applied sciences

MDPI

Article

A Fast Sparse Coding Method for Image Classification

Mujun Zang [1,*], Dunwei Wen [2], Tong Liu [1], Hailin Zou [1] and Chanjuan Liu [1]

1 School of Information and Electrical Engineering, Ludong University, Yantai 264025, China; liut@ldu.edu.cn (T.L.); zhl_8655@sina.com (H.Z.); luckycj80@sina.com (C.L.)
2 School of Computing and Information Systems, Athabasca University, Athabasca, AB T9S3A3, Canada; dunweiw@athabascau.ca
* Correspondence: zangmj@ldu.edu.cn; Tel.: +86-135-8980-3512

Received: 12 December 2018; Accepted: 29 January 2019; Published: 1 February 2019

Abstract: Image classification is an important problem in computer vision. The sparse coding spatial pyramid matching (ScSPM) framework is widely used in this field. However, the sparse coding cannot effectively handle very large training sets because of its high computational complexity, and ignoring the mutual dependence among local features results in highly variable sparse codes even for similar features. To overcome the shortcomings of previous sparse coding algorithm, we present an image classification method, which replaces the sparse dictionary with a stable dictionary learned via low computational complexity clustering, more specifically, a *k*-medoids cluster method optimized by *k*-means++. The proposed method can reduce the learning complexity and improve the feature's stability. In the experiments, we compared the effectiveness of our method with the existing ScSPM method and its improved versions. We evaluated our approach on two diverse datasets: Caltech-101 and UIUC-Sports. The results show that our method can increase the accuracy of spatial pyramid matching, which suggests that our method is capable of improving performance of sparse coding features.

Keywords: image classification; sparse coding; spatial pyramid matching; *k*-medoids; image feature

1. Introduction

Image classification is an important problem in computer vision. Pattern recognition and machine learning techniques have been widely applied in this field, which usually extract image features and then classify images according to the features. The setting of image features often significantly affects the performance of classification. Feature encoding methods, especially vector quantization (VQ) and sparse coding (SC), have attracted more and more attention because their well demonstrated performances.

The bag-of-features (BoF) representation methods [1–3] are important application of VQ and have been used in image classification and 1-D signal recognition [4]. This kind of representation is regarded as high-level feature because it represents an image as a vector of occurrence counts of a vocabulary built on the low-level feature extracted from subregions. The high-level feature concerns with the interpretation or classification of a scene, thus may achieve higher accuracy on most images. However, the BoF model disregards all information about the spatial layout of the features, thus may decrease the accuracy when classifying spatially sensitive images. To overcome this disadvantage, two extensions of the BoF model have been developed in the task of image classification: generative models [5–8] and spatial pyramid matching (SPM) [9].

Generative models are constructed by the latent variables learned from BoF, and classify images according to their intermediate semantics. They consider an image not only as a collection of codes [5,7] but also as a more complicated structure [10,11], which contains abundant image information and

is close to the way that human beings see the image. This can explain why generative models have gained growing attention of the research community.

Compared with generative models, SPM can achieve higher accuracy [12]. The SPM method [9] mines the spatial information through partitioning the image into increasingly finer spatial subregions and employs pyramid Match Kernel [1] to compare corresponding subregions. Typically, it partitions an image into 4^l subregions in different level ($l = 0, 1, 2$) and computes the BoF histogram within each of the 21 subregions. The BoF is a special case of SPM because SPM reduces to a standard BoF when $l = 0$. Many image classification tasks such as Caltech-101 [6] and 15-Scenes [9] have shown very promising performance of SPM.

Although the traditional SPM method works well for image classification, classifiers with nonlinear Mercer kernels, e.g., Chi-square kernel, are usually used to boost performance. Accordingly, the nonlinear SVM has a complexity $O(n^2 \sim n^3)$ in training and $O(n)$ in testing, where n is the number of training set images [13,14]. When the training set contains many samples, this complexity will cause many calculations, which implies a poor scalability for SPM's real-world applications. To overcome this drawback, Yang et al. [13] proposed an extension of SPM by using Sparse Coding (ScSPM) and achieved state-of-the-art performance on several benchmarks. The ScSPM computes a spatial-pyramid image representation by sparse coding (SC) instead of vector quantization (VQ).This method works better with employing linear classifiers, and thus reduces the computation of classification algorithm [14–16].

