

一种用于图像分类的语义增强线性编码方法

肖文华* 包卫东 陈立栋 王 炜 张茂军

(国防科技大学信息系统与管理学院 长沙 410073)

摘要: 针对传统编码模型中存在的编码歧义性问题, 该文提出一种考虑特征上下文的语义增强线性编码方法。首先, 通过学习局部邻域中特征共生关系矩阵来表示上下文信息。然后, 在编码过程中同时引入学习而得的上下文信息与特征上下文匹配权重得到语义增强编码模型。由于上下文信息与上下文匹配权重的功能, 使得此编码方法不仅丰富了编码的语义信息, 还能够有效避免噪声带来的影响。在3个基准数据集(Scene15, Caltech101以及 Caltech256)上充分的实验验证了该方法的有效性。

关键词: 图像分类; 特征编码; 上下文约束; 歧义性

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2015)04-0791-07

DOI: 10.11999/JEIT140743

A Semantic Enhanced Linear Coding for Image Classification

Xiao Wen-hua Bao Wei-dong Chen Li-dong Wang Wei Zhang Mao-jun

(Information System and Management College, National University of Defense Technology, Changsha 410073, China)

Abstract: Considering the ambiguity problem in the traditional feature coding model, a feature context-aware semantic enhanced linear coding method is proposed. At first, the context information is represented by the concurrence matrix learnt from local area of the features. Then, the context information and a context matching weight are introduced into the coding model to form a new semantic enhanced coding model. Owing to the functions of context information and the context matching weight, this model not only enriches the semantic meaning of coding, but also efficiently avoids the affects of noise. Experiments on the baselines (Scene15, Caltech101, and Caltech256) demonstrate the effectiveness of the proposed method.

Key words: Image classification; Feature coding; Context constraint; Ambiguity

1 引言

由于在机器人^[1], 卫星与遥感^[2,3], 医学^[4]等领域应用前景广泛, 自动图像分类技术已受到全世界学者们持续增长的关注。在最近的图像分类研究工作中, 特征袋^[5] (Bag of Feature, BoF)是最流行也是最有效的模型。BoF 主要包含以下通用步骤: 特征提取, 字典学习, 特征编码, 池处理以及分类。在以上步骤中, 特征编码联系了特征提取与池处理, 是方法的核心部分, 对图像分类的精度以及速度都有很大影响^[6]。由于特征编码的关键作用, 近年来对其编码策略的研究已受到越来越广泛的关注。

特征编码可以被认为是选择少量字典基并赋予相应的权重来表示特征的过程, 已有不少文章提出

了不同的编码方法。文献[5,7]采用向量量化(Vector Quantization, VQ)方法, 仅将特征编码到离其最近的基上, 简单有效但是会产生较大的量化误差。为了克服此缺点, 文献[8]提出一种更鲁棒的软向量量化 (Soft Vector Quantization, SVQ)方法, 将特征编码到所有字典基上, 并以距离的高斯核函数作为分配权重, 以减小量化误差。其更进一步的改进是局部软向量量化^[9](Localized Soft Vector Quantization, LSVQ), 不同于SVQ将权重分布在所有基上, 此方法将权重的分布限定在局部邻域。为了减小VQ方法存在的量化误差, 文献[10]提出了稀疏编码 (Sparse Coding, SC) 方法, 即通过稀疏约束将少量的字典基分配给特征, 但其缺点是计算复杂度高并且对相似特征编码时容易产生编码不一致问题。此后, 对SC的改进不断涌现。文献[11]提出了一种局部约束线性编码 (Locality-constrained Linear Coding, LLC)方法, 在保证稀疏性的同时克服SC方法复杂度高问题。文献[12]提出了一种拉普拉

2014-06-05 收到, 2014-08-29 改回

国家自然科学基金(61175006)和博士学科点专项科研基金(20134307110029)资助课题

*通信作者: 肖文华 wenhuaxiao@nudt.edu.cn

斯稀疏编码(Laplacian Sparse Coding, LSC)方法,解决了相似特征编码的结果不一致问题。然而,由于其中特征关系矩阵维数非常高,导致计算复杂度非常高。文献[13]提出了空间金字塔鲁棒稀疏编码(Spatial Pyramid Robust Sparse Coding, SP-RSC)方法,克服了编码时空间关系丢失问题,而且还提出了拉普拉斯仿射稀疏编码(Laplacian Affine Sparse Coding, LASC)方法^[14],将倾斜度与方向因子引入目标函数中,解决了成像时造成的特征判别性不强问题。

