

面向印花图案检索的特征融合方法研究

殷业瑜, 高家全, 李 莹

(南京师范大学计算机与电子信息学院/人工智能学院, 江苏 南京 210023)

[摘要] 伴随着纺织品 CAD 技术的发展,印花图案数量快速增长. 快速准确地在企业图库中查找到相似印花图案,对于帮助纺织企业极大程度降低成本,提升生产效率,具有重要意义. 本文针对印花图案检索问题,以 ResNet 为主干网络构建了一个基于特征融合的印花图案检索模型 PGLN(Pattern Global and Local feature Network). 在该模型中,将全局特征和局部特征进行融合,全局特征使用深度网络的池化特征图,高效地整合输入图像的显著特征;局部特征分支借由注意力机制,使用交互特征层来检测图像的显著性区域. 为了验证 PGLN 模型的有效性,本文在自主构建的印花图案数据集(Pattern)上对 PGLN 模型的检索效果进行了测试. 实验表明,与局部特征提取算法、全局特征提取算法以及融合特征算法相比,PGLN 模型在 Pattern 数据集检索任务上均取得了最优的表现.

[关键词] 印花图案,图像检索,特征提取,特征融合

[中图分类号] TP301 **[文献标志码]** A **[文章编号]** 1001-4616(2022)02-0118-08

Research on Image Feature Fusion Method for Pattern Image Retrieval

Yin Yeyu, Gao Jiaquan, Li Ying

(School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China)

Abstract: With the development of textile CAD technology, pattern images increases quickly. Finding similar pattern images in the corporate library quickly and accurately can help textile companies greatly reduce costs and improve production efficiency. Aiming at the problem of pattern image retrieval, this paper uses ResNet as the backbone network to construct a pattern retrieval model PGLN(Pattern Global and Local Feature Network) based on feature fusion. In this model, the global feature uses the pooled feature map of the deep network to efficiently integrate the salient features of the input image; the local feature branch uses the attention mechanism and uses the interactive feature layer to detect the salient area of the image. In order to verify the effectiveness of the PGLN model, this paper tests the retrieval effect of the PGLN model on the self-built pattern dataset(Pattern). Experiment results show that compared with local feature extraction algorithms, global feature extraction algorithms, and fusion feature algorithms, the PGLN model achieves the best performance for the pattern image retrieval task.

Key words: pattern image, image retrieval, feature extraction, feature fusion

自 20 世纪末,关于图像检索的研究就陆续展开. 一般来看,图像检索根据检索形式可分为两类:基于文本的图像检索^[1](Text Based Image Retrieval, TBIR)和基于内容的图像检索^[2](Content Based Image Retrieval, CBIR). TBIR 是通过文字信息检索与此关联的图片资源, CBIR 则是直接通过用户上传的图片检索关联图片. TBIR 的优点是查询方便快捷,缺点是其对图片的描述比较粗略,检索结果往往不够精确,用户需要浏览大量的图片才能最终筛选出自己想要的图片. 因此,检索性能更优的 CBIR 逐渐成为主流检索方法.

CBIR 检索技术主要包含特征提取、相似性度量、重排序 3 个核心步骤. 特征提取是提取图像的代表性特征以降低图像的维度,通常以 1 个特征向量的形式表示 1 张图片;度量学习是利用度量函数计算图像特征之间的距离,得到特征之间的相似度. 在训练网络时,将其作为损失函数,使得相似图像之间的距离更近,不相似的图像距离更远;而重排序则是利用额外的度量方式,进一步修正检索结果,从而提升检索效果.

收稿日期:2021-09-29.

基金项目:国家自然科学基金项目(61872422).

通讯作者:高家全,博士,教授,博士生导师,研究方向:高性能计算,大数据分析可视化. E-mail:springf12@163.com

图像特征提取的好坏会直接影响图像检索精度. 在 CBIR 发展早期,常用的底层特征包括颜色特征、纹理特征、形状特征以及空间关系特征等,这些底层特征通常局限性较强,往往只关注图像的某一方面特征. 为了弥补底层特征的不足,人们提出了直接通过手工设计的图像特征,主要有 SIFT^[3]、GIST^[4] 以及 HOG^[5] 等,这些特征在一定程度上弥补了早期底层特征的缺陷,但是随着图像数量的增多以及图像内容的多样化,手工特征也难以满足人类对于图像表示的需求,这些低层特征难以表示图像的高层语义.

