

Research Article

An Analysis and Application of Fast Nonnegative Orthogonal Matching Pursuit for Image Categorization in Deep Networks

Bo Wang, Jichang Guo, and Yan Zhang

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

Correspondence should be addressed to Bo Wang; neuwb@tju.edu.cn

Received 25 March 2015; Revised 15 June 2015; Accepted 18 June 2015

Academic Editor: Oscar Reinoso

Copyright © 2015 Bo Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonnegative orthogonal matching pursuit (NOMP) has been proven to be a more stable encoder for unsupervised sparse representation learning. However, previous research has shown that NOMP is suboptimal in terms of computational cost, as the coefficients selection and refinement using nonnegative least squares (NNLS) have been divided into two separate steps. It is found that this problem severely reduces the efficiency of encoding for large-scale image patches. In this work, we study fast nonnegative OMP (FNOMP) as an efficient encoder which can be accelerated by the implementation of QR factorization and iterations of coefficients in deep networks for full-size image categorization task. It is analyzed and demonstrated that using relatively simple gain-shape vector quantization for training dictionary, FNOMP not only performs more efficiently than NOMP for encoding but also significantly improves the classification accuracy compared to OMP based algorithm. In addition, FNOMP based algorithm is superior to other state-of-the-art methods on several publicly available benchmarks, that is, Oxford Flowers, UIUC-Sports, and Caltech101.

1. Introduction

In computer vision, image representation is a core topic for image understanding and processing. Over the past decade, sparsity has been implemented as one of the priors for a good encoder which makes the corresponding representations more useful when building classifiers [1]. In particular, it is suitable for categorization tasks as sparse representations are more likely to be separable in high dimensional spaces.

It is well known that the classical sparse coding with imposing l_1 norm regularization achieves impressive performance for face recognition, text classification, and robotic perception tasks [2–4], whereas orthogonal matching pursuit (OMP), the canonical greedy algorithm for sparse approximation, can commonly replace the l_1 relaxed algorithm due to its high efficiency in large-scale problems. While OMP as encoder shows simplicity and fast execution for many tasks, in practice it is not optimal in terms of stability. In other terms, such a greedy algorithm can augment small data variations which give rise to large deviations in terms of representations [5].

With the development of study on nonnegativity constraints in numerical analysis, nonnegative least squares (NNLS) and nonnegative matrix factorization (NMF), which are frequently used tools, have been applied in image processing and computer vision where the experiments show that enforcing a nonnegativity constraint can produce a much more accurate approximate solution [6]. Therefore, nonnegativity constraints can be employed to ameliorate the aforementioned instability of OMP. Furthermore, it is shown that nonnegative sparse coding is useful for modeling human vision systems on natural images in visual neuroscience [7]. More importantly, nonnegative sparse coding has also appeared in various other applications, such as motion extraction, text classification, and human action recognition [8–11].

On the other hand, current research on sparse representation learning falls into two groups which are dependent on manually designed descriptors, that is, SIFT [12, 13], and derive from pixel level via hierarchical structure, respectively [14, 15]. As a matter of fact, the latter is referred to as layerwise unsupervised training which advocates to

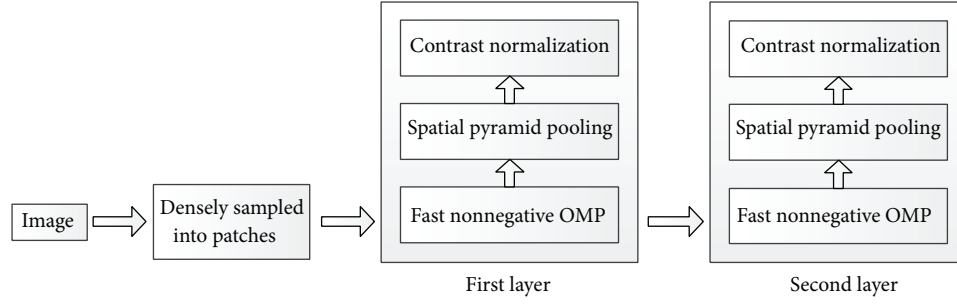


FIGURE 1: Hierarchical learning scheme adopted in this work.

build models from scratch instead of strong dependence on descriptors. A considerable amount of work is dedicated to learning this deep architecture. Specifically, deep belief nets and convolutional deep belief networks make use of stacked Restricted Boltzmann Machines (RBM) to learn high-level image features from low-level ones for recognition [16, 17]. Deconvolutional networks concentrating on high quality latent representations take advantage of a decoder-only model as opposed to the symmetric encoder-decoder of the RBM [18]. Deep autoencoders investigate the feasibility of building high-level features from only unlabeled data and obtain neurons that function as detectors for faces, human bodies, and cat faces [19]. Deep convolutional neural networks are capable of achieving record-breaking results on a highly challenging Imagenet dataset by using purely supervised learning [20]. It is a remarkable fact that a popular architecture based on multilayers matching pursuit encoders has achieved great success over the last few years [21, 22].