虽然,学者们对编码模型提出了诸多改进,但编码中存在的歧义性问题仍未得到解决。针对此问题,本文对 LLC 编码方法进行改进,通过将上下文约束引入编码过程中,提出一种语义增强线性编码(Semantic enhanced Linear Coding, SLC)方法。本文的主要贡献在于:(1)与传统方法通过上下文思想学习高层特征不同,本文将上下文思想引入特征编码中,并以共生矩阵对其进行描述,旨在解决编码歧义问题。(2)通过在 LLC 模型上引入上下文约束和权重匹配,提出了语义增强的线性编码模型。

2 编码歧义问题

编码歧义问题是由相似特征表示不同图像而产生的。如图 1 所示, A_1 与 B_1 分别代表两幅图像中的正方形角点,具有相似特征 F ,但代表不同语义(表示不同图像)。显然, A_1 与 B_1 应该采用不同的基进行编码以更好地区分这两幅图像。而当利用前述方法(如 VQ, SVQ, SC 以及 LLC)编码时会得到相似编码结果,达不到区分 A_1 与 B_1 的目的,本文称此问题为编码歧义问题。

3 本文方法

上下文思想是一种通过考虑特征邻域增强判别性的思想,本文亦采用此思想解决以上问题。如图 1, 当对 A_1 与 B_1 进行编码时考虑其邻近特征 A_2 与

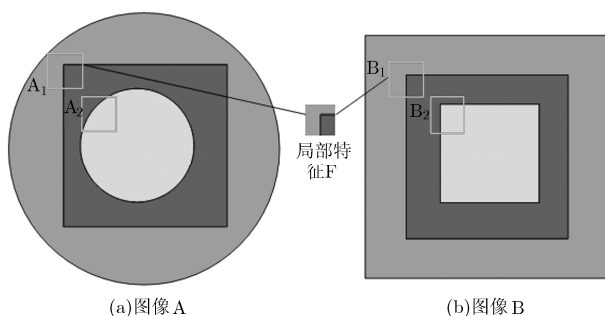


图 1 编码歧义示例

B_2 , 判别性显然增强了,从而有利于区分 A_1 与 B_1 。上下文思想已在机器视觉领域(如图像分类^[15]和动作识别^[16])广泛应用,本文与其不同之处在于:传统上下文思想主要利用上下文思想提取高层语义特征,而本文则将其引入特征编码过程中。另外,鉴于 LLC 编码的高效性及其二范数约束项与上下文约束有着良好的契合性,本文选择在此模型基础上进行改进。基于以上思想,本文所提方法主要由两个部分组成:描述上下文信息的共生矩阵构建以及语义增强线性编码模型的建立,下面对这两个方面作详细介绍。

3.1 共生矩阵的构建

本文通过构建特征间共生关系矩阵描述上下文信息。方法基于如此观测现象:对于一类特定场景或图像来说,在其局部临域,各图像常常表现出相似的共生模式。例如,在卧室场景中,床、衣柜和枕头往往呈现出共生关系。就特征层来讲,作者认为这种模式能由图像中局部临域中特征之间的共生关系反映。本节,提出一种新颖而简单的方法来描述这种关系,对于单副图像共生关系矩阵的构建流程如图 2 所示,其详细步骤所述为:

给定一训练集,首先利用 K 均值算法^[17](K-Means)构建大小为 D 的字典。对于一特定图像类,记 $\mathbf{X} = \{x_i \in R^M, i \in 1, 2, \dots, N_i\}$ 为从此类选取的训练特征集,其中 x_i 为特征描述子, M 为特征维数, N_i 为特征个数。然后,基于字典 D ,通过 K 近邻(K-Nearest Neighbor, K-NN)算法对这些特征进行标签。记 $f_i = \{x_i, l_i, p_i\}$ 为第 i 个特征,其中, l_i ($1 \leq l_i \leq D$) 对应于字典中基的索引号, $p_i = \{x_i, y_i\}$ 记录了此特征中心所处的像素位置。因此,所有的特征可以被分成 D 个集合,记此集合为 $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_D\}$,其中 $\mathbf{S}_i = \{f_1^i, f_2^i, \dots, f_{N_i}^i\}$ 包含了所有标签为 i 的特征, N_i 为 \mathbf{S}_i 中特征个数。为了获取局部特征间关系,定义特征 f_i 的上下文域为

