

doi: 10.11918/j.issn.0367-6234.2016.11.013

# 基于多描述子分层特征学习的图像分类

郭继昌, 王楠, 张帆

(天津大学 电子信息工程学院, 天津 300072)

**摘要:**为解决图像分类任务中词袋(Bag-of-Words)模型分类算法单一局部描述子信息缺失、特征量化误差较大、图像特征表现力不足等问题,提出一种基于多描述子分层特征学习的图像分类方法.结合尺度不变特征变换(SIFT)与形状核描述子(KDES-S)进行局部特征提取,并构建分层特征学习结构来减少编码过程中的量化误差,最后将图像特征分层归一化后进行线性组合并利用线性支持向量机(SVM)进行训练和分类.在Caltech-101、Caltech-256、Scene-15数据库上进行实验,结果表明:相比其他图像分类方法,本文方法在分类准确率上具有显著提升.

**关键词:** 图像分类; 分层特征学习; 分层归一化

**中图分类号:** TP391.4

**文献标志码:** A

**文章编号:** 0367-6234(2016)11-0083-07

## Image classification based on multi-descriptor hierarchical feature learning

GUO Jichang, WANG Nan, ZHANG Fan

(School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China)

**Abstract:** To address the problem that Bag-of-Words model still has several drawbacks such as the scarcity of information in single local descriptor, large quantization error and lack of representation upon image features in image classification tasks, an image classification method based on multi-descriptor hierarchical feature learning is proposed. Combing scale invariant feature transform (SIFT) and kernel descriptors-shape (KDES-S) features, a hierarchical structure is used to reduce quantization error in encoding process, which extracts local features. After that, image features in each layer are normalized respectively, the liner combination of which is the final feature representation for linear support vector machine (SVM) classifier. Experiments are conducted on datasets Caltech-101, Caltech-256 and Scene-15, and experimental results show that the proposed method improves the classification accuracy significantly in comparison with other algorithms.

**Keywords:** image classification; hierarchical feature learning; hierarchical normalization

图像分类作为图像理解的基础,在计算机视觉领域扮演着重要的角色.在图像分类中,词袋模型<sup>[1]</sup>经不断发展已被广泛应用于目标识别、场景分类等领域<sup>[2-3]</sup>,在应用中,其不足主要表现为:单一描述子造成信息缺失,特征编码过程中产生较大的误差进而影响图像特征表现力等.为了解决这些问题,国内外学者一直在对模型进行优化和改进.

词袋模型一般包括特征提取、视觉字典构建、特征编码、特征汇聚等四部分<sup>[4]</sup>.在特征提取中,单一SIFT描述子等由于信息不足,常遇到视觉词语的歧义性和同义性问题<sup>[5]</sup>.为解决这一问题,可以将各

种优秀的局部描述子取长补短,适当结合起来丰富局部特征<sup>[6]</sup>.文献[4]提出结合SIFT与Edge-SIFT特征来丰富图像特征的方法,文献[7]则结合SIFT与HOG描述子来全面描述图像.两种方法都提高了分类精度,但是仍存在不足,文献[4]的方法对于形变较大的目标分类效果欠佳,文献[7]中SIFT与HOG都属于基于梯度方向的描述子,且该方法对分类准确率的提升效果有限.

特征编码作为字典构建和特征汇聚中间的重要环节,编码结果直接影响图像特征的表现力.文献[8]用矢量量化(Vector Quantization)进行特征编码后容易产生较大的量化误差,为减小误差,文献[9]提出稀疏编码(Sparse Coding)法,但其编码稳定性较差.作为改进,文献[10]提出局部约束线性稀疏编码LLC(locality-constrained linear coding),强调了编码过程的局部性,并提高了编码稳定性.文献

收稿日期: 2016-04-28

基金项目: 国家重点基础研究计划(2014CB340400);

天津市自然科学基金(15JCYBJC15500)

作者简介: 郭继昌(1966—),男,博士,教授

通信作者: 郭继昌, jcguo@tju.edu.cn

[11]则利用图像文本和视觉信息对图像进行建模,通过模型训练对图像进行识别.近年来较多的研究集中于基于字典学习<sup>[12-13]</sup>(dictionary learning)的方法和训练多层的深度网络进行图像分类.其中基于字典学习的方法主要通过迭代更新字典来减小量化误差,而基于深度网络的方法通过多层的反复学习来获得高质量特征.文献[14]提出可学习感受野的深度网络,通过学习分类器和感受野来提高分类准确率.文献[15]构造多路径的深度特征学习方法,通过多路径特征的结合提高图像分类精度.特征编码阶段大多方法均是基于单一描述子,容易丢失图像中的显著特征.基于深度网络的方法直接从图像像素中学习特征,其复杂的网络结构对运算要求相对较高.尤为重要,学者们注意力主要集中在编码方法的研究,往往忽略了对编码特征的后续处理,但该过程是分类中是不可或缺的.

