Computer Science and Application 计算机科学与应用, 2022, 12(5), 1331-1340
Published Online May 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.125132

Hans汉斯

基于卷积块注意力模型的服装图像检索方法

宣益亮[1]，廖小飞[1,2*]，宿轩策[1]

[1]东华大学信息科学与技术学院，上海
[2]东华大学数字化纺织服装技术教育部工程研究中心，上海

收稿日期：2022年4月18日；录用日期：2022年5月17日；发布日期：2022年5月24日

摘　要

针对服装图像检索应用在日常场景下拍摄的服装图像难以避免各种噪声的干扰，如背景或遮挡，严重影响特征提取的准确性，导致检索精度较差等问题，提出一种基于卷积神经网络结合注意力机制的服装图像检索方法，即在ResNet50特征提取网络的基础上加入一种轻量级的通用注意力模块。通过对通道和空间两个独立维度提取特征图，提升在特征提取过程中服装区域的关注程度，压制背景区域，从而提高图像特征的表达能力。通过Triplet Loss损失函数进行网络训练，计算特征向量间的欧氏距离度量图像的相似性。所提方法在DeepFashion数据集上与其他检索方法进行了比较，结果表明该方法能够有效排除图像背景干扰，提高检索精度。

关键词

深度学习，服装检索，注意力机制，残差网络

Clothing Image Retrieval Method Based on Convolutional Block Attention Model

Yiliang Xuan[1], Xiaofei Liao[1,2*], Xuance Su[1]

[1]College of Information Science and Technology, Donghua University, Shanghai
[2]Engineering Research Center of Digitized Textile & Apparel Technology, Ministry of Education, Donghua University, Shanghai

Received: Apr. 18th, 2022; accepted: May 17th, 2022; published: May 24th, 2022

Abstract

For clothing image retrieval applications in daily scenes, it is difficult to avoid the interference of

*通讯作者。

文章引用: 宣益亮, 廖小飞, 宿轩策. 基于卷积块注意力模型的服装图像检索方法[J]. 计算机科学与应用, 2022, 12(5): 1331-1340. DOI: 10.12677/csa.2022.125132

various noises, such as background or occlusion, which seriously affects the accuracy of feature extraction, resulting in poor retrieval accuracy. A clothing image retrieval method based on convolutional neural network combined with attention mechanism is proposed, that is, a lightweight general attention module is added on the basis of ResNet50 feature extraction network. By extracting the feature map from two independent dimensions of channel and space, the attention of the clothing area in the process of feature extraction is increased, and the background area is suppressed, thereby improving the expressive ability of image features. The network is trained through the Triplet Loss function, and the Euclidean distance between the feature vectors is calculated to measure the similarity of the images. The proposed method is compared with other retrieval methods on the DeepFashion dataset, and the results show that the method can effectively eliminate the image background interference and improve the retrieval accuracy.

Keywords

Deep Learning, Clothing Retrieval, Attention Mechanism, Residual Network

1. 引言

如今，服装图像数据量呈突飞猛进的增长趋势，如何在线上对服装图像进行快速又准确地检索受到广大关注。为了满足人们在网上购物时可以准确地搜索到想要的服装，各种服装图像检索方法层出不穷。

服装图像检索方法可以划分成两大类：基于文本的服装图像检索方法和基于内容的服装图像检索方法。如今大多数电商平台使用的主要是基于文本的图像检索，如淘宝、京东等，它的优点是检索速度快且准确度高，但同时存在很多缺点。第一，在进行检索前先需要对数据库中的图像进行人工文本标注，但是随着服装图像数据量与日俱增，通过人工对其进行文本标注会耗费巨大的精力[1]。第二，人工标注还存在一定的主观性，即不同人对同一服装图像可能会有不一样的描述，这种主观因素会影响检索准确度。传统基于内容的服装图像检索一般提取图像的颜色和纹理等特征进行检索，如 ORB、HOG、SIFT [2]。传统方法只能提取服装图像的浅层特征，检索精度仍然不高，同时其缺少一定的学习能力，无法应用于如今大量的服装图像数据集[3]。

