

# 基于卷积神经网络和迁移学习的服装图像分类

高樱萍, 宋 丹, 陈玉婷

(湖南工程学院, 湖南 湘潭 411100)

**摘要:**针对传统的服装图像分类方法效率低、准确度不高等缺点,提出了一种基于 VGG16 和迁移学习的服装图像分类方法,主要采取预训练的 VGG16 网络模型,提供了更好的模型初始化,并在服装图像数据集上进行迁移学习,减少了对数据的依赖以及网络的训练时间;使用数据增强的手段来增加训练的数据量,提高了模型的泛化能力;在网络层中加入 BN 层来防止网络出现过拟合和梯度消失。结果表明,与其他传统的卷积神经网络模型相比,提出的模型有更高的准确率,能够更好地完成对服装图像的分类。

**关键词:** VGG16; 深度学习; 数据增强; 图像分类

**中图分类号:** TP391.4; TS941.19

**文献标识码:** A

**文章编号:** 1673-0356(2021)11-0048-05

随着服装电子商务的发展,越来越多的人选择网络进行线上购物,人们对服装图像识别技术的准确率的要求也在不断提高。如何更好地实现快速、准确的服装图像的分类成为当前服装研究的一个热点。传统的服装图像分类方法主要依据文本描述进行分类,提取图像的颜色、纹理、边缘等低级特征,花费了大量的人力物力且分类精度较低。

近年来,深度学习的发展为图像处理领域提供了一种新的思路,在机器学习中,图像分类的实质是对提取的图像特征进行分析,并利用神经网络将其分类的过程。卷积神经网络(Convolutional Neural Network, CNN)是目前最流行的深度神经网络学习算法之一,经过训练的卷积神经网络能够学习到图像中的特征,并且完成对图像特征的提取和分类。目前,使用卷积神经网络解决图像分类中的问题逐渐成为主流。

白琮等<sup>[1]</sup>提出了一种基于深度卷积神经网络 AlexNet 的二值化哈希图像分类方法,主要对原有的 AlexNet 网络模型进行了改进,在池化阶段采用最大一均值池化,并在全连接层采用最大值激活输出,使网络表达更精确的图像特征信息,同时在全连接层引入隐层来学习哈希编码,提高分类效率,该文提出的优化方法可以明显地提升深度卷积网络在大规模图像分类

任务上的性能。高妍等<sup>[2]</sup>针对图像分类中网络训练时间长、形变服装图像识别率低的问题,提出了一种改进 HSR-FCN 的服装图像识别分类算法,将 R-FCN 中的区域建议网络和 HyperNet 网络结合,使得 HSR-FCN 可以在更短的训练时间内达到更高的准确率。

Liu 等<sup>[3]</sup>提出了一个数据量大且标注完整的服装数据集 DeepFashion,并在此基础上提出了一种新的模型即 FashionNet,该模型通过共同预测服装属性和标签来学习服装特征,通过大量的试验证明了 FashionNet 的有效性和 DeepFashion 的实用性。谢小红等<sup>[4]</sup>为了提高服装图像的分类性能,提出了一种基于迁移学习的卷积神经网络服装图像分类方法,选取了 VGG16、VGG19、Inception\_v3 等 6 种经典的卷积神经网络在 DeepFashion 数据集上进行试验,结果表明,该方法能够提高模型的分类精度和时效性。

为了提高服装图像分类的准确率,提出了一种基于 VGG16 和迁移学习的服装图像分类方法,使用在 ImageNet 数据集上进行预训练的 VGG16 网络,然后采用迁移学习的方法,将模型迁移到对服装图像分类的问题中,并对网络进行微调,实现更高的准确率。

## 1 卷积神经网络 VGG16

VGG16 是牛津大学在 2014 年提出的一个模型<sup>[5]</sup>,它是一个比 AlexNet 网络更深层次的网络,在对图像进行识别与分类的时候,可以更准确地表达数据集的特征。VGG16 的基本结构包括输入层、卷积层、池化层、全连接层和输出层,卷积层利用卷积核进行卷积操作,提取输入图像的特征;池化层进行特征降维,

收稿日期:2021-05-14;修回日期:2021-05-26

基金项目:教育部人文社会科学研究项目(20YJA880045);湖南省研究生教育教改一般项目(湘教通[2019]293号);湖南省研究生科研创新项目(CX20190962);湖南工程学院研究生科研创新项目(YC2013)