Intuitively, an unsupervised hierarchical training manner combined with nonnegative sparse coding should be taken into account. According to the point of view proposed in [28], it is desirable to obtain good image representations on top of nonnegative sparsity. The 4-layer model is trained on a 24-core CPU and an Nvidia Tesla M2075 GPU for fast computing. As a result, this trained model based on ISTA algorithm layer-by-layer has shown slightly better performance with high computational configuration on object classification. In addition, the nonnegative OMP (NOMP) put forward by Lin and Kung [5] can be regarded as a more stable encoder in hierarchical architecture. However, NOMP is only applied to small-size images in the first layer of model and several complicated preprocessing steps are also needed for layer-1 as well as the sign-splitting technique. In spite of delivering competitive accuracy to some best known encoders, NOMP is actually not very efficient on account of separation of selecting and NNLS steps, which is verified on synthetic data in [30].

For this reason, by studying and analyzing efficient orthogonal matching pursuit with nonnegativity constraints called fast nonnegative OMP (FNOMP) in deep networks for full-size image categorization, we demonstrate benefits of the novel encoder. In this paper, firstly we compare computational efficiency of FNOMP encoder with NOMP encoder under different experiment conditions. Next, we

consider classification accuracy of FNOMP based algorithm on three categories of object and event datasets in comparison to OMP based deep learning models and other state-of-the-art approaches.

The main contribution of this paper is that we validate the computational time of novel FNOMP, which is significantly shorter than that of NOMP in encoding combined with dictionaries of different sizes and various sparsity levels. Then, it is shown that FNOMP based algorithm can obtain meaningful image representations and therefore is appropriate for full-size image classification in deep networks. Moreover, traditional preprocessing steps comprising mean subtraction, whitening, and sign-splitting are not applied in our method, which simplifies the whole process. Finally, it is found that image size has a great influence on classification accuracy.

The remainder of this paper falls into four sections. In Section 2, the definition of the hierarchical framework for categorization is given. In Section 3, the dictionary training and efficient OMP with nonnegativity constraints are presented. Then, in Section 4, details of our experimental results and analysis on several datasets are elaborated. Finally, in Section 5, the conclusion is drawn.

2. Hierarchical Learning in Deep Networks

Recently, it is desirable to propose fully automatic approaches which can replace those hand-designed descriptors. Meanwhile, a typical manner in machine learning has focused on learning good representations from unlabeled input data for higher-level tasks such as image categorization. More specifically, the hierarchical structures learn multilayer features by greedily training several layers, one layer at a time. For example, a 2-layer deep model which computes sparse codes with fast nonnegative OMP in each layer can be trained as shown in Figure 1.

As can be seen from Figure 1, the densely sampled image patches are computed with FNOMP for sparse codes in the first layer, which are converted as input for the second layer. Then, the higher image-level representations are provided by similar steps from the first layer.

In practice, as discussed in [15, 21], the deep network implementations are generally composed of four steps.

Given an image of n -by- n pixels with p channels, the pipeline can be illustrated in Figure 2.

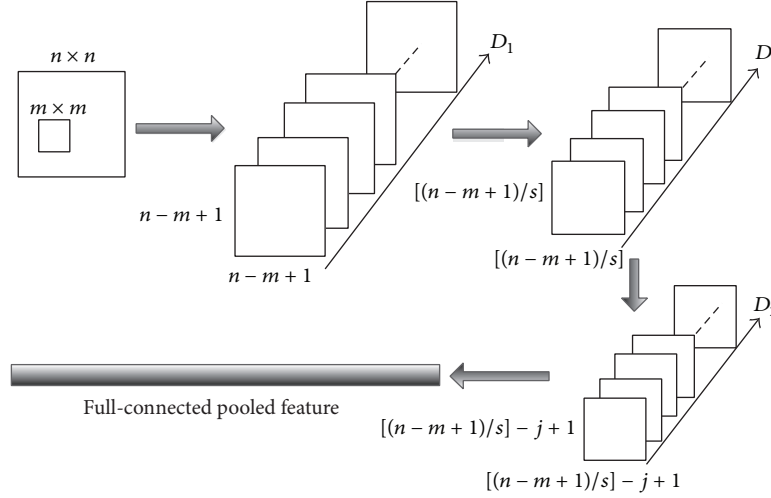


FIGURE 2: Deep architecture of computing representation.

- (i) m -by- m pixel receptive field with a step of one pixel between them is used for the first layer of features. After training the dictionary with D_1 filters for the first layer, we find that the image takes on a $(n-m+1)$ -by- $(n-m+1)$ -by- D_1 representation based on fast nonnegative OMP pattern.
- (ii) Max pooling strategy is employed over adjacent s -by- s spatial blocks; then, a $[(n-m+1)/s]$ -by- $[(n-m+1)/s]$ -by- D_1 pooled representation is generated.
- (iii) j -by- j pixel receptive field with a step of one pixel over the whole D_1 maps yields the second layer which is featured by j -by- j -by- D_1 dimension and the corresponding number of feature is $[(n-m+1)/s] - j + 1$ -by- $[(n-m+1)/s] - j + 1$. Akin to the dictionary training stage in the first step, the image finally obtains a $[(n-m+1)/s] - j + 1$ -by- $[(n-m+1)/s] - j + 1$ -by- D_2 representation by means of efficient OMP with nonnegativity constraints.
- (iv) Pyramid max pooling and contrast normalization are also applied to form final pooled representation.