随着深度学习的大放异彩,研究者各类图像数据集中训练网络,利用卷积神经网络 (Convolutional Neural Network, CNN) 学习并提取图像特征. CNN 上的参数通常通过训练海量的数据来学习,训练完成后各神经网络层得到最优的参数,从而提取到优异的图像特征. 利用深度 CNN 提取的特征在理解图像语义方面远胜于手工特征,并且可以利用神经网络中的卷积层、池化层以及全连接层等各层参数的设置来调整模型的拟合能力,使提取到的图像特征更科学,更能代表图片的高层语义. 从 2012 年 Krizhevsky 等^[6] 提出 AlexNet 模型后,研究者为了得到更优异的、更深层次的特征,设计的网络越来越深,如 GoogLeNet^[7]、VGG^[8] 等. 但是随着网络层次的加深,训练难以收敛,检索的精度反而开始下降,故而 He 等对 CNN 进行改进,提出了 ResNet^[9-13] 等模型解决网络的退化问题.

此外,随着图像检索在各个领域上的应用愈发广泛,对于检索的要求也随之提高. 单一特征仅仅从图像的一方面特点衡量图像,无法准确地描述出不同图像之间的区别. 因此,通过单一特征的图像检索无法满足精确检索的要求,针对印花图案的检索则更为复杂. 为了解决印花图案检索问题,本文提出一种通过融合多个特征以此表征图像进行检索的方案,即从多个特征中选取具有互补性的特征,利用合理的方法进行融合. 实验证明,通过融合形成的新特征可以从多个角度衡量图像,具有更好的表述力.

1 基于特征融合的印花图案检索模型 PGLN

本节将详细描述基于特征融合的印花图案检索模型 PGLN 的网络层次结构和训练流程.

图 1 展示了本文提出的基于特征融合的印花图案检索模型 PGLN. 模型利用卷积神经网络学习深度全局和局部特征,结合 GeM 池化和注意力机制检索相似图像. 第一阶段使用全局特征进行检索,以有效地选择最相似的图像. 第二阶段使用局部特征对检索结果进行重排序,从而进一步提高系统的检索精度.

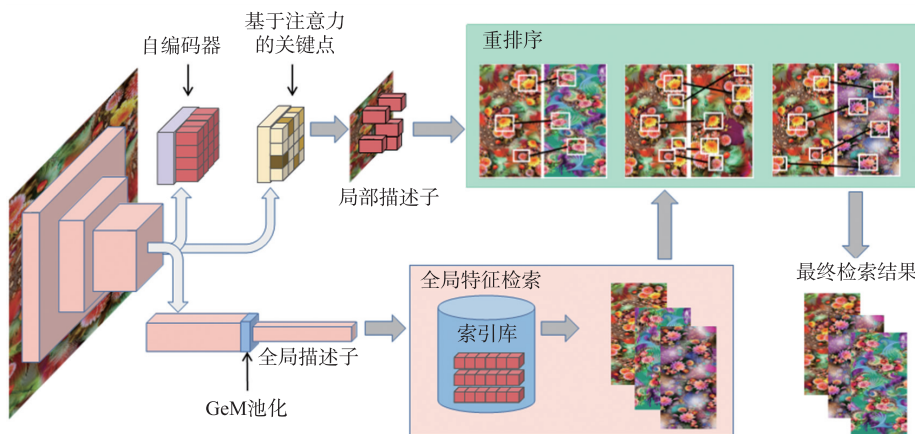


图 1 PGLN 模型结构

Fig. 1 The structure of PGLN model

1.1 全局特征提取

ResNet 网络结构能够有效缓解梯度消失的问题,其中 ResNet-50 模型更是兼顾准确性和运算量. 鉴于此,本文的 PGLN 网络结构使用 ResNet-50 作为主干网络.

