

# 非负局部约束线性编码图像分类算法

刘培娜<sup>1</sup> 刘国军<sup>1</sup> 郭茂祖<sup>1</sup> 刘扬<sup>1</sup> 李盼<sup>1</sup>

**摘要** 基于特征提取的图像分类算法的核心问题是如何对特征进行有效编码。局部约束线性编码 (Locality-constrained linear coding, LLC) 因其良好的特征重构性与局部平滑稀疏性, 已取得了很好的分类性能。然而, LLC 编码的分类性能对编码过程中的近邻数  $k$  的大小比较敏感, 随着  $k$  的增大, 编码中的某些负值元素与正值元素的差值绝对值也可能增大, 这使得 LLC 越来越不稳定。本文通过在 LLC 优化模型的目标方程中引入非负约束, 提出了一种新型编码方式, 称为非负局部约束线性编码 (Non-negative locality-constrained linear coding, NNLLC)。该模型一般采取迭代优化算法进行求解, 但其计算复杂度较大。因此, 本文提出两种近似非负编码算法, 其编码速度与 LLC 一样快速。实验结果表明, 在多个广泛使用的图像数据集上, 相比于 LLC, NNLLC 编码方式不仅在分类精确率上提高了近 1%~4%, 而且对  $k$  的选取具有更强的鲁棒性。

**关键词** 局部约束线性编码, 非负约束, 空间金字塔匹配, 图像分类

**引用格式** 刘培娜, 刘国军, 郭茂祖, 刘扬, 李盼. 非负局部约束线性编码图像分类算法. 自动化学报, 2015, 41(7): 1235–1243

**DOI** 10.16383/j.aas.2015.c140753

## Image Classification Based on Non-negative Locality-constrained Linear Coding

LIU Pei-Na<sup>1</sup> LIU Guo-Jun<sup>1</sup> GUO Mao-Zu<sup>1</sup> LIU Yang<sup>1</sup> LI Pan<sup>1</sup>

**Abstract** The most important issue of image classification algorithm based on feature extraction is how to efficiently encode features. Locality-constrained linear coding (LLC) has achieved the state-of-the-art performance on several benchmarks, due to its underlying properties of better construction and local smooth sparsity. However, the performance of LLC on image classification is sensitive to the number of neighbors, i.e., the value of  $k$ . With the increase of  $k$ , the absolute difference of some negative and positive elements may likely become larger and larger. This will make LLC more unstable. In this paper, a new coding scheme called non-negative locality-constrained linear coding (NNLLC) is proposed. It adds an extra non-negative constraint to the objective function of LLC. Generally, this new model can be solved by iterative optimization methods, however, such solutions are quite impractical due to high computational cost. Therefore, two fast approximation algorithms are proposed; more importantly, they and LLC have a similar computational complexity. To compare with LLC, the experiment results on several widely used image datasets demonstrate that NNLLC not only can improve the classification accuracy by nearly 1%~4%, but also is more robust on the selection of  $k$ .

**Key words** Locality-constrained linear coding (LLC), non-negative constraint, spatial pyramid matching (SPM), image classification

**Citation** Liu Pei-Na, Liu Guo-Jun, Guo Mao-Zu, Liu Yang, Li Pan. Image classification based on non-negative locality-constrained linear coding. *Acta Automatica Sinica*, 2015, 41(7): 1235–1243

图像自动分类系统是将一些已有类别标签的图像输入给计算机, 通过监督学习得到能够分类图像的模型, 将未知类别的图像输入给该分类模型, 得到该图像的分类。目前经典的图像分类技术包含两个主要部分: 词袋 (Bag of features, BoF)<sup>[1-3]</sup> 和空间金字塔匹配 (Spatial pyramid matching,

SPM)<sup>[4-5]</sup>。

BoF 将图像表示为局部特征的直方图, 该方法对特征的空间平移具有很强的鲁棒性, 但没有考虑特征的空间结构关系。为了弥补这一缺点, Lazebnik 等<sup>[5]</sup> 提出 SPM 方法, 将图像集中的每幅图像逐层地分割成更加精细的空间子区域, 计算每个子区域中的局部特征的直方图。通常, SPM 方法将一幅图像划分为三层, 每层子区域的数目为  $2^l \times 2^l$ ,  $l = 0, 1, 2$ 。

基于 BoF 的 SPM 方法在图像分类问题上已经取得了很好的分类结果<sup>[5]</sup>, 其整个流程主要包括: 输入图像, 对其进行特征提取、字典学习、特征编码、SPM、池化后得到最终编码, 将最终的编码向量输入到分类器中得到该输入图像的分类。

编码的有效性与鲁棒性对分类结果起着至关重要的作用。在不同的约束条件下, 学者们已经提出了

收稿日期 2014-10-28 录用日期 2015-03-03  
Manuscript received October 28, 2014; accepted March 3, 2015  
国家自然科学基金 (61171185, 61271346), 黑龙江省青年科学基金 (QC2014C071) 资助  
Supported by National Natural Science Foundation of China (61171185, 61271346) and Heilongjiang Province Science Foundation for Youths (QC2014C071)  
本文责任编辑 黄庆明  
Recommended by Associate Editor HUANG Qing-Ming  
1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001  
1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

多种重构特征描述子的编码方式, 如硬编码 (Vector quantization, VQ)<sup>[5]</sup>、改进的稀疏编码 (SPM based on sparse codes, ScSPM)<sup>[6]</sup> 和局部约束线性编码 (Locality-constrained linear coding, LLC)<sup>[7]</sup>. 传统的 SPM 使用 VQ 编码方式, 每个待重构的特征描述子的编码有且仅有一个非零元素, 该元素通过在字典中寻找待重构特征的最近邻得到. ScSPM 编码方式使用稀疏编码 (Sparse coding, SC)<sup>[8]</sup> 取代 VQ, 减小了重构误差. 但 ScSPM 得到的编码具有不稳定性, 即相似的特征描述子通过稀疏编码模型得到的编码有可能会差别很大. Yu 等<sup>[9]</sup> 指出局部性比稀疏性更加重要, 在 LLC 中, 使用视觉字典中的多个基表示一个特征描述子的方式更加精确, 且相似的特征描述子通过共享其局部的基, 得到相似的编码, 这使得 ScSPM 的不稳定性得到极大改善.