3. Sparse Coding with Efficient Nonnegative OMP

3.1. Dictionary Training. The gain-shape vector quantization for training in deep networks will be implemented throughout this work. Let S be a set of M -dimensional input signals; that is, $S = [s_1, \dots, s_N] \in R^{M \times N}$. Specifically, the dictionary is trained by using an alternating manner described as follows:

$$\begin{aligned} \min_{D, x_j} \quad & \sum_j \|s_j - Dx_j\|_2^2 \\ \text{s.t.} \quad & \|x_j\|_0 \leq T, \forall j, \|D_i\|_2^2 = 1, \forall i, \end{aligned} \quad (1)$$

where D_i indicates each column of dictionary D , $\|D_i\|_2^2 = 1$ makes each dictionary element normalized, $\|x_j\|_0$ is

the number of nonzero elements in x_j , and T is a sparsity constraint factor. For instance, OMP-1 will be used as a form of gain-shape vector quantization, and then it begins with $x_j = 0$ and greedily selects an element of x_j to be nonzero to minimize the residual reconstruction error at each iteration.

3.2. Efficient OMP with Nonnegativity Constraints. The standard nonnegative OMP (NOMP) can be applied to find an approximate solution to the following problem:

$$\begin{aligned} \min_x \quad & \|s - Dx\|_2 \\ \text{s.t.} \quad & \|x\|_0 \leq T, x_k \geq 0, \forall k, \end{aligned} \quad (2)$$

where NOMP computes codes with at most T nonzero elements and all elements are nonnegative. Generally, the pipeline of NOMP can be summarized as follows:

- (i) Firstly, the residual vector v is initialized as $v^{(0)} = s$ and iteration number l is set to be 1. In order to have the highest positive correlation with residual, the algorithm needs to choose the atom d_{r_l} ; that is $r_l = \arg \max_r \langle d_r, v^{(l-1)} \rangle$. When $r_l \leq 0$, the iteration will be terminated.
- (ii) Secondly, the nonnegative least squares (NNLS) will be served as a tool to approximate the coefficients of the selected atoms:

$$\begin{aligned} x^{(l)} = \arg \min_x \quad & \left\| s - \sum_{h=1}^l d_{r_h} x_{r_h} \right\|_2 \\ \text{s.t.} \quad & x_{r_h} \geq 0. \end{aligned} \quad (3)$$

- (iii) Finally, the new residual $v^{(l)} = s - Dx^{(l)}$ will be computed and the corresponding iteration number l will be incremented by 1.

However, the selection process and the NNLS are divided into two relatively independent stages. Accordingly, we

should take into account a more efficient algorithm which aggregates these two steps. Inspired by analysis of comparison of OMP based on matrix decomposition [31], we need to choose QR factorization, which provides the largest reduction of computational complexity when problem size increases and little numerical error will be accumulated in the inner products or the solution. Therefore, we can address the issue by computationally efficient decomposition fashion. In fact, OMP attempts to find the orthogonal projection at each iteration as follows:

$$\hat{x}_p := \arg \min_{x_p} \|s - D_p x_p\|_2, \quad (4)$$

where D_p and x_p are the subdictionary and coefficient vector, respectively, restricted to the support p . D_p^\dagger denotes Moore-Penrose pseudoinverse of D_p and $\hat{x}_p = D_p^\dagger s$.

Let $v_t := s - D x_t$ be the t th signal residual. At iteration t , inverting a $t \times t$ matrix for calculation has a complexity of $O(t^3)$ which is a heavy computing burden. Thus, QR factorization is applied to incorporate a matrix decomposition of the selected subdictionary.

The dictionary D which chooses t atoms can be factorized following $D_t = Q_t R_t$. The columns of D_t are grouped based on the iteration number t and d_w signifies the w , the selected atom. Actually, we can readily solve (4) which is replaced by another problem $\hat{z}_t := \arg \min_{z_t} \|s - Q_t z_t\|_2$ because the column span of Q_t and D_t is the same. Thus, $\hat{z}_t = Q_t^T y$ and Q_t is orthonormal; we can quickly find the solution by $\hat{x}_t = R_t^{-1} \hat{z}_t$. Therefore, the efficiency of the method will be heavily dependent on the calculation speed of Q , R , and R^{-1} .

According to Gram-Schmidt process which is a method for orthonormalising a set of vectors in an inner product space, we only need to keep Gram-Schmidt process to seek the last column of Q_{t+1} after the first t terms of D_{t+1} have been decomposed. Thus, $Q_{t+1} = [Q_t \ q_{t+1}]$. In order to find Q_{t+1} , firstly we need to find the orthogonal element to the span of Q_t and then normalize the corresponding orthogonal ones as follows:

$$q_{t+1} = \frac{g_{t+1}}{\|g_{t+1}\|_2}, \quad (5)$$

where $g_{t+1} = (I - Q_t Q_t^T) d_{t+1}$.

Similarly, R_{t+1} and R_{t+1}^{-1} can be updated, respectively, as follows:

$$R_{t+1} = \begin{bmatrix} R_t & \alpha \\ 0 & \beta \end{bmatrix} \quad (6)$$

$$R_{t+1}^{-1} = \begin{bmatrix} R_t^{-1} & -\frac{\alpha}{\beta} R_t^{-1} \\ 0 & \frac{1}{\beta} \end{bmatrix},$$

where $\alpha = Q_t^T d_{t+1}$ and $\beta = \|g_{t+1}\|_2$. While fast OMP is beneficial to QR factorization, this method may still be used to choose negative elements in x_t . Accordingly, we need to develop fast OMP with nonnegativity constraints.

