Deep Non-Negative Matrix Factorization Architecture Based on Underlying Basis Images Learning

Yang Zhao⬤, Huiyang Wang⬤, and Jihong Pei⬤, *Member, IEEE*

Abstract—The non-negative matrix factorization (NMF) algorithm represents the original image as a linear combination of a set of basis images. This image representation method is in line with the idea of "parts constitute a whole" in human thinking. The existing deep NMF performs deep factorization on the coefficient matrix. In these methods, the basis images used to represent the original image is essentially obtained by factorizing the original images once. To extract features reflecting the deep localization characteristics of images, a novel deep NMF architecture based on underlying basis images learning is proposed for the first time. The architecture learns the underlying basis images by deep factorization on the basis images matrix. The deep factorization architecture proposed in this paper has strong interpretability. To implement this architecture, this paper proposes a deep non-negative basis matrix factorization algorithm to obtain the underlying basis images. Then, the objective function is established with an added regularization term, which directly constrains the basis images matrix to obtain the basis images with good local characteristics, and a regularized deep non-negative basis matrix factorization algorithm is proposed. The regularized deep nonlinear non-negative basis matrix factorization algorithm is also proposed to handle pattern recognition tasks with complex data. This paper also theoretically proves the convergence of the algorithm. Finally, the experimental results show that the deep NMF architecture based on the underlying basis images learning proposed in this paper can obtain better recognition performance than the other state-of-the-art methods.

Index Terms—Non-negative matrix factorization, underlying basis images, deep factorization architecture, face recognition

✦

1 INTRODUCTION

IN the field of artificial intelligence currently, the data often has a higher dimension. Direct analysis and processing of such data lower the work efficiency and the results are affected by redundant information to a certain extent. Therefore, researchers have conducted a lot of studies on the dimension reduction sample representation. The non-negative matrix factorization algorithm is one of the effective dimensionality-reduced sample representation methods because of its excellent performance in the low-dimensional non-negative representation of samples.

First proposed by Lee *et al.*, the NMF algorithm represented the original image sample as a combination of a set of basis images [1], [2]. The original samples in the NMF algorithm can be reconstructed by the set of basis images, and the reconstruction coefficient is a new feature of the original sample. This sample representation based on the combination of the basis vectors has a very intuitive semantic interpretation, reflecting the concept of "parts constitute a whole" in human thinking. The NMF algorithm factorizes

a non-negative matrix into the product of a non-negative basis images matrix and a non-negative coefficient matrix. The matrix elements in the factorization process of NMF algorithm satisfy the non-negative characteristics, which makes it widely used in many practical data analysis, including image analysis with non-negative pixel values, text analysis with non-negative word frequency, etc. [3], [4], [5], [6], [7], [8], [9]. In order to improve its performance, many researchers are studying the NMF algorithm.

Hoyer presented an NMF algorithm with sparse constraints to improve the local characteristics of the basis images after factorization [10]. A sparse metric based on the relationship between l_1 norm and l_2 norm was proposed in this algorithm to measure the sparsity of the vector after factorization. Pascual-Montano *et al.* presented the non-smooth non-negative matrix factorization in order to obtain a basis image with better sample expression ability and stronger local feature representation. The algorithm introduces a "smoothing" matrix to construct a new factorization form. In this form, the sparseness is represented explicitly [11]. From the point of geometric, Cai *et al.* believe that data is usually sampled from low-dimensional manifolds embedded in high-dimensional space, and the graphic regularized non-negative matrix factorization (GNMF) algorithm is proposed [12]. The algorithm adds the graph regular term to the objective function, constraining the geometric structure of the new feature after factorization. Ding *et al.* relax the constraints of factorizations to make the algorithm not only applicable to the analysis of non-negative data but also the data with negative numbers. They only limit the coefficient matrix after factorization to be

• Y. Zhao and J. Pei are with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518160, China.
E-mail: zhaoyangmaths@163.com, jhpei@szu.edu.cn.
• H. Wang is with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518160, China.
E-mail: 2130090503@email.szu.edu.cn.

non-negative and propose the semi-non-negative matrix factorizations algorithm (SNMF) [13]. In 2018, Rousset *et al.* applied SNMF to pattern generalization in single-pixel imaging [14]. Although SNMF can adapt to more data, the interpretability of the algorithm is weaker because it contains negative values in the basis matrix when dealing with image intelligence analysis problems.

The NMF-based algorithms above are all linear algorithms, whereas the distribution characteristics of real data tend to be complex and nonlinear. It is generally difficult to obtain better performance by using linear methods. Some researchers have studied the nonlinear NMF algorithm [15], [16], [17], [18], [19]. Ioan Buciu *et al.* use kernel trick [15] and polynomial kernel functions to extend the linear NMF to a nonlinear method [16]. However, only polynomial kernel functions can be applied to this method. Stefanos Zafeiriou *et al.* proposed a projected gradient kernel non-negative matrix factorization (PGKNMF) algorithm, enabling the algorithm to use arbitrary kernel functions [17].

The existing NMF-based algorithm can achieve effects on the problem of the pattern recognition problem, but the structures of the methods above are all single-layered, while the human visual system uses a multilayered and nonlinear method to analyze images [20]. Deep learning based methods have been applied to many practical data analysis recently [21], [22]. Therefore, researchers focus largely on studying the multilayered NMF algorithm in recent years [23], [24], [25], [26], [27], [28]. Ahn *et al.* proposed a multiple NMF (MNMF) network structure in which the deep feature of the sample under the basis images is obtained by continuously performing non-negative factorization on the coefficient matrix [29]. Cichocki *et al.* propose a multilayer NMF algorithm with multi-start initializations, which is also a deep factorization of the coefficient matrix [30], [31]. Song *et al.* extended the nsNMF algorithm that can learn sparse features to an algorithm with a multilayered structure [32]. In 2017, Trigeorgis *et al.* extended the SNMF algorithm to a multilayer structure and proposed the deep semi NMF (DSNMF) algorithm [33]. The algorithm is achieved by minimizing the reconstruction error of the original sample when obtaining the factorization result of each layer, instead of minimizing the reconstruction error of the previous coefficient matrix. It reduces the reconstruction error of the original sample but increases the amount of calculation. The DSNMF algorithm relaxes the constraints on the factorized matrix so that the basis images matrix acquired by the factorization does not satisfy the non-negative property. It is not possible to consider each column vector in the basis matrix as an image. The DSNMF explains the factorization result from the perspective of clustering. It abandons the idea of representing primitive samples based on the combination of basis images in classical NMF and weakens the semantic interpretability of the algorithm. In 2018, Tong *et al.* proposed a deep discriminative and robust non-negative matrix factorization network method (DDRNMF) for obtaining differentiated, compact and robust data representations [34]. The algorithm achieves the data representation by minimizing an objective function with a soft label constraint. However, the deep architecture of these methods still performs the deep factorization on the coefficient matrix.

Nowadays algorithms based on NMF and deep NMF focus on the coefficient matrix, including deep factorization on the coefficient matrix [33], [34] and adding different regularization constraints on the coefficient matrix [35], [36]. The optimization of the basis images matrix is indirect in these methods. However, the actual extracted features are closely relevant to the basis matrix. Making the basis vectors with obvious part-based features is advantageous for classification [37]. In addition, there is no clear interpretability and purpose for the deep factorization on the coefficient matrix. In this paper, a novel deep NMF architecture based on the underlying basis images learning (UBIL) is proposed to extract features reflecting the deep localization characteristics of samples. The architecture obtains the underlying basis images by deep factorization on the basis images matrix. The motivation of the proposed architecture is discussed in detail in Section 2.

To realize the proposed deep NMF architecture based on the UBIL the deep non-negative basis matrix factorization (DNBMF) algorithm is proposed. DNBMF, a linear algorithm that includes a multilayered structure, is the most basic algorithm for implementing the proposed architecture. Then, the regularized deep non-negative basis matrix factorization (RDNBMF) algorithm is proposed. The objective function of the RDNBMF algorithm contains the regularization constraint term on the basis images matrix. This term makes a larger difference between the factorized basis images so the parts-based features reflected from the obtained underlying basis images matrix are more explicit, which is helpful for classification. Finally, the regularized deep nonlinear non-negative basis matrix factorization (RDNNBMF) algorithm is proposed to deal with more complex identification problems. The RDNNBMF algorithm is a nonlinear version of the RDNBMF algorithm. In the RDNNBMF algorithm, the original samples are projected into a high dimensional space by a nonlinear map, and deep non-negative factorization based on the UBIL is performed on the mapped samples in this high dimensional space.

The following are the main innovations and contributions of this proposed deep NMF architecture:

1. A novel deep non-negative matrix factorization architecture based on learning underlying basis images is proposed. The architecture obtains the underlying basis images representing the original samples by factorizing the basis images hierarchically. To the best of our knowledge, it is the first time that a multilayer NMF architecture based on the deep factorization of the basis images is proposed. The proposed deep NMF architecture based on the underlying basis images learning is more strongly interpretable.

2. A new objective function with direct constraints on the basis images is proposed. This paper theoretically explains how the regularization constraint implemented on the non-negative basis image matrix can encourage the basis images to have better sparse property. The sparse term based on the l_1 and l_2 norms may lead to low rather than sparse values, which can be avoided by the proposed regularizer can avoid this to a certain extent. With the proposed regularization term constraint, each basis vector reflects distinct parts-based features.

(a) The factorization process of the existing NMF-based methods.

(b) The factorization process of the existing deep NMF-based methods.

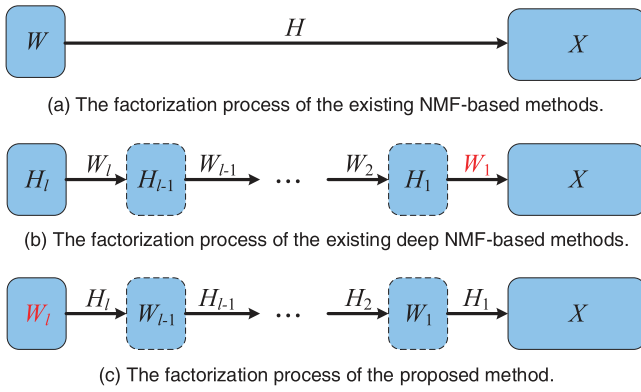(c) The factorization process of the proposed method.

Fig. 1. The factorizations of the classical shallow NMF algorithm, the existing deep NMF algorithm and the proposed deep NMF architecture based on the underlying basis images learning.

3. A nonlinear deep NMF algorithm based on UBIL is proposed to adapt to more complex data.
4. The convergence of the proposed algorithm is theoretically proved.

The rest of this paper is organized as follows. The second section details the motivation for the proposed architecture. The third section introduces the classic NMF algorithm and the existing deep NMF architecture. The fourth section presents the proposed deep NMF architecture based on the UBIL, which includes three algorithms DNBMF, RDNBMF and RDNNBMF. The fifth section is the theoretical proof of the convergence of the algorithm. The sixth section describes the experimental results and analysis. The seventh section concludes this paper.