LLC 对特征描述子进行编码时, 在保证平移不变性的约束条件下, 需要最小化特征的重构误差, 因此得到的编码中可能包含负值元素. 若编码中的负值元素与正值元素相差较大, 则会导致该编码方式的不稳定性. 将编码进行非负化, 也是近年来研究的一个重要方向. Hoyer<sup>[10]</sup> 将稀疏编码与非负矩阵分解相结合形成非负稀疏编码 (Non-negative sparse coding, NNSC), 这一方法的提出主要受到神经信息处理模型的影响, 主要用于信号处理框架中. 实验表明, 非负矩阵分解的约束对于学习数据的表示是非常重要的. Zhang 等<sup>[11]</sup> 应用非负稀疏编码、低秩与稀疏矩阵分解技术 (Non-negative sparse coding, low-rank and sparse matrix decomposition, LR-Sc + SPM) 构建图像分类框架, 在一些基准图像数据集上取得了比 LLC 更高的分类结果. Lin 和 Kung<sup>[12]</sup> 提出的非负正交匹配追踪算法 (Non-negative orthogonal matching pursuit, NOMP) 是在正交匹配追踪算法 (Orthogonal matching pursuit, OMP) 中引入一个简单的非负约束, 在一个多

层深度结构中使用 K-Means 算法进行特征学习、NOMP 进行编码表示. 这种学习框架不仅能够学习大尺寸特征字典, 而且能够抵抗噪声过拟合, 在 STL-10 数据集的图像分类测试中取得了最优的结果.

因此, 为了获得更好的分类性能, 并缓解 LLC 的不稳定性问题, 本文提出了非负局部约束线性编码 (Non-negative locality-constrained linear coding, NNLLC) 算法, 与传统算法相比, 本文的主要贡献如下:

1) 在 LLC 编码模型的目标方程中引入非负约束, 形成一种新型的 NNLLC 编码模型. 该编码方式具有较强的稳定性, 且编码的分类性能对近邻数  $k$  的选择具有较强的鲁棒性.

2) 对 NNLLC 编码模型的目标方程使用一般的迭代优化算法进行求解, 探索其解的特点与计算复杂度.

3) 提出两种求解 NNLLC 编码模型的近似算法. 近似算法不仅加速了编码过程, 而且在广泛使用的图像数据集上取得了比传统算法更好的分类结果.

本文第 1 节简单描述基于 SPM 的图像分类框架. 第 2 节介绍 NNLLC 的基本思想及其稳定性的理论分析. 在第 3 节, 本文提出两种近似算法, 以取代求解 NNLLC 编码模型的一般迭代优化算法, 有效地加速了编码过程. 第 4 节在多个广泛使用的图像数据集上, 给出本文算法与其他经典算法的分类精确率与编码时间的比较结果, 以及 NNLLC 稳定性的评测. 最后得出本文结论.

## 1 基于 SPM 的图像分类整体框架

基于 SPM 的图像分类整体框架的流程如图 1 所示. 首先, 检测输入图像上的特征点, 从每个特征点提取一个  $D$  维的特征描述子, 如 SIFT、HOG 等,

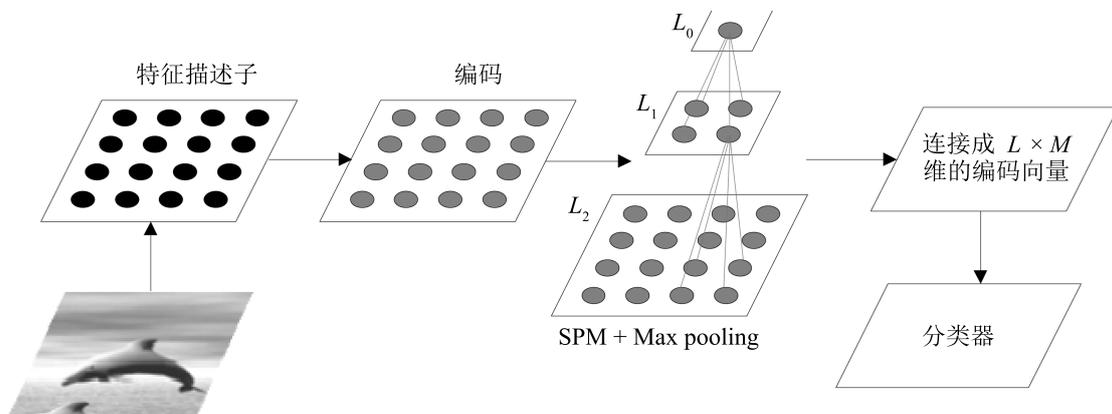


图 1 基于 SPM 的图像分类的整体框架

Fig. 1 The entire framework of image classification based on SPM

由黑色圆点表示; 其次, 重构每个特征描述子, 得到相应的  $M$  维编码向量, 由灰色圆点表示; 然后, 应用 SPM 方法将图像在空间上划分为三层, 分别为层  $L_0, L_1, L_2$ , 其中, 层  $L_l$  ( $l = 0, 1, 2$ ) 被划分为  $2^l \times 2^l$  个子区域. 对每层的每个子区域进行池化, 得到一个  $M$  维的编码向量, 则层  $L_l$  得到  $2^l \times 2^l$  个  $M$  维编码向量, 将该层的所有编码进行拼接, 得到该层的编码表示. 将三层的编码拼接起来, 形成一个  $L \times M$  维的编码向量, 该向量即是图像的最后编码, 其中子区域总个数为  $L = \sum_l 2^l \times 2^l = 2L$ . 最后, 将该输入图像的编码向量输入到分类器中进行分类, 输出类别信息.

本文的主要工作在“编码”这一步骤, 该阶段隐含了  $M$  维视觉字典的学习过程, 一般采用 K-Means 算法学习全体特征描述子上的视觉字典. 不同的编码方式得到的编码不同, 使用的分类器也就不相同, 如 LLC 编码得到的是线性编码, 只需要使用简单的线性分类器就能达到很好的图像分类效果<sup>[7]</sup>.