针对以上问题,本文提出了一种基于多描述子分层特征学习的图像分类方法,通过多描述子结合来解决特征提取阶段单一描述子信息缺失造成的问题,并构建一个两层的特征学习结构,学习过程中利用批正交匹配追踪法<sup>[16]</sup>BOMP(batch orthogonal matching pursuit)得到稀疏特征后结合空间金字塔结构<sup>[8]</sup>对特征进行空间汇聚(Pooling)和归一化处理,然后作为新的特征进行第二层的特征学习,最后将各层特征结合并利用线性SVM分类器分类.本文结合KDES-S<sup>[17]</sup>与SIFT描述子,利用KDES-S提取图像形状特征补充SIFT特征的信息,提出的分层特征学习方法利用了图像视觉结构的多面性<sup>[18]</sup>并在各层中进行特征归一化,根据图像结构合理分配归一化系数优化特征向量.三种方法相结合,在没有明显增加耗时的前提下提高了图像分类的准确率.

### 1 多描述子分层特征学习

图1为提出的多描述子分层特征学习方法的基本结构,主要包括多描述子提取和分层特征学习两部分.

#### 1.1 多描述子提取

在多描述子提取阶段,选用SIFT与KDES-S描述子相结合,在改善单一SIFT描述子信息不足的同时保证多描述子特征的鲁棒性.与SIFT描述子不同,KDES-S特征是一种基于局部二值模式匹配核的特征,它引入核信息获取图像的局部形状特征且具有较强的适用性,其核函数可以表示为

$$K_{\text{shape}}(P, Q) = \sum_{a \in P} \sum_{a' \in Q} \tilde{s}(a) \tilde{s}(a') k_b(b(a), b(a')) k_p(a, a')$$

式中:  $P$  和  $Q$  表示两个不同的图像块,  $a, a'$  分别表

示图像块中的像素,  $\tilde{s}(a)$  是  $a$  以为中心的  $3 \times 3$  邻域内像素的标准差,  $b(a)$  是二值化  $a$  邻域内像素差值的二元列向量,  $k_b(b(a), b(a')) = \exp(-\gamma_b \cdot \|b(a) - b(a')\|^2)$  用来测量两个像素点的空间距离.为降低计算复杂度,从匹配核中提取紧凑的低维特征并利用核主成分分析法<sup>[19]</sup>将特征维度降为性能最佳的200维.

相比文献[4]需从原图的边界图中提取Edge-SIFT特征,KDES-S直接从原图像中提取即可,大大减少了计算量;而且KDES-S引入核函数获取目标的形状特征,在完善特征的同时可保证特征较高的鲁棒性.在特征结合方式选择上,考虑到分层特征学习法可以充分利用结合后特征的多样性,在提取特征后先对特征进行结合然后再进行特征学习.