2 MOTIVATIONS

NMF and deep NMF-based algorithms represent a non-negative column vector x_i as a linear combination of basis vectors, i.e., $x_i = \sum_j w_j h_{ji}$, where w_j is a basis vector, h_{ji} is a linear combination coefficient. For a non-negative data matrix X, the above linear combination form can be written in a matrix form, i.e., $X = WH$. Every column of X constitutes the training samples. Every column of W constitutes the basis vector. Every column of H constitutes the reconstruction coefficient vector. Basically all NMF-based algorithms map the original samples into the new feature space for data analysis by the Moore-Penrose pseudo-inverse of the basis images matrix W. The basis images matrix directly affect the new features extracted by the algorithm [37]. Therefore, it is very important to accurately obtain the basis images matrix with distinct parts-based features for the performance of the algorithm. Moreover, the existing algorithms do not contain the explicit physical meaning and is not well interpretable. The deep factorization on the coefficient matrix makes the final projection matrix of the feature mapping still only related to the basis matrix obtained in the first layer, which is the shallow basis images. It is inconsistent with the idea of extracting deep features, which affects the effect of feature extraction.

Fig. 1 shows the factorization architecture of the classical shallow NMF algorithm, the existing deep NMF algorithm, and the proposed deep NMF architecture based on the underlying basis images learning (UBIL). Fig. 1a shows the factorization architecture based on the classical shallow

NMF algorithm. The original sample X is represented by a set of shallow basis images through a linear combination of coefficients $X = WH$. Fig. 1b shows the factorization architecture based on the existing deep NMF algorithm. This architecture obtains the underlying coefficient matrix H_l by deep factorization of the coefficient matrix. The factorization task in the ith layer is $H_{i-1} = W_i H_i$. The original sample matrix X is factorized into $X = W_1 W_2 \dots W_{l-1} W_l H_l$, where H_l is the new sample feature acquired by this architecture. However, in this architecture, only the basis matrix W_1 obtained by the first layer factorization is directly related to the original images, which reflects the shallow localization features. The factorization after the first layer is a deep factorization of the coefficient matrix representing the samples through these bases. The obtained new feature by this architecture cannot be considered as the true underlying feature. The coefficient matrix of each layer is essentially the weighting coefficient in the process of sample reconstruction, and there is no actual physical meaning about the factorization on the coefficient matrix. The optimization of the basis matrix is still indirect and the purpose of this deep factorization architecture is weak.

Regarding the above existing deficiency, this paper proposes a novel deep NMF algorithm architecture based on UBIL. Fig. 1c shows the proposed deep NMF architecture. The underlying basis images representing the original sample is obtained by hierarchical deep factorization of the basis images. The factorization task in the ith layer is $W_{i-1} = W_i H_i$. The factorization task of each layer in the proposed architecture is to find the bases that can represent the basis images of the previous layer. The factorization of each layer has a strong purpose and interpretability. The original sample is expressed as $X = W_l H_l H_{l-1} \dots H_2 H_1$, where W_l is the basis image matrix obtained by the final layer with distinct parts-based features to represent the original images.

Fig. 2 shows an illustration of the proposed deep NMF architecture in detail. The detailed architecture based on the UBIL is presented with a hierarchical form in Fig. 2a. Each column in the original sample matrix X represents an original image, and each column in the basis images matrix W_l represents a basis image of the same size as the original image. The original image is represented by a linear combination of a set of basis images. The local features that are reflected by the basis images obtained by the shallow factorization are not obvious [33]. Each shallow basis image may contain multiple local features, such as the forehead and cheek. By deep factorization of the basis image, it is possible to obtain a better underlying basis image that reflects the distinct parts-based features of the original image. For instance, the basis image after deep factorization can highlight the eyebrows, left cheeks, right cheeks, teeth, etc. of the face, respectively. The final factorization form of the proposed deep NMF architecture based on the UBIL proposed in this paper is shown in Fig. 2b. The original sample X is represented by the underlying basis images. The proposed deep NMF architecture based on UBIL continuously factorizes the basis images to obtain the underlying basis images with clear local features of the original samples. The factorization process of the proposed factorization architecture is strongly interpretable.
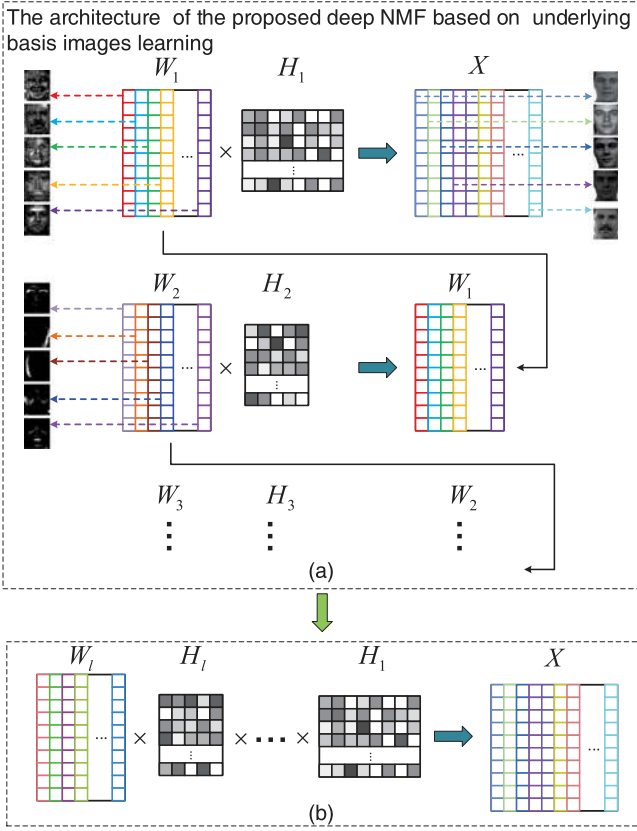
Fig. 2. The illustration of the proposed deep NMF architecture based on underlying basis images learning.

3 RELATED WORK

This section presents the classic NMF algorithm and the multilayer non-negative matrix factorization (MNMF) [29]. The classic NMF algorithm presents the most primitive idea of the NMF-based algorithm. MNMF is a deep matrix factorization method, which the existing deep NMF-based algorithms are based on. Let the original training sample be a non-negative vector of m dimensions $x_i \in R_+^m$, and the original training samples matrix X consists of n samples $X = [x_1, x_2, \ldots, x_n] \in R_+^{m \times n}$.

3.1 NMF

The classical NMF algorithm aims to decompose the original training sample $x_i \in R_+^m$ into a linear combination $x_i = \sum_{j=1}^{r} h_{ij} w_j$ of multiple basis images $w_j \in R_+^m$, where r is the number of basis images. h_{ij} is the coefficient of the jth basis image in the original sample x_i linearly represented by the basis images. For all training samples, there is a matrix form as $X = WH$. $W = [w_1, w_2, \ldots, w_r] \in R_+^{m \times r}$ is the basis images matrix composed of r basis images and $H = [h_1, h_2, \ldots, h_n] \in R_+^{r \times n}$ is the coefficient matrix of the linear representation. From the factorization form, it can be perceived that the training samples are represented by a set of basis vectors, which reflects the concept of "the whole consists of parts" in human thinking. The elements of the factorized basis vector are all non-negative, satisfying the non-negative nature of the image pixel points. To obtain the factorization, the algorithm minimizes the objective function:

$$J_{NMF}(W, H) = \frac{1}{2} \|X - WH\|_F^2, \ s.t. \ W \geq 0, \ H \geq 0. \quad (1)$$

Using the gradient descent method, the updated formulas for the optimization problem (1) can be obtained.

$$H^{(t+1)} = H^{(t)} \otimes (W^{(t)T} X) \oslash (W^{(t)T} W^{(t)} H^{(t)}), \quad (2)$$

$$W^{(t+1)} = W^{(t)} \otimes (XH^{(t)T}) \oslash (W^{(t)} H^{(t)} H^{(t)T}), \quad (3)$$

where $\otimes$ is the point multiplication operation and $\oslash$ is the point division operation.

Although the NMF algorithm and its variant algorithms have certain capabilities of feature extraction, the researchers believe that the underlying features of the sample cannot be obtained by shallow factorization [33]. Deep architectures have been widely applied particularly in image processing and analysis [23], [34], [38], since deep factorization provides high performance in data representation and efficient formulas behind deep learning [39]. However, there are still some challenges particularly in recognition [40]. The following section describes the MNMF, which represents the idea of the existing deep NMF algorithm.

3.2 Multilayer Non-Negative Matrix Factorization

The MNMF algorithm considers that deep factorization has better performance in data representation than shallow factorization. The MNMF constructs a deep factorization architecture based on the coefficient matrix as follows:

$$X = W_1 H_1, \quad (4)$$

$$H_1 = W_2 H_2, \quad (5)$$

$$\ldots, H_{l-1} = W_l H_l. \quad (6)$$

The original sample matrix X can be expressed as the product of $l + 1$ matrices.

$$X = W_1 W_2 \ldots W_l H_l, \quad (7)$$

where each element in the basis images matrix and the coefficient matrix is non-negative. The algorithm obtains the basis images matrix and coefficient matrix of the ith layer used to represent X by minimizing the following objective function:

$$J_{MNMF}(W_i, H_i) = \frac{1}{2} \|H_{i-1} - W_i H_i\|_F^2,$$
$$s.t. \ W_i \geq 0, \ H_i \geq 0. \quad (8)$$

The existing deep NMF algorithms are based on this multilayer factorization architecture [32], [33], [34]. The update formulas in each layer are

$$W_i^{(t+1)} = \arg\min_{W_i \geq 0} J_{MNMF}(W_i, H_i^{(t)}), \quad (9)$$

$$H_i^{(t+1)} = \arg\min_{H_i \geq 0} J_{MNMF}(W_i^{(t)}, H_i). \quad (10)$$

The existing algorithms for deep NMF with multilayered structure are based on the factorization architecture mentioned above, all of which perform deep factorization on the coefficient matrix to the best of our knowledge. The "partial" that makes up the original sample is in fact

obtained by factorizing only once. In this architecture, the basis images reflecting the local characteristics of the samples are the shallow basis images. The deep basis images reflecting deep local features of the original sample still cannot be obtained. Therefore, to extract the underlying basis images reflecting the local characteristics of the original samples, this paper proposes a deep NMF architecture based on the deep factorization of the basis images matrix to directly factorize the basis images. The Section 4 details the three proposed algorithms based on the deep NMF factorization architecture proposed in this paper.