$$C_i = \{f_j \mid \{x_j, y_j\} \in \Omega(p_i)\} \quad (1)$$

其中, $\Omega(p_i)$ 表示特征 f_i 的局部临域,本文以中心为 p_i ,半径为 r 的圆表示(如图 2 中的圆)。显然, f_i 的上下文域 C_i 包含了其局部临域 $\Omega(p_i)$ 中的所有特征。因而,对于集合 \mathbf{S}_i 中的第 j 个特征 S_{ij} ,在其上下文域中,可以统计得到一 D 维的向量 $\mathbf{h}_{ij} = [v_1^{ij}, v_2^{ij}, \dots, v_D^{ij}]$,其中 v_l^{ij} ($l = 1, 2, \dots, D$) 为 S_{ij} 上下文中标签为 l 的特征个数。对 \mathbf{S}_i 中所有特征得到的向量进行累加后,即可获得第 i 个字典基的临域分布直方图 \mathbf{h}_i

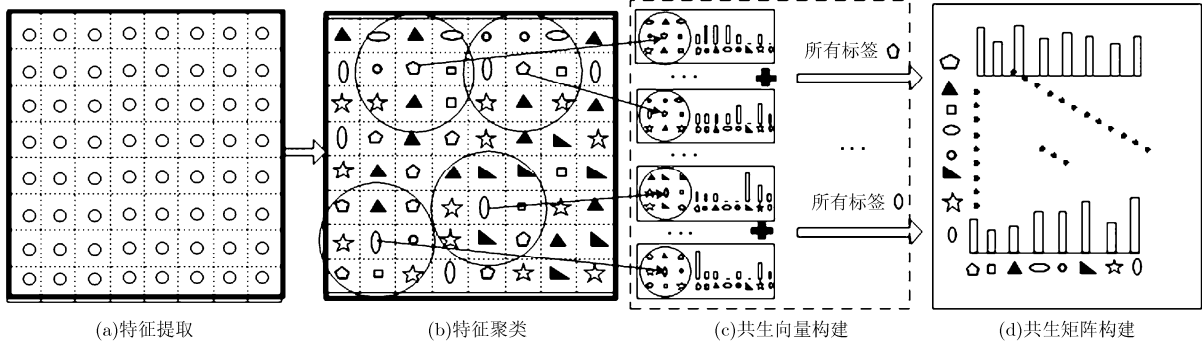


图 2 共生矩阵构造流程

$$h_i = \sum_{j=1}^{N_i} h_{ij} \quad (2)$$

式(2)中, 记 $h_i = [v_1^i, v_2^i, \dots, v_D^i]$, 且 v_j^i 描述了第 i 个字典基与第 j 个字典基之间的共生强度。本文将其归一化为 $[0,1]$, 因而 v_j^i 可反映 i 特征与 j 类特征在局部邻域的共现概率, 即代表了特征间的上下文关系。当对所有的 $S_i, i = 1, 2, \dots, D$ 重复以上步骤, 即可获得所有字典基的关系矩阵, 记为 $H = [h_1, h_2, \dots, h_D]$ 。为区分与测试图像构造的矩阵, 将此由训练图像产生的共生矩阵为模板矩阵, 而测试图像产生的共生矩阵为测试矩阵。如图 3(a)与图 3(b)所示, 以深度图表示了从 Caltech101 数据集^[18]中两个不同类“Background_google”与“accordion”构建的不同关系矩阵, 其灰度分布的显著区别说明了“Background_google”与“accordion”局部上下文的差异性。作者认为这就是上下文信息能解决编码歧义问题的原因所在。对所有类重复以上步骤后, 就可以构建所有类的关系矩阵, 记之为 $\{H^1, H^2, \dots, H^{nc}\}$, nc 为图像类别数。

3.2 语义增强线性编码

通过共生矩阵获得上下文信息之后, 将其集成到 LLC^[11]编码模型中并对其进行改进。记 $X = [x_1, x_2, \dots, x_N] \in \mathbf{R}^{M \times N}$ 为一从图像中提取的 M 维特征集合。给定一维度为 D 的字典 $B = [b_1, b_2, \dots, b_D] \in \mathbf{R}^{M \times D}$,