## 2 非负局部约束线性编码 (NNLLC)

用  $H$  表示从一幅图像中提取的  $D$  维的特征描述子集合, 如  $H = [h_1, h_2, \dots, h_N] \in \mathbf{R}^{D \times N}$ , 其中  $h_i \in \mathbf{R}^D$  是  $H$  中第  $i$  个特征描述子.  $B$  表示全部特征描述子上的视觉字典, 包含  $M$  个基, 如  $B = [b_1, b_2, \dots, b_M] \in \mathbf{R}^{D \times M}$ . 利用  $B$  对每个特征描述子  $h_i$  进行重构, 表示为  $h_i \approx B \times c_i$ , 可以使用不同的编码方式得到对应的编码  $c_i \in \mathbf{R}^M$ .

### 2.1 局部约束线性编码 (LLC)

Yu 等<sup>[9]</sup> 在理论上认为, 在某些情况下局部性比稀疏性更重要, 因为局部性一定会导致稀疏性, 反之却未必. LLC 将传统的稀疏编码方式进行改进, 通过加入局部性限制达到全局稀疏的目的, 其优化目标函数为

$$\min_C \sum_{i=1}^N \|h_i - Bc_i\|^2 + \lambda \|d_i \odot c_i\|^2 \quad (1)$$

$$\text{s.t. } \mathbf{1}^T c_i = 1, \forall i$$

其中,  $\mathbf{1}$  表示全部元素为 1 的向量, 约束条件  $\mathbf{1}^T c_i = 1$  保证编码的平移不变性, 局部正则化项  $\|d_i \odot c_i\|^2$  能够确保相似的特征描述子具有相似的编码.

### 2.2 近似 LLC 编码

式 (1) 的优化模型在求解  $c_i$  的过程中有如下规律: 待重构的特征描述子倾向于选择字典中距离其最近的基, 形成一个局部坐标系统. 对此, Wang 等<sup>[7]</sup> 提出一种简单的近似 LLC 编码方式来加速编码过程, 但最后的性能与式 (1) 的优化模型相差不

大. 近似 LLC 编码方式在保留局部特征的同时, 也保证了编码稀疏性的要求.

近似 LLC 的编码过程主要包括三个阶段: 字典学习、 $k$  近邻搜索和编码, 图 2 主要描述了后两个阶段.

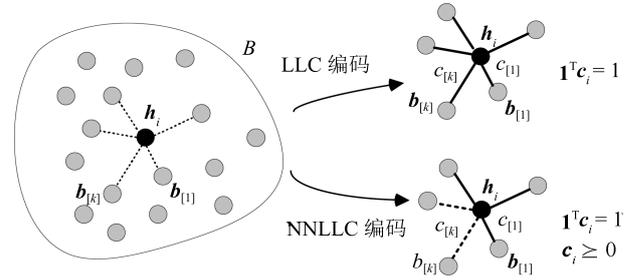


图 2 编码过程示意图

Fig. 2 The diagram of the encoding process

1) 字典学习. 从全部特征描述子中随机抽取一部分作为样本, 用于视觉字典学习, 使用 K-Means 算法学习字典  $B = [b_1, b_2, \dots, b_M] \in \mathbf{R}^{D \times M}$ .

2)  $k$  近邻搜索. 对于任意一个待编码的特征描述子  $h_i$ , 选取字典  $B$  中距离其最近的  $k$  个基, 形成子字典  $B_i = [b_{[1]}, b_{[2]}, \dots, b_{[k]}]$ .

3) 编码. 使用子字典  $B_i$  重构特征描述子  $h_i$ , 即优化式 (2) 的模型得到编码  $c_i \in \mathbf{R}^k$ .

$$\min_C \sum_{i=1}^N \|h_i - B_i c_i\|^2 \quad (2)$$

$$\text{s.t. } \mathbf{1}^T c_i = 1, \forall i$$

将最终编码  $c_i^* \in \mathbf{R}^M$  的第  $([1], \dots, [k])$  维的值设置为  $(c_{[1]}, \dots, c_{[k]})$ , 其他维的值设置为 0.

式 (2) 的优化模型可以使用拉格朗日乘子法求得解析解, 近似 LLC 编码能够将计算复杂度从 LLC 的  $O(M^2)$  减小到  $O(M + k^2)$ , 其中  $k \ll M$ .

### 2.3 非负局部约束线性编码 (NNLLC)

与近似 LLC 编码相比, 本文提出的非负局部约束线性编码 (NNLLC) 在其编码模型中加入了非负约束, 目标方程可由式 (3) 表示.

$$\min_C \sum_{i=1}^N \|h_i - B_i c_i\|^2 \quad (3)$$

$$\text{s.t. } \mathbf{1}^T c_i = 1$$

$$c_i \geq 0, \forall i$$

其中,  $c_i$  是重构特征描述子  $h_i$  后的非负编码, 等式约束保证编码的平移不变性, 不等式约束保证编码的非负性, 这要求每个特征描述子的重构编码位于其最近的  $k$  个基的凸包之内<sup>[13]</sup>. 在图 2 中, NNLLC

的编码  $\mathbf{c}_i$  中可能为 0 的元素由该图示部分的虚线表示 (NNLLC 与 LLC 编码的差别在第 2.4 节进行详细的分析描述)。

式 (3) 的优化模型无法直接求出解析解, 一般可采取迭代优化算法进行求解<sup>[14]</sup>. 实际应用中可采用 Matlab 优化工具包中的 Fmincon 函数来求解这样的非线性规划问题. 为了描述方便, 本文将使用拉格朗日乘子法求解式 (2) 的算法称为 LLC 算法, 求解得到的编码称为 LLC 编码, 将使用 Fmincon 函数优化式 (3) 的算法称为 NNLLC\_F 算法, 优化得到的编码称为 NNLLC\_F 编码。

## 2.4 NNLLC 的稳定性

将 NNLLC 与 LLC 优化模型的目标方程进行对比, 式 (3) 比式 (2) 增加了一个不等式约束条件  $\mathbf{c}_i \geq 0$ , 保证了编码的非负性. 在  $\mathbf{1}^T \mathbf{c}_i = 1$  与  $\mathbf{c}_i \geq 0$  这两个约束条件下, 式 (3) 得到的编码中的元素大小在  $[0, 1]$  之间, 而式 (2) 只包含约束条件  $\mathbf{1}^T \mathbf{c}_i = 1$ , 得到的编码中的元素大小在不可控的范围内变动, 这可能导致正值元素与负值元素相差较大。