As stated above in NOMP, the atom will be selected due to the highest positive correlation with residual. At iteration t , the approximation of s can be computed as follows:

$$\sum_{w=1}^t x_w d_w = \sum_{w=1}^t z_w q_w. \quad (7)$$

Then, according to (7), in the $(t+1)$ th iteration, we see

$$\sum_{w=1}^{t+1} z_w q_w = \sum_{w=1}^t z_w q_w + z_{t+1} q_{t+1} = \sum_{w=1}^t x_w d_w + z_{t+1} q_{t+1}. \quad (8)$$

For some unique γ_θ , $q_{t+1} = \sum_{\theta=1}^{t+1} \gamma_\theta d_\theta$. Thus, in the $(t+1)$ th iteration, we see

$$\begin{aligned} \sum_{w=1}^{t+1} z_w q_w &= \sum_{w=1}^t x_\theta d_\theta + \sum_{\theta=1}^{t+1} z_{t+1} \gamma_\theta d_\theta \\ &= \sum_{w=1}^t (x_w + z_{t+1} \gamma_w) d_w + z_{t+1} \gamma_{t+1} d_{t+1}. \end{aligned} \quad (9)$$

According to (9), as $z_{t+1} \gamma_{t+1}$ keeps positive, we can assure that all the x_w are nonnegative when z_w meets such condition as follows:

$$z_{t+1} \leq z^c := \begin{cases} \min_{w, \gamma_w < 0} \left| \frac{x_w}{\gamma_w} \right| & \exists w, \gamma_w < 0 \\ +\infty & \text{Otherwise.} \end{cases} \quad (10)$$

Then, we see $x_w + z_{t+1} \gamma_w \geq 0$.

Next, if the z_{t+1} of atom has the largest value or is shrunk by (10), the corresponding atom will be selected. But if the atom having the highest positive correlation does not comply with (10), the most possible solution should be recorded. The decisive criterion of the solution can be listed as follows:

| | | |
|------------------------------------|---|------|
| $z < 0$ | Terminate | |
| $z^{mp} < z^c < z$ | $z^{mp} \leftarrow z^c$, $b^{mp} \leftarrow b$, Terminate | |
| $z^c < z \leq z^{mp}$ | $z_{t+1} \leftarrow z^{mp}$, $b \leftarrow b^{mp}$, Terminate | (11) |
| $z^c \leq z^{mp} < z$ | $b \leftarrow b + 1$ | |
| $0 < z \leq z^c$, $z \leq z^{mp}$ | $z_{t+1} \leftarrow z^{mp}$, $b \leftarrow b^{mp}$, Terminate | |
| $0 < z \leq z^c$, $z > z^{mp}$ | $z_{t+1} \leftarrow z$, Terminate, | |

where z^{mp} signifies the most possible solution and z is the current possible one in the t th iteration for an internal loop. We can define $(\delta, \sigma) = \text{sort}_\downarrow(D^T v_t)$ where sort_\downarrow denotes the sorting operator in a descent order. The initialization of z can be described as $\delta(b)$, $b = 1$. Then, $\sigma(b)$ will be added to support p and update Q_t and R_t^{-1} after inner-loop termination. Therefore, the whole process of FNOMP can be summarized as follows in Algorithm 1.

As shown in Algorithm 1, more details about the difference between NOMP and FNOMP can be elaborated from two aspects. Firstly, although both algorithms are composed of two loops, that is, internal and external loops, the decision and update steps based on (11) make a difference to FNOMP which is implemented in the internal loop.

FNOMP

```

Initialization:  $v_0 = s, t = 0, p = z_0 = \emptyset$ 
While  $t < J$  &  $\max(D^T v_t) > 0$  do
  Let  $(\delta, \sigma) \leftarrow \text{sort}_1(D^T v_t)$ 
  Let  $b \leftarrow 1$ 
  Let  $b^{\text{mp}} \leftarrow 1$ 
  Let  $z^{\text{mp}} \leftarrow 0$ 
  While  $\sim \text{Terminate}$  &  $b < K$  do
     $z^c$  from (10)
    Let  $z \leftarrow q_{\sigma(b)}^T v_t : q_{\sigma(b)} = g / \|g\|_2$ 
     $g = (I - Q_t Q_t^T) d_{\sigma(b)}$ 
    Update based on (11)
  End while
  Let  $p \leftarrow p \cup \sigma(b)$ 
  Update  $Q$  and  $R^{-1}$ 
  Let  $z_{t+1} \leftarrow [z_t; z_{t+1}]$ 
  Let  $v_{t+1} \leftarrow v_t - z_{t+1} q_{t+1}$ 
  Let  $t \leftarrow t + 1$ 
End while
Output:  $x|_s \leftarrow R^{-1} z_J$ 

```

ALGORITHM 1: The pipeline of fast OMP with nonnegativity constraints.

Comparatively, NOMP needs nonnegative least squares to optimize the coefficients of the selected atoms, whereas both algorithms terminate when J is satisfied and the highest positive correlation with residual is less than or equal to zero in the external loop. Secondly, an analysis of time complexity incorporates difference between these two algorithms. The total computational cost of FNOMP mainly includes two parts which are $O(L(3M(J+1) + J^2))$ for internal loop and $O(JK \log(B))$ for sorting B largest coefficients, respectively, where J is the sparsity level, L indicates the iteration number of inner loop, M is the dimensionality of dictionary, and K signifies the number of atoms. Compared with NOMP, the total computational cost of NOMP is $O(PMJ^3)$, where P is the iteration number of inner loop.