The ScSPM method represents an image through the following three modules: (1) SIFT extraction; (2) sparse coding; and (3) spatial pooling. The sparse coding contains dictionary learning and feature quantization, which are the most important and govern the quality of image presentation. The ScSPM employs efficient sparse coding method [17] that adopts Lagrange dual to learn dictionary whose components are linear combinations of low-level features. The method has been shown experimentally to be much faster than first-order gradient descent methods [17], and is capable of generating state-of-the-art features in image classification. However, to further improve its performance, some problems need to be solved. First, it cannot effectively handle large training sets because the dictionary learning algorithm has high computational complexity [18,19]. Second, local features are dealt with separately, thus the mutual dependence among local features is ignored, which results in highly variable sparse codes even for similar features [16].

To solve these problems, we propose a self-representation sparse coding strategy that separates dictionary learning algorithm from sparse coding. Under this strategy, sparse dictionary is built with low-level feature extracted from some important subregions. This is different from existing sparse coding method, which learns a dictionary by employing linear combination of low level feature extracted from all subregions in training dataset. We employ cluster centers of k-medoids cluster to build our dictionary, because this cluster builds cluster centers by stably selecting local feature extracted from important subregions and can reduce the learning complexity of Lagrange dual. Therefore, the proposed method can solve the sparse coding's problem of high computational complexity and variable sparse codes for similar features.

Based on this strategy, we propose an image classification framework, as shown in Figure 1. Firstly, SIFT descriptors are extracted from each of subregions. Secondly, k-medoids cluster is applied to learn a dictionary from SIFT descriptors. Then, ScSPM method is employed to generate image feature. Finally, classification is performed to classify images according to image feature.