数据集 Caltech 101 上的部分特征描述子  $\mathbf{h}_1 \sim \mathbf{h}_3$  分别由 LLC 与 NNLLC\_F 算法进行重构, 得到的编码  $\mathbf{c}_i$  ( $i = 1, 2, 3$ ) 如图 3 所示, 其中  $\mathbf{c}_i$  是五元素向量 (在  $k$  近邻搜索中设置  $k = 5$ ), 五个元素分别表示为  $c_1, c_2, c_3, c_4, c_5$ . 在图 3 中, 横坐标表示编码中的五个元素, 纵坐标表示各个元素的大小, 柱状代表使用某种算法求得的编码的元素值大小, 柱状为空代表该元素的大小为 0. 图 3(a)~3(c) 分别表示特征描述子  $\mathbf{h}_1 \sim \mathbf{h}_3$  在 LLC、NNLLC\_F 算法下的编码大小比较。

从图中可以看出, LLC 编码的负值元素与正值元素相差较大, 如图 3(b) 与图 3(c). 在图 3(c) 中, 特征  $\mathbf{h}_3$  的 LLC 编码中的元素  $c_1$ , 其大小超过了 1.2, 与元素  $c_4$  相差约为 2, 这种同一编码中的元素

相差较大的情形将导致该编码方式的不稳定性. 而 NNLLC\_F 编码的元素大小都在  $[0, 1]$  内。

从数学角度分析, 使用字典  $B_i$  对特征  $\mathbf{h}_i$  进行重构, 有  $\mathbf{h}_i \approx B_i \times \mathbf{c}_i$ , 可以看作是对  $B_i$  进行加权求和来拟合  $\mathbf{h}_i$ ,  $\mathbf{c}_i$  是要求的加权系数. 在 LLC 中, 在满足等式约束  $\mathbf{1}^T \mathbf{c}_i = 1$  的情况下, 为了最小化拟合误差,  $\mathbf{c}_i$  中正值元素与负值元素可能相差很大. 这是因为负值加权系数对应的基在一定程度上是为了抑制正值加权系数对应的基. 也就是说, 在重构特征时, 编码中的负值元素对应的基降低了正值元素对应的基所占的比重. 在编码过程中,  $k$  近邻搜索选取的  $k$  值越大, 字典中的基的数目就越多, 得到的编码中的元素个数就越多, 正值元素与负值元素相差较大的情形就越容易出现, 该编码方式就越不稳定. 而在 NNLLC 中, 由于  $\mathbf{c}_i$  约束在  $[0, 1]$  范围内, 可将  $B_i \times \mathbf{c}_i$  看作是  $B_i$  的数学期望, 以此来拟合  $\mathbf{h}_i$ , 所以该编码方式存在合理的数学意义。

从图 3 可以看出, 后者中元素为 0 的位置正好是前者中元素为负数的位置 (大部分情况中如此). 这说明, 对于同一个特征, NNLLC 编码可以看作是将 LLC 编码中为负值的元素变为 0, 其余元素再进行归一化得到. 将编码中的负值元素置 0, 使得其对应的基在重构过程中不能够发挥作用, 使用剩余的基来对特征进行重构, 使得编码中的元素大小在可控的  $[0, 1]$  范围内, 这在减小重构误差的同时, 也使得编码更加稳定。

## 3 快速近似非负编码算法

对式 (3) 采用迭代优化的算法进行求解, 其速度比 LLC 慢得多. 例如, Fmincon 优化函数在 1s 内可以对几十个特征进行编码, 但一幅图像可提取出几千个特征, 对于有上万幅图像的数据集来说, 编码耗费的时间是实际应用中不能承受的. 因此, 一般的迭代优化算法使得非负编码的实用性大打折扣。

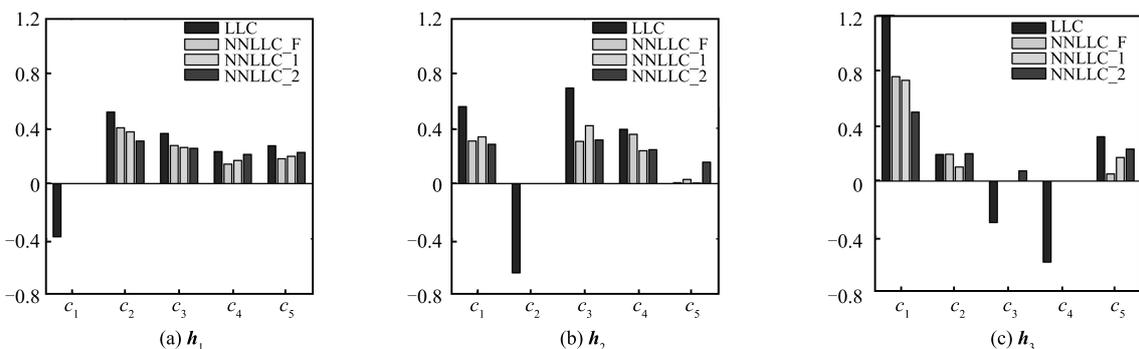


图 3  $k = 5$  时  $\mathbf{h}_1 \sim \mathbf{h}_3$  分别在四种编码算法上的编码大小比较

Fig. 3 The comparison of codes of LLC, NNLLC\_F, NNLLC\_1 and NNLLC\_2 for  $\mathbf{h}_1 \sim \mathbf{h}_3$  under  $k = 5$

为了得到非负编码, 并减小求解式 (3) 的计算复杂度, 本文提出两种近似非负编码算法 NNLLC.1 与 NNLLC.2. 将求解式 (3) 的过程分为两个步骤, 详见算法 1.