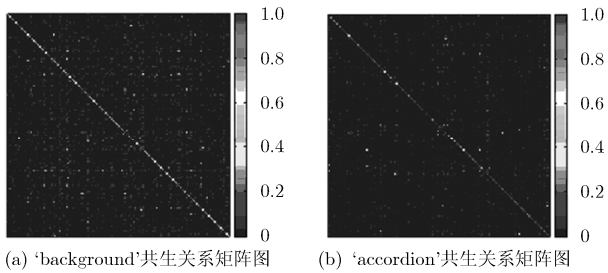


图 3 以深度图表示的关系矩阵

(其中 $b_i (i = 1, 2, \dots, D)$ 为第 i 个字典基)以及所有类的关系矩阵 $[H^1, H^2, \dots, H^{nc}]$, 则可根据模板矩阵 $H^p, (p = 1, 2, \dots, nc)$ 通过建立如式(3)所示的目标函数从而将上下文信息引入到编码过程中

$$c_i = \arg \min_c \left\{ \|wx_i - Bc\|_2^2 + \lambda \left\| \frac{\alpha \cdot \mathbf{d}f_i + 1}{\beta \cdot \mathbf{d}c_i + 1} \odot c \right\|_2^2 \right\} \quad (3)$$

s.t. $\mathbf{1}^T c / w = 1$

其中, $c_i \in \mathbf{R}^D$ 为 x_i 的编码结果, $\mathbf{d}f_i$ 表示 x_i 与 B 在特征空间的距离, 这与 LLC^[11]表示相同。不同之处在于 $\mathbf{d}c_i$ 和 w 的引入。 $\mathbf{d}c_i$ 表示 x_i 与 B 的连接权重。详细地,

$$\mathbf{d}c_i = \exp\left(\frac{\text{conn}(x_i, B)}{\sigma}\right) \quad (4)$$

其中, $\text{conn}(x_i, B) = [\text{conn}(x_i, b_1), \text{conn}(x_i, b_2), \dots, \text{conn}(x_i, b_D)]^T$, 且 $\text{conn}(x_i, b_j)$ 表示 x_i 与 b_j 之间的连接权重。 σ 为调整上下文空间局部权重因子下降速度的参数。 λ 为控制特征空间与上下文空间约束程度的参数。 α 代表了特征空间的局部约束权重而 β 代表了上下文空间的局部约束权重。实际上, α 与 β 可以通过参数 λ 进行调节。本文引入这两个参数的原因主要是为了在实验阶段研究 $\mathbf{d}f_i$ 与 $\mathbf{d}c_i$ 对模型的性能的影响。假若 x_i 的标签为 l , 则 $\text{conn}(x_i, b_j)$ 可以近似地计算如下:

$$\text{conn}(x_i, b_j) = H_{ij}^p \quad (5)$$

显然, H_{ij}^p 的值越大, x_i 与 b_j 的关系在 p 类中就越近, x_i 与 b_j 在上下文空间的距离也越近(由于 $\mathbf{d}c_i$ 代表了 x_i 与 b_j 在上下文空间的反距离); 反之亦然。从式(3)可知, x_i 与 B 在特征空间与上下文空间的距离同时决定了字典基的响应。因此, 具有不同上下文的相似特征可以通过此编码方法得到具有判别性的结果。

为降低噪声对模型性能的影响,式(3)中引入因子 w , 其测量了被编码特征 x_i 的上下文共生模式与模板矩阵 H^p 中相应共生模式的匹配程度。其详细计算过程如表 1 所示。即首先获得 x_i 上下文域 C_i , 然后对 C_i 所有特征从模板矩阵 H^p 获得与 x_i 的共生强度值, 并将获得的共生强度平均值作为当前特征 x_i 的上下文与第 p 个图像类的上下文拟合程度, 记为 w_i^p 。对每一 H^p ($p=1,2,\dots,nc$), 则可获得 $w_i^1, w_i^2, \dots, w_i^{nc}$ 。最后, 为将 w_i^p 归一化至 $[0, 1]$, 设 $w_i^p = \exp(2(w_i^p - \max_w))$, 其中 \max_w 代表 w_i^p ($p=1,2,\dots,nc$) 的最大值, n 为 x_i 上下文域特征个数。显然, 若 x_i 是从第 p 类的图像中提取的特征, 由于其局部上下文与第 p 类的上下文相似, 因而 w_i^p 的值会较大。否则, 由于其上下文差距较大, w_i^p 的值就会较小。另外, 从式(3)的解析解(见式(6)), 可以得出如下结论: w_i^p 的值越大, 编码系数的值也越大, 反之亦然。由于噪声特征常常不与任何模板的上下文匹配, 导致噪声特征编码系数会非常小。因而, 当利用 Max Pooling^[10]来获取最终图像表示时, 噪声则可被剔除。