2012 年，Krizhevsky 等人[4]提出 AlexNet 网络并应用在图像分类中，在 ILSRCV2012 图像分类竞赛中获得冠军，深度学习方法进入了学者们的视野。深度学习方法开始引起图像检索领域相关学者的关注，从此基于深度学习的图像检索方法成为了新的研究方向。随着如今深度学习的快速发展，研究者们发现利用卷积神经网络技术，在处理图像的过程中可以提取到丰富的图像特征[5]，有利于提高检索精度，因此利用卷积神经网络对图像提取特征的方法成为当前的主流[6]。由于背景噪音干扰、服装变形和遮挡等原因，导致提取的图像特征不准确，排除这些干扰因素是提升图像特征提取准确度的关键。大多数现有服装图像检索方法都是对整个图像进行特征提取，而不是对图像中服装部分区域进行特征提取，这样背景噪声会产生一定的干扰影响[7]。为了解决这个问题，研究人员在神经网络中添加了注意力机制，在识别出图像中需要关注的区域后对该区域重点进行特征提取，以此提高特征提取的准确率[8]。

本文提出一种新的服装图像检索方法，即以 ResNet50 为基础网络，将卷积块注意力模型(Convolutional Block Attention Module, CBAM) [9]引入到深度度量学习网络中，CBAM 是一个轻量级的通

用注意力模块，它可以无缝地添加到任何卷积神经网络中，与网络一起进行端到端的训练，并且可以不考虑开销。本文在 ResNet50 网络的第一层卷积和最后一层卷积加入卷积块注意力模型，使网络在提取图像特征过程中关注重要的特性并抑制不重要的特性，并通过 Triplet Loss 损失函数[10]进行端到端的网络训练，最后得到需要检索的图像和数据库中的图像特征向量间的欧氏距离，以此度量图像的相似度。通过对比实验表明，加入注意力机制来提取图像特征可以有效减少图像背景等因素的干扰，增强对服装区域特征的提取，检索精度与其他模型相比有较大的提升。

2. 相关工作

2.1. 服装检索

在如今的深度学习领域中，结合深度学习技术在图像分类，图像检索等方向有了广泛的应用，并诞生了许多经典网络模型，如 AlexNet [11]、VGG-16 [12]、GoogleNet [13]等。He [14]等在 2018 年提出一个适用于服装图像检索任务的基准模型 FashionNet，采用经典的分类网络 VGG-16 作为骨干结构，并对它作了修改，将其最后一个卷积层替换成三个采集不同信息的分支，在第 5 个卷积层后连接了 3 个分支网络结构，并同时用三个不同的损失函数训练这三个分支网络。将图像全局特征和关键点局部特征进行融合作为特征描述子，避免局部区域的干扰，提高了服装检索的精度。Lang 等[15]提出服装关键点引导的区域注意力机制，可以利用分布在服装图像各个关键位置的点位估计分支来预测服装的关键点，对服装图片进行多个区域的划分，服装图像检索分支和服装关键点估计分支都是用 HR-Net 主干网络，通过多级并联结构保证了高分辨率和多尺度特征。

2.2. 深度残差网络

在通常情况下，网络的深度对模型的性能至关重要，随着网络宽度和深度的加深，网络可以进行更加复杂的特征提取，性能也会随着得到提升，但是，无法避免的问题是在深度学习中，网络层数的增加导致网络准确度出现饱和，甚至出现下降，带来模型梯度爆炸、过拟合和梯度消失等问题，深度网络出现了退化现象[16]。针对这个问题，何凯明提出了残差网络，在 ILSVRC 和 COCO 2015 上的战绩非凡[17]。ResNet 网络参考 VGG19 经典神经网络，在它的基础上进行修改，并通过在网络的每两层间增加短路机制，添加了残差单元，这就形成了残差学习。残差网络根据不同的卷积层数，可分为 Resnet18、Resnet34、Resnet50、Resnet101、Resnet152 五种不同的类型[18]。