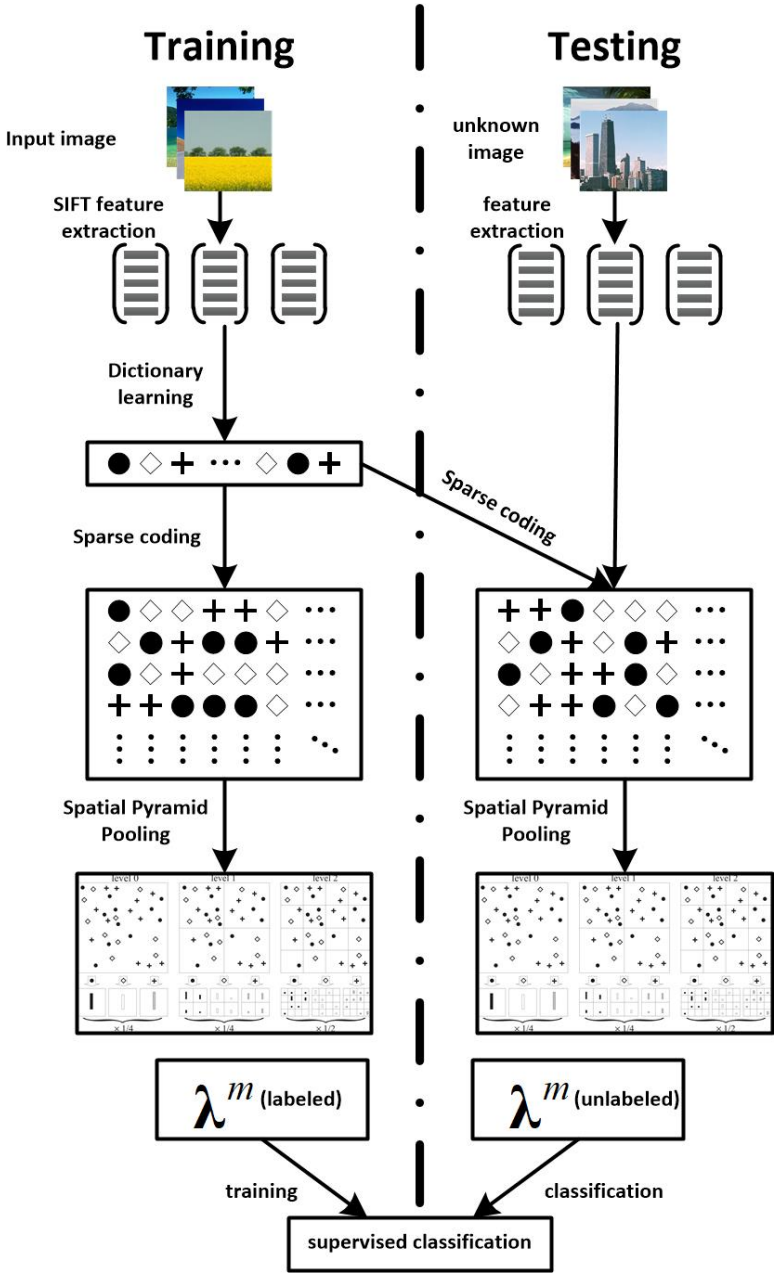

Figure 1. Flow chart of our framework.

2. Methods

2.1. SIFT Feature Extraction

The scale-invariant feature transform (SIFT) is employed as low-level feature in our framework. It is an algorithm in computer vision to detect and describe local regions in images. Four different ways of extracting local regions have been tested [5], which proved that the region descriptors of 128-dimensional SIFT [20] have more useful information and better robustness. These local features have been widely used in image classification [7,21,22]. Therefore, we employed a common setting of regions descriptors that adopt 128-dimensional gray level intensity of SIFT.

2.2. k-Medoids Dictionary Learning

The cluster samples are 128-dimensional SIFT descriptors that are extracted from locals of each image, and the *k*-means++ optimized *k*-medoids cluster [4] is employed as dictionary

learning algorithm, which chooses data points as centers (medoids or exemplars) and works with a generalization of the Manhattan Norm to define distance between data points.

The *k*-medoids dictionary learning algorithm is shown in Table 1. Firstly, *k*-means++ algorithm is adopted to initialize medoids. The first medoid is set by random extraction from all cluster samples. Then, iterations until *k* initialize medoids are selected; in each iteration, a new medoid is selected based on selecting probability of each sample. Secondly, selected medoids are employ as initialized medoids to apply *k*-medoids cluster algorithm. Finally, convergent cluster centers are adopted as dictionary for sparse coding.

Table 1. The algorithm of fast dictionary learning.

// Initialization
// Input all samples as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$
// Randomly select a sample to join the medoids set $\mathbf{M}_{set}$ as the first medoid $\mathbf{M}_{set}^{(1)}$
do $\mathbf{x}_{random} \Rightarrow \mathbf{M}_{set}$
// Iteratively search for other medoids
for $m = 2 : K$
// Allocation cluster numbers of candidate samples
for $j = 1 : N$
do $\tilde{p}_j = \arg\min_{p} \|\mathbf{x}_j - \mathbf{M}_{set}^{(p)}\|$
end
// Calculating the selecting probability of the candidate samples
for $j = 1 : N$
do $P_j = \dfrac{\|\mathbf{x}_j - \mathbf{M}_{set}^{(\tilde{p}_j)}\|}{\sum_{h=1}^{N} \|\mathbf{x}_h - \mathbf{M}_{set}^{(\tilde{p}_h)}\|}$
end
// Extracting a candidate sample as a medoid by probability
do $\mathbf{x}_{random\ j \sim \mathbf{P}_j} \Rightarrow \mathbf{M}_{set}$
end
// Learning dictionary
// Allocation cluster number
for $i = 1 : N$
do $w_i = \arg\min_{k} \|\mathbf{x}_i - \mathbf{M}_{set}^{(k)}\|$
end
// Update dictionary
$\mathbf{M}_{oldset} = \mathbf{M}_{set}$
for $k = 1 : K$
do $\mathbf{M}_{set}^{(k)} = \arg\min_{\mathbf{x}_i} \sum_{j:w_j=k} \|\mathbf{x}_i - \mathbf{x}_j\|$
end
// To determine whether convergence
if $\mathbf{M}_{set} = \mathbf{M}_{oldset}$
do **"Output $\mathbf{M}_{set}$ as medoids"**
else **Go to the step "// Allocation cluster number"**
end

