

姚砺, 朱子卿, 万燕. 基于注意力机制与属性解纠缠的服装图像检索研究[J]. 智能计算机与应用, 2026, 16(3): 187-193.
DOI:10.20169/j.issn.2095-2163.24042502

基于注意力机制与属性解纠缠的服装图像检索研究

姚 砺, 朱子卿, 万 燕

(东华大学 计算机科学与技术学院, 上海 201620)

摘要: 电子商务的发展催生了服装时尚购物领域个性化搜索需求的增加, 多模态服装图像检索是解决此需求的关键任务。本文关注多模态检索中带有属性操作的服装图像检索任务, 并提出了一个服装属性操作检索模型。模型在残差网络(ResNet)基础上添加特征解纠缠模块提取服装的属性解纠缠特征, 以使特征便于属性操作; 引入属性引导的卷积注意力模块(CBAM), 以提高特征提取能力; 同时采用改进的交叉熵与三元损失函数提升模型在多标签数据集上的表现。在此基础上引入特征记忆模块存储属性解纠缠特征, 将属性操作利用特征矩阵变换为残差特征并与服装特征融合进行相似性检索。最后在公开数据集上的实验验证了论文相关改进在检索准确度上的提升与相对于其他方法的优越性。

关键词: 深度学习; 图像特征提取; 服装图像检索; 注意力机制; 特征解纠缠

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2026)03-0187-07

Research on clothing image retrieval based on attention mechanism and attribute disentanglement

YAO Li, ZHU Ziqing, WAN Yan

(College of Computer Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: The development of e-commerce has led to an increased demand for personalized search in the fashion retail sector, making multimodal clothing image retrieval a pivotal task. This paper addresses the task of clothing image retrieval with attribute manipulations in multimodal retrieval scenarios and proposes a clothing attribute manipulation retrieval model. Building upon the Residual Network (ResNet), the model incorporates a feature disentanglement module to extract disentangled attribute features, facilitating attribute manipulations. An attribute-guided Convolutional Block Attention Module (CBAM) is introduced to enhance feature extraction capabilities. Furthermore, improvements in cross-entropy and triplet loss functions are employed to enhance model performance on multi-label datasets. Building on this, a feature memory module is introduced to store disentangled attribute features, transforming attribute manipulations into residual features and fusing them with clothing features for similarity retrieval. Finally, experiments on public datasets validate the proposed enhancements in retrieval accuracy and demonstrate their superiority over alternative methodologies.

Key words: deep learning; image feature extraction; clothing image retrieval; attention mechanism; feature disentanglement

0 引言

服装图像检索是服装搜索的主要手段之一, 优化服装图像的检索, 对提高消费者消费体验和推动服装产业设计与销售的智能化、自动化都有一定的应用层面价值^[1]。相比于基于文本的图像检索与基于内容的图像检索, 将服装图像与其他模态信息(如特定服装属性、相关文本、用户偏好信息等)相结合的多模态图像检索能更好满足用户灵活多变与

个性化的检索需求^[2]。

本文研究的服装属性操作检索即是多模态检索任务之一, 可让用户在输入查询图像的基础上提供额外的属性修改, 从而将查询图像的某个属性变为自己更符合需求的属性^[3]。图1示例中输入查询图像“蓝色T恤连衣裙”, 并添加“颜色改变为灰色”的属性操作, 通过检索模型与服装图片库中服装比对, 得到第一个检索结果图像“灰色T恤连衣裙”, 即为所预期的目标图像。

作者简介: 朱子卿(1999—), 男, 硕士研究生, 主要研究方向: 图像处理; 万燕(1970—), 女, 博士, 教授, 硕士生导师, 主要研究方向: 图像处理。

通信作者: 姚 砺(1967—), 男, 博士, 副教授, 硕士生导师, 主要研究方向: 软件测试技术, 图像处理等。Email: yaoli@dhu.edu.cn。

