

# 基于“点到面”策略的图像检索

顾广华<sup>1,2</sup> 霍文华<sup>1,2</sup> 任贤龙<sup>1,2</sup> 刘江涛<sup>1,2</sup> 苏明月<sup>1,2</sup>

(1. 燕山大学信息科学与工程学院, 河北秦皇岛 066000; 2. 河北省信息传输与信号处理重点实验室, 河北秦皇岛 066000)

**摘要:** 图像检索是计算机视觉领域的一个重要分支。其主要目的是从图像数据库中找出与查询图像相似的语义图像。传统的图像检索方法是在查询图像和数据库图像之间进行“点到点”检索。但是, 单个查询图像包含的类别提示较少, 即类别信息较弱, 使得检索结果并不理想。为了解决这个问题, 本文提出了一种基于“点到面”的类别检索策略来扩展一个图像(点)到一个图像类别(面), 这意味着从单个查询图像到整个图像类别的语义扩展。该方法挖掘了查询图像的类别信息。在两个常用的数据集上对所提出方法的性能进行了评估。实验表明, 该方法可以显著提高图像检索的性能。

**关键词:** 图像检索; 类别信息; 相似性度量; “点到面”策略

**中图分类号:** TN919.81 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2020.09.011

**引用格式:** 顾广华, 霍文华, 任贤龙, 等. 基于“点到面”策略的图像检索[J]. 信号处理, 2020, 36(9): 1464-1470. DOI: 10.16798/j.issn.1003-0530.2020.09.011.

**Reference format:** Gu Guanghua, Huo Wenhua, Ren Xianlong, et al. “Point-to-Flat” Strategy-Based Image Retrieval [J]. Journal of Signal Processing, 2020, 36(9): 1464-1470. DOI: 10.16798/j.issn.1003-0530.2020.09.011.

## “Point-to-Flat” Strategy-Based Image Retrieval

Gu Guanghua<sup>1,2</sup> Huo Wenhua<sup>1,2</sup> Ren Xianlong<sup>1,2</sup> Liu Jiangtao<sup>1,2</sup> Su Mingyue<sup>1,2</sup>

(1. School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066000, China;

2. Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao, Hebei 066000, China)

**Abstract:** Image retrieval is an important branch of computer vision. Its main purpose is to find the similar semantic images to the query image from the image database. The traditional image retrieval method is a “point-to-point” retrieval between the query image and the database. However, a single query image contains fewer category hints, that is, the category information is weaker, and the retrieval results are not satisfactory. In this paper, the “point-to-flat” category-based retrieval strategy is proposed to extend an image (point) to an image category (flat), which means the semantic extension from the individual query image to the whole image category. The proposed method mines the category information of the query image. The performance of the proposed method is evaluated on two commonly used databases. The experimental results demonstrate that the proposed method can significantly improve the performance of image retrieval.

**Key words:** image retrieval; category information; similarity metric; “point-to-flat” strategy

## 1 引言

随着搜索引擎和社交网络中多媒体数据的快速增长, 从一个庞大的图像数据库中快速找到一个

特定的图像是至关重要的。基于内容的图像检索(CBIR)是一项极具挑战的计算机任务, 并且得到了长期的研究关注<sup>[1]</sup>。给定一个包含特定实例(例如特定目标、场景、建筑等)的查询图像, 图像检索旨

在从数据库图像中找到包含相同实例的图像<sup>[2]</sup>。传统的基于内容的图像检索系统主要使用颜色、形状、纹理等底层视觉特征。这些特征的最大问题是无法处理语义鸿沟<sup>[3]</sup>(即机器从低级视觉特征获得的相似性与人类从高级语义特征获得的相似性之间的差距)。即使在 GIST、SIFT 等人为特征上有所改进,仍不能很好地找到相似的语义图像。因此,尽管已经提出了一系列关于图像检索的技术,但是由于语义上的差距,它仍然是一个开放且具有挑战性的问题。随着深度学习<sup>[4]</sup>在计算机视觉领域取得的重大突破,基于深度卷积神经网络(convolutional neural network, CNN)的图像描述已经广泛应用于图像分类<sup>[5]</sup>、图像分割<sup>[6-7]</sup>、目标检测<sup>[8-9]</sup>和图像检索<sup>[10-11]</sup>等领域<sup>[12]</sup>,且都取得了良好的效果。因此,近年来图像检索的研究主要是利用深度卷积神经网络进行图像表示。