3. 基于卷积块注意力模型的服装图像检索方法

本节将详细介绍基于残差网络(ResNet50)和卷积块注意力模型(CBAM)的服装图像检索方法。注意力机制是模仿人类视觉的一种方法，主要分为三种模型：通道域注意力、空间域注意力、空间和通道混合的注意力。

本文提出的服装图像检索方法以 ResNet50 为主干网络，将卷积块注意力模型添加到主干网络的各卷积层之间，整体网络结构如图 1 所示。CBAM 是一个轻量级的通用注意力模块，它可以无缝地添加到任何卷积神经网络中，因此在 ResNet50 主干网络的第一层卷积层和最后一层卷积层后加入卷积块注意力模型，为了使用预训练参数，不能改变 ResNet 的网络结构，所以 CBAM 不能加在卷积层里面，因为加进去网络结构发生了变化不能用预训练参数，加在最后一层卷积和第一层卷积不改变网络，可以用预训练参数。通过加入注意力机制使网络在提取特征时关注重要特征，并抑制不必要特征，通过 Triplet Loss 损失函数进行端到端的网络训练，最后得到被检索图像与数据库图像特征向量之间的欧氏距离，以此来度量图像之间的相似度。

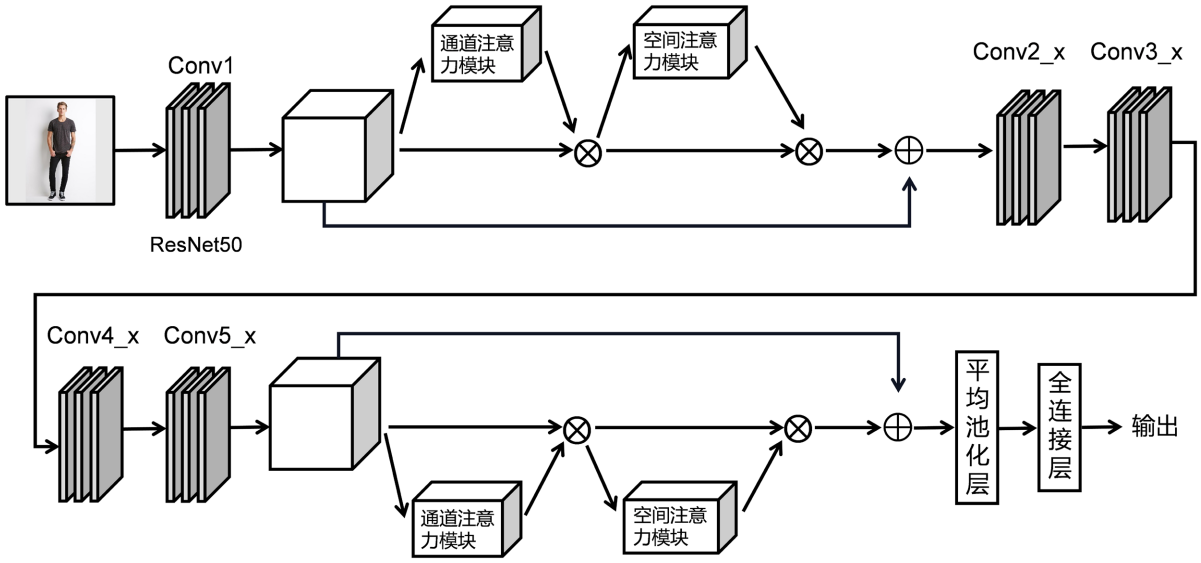

Figure 1. Overall network architecture
图 1. 整体网络架构

3.1. 卷积块注意力模型

卷积块注意力模型(Convolutional Block Attention Module, CBAM)，包括通道注意力模块和空间注意力模块。它可以添加到前馈卷积神经网络中，使网络关注图像中需要关注部分所在区域的特征，抑制背景和不需要注意区域的特征。它的原理可以简单理解为输入一个图像特征图，CBAM 模型会沿着两个独立的维度，即通道维度和空间维度依次进行注意力图的推断学习，用输入的特征图与输出的注意力特征图相乘，并进行自适应的特征向量融合。由于 CBAM 是轻量级的通用注意力模块，因此可以忽略该模块的开销从而将其无缝添加到经典卷积网络中，并且与网络同步进行端到端训练。从文献[9]中的实验中得到的结果证明了神经网络中加入该卷积块注意力模型应用在分类和目标检测等任务中都能取得不错的效果，因此本文将其应用在服装图像检索任务中。