收稿日期: 2024-04-25

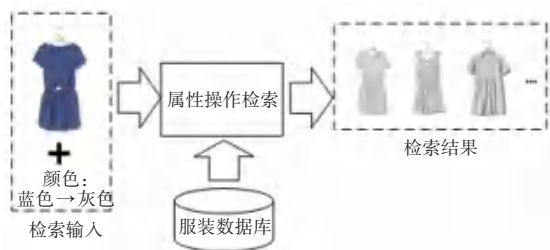


图1 服装属性操作检索示例

Fig. 1 Example of clothing attribute manipulation retrieval

在这一任务的早期研究中, Hardoon 等学者^[4]利用传统的多视图嵌入方法之一的典型相关分析(Canonical Correlation Analysis, CCA)来将图像与操作文本映射到同一子空间中。Zhao 等学者^[5]提出了一个记忆增强属性操作网络(AMNet),利用双路径网络提取图像特征,并采用记忆模块对应的属性原型特征修改对应特征来检索图片库中的目标服装图像。此后,Kenan 等学者^[6]构建了时尚搜索网络FashionSearchNet,通过弱监督定位方法来提取图像中的关键属性区域以更好定位修改的属性位置,并提升图像搜索的准确率。

服装特征的解纠缠方法、即将服装特征按照类别、颜色、纹理等属性解纠缠成对应特征以表示不同维度的语义信息能够提高特征的可解释性、可操作性,契合服装属性操作检索的需求^[7]。服装领域的相关研究还在起步阶段:如 Saha 等学者^[8]通过将投影层的权重矩阵初始化为各属性类型的对角矩阵来扩展相关网络以提高特征分类性能。Yildirim 等学者^[9]则提出了一种基于条件选择的生成对抗网络

来生成具有颜色、纹理和形状特征控制的服装图像以提高其在图像生成任务上的表现。Lin 等学者^[10]提出了一种基于类别的子空间处理网络,利用子空间权重与属性向量结合来选择适当的子空间嵌入以便进行属性操作。

上述研究较好解决了属性操作与图像特征结合、服装特征解纠缠与属性特征操作的问题,但没有针对性解决服装图像“类间方差小,类内方差大”与多标签服装属性分布和关联信息的提取等问题。为此,本文提出了基于属性引导的注意力机制与特征解纠缠网络的服装属性操作检索模型以解决上述问题。

1 总体方法设计

为了实现服装图像属性操作检索,模型分别需要实现以下子目标:

(1) 读取输入的服装图像,提取包含服装属性语义信息的特征用于后续属性操作与检索。

(2) 将对属性的操作与服装特征相结合,以修改对应属性信息并将新属性的语义信息更新至服装检索特征中。

(3) 利用包含属性操作的语义特征检索目标图像,需要构造服装图片库对应的服装特征库以便检索特征与其比对得到相似图像作为检索结果。

为分别实现以上子目标,设计的总体架构如图2所示,主要由属性解纠缠特征提取模块、属性特征记忆模块与特征检索部分组成。

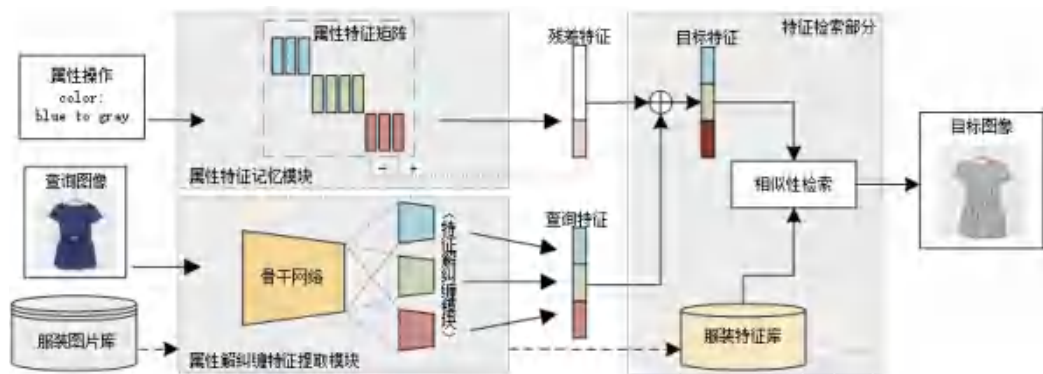


图2 服装图像检索模型架构

Fig. 2 Architecture of clothing image retrieval model

属性解纠缠特征提取模块负责提取服装图像的属性解纠缠特征。模块接收输入的查询图像,先通过结合注意力机制的骨干网络提取图像全局特征,然后经过每个服装属性对应的特征解纠缠模块从全局特征提取出对应的解纠缠属性特征。最后,模块