近年来,哈希方法因其速度快和存储优势而被广泛应用于大规模图像检索。哈希是学习一种保留图像相似性的近位表示方法,使相似的图像可以用相似的二进制哈希码表示。在图像检索任务<sup>[13-14]</sup>中,现有的哈希方法大致可以分为两类:独立于数据的方法和依赖于数据的方法。局部敏感哈希(LSH)<sup>[15]</sup>是一种比较有代表性的独立于数据的方法,它利用随机线性投影将附近的数据映射成相似的二进制代码。由于数据独立哈希方法的局限性,近年来的哈希方法尝试利用各种机器学习技术,在给定数据集的基础上学习更有效的哈希函数。数据依赖方法的代表性工作包括迭代量化二进制自动编码哈希(BAE)<sup>[16]</sup>,随机生成哈希

(SGH)<sup>[17]</sup>,非对称离散图哈希(ADGH)<sup>[18]</sup>等。哈希方法更注重计算效率,它将数据映射到一个汉明空间,在这个空间中,图像检索将是快速的。由于图像表示被编码为二进制代码,容易造成信息丢失,检索的准确性通常会略有下降。本文提出的方法是一种实值表示学习方法。

对于实值表示学习,传统的图像检索是在查询图像和数据库之间进行“点到点”(point-to-point, PTP)检索<sup>[19]</sup>。首先从查询图像和数据库图像中提取特征,分别生成图像的特征描述。然后利用距离度量测量特征之间的相似性,实现传统的图像检索。然而,单个图像包含较少的类别信息,也就是说,只有一个图像无法表示图像类别的特征。例如,办公室类别可能包含计算机、打印机、扫描仪、书桌、文件夹等。但是并非所有办公室图像都包含上述所有对象。因此,传统的 PTP 检索无法获得令人满意的性能。

为了解决这一问题,本文在传统 PTP 检索的基础上提出了一种新的图像检索方法——基于“点到面”(point-to-flat, PTF)策略的图像检索。该方法挖掘了查询图像的类别信息,实现了从单个图像到整个图像类别的语义扩展。除此之外,提出的 PTF 方法训练成本低,因此在实践中更有效。

## 2 基于“点到面”策略的图像检索

### 2.1 特征学习

本文采用 VGGNet-D 网络结构作为算法的基本组成部分。表 1 显示了该网络的详细配置。

表 1 VGGNet-D 配置

Tab. 1 VGGNet-D configurations

Conv1	Conv2	Conv3	Conv4
64×3×3	64×3×3	128×3×3	128×3×3
Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1
Conv5	Conv6	Conv7	Conv8
256×3×3	256×3×3	256×3×3	512×3×3
Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1
Conv9	Conv10	Conv11	Conv12
512×3×3	512×3×3	512×3×3	512×3×3
Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1	Stride 1, Pad 1
Conv13	FC14	FC15	FC16
512×3×3	1×1×4096	1×1×4096	1×1×1000
Stride 1, Pad 1	dropout	dropout	softmax

VGGNet 通过反复叠加  $3 \times 3$  个小卷积核和  $2 \times 2$  个最大池化层,成功构建了一个 16~19 个深度层的卷积神经网络。具体来说,本文选择 VGGNet-16 (VGGNet-D) 模型提取图像特征。这个 VGGNet-D 模型由 13 个卷积层和 3 个全连通层组成。FC14 或 FC15 的输出通常作为 CNN 的特征描述。本文的主要目的是使用深度神经网络进行特征学习,但是如何设计不同的神经网络并不是本文的重点。其他深度神经网络也可用于执行本文的 PTF 模型的特征学习,这里,本文只是用 VGGNet-D 来说明模型的有效性。

## 2.2 PTF 算法实现

本文将提出的 PTF 策略分为两类:无监督 PTF (N-PTF) 和监督 PTF。N-PTF 没有监督信息来训练分类器。监督 PTF 仅使用类别标签进行简单分类,而在模型训练过程中没有使用任何监督信息。因此,称该方法为弱监督 PTF (W-PTF)。提出的 N-PTF 的算法流程图如图 1 所示。

