Pattern Recognition 75 (2018) 282–291

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

# Nonlinear dictionary learning with application to image classification

Junlin Hu*, Yap-Peng Tan

*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, Singapore*

ARTICLE INFO

*Article history:*
Received 11 August 2016
Revised 29 December 2016
Accepted 3 February 2017
Available online 4 February 2017

*Keywords:*
Nonlinear dictionary learning
Sparse coding
Neural network
Image classification

ABSTRACT

In this paper, we propose a new nonlinear dictionary learning (NDL) method and apply it to image classification. While a variety of dictionary learning algorithms have been proposed in recent years, most of them learn only a linear dictionary for feature learning and encoding, which cannot exploit the nonlinear relationship of image samples for feature extraction. Even though kernel-based dictionary learning methods can address this limitation, they still suffer from the scalability problem. Unlike existing dictionary learning methods, our NDL employs a feed-forward neural network to seek hierarchical feature projection matrices and dictionary simultaneously, so that the nonlinear structure of samples can be well exploited for feature learning and encoding. To better exploit the discriminative information, we extend the NDL into supervised NDL (SNDL) by learning a class-specific dictionary with the labels of training samples. Experimental results on four image datasets show the effectiveness of the proposed methods.



## 1. Introduction

In recent years, sparse representation has been widely studied in signal processing and machine learning [1,2], and it has also been successfully applied to various computer vision applications such as image denoising [3], face recognition [4–7], facial analysis [8–11], image classification [12–15], and visual tracking [16]. The basic assumption of sparse representation is that one signal can be well approximated by a linear combination of a small number of atoms (or basis) from an over-complete dictionary. Generally, the dictionary plays an important role in sparse modeling, and its quality can heavily affect the performance of sparse representation [17]. Therefore, dictionary learning is a basic element to sparse representation. Instead of using a predefined dictionary (e.g., various wavelets), recent advances in dictionary learning have shown that learning a desirable dictionary from the training data itself can usually yield good results for numerous image and video analysis tasks [3,12,13,18–24].

There have been a number of dictionary learning methods proposed in recent years [2,19–21,25]. These dictionary learning methods can be roughly divided into two categories: *unsupervised* and *supervised*. Unsupervised dictionary learning methods aim to learn an over-complete dictionary by minimizing the reconstruction error of a set of signals with sparsity constraints [13,20,26], which have shown good performance in some visual reconstruction and clustering tasks [3,22] (e.g., image denoising, image restoration, and video clustering). For the second category, supervised dictionary learning methods usually learn a compact and discriminative dictionary by exploiting the label or side information of the training data, which complements a discriminative term to the reconstruction error and optimizes the objective functions for different settings [18,19,21,27–30]. Methods in this category are usually applied to various visual recognition tasks [13,25,31]. To promote the discriminative capacity of the learned dictionary, there are several strategies that can be used for this discriminative term such as introducing a classifier on sparse coefficients [19,21,27], learning a structured dictionary using the incoherence constraint between class-specific dictionaries [22,32] and the Fisher discrimination criterion [23] on sparse coefficients. Generally, the dictionary learned by supervised methods can achieve good performance for many visual applications [21,31].

Most existing dictionary learning algorithms, however, usually learn a linear dictionary for feature learning and encoding in the original space such that they cannot capture the nonlinear structure of data points. To address this nonlinearity problem, the kernel trick is often adopted to map the data points into a high-dimensional space and then learn a dictionary in this transformed space using existing dictionary learning approaches [25,33]. However, the kernel-based methods cannot explicitly obtain the nonlinear mapping function and often suffer from the scalability problem. Different from these methods, in this paper, we propose a nonlinear dictionary learning (NDL) method to seek hierarchical nonlinear projection matrices and dictionary simultaneously via a feed-forward neural network, so that the nonlinear structure of samples can be well exploited. To better exploit the discriminative