对每一个 $H^p, p=1,2,\dots,nc$, 通过以上方法对特征 x_i 进行编码, 得到编码系数 $[c_i^1, c_i^2, \dots, c_i^{nc}]$ 。其中 c_i^p

代表对应于第 p 类的关系矩阵 H^p 的编码系数。给定一个拥有 N 个描述子 $X = \{x_i \in R^M, i=1, 2, \dots, N\}$ 的图像, 对应于关系矩阵 $\{H^1, H^2, \dots, H^{nc}\}$ 的编码系数为 $c = [c_1^1, c_1^2, \dots, c_1^{nc}, c_2^1, c_2^2, \dots, c_2^{nc}, \dots, c_N^1, c_N^2, \dots, c_N^{nc}]$ 。由于 Max Pooling 被广泛应用于模式识别任务^[10-13]中, 并且被证明与大脑视觉皮层的特性一致, 本文在系数矩阵 c 上进行 Max Pooling 操作, 得到图像的最终表示。

4 模型分析

此节对模型性能进行了分析并阐述了其优点。

4.1 模型进化

当控制参数 w, α 以及 β 时, 本文编码模式会退化成两种特殊的情况。(1)当 $w=1$ 以及 $\alpha=1, \beta=0$ 时, 模型只考虑了在特征空间的局部约束性, 因而 SLC 退化成 LLC 模型。(2)当 $w=1$ 以及 $\alpha=0, \beta=1$ 时, SLC 退化为只考虑上下文空间局部约束的编码模型。

4.2 模型优点

与传统的 VQ, SVQ, SC, LSC 以及 LLC 等编码方式相比, SLC 编码模式有以下优点。

(1)避免了编码的歧义性。SLC 同时在特征空间和上下文空间进行局部约束编码。在特征空间的局部性保证了重构误差的精度, 而在上下文空间的局部性确保了编码的语义性。因此, 通过将上下文信息引入到编码模型中, 本文在解决编码的歧义性问题同时又保证了重构精度。

(2)有去噪能力。对此, 分别从 Caltech101 选取了类别 Wild_cat, Panda, pagoda 以及从 Caltech256 选取了类别 American-flag, Sunflower-101 进行实验。实验时每隔 8 个像素进行稠密采样提取局部特征, 并通过表 1 中算法 1 计算每个特征对应 w 值。结果如图 4 所示, 圆圈代表在条件 $w > 0.4$ 下保存的特征。通常, 每幅图片中都有许多特征(比如从图 4(a)美国

表 1 因子 w 计算过程

算法 1	因子 w 计算过程
输入:	x_i, l, H^p (l 为特征 x_i 的标签)
输出:	w_i^p
(1)	初始化: $w_i^p = 0$
(2)	通过式(1)计算 x_i 上下文域 C_i
(3)	For each f_j in C_i
(4)	获得特征 f_j 的标签 l_j
(5)	$w_i^p = w_i^p + H^p(l, l_j)$
(6)	End For
(7)	$w_i^p = w_i^p / C_i $ ($ C_i $ 代表上下文域特征个数)