4. Experiments and Analysis

In this section, we apply FNOMP based method on three widely used datasets, that is, Oxford Flowers, UIUC-Sports, and Caltech101. Compared with standard NOMP with different dictionary sizes, computational costs of FNOMP will be shown in the first part. Without traditional preprocessing steps, the algorithm based on FNOMP will be used for training 2-layer deep models and compared with several state-of-the-art methods in terms of classification accuracy in the second part. The configurations of our PC are Intel Core i5 quad core CPU and frequency is 3.1 GHZ, 16 GB RAM, Windows 7 64-bit operating system. All codes are written in Matlab.

4.1. Comparison of Computation Costs. In this part, we run the efficient NOMP encoder on 6×6 patch size over dense grid with step size of one pixel and the corresponding dimension

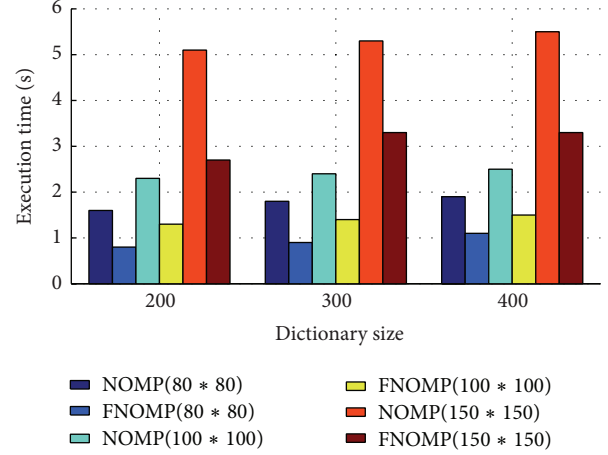


FIGURE 3: Comparison of the computation costs with different dictionary size.

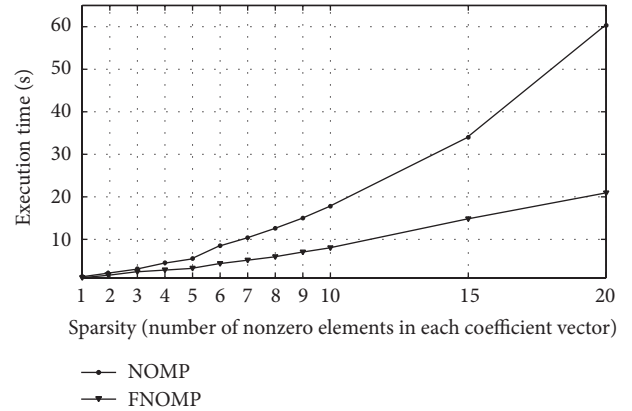


FIGURE 4: Comparison of the computation costs of the nonnegative OMP.

of atoms will be fixed at 108. The size of overcomplete dictionary will be increasing from 200 to 400 and the sparsity level is set to be at 5 in the first experiment. To study the computation time for encoding in practice, we will resize the test image to various ratios which are no larger than 80×80 , 100×100 , and 150×150 , respectively. Figure 3 illustrates that the running time of NOMP of three sets is consistently longer than that of FNOMP algorithm. On average, the computation cost of this novel method decreases by 42% which implies it can be applied to the full-size datasets using medium size of images.

In the second experiment, we will keep the sparsity level ranging from 1 to 20 and the size of dictionary will be fixed at 400, while the image size is fixed to be 150×150 . A comparison of computation costs of NOMP and FNOMP is shown in Figure 4. As the sparsity level increases, the execution time of standard NOMP apparently rises at a faster rate than that of FNOMP. In particular, the computation time will mount by more than 50% when sparsity is greater than 8 and will shoot up to 60 seconds when 20 nonzero elements are in coefficient vectors.