* Corresponding author.
*E-mail addresses:* jhu007@e.ntu.edu.sg (J. Hu), eyptan@ntu.edu.sg (Y.-P. Tan).

http://dx.doi.org/10.1016/j.patcog.2017.02.009

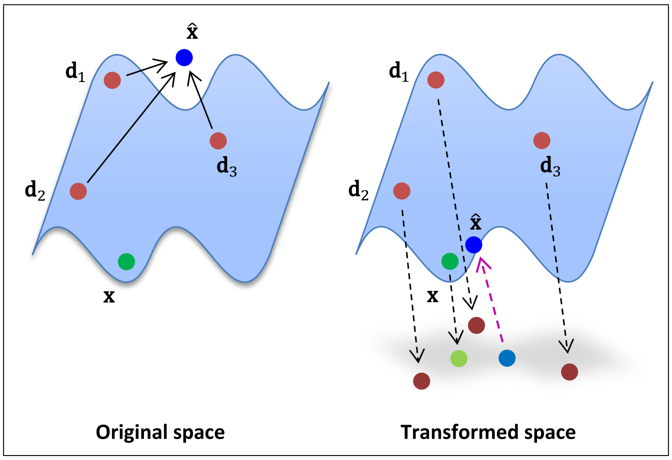


**Fig. 1.** A toy example illustrates how NDL captures the nonlinearity of data points. In the **original** space, $\hat{\mathbf{x}}$ (blue point) is the linear approximation of the data point $\mathbf{x}$ (green) over three atoms $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_3$ of a dictionary (red points), and it is not on the manifold $\mathcal{M}$. The NDL method maps both the $\mathbf{x}$ and atoms to a new space via a feed-forward neural network and then computes the linear combination on this **transformed** space so that the representation $\hat{\mathbf{x}}$ of the $\mathbf{x}$ is on the manifold $\mathcal{M}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

information, we extend the NDL into supervised NDL (SNDL) by learning a class-specific dictionary with the label of training samples. Fig. 1 shows a toy example illustrating how the proposed NDL captures the nonlinearity of data points.

## 2. Related work

### 2.1. Deep learning

Deep learning has been a popular research topic in the communities of machine learning and computer vision due to its excellent performance in various visual tasks such as face recognition [34–42], image classification [43,44], action recognition [45] and visual tracking [46]. A variety of deep learning methods have been proposed to directly learn rich hierarchical feature representations from raw data [47]. Representative deep learning models are deep belief networks [48], deep convolutional neural networks (CNN) [44], deep stacked auto-encoder [45], and deep metric network (or siamese network) [49–53]. However, to our best knowledge, little attempt has been made on coordinating deep learning and dictionary leaning to utilize the merits of deep learning. In this work, we introduce a nonlinear dictionary learning approach to learn several hierarchical nonlinear transformations and the desired dictionary simultaneously by integrating deep learning and dictionary leaning into a unified framework.

### 2.2. Dictionary learning

Learning a desired dictionary from the training data for sparse representation has attracted considerable attention in the field of computer vision [2,21], and many advanced dictionary learning algorithms have been proposed in recent years. The K-SVD method [26] is a representative unsupervised dictionary learning approach, which learns an over-complete dictionary from natural images in an iterative fashion. Based on K-SVD, a discriminative term is usually added to the reconstruction error to obtain discriminative dictionaries [19,21,23,54–56]. For example, Mairal et al. [21,27] proposed to learn a shared dictionary for all classes and discriminative class models on sparse coefficients simultaneously. Jiang et al. [19] introduced the label consistent K-SVD algorithm to balance reconstructive and discriminative power of the learned dictionary. To capture the nonlinearity of data, kernel-based dictionary learning approaches [25,33] employ kernel trick to first map the data from the original space into another space and then utilize well-known dictionary learning methods in this space. More recently, deep sparse coding (DeepSC) [57] extends sparse coding to a multi-layer architecture, where it employed an approach called dimensionality reduction by learning an invariant mapping (DrLIM) [49] to learn a parametric mapping in the siamese network. Different from the DeepSC and hierarchical sparse coding [58] methods which treat the output of sparse coding as the input of the next layer, our NDL method employs a feedforward neural network to learn a nonlinear mapping and dictionary simultaneously, so that the nonlinear structure of data points can be well exploited.