给定一幅图像,本文利用 ResNet-50 进行特征提取,得到两个特征图: $S \in R^{H_s \times W_s \times C_s}$ 和 $D \in R^{H_d \times W_d \times C_d}$, 分别表示较浅的和较深的激活图,其中 H, W, C 分别表示通道的高度、宽度和数量. 对于常见的卷积网络, $H_d \leq H_s, W_d \leq W_s$ 和 $C_d \geq C_s$, 越深层次的特征图越小,通道数量更多, $s_{h,w} \in R^{C_s}$ 和 $d_{h,w} \in R^{C_d}$ 表示 h, w 位置的特征图. 对于一般的网络设计,这些特征值都是非负的,因为它们是在 ReLU 非线性之后得到的.

为了将深度激活图聚合为一个全局特征,本文采用了广义平均池 (GeM), 它有效地加权了每个特征

的贡献. 全局特征学习的另一个关键组成部分是对聚合表示进行白化; 本文通过一个全连接层 $F \in R^{C_F \times C_D}$ 将它集成到本文的模型中, 其学习到的偏置为 $b_F \in R^{C_F}$. 这两个部分整合成一个总结了整个图像的判别内容的全局特征 $g \in R^{C_F}$:

$$g = F \times \left(\frac{1}{H_D W_D} \sum_{h,w} d_{h,w}^p \right)^{1/p} + b_F, \quad (1)$$

式中, p 为广义平均池化的超参数. $d_{h,w}^p$ 为特征映射图中的每一个特征.

1.2 局部特征提取

对于局部特征, 只选择相关的区域进行匹配是很重要的. 这可以通过增加注意模块 M 来实现, 该模块的目标是预测具有判别性的局部区域. 设为 $A = M(S)$, 其中 M 是一个小的卷积网络, $A \in R^{H_S \times W_S}$ 表示与 S 特征相关联的注意力得分图.

此外, CNN 网络提取出的图像局部特征数以万计, 容易导致维度过高, 所以必须对它们进行紧凑的表示. 为此, 本文增加一个小的卷积自编码器 (AE) 模块, 它负责学习合适的低维表示. 局部描述符为 $L = T(S)$, 其中 $L \in R^{H_S \times W_S \times C_T}$, T 是自编码器的编码部分, 对应为一个带有 C_T 滤波器的 1×1 卷积层. 与 S 相反, 局部描述符 L 不被限制为非负的.

因此, 每个提取的位置为 h, w 的局部特征都用一个局部描述符 $l_{h,w} \in L$ 来表示, 其对应的关键点检测分数 $a_{h,w} \in A$. 全局描述符和局部描述符经过 $L2$ 正则化为 \hat{g} 和 $\hat{l}_{h,w}$.

1.3 损失函数

对于全局特征的学习, 本文采用一个结合全局损失和局部损失的新损失函数, 融合了 softmax loss 和 cross-entropy loss. 此外还添加了 ArcFace margin, 它通过诱导较小的类内方差, 在全局特征学习方面表现出了良好的效果.

经 Arcface 调整后的余弦相似度可以表示为:

$$AF(u, c) = \begin{cases} \cos(\arccos(u) + m), & \text{if } c = 1, \\ u, & \text{if } c = 0, \end{cases} \quad (2)$$

式中, u 为余弦相似度, m 为 ArcFace margin, c 表示对应标签是否为 ground-truth 类, 值为 1 表示是同一类, 值为 0 则非同一类. 在这种情况下, 用 softmax 归一化计算的 cross-entropy loss 可以表示为:

$$L_g(\hat{g}, y) = -\log \left(\frac{\exp(\gamma \times AF(\hat{\omega}_k^T \hat{g}, 1))}{\sum_n \exp(\gamma \times AF(\hat{\omega}_n^T \hat{g}, y_n))} \right), \quad (3)$$

式中, γ 是一个可学习的标量, $\hat{\omega}_i$ 指类别 i 的 $L2$ 正则化权重, y 为 one-hot 标签向量, k 为各真实类别.