2.3. Sparse Coding

An efficient sparse coding algorithm [17] is employed as mid-level feature, which represents images as a set of codes according to its 128-dimensional SIFT descriptors that are extracted from locals of each image. This sparse coding algorithm has a dictionary learning step and an encoding step, and is based on iteratively solving two convex optimization problems: an L1-regularized least squares problem and an L2-constrained least squares problem. In our framework, the dictionary learning step is replaced by the proposed *k*-medoids dictionary learning method, and existing encoding step is employed to generate the image coding based on dictionary learned by our method.

2.4. ScSPM Feature Building and Classification

The ScSPM model [13] is employ for generalizing vector quantization to sparse coding followed by multi-scale spatial max pooling and employed a linear SPM kernel based on SIFT sparse codes. For any image represented by a set of descriptors, we can compute a single feature vector based on the descriptors' codes generated by sparse coding algorithm. Let U be the code obtained via sparse coding algorithm; ScSPM feature is computed by the histogram pooling method:

$$z = \frac{1}{M} \sum_{m=1}^{M} u_m \tag{1}$$

In ScSPM, an approach of using linear SVMs based SC of SIFT is advocated. Let U be the result of applying the sparse coding Equation (1) to a descriptor set; assuming the dictionary to be pre-learned and fixed, image feature is computed by a pre-chosen pooling function:

$$z = F(\mathrm{U}) \tag{2}$$

where the pooling function F is defined on each column of U, and each column of U corresponds to the responses of all the local descriptors to one specific item in dictionary. The pooling function F is a max pooling function on the absolute sparse codes:

$$z_j = \max\left\{|u_{1j}|, |u_{2j}|, \cdots, |u_{Mj}|\right\} \tag{3}$$

where z_j is the jth element of z, u_{ij} is the matrix element at ith row and jth column of U, and M is the number of local descriptors in the region.

Let image I_i be represented by z_i, a simple linear SPM kernel, as shown in Equation (4), are employed in SVM, so that image can be labeled by SVM training and classification algorithm.

$$k(z_i, z_j) = z_i^{\mathrm{T}} z_j = \sum_{l=0}^{2} \sum_{s=1}^{2^l} \sum_{t=1}^{2^l} \left\langle z_i^l(s,t), z_j^l(s,t) \right\rangle \tag{4}$$

where $\langle z_i, z_j \rangle = z_i^{\mathrm{T}} z_j$, and $z_i^l(s,t)$ is the max pooling statistics of the descriptor sparse codes in the (s,t)th segment of image I_i in the scale level l.

3. Results

We evaluated our approach on two diverse datasets: Caltech-101 [6] and UIUC-Sports [23]. All experiments were repeated ten times with different, randomly selected training and testing images, and the final results were averaged for classification accuracy. Multi-class classification was done with a simple linear SVM classifier trained using the one-versus-all rule: a classifier was learned to separate each class from the rest, and a test image was assigned the label of the classifier with the highest response.

3.1. Caltech-101 Dataset

The Caltech-101 dataset contains 101 classes (including animals, vehicles, flowers, etc.) and totals 9144 images with high shape variability, which are provided by Fei-Fei Li [6]. The number of images per category varies from 31 to 800. Most images are medium resolution, i.e., about 300 × 300 pixels. We followed the common experiment setup for Caltech-101, i.e., training on 15 and 30 images per category, denoted as *15 training* experiment and *30 training* experiment, and testing on the rest.