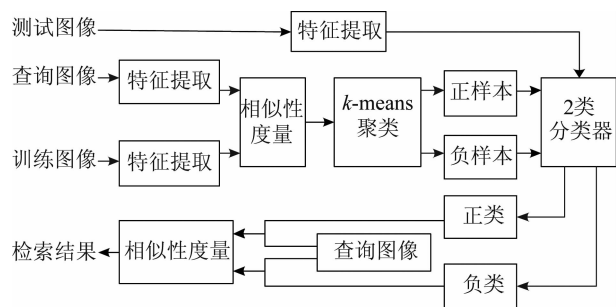


图 1 N-PTF 算法流程图

Fig. 1 Flowchart of N-PTF algorithm

本文将两个图像之间的相似性度量定义为两个相应特征向量的距离度量。在图 1 中,特征提取是通过 VGGNet-D 模型执行的,而相似性度量是通过相关距离进行的。距离越小意味着两个图像的相似度越高。该算法的具体步骤概述如下:

(1) 首先,该算法利用 VGGNet-D 模型中的 fc15 层提取查询图像和训练图像的特征,因此每幅图像的特征都是一个 4096 维的向量。计算每个训练图像与查询图像之间的特征距离并进行排序。

(2) 然后,在距离度量的前  $N$  个序列图像上进行  $k$ -means 均值聚类,以便在同一聚类中收集相似的样本。计算查询图像与  $k$  个聚类中心之间的距离,并选取离查询图像最近的一簇聚类中心图像作为正样本(共有  $N_1$  幅图像),从距离度量  $N+1$  序列之后的图

像中随机选择  $N_2$  个图像作为负样本,然后利用选取的正负本来训练 2 类分类器。支持向量机(SVM)模型将区间最大化问题转换为凸二次规划问题。它可以解决一系列模式识别问题,例如小样本,非线性和高数据维度。因此,本文将 SVM 模型作为 2 类分类器。其他分类器模型也可以代替 SVM 模型,但对不同的分类器选择进行研究并不是本文的重点。

(3) 最后,将测试样本通过训练好的 2 类分类器,可以将测试样本分为正类和负类。这样,分类后的正样本构成一个“面”,包含对应查询图像(“点”)所隐含的特定“类别信息”。也就是说,类别信息由与查询图像具有相同类别标签的多个图像表示。随后对分类后的正类图像和查询图像之间的特征相似性进行度量并排序,对分类后的负类图像也执行相同的操作。最终,将两个排序后的图像序列进行级联以获得图像检索结果。

本文提出的 W-PTF 方法是在 N-PTF 上进行改进的,它们的主要区别在于有无标签信息的使用。在 W-PTF 中,利用 VGGNet-D 模型提取特征后,计算每个训练图像与查询图像之间的特征距离,并对训练样本进行排序。由于训练集的类别标签是已知的,因此本文对排名前  $r$  (top- $r$ ) 的图像进行计数以找到出现最频繁的类别标签。具有这种标签的训练图像被视为正样本,而其他图像则被视为负样本。接下来的算法流程与 N-PTF 方法一样。为了更直观地表达提出的 W-PTF 方法,图 2 展示了该方法的可视化过程。

从图 1 和图 2 可以清楚地看出,所提出的 PTF 方法将单个图像的查询信息丰富到查询图像所属的图像类别中。也就是说,将单一的基于图像的检索转换为基于语义分类的检索。基于语义分类的检索在图像检索任务中具有更大的意义,在接下来的实验中将验证该方法的有效性。

## 3 实验结果与分析

### 3.1 实验设置

本文在两个广泛使用的基准数据集上进行了实验: Caltech256<sup>[20]</sup> 和 Paris<sup>[21]</sup>。Caltech256 数据集包含与 256 个对象类别相关联的 30607 个图像,每个类别包含至少 80 张图像,最多 827 张图像。原始的 Caltech-101<sup>[22]</sup> 是通过选择一组对象类别,从谷歌图像中下载示例,然后手动筛选不符合类别的所有