为了训练局部特征, 本文使用了两个损失函数. 首先, 第一个是 MSE loss 损失, 使得自编码器良好地重建 S . 以 $S' = T'(L)$ 作为 S 的重构版本, 两者维数相同, 其中 T' 为一个带有 C_T 滤波器的 1×1 卷积层, 然后是 ReLU. 损失如式 4 所示:

$$L_r(S', S) = \frac{1}{H_S W_S C_S} \sum_{h,w} \|s'_{h,w} - s_{h,w}\|^2. \quad (4)$$

第二个是 cross-entropy loss, 促使注意力模块选择更具判别性的局部特征, 通过注意力权重 $a_{h,w}$ 对重构特征 S' 进行池化, 如式 (5) 所示:

$$a' = \sum_{h,w} a_{h,w} s'_{h,w}. \quad (5)$$

然后使用一个标准的 softmax-cross-entropy loss:

$$L_a(a', k) = -\log \left(\frac{\exp(v_k^T a' + b_k)}{\sum_n \exp(v_n^T a' + b_n)} \right), \quad (6)$$

式中, v_i 和 b_i 分别指第 i 类的分类器的权重和偏差, k 为各真实类别. 该损失函数的目的是为了使得判别性特征的注意权重变大. 综上, PGLN 网络结构的损失为 $L_g + \lambda L_r + \beta L_a$.

考虑到 CNN 网络提取出的图像局部特征数以万计, 容易导致维度过高, 所以必须对图像特征进行紧凑的表示. 为此, 本文增加一个小的 AE 卷积自编码器模块, 它负责学习合适的低维表示.

$$L=T(S), L \in R^{H_S \times W_S \times C_T}, \quad (7)$$

式中, T 是自编码器的编码部分, 对应为一个带有 C_T 滤波器的 1×1 卷积层. 与 S 相反, 局部描述符 L 不被限制为非负的.

2 实验结果与分析

为了衡量 PGLN 模型在印花图案检索任务上的检索效果, 验证特征融合方式的有效性, 本文采用 Pattern 数据集中的 Design 设计图子集进行实验, 将 PGLN 模型与局部特征提取算法、全局特征提取算法以及特征融合算法分别进行了比较.

2.1 特征提取算法

图像检索主要包含特征提取、相似性度量、重排序这 3 个核心模块. 在特征提取阶段, 提取出的特征优劣将直接影响图像检索的精度. 为了提高图像检索性能, 需要两种类型的图像表示: 全局特征和局部特征. 一方面, 全局特征通常被称为“全局描述符”, 它通常以一种紧凑的表示总结图像的内容. 另一方面, 局部特征包含特定图像区域的描述符和几何信息. 总的来说, 全局特征的查全率更高, 而局部特征的查准率更高. 全局特征可以学习非常不同姿态的相似性, 而局部特征无法找到相应关系; 与此同时, 基于局部特征的几何验证所提供的分数通常能很好地反映图像相似性, 比全局特征距离更可靠. 然而, 这两种特征在特征表示方面具有显著差异, 因此, 本文采用特征融合方法, 克服传统特征提取方法较为片面、底层特征拘泥于图像细节、深层特征只考虑图像整体等局限性.

2.2 实验数据集和实验方法

在实验中, 本文采用的数据集为 Pattern 数据集中的 Design 设计图子集, 图 2 展示了实验数据集的部分图片.