4 DEEP NMF ARCHITECTURE BASED ON UNDERLYING BASIS IMAGES LEARNING

The existing deep NMF algorithm performs deep factorization on the coefficient matrix. However, in such a factorization architecture, only the basis matrix obtained by the first layer factorization is directly related to the original images, which reflects the shallow localization features. The factorization in the latter layers is a deep factorization of the coefficient matrix representing the samples through the shallow bases. The obtained new feature by this architecture cannot be considered as the true underlying feature. The optimization of the basis matrix is still indirect and the interpretability of this deep factorization architecture is weak. Therefore, a deep NMF architecture based on the UBIL is proposed to extract the real underlying basis images. In this deep factorization architecture, the factorization process of each layer contains explicit interpretability. The proposed factorization architecture consists of l layers with the following factorization process.

$$X = W_1 H_1, \tag{11}$$

$$W_1 = W_2 H_2, \tag{12}$$

$$\dots, W_{l-1} = W_l H_l, \tag{13}$$

where $X \in R^{m \times n}$ is the training samples matrix, $W_i = [w_1^i, w_2^i, \dots, w_{r_i}^i] \in R_+^{m \times r_i}$ is the basis image matrix of the ith layer and H_i is the coefficient matrix of the ith layer. Formula (11) represents the factorization process of the first layer. The original samples matrix X is factorized into the product of the basis images matrix W_1 and the coefficient matrix H_1. Formula (12) represents the factorization process of the second layer. The basis images matrix obtained by the factorization of the previous layer is further factorized into the product of the basis images matrix W_2 and the coefficient matrix H_2. A further factorization on the basis images matrix is intended to find the more localized basis images that can characterize the basis images of the previous layer. According to the factorization mode of the second layer, the basis images matrix W_i is continuously factorized until the underlying basis images W_l with deep localization features are obtained. Finally, the original sample image is represented by the underlying basis images

$$X = W_l H_l H_{l-1} \dots H_2 H_1. \tag{14}$$

To implement the proposed deep NMF architecture based on the UBIL, the DNBMF algorithm is proposed in Section 4.1, which is the most basic algorithm for implementing the

proposed architecture. The RDNBMF algorithm is proposed in Section 4.2 to obtain the basis images with more localized characteristics. In Section 4.3, the RDNBMF algorithm is extended to a nonlinear version, which enables the algorithm to deal with the more complex data.

4.1 Deep Non-Negative Basis Matrix Factorization Algorithm

For the sample matrix $X \in R^{m \times n}$, to obtain the underlying basis images matrix W_l of the sample images, the hierarchical factorization structure is designed to obtain the underlying basis images matrix W_l of the sample images.

In the first layer, the original sample image x_i is represented as a linear combination $x_i = \sum_j w_j^1 h_{ij}$ of a set of basis images w_j^i. It can be written in the matrix form $X = W_1 H_1$, where $W_1 \in R_+^{m \times r_1}$ is the basis images matrix containing r_1 basis images, and $H_1 = [h_1^1, h_2^1, \dots, h_n^1] \in R_+^{r_1 \times n}$ is the coefficient matrix of the first layer. In order to obtain the underlying basis images matrix W_l, this paper designs a hierarchical optimization method, which in turn optimizes basis images matrix W_l and coefficient matrices $H_i, i = 1, \dots, l$. In the optimization of the ith layer, it is necessary to solve the following optimization problem:

$$\min_{W_i, H_i} J_{DNBMF}^{(i)}(W_i, H_i) = \min_{W_i, H_i} \frac{1}{2} \|X - W_i H_i \cdots H_1\|_F^2 \tag{15}$$
$$s.t. \ W_i \geq 0, \ H_i \geq 0.$$

When optimizing the basis images matrix of the ith layer and its coefficient matrix, the coefficient matrixes of the front layer $H_1, H_2, \dots, H_{i-1}$ is known. To simplify the representation, the optimization problem (15) can be expressed as

$$\min_{W_i, H_i} J_{DNBMF}^{(i)}(W_i, H_i) = \min_{W_i, H_i} \frac{1}{2} \|X - W_i H_i \Lambda_{i-1}\|_F^2 \tag{16}$$
$$s.t. \ W_i \geq 0, \ H_i \geq 0,$$

where $\Lambda_{i-1} = H_{i-1} \dots H_1$, and when $i = 1$, $\Lambda_0 = I_n$ is the identity matrix. The above optimization problem can be transformed into two optimization sub-problems to optimize W_i, H_i, respectively, while fixing the other one. The gradient descent method is used to optimize the solution. The solution process is referred to Supplementary, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2019.2962679. The final update formula can be obtained

$$H_i^{(t+1)} = H_i^{(t)} \otimes \frac{W_i^{(t)T} X \Lambda_{i-1}^T}{W_i^{(t)T} W_i^{(t)} H_i^{(t)} \Lambda_{i-1} \Lambda_{i-1}^T} \tag{17}$$

$$W_i^{(t+1)} = W_i^{(t)} \otimes \frac{X \Lambda_{i-1}^T H_i^{(t)T}}{W_i^{(t)} H_i^{(t)} \Lambda_{i-1} \Lambda_{i-1}^T H_i^{(t)T}}, \tag{18}$$

where $\otimes$ denotes the point multiplication operation and $/$ is the element-wise matrix division. The algorithm step of the DNBMF algorithm is given in Algorithm 1. After iteratively optimizing the underlying basis images matrix W_l and coefficient matrices $H_i, i = 1, 2, \dots, l$) of the samples, the sample x can be directly mapped by the Moore-Penrose pseudo-inverse $W_n^\dagger$ of the underlying basis images to obtain new feature $W_n^\dagger x$.

Algorithm 1. The Proposed DNBMF

Input:
The training samples matrix $X \in R^{m \times n}$, the number of layers l, the number of basis images per layer $r_i, i = 1, 2, \ldots, l$, the iterations t.

Output:
The underlying basis images matrix W_l, coefficient matrices $H_i, i = 1, 2, \ldots, l$.
Initialize $\{W_i, H_i\}_{i=1}^{l}$
for $i = 1 : l$ **do**

$$\Lambda_{i-1} = \begin{cases} I_n, & i = 1 \\ H_{i-1} \ldots H_1, & i \neq 1 \end{cases}$$

for t = 1: item **do**

$$H_i^{(t+1)} = H_i^{(t)} \otimes \frac{W_i^{(t)T} X \Lambda_{i-1}^T}{W_i^{(t)T} W_i^{(t)} H_i^{(t)} \Lambda_{i-1} \Lambda_{i-1}^T}$$

$$W_i^{(t+1)} = W_i^{(t)} \otimes \frac{X \Lambda_{i-1}^T H_i^{(t)T}}{W_i^{(t)} H_i^{(t)} \Lambda_{i-1} \Lambda_{i-1}^T H_i^{(t)T}}.$$

end for
end for

4.2 The Regularized Deep Non-Negative Basis Matrix Factorization Algorithm

In the study of single-layer NMF algorithms, some researchers believe that it is helpful for classification to make the basis vectors sparse with distinct parts-based features [37]. This paper considers adding a regularization term to the objective function of each layer and the regularized deep non-negative basis matrix factorization (RDNBMF) algorithm is proposed. With the proposed regularization term constraint, each basis vector reflects distinct parts-based features. Meanwhile, this regularizer produces a better sparse effect. The existing sparse term based on the l_1 and l_2 norms may lead to low rather than sparse values, and the proposed regularizer can avoid this to a certain extent. The objective function of the RDNBMF in the ith layer is

$$J_{RDNBMF}^{(i)}(W_i, H_i) = \frac{1}{2} \|W_{i-1} - W_i H_i\|_F^2 - \frac{\alpha}{2} Tr(S_{W_i}), \tag{19}$$

where $W_i \geq 0$, $H_i \geq 0$ and $S_{W_i} = \sum_{j=1}^{r_i}(w_j^i - \bar{w}^i)(w_j^i - \bar{w}^i)^T$ is the total scatter matrix of the basis images at layer i. The objective function can be further written as follow, and the derivation process is given in the Supplementary, available online

$$J_{RDNBMF}^{(i)}(W_i, H_i) = \frac{1}{2} \|W_{i-1} - W_i H_i\|_F^2 + \frac{\alpha}{2r_i^2} \sum_{p \neq q} w_p^{iT} w_q^i \\ - \frac{\alpha(r_i - 1)}{2r_i^2} \sum_j w_j^{iT} w_j^i. \tag{20}$$

When the objective function $J_{RDNBMF}^{(i)}(W_i, H_i)$ is minimized, $\sum_{p \neq q} w_p^{iT} w_q^i$ has a smaller value and $\sum_j w_j^{iT} w_j^i$ has a larger value. Because each element of the basis vectors w_j^i is non-negative, the minimum value of $\sum_{p \neq q} w_p^{iT} w_q^i$ is 0. The effect of $\sum_{p \neq q} w_p^{iT} w_q^i = 0$ on the value of the elements in the basis vectors is that non-zero elements in different basis vectors w_j^i are at different positions. The reward term $\sum_j w_j^{iT} w_j^i$ does not degrade the basis vectors w_j^i to the zero vector, and at the same time enhances the non-zero elements of the basis vector at different positions from the other basis

vector. Therefore, the effects of the proposed regularization term lead to greater differences among different basis vectors, and avoid the situation where the values of elements in the basis vector are low rather sparse values.

The detailed solution process for the optimization problem (19) is given in the Supplementary, available online. The iteration formulas of the optimization problem (19) is

$$H_i^{(t+1)} = H_i^{(t)} \otimes \frac{W_i^{(t)T} X}{W_i^{(t)T} W_i^{(t)} H_i^{(t)}}, \tag{21}$$

$$W_i^{(t+1)} = W_i^{(t)} \otimes \frac{W_{i-1} H_i^{(t)T} + \alpha W_i^{(t)}}{W_i^{(t)} H_i^{(t)} H_i^{(t)T} + \alpha W_i^{(t)} A_i}, \tag{22}$$

where $A_i = \frac{1}{r_i} \mathbf{1}_{r_i \times r_i}$, $\mathbf{1}_{r_i \times r_i}$ is an $r_i \times r_i$ matrix of ones and r_i is the number of basis images of the ith layer. The RDNBMF algorithm step is given in Algorithm 2.

Algorithm 2. The Proposed RDNBMF

Input:
The training samples matrix $X \in R^{m \times n}$, the number of layers l, the number of basis images per layer $r_i, i = 1, 2, \ldots, l$, the regularization factor α, the iterations t.

Output:
The underlying basis images matrix W_l, coefficient matrices $H_i, i = 1, 2, \ldots, l$.
Initialize $\{W_i, H_i\}_{i=1}^{l}$
for $i = 1 : l$ **do**

$A_i = \frac{1}{r_i} \mathbf{1}_{r_i \times r_i}$

for t = 1: item **do**

$$H_i^{(t+1)} = H_i^{(t)} \otimes \frac{W_i^{(t)T} X}{W_i^{(t)T} W_i^{(t)} H_i^{(t)}},$$

$$W_i^{(t+1)} = W_i^{(t)} \otimes \frac{W_{i-1} H_i^{(t)T} + \alpha W_i^{(t)}}{W_i^{(t)} H_i^{(t)} H_i^{(t)T} + \alpha W_i^{(t)} A_i}.$$

end for
end for

4.3 The Regularized Deep Nonlinear Non-Negative Basis Matrix Factorization Algorithm

Since most of the actual data is complex and nonlinear, the factorization of the original sample data by linear methods cannot be effectively described [33]. The human visual system adopts a hierarchical and nonlinear method when analyzing images [20]. From a mathematical point of view, the model's expressibility can be enhanced by nonlinear functions and a better sample representation is provided [41]. Therefore, a regularized deep nonlinear non-negative basis matrix factorization algorithm (RDNNBMF) is proposed. In RDNNBMF, the original samples and the basis images of each layer are projected into a high-dimensional space by a nonlinear mapping, and the preimages of the underlying basis images are optimized in this high-dimensional space.