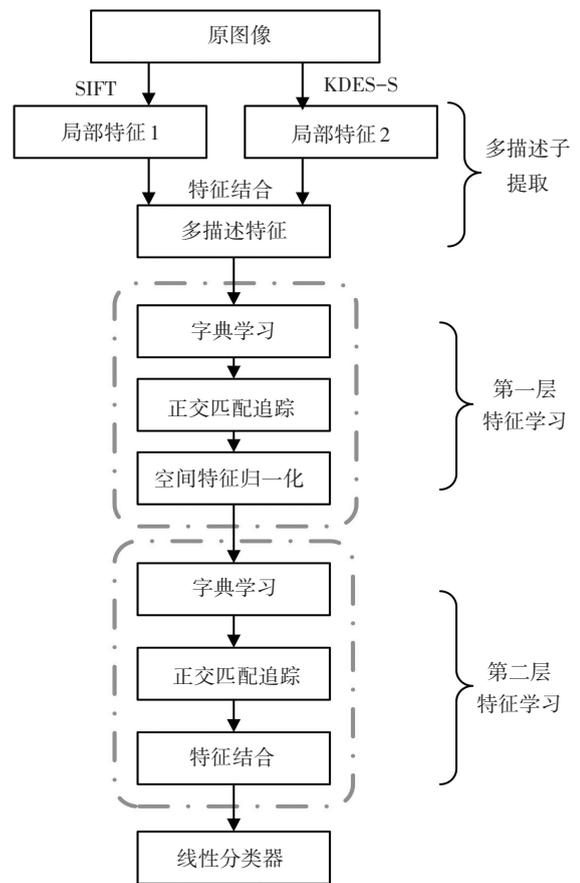


图1 多描述子分层特征学习结构

Fig.1 Structure of multi-descriptor hierarchical feature learning

#### 1.2 分层特征学习

选用MI-KSVD<sup>[15]</sup>(mutual incoherence KSVD)进行字典学习,并结合高效的BOMP方法在获取稀疏编码的同时提高编码效率.得到编码向量后对其进行空间结构划分并进行特征汇聚,根据不同的空间结构分配特征归一化系数进行空间特征归一化,充分利用图像的空间信息.借鉴深度学习的思想,对第一层学习后的特征再次进行学习得到丰富有效

的图像特征. 由于第二层学习建立在第一层基础上, 其特征从深层上利用了图像的空间信息, 对局部形变更具鲁棒性, 且两层学习后特征冗余也得到降低. 照此结构, 可以构建多层的特征学习结构, 学习图像的深层特征, 为了减少计算量, 综合分析各层分类效果后最终选用两层的结构.

第一层的目的是学习多描述子特征, 得到稀疏的编码向量. 重点是对编码特征的后续处理, 为充分利用图像的空间信息, 增加特征的稳定性和局部不变性, 对学习到的特征先以  $16 \times 16$  的块为单位进行空间划分, 如图 2 所示. 并分别对各单元特征进行最大值汇聚 (Max Pooling) 得到汇聚特征:

$$F(C) = \left[ \max_{j \in C} |c_{j1}|, \max_{j \in C} |c_{j2}|, \dots, \max_{j \in C} |c_{jm}|, \dots, \max_{j \in C} |c_{jB}| \right]. \quad (1)$$

式中:  $B$  为视觉单词的个数,  $c_{jm}$  表示单元中第  $j$  个稀疏编码向量  $c_j$  的第  $m$  个元素.

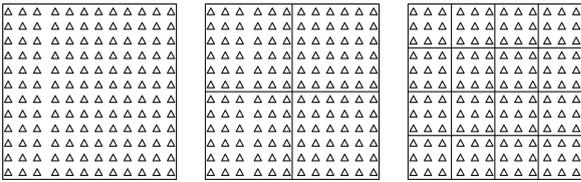


图 2 3 层空间金字塔结构

Fig.2 Three-layer structure of spatial pyramid

则一个图像块  $P$  的特征表示为

$$F_p = [F(C_1^p), \dots, F(C_s^p), \dots, F(C_S^p)],$$

其中  $F(C_s^p)$ ,  $s \in S$  表示由式(1)得到的第  $s$  个空间单元的特征,  $S$  表示空间单元的总数. 利用层次化的空间结构将各单元汇聚后的特征分别归一化处理:

$$\tilde{F}_s = k_s \frac{F_s}{\|F_s\|_p} \quad \forall s = 1, 2, \dots, S. \quad (2)$$

其中  $\|F\|_p = \left( \sum_i F_i^p \right)^{\frac{1}{p}}$  是  $p$  范数<sup>[19]</sup>,  $k_s$  为各空间特征分配的归一化系数, 这里按单元的像素数量正比分配, 如图 2 所示的空间结构, 令  $k_s$  满足比例关系  $k_1 : k_2 : k_3 = 1 : 1/4 : 1/16$ . 需说明的是, 这里各单元特征的分别归一化虽计算简单却可以突出小区域特征的表现力, 尤其归一化参数的引入, 增加了处理后特征的代表性.

