Multipath Sparse Coding Using Hierarchical Matching Pursuit

Yufeng Jiang

Abstract

Complex real-world signals, such as imaes, contain discriminative structures that differ in many aspects including scale, invariance and data channel. While progress in deep learning shows the importance of learning features through multiple layers, it is equally important to learn features through multiple paths. Authours propose Multipath Hierarchical Matching Pursuit (M-HMP), a novel feature learning architecture that combines a collection of hierarchical sparse features for image classification to capture multiple aspects of discriminative structures. The blocks builded by authors are MI-KSVD, a codebook learning algorithm that balances the reconstruction error and the mutual incoherence of the codebook, and bath orthogonal matching pursuit (OMP). They apple them recursively at varying layers and scales.

1. Introduction

Images are high dimensional signals that change dramatically under varying scales, viewpoints, lighting conditions and scene layouts. How to extract features that are robust to these changes is an important question in computer vision. Traditionally, people rely on designed features such as SIFT. SIFT can be understood and generalized as a way to go from pixels to patch descriptors [1], but it is a challanging task to design good features because it requires deep domain knowledge and adapts to new settings.

One crucial problem that is often overlooked in image feature learning is the multi-facet nature of visual structures: discriminative structures may appear at varying scales with varying amouts of spatial and appearance invariance. In this paper, authors propose Multipath Hierarchical Matching Pursuit (M-HMP), which builds on the single-path Hierarchical Matching Pursuit approach to learn and combine recursive sparse coding through many pathways on multiple bags of pathes of varying size. It alse can be learned by encoding each patch through multiple paths with a varying number of layers just as shown at Fig. 1cited at [2]. Their M-HMP approach is generic and can adapt to new tasks, new sensor data or new feature learning and coding algorithms.

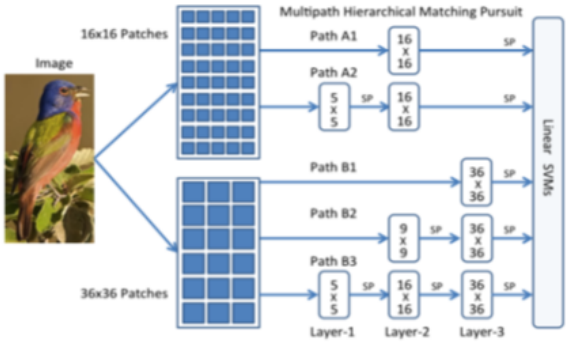

Figure 1. Architecture of multipath sparse coding. Image patches of different sizes (here, 16×16 and 36×36) are encoded via multiple layers of sparse coding. Each path corresponds to a specific patch size and number of layers (numbers inside boxes indicate patch size at the corresponding layer and path). Spatial pooling, indicated by SP, is performed between layers to generate the input features for the next layer. The final layer of each path encodes complete image patches and generates a feature vector for the whole image via another spatial pooling operation. Path features are then concatenated and used by a linear SVM for object recognition.

2. Multipath Sparse Coding

This section provides an overview of their Multipath Hierarchical Matching Pursuit (M-HMP) approach. They propose a novel codebook learning algorithm, MI-KSVD, to maintain mutual incoherence of the codebook and discuss how multi-layer sparse coding hierarchies for images can be built from scratch and how multipath sparse coding helps capture discriminative structures of varying characteristics.

2.1. Codebook Learning with Mutual Incoherence

The key idea of sparse coding is to represent data as sparse linear comnbinations of codewords selected from a codeboook [4]. The standard sparse coding approaches learrn the codebook $\mathbf{D} = [d_1, \ldots, d_m, \ldots, d_M] \in \mathbf{R}^{H \times M}$ and the associated sparse codes $\mathbf{X} = [x_1, \ldots, x_n, \ldots, x_N] \in \mathbf{R}^{M \times N}$ from a matrix $\mathbf{Y} = [y_1, \ldots, y_n, \ldots, y_N] \in \mathbf{R}^{H \times N}$ of observed data by minimizing the reconstruction error [2]

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 \quad (1)$$
$$s.t.\ \forall m,\ ||d_m||_2 = 1\ and\ \forall_n,\ ||x_n||_0 \leq \mathbf{K}$$

where $\mathbf{H}$,$\mathbf{M}$ and N are the dimensionality of the codewords, the size of the codebook and the number of training samples. Respectively, $||\cdot||_F$ denotes the Frobenius norm, the zero-norm $||\cdot||_0$ counts non-zero entries in the sparse codes x_n, and $\mathbf{K}$ is the sparsity level controlling the number of the non-zero entries.

When sparse coding is applied to object recognition, the data matrix $\mathbf{Y}$ consists of raw patches randomly sampled from images. In order ot balance the roles of different types of image patches, it is desirable to maintain large mutual incoherence during the codebook learning phase. On the other hand, theoretical results on sparse coding [3] have also indicated that it is much easier to recover the underlying sparse codes of data when mutual incoherence of the codebook is large. This motivaties us to balance the reconstruction error and the mutual incoherence of the codebook [2]

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{DX}||_F^2 + \lambda \sum_{i=1}^{\mathbf{M}} \sum_{j=1,j\neq i}^{\mathbf{M}} |d_i^\top d_j| \quad (2)$$
$$s.t.\ \forall m,\ ||d_m||_2 = 1\ and\ \forall n,\ ||x_n||_0 \leq \mathbf{K}$$

Here, the mutual coherence $\lambda \sum_{i=1}^{\mathbf{M}} \sum_{j=1,j\neq i}^{\mathbf{M}} |d_i^\top d_j|$ has been included in the objective function to encourage large mutual incoherence where $\lambda \geq 0$ is a tradeoff parameter.

References

[1] L. Bo, X. Ren, and D. Fox. Kernel descriptors for visual recognition. In *NIPS*, 2010. 1

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

[3] E. Candes and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007. 2

[4] 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