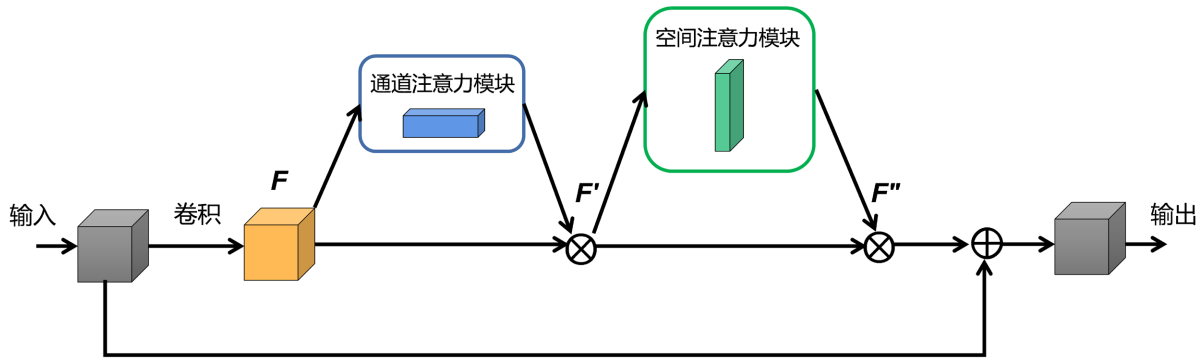

Figure 2. Convolutional block attention model
图 2. 卷积块注意力模型

如图 2 所示，利用深度卷积神经网络处理输入的服装图像后，将得到的图像特征 $F \in \mathbb{R}^{C\times H\times W}$ 作为卷积块注意力模型的输入，卷积块注意力模型将会依次通过通道和空间注意力模块生成一维的通道注意力图 $M_c \in \mathbb{R}^{C\times 1\times 1}$ 和二维的空间注意力图 $M_s \in \mathbb{R}^{1\times H\times W}$ 。整个模型对输入的图像特征的处理过程可以表示为：

$$\begin{aligned} \boldsymbol{F}' &= \boldsymbol{M}_c(\boldsymbol{F}) \otimes \boldsymbol{F} \\ \boldsymbol{F}'' &= \boldsymbol{M}_s(\boldsymbol{F}') \otimes \boldsymbol{F}' \end{aligned} \tag{1}$$

表达式中的$\otimes$表示对应元素相乘。在进行元素相乘时，注意力图的扩展方式为：通道注意力图的值沿着空间维度扩展，空间注意力图则沿着通道维度扩展。图 3 是卷积块注意力模块中通道和空间注意力模块具体处理过程。