第二层的输入是第一层的输出特征, 本层的目的是通过对第一层特征的进一步学习充分利用图像的空间信息, 得到更加丰富有效的图像特征表示. 与第一层不同, 学习到稀疏特征后本层对全图像进行空间金字塔划分, 如图 3 所示. 然后分别对每个单元进行最大值汇聚:

$$\bar{F}(C) = \left[ \max_{j \in C} |z_{j1}|, \dots, \max_{j \in C} |z_{jU}|, \max_{j \in C} |F_{j1}|, \dots, \max_{j \in C} |F_{jV}| \right]. \quad (3)$$

式中:  $z_j$  为第二层得到的稀疏编码,  $F_j$  为第一层产生的块特征.  $U, V$  分别为两种特征的维度. 由式(3)可以看出, 第二层特征汇聚结合了本层的稀疏编码  $z_j$  和第一层的块特征编码  $F_j$ , 结合的特征包含了第一层的细密纹理和第二层的粗糙纹理, 比单层特征更丰富. 各单元的特征结合后得到全图的特征:

$$\bar{F}_l = [\bar{F}(C_1^p), \dots, \bar{F}(C_s^p), \dots, \bar{F}(C_S^p)].$$

利用式(2)对各单元的特征分别归一化后相结合, 构成最终的图像特征:

$$\tilde{F} = [\tilde{F}_1, \dots, \tilde{F}_s, \dots, \tilde{F}_S].$$



图 3 全图空间金字塔结构

Fig.3 Spatial pyramid structure of the whole image

在两层的特征学习过程中, 每一层分别学习本层的字典并根据字典进行稀疏编码. 图 4 为在 Caltech-101 数据集上利用训练图像学习到字典的部分图示. 左右两图分别对应第一层和第二层的字典, 对比两层的字典可以发现, 第二层的字典相比第一层的字典更加细致. 这是因为该字典是在第一层特征的基础上学习得到的, 利用该字典得到的稀疏特征也将更具辨别力.

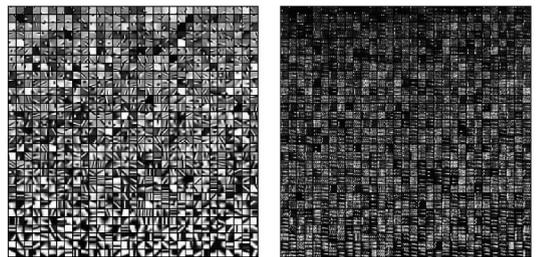


图 4 特征学习阶段的字典

Fig.4 Dictionaries in feature learning phase

SIFT 与 KDES-S 描述子相结合的方法, 结合后特征的维度由 128 维增加到了 328 维, 相比单一特征的方法, 本文方法在计算复杂度上增加了近 1.5 倍, 但是相比于其他基于多描述子结合的方法, 如文献[4]和文献[7]的方法, 本文方法在计算复杂度上没有明显增加. 两层结构中的分层空间特征归一化利用图像空间分层结构, 合理分配权重系数, 对各层特征归一化处理, 不仅可以降低大区域特征对小区域特征的影响还突出了具有辨别力的特征, 而大多数研究者往往忽视了这一工作. 该方法直接利用分

层结构,在几乎没有增加计算量消耗的同时增加了图像特征的表现力,提高了分类准确率.与设计复杂的特征编码算法相比,该方法更加简单可行.

## 2 实验分析

为了验证方法的有效性,首先在 Caltech-101 数据库上分别进行多描述子提取和分层特征学习实验,然后在 Caltech-256 和 Scene-15 数据库上进行整体实验分析.多描述子提取时,SIFT 与 KDES-S 均采用  $16 \times 16$  的图像块,采样间隔均设置为 8 像素.分层特征学习阶段,字典大小设置为 1024,层数设置为 2,空间金字塔划分采用图 2 和图 3 结构.分类阶段直接使用台湾大学开发的 LIBLINEAR-SVM<sup>[20]</sup> 软件包对图像进行训练和测试.一次实验中,分类准确率计算方法为  $Acc = n/N$ ,其中  $n$  为预测正确的图片张数, $N$  为参与测试的总图片的张数.分别在每类中统计,分类结果取 10 次实验的平均值.

### 2.1 多描述子提取

考虑到文献[4]中多描述子方法在 Caltech-101 数据库上的分类效果优于文献[7],本实验直接选择与前者进行比较.为了比较,特征编码均采用 LLC,每类图像的训练样本数设为 30,剩余为测试样本.表 1 为采样间隔设置为 6~10 像素时,不同描述子的分类准确率.