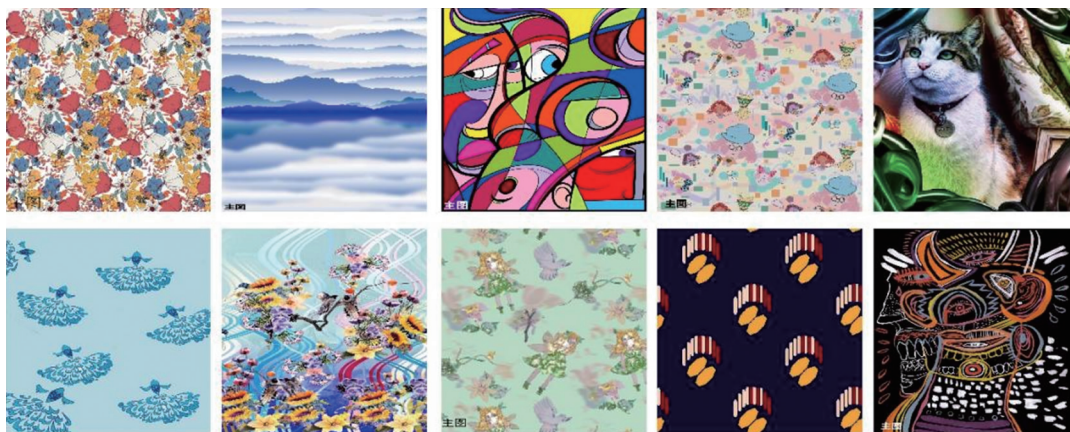


图 2 Pattern 数据集中的 Design 设计图子集

Fig. 2 A subset of the Pattern data set named Design data set

设计图 (Design): 包含组成元素、图案风格、排列方式等 3 种类别标签, 共计 29 种标签, 例如花卉、田园风、单独图样等. 数据规模为 20 000 张图片, 主要是由专业设计人员利用 CAD 软件设计得到的高清设计图. 图片通常较大, 大小一般在 1M~10M 之间, 清晰度高.

在实验中, 各项基本参数的设置如下: PGLN 模型使用的主干网络为 ResNet-50, 模型由预先训练的 ImageNet 权重初始化, 根据检索任务进行了微调. 其中, 从 conv4 输出中得到较浅的特征映射 S 作为局部特征, 从 conv5 输出中得到较深的特征映射 D 作为全局特征. D 的通道数是 $C_D = 2048$; GeM 池化使用参数 $p=3$. 之后使用白化全连通层, 产生一个维度为 2 048 的全局特征 $C_F = 2048$. S 的通道数是 $C_S = 1024$; 自动编码器模块学习这个特征映射的降维 $C_T = 128$. 注意力网络 M 有 2 个卷积层, 卷积核大小为 1×1 ; 第一层使用 ReLU 作为激活函数, 第二层使用 Softplus 作为激活函数. 检索的评价指标, 选择平均检索精度 mAP , 该项指标能够较为全面地反映检索的准确率.

2.3 对比实验分析

实验 1: 为了验证 PGLN 算法在印花检索任务上的有效性, 我们将 PGLN 与 SIFT、DELFI (DEep Local

Feature) 等局部特征提取算法进行了对比。

目前较为常用的局部特征包括基于手工的传统特征与基于深度学习的局部特征。其中,传统局部特征提取算法包括 SIFT 算法、LBP 算法等,基于深度学习的局部特征提取包括 DELF 算法等。在实验中,本文以 Design 数据集为例,将这 3 种局部特征提取算法与 PGLN 算法进行了对比。

从表 1 可以看出,针对 Design 数据集,SIFT 算法、LBP 算法的检索精度非常低,分别为 48.63% 和 49.32%。DELF 算法的检索精度略高于二者,达到了 56.79%。在该数据集上,PGLN 模型的检索精度最高,达到了 66.73%。由此可以看出,一方面,无论是哪种局部特征提取算法,提取出的局部特征往往只能表征图像的局部细节,并没有深入挖掘特征的全局语义信息。因此,仅使用局部特征提取算法来提取印花图案的特征是远远不够的,本文的 PGLN 模型则结合了全局特征提取算法,有效提升了检索的精度。另一方面,Design 数据集为 Pattern 数据集中的设计图子集,图片的内容最为丰富多样且标签数众多,检索的难度较高。4 种算法在该数据集上的检索精度尚未达到工业应用的检索要求。

本文从 Design 数据集上选取了一张印花图片作为查询图片,对比了几种局部特征提取算法的检索效果,检索结果如图 3 所示。

表 1 不同局部特征提取算法在 Design 数据集上的检索结果
Table 1 Retrieval results of different local feature extraction algorithms on the Design data set

| 种类 | 模型 | 准确率/% |
|-------------|----------------------|-------|
| Traditional | SIFT ^[3] | 48.63 |
| | LBP | 49.32 |
| Deep | DELF ^[10] | 56.79 |
| | PGLN | 66.73 |