The original sample matrix $X \in R^{m \times n}$ is mapped to $\varphi(X) = [\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)] \in R^{M \times n}$ by a non-linear map φ. The basis images $\hat{w}_j^i \in R_+^{m \times 1}, j = 1, \ldots, r_i$ are mapped to $\varphi(\hat{w}_j^i) \in R_+^{M \times 1}, j = 1, \ldots, r_i$, where $M \gg m$. By factorizing the basis images of each layer, there is

$$\varphi(\hat{W}_{i-1}) = \varphi(\hat{W}_i)\hat{H}_i, \tag{23}$$

where $\hat{W}_i = [\hat{w}_1^i, \hat{w}_2^i, \ldots, \hat{w}_{r_i}^i] \in R_+^{m \times r_i}$ and $\varphi(\hat{W}_i) = [\varphi(\hat{w}_1^i), \varphi(\hat{w}_2^i), \ldots, \varphi(\hat{w}_{r_i}^i)] \in R_+^{M \times r_i}$. The mapped original samples matrix is deeply factorized by the mapped basis images matrix in a hierarchical architecture, which can be expressed as

$$\varphi(X) = \varphi(\hat{W}_i)\hat{H}_i\hat{H}_{i-1}\ldots\hat{H}_1. \tag{24}$$

The algorithm also directly factorizes the basis images of each layer. To obtain the final factorization form in (24), it is necessary to solve the following optimization problem with regularization by hierarchically optimizing the basis images matrix of each layer

$$\min_{\hat{W}_i \geq 0, \hat{H}_i \geq 0} J_{RDNNBMF}^{(i)}(\hat{W}_i, \hat{H}_i)$$
$$= \min_{\hat{W}_i \geq 0, \hat{H}_i \geq 0} \frac{1}{2}\|\varphi(\hat{W}_{i-1}) - \varphi(\hat{W}_i)\hat{H}_i\|_F^2 - \frac{\alpha}{2}Tr(S_{\hat{W}_i}^\varphi), \tag{25}$$

where $S_{\hat{W}_i}^\varphi$ is the total scatter matrix $S_{\hat{W}_i}^\varphi = \sum_{j=1}^{n_i}(\varphi(\hat{w}_j^i) - \varphi(\bar{w}^i))(\varphi(\hat{w}_j^i) - \varphi(\bar{w}^i))^T$ of the ith layer basis images matrix in the high dimensional space. In general, it is difficult to determine the explicit expression of the nonlinear function φ. According to the kernel trick [15], the inner product between samples after nonlinear mapping can be represented by the kernel function $k(x_i, x_j) = \langle\varphi(x_i), \varphi(x_j)\rangle$ of the original samples. Gaussian radial basis functions $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2/t)$ are used as kernel functions in this paper. The solution process is referred to Supplementary, available online.

The iterative formulas for the basis images matrix and the coefficient matrix are shown as (26) and (27), where $A_i = \frac{1}{r_i}\mathbf{1}_{r_i \times r_i}$, $\mathbf{1}_{r_i \times r_i}$ is an $r_i \times r_i$ matrix of ones and r_i is the number of basis images of the ith layer. The algorithm steps of the proposed RDNNBMF are given in Algorithm 3.

Algorithm 3. The Proposed RDNNBMF

Input:
  The training samples matrix $X \in R^{m \times n}$, the number of layers l, the number of basis images per layer $r_i, i = 1, 2, \ldots, l$, the regularization factor α, the iterations t.

Output:
  The underlying basis images matrix $\hat{W}_l$, coefficient matrices $\hat{H}_i, i = 1, 2, \ldots, l$.
  Initialize $\{\hat{W}_i, \hat{H}_i\}_{i=1}^l$
  for $i = 1 : l$ **do**
    $A_i = \frac{1}{r_i}\mathbf{1}_{r_i \times r_i}$
    for t = 1: item **do**
      Updating $\hat{H}_i$ according to formula (26).
    Updating $\hat{W}_i$ according to formula (27).
    end for
  end for

When the training is completed, the mapped samples are expressed in the form of the formula (24) by the mapped basis images matrix and the coefficient matrix. By multiplying $\varphi(\hat{W}_l)^T$ on both sides of the formula (24), there is

$$\varphi(\hat{W}_l)^T\varphi(X) = \varphi(\hat{W}_l)^T\varphi(\hat{W}_l)\hat{H}_l\hat{H}_{l-1}\ldots\hat{H}_1$$
$$\Rightarrow K_{\hat{W}_lX} = K_{\hat{W}_l\hat{W}_l}H_{RDNNBMF}.$$

The new features $H_{RDNNBMF}$ of the samples extracted by the proposed RDNNBMF based on the underlying base image are

$$H_{RDNNBMF} = K_{\hat{W}_l\hat{W}_l}^\dagger K_{\hat{W}_lX}, \tag{28}$$

where $[K_{\hat{W}_l\hat{W}_l}]_{pq} = \exp(-\|\hat{w}_p^l - \hat{w}_q^l\|^2/t)$, $[K_{W_lX}]_{pq} = \exp(-\|\hat{w}_p^l - x_q\|^2/t)$ and $K_{\hat{W}_l\hat{W}_l}^\dagger$ is the Moore-Penrose pseudo-inverse of $K_{\hat{W}_l\hat{W}_l}$.

5 THE CONVERGENCE OF THE PROPOSED ALGORITHM

This section theoretically proves the convergence of the proposed RDNBMF algorithm. The convergence of Algorithms 1 and 3 proves similar. In order to prove that the optimization problem (19) is convergent under the iterative formulas (21) and (22), the theory of the auxiliary function is needed [2]. The following is a brief introduction of the definition of the auxiliary function and its properties.

Definition 1. $G(w, w')$ is an auxiliary function for $J(w)$ if, for any vector $w, w' \in R^m$, there are $G(w, w') \geq J(w)$ and $G(w, w) = J(w)$.

Lemma 1. $J(w)$ is non-increasing with the iteration formula $w^{(t)} = \arg\min_w G(w, w^{(t-1)})$, if $G(w, w')$ is an auxiliary function for $J(w)$.

The objective function in the optimization problem (19) is able to be treated as a function of the basis image w_b^i and can be re-expressed as

$$J(w_b^i) = \frac{1}{2}\|W_{i-1} - W_iH_i\|_F^2 - \frac{\alpha}{2}Tr(S_{W_i}), \tag{29}$$

where w_b^i is the bth basis image of the ith layer, $W_i = [w_1^i, w_2^i, \ldots, w_{r_i}^i]$ is the basis images matrix of the ith layer. Because the convergence proof process for each layer is similar, the convergence of the first layer is proved in this section. In order to facilitate the observation, the

$$\hat{H}_i^{(t+1)} = \hat{H}_i^{(t)} \otimes \frac{K_{\hat{W}_i^{(t)}\hat{W}_{i-1}^{(t)}}}{K_{\hat{W}_i^{(t)}\hat{W}_i^{(t)}}\hat{H}_i^{(t)}}, \tag{26}$$

$$\hat{W}_i^{(t+1)} = \hat{W}_i^{(t)} \otimes \frac{\hat{W}_{i-1}(\hat{H}^T \otimes K_{\hat{W}_{i-1}\hat{W}_i^{(t)}}) + \hat{W}_i^{(t)} \otimes (\mathbf{1}_{m \times 1}diag((\hat{H}\hat{H}^T + \alpha A_i)K_{\hat{W}_i^{(t)}\hat{W}_i^{(t)}})) + \alpha\hat{W}_i^{(t)}K_{\hat{W}_i^{(t)}\hat{W}_i^{(t)}}}{\hat{W}_i^{(t)} \otimes (\mathbf{1}_{m \times 1}diag(\hat{H}K_{\hat{W}_{i-1}\hat{W}_i^{(t)}} + \alpha K_{\hat{W}_i^{(t)}\hat{W}_i^{(t)}})) + \hat{W}_i^{(t)}((\hat{H}\hat{H}^T + \alpha A_i) \otimes K_{\hat{W}_i^{(t)}\hat{W}_i^{(t)}})}. \tag{27}$$

layer mark is omitted in the proof process. In the first layer $W_0 = X$, then the formula (29) can be rewritten as

$$J(w_b) = \frac{1}{2}\|X - WH\|_F^2 - \frac{\alpha}{2}Tr(S_W). \tag{30}$$

The following Theorem 1 proves that the $G(w_b, w_b^{(t)})$ given by the formula (31) is an auxiliary function for $J(w_b)$.

Theorem 1. $G(w_b, w_b^{(t)})$ is an auxiliary function for $J(w_b)$, $b = 1, 2, \ldots, r$. $G(w_b, w_b^{(t)})$ is given by

$$G(w_b, w_b^{(t)}) = J(w_b^{(t)}) + (w_b - w_b^{(t)})^T \nabla J(w_b^{(t)})$$
$$+ \frac{1}{2}(w_b - w_b^{(t)})^T L(w_b^{(t)})(w_b - w_b^{(t)}), \tag{31}$$

where w_b, $w_b^{(t)}$ are the bth columns of the matrices W, $W^{(t)}$, respectively. $L(w_b^{(t)})$ is a diagonal matrix. The diagonal element of $L(w_b^{(t)})$ is $[L(w_b^{(t)})]_{aa} = [(W^{(t)}B_b + \alpha W^{(t)}A_b)]_a / [w_b^{(t)}]_a$, where $A = \frac{1}{r}\mathbf{1}_{r\times r}$, $B = HH^T$ and A_b, B_b represent the bth columns of the matrices A, B, respectively.