将属性特征拼接作为图像的最终特征(即查询特征)输出,用于后续的检索任务。

属性特征记忆模块主要负责将输入的属性操作编码为残差特征用于与查询图像的特征结合。模块维护的属性特征矩阵用于记忆每个属性所有属性值

的代表属性特征。当模块以操作向量的形式接收输入的属性操作后, 会相应地将代表属性与目标属性的特征结合计算出与图像特征形状相同的残差特征。

特征检索部分负责结合特征并用其检索出最终图片结果。服装特征库的特征由服装图片库中图片通过特征提取模块生成。通过残差特征与目标特征相加生成目标特征与服装特征库中的特征比较相似度, 通过聚类检索的方式查找到特征最相似的目标图片, 即最终模型输出的检索结果。

2 基于属性引导注意力的特征提取网络

服装特征提取需要排除背景姿势光线等信息噪音以准确识别服装类别纹理等信息, 并将不同属性信息映射于各自特征子空间中。本文提出的基于属性引导注意力的特征提取网络是基于残差网络 ResNet50^[11] 的结构改进的。改进后的特征提取网络如图 3 所示。由图 3 可知, 首先将 ResNet 末尾全连接层替换为属性各自的特征解纠缠模块 (Attribute Feature Disentanglement block, AFD block), 以将不同属性的特征解耦。其次, 在网络底层与每个 AFD 模块前引入通道与空间注意力模块 (C&S Attention block) 以学习不同属性在特征通道与空间上的偏好。最后, 优化了损失函数, 采用加权交叉熵损失函数与兼容多标签形式数据的三元组损失以更好地学习多标签服装图像的属性语义信息。

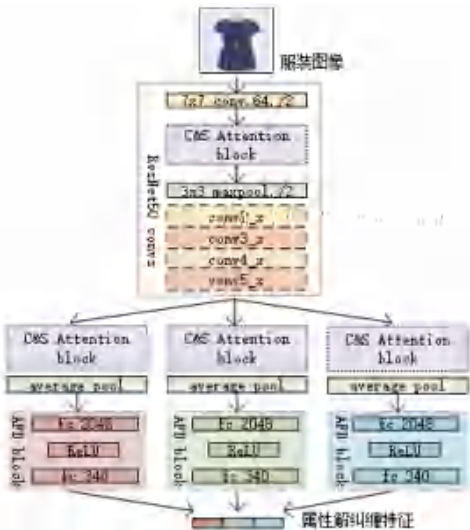


图 3 改进后的特征提取网络

Fig. 3 Improved feature extraction network

2.1 残差网络的属性解纠缠改进

由于服装属性操作检索需要将属性改变的语义信息作用于图像特征同时不影响其他属性, 这需要提取的服装图像特征的属性之间能足够解纠缠, 以

便互不影响语义操作。特征提取骨干网络残差网络 ResNet50 通过残差学习机制能缓解梯度消失问题, 但不具备解纠缠形式的特征提取功能。故本文为每个属性构建了参数独立的 AFD 模块, 属性特征解纠缠的网络得到的特征表示如下。

对于服装图像 I , 其在属性 a 上的解纠缠特征 $f_d(I, a)$ 的数学公式为:

$$f_d(I, a) = AFD_a(f_g(I)) \quad (1)$$

其中, $f_d(I, a)$ 表示骨干网络提取的全局特征, $AFD_a(\cdot)$ 表示属性 a 对应的 AFD 模块。

最终得到的属性解纠缠特征 $f(I)$ 的数学公式为:

$$f(I) = f_d(I, A) = [f_d(I, a_1), f_d(I, a_2), \dots, f_d(I, a_N)] \quad (2)$$

其中, $f(I)$ 、即 $f_d(I, A)$ 表示属性集 A 所有属性解纠缠特征的拼接。通过利用 2 个全连接层将全局特征降维并用激活函数选择性地舍弃全局信息的语义信息, AFD 模块能够从全局特征中提取出对应属性的特征。

2.2 属性引导的注意力模块

针对服装图像属性细节多样、背景姿势等干扰信息大的特点, 需要引入注意力机制来引导网络关注任务相关区域。本文引入结合通道与空间的注意力机制, 具体结构设计采用了卷积块注意力模块 (CBAM)^[12] 的思想, 通道与空间注意力机制模块如图 4 所示, CBAM 将通道与空间注意力依次连接以更好结合不同维度的权重信息。