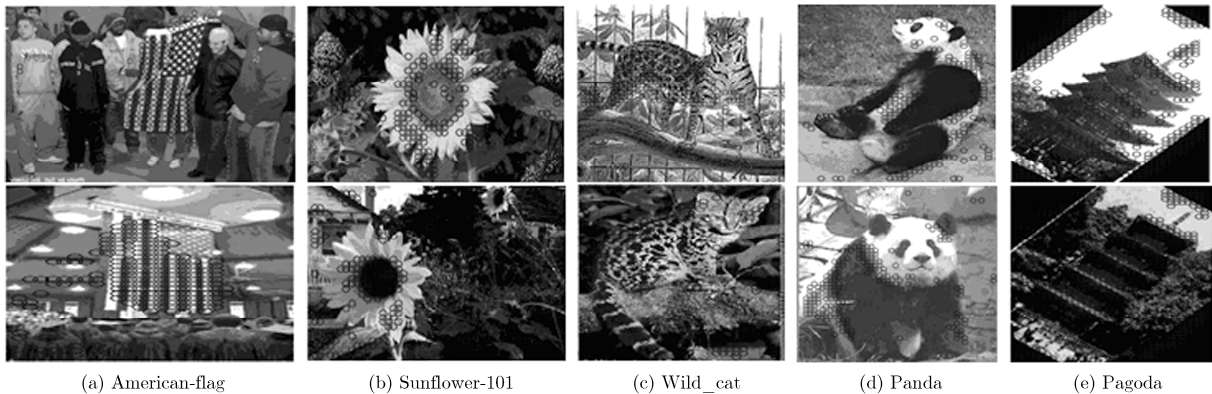


图 4 w 的去噪功能实验结果

国旗复杂背景中提取的特征)不仅对描述图像无用而且还影响了图像描述的准确性,本文称这些特征为噪声。由于这些噪声的上下文不与任何模板上下文匹配(w 值很小),当利用 SLC 编码时,会导致其编码系数非常小。进一步,当采用 Max Pooling 策略获取最终图像表示时,将导致这些噪声不会对最终的图像表示有任何影响。如图 4 所示,大部分描述背景的特征去除了而每类描述目标本身的特征保留了。

(3) 快速系数求解。由于其目标函数的凸性,SLC 继承了 LLC 的独特优点:目标函数具有解析解。通过推导可得其解析解为

$$c_i = w(\mathbf{1}^T \Psi \mathbf{1})^{-1} (\Psi)^{-1} \mathbf{1} \quad (6)$$

其中

$$\Psi = 2(Q + \lambda \text{diag}(\mathbf{d}_i)^2), Q = (x_i \mathbf{1}^T - B)^T (x_i \mathbf{1}^T - B)$$

其他变量意义与上文相同,且

$$d_i = \frac{\alpha \cdot \mathbf{d} \mathbf{f}_i + 1}{\beta \cdot \mathbf{d} \mathbf{c}_i + 1} \quad (7)$$

此外,与 LSC 对所有特征同时进行编码不同,SLC 对特征独立地进行编码且能保持他们的相互关系。这使得在分布式计算模型 Mapreduce^[19]下,即使处理海量数据,都能达到实时的速度。因而,对实际应用有非常重要的意义。

5 实验及结果分析

本节在 3 个广泛采用的图像数据库 Scene15^[7], Catech101^[19] and Caltech256^[20]上进行实验以验证所提方法的有效性。首先,通过比较 LLC 与 SLC 来验证考虑上下文信息的有效性。然后,将 SLC 在 3 个数据库上的测试结果与当前最好结果进行了对比。

5.1 实验设置

为实验对比时公平起见,在所有进行的实验中,均采用以下相同的实验设置。对于所有数据集,首先缩放图像使其宽与高不超过 300 像素。在所有数

据集上采用单层的 16×16 块,间隔为 8 像素方式提取稠密尺度不变特征转换(Scale-Invariant Feature Transform, SIFT)特征^[21]。通过随机选择约 10^5 特征并利用 K-Means^[17]训练得到字典。在数据集 Caltech101, Caltech256, Scene15 上分别在每类中随机选择 30, 30, 100 幅图像作为训练数据,其他剩余的作为测试数据。为了获得最终的图像表示,采用 3 层的空间金字塔^[7]以及 Max Pooling 方法^[10]。在分类阶段采用 Linear SVM package^[22],所有的实验都进行了 10 次,并以平均值作为最终的实验结果。

5.2 上下文信息的有效性

为了验证本文考虑上下文因子的有效性,本文基于 LLC^[11]作者提供的代码对其重新实现并与原有算法进行对比。此对比实验在 Caltech101 上进行,实验中,通过多次试验分析选择最佳参数组合($\alpha = 0.6, \beta = 0.4, r = 17$)。实验将 LLC 方法误分率最大的 10 类与本文 SLC 算法进行了详细对比。如图 5, SLC 相比 LLC 在大多数图像类上都显著(在“鸭嘴兽”上最大改进了 12%)提高了分类准确率,甚至在易混淆的类如“龙虾”,“蟹”以及“小龙虾”亦是如此。作者认为此改进主要是由于上下文信息的考虑。因为,易混淆的类在细节上可能很相似,对于 LLC 编码来说,编码歧义性问题未能解决,因而会产生较大的误分率;然而,对 SLC 编码方法来说,由于其考虑上下文解决了编码歧义性问题,使易混淆的类更易区分。

5.3 与其他算法对比

本节,将本文方法与一些已发布的方法在 3 个数据集上进行对比。比较主要聚焦于以下两种编码策略:LSVQ 与 LLC,因为这两种编码方法是当前综合性能最好方法的代表。需要提及的是文献[9,11]所获得的结果是在与本文不同设置下产生的。比如,