Proof. According to the definition of $J(w_b)$ and $G(w_b, w_b^{(t)})$ in formulas (30) and (31), it is easy to know that $G(w_b, w_b) = J(w_b)$. Then just need to prove that $G(w_b, w_b^{(t)}) \geq J(w_b)$. The function $J(w_b)$ can be written as a Taylor expansion of the point $w_b^{(t)}$

$$J(w_b) = J(w_b^{(t)}) + (w_b - w_b^{(t)})^T \nabla J(w_b^{(t)})$$
$$+ \frac{1}{2}(w_b - w_b^{(t)})^T \nabla^2 J(w_b^{(t)})(w_b - w_b^{(t)}), \tag{32}$$

where $\nabla J(w_b^{(t)})$ is the first-order partial derivative of the function $J(w_b^{(t)})$ with respect to $w_b^{(t)}$

$$\nabla J(w_b^{(t)}) = W^{(t)}B_b - XH^{(t)T}e_b - \alpha w_b^{(t)} + \alpha W^{(t)}A_b,$$

where $A = \frac{1}{r}\mathbf{1}_{r\times r}$, $B = HH^T$ and e_b is a column vector whose element in row b is 1 and the other elements are 0. The second-order partial derivative of the function $J(w_b^{(t)})$ with respect to $w_b^{(t)}$ is $\nabla^2 J(w_b^{(t)}) = (B_{bb} - \alpha + \alpha A_{bb})I$, where A_{bb} and B_{bb} are elements on the bth row and the bth column of A and B, respectively. The matrix I is an identity matrix. It is easy to know from formulas (31) and (32), if the inequality $G(w_b, w_b^{(t)}) \geq J(w_b)$ is true, just need to prove that

$$(w_b - w_b^{(t)})^T (L(w_b^{(t)}) - \nabla^2 J(w_b^{(t)}))(w_b - w_b^{(t)}) \geq 0. \tag{33}$$

That means we should prove that $L(w_b^{(t)}) - \nabla^2 J(w_b^{(t)})$ is a positive semi-definite matrix. According to the definition of $L(w_b^{(t)})$ and $\nabla^2 J(w_b^{(t)})$, the definition of the element in the qth column and the pth row is $[L(w_b^{(t)})]_{pq} = \delta_{pq}[W^{(t)}B_b + \alpha W^{(t)}A_b]_p / [w_b^{(t)}]_p$, $[\nabla^2 J(w_b^{(t)})]_{pq} = \delta_{pq}([B]_{bb} - \alpha + \alpha[A]_{bb})$, where δ_{pq} is defined as

$$\delta_{pq} = \begin{cases} 1, & p = q, \\ 0, & p \neq q. \end{cases} \tag{34}$$

Let $U \equiv L(w_b^{(t)}) - \nabla^2 J(w_b^{(t)})$, for any vector v, there is

$$v^T U v = \sum_{pq} v_p([L(w_b^{(t)})]_{pq} - [\nabla^2 J(w_b^{(t)})]_{pq})v_q$$
$$= \sum_{pq} v_p(\delta_{pq}[L(w_b^{(t)})]_{pp} - [\nabla^2 J(w_b^{(t)})]_{pq})v_q$$
$$= \sum_{pq} v_p \left(\delta_{pq} \left[\frac{W^{(t)}B_b + \alpha W^{(t)}A_b}{w_b^{(t)}} \right]_p \right) v_q$$
$$- \sum_{pq} v_p(\delta_{pq}(B_{pp} - \alpha + \alpha A_{pp}))v_q$$
$$= \sum_p \left(\frac{\sum_{i\neq b}([w_i^{(t)}]_p[B]_{ib} + \alpha[w_i^{(t)}]_p[A]_{ib})}{[w_b^{(t)}]_p} + \alpha \right) v_p^2. \tag{35}$$

Hence, $U = L(w_b^{(t)}) - \nabla^2 J(w_b^{(t)})$ is positive semi-definite. Then $G(w_b, w_b^{(t)}) \geq J(w_b)$. According to Definition 1, $G(w_b, w_b^{(t)})$ is an auxiliary function for $J(w_b)$. Then Theorem 1 is proved. □

As can be seen from Lemma 1, the objective function $J(w_b)$ is non-increasing under the iterative formula $w^{(t)} = \arg\min_w G(w, w^{(t-1)})$. Let $\nabla G(w_b) = \nabla J(w_b^{(t)}) + L(w_b^{(t)})(w_b - w_b^{(t)}) = 0$. With some line algebra, the iterative formula for w_b is

$$w_b^{(t+1)} = w_b^{(t)} \otimes \frac{XH^{(t)T}e_b + \alpha w_b^{(t)}}{W^{(t)}B_b + \alpha W^{(t)}A_b}. \tag{36}$$

The iterative formula can be written in the form of a matrix with layer information as follows:

$$W_i^{(t+1)} = W_i^{(t)} \otimes \frac{W_{i-1}H_i^{(t)T} + \alpha W_i^{(t)}}{W_i^{(t)}H_i^{(t)}H_i^{(t)T} + \alpha W_i^{(t)}A_i}. \tag{37}$$

The iterative formula (37) for W_i can be obtained, which is the same as the iterative formula (22) derived in Section 4.2. Therefore, the objective function $J_{RDNBMF}^{(i)}(W_i, H_i)$ is non-increasing under the iteration formula (37). Similarly, it can be proved that the objective function $J_{RDNBMF}^{(i)}(W_i, H_i)$ is non-increasing with the iterative formula for H_i. Algorithms 1 and 3 can be proved to be convergent under their respective iterative formulas in a similar way. The experimental results about the convergence of the proposed algorithm are given in the Supplementary, available online.

6 EXPERIMENTAL RESULTS AND ANALYSES

To evaluate the performance of the proposed deep NMF architecture based on the UBIL this section is organized in sequence of the following four parts. Section 6.1 introduces the databases and the experimental parameters setting of the proposed algorithms and the comparison algorithms. The experiments assessing the recognition performance of the proposed three algorithms and 10 state-of-the-art comparison algorithms are carried out on 6 publicly available databases in Section 6.2. The algorithmic performance on different number of underlying basis images is presented in Section 6.3. The basis images extracted by different algorithms are analyzed intuitively in Section 6.4.
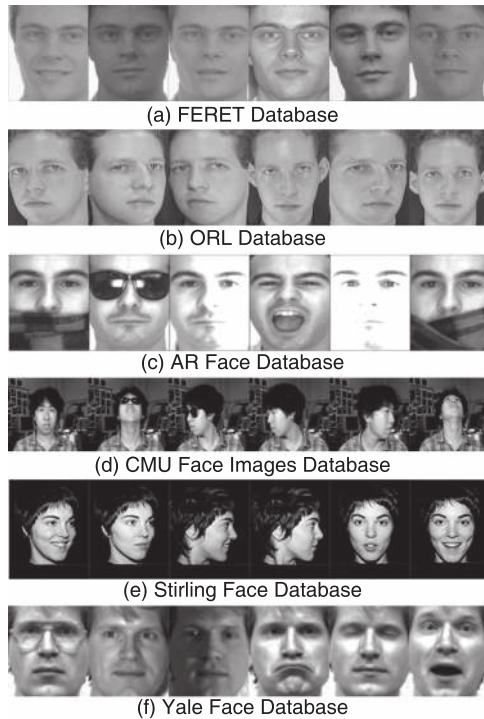
(a) FERET Database

(b) ORL Database

(c) AR Face Database

(d) CMU Face Images Database

(e) Stirling Face Database

(f) Yale Face Database

Fig. 3. Part face images from the 6 face databases.

6.1 Databases and Experimental Parameters Setting

Fig. 3 shows part face images form FERET, ORL, AR, CMU, Stirling and Yale face databases. The brief description of the 6 publicly available databases used in the experiments is given as follow.

FERET Database. This database comprises of 720 facial images of 120 subjects [42], [43]. Each person has 6 images with different illumination conditions, facial expressions, poses and ages. The resolution of each image is 30×25.

ORL Database. There are 400 grayscale facial images from 40 persons, 10 of each. The images change in pose and expression. The resolution of each image is 30×25.

AR Face Database. This database comprises of 3094 facial images of 119 persons [44]. Each person has 26 images with different illumination conditions, facial expressions and with/without sunglasses or scarves. The resolution of each image is 30×25.

CMU Face Images Database. There are 640 grayscale facial images from 20 people. Each one has 32 images with the difference of angles, expressions and with/without sunglasses. The resolution of each image is 32×34.

Stirling Face Database. There are 297 grayscale facial images from 33 people, including 18 females and 15 males. Each one has 9 images with different angles, expressions. The resolution of each image is 88×68.

Yale Face Database. There are 165 grayscale facial images with 15 persons. Each person has 11 images. The images demonstrate variations in illumination condition, pose, facial expression, and with/without glasses. The resolution of each image is 27×27.

Principal Component Analysis (PCA) [45], Kernel Principal Component Analysis (KPCA) [46], Direct Linear Discriminate Analysis (DLDA) [47], NMF [1], MNMF [29], PNMF [16], KNMF-RBF [17], SNMF [13], DSNMF [33] and DDRNMF [34] are chosen as comparison algorithms. In PCA, the eigenvector

whose eigenvalue is large than 0.01 is kept. In KPCA, the kernel function is radial basis function $k(x,y) = \exp(-\|x-y\|^2/t)$, the eigenvector whose eigenvalue is large than $1e-6$ is kept. In DLDA, the threshold value of the eigenvalue is $1e-2$. In NMF, the number of basis images is 100. In PNMF, the kernel function is polynomial function $k(x,y) = \langle x,y \rangle^d$, the parameter $d \neq 1$, the number of basis images is 90. In KNMF-RBF, the kernel function is radial basis function, the number of basis images is set as 60. In SNMF, the number of basis images is set as 100 on ORL database, 60 on other databases. For other comparison algorithms MNMF, DSNMF, DDRNMF and the proposed DNBMF, RDNBMF and RDNNBMF with deep structure, which are comprised of two reconstruction layers, the number of underlying basis images of those algorithms are shown in Table 1. The value of the Gaussian RBF parameters t is tried between $1e+2$ to $1e+4$, and the range of the regularization coefficient α from the proposed RDNBMF and RDNNBMF is in (0,1). In all methods, the maximum number of iterations is set to 1000. The important parameters setting of all the comparison algorithms are based on the algorithmic performance.

6.2 Experiments on Recognition Performance

In order to evaluate the performance of the proposed deep NMF architecture based on the UBIL and the recognition performance of the proposed three algorithms, we experiment on 6 publicly available databases. All images of the same person are taken as one class. The training number (TN) represents the number of images per person used for training. Different training numbers are chosen randomly on each database. The rest are chosen as testing samples. Three popular metrics that are used to assess the results of recognition experiments are accuracy (ACC), receiver operating characteristic (ROC) curve, and the area under curve (AUC) [48], [49]. The ROC curve is a 2-dimensional graph in which the true positive rate (TPR) is plotted on the Y-axis and the false positive rate (FPR) is plotted on the X-axis. TPR is the ratio of the number of positive samples being classified correctly to the number of all positive samples. FPR is the ratio of the number of negative samples classified as positive samples to the number of all negative samples. The area under the ROC curve is the AUC score. The closer the ROC curve to the upper left corner (0,1), the better the classification performance. The larger the area of the area under the ROC curve (AUC), the better the classification performance. All of the experiments are repeated 10 times with the same experimental conditions, and the average value is recorded as the final result. The nearest neighbor classifier is used for classification.