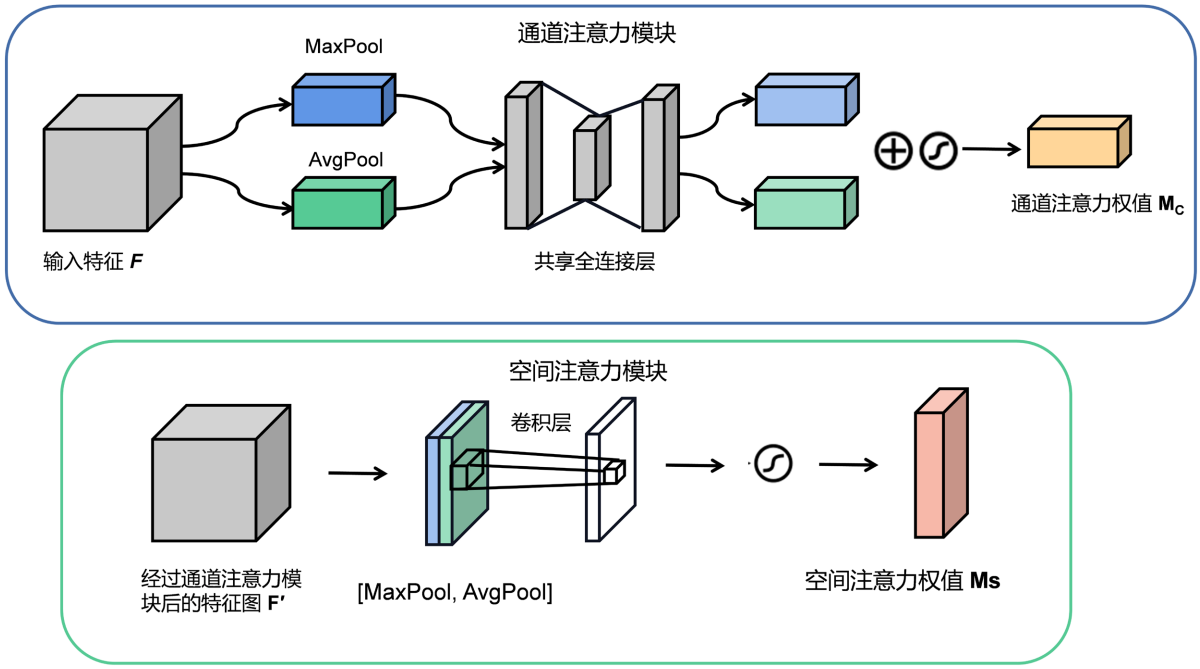

Figure 3. Flowchart of two attention modules
图 3. 两个注意力模块流程图

在通道注意力模块中，将输入的图像特征在空间维度上进行压缩，采取的压缩方式包括平均值池化(Average Pooling)和最大值池化(Max Pooling)。图像特征在空间维度上分别经过平均值池化和最大值池化聚合得到两个空间特征描述子：F_{avg}^{c} 和 $F_{\max}^{c}$。然后将这两个空间特征描述子送到一个共享网络，压缩输入特征图的空间维数，逐元素求和合并，以产生通道注意力图。通道注意力过程可以表达为：

$$\begin{aligned} \boldsymbol{M}_c(\boldsymbol{F}) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}(\boldsymbol{F})\right) + \mathrm{MLP}\left(\mathrm{MaxPool}(\boldsymbol{F})\right)\right) \\ &= \sigma\left(\boldsymbol{W}_1\left(\boldsymbol{W}_0\left(\boldsymbol{F}_{avg}^{c}\right)\right) + \boldsymbol{W}_1\left(\boldsymbol{W}_0\left(\boldsymbol{F}_{\max}^{c}\right)\right)\right), \end{aligned} \tag{2}$$

其中σ表示 sigmoid 函数，$\boldsymbol{W}_0 \in \mathbb{R}^{C/r \times C}$，$\boldsymbol{W}_1 \in \mathbb{R}^{C \times C/r}$。

同样，在空间注意力模块中是对图像特征在通道维度上进行压缩，在通道维度分别进行了平均值池化和最大值池化。聚合得到两个二维的通道特征描述子：$F_{avg}^{s} \in \mathbb{R}^{1 \times H \times W}$ 和 $F_{\max}^{s} \in \mathbb{R}^{1 \times H \times W}$，然后通过一个标准卷积层将这些特征连接并卷积起来，生成二维空间注意图。空间注意力过程可以表达为：

$$\begin{aligned} \boldsymbol{M}_s(\boldsymbol{F}) &= \sigma\left(f^{7\times7}\left(\left[\mathrm{AvgPool}(\boldsymbol{F}); \mathrm{MaxPool}(\boldsymbol{F})\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[\boldsymbol{F}_{avg}^{s}; \boldsymbol{F}_{\max}^{s}\right]\right)\right) \end{aligned} \tag{3}$$