作者简介:高樱萍(1997-),女,湖南永州人,硕士研究生,研究方向为图像处理。

降低网络复杂度,最后的全连接层和输出层对结果进行分类,VGG16网络架构如图1所示。

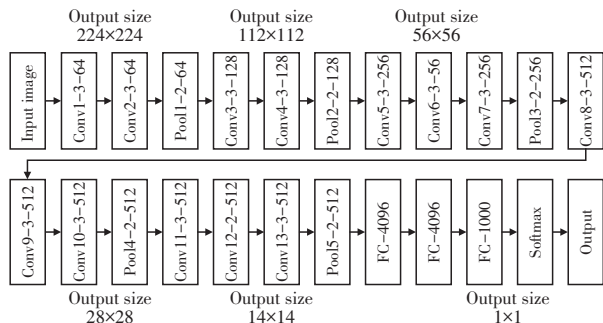


图1 VGG16网络结构及参数

图像经过预处理后进入卷积层,卷积层输入的图片尺寸为 $224 \times 224 \times 3$ ,在第一次卷积层时经64个 $3 \times 3$ 的卷积核作两次卷积后尺寸变为 $224 \times 224 \times 64$ ,池化层对输入的图片进行压缩,池化后的尺寸为 $112 \times 112 \times 64$ ,接着继续作卷积以及池化操作,后面几层卷积核的数量分别为128、256、512、512,最后得到 $7 \times 7 \times 512$ 的特征图输入到全连接层。与AlexNet网络不同的是VGG16使用的是三个 $3 \times 3$ 的卷积层,一方面,用三个非线性单元代替前一个单元,使决策函数更具判别性;另一方面,这样可以减少训练参数。

## 2 基于VGG16和迁移学习的服装图像分类

### 2.1 数据增强

在神经网络中,训练过程中分配的数据集是决定神经网络性能的因素之一,当分配的数据集数量较少时,此时神经网络的精度要低于使用大量数据集训练的神经网络。因此需要对现有的数据集进行数据增强,以增加多样性。数据增强是一种重要的图像预处理技术,主要用来产生更多的训练数据以及减少模拟过拟合<sup>[6]</sup>。数据增强可以表示为:

$$\phi: X \rightarrow Y \quad (1)$$

其中 $X$ 为原始训练集, $Y$ 为 $X$ 的增广集,通过数据增强后的训练集可以表示为:

$$X' = X \cup Y \quad (2)$$

其中 $X'$ 包含原始训练集和经过数据增强后得到的数据集。

数据增强的方式主要有翻转、平移、旋转、噪声注入等,考虑到服装图像的特性,本文主要采用水平翻转和注入高斯噪声两种数据增强方式。

#### 2.1.1 翻转

在对图像的处理中,对图像进行翻转是最常用的

方法之一,这种增强是最容易实现的,并且已经证明在CIFAR-10和ImageNet等数据集上很有用。翻转分为水平翻转与垂直翻转,较为常用的是水平翻转,水平翻转一般不会大幅度改变检测目标的整体结构。对图像进行水平翻转操作,效果图如图2所示。



图2 水平翻转

#### 2.1.2 噪声注入

训练神经网络时,如果数据集比较少,往往会出现过拟合,因此可以通过在训练期间添加噪声提高模型的泛化能力和容错率。噪声注入就是在原图中引入噪声,生成新的图片,Moreno-Barea等对UCI存储库<sup>[7]</sup>中的9个数据集注入噪声进行测试,试验结果证明向图像中添加噪声可以帮助神经网络学习更稳健的特征。对图像注入高斯噪声,效果图如图3所示。



图3 加入高斯噪声

### 2.2 迁移学习

VGG16网络在图像分类中具有较好的分类能力,但在模型训练过程中耗时较长,在小数据集中获得的准确率不高,而迁移学习的应用可以有效地解决这个问题。

迁移学习(Transfer Learning)是机器学习领域的一个重要分支,其工作原理是在ImageNet这样的大数

数据集上训练网络,然后在新的分类任务中使用这些权重作为初始权重,简单来说就是把已训练好的模型的参数迁移到新的模型上来帮助训练新模型。采用迁移学习的方法可以减少对数据的依赖,提高模型的泛化能力。迁移学习的过程如图4所示。

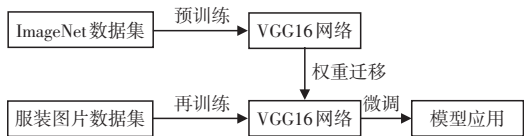


图4 迁移学习过程

文中使用的迁移学习模型分为两部分:预训练网络和迁移网络。预训练模型的参数从ImageNet数据集中获取,然后迁移到服装图像分类任务中。预训练模型的网络层主要用于特征提取和数据获取,帮助训练收敛,减少训练时间。

### 2.3 VGG16 网络优化

虽然VGG16网络模型具有识别准确的优点,但随着网络结构的加深,训练过程中模型参数的数量和计算的复杂性也增加了,导致训练时间长,训练效率低。为了保持模型提取的主要特征,提高模型的训练率,减少模型的训练时间,主要采用以下方法对VGG16模型进行改进。