TABLE 1
The Number of Underlying Basis Images of the Algorithms With Multi-Layer Structure

	ORL	Yale	FERET	AR	CMU	Stirling
MNMF	30	40	60	30	50	30
DSNMF	90	110	110	90	90	100
DDRNMF	120	120	130	120	130	120
DNBMF	40	60	90	90	90	90
RDNBMF	100	60	140	140	60	140
RDNNBMF	60	70	140	60	140	60

TABLE 2
Mean ACC (%) Versus TN on the FERET Dataset

TN	2	3	4	5
PCA	36.88 (1.87)	40.04 (2.53)	38.00 (2.36)	40.67 (2.38)
KPCA	49.00 (2.04)	56.78 (2.14)	59.42 (2.58)	61.75 (3.20)
DLDA	60.17 (1.62)	78.36 (1.56)	83.75 (1.97)	87.17 (3.00)
NMF	70.81 (1.32)	80.61 (1.00)	83.75 (1.33)	83.33 (1.67)
MNMF	57.13 (2.76)	61.75 (2.69)	65.63 (1.79)	66.58 (5.10)
PNMF	66.08 (0.78)	76.17 (1.14)	79.79 (1.22)	75.17 (1.92)
KNMF-RBF	37.13 (1.89)	41.53 (3.25)	41.88 (2.54)	47.67 (4.69)
SNMF	59.33 (1.16)	67.00 (1.19)	69.79 (1.08)	73.67 (3.91)
DSNMF	63.13 (1.84)	69.86 (1.12)	73.04 (1.72)	73.83 (2.46)
DDRNMF	68.10 (2.34)	74.94 (2.21)	78.25 (2.80)	81.50 (3.31)
DNBMF	65.67 (2.11)	71.75 (1.66)	75.88 (2.11)	77.50 (2.89)
RDNBMF	75.19 (0.51)	82.89 (1.24)	86.69 (1.01)	88.17 (1.16)
RDNNBMF	**75.54** (1.27)	**84.08** (0.45)	**88.17** (0.77)	**93.25** (1.54)

Table 2 shows the recognition results on FERET database. The best accuracy of each column is in bold. It is obvious that the recognition rate increases with the increasing TN. This means that the recognition performance of algorithm enhances as TN increasing. Experimental results show that the proposed DNBMF has better recognition performance than the other two deep factorization algorithms MNMF and DSNMF. The proposed DNBMF is based on the deep factorization of the basis images, while the MNMF and DSNMF are based on deep factorization of the coefficient matrix. In particular, the only difference between the proposed linear algorithm DNBMF and MNMF is that the object of deep factorization is different. This shows that the proposed architecture based on the deep factorization of the basis images can obtain better sample feature representation, which is more conducive to the recognition task. It can be observed that in this database, NMF has higher recognition performance than the proposed linear method DNBMF and the RDNBMF algorithm with regularization constraints on the basis images has higher performance than the NMF algorithm. This is because there is no special constraint on the basis images in the DNBMF algorithm, and the performance of the deep basis images obtained by the multi-layer factorization algorithm without regular guidance on the basis images for the complex FERET database may be degraded layer by layer. For the proposed RDNBMF, it adds regular guidance on the expression ability of the basis images in the learning process. The basis images obtained in this way is more purposeful, and the recognition performance can be significantly improved compared with NMF. This also shows that adding regularization constraints on the basis images during the learning process is beneficial to the recognition performance.

The experimental results show that our proposed algorithms significantly outperform all the compared algorithms. For the proposed deep NMF architecture based on the UBIL, the recognition performance of linear algorithm DNBMF is not particularly outstanding, but the performance of linear algorithm RDNBMF has got a significant improvement. This is because in RDNBMF the basis images in each layer are obtained by the objective functions with regularization constraint, which makes a greater difference between the learned basis images. It demonstrates that the basis images learned by RDNBMF are helpful for the recognition task. Noted that the nonlinear algorithm RDNNBMF outperforms linear algorithms DNBMF and RDNBMF because the nonlinear method can use the nonlinear interface to classify the samples. In general, the nonlinear method has better performance in dealing with the nonlinear distribution data [50].

To further evaluate the performance of the deep NMF architecture and the proposed algorithms, ROC curves and the AUC scores of the corresponding curves are given in the cases of $TN = 3$ and $TN = 4$ on FERET database, shown in Figs. 4a and 4b. The figures show that the AUC score of
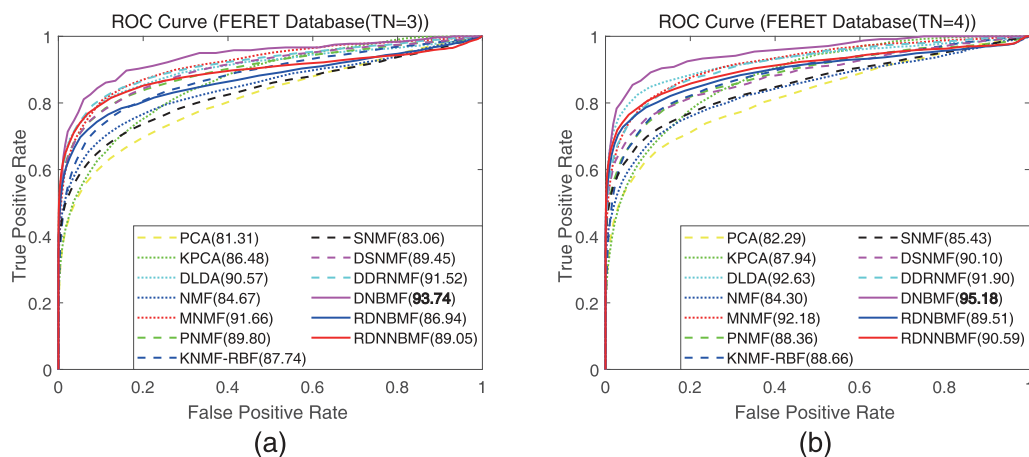

Fig. 4. The ROC curves on FERET database. The DNBMF, RDNBMF and RDNNBMF are the proposed algorithms. The AUC scores is shown in parentheses.

TABLE 3
Mean ACC (%) Versus TN on the ORL Dataset

TN	2	3	4	5	6	7	8	9
PCA	84.00	87.71	89.54	90.25	92.06	94.00	94.00	95.25
	(1.94)	(2.51)	(2.23)	(1.60)	(2.28)	(2.32)	(2.69)	(1.84)
KPCA	84.41	88.54	90.92	92.90	93.75	95.00	**95.88**	97.50
	(3.23)	(2.84)	(1.81)	(2.09)	(2.23)	(1.76)	(1.96)	(2.36)
DLDA	78.38	86.75	91.04	92.75	93.13	93.75	95.50	94.75
	(2.92)	(2.32)	(1.69)	(1.51)	(1.25)	(1.53)	(1.79)	(2.19)
NMF	82.03	82.25	84.25	90.10	89.56	95.33	94.38	94.25
	(1.13)	(1.61)	(0.62)	(1.13)	(2.36)	(1.63)	(1.69)	(2.06)
MNMF	72.81	79.64	84.50	87.20	90.63	89.83	93.50	97.25
	(4.03)	(1.60)	(1.59)	(1.01)	(1.47)	(2.18)	(1.75)	(2.75)
PNMF	79.38	87.21	88.21	90.80	91.88	94.33	95.75	95.50
	(0.96)	(1.03)	(0.86)	(0.63)	(0.78)	(0.66)	(0.87)	(1.05)
KNMF-RBF	84.22	88.25	91.08	92.65	92.06	94.08	94.13	96.00
	(2.93)	(2.64)	(1.46)	(1.99)	(0.84)	(2.17)	(1.32)	(3.57)
SNMF	68.81	75.04	80.04	89.10	90.75	90.42	90.13	94.25
	(1.53)	(2.90)	(2.25)	(2.54)	(1.44)	(1.32)	(1.38)	(1.69)
DSNMF	69.97	80.64	86.79	88.10	91.50	92.92	93.50	92.00
	(2.30)	(2.28)	(1.84)	(2.07)	(1.29)	(1.43)	(1.75)	(2.58)
DDRNMF	78.81	83.46	88.17	91.00	91.19	92.50	93.88	94.00
	(3.22)	(1.77)	(2.28)	(2.32)	(1.68)	(2.55)	(2.32)	(2.93)
DNBMF	81.47	85.36	87.46	92.20	**93.88**	94.50	94.63	96.75
	(1.55)	(1.57)	(1.86)	(1.49)	(1.34)	(1.48)	(2.13)	(1.21)
RDNBMF	82.13	88.14	90.33	**93.00**	92.88	**95.42**	95.50	**99.25**
	(2.38)	(1.61)	(1.36)	(1.00)	(1.03)	(1.13)	(1.58)	(1.21)
RDNNBMF	**84.66**	**89.43**	**91.17**	92.75	93.69	94.83	94.38	95.75
	(2.20)	(1.70)	(1.71)	(1.78)	(2.68)	(1.88)	(1.47)	(2.65)

Table 3 shows the experimental results on the ORL database, and Fig. 5a shows the results visually. The experimental results show that some linear algorithms can achieve good recognition performance. This is because the images from the ORL database are affected slightly by shooting angle, illumination, and occlusion, so the distribution of data is relatively simple. In this case, some linear algorithms can also obtain competitive recognition performance, especially the shallow NMF algorithm. In general, the best recognition performance with different numbers of training samples is given by the proposed algorithms. DNBMF has better recognition performance than MNMF in all cases on the ORL database, which means that the proposed deep NMF architecture based on the UBIL can obtain more useful basis images for feature representation that are beneficial for improving recognition performance.

To further evaluate the performance of the proposed deep NMF architecture and the proposed algorithms, the ROC curves in the cases of $TN = 4$ and $TN = 6$ on the ORL database, and the AUC scores of the corresponding curves are shown in Figs. 5b and 5c, respectively. From the results in figures, it can be known that the AUC score of the proposed RDNNBMF is the highest 97.5 percent in the case of $TN = 4$, and DNBMF is the highest 98.90 percent in the case of $TN = 6$. This validates that the proposed deep NMF architecture is more suitable for recognition tasks.

The experimental results on AR face database are shown in Table 4. Fig. 5d visually shows a line graph of the experimental results. The number of the classes is large in this face database. The changes of face images are more complex, including excessive exposure, large areas of occlusion, and dramatic expression changes. Moreover, the data used in this experiment has never been filtered, so the recognition performance of some algorithms on this database is particularly

DNBMF is the highest, which demonstrates that the proposed deep NMF architecture is more suitable for recognition tasks.
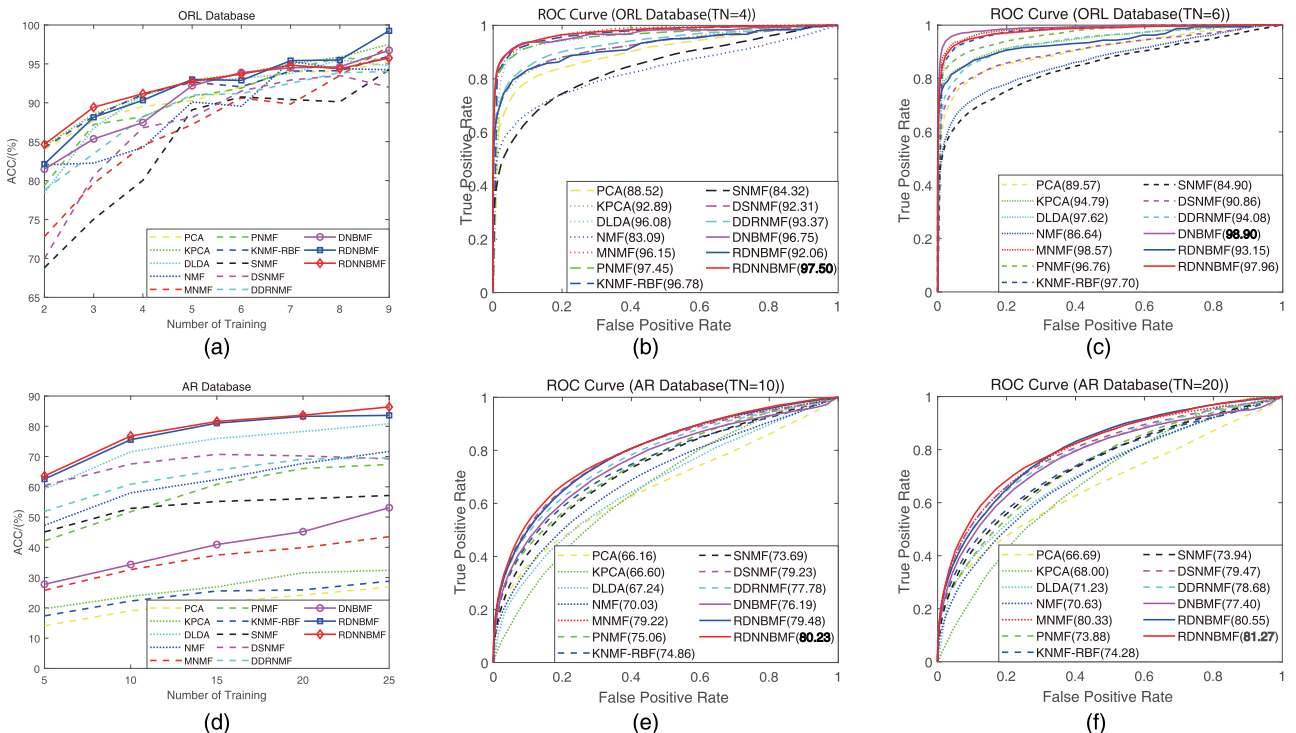

