Multipath Sparse Coding Using Hierarchical Matching Pursuit

Yufeng Jiang

1. Related Work

In the past few years, a growing amount of research on visual recognition has focused on learning rich features using unsupervised and supervised hierachical architectures.

Deep Networks: Deep belief nets [2] learn a hierarchy of features, layer by layer, using the unsupervised restricted Boltzmann machine. The learned weights are then further adjusted to the current task using supervised information. To make deep belief nets applicable to full-size images, convolutional deep belief nets [4] use a small receptive field and share the weights between the hidden and visible layers among all locatios in an image. Deconvolutional networks [7] convolutionally decompose images in an unsupervised way under a sparsity constraint. Deep convolutional neural networks [3] won the ImageNet Large Scale Visual Recognition Challenge 2012 and demonstrated their potential for training on large, labeled datasets.

Sparse Coding: For many years, sparse coding [6] has been a popular tool for modeling images. Sparse coding on top of raw pathes or SIFT features has achieved state-of-the-art performance on face recognition, texture segmentation [5]. Very recently, multi-layer sparse coding networks including hierarchial sparse coding and hierarchical matching pursuit have been proposed for building multiple level features from raw sensor data. Such networks learn codebooks at each layer in an unsupervised way such that image patches or pooled features can be represented by a sparse, linear combination of codebook entries.

2. Multipath Sparse Coding

Authors propose MI-KSVD to solve the above optimization by adapting the well-known KSVD algorithm. MI-KSVD decomposes the above optimization into two subproblems, **Encoding** and **Codebook Update**, and solves them in an alternating manner. During each iteration, the current codebook D is used to encode the data Y by computing the sparse code matrix X. Then, the codewords of the codebook are updated one at a time, resulting in a new codebook.

Encoding: Given a codebook D, the encoding problem is to find the sparse code x of y, leading to the following optimization [1]

$$\min_{x} ||y - \mathbf{D}x||^2 \qquad \textit{s.t.} \ \ ||x||_0 \leq \mathbf{K} \tag{1}$$

Here, orthogonal matching pursuit (OMP) is used to compute the sparse code x due to its efficiency and effectiveness. OMP selects the codeword best correlated with the current residual at each iteration, which is the reconstruction error remaining after the codewords chosen thus far are subtracted. At the first iteration, this residual is exactly the observation y. Once a new codeword is selected, the observation is orthogonally projected onto the span of all the previously selected codewords and the residual is recomoputed. The procedure is repeated until the desired sparsity level K is rearched.

Codebook Update: Given the sparse code matrix X, the codewords d_m are optimized sequentially. In the m-th step, the m-th codeword and its sparse codes can be computed by minimizing the residual matrix and the mutual coherence corresponding to that codeword [1]

$$\min_{d_m} ||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2 + \lambda \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} |d_i^\top d_j| \tag{2}$$
$$\textit{s.t.} \ ||d_m||_2 = 1$$

Removing the constant terms, the above optimization problem can be simplified to [1]

$$\min_{d_m} \{\bar{x}_m^\top \bar{x}_m d_m^\top d_m - 2\mathbf{R}_m \bar{x}_m + \lambda \sum_{j=1, j\neq m}^{M} |d_j^\top d_m|\} \tag{3}$$
$$\textit{s.t.} \ ||d_m||_2 = 1$$

where $\bar{x}_i^\top$ are the rows of $\mathbf{X}$, and $\mathbf{R}_m = \mathbf{Y} - \sum_{i\neq m} d_i \bar{x}_i^\top$ is the residual matrix for the m-th codeword. This matrix contains the differences between the observations and their approximations using all other codewords and their sparse codes. To avoid introducint new non-zero entries in the sparse code matrix $\mathbf{X}$, the update process only considers observations that use the m-th codeword. They solve Eq. 2 by standard gradient descent with initialization $d_m^0 = \frac{\mathbf{R}_m \bar{x}_m}{||\mathbf{R}_m \bar{x}_m||_2}$, which is optimal when ignoring the mutual incoherence penalty.

	Test Accuracy	AMC	Training Time
KSVD	71.9	0.153	6126.1s
MI-KSVD	73.1	0.118	6713.8s

Table 1. MI-KSVD and KSVD on Caltech-101. Both of them are stopped after 100 iterations that are sufficient for convergence.

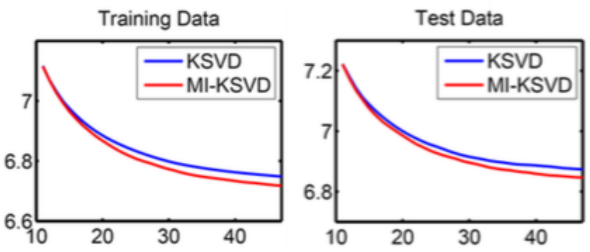

Figure 1. Mean square error as a function of iterations. Training and test data consists of 1,000,000 36×36 image patches and 100,000 36×36 image patches sampled from images, respectively.

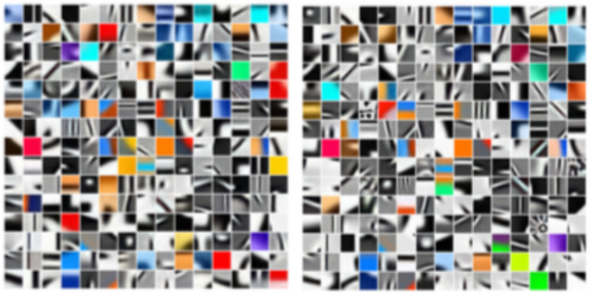

Figure 2. Learned codebooks by KSVD $(left)$ and MI-KSVD $(right)$ on Caltech-101. 225 codewords randomly selected from 1000 codewords are shown.

In practice, authors find that MI-KSVD converges to good codebooks for a wide range of initializations. Fig. 1 used at [1].compares the reconstruction error of the proposed MI-KSVD and KSVD on both training data and test data. This is remarkable since the additional penalty usually increase the reconstruction error. The codebook learned by MI-KSVD has average mutual coherence (AMC) $\frac{1}{M(M-1)}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M}|d_i^\top d_j| = 0.118$, substantially smaller than 0.153 yielded by KSVD.

2.1. MI-KSVD

They compare KSVD and MI-KSVD for a one-layer HMP network with image patches of size 36×36 on the Caltech-101 dataset. Authors choose a one-layer HMP due to the convenience of showing the learned codebooks. They learn codebooks of size 1000 with sparsity level 5 on 1,000,000 sampled 36×36 raw patches. The tradeoff parameter λ is chosen by performing five-fold cross validation on the training set.

They visualize the codebooks learned by KSVD and MI-KSVD in Fig. 2 shown at [1].. First of all, the learned dictionaries have very rich appearahces and include uniform colors of red, green and blue, transition codewords between different colors, gray and color edges, double gray and color edges, and so on. They compare KSVD and MI-KSVD in Tab. 1 shown at [1]. in terms of test accuracy, training time and average mutual coherence (AMC). As can been seen, MI-KSVD leads to higher test accuracy and lower average mutual coherence than KSVD, with comparable training time.

References

[1] L. Bo, X. Ren, and D. Fox. Multipath sparse coding using hierarchical matching pursuit. In *CVPR*, 2013. 1, 2

[2] G. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006. 1

[3] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012. 1

[4] H. Lee, R. Grosse, R. Ranganath, and A. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *ICML*, 2009. 1

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *CVPR*, 2008. 1

[6] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996. 1

[7] M. Zeiler, G. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *ICCV*, 2011. 1