在  $k = 5$  时, 数据集 Caltech 101 上的特征描述子  $\mathbf{h}_1 \sim \mathbf{h}_3$  分别使用 NNLLC.1 与 NNLLC.2 算法进行重构, 得到相应的编码, 如图 3 所示. 在图 3 中, 图 3(a)~3(c) 分别表示特征描述子  $\mathbf{h}_1 \sim \mathbf{h}_3$  在四种算法 LLC、NNLLC.F、NNLLC.1、NNLLC.2 中的编码大小比较. 可以看出, NNLLC.1 和 NNLLC.2 编码与 NNLLC.F 编码相差不大, 三种编码中的元素大小都控制在  $[0, 1]$  之间, 不会出现相差很大的情况, 这也说明了两种近似算法能够保证 NNLLC 的稳定性. NNLLC.1、NNLLC.2 与 NNLLC.F 在  $\mathbf{h}_1 \sim \mathbf{h}_3$  上的重构误差也相差不大, 这也印证了近似非负编码算法提高图像分类效果的原因.

NNLLC.1 与 NNLLC.2 算法的计算复杂度都为  $O(M + k^2 + k)$ , 与 LLC 算法的计算复杂度  $O(M + k^2)$  相同, 其中  $k \ll M$ , 故 NNLLC.1 和 NNLLC.2 与 LLC 的编码过程一样快速.

## 4 实验及结果分析

本文实验的参数设置与文献 [7] 保持一致. 对于数据集中的彩色图像, 先将其转化为灰色图像, 并在保持原图像的高宽比例不变的前提下, 重新调整每幅图像的大小在  $300 \text{ 像素} \times 300 \text{ 像素}$  内. 将图像按照每 6 个像素划分为  $16 \text{ 像素} \times 16 \text{ 像素}$  的 patch 块, 从每个 patch 块提取一个 SIFT 特征描述子. 整个实验中只使用单一的 SIFT 特征作为输入, 每个 SIFT 特征是  $D = 128$  维的向量.

### 算法 1. NNLLC 近似编码

**输入.** 待重构的特征描述子  $\mathbf{h}$  及  $k$  近邻搜索后得到的子字典  $B$ .

**输出.**  $\mathbf{h}$  的非负编码  $\tilde{\mathbf{c}}$ .

**步骤 1.** 只考虑  $\mathbf{1}^T \mathbf{c} = 1$ , 求解拉格朗日对偶形式  $L(\mathbf{c}, \lambda) = \|\mathbf{h} - B\mathbf{c}\|^2 + \lambda(1 - \mathbf{1}^T \mathbf{c})$ , 得到解析解  $\mathbf{c} = [c_1, c_2, \dots, c_k]^T$ , 即  $\mathbf{c} = B^\dagger(B^T \mathbf{h} + \lambda \mathbf{1})$ , 其中,  $B^\dagger = (B^T B)^{-1}$ ,  $\lambda = \alpha/\beta$ ,  $\alpha = 1 - \sum_{jk} B^\dagger_{jk}(\mathbf{h}^T \mathbf{b}_{[k]})$ ,  $\beta = \sum_{jk} B^\dagger_{jk}$ .

**步骤 2.** 为了满足  $\mathbf{c} \succeq 0$ , 将步骤 1 得到的  $\mathbf{c}$  进行两种非负处理, 得到编码  $\tilde{\mathbf{c}} = [\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_k]^T$ .

1) NNLLC.1. 若  $\mathbf{c}$  中的元素有负值, 找出负值元素并置 0, 然后对  $\mathbf{c}$  进行归一化; 否则不做处理. 即

$$\tilde{c}_i = \frac{\max\{0, c_i\}}{\sum_{c_i} \max\{0, c_i\}}, \quad i = 1, 2, \dots, k$$

2) NNLLC.2. 若  $\mathbf{c}$  中的元素有负值, 找出负值绝对值最大的那个元素, 将  $\mathbf{c}$  加上该元素的绝对值后进行归一化; 否则不做处理. 即

$$\tilde{c}_i = \frac{c_i + |\min\{0, c_{i^*}\}|}{\sum_{c_i} c_i + |\min\{0, c_{i^*}\}|}, \quad i^* = \arg \min_i c_i, \\ i = 1, 2, \dots, k$$

在字典学习的过程中, 字典大小设置为  $M = 1024$ , 由于从整个数据集上提取的 SIFT 特征数目庞大, 全部用于字典学习耗费时间太长, 实验中随机采样 20 万个 SIFT 特征作为训练数据, 约占总数的  $1/70$ .

在获得每个 SIFT 的编码后, 应用 SPM 将每幅图像在空间划分为三层, 每层的子区域数目分别为  $4 \times 4, 2 \times 2, 1 \times 1$ . 然后, 每层子区域使用最大池化算法<sup>[6]</sup>后拼接起来获得该层的编码, 最后在层与层之间同样使用拼接的操作, 这样就得到每幅图像的最后编码向量表示, 即  $L \times M = 21 \times 1024$  维的向量.

在获得每幅图像的编码后, 使用 liblinear<sup>[15]</sup> 工具包进行线性分类. 根据经验设置该线性分类器的参数为 “-s 2 -c 5 -q”, 即为  $l_2$  规约  $l_2$  损失函数. 所有实验在一个 16 GB 内存、2.0 GHz 双核 CPU 的 Linux 服务器上运行.

实验内容主要包括四个部分, 首先比较不同的编码算法在三个常用图像数据集上的分类结果. 其次, 将 LLC 与 NNLLC.F、NNLLC.1、NNLLC.2 在同一数据集上的编码时间进行对比. 然后在 Caltech 101 数据集上, 在不同的编码方式下比较不同的参数  $k$  对分类结果的影响, 并对 NNLLC 的稳定性进行评测.

### 4.1 图像分类结果的比较

为了验证 NNLLC 的有效性, 本文在三个广泛使用的图像数据集 15 Scenes<sup>[5]</sup>、Caltech 101<sup>[16]</sup>、Caltech 256 上测试图像的分类精确率并进行对比. 下面对三个图像数据集进行简要介绍.

1) 15 Scenes. 包含 15 个自然场景、4485 幅图像, 其中每个类别的图像数目在 200 到 400 不等. 这 15 个自然场景差别非常大, 从厨房到山峰, 从沙滩到高速公路, 等等.

2) Caltech 101. 包含 101 个类别、9144 幅图像, 其中每个类别的图像数目在 31 到 800 不等, 包括动物、花朵、飞机等在形状上差别比较大的类别.