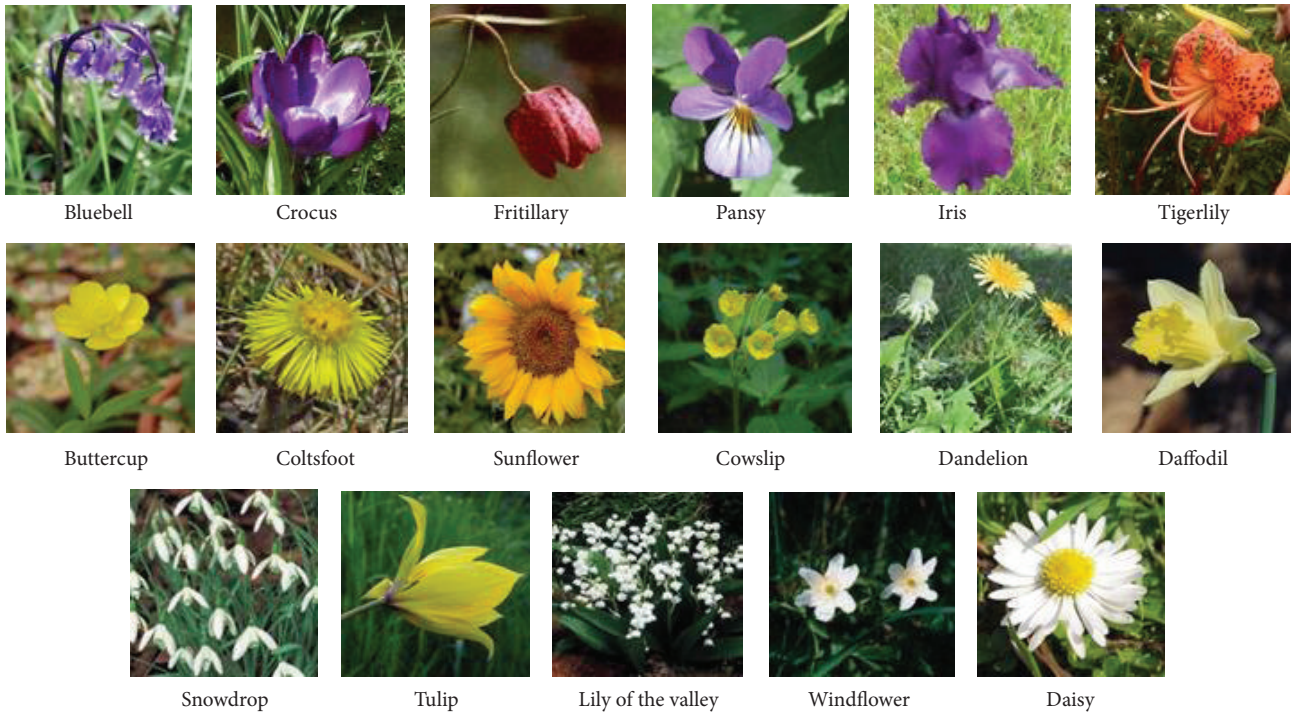


FIGURE 5: Sample images for the Oxford Flowers dataset.

4.2. Comparison of Classification Accuracy

4.2.1. Oxford Flowers Categorization. The Oxford Flowers dataset contains 1360 images with 17 different categories of flowers and each class has 80 images. The issue of similarity of two different classes is challenging and the intraclass variation is sometimes greater than that of the interclass between two species. According to [14], we follow the standard experimental settings for evaluation, that is, 60 random images are employed for training. Specifically, the receptive field size for max pooling is set to be 4 and 4×4 patch size is exploited for the second layer. Besides, the dictionary size is fixed at 400 and 1600 for the first and second layers, respectively. All images are kept as RGB type and resized to be no larger than 100×100 and 150×150 . We obtain average classification accuracy over 10 trials. As shown in Table 1, the classification accuracy of FNOMP based deep learning method is far above the HSSL which leverages a hierarchical model comprised of sparse coding, saliency pooling, and local grouping. As a typical one, Ito's method called color-CoHOG and CoHD, respectively, developing heterogeneous features based on cooccurrence is outperformed by FNOMP based approach. More importantly, image size has a great influence on final accuracy according to the results. Figure 5 shows some examples of this dataset.

4.2.2. UIUC-Sports Categorization. UIUC-Sports can be regarded as a statistic event category dataset which consists of 8 sport categories, for example, bocce, polo, rock climbing, and snowboarding. The total number of images is 1579 and 137~250 in each class. This dataset is quite challenging due to

TABLE 1: Classification accuracy (%) comparison on Oxford Flowers dataset.

| Algorithms | Classification accuracy |
|----------------------------|-------------------------|
| HSSL [14] | 76.2 ± 3.8 |
| Color-CoHOG [23] | 78.89 ± 1.19 |
| CoHD [23] | 84.24 ± 1.07 |
| FNOMP (100×100) | 83.1 |
| FNOMP (150×150) | 85.7 |

variations of poses and sizes across each category with cluttered backgrounds. According to the common experimental setting, we choose 70 images for training and 60 for testing at random per category. Figure 6 gives example images from classes of UIUC-Sports.

As is mentioned above, the settings of the experiment are the same as the previous one. The results from Table 2 indicate that FNOMP based method significantly outperforms the object bank (OB) and SIFT-based single layer sparse coding (SIFT + SC), respectively. Meanwhile, FNOMP based scheme can achieve highly competitive performance compared with the algorithm only using nonnegative sparse coding and spatial pyramid matching (Sc + SPM), adapted Gaussian models (AGM), and soft-assignment coding (SAC) approach. Similarly, it is found that larger image size can enhance performance by a large margin.

4.2.3. Caltech101 Categorization. This is a challenging dataset for object recognition task, comprising 9144 images in 102 classes. The number of images per category varies from 31 to



FIGURE 6: Sample images for the UIUC-Sports dataset.