其中σ表示 sigmoid 函数，$f^{7\times7}$ 表示卷积层的卷积核大小为 7 × 7。

3.2. 相似性度量

服装图像通过神经网络提取图像特征后，需要将输出得到的特征向量和数据库中图像的特征向量进行计算比较。相似性度量就是综合评价两个事物之间相似程度，越相似的两个图像，它们的相似性度量越大；越不相似的两个事物，它们的相似性度量越小，通常一个好的距离度量方法往往就能决定算法最后结果的好坏。本文采用的度量公式为欧氏距离，即直接计算两个向量的直线距离。n 维特征向量 $a(x_1,x_2,\cdots,x_n)$ 和 $b(y_i,y_2,\cdots,y_n)$ 的欧式距离公式为：

$$dist(a,b)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2} \tag{4}$$

4. 实验与讨论

4.1. 数据集

为了验证本文提出的网络模型的性能，选择了在服装图像检索领域较为常用的 DeepFashion 数据集[18]进行实验。DeepFashion 数据集是香港中文大学提出的一个大规模服装图像数据集，其中包含 80 多万张不同场景、不同角度的服装图像还有买家秀的照片，图像包含有类别、属性、关键点和特征点等信息，一共分为四个大类。面向四种不同预测任务子数据集的服装数据集，分别是服装图像分类、同域服装图像检索、跨域服装图像检索、服装图像关键点检测。从其子集中抽取 4 万训练集、1 万测试集和 1 万验证集进行试验，其中包含 30 种类别的图片。

4.2. 评价指标

在图像检索中评价检索精度比较常用的是 top-k 和平均查准率均值(mAP)，以此作为实验中检索性能评价指标。

1) Top-K 精度：对前 k 张检索的图像中，于查询图像相似的样本占的百分比。该衡量指标可用 P_k 表示，公式中 N_r 代表检索到相似图像样本数，k 表示检索样本总数。

$$P_k = N_r/k \tag{5}$$

2) 平均查准率均值：在不同的 top-k 精度下，为了综合比较算法模型的性能，常使用平均查准率均值作为最终性能评价指标。该指标计算如式下所示，其中，X_q 表示第 q 个被查询图像，N_r 表示检索到相似图像样本个数，M 表示待检索的图像样本个数。

$$AP(X_q)=\frac{1}{N_r}\sum_{k=1}^{N_r}P_k \tag{6}$$

$$mAP=\frac{1}{M}\sum_{q=1}^{M}AP(X_q) \tag{7}$$

4.3. 实验结果与分析

为了判断结合卷积块注意力模型的服装图像检索方法的性能优越性，本文展开了一系列对比实验。文献[10]中提出的注意力模型属于通道注意力模型，为了验证本文提出网络模型的优势，将只包含通道注意力模块的检索方法和基于卷积块注意力模型的检索方法进行对比实验，将卷积块注意力模型表示为 CBAM，通道注意力模型表示为 SE。

对比实验中所有的网络结构所使用的损失函数都是 Triplet loss，使用的主干网络 ResNet50 是预先在 ImageNet 数据集中训练过的，将卷积块注意力模型加入到 ResNet50 上时，使用的也是预训练过的

ResNet50 + CBAM 网络。以 ResNet50 为主要网络加入不同的注意力模块进行检索性能的对比，根据检索出的前 K (= 5, 10, 20)张图像来计算准确率，实验对比结果如表 1 所示。

Table 1. Retrieval mAP comparison of ResNet50 combined with different attention modules
表 1. ResNet50 结合不同注意力模块的检索 mAP 比较

方法	Top5 (%)	Top10 (%)	Top20 (%)
ResNet50	88.7	82.4	68.2
ResNet50 + SE	90.1	85.1	69.4
ResNet50 + CBAM	92.5	89.8	71.3