3) Caltech 256. 包含 256 个类别、30607 幅图像, 其中每个类别至少包含 80 幅图像. 该图像集扩展了 Caltech 101 数据集, 类别的变化性更大, 其中包含了蝙蝠、美国国旗等一些类别.

与文献 [6] 和文献 [7] 相同, 本实验分别使用 15 Scenes 图像集的每个类别中的 10, 20, 40, 60, 80, 100 幅图像, Caltech 101 图像集的每个类别中

的 5, 10, 15, 20, 25, 30 幅图像, Caltech 256 图像集 的每个类别中的 15, 30, 45, 60 幅图像作为训练数据, 剩余的图像作为测试数据. 设置最近邻数  $k = 5$ , 将本文的 NNLLC\_F 算法、NNLLC\_1 算法和 NNLLC\_2 算法与一些现有的传统算法在 15 Scenes、Caltech 101、Caltech 256 数据集上进行图像分类精确率的比较, 结果分别详见表 1、表 2 和表 3.

可以看出, 随着训练图像数目的增大, 分类性能也在逐渐地提高. 整体看来, NNLLC 算法在分类性能上要优于其他传统算法. NNLLC\_1、NNLLC\_2

两种近似算法与 NNLLC\_F 的分类性能相差不大, 然而它们在分类精确率上比 LLC 都高出了近 1%~4%. 尤其当训练图像数目较小时, 本文算法的优势更加明显. 这是因为 NNLLC 算法得到的特征编码比 LLC 更加有效, 只需要较少的训练数据就能达到比较好的分类结果. 增加训练数目能够更好地训练分类器, 所以训练数目增大, NNLLC 算法的分类精确率提高的幅度变小.

#### 4.2 编码时间的比较

对于某一数据集, 图像分类精确率固然是衡量

表 1 15 Scenes 图像集上的分类结果 (%)

Table 1 The comparison of classification accuracy on 15 Scenes dataset (%)

训练图像数目	10	20	40	60	80	10
ScSPM <sup>[6]</sup>	—	—	—	—	—	80.28
LLC	64.22	70.15	75.51	77.86	79.05	80.31
NNLLC_F	66.48	72.95	77.41	79.37	80.97	81.75
NNLLC_1	67.56	73.72	77.88	80.16	80.35	81.79
NNLLC_2	68.12	72.20	77.32	79.78	80.87	81.83

表 2 Caltech 101 数据集上的分类结果 (%)

Table 2 The comparison of classification accuracy on Caltech 101 dataset (%)

训练图像数目	5	10	15	20	25	30
Lazebnik <sup>[5]</sup>	—	—	56.40	—	—	64.60
Zhang 等 <sup>[17]</sup>	46.6	55.8	59.1	62.0	—	66.20
Boiman 等 <sup>[18]</sup>	—	—	65.0	—	—	70.40
Balasubramanian 等 <sup>[19]</sup>	—	—	—	—	—	77.54
ScSPM <sup>[6]</sup>	—	—	67.00	—	—	73.20
LLC	55.14	64.69	69.77	73.38	75.68	78.26
NNLLC_F	58.62	67.66	72.38	75.69	77.69	80.08
NNLLC_1	58.48	68.30	72.76	75.80	77.72	79.19
NNLLC_2	59.25	68.50	72.69	75.54	78.00	79.65

表 3 Caltech 256 数据集上的分类结果 (%)

Table 3 The comparison of classification accuracy on Caltech 256 dataset (%)

训练图像数目	15	30	45	60
ScSPM <sup>[6]</sup>	27.73	34.02	37.46	40.14
Gemert 等 <sup>[20]</sup>	—	27.17	—	—
LLC	30.12	36.28	40.21	43.06
NNLLC_F	33.12	39.25	42.73	45.56
NNLLC_1	33.43	39.48	43.12	45.67
NNLLC_2	33.54	39.56	43.17	45.79

图像分类性能的一个重要标准, 但整个实验消耗的时间也是一个重要的衡量标准. 如图 1 所示, 对于从开始编码到得到一个  $L \times M$  维的编码向量这一过程, 本文实验将其耗费的时间称为编码时间. 对于 NNLLC\_F、NNLLC\_1、NNLLC\_2 与 LLC 来说, 为了统计方便, 本文只比较它们的编码时间. 编码是在某个数据集的全部特征描述子上进行的, 且 15 Scenes 提取出约 728 万个特征, Caltech 101 数据集提取出约 1240 万个特征, Caltech 256 提取出约 4472 万个特征.

表 4 给出了  $k = 5$  时 NNLLC\_F、NNLLC\_1、NNLLC\_2 与 LLC 分别在 15 Scenes、Caltech 101、Caltech 256 数据集上的编码时间比较. 可以看出, 数据集越大, 提取出的特征数越多, 编码时间就越长. 同样地,  $k$  的取值越大, 编码时间也就越长. NNLLC\_1、NNLLC\_2 与 LLC 在相同数据集上的编码时间几乎相同, 却比 NNLLC\_F 的编码时间小得多.

NNLLC\_1 和 NNLLC\_2 两种近似算法不仅与 LLC 在编码速度上一样快, 而且能够达到与 NNLLC\_F 相近的图像分类精确率. 由此看来, 两种近似算法不但保留了快速编码的特性, 而且能够提高分类性能.

### 4.3 参数 $k$ 的选择与编码方式的稳定性的关系

随着编码过程中的近邻数  $k$  的增大, LLC 编码方式越来越不稳定, 而 NNLLC 能够显著缓解这一问题. 为了探索参数  $k$  与两种不同编码方式的稳定性之间的关系, 本文首先比较了 LLC 与 NNLLC 的分类性能在不同  $k$  值下的变化趋势, 对 NNLLC 的稳定性进行了评测, 然后分析了不同的  $k$  值对 NNLLC 的分类性能的影响, 对  $k$  值进行合理的选择.

#### 4.3.1 NNLLC 稳定性的评测

为了验证 NNLLC 编码在分类性能上对参数  $k$  的选择的鲁棒性, 本文在 Caltech 101 图像数据集上测试了 LLC 与 NNLLC\_1、NNLLC\_2 在不同近邻数  $k$  值上的分类性能. 图 4 给出了训练样本大小分别选取 5, 15 与 30 时, 三种算法的分类精确率在 2, 5, 10, 20, 30, 40, 50 这 7 种不同  $k$  值上变化的曲

线.