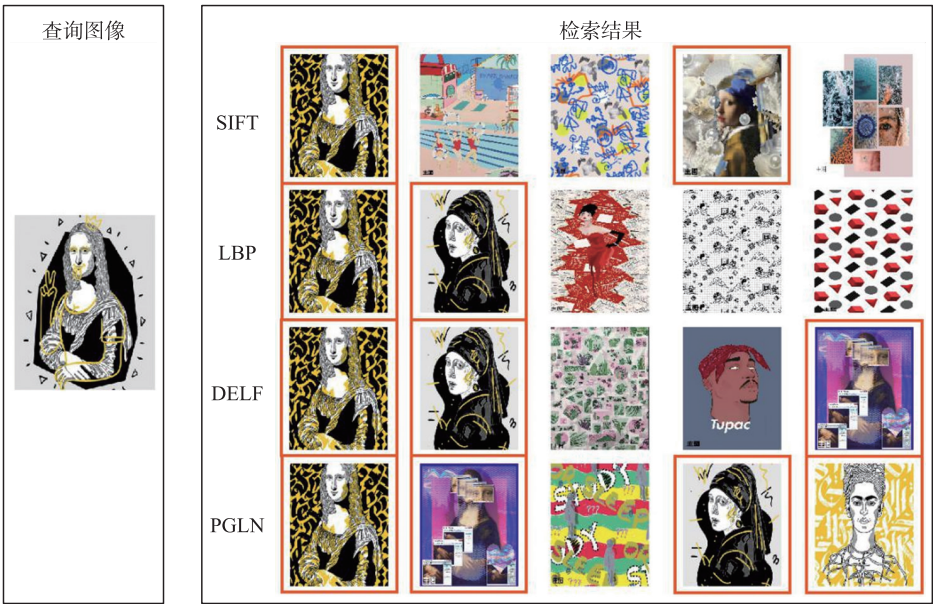


图 3 不同局部特征提取算法的检索结果

Fig. 3 Retrieval results of different local feature extraction algorithms

本文选择了一张具有“人”“头像”“三角”“抽象”等 4 个标签的印花图片作为查询图像,红框内为检索正确的图片。从图 3 来看,SIFT 算法与 LBP 算法的检索效果较差,检索出的前 5 张图片中仅有 2 张与查询图像相似。作为深度局部特征提取算法,DELF 算法的检索效果要优于 SIFT 和 LBP 这两种传统局部特征提取算法,检索出的前 5 张图片中有 3 张与查询图像相似。PGLN 模型则展现出更为精确的检索效果,检索出的前 5 张图片中仅有 1 张与查询图像不相似。

实验 2:为了验证 PGLN 算法相较于全局特征提取算法的优势,我们将 PGLN 与 HOG、VGG、ResNet 等全局特征提取算法进行了对比。

常用的全局特征包括传统全局特征和深度全局特征。传统全局特征包括 Color Histogram 特征、HOG 特征和 Gabor 特征等。Color Histogram 特征体现图像的颜色分布,HOG 特征更关注图像中目标的形状轮廓,Gabor 特征则常常用来描述图像的全局纹理信息。深度全局特征包括 VGG、ResNet 等深度网络模型提取出的图像特征。

表 2 为不同全局特征提取算法在 Design 数据集上的检索结果. 从特征提取方式来看,3 种传统全局特征提取算法在 Design 数据集上的检索精度均未达到 50%,VGG、ResNet 和 PGLN 这 3 种深度全局特征提取算法的检索精度则均在 55%以上,其中 PGLN 模型取得了高达 66.73%的检索精度. 因此,深度全局特征提取算法在该数据集上的检索优势极为明显. 同时,深度网络的层数越深,其提取到的全局图像特征可能更具代表性,检索精度也越高.

为了直观地展示不同全局特征提取算法的检索效果,本文选取了一张标签为“文字”“抽象”的印花图片作为查询图像,红框内为检索正确的图片. 图 4 中,Color Histogram 特征的检索效果较差,只检索出了 1 张与查询图像相似的图片,其余 4 张图片与查询图片只在颜色分布上较为相似. Gabor 特征和 VGG-19 模型均检索出了 2 张相似图片,ResNet-50 则检索出了 3 张相似图片. PGLN 的检索效果最佳,检索出了 4 张相似图片.

实验 3:为了证明 PGLN 算法先用全局特征检索、再用局部特征排序的特征融合策略的有效性,我们将 PGLN 与目前主流的特征融合方法如 FashionNet 等进行了对比.

目前主流的特征融合网络包括 U-net、FashionNet、HyperNet、ExFuse 等,本文在 Design 数据集上验证了各个网络的性能.