从表中可以发现，在相同的 Top-K 下，不同的算法模型进行检索得出的 mAP 由低到高依次为：Resnet50、Resnet50 + SE、ResNet50 + CBAM。从中可以看出，在 ResNet50 网络上添加通道注意力模块和添加卷积注意力模块都能对检索性能带来提升效果。不添加注意力模块的模型检索 mAP 最低，因为这种方法只关注了服装图像的全局特征，没有对需要特别关注的服装区域进行重点特征提取，提取特征的不精准，导致检索 mAP 过低；ResNet50 + SE 加入了通道注意力模块，对准确度确实有一定的提升，但是单单加入通道注意力模块对性能提升并不是很明显；本文提出的 ResNet50 + CBAM 模型对检索性能的提升最大，同时加入通道注意力和空间注意力，在提取图像特征过程中关注重要的特性并抑制不必要特性，提高图像特征的表达能力。

为了比较在不同的网络中加入卷积块注意力模型对检索性能的影响，进行了以下实验，分别在经典的 VGG16 和 GoogleNet 深度神经网络中加入 CBAM 注意力机制，并与本文的 ResNet50 + CBAM 模型进行比较，实验结果如表 2 所示，结果表明在 ResNet50 网络中加入注意力模块的检索性能比在 VGG16 和 GoogleNet 中加入注意力模块更优越。

Table 2. Comparison of Retrieval mAPs in different backbone networks combined with CBAM
表 2. 不同主干网络结合 CBAM 的检索 mAP 比较

方法	Top5 (%)	Top10 (%)	Top20 (%)
VGG16 + CBAM	81.5	79.1	59.2
GoogleNet + CBAM	83.6	81.2	62.4
ResNet50 + CBAM	92.5	89.8	71.3

图 4 展示了部分 top-5 查询示例，第 1 列表示待检索的服装图像，其他列表示 top-5 的查询结果，其中检索结果中和待检索图像相似的由红色框标出。

文献[19]中提出了类激活热力图(Grad-CAM)，可以可视化显示网络中卷积层的激活情况，不需要修改网络架构。对图像的哪个区域越关注，在可视化结果中这个区域的热力就越高。将服装图像输入到 ResNet50、ResNet50 + SE 和 ResNet50 + CBAM 模型中，用最后一个卷积层的输出来计算类激活热力图。图 5 给出了三个模型的热力图可视化结果。从图中可以明显看出，加入注意力机制可以增强网络提取特征能力，可以使神经网络提取的图像特征与输入图像中需要重点关注区域相关性更高，同时抑制背景区域的干扰噪音，从图中可以看出 ResNet50 + CBAM 模型对需要关注的目标区域的关注程度比 ResNet50 + SE 模型更高一些。

Figure 4. Search results
图 4. 检索结果

Figure 5. Class activation heatmaps for the three models
图 5. 三种模型的类激活热力图

5. 结语

本文提出的一种卷积神经网络结合卷积块注意力模型的服装图像检索方法，主要思想是以 ResNet50 网络为主干网络，在其第一层和最后一层卷积层中嵌入注意力模型，结合注意力机制提取图像需要关注区域的重要特征，提高服装检索的准确度。在 DeepFashion 数据集上，与其他方法进行对比，通过实验证明本方法可以有效减少图像背景等因素的干扰，提升服装图像检索精度。下一步可以增加服装图像数据量或者多种类的服装图像数据集进行实验验证，并优化网络模型，从而进一步提高检索精度。

基金项目

中央高校基本科研业务费专项基金(15D110422)资助。

参考文献

[1] Goei, K., Hendriksen, M., de Rijke, M., *et al*. (2021) Tackling Attribute Fine-Grainedness in Cross-Modal Fashion Search with Multi-Level Features. *SIGIR* 2021 *Workshop on eCommerce*, Montreal, 15 July 2021.