在图 4 中, 横坐标表示参数  $k$  的大小, 纵坐标表示平均分类精确率. 对 LLC 算法的分类精确率的变化曲线进行观察, 可以看出, LLC 编码的分类性能随着  $k$  的增大而明显减小, 这说明 LLC 编码的分类性能对  $k$  的选择具有很强的敏感性. 对 NNLLC\_1、NNLLC\_2 算法的分类精确率的变化曲线进行观察, 可以看出,  $k$  值在 5~20 之内, 分类精确率基本相同,  $k$  取值大于 20 后, 分类精确率开始轻微下降. 整体看来, NNLLC\_1 编码与 NNLLC\_2 编码的变化曲线比较平稳, 即使  $k$  增长到较大的数值, 分类性能只是轻微下降, 基本保持在稳定的范围内. 这说明 NNLLC 编码的分类性能对  $k$  值的选择具有较强的鲁棒性.

#### 4.3.2 参数 $k$ 的选择

NNLLC 编码方式的稳定性使得其分类性能对参数  $k$  的选择具有较强的鲁棒性, 为了探究不同的参数  $k$  与分类性能之间的变化关系, 本文在 Caltech 101 图像数据集上测试了 NNLLC\_1 与 NNLLC\_2 取不同  $k$  值时对分类性能的影响. 图 5 给出了在不同训练样本大小下, 分类精确率在 2, 5, 10, 20, 40, 50 这 6 种不同  $k$  值下变化的曲线.

在图 5 中, 横坐标表示每个类别中训练图像的数目, 纵坐标表示平均分类精确率. 通过对 6 种不同  $k$  值下的分类精确率的变化曲线进行观察, 较小的  $k$  能够取得更高的分类精确率, 这正是本文所期望的, 因为  $k$  取值越小, 计算复杂度越小, 编码速度越快.  $k = 5$  比  $k = 2$  时的分类精确率有所提高, 且  $k$  取值在 5~20 内的分类精确率变化不大.  $k$  取值大于 20 后, 分类精确率开始轻微下降. 为了加快编码速度, 可以选择较小的  $k$  值.

## 5 结论

本文通过在 LLC 编码模型的目标方程中引入非负约束, 提出了一种新型的编码方式 NNLLC. 相比于 LLC, 该编码方式不仅具有较强的稳定性, 而且对参数的选择具有更强的鲁棒性. 同时, 与使用一般的迭代优化算法求解 NNLLC 的编码模型相比, 两种近似非负编码算法能够显著地缩减编码时间. 实验结果表明, 在多个广泛使用的图像数据集上, 相比

表 4 本文算法与 LLC 在三个数据集上的编码时间比较 (小时)

Table 4 The encoding time comparison of our methods and LLC on 15 Scenes, Caltech 101 and Caltech 256 (hour)

编码算法	LLC	NNLLC_F	NNLLC_1	NNLLC_2
15 Scenes	0.93	65.75	0.94	0.96
Caltech 101	1.92	153	1.95	1.98
Caltech 256	6.50	403.36	6.58	6.68

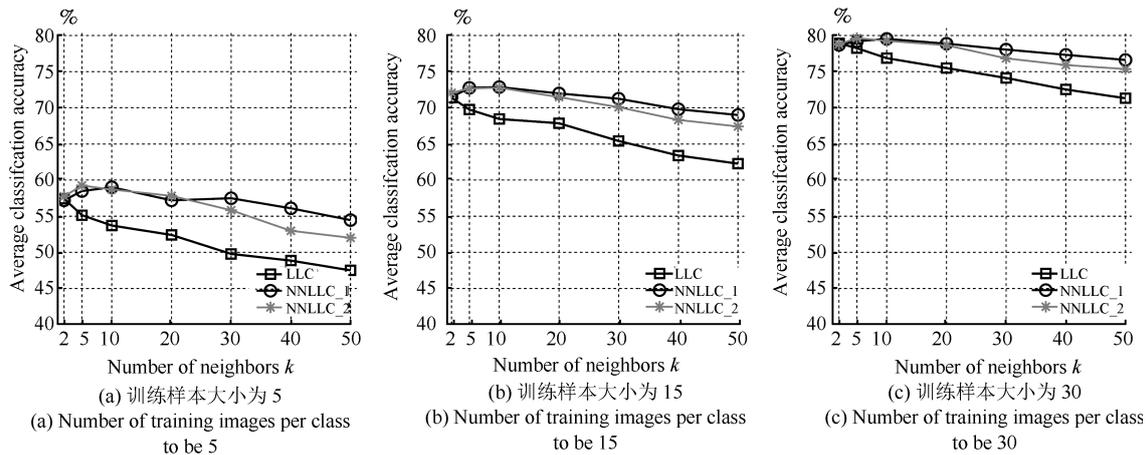


图 4 在 Caltech 101 数据集不同的训练样本大小下, LLC 与 NNLLC.1、NNLLC.2 在不同近邻数  $k$  上的分类精确率比较  
Fig. 4 The comparison of classification accuracy of LLC, NNLLC.1 and NNLLC.2 for each  $k$  on Caltech 101 dataset

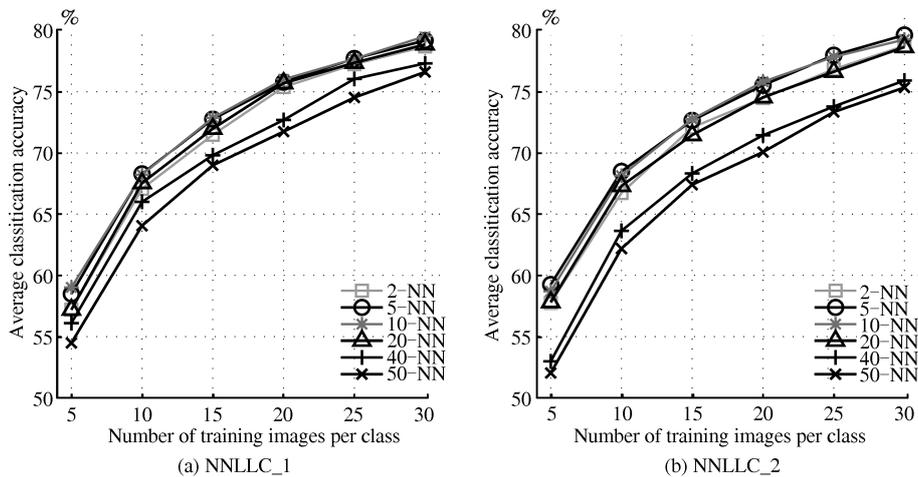


图 5 在 NNLLC.1 与 NNLLC.2 下, 不同近邻数  $k$  在 Caltech 101 数据集各个类别上的分类精确率比较  
Fig. 5 The comparison of classification accuracy of each  $k$  for each number of training images per class on Caltech 101 dataset