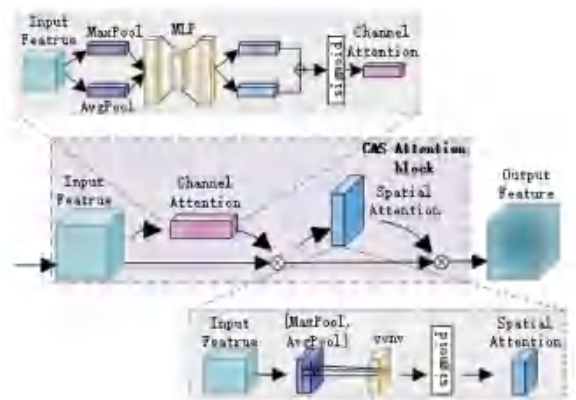


图 4 通道与空间注意力机制模块

Fig. 4 Channels and spatial attention block

本文在网络底层与每个特征解纠缠模块前添加 CBAM 模块以构建属性引导的注意力, 提高了网络对服装图像低层次特征的提取, 同时网络能够更好学习不同属性特征在特征通道与空间上的偏好, 有利于各自属性特征的提取。

2.3 损失函数优化

为了更好地学习多标签服装图像的属性语义信息,并且使不同的属性解纠缠特征能在对应的属性语义空间中合理分布,本文采用加权交叉熵损失函数与兼容多标签形式数据的三元组损失函数作为网络训练的损失函数,损失函数的计算公式为:

$$\text{Loss} = w_1 \times L_{\text{WCE}} + w_2 \times L_T \quad (3)$$

其中, L_{WCE} 表示正负样本加权平衡的交叉熵损失函数; L_T 表示三元组损失函数; w_1 与 w_2 分别表示2个损失函数的系数。

加权交叉熵损失函数的定义公式为:

$$L_{\text{WCE}} = \sum_{i=1}^D (w_p \cdot a_i \log p(a_i | I_i) + w_n \cdot (1 - a_i) \log(1 - p(a_i | I_i))) \quad (4)$$

其中, D 表示训练样本的数量; a_i 表示图像 I_i 的属性标签; w_p 与 w_n 分别表示正样本系数与负样本系数,由属性正负标签数量的比值决定。

通过对正样本和负样本的贡献进行不同的加权,加权交叉熵损失函数防止模型偏向于多数类别,并确保学习到的特征对所有类别都有区分度。

三元组损失函数^[13]能使相同属性的服装在其属性子空间趋向靠近,不同属性的图像特征趋向远离,为了训练不同属性子空间上的特征分布与图像特征整体的特征分布,本文提出了属性引导的三重损失与全局三重损失结合的重重损失,具体计算公式为:

$$L_T = \frac{1}{N_A} \sum_{i=1}^{N_A} L_T(a_i) + L_{\text{GT}} \quad (5)$$

其中, N_A 表示属性数量; $L_T(a_i)$ 表示在属性 a_i 子空间上的三元组损失函数; L_{GT} 表示全局三元组损失函数。

通过对三元组损失函数的学习,网络能在各自属性子空间有效捕捉服装图像之间的相似性和相异性,同时保证服装全局特征的合理分布。

3 基于特征解纠缠的服装属性操作检索方法

3.1 属性特征记忆模块

为使属性操作能根据不同的目标属性查询图像的特征,本文提出属性特征记忆模块来储存属性代表特征并辅助属性操作检索任务。储存属性代表特征的方法在服装图像领域已有一些实践^[8-9,14-15],本文受文献[14]启发,采用伪对角矩阵形式储存属性特征,矩阵的具体表示公式为:

$$\mathbf{M}_F(a_i) = [f_d(a_1)^T, f_d(a_2)^T, \dots, f_d(a_{M_i})^T] \quad (6)$$

$$\mathbf{M}_F = \begin{bmatrix} \mathbf{M}_F(a_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_F(a_2) & & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}_F(a_N) \end{bmatrix} \quad (7)$$

其中,属性 a_i 的属性特征子矩阵 $\mathbf{M}_F(a_i)$ 表示一个 $D_i \times M_i$ 形状的矩阵 (D_i 为解纠缠特征维度, M_i 为属性取值个数),由该属性所有属性值代表的解纠缠特征 $f_d(a_i)$ 拼接而成。而总的属性特征矩阵 \mathbf{M}_F 则为 $\sum D_i \times \sum M_i$ 形状的伪对角矩阵,对角线上依次是每个属性对应的属性特征子矩阵 $\mathbf{M}_F(a_i)$,其他位置则用零矩阵填充。属性代表特征 $f_d(a_i)$ 对矩阵十分关键,受文献[8]启发,本文采用对应属性值图像的解纠缠特征的均值作为矩阵初始化的值。