表 1 不同采样间隔下各描述子分类准确率

采样间隔	描述子			SIFT+描述子	
	SIFT	Edge-SIFT	KDES-S	Edge-SIFT	KDES-S
6	72.88	72.83	73.81	74.55	74.88
7	73.78	73.71	73.70	74.81	75.51
8	73.65	73.60	73.89	75.79	76.28
9	72.51	72.95	73.91	75.21	75.47
10	72.49	72.90	72.98	74.58	75.31

由表 1 的数据可以看出,基于多描述子的分类准确率都高于单一描述子,且本文的多描述子分类效果优于文献[4].在采样间隔为 8 时,多描述子分类准确率最高,因此后面实验采样间隔均取 8 像素.这一组实验中,相比文献[4],本文方法并没有明显优势,但是从描述子提取上考虑,不必单独生成原图

像的边界图再进行提取,减少计算消耗.

为了进一步比较分析,取数据库中分类结果差异较大的 5 个子类分别用 SIFT、Edge-SIFT、KDES-S 分类,结果如图 5 所示.

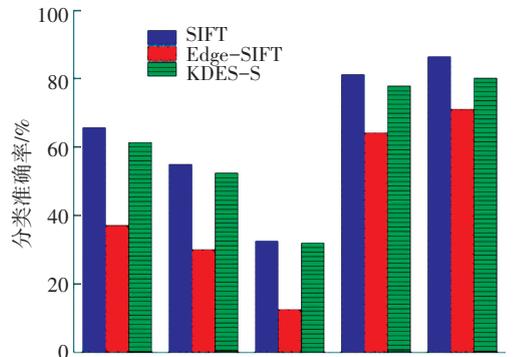


图 5 不同描述子在各子类的分类准确率

Fig.5 Classification accuracy of different descriptors on subcategories

从图 5 可以看出,5 个子类中,SIFT 与 Edge-SIFT 的分类准确率相差均较大,相比之下 KDES-S 则较为稳定.Edge-SIFT 在子类中分类准确率的过低会影响多描述子的分类效果,而 KDES-S 描述子则改善了这一问题,尤其在面对边界形变较大的图像,本文方法分类效果更好.

为进一步分析 SIFT 与 KDES-S 描述子结合对分类效果的影响,对上面 5 个子类图像,分别进行实验分析,得到表 2 所示的统计结果.

每类图像的数量在类名后的括号中给出,表中每一行分别表示满足条件的图像数量(如第一个数据 2 表示 Water lily 中,单独用 SIFT 分类出错,单独用 KDES-S 出错且 SIFT 与 KDES-S 结合分类正确的图片数量.第二行第二列数据 6 表示 Wild cat 中,单独用 SIFT 分类正确,用 KDES-S 分类错误,但用 SIFT 与 KDES-S 结合分类正确的图片数量.实验中没有 SIFT 或 KDES-S 单独分类正确但二者结合分类错误的图片,所以在表中没有列出这一情况).5 个子类中 SIFT 与 KDES-S 分别分类正确的图像说明本文采用的两种特征各具优势,具备互补性.而两种特征均分类错误但用二者结合后可以正确分类,这充分说明采用的结合方法可以利用两种特征进行互补,增强局部特征的表现力.

表 2 不同方法下的分类结果

Tab.2 Classification results on different methods

方法			类别				
SIFT	KDES-S	SIFT+KDES-S	Water lily (37)	Wild cat (34)	Crocodile (50)	Ferry (67)	Hedgehog (54)
错误	错误	正确	2	3	2	4	4
正确	错误	正确	4	6	2	5	3
错误	正确	正确	2	2	1	4	3
正确	正确	正确	18	24	15	53	42

为比较两种描述子结合方式, 分别利用 LLC 和分层特征学习结合两种特征结合方式在 Caltech-101 数据库上进行实验, 实验结果见表 3.

从分类结果可以发现, 先结合的特征联合分层特征学习获得了最佳的分类效果. 这是因为分层特征学习结构在编码过程中充分利用了结合后的特征, 得到更加丰富的图像特征, 这也表明采用的多描述子方法和分层特征学习方法能很好地相结合, 共同提高分类准确率.

表 3 不同多特征结构下 Caltech-101 的分类准确率

编码方法	先结合	后结合
LLC	76.28	76.48
分层特征学习	78.82	77.20

### 2.2 分层特征学习

该实验主要验证分层特征学习结构中层数和空间特征归一化对分类结果的影响. 实验分别在 Caltech-101 和 Scene-15 数据库上测试, 结果如图 6 所示. 实验构建的最大层数为 4, 其中第一、二层特征块大小设为  $16 \times 16$ , 第三、四层特征块大小分别为  $32 \times 32$ 、 $64 \times 64$ , 其他设置不变. Caltech-101 和 Scene-15 数据库中每一类训练图像分别设置为 30 和 100.