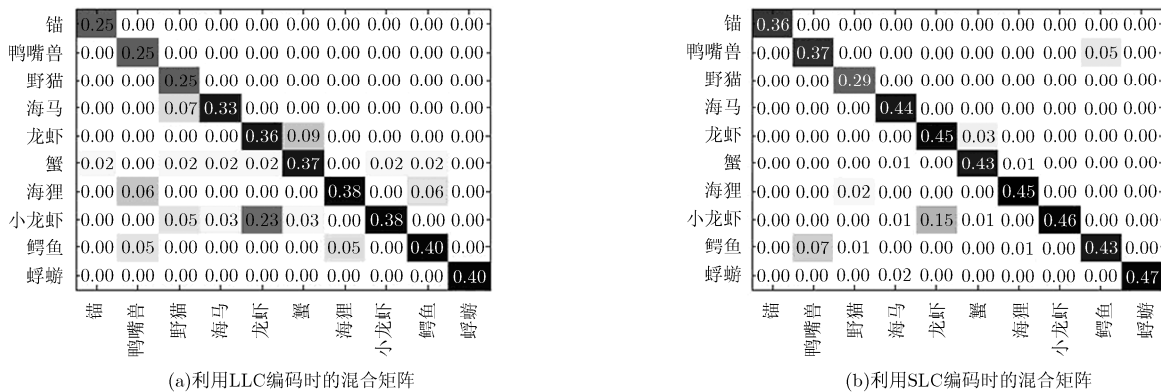


图 5 在最易错分 10 类的详细对比

表 2 与已发布的编码方法对比(%)

方法	Cal.101(#30)	Cal.256(#30)	Cal.256(#60)	Scene15(#100)
SC ^[10]	73.20+0.54	34.02+0.35	40.14+0.91	80.28+0.93
LSC ^[12]	NA	35.74+0.10	40.32+0.32	89.78+0.40
LLC ^[11]	73.44+NA	41.19+NA	47.68+NA	NA
LSVQ ^[9]	74.21+0.81	NA	NA	82.70+0.39
SP-RSC ^[13]	NA	41.87+0.50	NA	83.67+0.49
LASC ^[14]	NA	44.20+0.72	49.95+0.52	90.36+0.63
LLC(本文)	72.32+0.91	40.32+0.26	46.56+0.78	81.73+0.75
LSVQ(本文)	72.58+0.79	38.51+0.42	43.10+0.11	83.08+0.56
SLC(本文)	74.90+0.44	42.37+0.38	49.45+0.67	81.89+0.54

注: #30 代表训练样本的个数为 30。NA 表示相应文章中无此项数据

LLC^[11]方法中每个 8 像素提取多尺度特征, LSVQ 方法^[9]中采用 mix-order Max-Pooling, 并且字典大小也不尽相同。为公平比较, 作者重写了这两种方法并按照本文的设置进行实验。同时, 也与这些方法的其他实现方式进行了对比以更充分地评估本文方法。在此试验中, 在所有数据集上采用相同的参数设置 $\alpha = 0.6, \beta = 0.4, r = 17$ 。根据 LLC 的设置方法, 实验利用 K-Means 方法分别在 Scene15, Caltech101 和 Caltech256 训练大小为 1024, 2048 和 4096 的字典。如表 2 所示, 本文 SLC 算法在 3 种不同数据库不同设置下都优于大部分已发布的方法, 部分设置只比文献[14]稍差。但是, 由于文献[14]包含复杂的拉普拉斯矩阵运算, 本文方法相比文献[14]在计算速度上有较大优势。同时, 注意到本文的 LLC 版本结果比文献[11]中的 LLC 结果稍差。这可能是由于本文采用单尺度的特征而文献[11]中采用多尺度特征的原故。同样值得注意的是本文实现的 LSVQ 在 Scene15 上获得了比原始 LSVQ^[9]更好的效果, 这可能归功于本文所采用字典大于原始文献[9]所采用的字典。另外, 本文方法在数据集 Caltech101 与 Caltech256 上优于所有列出的大部分方法, 而在数据集 Scene15 并不是表现最好的。就此方面而言, 本文方法更适合于目标识别而非场景图像分类。作者认为可能是由于这样一个事实: 在局部区域中对象图像通常比场景图像具有更相似的上下文, 因而在对象图像中考虑上下文信息比场景图像中考虑上下文更有效。尽管如此, 在数据集 Scene15 上, 本文 SLC 方法计算更快, 并比 LLC 方法有所提高而且达到了与大部分列出的方法相近的结果。

