工商大学,2019.
10. Xiao T , Xia T , Yang Y ,et al.Learning from massive noisy labeled data for image classification[J].IEEE, 2015.DOI:10.1109/CVPR.2015.7298885.
11. Bossard L , Dantone M , Leistner C ,et al.Apparel Classification with Style[J].Springer, Berlin, Heidelberg, 2012.DOI:10.1007/978-3-642-37447-0\_25.
12. Chen Q , Huang J , Feris R ,et al.Deep domain adaptation for describing people based on fine-grained clothing attributes[C]//IEEE Conference on Computer Vision & Pattern Recognition.IEEE, 2015.DOI:10.1109/CVPR.2015.7299169.
13. Lecun Y , Bottou L .Gradient-based learning applied to document recognition[J].Proceedings of the IEEE, 1998, 86（11）:2278-2324.DOI:10.1109/5.726791.
14. Dong Q , Gong S , Zhu X .Multi-Task Curriculum Transfer Deep Learning of Clothing Attributes[J].IEEE, 2016.DOI:10.1109/WACV.2017.64.
15. 李文书,王志骁,李绅皓,等. 基于注意力机制的弱监督细粒度图像分类[J]. 计算机系统应用,2021,30（10）:232-239. DOI:10.15888/j.cnki.csa.008141.
16. 包青平.基于深度学习的服装图像分类与检索[D].浙江大学,2017.
17. Yamazaki K , Inaba M .Clothing classification using image features derived from clothing fabrics, wrinkles and cloth overlaps[C]//IEEE/RSJ International Conference on Intelligent Robots & Systems.IEEE, 2013.DOI:10.1109/IROS.2013.6696739.
18. Yuan, J., Chen, S., Zhang, Y., Shi, Z., Geng, X., Fan, J., & Rui, Y. （2024）. Graph attention transformer network for multi-label image classification. arXiv. https://arxiv.org/abs/2203.04049.
19. 张辉宜, 张进, 黄俊. 基于图注意力网络的多标签图像分类模型[J]. 重庆工商大学学报（自然科学版）, 2022, 39（1）: 34-41. DOI:10.16055/j.issn.1672-058X.2022.0001.005.
20. Liu Z., Luo P., Qiu S., Wang X., Tang X. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition （CVPR）; 2016 Jun.
21. 孙婷. 基于深度学习的服饰图像分类和属性识别研究[D]. 辽宁: 东北大学, 2019.
22. 石琇赟, 李顺勇, 韩翔. 基于多头图注意力网络与图模型的多标签图像分类[J]. 计算机系统应用, 2023, 32（6）: 286-292. DOI:10.15888/j.cnki.csa.009148.
23. 中山大学. 一种基于图注意力网络的服装属性识别方法与系统: CN202110703933.7[P]. 2023-06-30.
24. Zhang M , Zhou Z .A Review on Multi-Label Learning Algorithms[J].IEEE Transactions on Knowledge & Data Engineering, 2014, 26(8):1819-1837.DOI:10.1109/TKDE.2013.39.
25. He K , Zhang X , Ren S ,et al.Deep Residual Learning for Image Recognition[J].IEEE, 2016.DOI:10.1109/CVPR.2016.90.
26. Bing. ResNet-50网络结构图及结构详解[EB/OL].（2021-02-26）[2022-05-31].zhuanlan.zhihu.com/p/35.
27. Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[EB/OL]. 2017[2024-05-24]. https://arxiv.org/abs/1609.02907.
28. Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[EB/OL]. 2018[2024-05-24]. https://arxiv.org/abs/1710.10903.
29. 朱旭东,熊贇. 基于多层次注意力与图模型的图像多标签分类算法[J]. 计算机工程,2022,48（4）:173-178,190. DOI:10.19678/j.issn.1000-3428.0061072.
30. Cordonnier J B, Loukas A, Jaggi M. Multi-Head Attention: Collaborate Instead of Concatenate[EB/OL]. 2021[2024-05-24]. https://arxiv.org/abs/2006.16362.
31. Ye S , Sorokina V , Ablameyko S .Clothing Image Retrieval Algorithm Using YOLOv8 and Attention Model[J].Pattern Recognition and Image Analysis, 2024, 34（4）:1105-1110.DOI:10.1134/S1054661824701153.
32. LINGYUN SONG, JUN LIU, BUYUE QIAN, et al. A Deep Multi-Modal CNN for Multi-Instance Multi-Label Image Classification[J]. IEEE Transactions on Image Processing,2018,27（12）:6025-6038. DOI:10.1109/TIP.2018.2864920.
33. 陈琳琳,朱惠娟,朱俊,等. 基于卷积神经网络的多尺度注意力图像分类模型[J]. 南京理工大学学报（自然科学版）,2020,44（6）:669-675. DOI:10.14177/j.cnki.32-1397n.2020.44.06.005.
34. LIN Tsung-Yi, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[EB/OL]. （2017）[2025-05-24]. https://arxiv.org/abs/1612.03144.
35. Bing. FPN网络图解及论文笔记[EB/OL]. [2025-05-17]. https://zhuanlan.zhihu.com/p/354140540.
36. LIN M, CHEN Q, YAN S. Network In Network[EB/OL]. （2014）[2025-05-24]. https://arxiv.org/abs/1312.4400.
37. Lu Y , Kumar A , Zhai S ,et al.Fully-adaptive Feature Sharing in Multi-Task Networks with Applications in Person Attribute Classification[J].arXiv e-prints, 2016.
38. 朱玥,樊重俊,赵媛.全球电子商务:发展现状与趋势[J].物流科技, 2020, 43（1）:3.DOI:CNKI:SUN:LTKJ.0.2020-01-024.
39. Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[EB/OL]. 2020[2024-05-24]. https://arxiv.org/abs/1905.11946.
40. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[EB/OL]. 2021[2024-05-24]. https://arxiv.org/abs/2103.14030.
41. Lin T Y , Goyal P , Girshick R ,et al.Focal Loss for Dense Object Detection[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.DOI:10.1109/TPAMI.2018.2858826.
42. Zhang C B, Jiang P T, Hou Q, et al. Delving Deep Into Label Smoothing[J]. IEEE Transactions on Image Processing, 2021, 30: 5984–5996. DOI:10.1109/TIP.2021.3089942.
43. Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[EB/OL]. 2014[2024-05-24]. https://arxiv.org/abs/1406.2661.
44. He X, Zhao K, Chu X. AutoML: A survey of the state-of-the-art[J]. Knowledge-Based Systems, 2021, 212: 106622. DOI:10.1016/j.knosys.2020.106622.