本文为了属性特征记忆模块的训练引入了特征一致性损失与矩阵对角化损失。属性特征记忆模块机制如图5所示,特征一致性损失考察特征矩阵对应属性拼接而成的特征 f_c 、目标特征 $f'(q)$ 与目标图像经过网络提取到的特征 $f(t)$ 三者的 $L1$ 范式距离,以维持属性特征记忆模块、目标特征与特征网络提取的图像特征三者之间的语义一致性。同时矩阵对角化损失通过计算非对角元素的绝对值使其接近于零,来维持非对角元素的稀疏性,保证属性特征间的解纠缠。

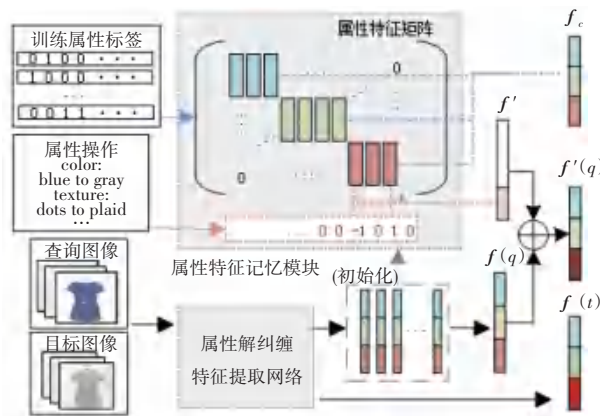


图5 属性特征记忆模块机制

Fig. 5 Mechanism of attribute feature memory module

3.2 属性操作检索实现

属性操作检索实现具体过程如图6所示。

属性操作可以表征为一个属性操作独热向量 \mathbf{v}_m ,其中待修改的属性值位置与修改后的属性值位置分别取值-1与1,其余位置取值皆为0。首先, \mathbf{v}_m 通过与属性特征矩阵 \mathbf{M}_F 相乘计算得到残差特征向量 \mathbf{f}' 。其次,查询图像经过特征提取网络得到的特

征 $f(q)$, 与残差特征 f' 合并得到用于检索的目标特征 $f'(q)$ 。最后, 目标特征 $f'(q)$ 与服装特征库中的特征进行相似性比较, 寻找与目标特征在欧式距离上最近的一组图片特征, 作为属性操作检索的最终检索结果。

力的提升, 进行分类任务上的消融实验, 实验结果见表 2。

表 1 特征解纠缠骨干网络对比实验

Table 1 Comparative experiments on feature disentanglement backbone networks

结合属性解纠缠的骨干网络	准确率/%
AlexNet ^[17] w/FD	51.8
VGG16 ^[18] w/FD	53.8
ResNet34 w/FD	52.5
ResNet50 w/FD	54.7

表 2 特征提取网络消融实验

Table 2 Feature extraction network ablation experiment

模型	准确率/%
ResNet50 (baseline)	54.7
ResNet50 + AD attention	56.3
ResNet50 + W&T loss	59.8
ResNet50 + AD attention + W&T loss	60.3

由实验结果可知, 两者的改动均能提高网络模型的特征提取与分类能力。

最后, 本文研究采用了梯度加权的类激活热力图 (Grad-CAM)^[19] 来可视化属性引导的注意力机制的效果, 如图 7 所示。由图 7 可知, 无论是对于标准的模特图片、还是复杂的用户图片, 注意力机制都能很好地排除干扰以关注图片中的服装部位。同时, 不同属性的注意力机制模块关注区域也有一定差异, 聚焦对应的属性特征。

4.2 属性检索实验

为验证属性解纠缠引导的特征提取网络上的改进是否能有效提升服装属性操作检索任务的性能, 对本文提出的检索模型进行了服装图像检索任务上的消融实验, 实验的评估指标为 Top - K Accuracy 与归一化折损累积增益 NDCG^[20]。其中 Top - K Accuracy 指标能反映模型前 K 个搜索结果中存在符合查询条件的图像的平均概率, 而 NDCG 则通过考虑列表中每个项目的相关性和位置来衡量项目排名列表的质量。消融实验的结果见表 3。