深层神经网络的训练过程中,随着前一层参数的变化,每一层输入的分布也会发生变化,训练速度会变慢,并使训练带有饱和和非线性的模型变得非常困难,而在网络中加入BN(Bath Normalization)层可以解决这个问题<sup>[9]</sup>。

BN层是一种规范化神经网络的激活方法,该方法可以加快网络的收敛速度,使网络训练变得容易。BN是对同一层的输入信号进行归一化,即对与每个神经元做归一化处理,公式如式(2)所示。

$$\hat{x} = \frac{X - E(x)}{\sqrt{Var(x) + \epsilon}} \quad (3)$$

式中, $\hat{x}$ 为网络归一化的激活值, $x$ 为网络某一层的激活值, $E(x)$ 为平均值, $Var(x)$ 为方差, $\epsilon$ 为极小值。

## 3 试验和分析

### 3.1 数据集

本试验中使用的服装图片均来自香港中文大学的DeepFashion数据集,从中选取了640张包含8类服装

类型的服装图片制作自己的数据集,包含的服装类型标签为T恤、连衣裙、毛衣、短裙、牛仔裤、西服外套、无袖连衣裙、无袖上衣,每类各80张。为了减少训练阶段图像集的过拟合,采用水平翻转以及添加高斯噪声对原始图像进行处理,使得最后每种类型的样本集扩展到240张,总样本集为2560张,其中80%的数据集被划分到训练集,用于模型训练和参数学习,20%作为验证数据集,用于对模型进行优化。随机选取的部分数据集中的图片如图5所示。



图5 试验用到的数据集中的部分图片

### 3.2 对比试验结果

本次试验主要是基于VGG16网络模型,采用迁移学习的方法,利用在ImageNet上训练完成的VGG16模型进行训练。为了验证提出的模型的有效性,对试验中选取的VGG16网络模型以及AlexNet模型进行对比,3个模型在文中的服装图像数据集中进行训练,将训练数据保存并绘制成如图6所示的曲线。3个模型的平均精确度见表1。



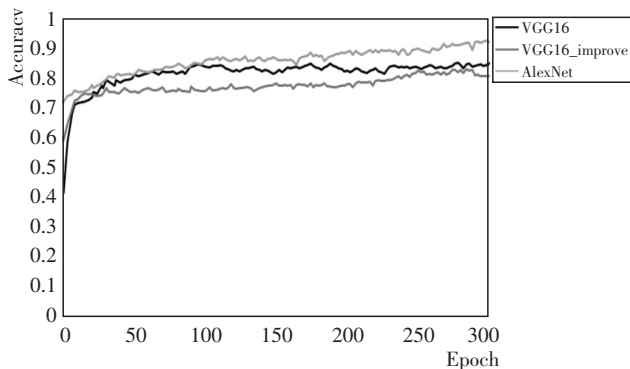


图6 VGG16、VGG16\_improve、AlexNet 模型精度比较

从图6中可以看出, AlexNet 模型与改进的 VGG16 模型初始精度较高, 随着训练的开始, 准确度开始呈现上升趋势。进行迁移学习的 VGG16\_improve 模型的准确率最高, 其次是 VGG16 模型, AlexNet 模型的准确度最低。从表1中可以看出, 未进行迁移学习的模型在数据集上的准确率低于 90%, 而进行迁移学习的 VGG16\_improve 模型的准确率达到了 93.02%。

表1 VGG16、VGG16\_improve、AlexNet 模型的平均精确度

模型	准确率/%
VGG16_improve	93.02
VGG16	84.43
AlexNet	81.85

### 3.3 测试结果

为了进一步测试该模型的准确率, 从 DeepFashion 数据集中另外选取 400 张图片组成测试集, 每类图片 50 张, 将测试集分别用 VGG16\_improve、VGG16、AlexNet 进行测试, 测试结果见表2。从表中可以看出, VGG16\_improve 模型的泛化能力在 3 个模型中是最好的, 平均准确率达到 92.25%。

表2 不同网络模型对不同类别的精确度

单位: %

	VGG16_improve	VGG16	AlexNet
T恤	96	84	80
连衣裙	90	82	76
毛衣	92	82	78
短裙	94	84	86
牛仔裤	94	86	84
西服外套	90	80	80
无袖连衣裙	88	78	78
无袖上衣	94	82	82
平均准确率	92.25	82.25	80.5

从表2中我们可以看出, 在 8 类服装图像中, VGG16\_improve 均取得了最好的分类准确率, 其中有

7 类服装图像的分类准确度都达到了 90% 以上。VGG16 与 AlexNet 两个模型中, 分类效果较好的是 VGG16 模型, 有四个类别的分类准确度都比 AlexNet 模型高, 另外有三类服装图像的分类准确率表现相当。