Fig. 5. Recognition results (ACC) and ROC curves on ORL and AR databases. The DNBMF, RDNBMF, and RDNNBMF are the proposed algorithms. The AUC scores are given in parentheses.

TABLE 4
Mean ACC (%) Versus TN on the AR Face Dataset

TN	5	10	15	20	25
PCA	14.07 (0.88)	19.04 (0.80)	21.73 (1.06)	24.24 (1.41)	26.89 (5.07)
KPCA	19.74 (1.06)	23.85 (0.89)	26.94 (1.15)	31.61 (1.85)	32.44 (5.85)
DLDA	59.29 (0.87)	71.59 (0.96)	75.98 (0.60)	78.28 (1.11)	80.84 (3.79)
NMF	47.27 (0.53)	58.02 (1.11)	62.38 (1.73)	67.72 (1.26)	71.68 (3.45)
MNMF	25.81 (1.54)	32.64 (2.10)	37.42 (1.87)	39.92 (1.36)	43.53 (4.15)
PNMF	42.14 (0.42)	51.77 (0.57)	60.82 (0.66)	66.05 (0.91)	67.39 (3.33)
KNMF-RBF	17.37 (1.04)	22.26 (1.03)	25.58 (1.74)	25.94 (1.58)	28.82 (4.06)
SNMF	45.10 (0.55)	52.86 (0.82)	55.16 (0.81)	56.04 (1.96)	57.14 (3.12)
DSNMF	60.62 (1.33)	67.73 (1.53)	70.93 (1.70)	70.43 (1.47)	69.45 (3.47)
DDRNMF	52.20 (1.20)	61.17 (1.28)	65.80 (2.04)	69.45 (1.81)	69.87 (3.54)
DNBMF	27.79 (0.82)	34.33 (1.37)	40.92 (1.41)	45.18 (1.63)	53.11 (2.93)
RDNBMF	62.55 (0.78)	75.57 (0.55)	81.05 (0.42)	83.28 (0.87)	83.61 (2.10)
RDNNBMF	**63.60** (0.80)	**76.79** (0.75)	**81.60** (0.87)	**83.66** (0.90)	**86.39** (1.89)

TABLE 5
Mean ACC (%) Versus TN on the CMU Face Images Dataset

TN	5	10	15	20	25	30
PCA	81.89 (2.65)	94.68 (1.56)	96.06 (1.23)	97.13 (1.03)	97.00 (2.00)	94.50 (2.11)
KPCA	90.15 (2.64)	93.82 (1.66)	94.41 (0.90)	95.79 (1.17)	96.07 (1.51)	96.50 (4.09)
DLDA	89.39 (1.50)	94.52 (1.40)	95.68 (1.44)	96.71 (0.90)	96.43 (1.12)	95.75 (3.87)
NMF	93.22 (0.34)	95.80 (0.57)	95.68 (0.52)	96.13 (0.35)	96.57 (1.34)	98.00 (1.58)
MNMF	90.28 (2.30)	93.66 (0.99)	93.97 (1.27)	94.50 (2.08)	93.64 (2.72)	95.50 (4.38)
PNMF	92.17 (0.57)	94.18 (0.57)	95.21 (0.31)	97.08 (0.34)	97.93 (0.22)	97.50 (0.00)
KNMF-RBF	79.46 (5.58)	86.18 (2.85)	88.94 (3.17)	89.50 (2.90)	90.86 (2.30)	91.00 (4.44)
SNMF	77.13 (3.64)	93.66 (1.36)	95.26 (0.64)	96.58 (0.51)	97.29 (0.45)	97.25 (0.79)
DSNMF	81.89 (2.41)	94.68 (1.17)	96.06 (0.70)	**97.13** (0.72)	97.00 (1.11)	94.50 (1.97)
DDRNMF	92.87 (2.10)	95.52 (0.68)	95.56 (1.08)	95.75 (0.92)	96.43 (1.17)	95.75 (2.90)
DNBMF	91.46 (1.08)	92.25 (1.13)	96.03 (0.81)	94.92 (0.67)	97.21 (0.79)	95.00 (0.00)
RDNBMF	91.37 (0.86)	95.00 (0.68)	**96.82** (1.07)	97.04 (0.31)	**97.93** (0.79)	**98.50** (2.11)
RDNNBMF	**94.20** (1.35)	**96.52** (0.67)	96.29 (0.97)	96.50 (1.20)	96.79 (1.59)	97.25 (2.49)

low. The proposed deep nonlinear algorithm RDNNBMF outperforms the shallow nonlinear algorithm PNMF. This demonstrates that deep factorization can obtain the basis image which is more conducive to feature representation and make the algorithm get better recognition performance. The performance of the proposed DNBMF is better than MNMF. This is mainly due to the fact that the deep factorization of DNBMF and MNMF is essentially different. The MNMF algorithm continuously factorizes the coefficient matrix and expresses the original data as $X = W_1 W_2 \cdots W_l H_l$. In fact, the factorization method only has the basis W_1 obtained by the first layer as a local feature with respect to the original sample. The proposed DNBMF is to represent the original data as $X = W_l H_l \cdots H_2 H_1$ by deep factorization of the basis matrix, and each column of W_l is the basis vector of the original sample factorized in the lth layer. The proposed factorization

architecture has clear interpretability and purpose, and can obtain a set of underlying basis images reflecting distinct parts-based features, which is more helpful to classify.

The ROC curves are plotted in the cases of $TN = 10$ and $TN = 20$ on AR face database, and the AUC scores of the corresponding curves, shown in Figs. 5e and 5f. The figures show that the AUC scores of the proposed RDNNBMF are the highest 80.23 and 81.27 percent in the case of $TN = 10$ and $TN = 20$, respectively.

Fig. 6 and Tables 5, 6, and 7 show the experimental results on CMU, Stirling and Yale face databases. Similarly, the proposed linear algorithm DNBMF based on basis images learning outperforms the linear algorithm MNMF based on the deep coefficient matrix factorization in most cases. Meanwhile, the recognition performance of the proposed nonlinear deep factorization algorithm RDNNBMF is better than other
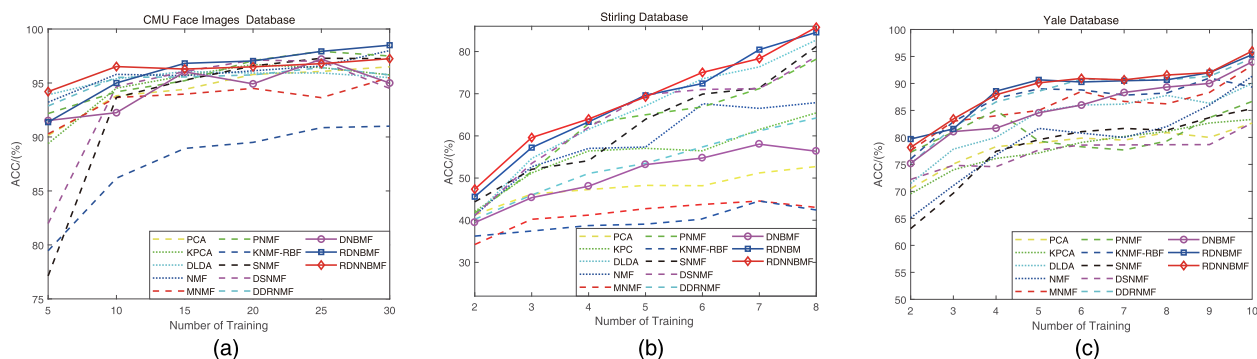

Fig. 6. Recognition results (ACC) on CMU, Stirling, and Yale face databases. The DNBMF, RDNBMF, and RDNNBMF are the proposed algorithms.

TABLE 6
Mean ACC (%) Versus TN on the Stirling Face Dataset

TN	2	3	4	5	6	7	8
PCA	41.26 (3.96)	46.21 (2.77)	47.27 (2.21)	48.26 (4.23)	48.18 (5.11)	51.21 (6.99)	52.73 (8.72)
KPCA	42.08 (3.37)	51.21 (3.85)	56.42 (3.89)	57.05 (7.10)	56.46 (5.58)	61.52 (7.90)	65.45 (6.58)
DLDA	41.17 (2.34)	54.65 (5.28)	61.58 (4.62)	67.12 (4.15)	73.54 (3.95)	76.36 (2.69)	82.73 (5.16)
NMF	41.17 (1.28)	52.68 (1.77)	57.03 (1.41)	57.35 (2.90)	67.39 (3.93)	66.52 (3.81)	67.88 (4.33)
MNMF	34.24 (2.54)	40.71 (4.36)	41.21 (2.32)	42.73 (2.58)	43.74 (6.93)	44.55 (5.31)	43.03 (4.91)
PNMF	41.69 (1.12)	51.52 (1.26)	62.91 (0.94)	65.00 (1.12)	66.87 (3.33)	71.21 (2.57)	78.18 (2.78)
KNMF-RBF	36.23 (6.71)	37.48 (3.67)	38.73 (3.52)	38.20 (3.18)	40.18 (5.36)	44.55 (5.06)	42.42 (10.20)
SNMF	44.37 (2.34)	52.07 (2.13)	54.12 (3.38)	63.94 (3.14)	69.90 (3.43)	71.36 (2.62)	81.21 (2.78)
DSNMF	41.13 (1.21)	53.48 (1.61)	62.00 (1.85)	**69.77** (1.36)	71.01 (5.22)	71.21 (4.95)	78.79 (5.53)
DDRNMF	40.22 (4.49)	45.96 (4.22)	51.09 (5.54)	53.48 (3.96)	57.37 (4.03)	61.21 (6.40)	64.24 (5.11)
DNBMF	39.44 (3.65)	45.35 (4.65)	48.06 (3.74)	53.18 (1.92)	54.75 (5.69)	58.03 (5.35)	56.36 (6.26)
RDNBMF	45.58 (1.73)	57.22 (1.70)	63.33 (1.52)	69.62 (3.18)	72.42 (2.18)	**80.45** (3.15)	84.55 (3.63)
RDNNBMF	**47.36** (1.95)	**59.60** (0.82)	**64.00** (1.57)	69.24 (2.38)	**75.05** (3.06)	78.33 (3.35)	**85.76** (3.44)