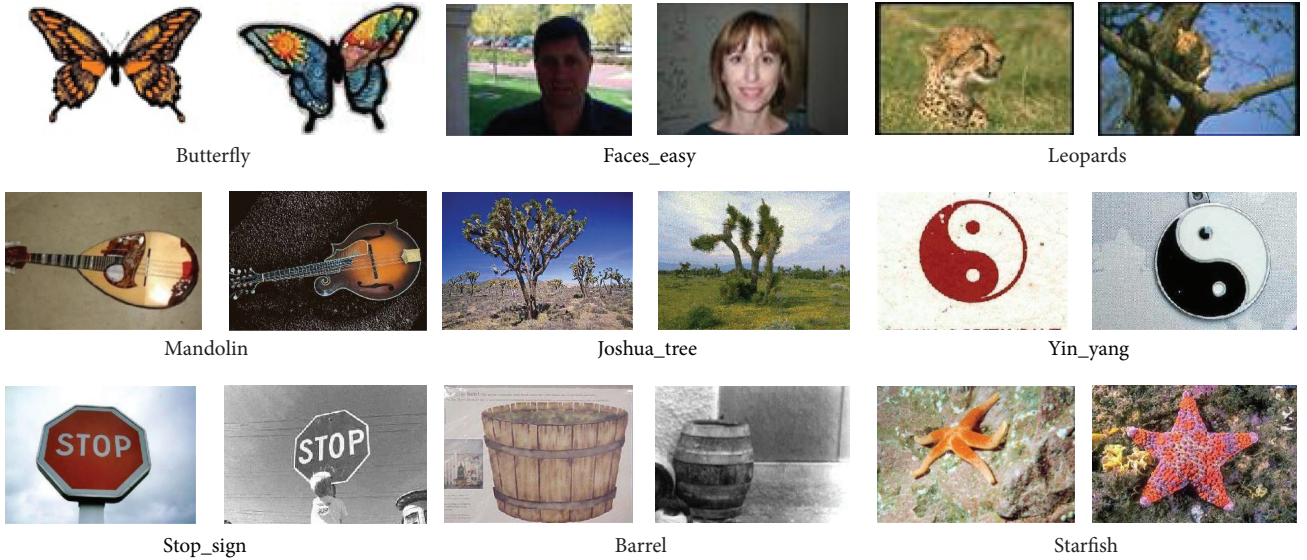


FIGURE 7: Sample images for the Caltech101 dataset.

TABLE 2: Classification accuracy (%) comparison on UIUC-Sports dataset.

| Algorithms | Classification accuracy |
|-------------------|-------------------------|
| OB [24] | 76.3 |
| SIFT+SC [21] | 82.7 |
| Sc+SPM [25] | 83.77 ± 0.97 |
| AGM [26] | 84.4 |
| SAC [27] | 84.56 ± 1.5 |
| FNOMP (100 × 100) | 82.8 |
| FNOMP (150 × 150) | 84.9 |

800. In addition to the background class, the remaining classes are composed of vehicles, flowers, animals, and so forth. Some sample images of Caltech101 are shown in Figure 7. Following the common experiment setup for Caltech101, we train on 30 images and test the rest. In the same way, we repeat the experiments 10 times with other experimental settings

identical to the previous one. As can be seen from Table 3, the performance of FNOMP based algorithm is marginally better than ScSPM and LLC which are both SIFT-based algorithms. As to other hierarchical models, FNOMP based pattern outperforms deconvolutional networks (DN) by about 9% and deconvolutional networks with both nonnegative sparsity and selectivity (DNNS) by about 3%. However, DNNS employs the combined features of 1st and 4th layer from the model trained with both properties which have more complex deep architecture. Typically, hierarchical sparse coding (HSC) jointly learns two codebooks which are more complicated, while the algorithm combined with FNOMP as encoder can outperform HSC by 1.9%. Interestingly, the performance of low-rank nonnegative sparse coding (LR-Sc + SPM) is extremely close to the FNOMP based method, which adopts different strategies of nonnegativity constraints.

Finally, we compare the performance of FNOMP as encoder with OMP in deep networks with the same dictionary training scheme. Specifically, the patch size over dense

TABLE 3: Classification accuracy (%) comparison on Caltech101 dataset.

| Algorithms | Classification accuracy |
|--------------------------|-------------------------|
| DN [18] | 66.9 \pm 1.1 |
| ScSPM [12] | 73.20 |
| DNNS [28] | 73.3 \pm 0.4 |
| LLC [13] | 73.44 |
| HSC [29] | 74.0 |
| LR-Sc+SPM [25] | 75.68 \pm 0.89 |
| FNOMP (100 \times 100) | 73.6 |
| FNOMP (150 \times 150) | 75.9 |

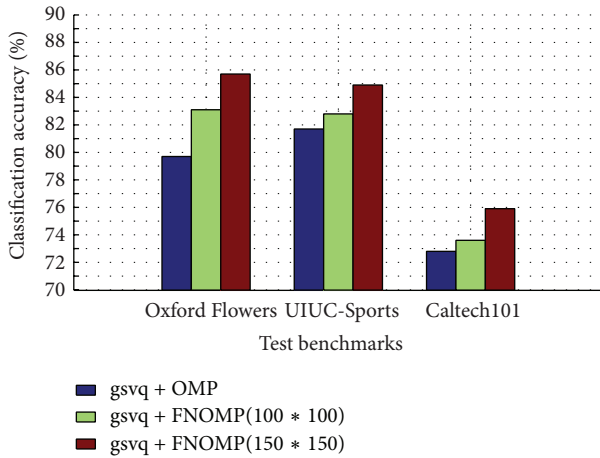


FIGURE 8: Comparison of classification accuracy using gsvq for dictionary training.

grid with step size of one pixel is still adopted and the dimension of atoms remains unchanged. The dictionary size is set to be at 400 and 1600 for the first and second layers, respectively. We can see from Figure 8, in this trial, FNOMP shows more powerful performance than OMP on three different benchmarks. Using gsvq for dictionary training, we find FNOMP based algorithm considerably increases the classification accuracy by around 6%, 3.2%, and 3.1%, respectively.