## 4 结语

针对传统服装图像分类方法效率低、准确度不高的问题, 提出了一种基于 VGG16 和迁移学习的服装图像分类方法。用迁移学习的方法把在 ImageNet 数据集上训练的 VGG16 网络迁移到自己的数据集上进行微调, 并通过数据增强、在网络中添加 BN 层来防止模型过拟合, 提高模型的泛化能力。通过对比试验表明, 文中提出的模型的平均准确率达到 93.02%, 能够对服装图像进行有效分类。

### 参考文献:

- [1] 白琮, 黄玲, 陈佳楠, 等. 面向大规模图像分类的深度卷积神经网络优化[J]. 软件学报, 2018, 29(4): 1029-1038.
- [2] 高妍, 王宝珠, 郭志涛, 等. 改进 HSR-FCN 的服装图像识别分类算法研究[J]. 计算机工程与应用, 2019, 55(16): 144-149.
- [3] LIU Z, LUO P, QIU S, *et al.* DeepFashion: Powering robust clothes recognition and retrieval with rich annotations [C]// Computer Vision and Pattern Recognition. IEEE, 2016.
- [4] 谢小红, 陆建波, 李文韬, 等. 基于迁移学习的服装图像分类模型研究[J]. 计算机应用与软件, 2020, 37(9): 88-93.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv Preprint ArXiv: 1409.1556, 2014.
- [6] GU S, PEDNEKAR M, SLATER R. Improve image classification using data augmentation and neural networks [J]. SMU Data Science Review, 2019, 2(2): 1.
- [7] DUA D, KARRA T E. UCI machine learning repository. Irvine: University of California, 2017. <http://archive.ics.uci.edu/ml>.
- [8] QUATTONI A, COLLINS M, DARRELL T. Transfer learning for image classification with sparse prototype representations [C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.
- [9] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]// International conference on machine learning. PM-

- LR, 2015: 448—456.
- [10] 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(6): 1 229—1 251.
- [11] WANG H. Garbage recognition and classification system based on convolutional neural network vgg16[C]//2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AE-MCSE). IEEE, 2020: 252—255.
- [12] CHEN X, YANG H, YU J, *et al.* A feature extraction strategy of fire hole video based on VGG16 and migration learning[C]//2018 Chinese Automation Congress (CAC). IEEE, 2018: 2 715—2 720.
- [13] 厉 智, 孙玉宝, 王 枫, 等. 基于深度卷积神经网络的服装图像分类检索算法[J]. 计算机工程, 2016, 42(11): 309—315.
- [14] TAMMINA S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images[J]. International Journal of Scientific and Research Publications, 2019, 9(10): 143—150.
- [15] LIU J, WANG M, BAO L, *et al.* Classification and recognition of turtle images based on convolutional neural network[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020, 782(5): 052044.

## Clothing Image Classification Based on VGG16 and Transfer Learning

GAO Ying-ping, SONG Dan, CHEN Yu-ting

(Hunan Institute of Engineering, Xiangtan 411100, China)

**Abstract:** Aiming at the shortcomings of traditional image classification methods such as low efficiency and low accuracy, a clothing image classification method based on VGG16 and migration learning was proposed. The pre-trained VGG16 network model was mainly adopted, which provided better model initialization. Migration learning on image data sets was performed to reduce the dependence on data and the training time of the network. Data enhancement methods were used to increase the amount of training data and improve the generalization ability of the model. The BN Layer was added to the network layer to prevent over fitting and gradient disappearance of the network. Experimental results showed that compared with other traditional convolutional neural network models, the model had a higher accuracy rate and could better classify clothing images.

**Key words:** VGG16; deep learning; data enhancement; image classification

(上接第 44 页)

## Factors Influencing Apparel Recycling Intention in Omni-channel Environment

SHAO Peng, LIU Xiao-pei

(School of Management, Xi'an Polytechnic University, Xi'an 710048, China)

**Abstract:** Sustainable and omni-channel development were two important trends in apparel industry. Based on the data collected from the questionnaire survey, the influence factors of impulsive consumption and apparel disposal recycling intention in omni-channel retailing environment were studied. The main findings were as follows: channel integration quality had a significant negative impact on impulsive consumption, and channel integration exerted significant positive effects on used apparel recycling intention. Materialism had a significant positive impact on impulsive consumption, but there was no significant relationship between materialism and the used apparel recycling intention. Environmental attitude exerted significant positive effects on used apparel recycling intention, but environmental attitude had no significant impact on impulsive consumption.

**Key words:** apparel; environmental attitude; recycling intention; sustainable; impulsive consumption