由图 6 可以看出, 随着层数的增加, 分类准确率有所提高, 层数由 1 增加到 2 时分类效果的提升最为明显, 这是对第一层经归一化处理后特征进行学习的结果. 之后随着字典层数的增加分类准确率的提升效果开始逐渐削弱, 为了降低计算消耗, 最终在其他实验中采用两层结构.

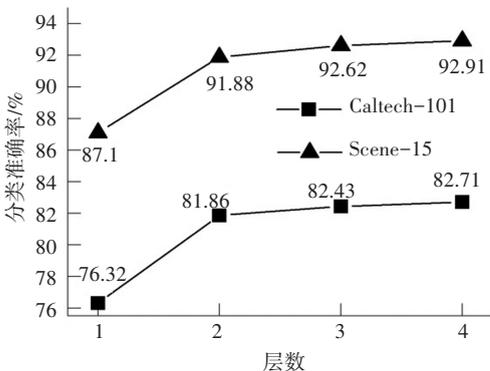


图 6 不同层特征学习分类结果

Fig.6 Classification results on different feature layer

比较以上两组实验, 同为 Caltech-101 数据库上两层的特征学习, 表 2 中分类结果 78.82% 和图 6 结果 81.86% 相差较大, 这是因为前者是用未分层的归一化方法得到的, 而后者是本文分配归一化系数

的方法得到的. 由此可得, 提出的方法在 Caltech-101 数据库上获得了 3.04% 的提升, 需要注意的是, 该方法结合分层结构, 只需对各层图像特征归一化处理, 对计算复杂度的增加基本可以忽略.

为进一步验证分层归一化系数  $k_i$  对分类结果的影响, 分别在两个数据库上对未分配归一化系数和分配系数的两种分层归一化结构比较, 结果见表 4.

表 4 不同归一化方法的分类准确率

数据集	未分配系数	分配系数
Caltech-101	79.92	81.86
Scene-15	89.33	91.88

观察实验结果可以发现, 分层归一化系数的引入在两个数据库上均有约 2% 的准确率提升, 这也验证了前面提到的分配系数可以进一步提高特征表现力, 同样表明了特征处理这一过程的必要性.

### 2.3 整体实验分析

前两部分实验已充分验证了本文方法在 Caltech-101 数据库上的有效性, 因此, 目标分类数据库改用更具挑战的 Caltech-256, 场景分类则选用应用较多的 Scene-15 数据库.

#### 2.3.1 Caltech-256

Caltech-256 数据库相当于 Caltech-101 数据库的扩展, 它包含 256 个目标类别和一个背景类别共计 30607 张图像. 由于图像类别和数目的增加, 实验中字典大小增加为 2048. 每一类训练图像分别随机取 15、30、45、60, 其他设置与上面实验一致, 结果见表 5.

表 5 Caltech-256 数据库中不同方法的分类准确率

方法	不同训练图像下平均分类准确率/%			
	15	30	45	60
GPP <sup>[4]</sup>	36.35	45.07	48.02	50.33
LLC <sup>[10]</sup>	34.36	41.19	45.31	47.68
IMFSC <sup>[7]</sup>	35.01	42.26	45.97	47.84
CRBM <sup>[21]</sup>	35.09	42.05	45.69	47.94
本文方法	39.61	49.23	53.21	56.27

由表 5 的数据可以看出, 采用的方法在图像类别和数量增加、目标更为复杂的情况下, 仍能得到理想的分类结果. 对比最新的文献[4, 7] 基于多描述子的方法, 本文方法分类效果更佳, 这不仅与所选特征有关, 更多的是分层学习结构和对特征进行空间归一化的影响. 相比文献[21] 基于深度学习方法, 本文算法具有较大提升, 这表明提出的多描述子与分层特征学习能很好地结合共同提高特征表现力. 由表 5 数据还可以发现, 随着训练图像的增加, 本算法优势更加明显, 这表明提取的特征更为丰富有效,

随着训练样本增加更能体现出优势.

### 2.3.2 Scene-15

Scene-15 数据库包含 15 个室外场景类别共

4485 张图像,是一个被广泛使用的场景识别数据库,图 7 是该数据集部分示例图片. 实验中训练图像设为 100,分类结果见表 6.