Table 2 shows the comparison between our approach and other SPM based features. Our approach could classify images at the highest accuracy on both *15 training* experiment and *30 training* experiment. Especially, the accuracy of our method is much higher than that methods based on SPM

technology, such as SPM [1], ScSPM [13], and LLC [14]. Different from existing SPM based methods, our method adopts the proposed dictionary learning strategy; the results prove that this strategy yields an observable performance boost for SPM. Table 3 shows the comparison between our approach and the state-of-the-art methods. Our approach could classify images at the highest accuracy. Although the computational complexity of our approach is much lower than deep learning methods, the accuracy of our method is higher than deep learning method reported in [24] and equal to deep learning method reported in [25], which shows advantage of our method in both reducing computational cost and increasing accuracy.

Table 2. Classification accuracy comparison between our approach and other SPM based features on Caltech-101 dataset.

Algorithms	**15 Training**	**30 Training**
SPM [1]	56.4%	66.4%
ScSPM [13]	67.0%	73.2%
LLC [14]	65.4%	73.4%
Local Pooling [26]	-	77.3%
GLP [27]	70.3%	82.7%
DASDL$_p$ [28]	-	75.5%
N^3SC encoder [29]	67.5%	73.9%
FScSPM (Our Approach)	**76.3%**	**84.8%**

Table 3. Classification accuracy comparison between our approach and the state-of-the-art on Caltech-101 dataset.

Algorithms	**15 Training**	30 **Training**
HVFC-HSF [30]	70.7%	78.7%
CLGC(RGB-RGB) [31]	-	72.6%
CSAE [24]	64.0%	71.4%
Hybrid-CNN [25]	-	**84.8%**
FScSPM (Our Approach)	**76.3%**	**84.8%**

3.2. UIUC-Sports Dataset

This dataset is composed of eight complex event classes provided by Li-Jia Li and Fei-Fei Li [23], and contains 1579 color images of different sizes. There are 194 *rock climbing*, 200 *badminton*, 137 *bocce*, 236 *croquet*, 182 *polo*, 250 *rowing*, 190 *sailing*, and 190 *snowboarding* images. We followed their experimental setting in Object Bank [32] by using 70 randomly drawn images from each class for training and 60 for testing.

Table 4 shows the comparison between our approach and other SPM based features. Our approach could classify images at the highest accuracy. Our method could increase the accuracy by 1.4% compared to LScSPM and 4.4% Equation Local Soft Assignment based on SPM technology, which proves that the proposed strategy boosts the performance of SPM. Table 5 shows the comparison between our approach and state-of-the-art methods. Our approach could increase the accuracy compared to feature extraction models, such as LLKc [33], N^3SC encoder [29], and CLGC(RGB-RGB) [31]. Although the accuracy of our approach was lower than deep learning models, such as Hybrid-CNN [25], TPN-FS [34], and DeepSCNet [35], it could reduce computational cost compared to deep learning methods, which suggest that it is still useful in real-time processing systems.

Table 4. Classification accuracy comparison between our approach and other SPM based features on UIUC-Sports dataset.

Algorithms	Avg. Accuracy
LScSPM [16]	85.3%
MR-BoF [36]	85.1%
HILLC+SPM [37]	85.0%
SNDL [38]	85.2%
FScSPM (Our Approach)	**86.7%**

Table 5. Classification accuracy comparison between our approach and state-of-the-art on UIUC-Sports dataset.

Algorithms	Avg. Accuracy
Local Soft Assignment [39]	82.3%
LLKc [33]	86.4%
N^3SC encoder [29]	85.5%
CLGC(RGB-RGB) [31]	86.4%
MUSIC [40]	81.8%
OB2014 [41]	82.3%
Hybrid-CNN [25]	**94.2%**
TPN-FS [34]	**95.2%**
DeepSCNet [35]	**87.1%**
FScSPM (Our Approach)	**86.7%**

4. Discussion

Image classification is an important research field in computer vision. The ScSPM method, which represents an image through SIFT extraction, sparse coding, and spatial pooling, has shown its advantage in this field. However, the dictionary learning involved in sparse coding demands high computational complexity and is very unstable. To solve these problems, in this paper, we propose an image classification framework, whose dictionary learning and feature quantization are developed as two independent parts of sparse coding. Firstly, SIFT descriptors are extracted from each of subregions. Secondly, *k*-medoids cluster is applied to learn a dictionary from SIFT descriptors. Then, ScSPM method is employed to generate image feature. Finally, classifier is applied to classify images according to image feature.