由实验结果可知, 提出的 2 个改进在基线模型 ResNet 上均有正向的提升。

为验证本文提出的服装属性操作检索网络的优越性, 本文选取了只使用图像检索的基线与几个同类方法的服装属性操作检索模型在相同数据集上进行对比试验, 评估指标同上, 实验结果见表 4。由表 4 可以看出本文提出的方法相比同类方法在 Top - K Accuracy 指标与 NDCG@30 上均有一定的提升。

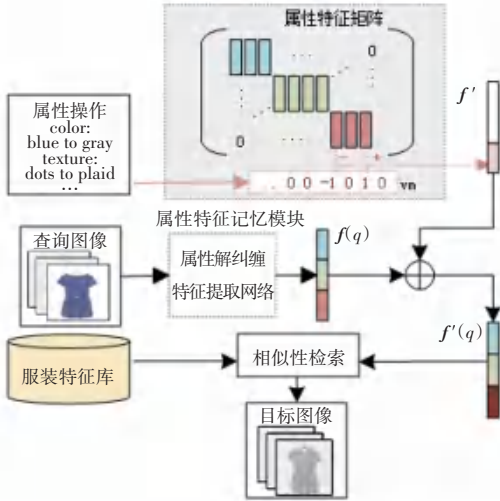


图 6 属性操作检索实现

Fig. 6 Implementation of attribute manipulation retrieval

4 实验与分析

本文实验均在 Ubuntu 18.04 LTS 环境中运行, CUDA 版本为 12.2, 模型基于 Pytorch 深度学习框架实现。实验采用公开服装数据集 DeepFashion^[16] 的类别属性基准子数据集, 在分类任务实验上选择类别、纹理与形状三种属性标签, 并按照 4:1 的比例划分共 113 948 张图片为训练集与测试集。在属性操作检索任务实验上, 随机选取并构建了“查询图片、属性操作、目标图片、负面图片”的带属性操作的三元组训练样本 780 336 组训练三元组与 204 221 组测试三元组。

4.1 特征提取网络实验

为了考察不同骨干网络的属性解纠缠特征结构的特征提取能力, 本文首先实验并评估不同骨干网络引入 AFD 模块后输出属性解纠缠特征在数据集上的分类能力, 采用评估指标为分类准确率 (Accuracy), 实验结果见表 1。表 1 中, w/FD 表示骨干网络结合了属性解纠缠模块结构。

由实验结果可知, ResNet50 相比其他网络准确率更高, 更加适合作为属性解纠缠的特征提取网络的骨干网络。

为了验证改进的属性引导注意力机制 (AD attention) 与标签平滑交叉熵损失函数和局部-全局三元组损失函数 (W&T loss) 的对模型特征提取能

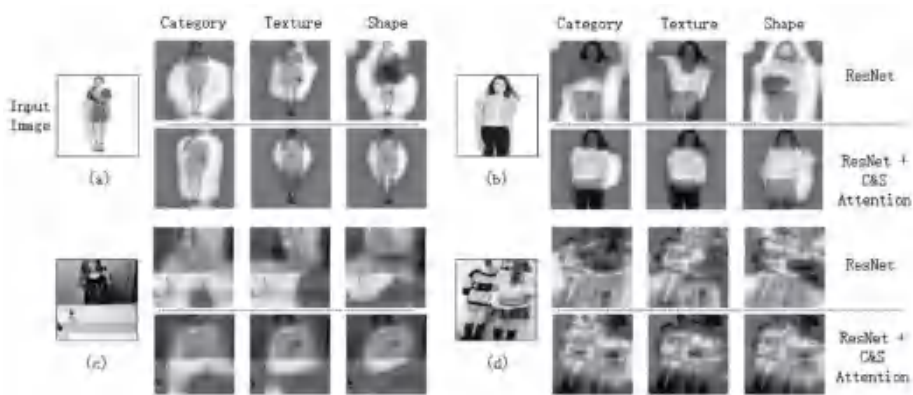


图 7 注意力机制的激活热力对比图

Fig. 7 Activation heatmap comparison of attention mechanisms

表 3 属性操作检索消融实验结果

Table 3 Results for ablation experiments retrieved by attribute manipulation

模型	Top-10 acc/%	Top-20 acc/%	Top-30 acc/%	Top-50 acc/%	NDCG@ 30
ResNet50(baseline)	20. 4	25. 5	28. 7	33. 0	0. 330
ResNet + AD attention	21. 8	27. 4	30. 5	34. 7	0. 332
ResNet + W&T loss	21. 4	26. 8	30. 1	34. 2	0. 328
ResNet + AD attention + W&T loss	24. 4	29. 0	32. 0	35. 6	0. 335