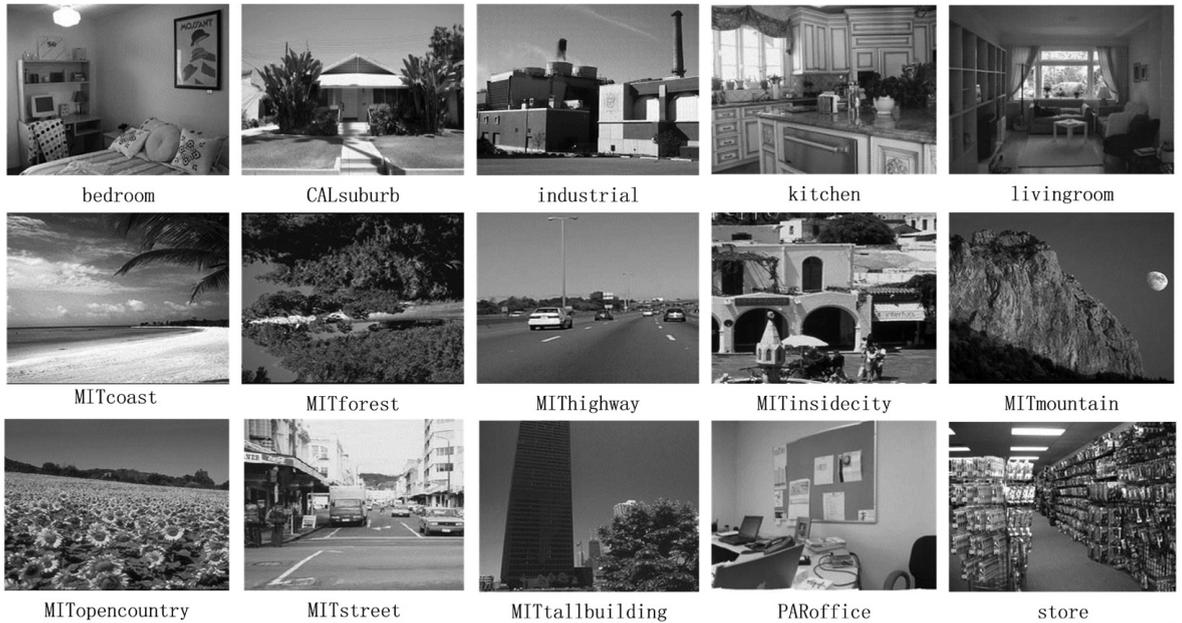


图 7 Scene-15 部分图片

Fig.7 Some images of Scene-15 datasets

表 6 Scene-15 数据库中不同方法的分类准确率

Tab.6 Classification accuracy of different methods on Scene-15 %

方法	分类准确率	方法	分类准确率
GPP <sup>[4]</sup>	85.13	KDES-A <sup>[17]</sup>	86.70
LLC <sup>[10]</sup>	82.34	LCKSVD <sup>[13]</sup>	90.40
SPM <sup>[8]</sup>	81.40	本文方法	91.88
IMFSC <sup>[7]</sup>	83.12		

由表 6 数据可以看出,算法在场景分类任务中也具有较好的分类效果. 相比最近的文献[4,7]中的方法,本文方法分类准确率要高 6%~8%,相比文献[8,10,17]方法提升更为明显. 结合 Caltech-256 数据库的实验分析,一个原因是训练样本的增多,本算法提取的特征更能体现其有效性,另一个原因本文是分层特征学习与空间特征归一化相结合得到的图像特征更加丰富稳定. 与文献[13]方法相比提升不是特别明显,但是本文带系数的分层特征归一化与分层特征学习结构相结合的方法更为简单,且所用归一化方法计算量小,耗时较少.

## 3 结 论

结合深度学习思想和多特征提取方法,提出一种基于多描述子分层特征学习的图像分类方法. 在分别提取图像的局部特征并将特征结合后采用基于字典学习和批正交匹配追踪的方法对特征进行编码. 这一过程中,构建一个两层的结构,分别对各层

特征进行学习和归一化处理,最后将两层的特征连接起来作为图像的最终表示并用线性 SVM 分类器进行分类. 通过实验验证,该方法在目标分类和场景分类任务中均具有较好的性能,且对于多类别多数量的数据集,仍具有较好的鲁棒性. 当训练样本较少时该方法的分类结果不够理想,这是需要进一步研究和改进的地方.