We evaluated our approach on two diverse datasets: Caltech-101 and UIUC-Sports. The results show that our method can increase accuracy of spatial pyramid matching, which suggests that our method can improve performance of sparse coding features.

Although the proposed method has improved the classification performance, similar categories may be misclassified mutually. To improve this situation, in our future work, we will try to use more efficient low-level features, and apply multiple stage classifiers for boosting the classify performance.

Author Contributions: Conceptualization, M.Z. and D.W.; methodology, M.Z., D.W. and H.Z.; software, T.L.; validation, M.Z., T.L. and C.L.; formal analysis, H.Z.; investigation, C.L.; resources, M.Z.; data curation, T.L.; writing—original draft preparation, M.Z.; writing—review and editing, D.W. and H.Z.; visualization, T.L. and C.L.; supervision, D.W.; project administration, M.Z.; and funding acquisition, M.Z.

Funding: This research was funded by the National Natural Science Foundation of China under Grant No.61741311, National Natural Science Foundation of China under Grant No.61702249, and Natural Science Foundation of Shandong Province under Grant No.ZR2017PF010.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grauman, K.; Darrell, T. The pyramid match kernel: Discriminative classification with sets of image features. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2, pp. 1458–1465.
2. Zhang, J.; Marszałek, M.; Lazebnik, S.; Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vis.* **2007**, *73*, 213–238. [CrossRef]
3. Wu, J.; Rehg, J.M. Beyond the euclidean distance: Creating effective visual codebooks using the histogram intersection kernel. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 9 September–2 October 2009; pp. 630–637.
4. Liu, T.; Si, Y.; Wen, D.; Zang, M.; Lang, L. Dictionary learning for VQ feature extraction in ECG beats classification. *Expert Syst. Appl.* **2016**, *53*, 129–137. [CrossRef]
5. Fei-Fei, L.; Perona, P. A bayesian hierarchical model for learning natural scene categories. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 524–531.
6. Fei-Fei, L.; Fergus, R.; Perona, P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.* **2007**, *106*, 59–70. [CrossRef]
7. Chong, W.; Blei, D.; Li, F.F. Simultaneous image classification and annotation. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1903–1910.
8. Zang, M.; Wen, D.; Wang, K.; Liu, T.; Song, W. A novel topic feature for image scene classification. *Neurocomputing* **2015**, *148*, 467–476. [CrossRef]
9. 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, NY, USA, 17–22 June 2006; pp. 2169–2178.
10. Niu, Z.; Hua, G.; Gao, X.; Tian, Q. Context aware topic model for scene recognition. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2743–2750.
11. Wang, X.; Grimson, E. Spatial latent dirichlet allocation. In Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1577–1584.
12. Zang, M.; Wen, D.; Liu, T.; Zou, H.; Liu, C. A pooled Object Bank descriptor for image scene classification. *Expert Syst. Appl.* **2018**, *94*, 250–264. [CrossRef]
13. Yang, J.; Yu, K.; Gong, Y.; 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, USA, 20–25 June 2009; pp. 1794–1801.
14. Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained linear coding for image classification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
15. Yu, K.; Zhang, T.; Gong, Y. Nonlinear learning using local coordinate coding. In Proceedings of the Twenty-Third Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; pp. 2223–2231.
16. Gao, S.; Tsang, I.W.H.; Chia, L.T.; Zhao, P. Local features are not lonely—Laplacian sparse coding for image classification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3555–3561.
17. Lee, H.; Battle, A.; Raina, R.; Ng, A.Y. Efficient sparse coding algorithms. In Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 4–7 December 2006; pp. 801–808.
18. Bottou, L.; Bousquet, O. The tradeoffs of large scale learning. In Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 161–168.

19. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G. Online dictionary learning for sparse coding. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 689–696.
20. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
21. Quelhas, P.; Monay, F.; Odobez, J.M.; Gatica-Perez, D.; Tuytelaars, T.; Van Gool, L. Modeling scenes with local descriptors and latent aspects. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 1, pp. 883–890.
22. Bosch, A.; Zisserman, A.; Muñoz, X. Scene classification via pLSA. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 517–530.
23. Li, L.J.; Fei-Fei, L. What, where and who? Classifying events by scene and object recognition. In Proceedings of the IEEE International Conference on Computer Vision, io de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
24. Luo, W.; Li, J.; Yang, J.; Xu, W.; Zhang, J. Convolutional sparse autoencoders for image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3289–3294. [CrossRef] [PubMed]
25. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning deep features for scene recognition using places database. In Proceedings of the Twenty-eighth Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 487–495.
26. Boureau, Y.L.; Le Roux, N.; Bach, F.; Ponce, J.; LeCun, Y. Ask the locals: Multi-way local pooling for image recognition. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2651–2658.
27. Feng, J.; Ni, B.; Tian, Q.; Yan, S. Geometric ℓ_p-norm feature pooling for image classification. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2609–2704.
28. Yang, M.; Chang, H.; Luo, W. Discriminative analysis-synthesis dictionary learning for image classification. *Neurocomputing* **2017**, *219*, 404–411. [CrossRef]
29. Zhang, S.; Wang, J.; Shi, W.; Gong, Y.; Xia, Y.; Zhanga, Y. Normalized Non-Negative Sparse Encoder for Fast Image Representation. *IEEE Trans. Circuits Syst. Video Technol.* **2018**. [CrossRef]
30. Lin, G.; Fan, C.; Zhu, H.; Miu, Y.; Kang, X. Visual feature coding based on heterogeneous structure fusion for image classification. *Inf. Fus.* **2017**, *36*, 275–283. [CrossRef]
31. Kabbai, L.; Abdellaoui, M.; Douik, A. Image classification by combining local and global features. *Vis. Comput.* **2018**, 1–15. doi:10.1007/s00371-018-1503-0. [CrossRef]
32. Li, L.J.; Su, H.; Fei-Fei, L.; Xing, E.P. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In Proceedings of the Twenty-fourth Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; pp. 1378–1386.
33. Liu, Q.; Liu, C. A novel locally linear KNN method with applications to visual recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2010–2021. [PubMed]
34. Bai, S.; Li, Z.; Hou, J. Learning two-pathway convolutional neural networks for categorizing scene images. *Multimed. Tools Appl.* **2017**, *76*, 16145–16162. [CrossRef]
35. Zhang, S.; Wang, J.; Tao, X.; Gong, Y.; Zheng, N. Constructing deep sparse coding network for image classification. *Pattern Recognit.* **2017**, *64*, 130–140.
36. Zhou, L.; Zhou, Z.; Hu, D. Scene classification using a multi-resolution bag-of-features model. *Pattern Recognit.* **2013**, *46*, 424–433. [CrossRef]
37. Chen, H.; Xie, K.; Wang, H.; Zhao, C. Scene image classification using locality-constrained linear coding based on histogram intersection. *Multimed. Tools Appl.* **2018**, *77*, 4081–4092. [CrossRef]
38. Hu, J.; Tan, Y.P. Nonlinear dictionary learning with application to image classification. *Pattern Recognit.* **2018**, *75*, 282–291. [CrossRef]
39. Liu, L.; Wang, L.; Liu, X. In defense of soft-assignment coding. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2486–2493.

40. Li, L.J.; Zhu, J.; Su, H.; Xing, E.P.; Fei-Fei, L. Multi-level structured image coding on high-dimensional image representation. In Proceedings of the Asian Conference on Computer Vision, Daejeon, Korea, 5–9 November 2012; pp. 147–161.
41. Li, L.J.; Su, H.; Lim, Y.; Fei-Fei, L. Object bank: An object-level image representation for high-level visual recognition. *Int. J. Comput. Vis.* **2014**, *107*, 20–39. [CrossRef]