## 3. Nonlinear dictionary learning

We briefly introduce some background of the dictionary learning methods, and then focus on formulating nonlinear dictionary learning method in an unsupervised manner.

### 3.1. Background

Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_K] \in \mathbb{R}^{r\times K}$ be a dictionary that contains $K$ atoms. In the classical sparse coding task, a signal $\mathbf{x} \in \mathbb{R}^r$ can be sparsely represented by a linear combination of a few atoms from the dictionary $\mathbf{D}$ as:

$$\mathbf{x} \approx \mathbf{D}\ \mathbf{a} = a_1\mathbf{d}_1 + a_2\mathbf{d}_2 + \cdots + a_K\mathbf{d}_K, \tag{1}$$

where $\mathbf{a} = [a_1, a_2, \cdots, a_K]^T \in \mathbb{R}^K$ is a sparse coefficient vector. The sparse coding with a $\ell_1$ regularization problem is usually solved to obtain an optimal sparse solution $\mathbf{a}$:

$$\min_{\mathbf{a}} \ \ \|\mathbf{x} - \mathbf{D}\ \mathbf{a}\|_2^2 + \lambda_1\|\mathbf{a}\|_1, \tag{2}$$

where $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_1$ denote $\ell_2$ and $\ell_1$ norms of vector $\mathbf{x}$ respectively, and $\lambda_1$ is a positive regularization parameter. The first term in (2) is the reconstruction error, and the second term is the sparsity penalty.

To learn a dictionary from the training set of $N$ samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{r\times N}$, dictionary learning (DL) algorithms [20,21] simply minimize the following empirical cost function over both a dictionary $\mathbf{D}$ and a sparse matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \in \mathbb{R}^{K\times N}$ as:

$$\begin{aligned} &\min_{\mathbf{D},\ \mathbf{A}} \ \frac{1}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i - \mathbf{D}\ \mathbf{a}_i\|_2^2 + \lambda_1\|\mathbf{a}_i\|_1\right) \\ &\text{s.t.} \ \ \ \|\mathbf{d}_i\|_2^2 \le 1, \ \ \forall\ i = 1, 2, \cdots, K, \end{aligned} \tag{3}$$

where vector $\mathbf{a}_i$ is the $i$th column of $\mathbf{A}$ and the sparse coefficient of sample $\mathbf{x}_i$ over $\mathbf{D}$. The constraint $\{\|\mathbf{d}_i\|_2^2 \le 1\}_{i=1}^{K}$ aims to prevent $\mathbf{D}$ from being arbitrarily large because it would cause very small values of the $\mathbf{A}$. The problem (3) can be solved by several works such as the feature-sign search algorithm [20] and gradient descent method.

### 3.2. Nonlinear dictionary learning

Unlike most existing dictionary learning methods which only learn a linear dictionary, our NDL employs a feed-forward neural network to exploit the nonlinearity of samples for dictionary learning. Fig. 2 illustrates the basic idea of the proposed NDL method.

Consider a feed-forward neural network with $2M$+1 layers, which includes two parts: encoding and decoding, and there are $r^{(m)}$ neurons in layer $m$, holding $r^{(2M-m)} = r^{(m)}$ for all $m = 0, 1, \cdots, 2M$. The representation $\mathbf{h}^{(m)}$ of an input $\mathbf{x} \in \mathbb{R}^{r^{(0)}}$ in the layer $m$ ($m \ge 1$) is represented as:

$$\mathbf{h}^{(m)} = \varphi(\mathbf{z}^{(m)}) \in \mathbb{R}^{r^{(m)}}, \tag{4}$$