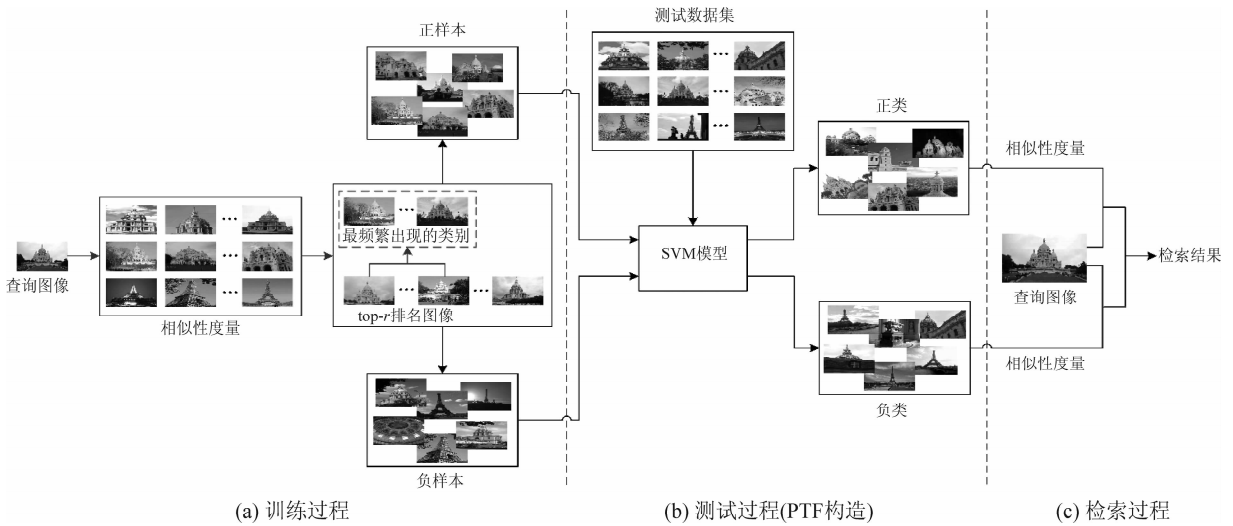


图 2 W-PTF 方法的可视化

Fig. 2 Shows the W-PTF in visualization

图像来收集的。Caltech256 分类数据集是从 Caltech-101 数据集开发的。

Paris 数据集包含来自 Flickr 的 6412 张高分辨率图像,这些图像是通过查询“巴黎埃菲尔铁塔”或“巴黎金字塔”等巴黎著名地标的相关文本标签获得的。原始的注释(标记)是手动执行的,由 11 个地面真实标签组成。给定一个地标查询图像,目标是检索描述相同地标的数据库图像。

对于 Caltech256 数据集,大多数类别包含的图像在 80 到 150 之间。为了增加每类图像的数量,我们对每个类别随机采样 80 张图像,然后翻转并裁剪这些图像以使每类图像数量达到 320。对于 Caltech256 数据集的每个类别,随机选择 10 张图像作为查询图像,50 张图像作为测试图像,其余 260 张图像作为训练图像。也就是说,查询集、测试集和训练集中分别有 2560 张图像、12800 张图像和 66560 张图像。对于 Paris 数据集,类似于 Caltech256 数据集,我们从每个类别中随机抽取 260 张图像作为实验数据集,其中查询、测试和训练图像的数量分别选择为 10、50 和 200。即,查询集、测试集和训练集中分别有 110 张图像 550 张图像和 2200 张图像。

所有实验均采用 MatConvNet<sup>[23]</sup> 实现。工具箱中有很多 MatConvnet 版本,本实验选用 MatConvnet-1.0-beta25 进行实验。所有实验重复五次,每次使用不同的随机选择的查询、训练和测试图像。报告

的结果是五次实验的平均结果。

### 3.2 评估标准

图像检索的目的是找出与查询图像相似的图像,因此评价指标必须考虑检索到的相似图像的数量和排序水平。为了定量地测量所提出的方法和其他比较方法,本文采用两种广泛使用的评价指标来评价图像检索质量:不同比特数的平均准确率均值(mean average precision, mAP)和不同检索返回样本数的查准率-查全率(Precision-Recall)曲线,简称 P-R 曲线。

查准率反映了检索系统返回相关信息的正确能力。定义为:

$$\text{Precision} = \frac{tp}{k} * 100\% \quad (1)$$

其中  $tp$  为检索到的正样本总数,  $k$  为检索到的所有样本总数。

查全率反映了检索系统返回相关信息的完整性能力。定义为:

$$\text{Recall} = \frac{tp}{n} * 100\% \quad (2)$$

其中  $n$  为所有正样本的总数。

查准率仅考虑检索到的图像数量,而没有考虑其排名顺序。在图像检索系统中,检索到的图像的排序也是必要的。为了兼顾图像检索的精度和检索的先后顺序,我们还采用 mAP 来评估本文的实验。mAP 计算如下:

$$\text{mAP}(q) = \frac{1}{q} \sum_{i=1}^q \frac{1}{n_i} \sum_{j=1}^{n_j} \frac{j}{R_j} \quad (3)$$