5. Conclusion

In this paper, we have studied fast nonnegative OMP as an encoder in deep networks to obtain meaningful image representations. Impressive research results are obtained with FNOMP in terms of both computational efficiency and classification accuracy. It is found that FNOMP performs significantly faster than the standard NOMP with medium size of images in practice. In particular, the computation cost of NOMP becomes 2 times or more than that of FNOMP as the sparsity level increases. In addition, we have conducted further studies on three widely used benchmarks for image classification tasks. The experimental results show that FNOMP performs better than SIFT-based single layer

sparse coding, hierarchical feature learning, and other state-of-the-art methods on Oxford Flowers, UIUC-Sports, and Caltech101 datasets. Furthermore, with the same dictionary training approach, we find that FNOMP is superior to OMP in terms of classification accuracy. More importantly, it is also found that image size has a great influence on classification accuracy according to the results.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Research Fund for the Doctoral Program of Higher Education of China (no. 20120032110034) and the National Program on Key Basic Research Project (no. 2014CB340403).

References

- [1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [2] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry, and Y. Ma, "Fast ℓ_1 -minimization algorithms for robust face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3234–3246, 2013.
- [3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [4] H. Lee, R. Raina, A. Teichman, and A. Y. Ng, "Exponential family sparse coding with applications to self-taught learning," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, pp. 1113–1119, July 2009.
- [5] T. H. Lin and H. T. Kung, "Stable and efficient representation learning with nonnegativity constraints," in *Proceedings of the 31st International Conference on Machine Learning (ICML '14)*, pp. 1323–1331, Beijing, China, June 2014.
- [6] D. Chen and R. J. Plemmons, "Nonnegativity constraints in numerical analysis," in *Proceedings of the Symposium on the Birth of Numerical Analysis*, pp. 109–140, Leuven, Belgium, 2009.
- [7] P. O. Hoyer, "Modeling receptive fields with non-negative sparse coding," *Journal of Neurocomputing*, vol. 52–54, pp. 547–552, 2003.
- [8] T. Guthier, V. Willert, A. Schnall, K. Kreuter, and J. Eggert, "Non-negative sparse coding for motion extraction," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '13)*, pp. 1–8, Dallas, Tex, USA, August 2013.
- [9] W. Zheng and Y. Qian, "Non-negative sparse semantic coding for text categorization," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR'12)*, pp. 409–412, Tsukuba, Japan, November 2012.
- [10] S. M. Amiri, P. Nasiopoulos, and V. C. M. Leung, "Non-negative sparse coding for human action recognition," in *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 1421–1424, Orlando, Fla, USA, October 2012.

- [11] Y. Chen, Y. Zhao, and A. Cai, "Recognizing human actions based on sparse coding with non-negative and locality constraints," in *Proceedings of the IEEE International Conference on Visual Communications and Image Processing (VCIP '13)*, pp. 1–6, Kuching, Malaysia, November 2013.
- [12] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1794–1801, Miami, Fla, USA, June 2009.
- [13] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 3360–3367, San Francisco, Calif, USA, June 2010.
- [14] J. Yang and M. Yang, "Learning hierarchical image representation with sparsity, saliency and locality," in *Proceedings of the 22nd British Machine Vision Conference (BMVC '11)*, pp. 19.1–19.11, Dundee, UK, August 2011.
- [15] A. Coates and A. Y. Ng, "Selecting receptive fields in deep networks," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '11)*, vol. 24, pp. 2528–2536, Granada, Spain, September 2011.
- [16] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Journal of Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [17] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th International Conference on Machine Learning (ICML '09)*, pp. 609–616, Montreal, Canada, June 2009.
- [18] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2528–2535, San Francisco, Calif, USA, June 2010.
- [19] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8595–8598, Vancouver, Canada, May 2013.
- [20] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '12)*, vol. 25, pp. 1106–1114, Lake Tahoe, Nev, USA, December 2012.
- [21] L. Bo, X. Ren, and D. Fox, "Hierarchical matching pursuit for image classification: architecture and fast algorithms," in *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS '11)*, pp. 1–9, Granada, Spain, December 2011.
- [22] L. Bo, X. Ren, and D. Fox, "Multipath sparse coding using hierarchical matching pursuit," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 660–667, Portland, Ore, USA, June 2013.
- [23] S. Ito and S. Kubota, "Object classification using heterogeneous co-occurrence features," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, pp. 209–222, Heraklion, Greece, September 2010.
- [24] L.-J. Li, H. Su, E. P. Xing, and L. Fei-Fei, "Object Bank: a high-level image representation for scene classification and semantic feature sparsification," in *Proceedings of the 24th Annual Conference on Neural Information Processing Systems 2010 (NIPS' 10)*, vol. 23, p. 7, Vancouver, Canada, December 2010.
- [25] C. Zhang, J. Liu, Q. Tian, C. Xu, H. Lu, and S. Ma, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1673–1680, Providence, RI, USA, June 2011.
- [26] M. Dixit, N. Rasiwasia, and N. Vasconcelos, "Adapted gaussian models for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 937–943, Providence, RI, USA, June 2011.
- [27] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2486–2493, IEEE, Barcelona, Spain, November 2011.
- [28] B. Liu, J. Liu, X. Bai, and H. Lu, "Regularized hierarchical feature learning with non-negative sparsity and selectivity for image classification," in *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR '14)*, pp. 4293–4298, IEEE, Stockholm, Sweden, August 2014.
- [29] K. Yu, Y. Lin, and J. Lafferty, "Learning image representations from the pixel level via hierarchical sparse coding," in *2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pp. 1713–1720, usa, June 2011.
- [30] M. Yaghoobi, D. Wu, and M. E. Davies, "Fast non-negative orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1229–1233, 2015.
- [31] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO '12)*, pp. 220–224, Bucharest, Romania, August 2012.