表 2 不同全局特征提取算法在 Design 数据集上的检索结果
Table 2 Retrieval results of different global feature extraction algorithms on the Design data set

| 种类 | 模型 | 准确率/% |
|------|--------------------|-------|
| 传统特征 | Color Histogram | 43.21 |
| | HOG ^[5] | 49.13 |
| | Gabor | 47.52 |
| 深度特征 | VGG-16 | 56.55 |
| | VGG-19 | 58.64 |
| | ResNet-18 | 58.71 |
| | ResNet-34 | 59.68 |
| | ResNet-50 | 60.24 |
| | PGLN | 66.73 |

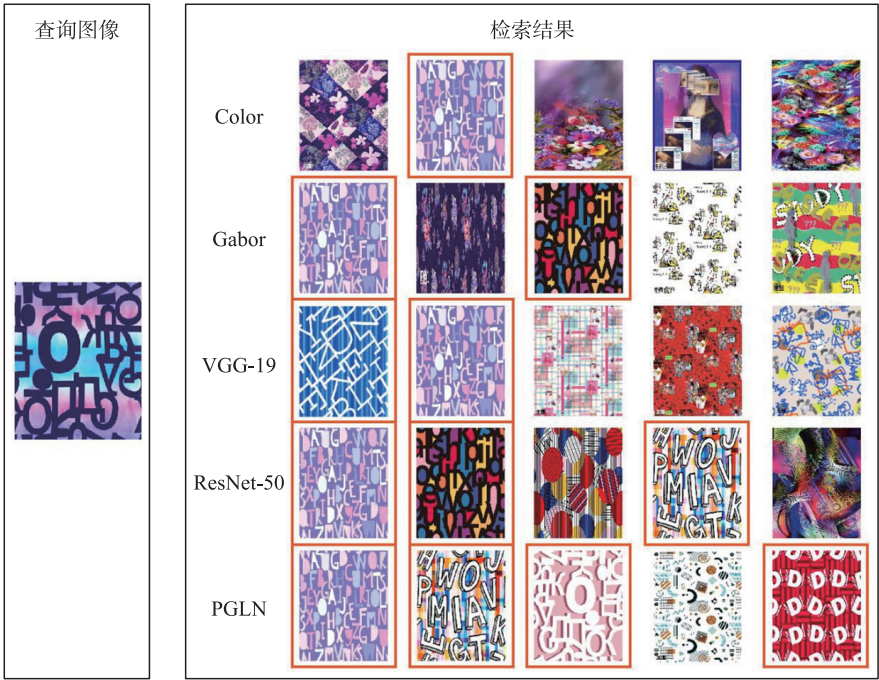


图 4 不同全局特征提取算法的检索结果

Fig. 4 Retrieval results of different global feature extraction algorithms

从表 3 可以看出,PGLN 模型在所有的特征融合方法中取得了最高的检索精度. 这表明,先使用全局特征进行常规检索,再使用局部特征对检索结果重排序的特征融合方式针对印花图案检索效果较好.

此外,本文选取了一张印花图片作为查询图像进行了测试,对比了几种特征融合算法的效果.

图 5 展示了 5 种特征融合算法的检索结果,红框内为检索正确的图片. 查询图像的标签为“花卉”“田

表 3 不同特征融合算法在 Design 数据集上的检索结果
Table 3 Retrieval results of different feature fusion algorithms on the Design data set

| 模型 | 准确率/% | 模型 | 准确率/% |
|----------------------------|-------|--------|-------|
| FPN ^[11] | 61.24 | ExFuse | 62.57 |
| FashionNet ^[12] | 62.96 | PGLN | 66.73 |
| HyperNet ^[13] | 60.41 | | |

园风”。从图中可以看出,FPN、FashionNet、HyperNet、ExFuse 等算法在印花图案检索任务上取得了一定的效果,检索出的前 5 张图片中均包含了 2~3 张正确相似图片;PGLN 模型则取得了最优的检索效果,检索出的前 5 张图片包含了 4 张正确相似图片. 这表明本文的 PGLN 特征融合方法在印花图案检索上具有显著的优势,先通过全局特征进行检索,后通过局部特征重排序的方法能够充分利用图像特征,提升检索准确率.