其中  $q$  为查询图像的数量,  $n_i$  为第  $i$  个查询图像返回的正类图像的总数。  $R_j$  表示所有返回图像的排序顺序中第  $j$  个正类图像的索引。

### 3.3 相似性度量分析

图像检索的两个关键问题是特征描述和相似性度量。对于特征描述, 本文均采用 VGGNet-D 进行特征提取。对于相似度量问题, 常见的距离度量包括欧氏距离、相关距离、切比雪夫距离、马氏距离、汉明距离等。为了比较各种相似性度量方法的准确性, 本文分别采用相关距离, 欧氏距离和切比雪夫距离在传统 PTF 方法上进行了实验。该实验分别从 Caltech256 和 Paris 数据集中为每个类别随机选择 10 张和 50 张图像作为查询和测试样本。表 2 显示了两个数据集上三种距离度量方法的 mAP 结果。

表 2 两个数据集上三个距离度量的 mAP 结果

Tab. 2 mAP results of three distance metrics on two databases

数据集	切比雪夫距离	欧氏距离	相关距离
Caltech256	18.49%	37.15%	38.76%
Paris	34.73%	43.11%	44.35%

欧氏距离因其简单性而被广泛使用, 但它不能反映高维空间中两个特征之间的相关性。切比雪夫距离是空间中两点之间的绝对距离, 与该点的坐标直接相关, 与各分量之间的相关性无关。而相关距离考虑了样本之间的相关性。表 2 中的实验结果也表明, 相关距离优于其他两种方法。因此, 在接下来的实验中, 均选择相关距离进行距离度量。

### 3.4 N-PTF 实验分析

对于 N-PTF 方法, 首先对查询与训练图像之间的距离度量结果进行排序。对于 Caltech256 数据集, 本文选取相似度最高的前 260 幅图像进行  $k$ -means 聚类, 得到  $k$  个特征中心。实验中算法的参数  $k$  设置为 3。选择一组聚类中心距离查询图像最近的图像作为正样本, 并从没有参与聚类的图像中随机选择另外 260 幅训练图像作为负样本。与 Caltech256 相似, 对于 Paris 数据集, 本文选取  $k$ -means 聚类后相似度最高的 200 幅图像作为正样本, 并从其余的训练图像中随机选取 200 幅图像作为负样本。

为了验证  $k$ -means 的有效性, 对是否使用  $k$ -means 的效果进行了实验比较。该实验在不使用  $k$ -means 的情况下, 根据相似度排序, 直接从训练图像中选取正样本和负样本。在 Caltech256 中, 分别选取前 40 张训练图像和随机选取 260 张训练图像作为正样本和负样本来训练分类器模型。对于 Paris 数据集, 选取的正样本和负样本数量分别为 40 和 200 张。表 3 显示了该实验在两个数据集上的 mAP 结果。

表 3 N-PTF 的 mAP (%) 结果

Tab. 3 mAP (%) results for N-PTF

方法	Caltech256	Paris
无 $k$ -means	43.35	52.03
有 $k$ -means	45.71	54.47

聚类的目的是将相似的样本聚在一起, 同时分离不同的样本。直观的想法是同一聚类中的相似度应该更高, 而聚类之间的相似度应该更低, 这提高了相似样本之间的相似度, 使检索性能更好。从表 3 可以看出, 使用  $k$ -means 方法的 N-PTF 在 Caltech256 和 Paris 数据集上要优于未使用  $k$ -means 的 mAP 结果高大约 2.4%。因此, 在接下来的对比实验中, N-PTF 方法采用  $k$ -means 聚类的结果。

### 3.5 W-PTF 实验分析

虽然 N-PTF 方法使用  $k$ -means 聚类来提高同构聚类的准确性, 但所选的正样本仍与一些负样本混合, 因此训练的分类器也不是最优的。为了使模型训练更加准确, 本文对 N-PTF 方法进行了优化, 提出了 W-PTF 方法。在 W-PTF 方法中首先对查询图像和训练图像的相似性度量进行排序。然后选择前  $r$  个序列 (top- $r$ ) 图像中出现最频繁的一个图像类别作为正样本, 而将其他类别的图像作为负样本。为了对参数  $r$  进行敏感度分析, 该实验比较了不同 top- $r$  的检索结果。表 4 列出了 W-PTF 方法在 Caltech256 和 Paris 上的 mAP 结果。