6 结束语

为缓解编码的歧义性问题, 本文通过考虑上下文信息提出了一种新颖的改进 LLC 的 SLC 方法。

由于考虑了上下文信息以及特征匹配权重 w , 该方法不仅能帮助解决编码歧义性问题, 而且能够克服噪声特征所产生的数据变形问题, 从而能提高分类准确率。在 Caltech101 与 Caltech256 数据集上的实验结果证明了本文方法的有效性。同时, 在 Scene15 数据集上实验结果表明, 相比场景图像分类来说, 此方法更适用于目标图像分类。

参考文献

- [1] Espinace P, Kollar T, Roy N, *et al.* Indoor scene recognition by a mobile robot through adaptive object detection[J]. *Robotics and Autonomous Systems*, 2013, 61(9): 932-947.
- [2] 蒋飞云, 孙锐, 张旭东, 等. 基于二次表示的空间目标图像分类[J]. 电子与信息学报, 2013, 35(5): 1247-1251.
Jiang Fei-yun, Sun Rui, Zhang Xu-dong, *et al.* Space target image categorization based on the second representation[J]. *Journal of Electronics & Information Technology*, 2013, 35(5): 1247-1251.
- [3] Zhang Y, Zheng X W, Liu G, *et al.* Semi-supervised manifold learning based multigraph fusion for high-resolution remote sensing image classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(2): 464-468.
- [4] Zhang Y, Zhang B, Coenen F, *et al.* One-class kernel subspace ensemble for medical image classification[J]. *EURASIP Journal on Advances in Signal Processing*, 2014, (17): 1-13.
- [5] Csurka G, Dance C, Fan L, *et al.* Visual categorization with bags of keypoints[C]. Proceedings of the European Conference on Computer Vision on Statistical Learning in Computer Vision, Prague, Czech Republic, 2004: 59-74.
- [6] Huang Y, Wu Z, Wang L, *et al.* Feature coding in image classification: a comprehensive study[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(3): 493-506.
- [7] Lazebnik S, Schmid C, and Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene

- categories[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 2169–2178.
- [8] Gemert J C, Geusebroek J, Veenman C J, *et al.* Kernel codebooks for scene categorization[C]. Proceedings of the European Conference on Computer Vision, Marseille, France, 2008(5304): 696–709.
- [9] Liu L, Wang L, and Liu X. In defense of soft-assignment coding[C]. Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 2011: 2486–2493
- [10] Yang J, Yu K, Gong Y, *et al.* Linear spatial pyramid matching using sparse coding for image classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, 2009: 1794–1801.
- [11] Wang J Yang J, Yu K, *et al.* Locality-constrained linear coding for image classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 3360–3367.
- [12] Gao S, Tsang I W, Chia L, *et al.* Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 92–104.
- [13] Zhang C, Wang S, Huang Q, *et al.* Image classification using spatial pyramid robust sparse coding[J]. *Pattern Recognition Letters*, 2013, 34(9): 1046–1052.
- [14] Zhang C, Wang S, Huang Q, *et al.* Laplacian affine sparse coding with tilt and orientation consistency for image classification[J]. *Journal of Visual Communication and Image Representation*, 2013, 24(7): 786–793.
- [15] Qi G J, Hua X S, Rui Y, *et al.* Image classification with Kernelized spatial-context[J]. *IEEE Transactions on Multimedia*, 2010, 12(4): 278–287.
- [16] Wu X, Xu D, and Duan L. Action recognition using context and appearance distribution features[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2011: 489–496.
- [17] Sivic J and Zisserman A. Video google: a text retrieval approach to object matching in videos[C]. Proceedings of the International Conference on Computer Vision, Nice, France, 2003: 1470–1477.
- [18] Li F, Fergus R, and Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[J]. *Computer Vision and Image Understanding*, 2007, 10(1): 59–70.
- [19] Dean J and Ghemawat S. MapReduce: simplified data processing on large clusters[J]. *Communications of the ACM*, 2008, 51(1): 107–113.
- [20] Griffin G, Holub A, and Perona P. Caltech-256 Object Category Dataset[R]. Technical Report, California Institute of Technology, 2007.
- [21] Lowe D G. Distinctive image features from scale invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110.
- [22] Fan R E, Chang K W, Hsieh C J, *et al.* LIBLINEAR—a library for large linear classification[OL]. <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>. 2014.7.
- 肖文华：男，1988年生，博士生，研究方向为多媒体技术、指挥信息系统。
- 包卫东：男，1971年生，教授，博士生导师，研究方向为多媒体技术、指挥信息系统。
- 陈立栋：男，1985年生，讲师，研究方向为多媒体技术。