表 4 同类模型对比实验结果

Table 4 Results for comparative experiments with similar models

模型	Top-10 acc/%	Top-20 acc/%	Top-30 acc/%	Top-50 acc/%	NDCG@ 30
Image only	2. 3	4. 1	5. 6	8. 1	0. 318
AMNet ^[5]	14. 1	19. 4	22. 9	27. 6	0. 282
Adde-m ^[14]	23. 0	27. 7	30. 9	35. 1	0. 329
本文模型	24. 4	29. 0	32. 0	35. 6	0. 335

本文服装属性操作检索模型的检索示例如图 8 所示,左边为输入的查询图像与属性修改操作,右边是模型根据查询输出的检索结果。在检索结果前 5

个中,深色框中的结果为符合预期条件的检索结果。可以看出在保证服装其他属性基本不变的情况下,模型能够检索到修改特定属性的服装结果。



图 8 属性操作检索结果示例

Fig. 8 Example of attribute manipulation search results

5 结束语

本文针对服装图像背景姿势等无关信息多样复杂、属性标签多样,以及属性操作检索任务中属性操作与图像特征结合与服装特征解纠缠的问题,提出了结合注意力机制与属性特征解纠缠的服装属性操作检索模型,通过对特征提取网络与损失函数等改进使模型更好地实现服装图像属性操作检索任务。并通过实验验证了模型的优越性与其在属性操作检索任务上的应用价值。

参考文献

- [1] 程丽文,刘晴晴,武年钰,等. 新零售背景下服装品牌消费体验提升策略研究[J]. 中国市场,2021(16):104-106.
- [2] CHENG Wenhuan, SONG Sijie, CHEN C Y, et al. Fashion meets computer vision: A survey [J]. ACM Computing Surveys (CSUR), 2021, 54(4):72.
- [3] ISLAM S M, JOARDAR S, SEKH A A. A survey on fashion image retrieval[J]. ACM Computing Surveys, 2024, 56(6): 25.
- [4] HARDOON D R, SZEDMAK S, SHAWETAYLOR J. Canonical correlation analysis: An overview with application to learning methods[J]. Neural Computation, 2004, 16(12): 2639-2664.
- [5] ZHAO Bo, FENG Jiashi, WU Xiao, et al. Memory-augmented attribute manipulation networks for interactive fashion search [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1520-1528.
- [6] KENAN E A, KASSIM A A, LIM J H, et al. Learning attribute representations with localization for flexible fashion search [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7708-7717.
- [7] DIVITIIS D L, BECATTINI F, BAECCHI C, et al. Disentangling features for fashion recommendation [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(1s): 1.
- [8] SAHA A, NAWHAL M, KHAPRA M M, et al. Learning disentangled multimodal representations for the fashion domain [C]//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ: IEEE, 2018: 557-566.
- [9] YILDIRIM G, SEWARD C, BERGMANN U. Disentangling multiple conditional inputs in GANs [J]. arXiv preprint arXiv, 1806.07819, 2018.
- [10] LIN Y L, TRAN S, DAVIS L S. Fashion outfit complementary item retrieval [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 3311-3319.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 3-19.
- [13] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 815-823.
- [14] HOU Yuxin, VIG E, DONOSER M, et al. Learning attribute-driven disentangled representations for interactive fashion retrieval [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 12147-12157.
- [15] SCARAMUZZINO G, BECATTINI F, DEL B. Attribute disentanglement with gradient reversal for interactive fashion retrieval [J]. Pattern Recognition Letters, 2023, 172: 203-212.
- [16] LIU Ziwei, LUO Ping, QIU Shi, et al. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1096-1104.
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. Nevada, USA: NIPS Foundation, 2012: 1097-1105.
- [18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv, 1409.1556, 2014.
- [19] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 618-626.
- [20] WANG Yining, WANG Liwei, LI Yuanzhi, et al. A theoretical analysis of NDCG type ranking measures [J]. arXiv preprint arXiv, 1304.6480, 2013.