表 4 W-PTF 的 mAP (%) 结果

Tab. 4 mAP (%) results for W-PTF

Top- $r$	Caltech256	Paris
$r=1$	94.10	88.52
$r=20$	78.96	83.47
$r=40$	76.14	81.59
$r=60$	75.01	79.26
$r=80$	74.15	78.61

从表4可以看出,W-PTF方法的性能大大优于N-PTF方法。此外,如果 $r$ 的值越小,则mAP结果越好。这表明,如果 $r$ 越小,与查询图像具有相同类别标签的图像在top- $r$ 图像中出现的频率越高,这使分类器训练得更好,检索性能也就越好。特别是,当 $r=1$ 时,mAP在Caltech256上可达到94.10%,在Paris上可达到88.52%,分别比N-PTF在Caltech256和Paris数据集上高48.4%和34.1%。

为了直观进行比较,本文将上述三种方法PTP,N-PTF和W-PTF的检索性能通过如图3所示的P-R曲线进行了评估。在实验中,将返回的检索图像数量设置为10到100,每次递增10,即[10,20,30,...,90,100]。这里的PTP和PTF方法都利用相关距离进行相似性度量。对于W-PTF方法,采用 $r=1$ 的实验设置。

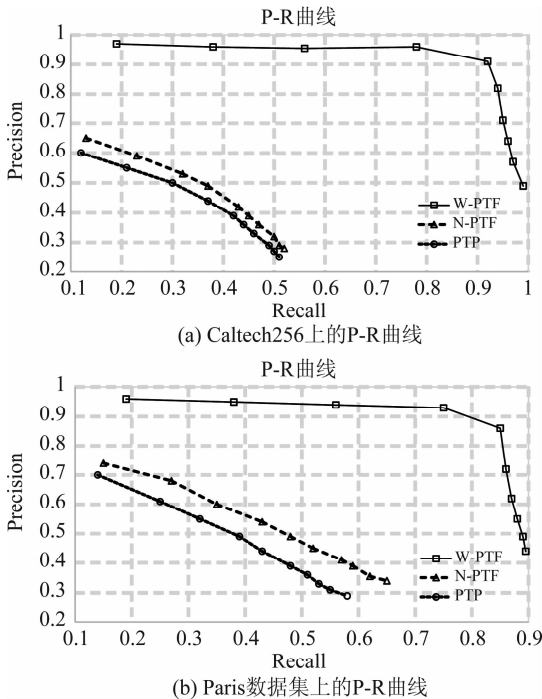


图3 两个数据集上的P-R曲线

Fig.3 Precision-Recall curves on two databases

从图3所示的P-R曲线结果可以很明显地得出结论,所提出的方法在两个数据集上的性能明显优于传统的PTP方法,这与实验在mAP上观察到的结果是一致的。由于单个查询图像中所表达的信息是不完整的,可能只包含部分信息,不能准确反映所检索图像的整体信息。因此,PTP方法的实验结果不尽如人意。在N-PTF和W-PTF实验中,使用SVM分类器将单个查询图像扩展为一组具有相同类别标签的图像,丰富了查询图像的语义信息。这

里,分类器模型训练的性能在整个检索过程中起着至关重要的作用。对于N-PTF,为了提高SVM的性能,该算法在距离度量排序后加入 $k$ -means聚类,使输入分类器的正样本和负样本更加准确。在W-PTF中,利用数据集的标签信息用来选择正负样本。与N-PTF相比,W-PTF方法实现了更卓越的图像检索性能。这表明借助标签信息选择的正样本和负样本更接近真实的正样本和负样本。它进一步说明了本文提出的框架的优越性。

## 4 结论

在本文中,我们提出了一种新的“点到面”(PTF)策略进行图像检索,它将单个查询图像扩展到与查询图像具有相同标签的图像类别。该扩展丰富了查询图像的语义信息,提高了图像检索的性能。所提出的方法试图挖掘类别信息,并将基于图像的检索转换为基于类别的检索。在Caltech256和Paris数据集上的实验结果表明,所提出的PTF方法在图像检索任务上的性能优于其他方法。