于 LLC, NNLLC 编码方式不仅在分类精确率上提高了近 1%~4%, 而且对  $k$  的选取具有更好的鲁棒性。

## References

- Sivic J, Zisserman A. Video google: a text retrieval approach to object matching in videos. In: Proceedings of the 9th IEEE International Conference. Nice, France: IEEE, 2003. 1470–1477
- Csurka G, Dance C R, Fan L X, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: Proceedings of the 2004 ECCV International Workshop on Statistical Learning in Computer Vision. Grenoble, France: ECCV, 2004. 1–22
- Zhao Zhong-Qiu, Ji Hai-Feng, Gao Jun, Hu Dong-Hui, Wu Xin-Dong. Sparse coding based multi-scale spatial latent semantic analysis for image classification. *Chinese Journal of Computers*, 2014, **37**(6): 1251–1260
- Li Qian-Qian. Image Classification Research of Improved Non-negative Sparse Coding [Master dissertation], Nanjing University of Science and Technology, China, 2014. (李钱钱. 基于改进的非负稀疏编码图像分类研究 [硕士学位论文], 南京理工大学, 中国, 2014.)
- Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 2169–2178
- Yang J C, Yu K, Gong Y H, Huang T. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009. 1794–1801

- 7 Wang J J, Yang J C, Yu K, Lv F J, Huang T S, Gong Y H. Locality-constrained linear coding for image classification. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3360–3367
- 8 Yang J C, Wang J P, Huang T. Learning the sparse representation for classification. In: Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (ICME). Barcelona, Spanish: IEEE, 2011. 1–6
- 9 Yu K, Zhang T, Gong Y H. Nonlinear learning using local coordinate coding. In: Proceedings of the 2009 Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2009. 2223–2231
- 10 Hoyer P O. Non-negative sparse coding. In: Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing. Falmouth, USA: IEEE, 2002. 557–565
- 11 Zhang C J, Liu J, Tian Q, Xu C S, Lu H Q, Ma S D. Image classification by non-negative sparse coding, low-rank and sparse decomposition. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 1673–1680
- 12 Lin T H, Kung H T. Stable and efficient representation learning with non-negativity constraints. In: Proceedings of the 31st International Conference on Machine Learning. Beijing, China: JMLR W & CP, 2014. 1323–1331
- 13 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, **290**(5500): 2323–2326
- 14 Goldfarb D, Idnani A. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, 1983, **27**(1): 1–33
- 15 Fan R E, Chang K W, Hsieh C J, Wang X R, Lin C J. Liblinear: a library for large linear classification. *The Journal of Machine Learning Research*, 2008, **9**: 1871–1874
- 16 Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 2007, **106**(1): 59–70
- 17 Zhang H, Berg A C, Maire M, Malik J. SVM-KNN: discriminative nearest neighbor classification for visual category recognition. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 2126–2136
- 18 Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska: IEEE, 2008. 1–8
- 19 Balasubramanian K, Yu K, Lebanon G. Smooth sparse coding via marginal regression for learning sparse representations. In: Proceedings of the 30th International Conference on Machine Learning. Edinburgh, Scotland: ICML, 2012. 1326–1334
- 20 van Gemert J C, Geusebroek J M, Veenman C J, Smeulders A W M. Kernel codebooks for scene categorization. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France: ECCV, 2008. 696–709



刘培娜 哈尔滨工业大学计算机科学与技术学院硕士研究生. 2013 年获得内蒙古大学计算机学院学士学位. 主要研究方向为计算机视觉与机器学习.

E-mail: liupeina@hit.edu.cn

(LIU Pei-Na Master student at the School of Computer Science and Technology, Harbin Institute of Technology.

She received her bachelor degree from Inner Mongolia University in 2013. Her research interest covers computer vision and machine learning.)



刘国军 博士, 哈尔滨工业大学计算机科学与技术学院讲师. 主要研究方向为计算机视觉, 图像处理与模式识别.

E-mail: hitliu@hit.edu.cn

(LIU Guo-Jun Ph.D., lecturer at the School of Computer Science and Technology, Harbin Institute of Technology. His research interest covers

computer vision, image processing, and pattern recognition.)



郭茂祖 博士, 哈尔滨工业大学计算机科学与技术学院教授. 主要研究方向为机器学习与数据挖掘, 计算生物学与生物信息学, 计算机视觉与图像处理. 本文通信作者.

E-mail: maozuguo@hit.edu.cn

(GUO Mao-Zu Ph.D., professor at the School of Computer Science and

Technology, Harbin Institute of Technology. His research interest covers machine learning and data mining, computational biology and bioinformatics, computer vision, and image processing. Corresponding author of this paper.)



刘扬 博士, 哈尔滨工业大学计算机科学与技术学院副教授. 主要研究方向为机器学习, 图像处理与计算机视觉.

E-mail: yliu76@hit.edu.cn

(LIU Yang Ph.D., associate professor at the School of Computer Science and Technology, Harbin Institute of Technology. His research interest

covers machine learning, image processing, and computer vision.)



李盼 2014 年获得哈尔滨工业大学计算机科学与技术学院硕士学位. 主要研究方向为计算机视觉与机器学习.

E-mail: lipan@hit.edu.cn

(LI Pan Received his master degree from the School of Computer Science and Technology, Harbin Institute of Technology in 2014. His research interest

covers computer vision and machine learning.)