## 参 考 文 献

- [1] CSURKA G, DANCE C, FAN Lixin, et al. Visual categorization with bags of keypoints[C]// Workshop on Statistical Learning in Computer Vision in ECCV. Berlin: Springer, 2004: 1-22.
- [2] ZHANG Shiliang, QI Tian, HUA Gang, et al. Generating descriptive visual words and visual phrases for large-scale image applications[J]. IEEE Transactions on Image Processing, 2011, 20(9): 2664-2677.
- [3] LIU Lingqiao, WANG Lei, LIU Xinwang. In defense of soft-assignment coding[C]// 2011 International Conference on Computer Vision. Barcelona: IEEE, 2011: 2468-2493.
- [4] XIE Lingxi, QI Tian, WANG Meng, et al. Spatial pooling of heterogeneous features for image classification[J]. IEEE Transaction on Image Processing, 2014, 23(5): 1994-2008.
- [5] YUAN Junsong, WU Ying, YANG Ming. Discovery of collocation patterns: From visual words to visual phrases[C]// IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007: 1-8.
- [6] 许允喜, 陈方. 局部图像描述符最新研究进展[J]. 中国图象图形学报, 2015, 20(9): 1133-1150. DOI:10.11834/jig.20150901. XU Yunxi, CHEN Fang. Recent advances in local image descriptor[J]. Journal of Image and Graphics, 2015, 20(9): 1133-1150.

DOI:10.11834/jig.20150901.

- [7] 罗会兰, 郭敏杰, 孔繁胜. 集成多特征与稀疏编码的图像分类方法[J]. 模式识别与人工智能, 2014, 27(4): 345-355.  
LUO Huilan, GUO Minjie, KONG Fansheng. Image Classification Method by Combining Multi-features and Sparse Coding[J]. Pattern Recognition & Artificial Intelligence, 2014, 27(4): 345-355.
- [8] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features; spatial pyramid matching for recognizing natural scene categories [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ; IEEE, 2006: 2169-2178.
- [9] YANG Jianchao, YU Kai, GONG Yihong, et al. Linear spatial pyramid matching using sparse coding for image classification [C]// IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL; IEEE, 2009: 1794-1801.
- [10] WANG Jinjun, YANG Jianchao, YU Kai, et al. Locality-constrained linear coding for image classification [C]// IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA; IEEE, 2010: 3360-3367.
- [11] 段喜萍, 刘家锋, 王建华, 等. 一种语义级文本协同图像识别方法[J]. 哈尔滨工业大学学报, 2014, 46(3): 49-53.  
DUAN Xiping, LIU Jiafeng, WANG Jianhua, et al. A collaborative image recognition method based on semantic level of text[J]. Journal of Harbin Institute of Technology, 2014, 46(3): 49-53.
- [12] AHARON M, ELAD M, BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.
- [13] JIANG Zhuolin, LIN Zhe, DAVIS L S. Label Consistent K-SVD: Learning a discriminative dictionary for recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2651-2664.
- [14] 王博, 郭继昌, 张艳. 基于深度网络的可学习感受野算法在图像分类中的应用[J]. 控制理论与应用, 2015, 32(8): 1114-1119.  
WANG Bo, GUO Jichang, ZHANG Yan. Learnable receptive fields scheme in deep networks for image categorization[J]. Control Theory & Application, 2015, 32(8): 1114-1119.
- [15] BO Liefeng, REN Xiaofeng, FOX D. Multipath sparse coding using hierarchical matching pursuit [C]// IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR; IEEE, 2013: 660-667.
- [16] RUBINSTEIN R, ZIBULEVSKY M, ELAN M. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit[J]. Cs Technion, 2008, 40(8): 1-15.
- [17] BO Liefeng, REN Xiaofeng, FOX D. Kernel descriptors for visual recognition[J]. Advances in Neural Information Processing Systems, 2010: 244-252.
- [18] XIE Lingxi, QI Tian, ZHANG Bo. Simple techniques make sense: feature pooling and normalization for image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(7): 1251-1264.
- [19] FENG Jiashi, NI Bingbing, QI Tian, et al. Geometric  $\ell_p$ -norm feature pooling for image classification [C]// IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI; IEEE, 2011: 2697-2704.
- [20] FAN Rongen, CHANG Kaiwen, HSIEH C J, et al. LIBLINEAR: a library for large linear classification[J]. Journal of Machine Learning Research, 2008, 9(12): 1871-1874.
- [21] SOHN K, JUNG D Y, LEE H, et al. Efficient learning of sparse, distributed, convolutional feature representations for object recognition [C]// IEEE International Conference on Computer Vision. Barcelona; IEEE, 2011: 2643-2650.

(编辑 王小唯 苗秀芝)