## 参考文献

- [1] Zheng Liang, Yang Yi, Tian Qi. SIFT meets CNN: A decade survey of instance retrieval [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (5): 1224-1244.
- [2] Bhattacharyya S, Bhaumik H, De S, et al. Intelligent Analysis of Multimedia Information [M]. Hershey: IGI Global, 2017, 40(5): 143-180.
- [3] Datta R, Joshi D, Jia L I, et al. Image Retrieval: Ideas, Influences, and Trends of the New Age [J]. ACM Computing Surveys, 2008, 40(2): 35-94.
- [4] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [5] Christian S, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.
- [6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [7] Szegedy C, Toshev A, Erhan D. Deep Neural Networks for Object Detection [C] // Advances in Neural Informa-

- tion Processing Systems. NIPS, 2013; 2553-2561.
- [8] Li Haoxiang, Lin Zhe, Shen Xiaohui, et al. A convolutional neural network cascade for face detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015; 5325-5334.
- [9] Sun Y, Wang X, Tang X. Deep Learning Face Representation by Joint Identification-Verification [J]. Advances in Neural Information Processing Systems, 2014, 27: 1966-1988.
- [10] Cimpoi M, Maji S, Vedaldi A. Deep convolutional filter banks for texture recognition and segmentation [J]. International Journal of Computer Vision, 2016, 118(1): 65-94.
- [11] Ghodrati A, Diba A, Pedersoli M, et al. DeepProposal: Hunting Objects by Cascading Deep Convolutional Layers [J]. International Journal of Computer Vision, 2017, 124(2): 115-131.
- [12] Bruna J, Szlam A, Lecun Y. Signal Recovery from Pooling Representations [J]. Statistics, 2013; 307-315.
- [13] Li Jinxing, Zhang Bob, Lu Guangming, et al. Dual Asymmetric Deep Hashing Learning [J]. IEEE Access, 2019, 7: 113372-113384.
- [14] Munjal M N, Bhatia S. A Novel Technique for Effective Image Gallery Search using Content Based Image Retrieval System [C] // International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). Faridabad, India: IEEE, 2019; 25-29.
- [15] Andoni A, Indyk P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions [J]. Communications of the ACM, 2008, 51(1): 117-122.
- [16] Carreira-Perpinan M A, Raziperchikolaei R. Hashing with binary autoencoders [C] // IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015; 557-566.
- [17] Dai Bo, Guo Ruiqi, Kumar S, et al. Stochastic Generative Hashing [J]. Proceedings of the 34th International Conference on Machine Learning, 2017, 70: 913-922.
- [18] Shi Xiaoshuang, Xing Fuyong, Xu Kaidi, et al. Asymmetric discrete graph hashing [C] // In Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2017; 2541-2547.
- [19] Li Zhao, Lu Wei, Xing Weiwei, et al. Image retrieval based on CNN visual features [J]. Journal of Beijing University of Posts and Telecommunications, 2015, 38(B06): 103-106.
- [20] Gregory G, Alex H, Pietro P. Caltech-256 object category dataset [R]. Technical Report 7694, California Institute of Technology, 2007.
- [21] James P, Ondrej C, Michael I, et al. Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2008; 1-8.
- [22] Li Feifei, Fergus R, Perona P. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories [C] // Conference on Computer Vision and Pattern Recognition Workshop. Washington, DC, USA: IEEE, 2004; 178-178.
- [23] Vedaldi A, Lenc K. MatConvNet: Convolutional Neural Networks for MATLAB [J]. Proceedings of the 23rd ACM International Conference on Multimedia, 2015; 689-692.

### 作者简介



**顾广华** 男, 1979年生, 河南濮阳人。燕山大学信息科学与工程学院教授, 博士, 主要研究方向为图像检索、图像分类和图像识别。

E-mail: guguanghua@ysu.edu.cn



**霍文华** 女, 1995年生, 山西汾阳人。燕山大学信息科学与工程学院硕士研究生, 主要研究方向为图像检索和深度哈希。

E-mail: 13273361920@163.com



**任贤龙** 男, 1996年生, 江西赣州人。燕山大学信息科学与工程学院硕士研究生, 主要研究方向为行为识别。

E-mail: m15032358583@outlook.com



**刘江涛** 男, 1994年生, 山西原平人。燕山大学信息科学与工程学院硕士研究生, 主要研究方向为图像检索和深度哈希。

E-mail: ysljt0815@163.com



**苏明月** 女, 1996年生, 河北沧州人。燕山大学信息科学与工程学院硕士研究生, 主要研究方向为深度哈希。

E-mail: smy15076062971@163.com