[2] Li, S., Wang, Z. and Zhu, Q. (2020) A Research of ORB Feature Matching Algorithm Based on Fusion Descriptor. 2020 *IEEE* 5*th Information Technology and Mechatronics Engineering Conference* (*ITOEC*), Chongqing, 12-14 June 2020, 417-420. https://doi.org/10.1109/ITOEC49072.2020.9141770

[3] Yang, M., He, D., Fan, M., *et al*. (2021) DOLG: Single-Stage Image Retrieval with Deep Orthogonal Fusion of Local and Global Features. 2021 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 11752-11761. https://doi.org/10.1109/ICCV48922.2021.01156

[4] Li, S., Wang, L., Li, J. and Yao, Y. (2021) Image Classification Algorithm Based on Improved AlexNet. *Journal of Physics*: *Conference Series*, **1813**, Article ID: 012051. https://doi.org/10.1088/1742-6596/1813/1/012051

[5] D'Innocente, A., Garg, N., Zhang, Y., *et al*. (2021) Localized Triplet Loss for Fine-Grained Fashion Image Retrieval. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Nashville, 19-25 June 2021, 3905-3910. https://doi.org/10.1109/CVPRW53098.2021.00435

[6] Sharma, V., Murray, N., Larlus, D., *et al*. (2021) Unsupervised Meta-Domain Adaptation for Fashion Retrieval. 2021 *IEEE Winter Conference on Applications of Computer Vision* (*WACV*), Waikoloa, 3-8 January 2021, 1347-1356. https://doi.org/10.1109/WACV48630.2021.00139

[7] Su, H., Wang, P., Liu, L., *et al*. (2020) Where to Look and How to Describe: Fashion Image Retrieval with an Attentional Heterogeneous Bilinear Network. *IEEE Transactions on Circuits and Systems for Video Technology*, **31**, 3254-3265.

[8] Tesfaye, A.L. and Pelillo, M. (2018) Multi-Feature Fusion for Image Retrieval Using Constrained Dominant Sets. *Image and Vision Computing*, **94**, Article ID: 103862.

[9] Woo, S., Park, J., Lee, J.Y., *et al*. (2018) CBAM: Convolutional Block Attention Module. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[10] Kuang, Z., Gao, Y., Li, G., *et al*. (2019) Fashion Retrieval via Graph Reasoning Networks on a Similarity Pyramid. 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 3066-3075. https://doi.org/10.1109/ICCV.2019.00316

[11] Barz, B. and Denzler, J. (2021) Content-Based Image Retrieval and the Semantic Gap in the Deep Learning Era. *International Conference on Pattern Recognition*, Springer, Cham, 245-260. https://doi.org/10.1007/978-3-030-68790-8_20

[12] Sun, Y., Wong, W.K. and Zou, X. (2021) A Multi-Task Model for Multi-Attribute Fashion Recognition and Retrieval. *AATCC Journal of Research*, **8**, 105-116.

[13] Hu, J., Shen, L., Albanie, S., *et al*. (2019) Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7132-7141. .

[14] He, T. and Hu, Y. (2018) FashionNet: Personalized Outfit Recommendation with Deep Neural Network. arXiv:1810.02443.

[15] Lang, Y., He, Y., Yang, F., *et al*. (2020) Which Is Plagiarism: Fashion Image Retrieval Based on Regional Representation for Design Protection. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 14-19 June 2020, 2595-2604.

[16] Li, Z., Liu, F., Yang, W., *et al*. (2021) A Survey of Convolutional Neural Networks: Analysis, Applications, and Pros-

pects. *IEEE Transactions on Neural Networks and Learning Systems*, 1-21.

[17] Morelli, D., Cornia, M. and Cucchiara, R. (2021) FashionSearch++: Improving Consumer-to-Shop Clothes Retrieval with Hard Negatives. *Italian Information Retrieval Workshop*, Bari, 13-15 September 2021.

[18] Ge, Y., Zhang, R., Wu, L., *et al*. (2019) DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 5332-5340. https://doi.org/10.1109/CVPR.2019.00548

[19] Selvaraju, R.R., Cogswell, M., Das, A., *et al*. (2020) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, **128**, 336-359. https://doi.org/10.1007/s11263-019-01228-7