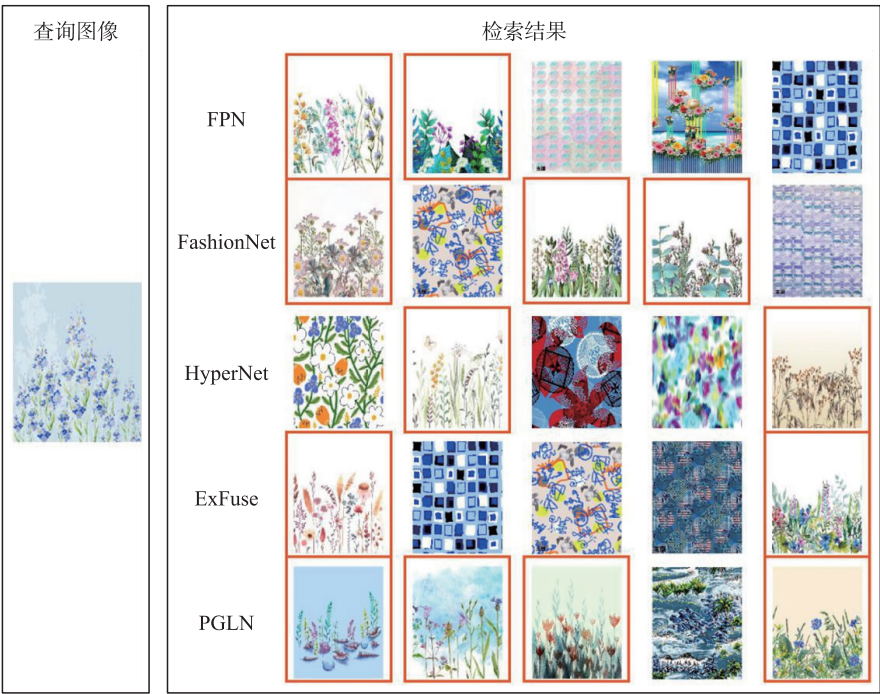


图 5 不同特征融合算法的检索结果

Fig. 5 Retrieval results of different feature fusion algorithms

3 结论

本文针对单一特征对图像的特征能力不足的问题,提出一种融合全局特征和局部特征的网络模型 PGLN,即先通过全局特征进行图像检索,再利用局部特征对检索结果进行排序. 介绍了 PGLN 的网络框架的各部分细节,包括网络的各部分组成、全局特征提取、局部特征提取、损失函数及降维模块. 最后与单一局部特征、单一全局特征以及特征融合等其他特征提取方法进行对比. 结果表明,本文提出的 PGLN 模型,无论与单一局部特征或全局特征相比,针对印花图案的检索效果都是最佳的. 同时,相比于其他的特征融合方式,本文的融合方式也取得了最优表现,证明了先用全局特征检索、再用局部特征排序的特征融合方法的有效性.

[参考文献]

[1] FU X,ZHAO Y,WEI Y,et al. Rich features embedding for cross-modal retrieval;a simple baseline[J]. IEEE transactions on multimedia,2020,22(9):2354-2365.

[2] CEYHUN C,HASAN S B. Content based image retrieval with sparse representations and local feature descriptors;a comparative study[J]. Pattern recognition,2017,68:1-13.

[3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision,2004,60(2):91-110.

[4] OLIVA A,TORRALBA A. Modeling the shape of the scene;a holistic representation of the spatial envelope[J]. International journal of computer vision,2001,42(3):145-175.

[5] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,San Diego,CA,USA:IEEE Computer Society,2005:886-893.

- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]//The 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, Nevada, USA, ACM, 2012: 1106–1114.
- [7] SZEGEDY C, WEI L, JIA Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA: IEEE Computer Society, 2015: 1–9.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//The 3rd International Conference on Learning Representations, San Diego, CA, USA: IEEE, 2015: 45–52.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: IEEE Computer Society, 2016: 770–778.
- [10] HYEONWOO N, ANDRE A, JACK S, et al. Large-scale image retrieval with attentive deep local features[C]//2017 IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017: 3476–3485.
- [11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA: IEEE, 2017: 936–944.
- [12] LIU Z, LUO P, QIU S, et al. Deep Fashion: powering robust clothes recognition and retrieval with rich annotations[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: IEEE, 2016: 1096–1104.
- [13] KONG T, YAO A, CHEN Y, et al. HyperNet: towards accurate region proposal generation and joint object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: IEEE, 2016: 845–853.

[责任编辑: 顾晓天]