TABLE 7
Mean ACC (%) Versus TN on the Yale Face Dataset

TN	2	3	4	5	6	7	8	9	10
PCA	70.52 (3.06)	75.08 (1.78)	78.29 (2.69)	79.00 (2.94)	79.87 (3.17)	79.50 (3.77)	80.89 (6.04)	80.00 (6.09)	82.67 (6.44)
KPCA	69.56 (2.87)	74.00 (3.21)	76.10 (3.58)	77.11 (4.42)	79.07 (3.67)	80.17 (2.28)	81.11 (2.62)	82.67 (6.25)	83.33 (12.27)
DLDA	71.48 (6.52)	77.83 (5.03)	80.00 (5.61)	84.78 (3.01)	86.00 (3.73)	86.17 (3.93)	87.78 (5.26)	83.66 (5.08)	90.00 (7.20)
NMF	65.11 (3.46)	71.08 (2.33)	76.76 (2.55)	81.67 (3.07)	80.80 (2.10)	80.00 (2.60)	82.00 (2.44)	86.00 (4.66)	91.33 (6.32)
MNMF	77.56 (3.29)	83.00 (1.93)	84.00 (2.57)	85.00 (3.36)	88.53 (3.89)	86.67 (4.30)	86.22 (6.00)	88.33 (5.27)	93.33 (5.44)
PNMF	77.26 (1.48)	81.00 (1.40)	85.05 (2.73)	79.33 (7.26)	78.27 (6.19)	77.67 (8.32)	79.33 (3.93)	83.67 (3.67)	86.67 (5.44)
KNMF-RBF	76.00 (4.05)	82.92 (4.54)	87.24 (2.43)	89.00 (3.90)	88.80 (2.28)	87.83 (3.52)	88.22 (4.69)	91.00 (5.89)	89.33 (7.82)
SNMF	63.11 (3.79)	69.67 (3.95)	77.43 (2.88)	80.11 (2.15)	81.07 (3.54)	81.67 (4.23)	81.33 (3.95)	83.67 (3.67)	85.33 (5.26)
DSNMF	72.22 (3.78)	74.83 (3.82)	74.57 (1.42)	77.67 (4.46)	78.33 (2.95)	78.83 (4.65)	78.11 (4.50)	78.67 (5.02)	82.67 (4.66)
DDRNMF	75.11 (5.38)	81.92 (4.07)	86.57 (2.60)	88.56 (3.06)	90.80 (4.51)	90.67 (2.85)	90.67 (4.78)	91.33 (6.52)	94.67 (5.26)
DNBMF	75.19 (3.62)	81.08 (1.93)	81.71 (3.35)	84.56 (2.64)	86.01 (3.94)	88.33 (2.36)	89.33 (3.60)	90.00 (4.97)	94.00 (3.78)
RDNBMF	**79.70** (2.54)	81.58 (2.34)	**88.57** (2.69)	**90.67** (2.04)	90.27 (2.27)	90.51 (2.94)	90.67 (1.75)	92.00 (3.58)	95.33 (3.22)
RDNNBMF	78.15 (3.26)	**83.42** (2.92)	87.90 (3.73)	90.11 (3.37)	**90.93** (4.52)	**90.97** (3.30)	**91.56** (3.89)	**92.01** (5.02)	**96.03** (3.44)

shallow nonlinear algorithms, such as KPCA, PNMF, and KNMF-RBF, on the above three face databases. This shows that the deep factorization algorithm can obtain the basis images which are more suitable for feature representation and it can improve the recognition performance of the algorithm. The best recognition performance on these three databases is also given by the algorithm based on the proposed factorization architecture.

Its worth noting that in some databases with complex changes (such as FERET and AR databases), the algorithm based on the proposed deep factorization architecture has a very significant improvement in recognition performance. The recognition performance of proposed RDNNBMF on the FERET dataset is $3.47 - 9.92$ percent better than the best comparison algorithm with different conditions. The recognition performance of proposed RDNNBMF on the AR face dataset is $4.31 - 5.62$ percent better than the best comparison algorithm with different conditions. This shows that the proposed deep factorization algorithm based on the UBIL is more conducive to the face recognition task with complex interference.

In general, the proposed deep factorization algorithms based on the UBIL outperform the deep factorization algorithms based on coefficient matrix. This reveals that the deep NMF architecture based on the UBIL is more beneficial for the improvement of recognition performance. The algorithm can obtain the true underlying basis images which can reflect the local characteristics of the samples by deep factorization on the basis images matrix. The feature of the sample responding to these bases can more accurately reflect the

difference in local characteristics. Meanwhile, the proposed RDNBMF with basis images regularization constraint outperforms the proposed DNBMF without basis images regularization constraint. This also shows that basis images regularization constraint is beneficial for feature extraction. In most cases, the RDNBMF algorithm with the regularization constraint in the basis images matrix performs better than the DDRNMF algorithm with the regularization constraint in the coefficient matrix. It can be observed that the basis images constrained and optimized directly are more conducive to the recognition task than those constrained and optimized indirectly. These results validate that the proposed deep NMF architecture based on the UBIL is more reasonable in recognition tasks, and it is necessary to improve the recognition performance of the algorithm.

6.3 Experiments on Different Number of Underlying Basis Images

The effect on the number of basis images is analyzed in this section. The experiment is performed on Yale, FEREF and Stirling face databases with different numbers of basis images. In this experiment, all the other parameters are set the same except the number of underlying basis images r. Fig. 7 shows the recognition results of the proposed algorithms on those three face databases in the case of differents number of basis images. ACC scores are shown in Tables 8, 9, and 10, and the mean square deviation (MSD) of 10 times experiments is given in parentheses. The experimental results show that the recognition performance of the algorithm is first enhanced then weakened with the increase of
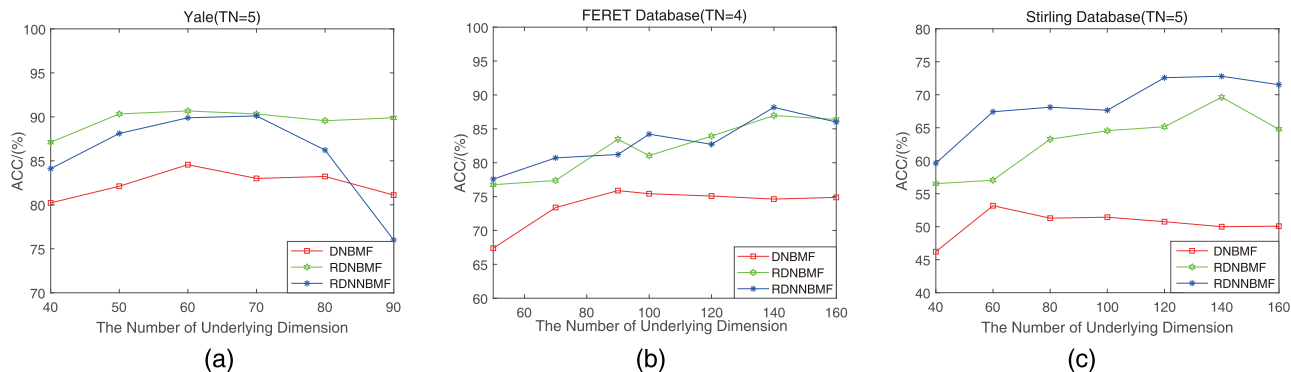
Fig. 7. The recognition results of the proposed algorithms on different numbers of basic images.

the numbers of underlying basis images. Because few underlying basis images are not sufficient to represent the original images, the recognition performance of the algorithm is weak. However, too many underlying basis images lead to excessively detailed local feature representation, resulting in redundancy in the extracted features that affects recognition performance relatively. Therefore, it is advantageous to select an appropriate number of underlying basis images for recognition performance.

In addition, it can be seen from the experimental results that on the Yale face database, as shown in Table 8 and Fig. 7a, when the number of underlying basis images is the same, the performance of the proposed RDNBMF algorithm is better in most cases. The best recognition performance of the nonlinear method RDNNBMF is 90.11 percent (r = 60) and the best recognition performance of the RDNBMF algorithm is 90.67 percent (r = 70), which are basically equal. On the FERET database, when the number of underlying basis images is the same, the proposed nonlinear method RDNNBMF has better recognition performance in most cases, as shown in Table 9 and Fig. 7b. The best recognition

result of RDNNBMF is 88.17 percent, which is the best performance among the three algorithms on the FERET database. On the Stirling database, while the number of underlying basis images is the same, the recognition performance of the proposed nonlinear algorithm RDNNBMF is better than the other two algorithms, as shown in Table 10 and Fig. 7c. The above results are mainly due to the different characteristics of the data in the three databases.

On the Yale face database, the change of the face images in angle and occlusion is very weak—only the illumination and the slight facial expression change. Because the data in this database is not very complicated, the basis images extracted by the linear method can better describe the common local characteristics of the samples. The linear RDNBMF method can obtain the recognition performance similar to the nonlinear method RDNNBMF. In addition to changes in the illumination and expression changes, the face images in the FERET database have change in posture and age. The images in this database are more complex than those in the Yale face database. The linear approach to dealing with identification problems on FERET database is not better than nonlinear methods generally. In the Stirling database, there is a sharp change in the angle of the images, which makes the linear method difficult to extract the local features of the samples. The nonlinear method has better recognition performance than the linear method on this data.

The proposed RDNBMF algorithm performs better than the DNBMF in almost all cases, which validates that the basis images obtained by the algorithm restraining the basis images directly is more beneficial to improve the recognition performance. It can be noticed from the experimental results that when the number of underlying basis images is within a certain range, the recognition rate fluctuates within

TABLE 8
Mean ACC (%) Versus the Number of Underlying Basis Images on Yale Face Dataset (TN=5)

r	40	50	60	70	80	90	100
DNBMF	80.22 (3.17)	82.11 (3.80)	84.56 (2.51)	83.00 (3.48)	83.22 (4.11)	81.11 (4.47)	78.67 (3.08)
RDNBMF	87.11 (2.86)	90.33 (1.80)	**90.67** (1.94)	90.33 (1.99)	89.56 (1.59)	89.89 (2.70)	88.33 (2.73)
RDNNBMF	84.11 (4.39)	88.11 (2.05)	89.89 (3.80)	90.11 (3.20)	86.22 (6.96)	76.00 (5.82)	65.67 (2.19)

TABLE 9
Mean ACC (%) Versus the Number of Underlying Basis Images on FERET Face Dataset (TN=4)

r	50	70	90	100	120	140	160
DNBMF	67.38 (2.47)	73.38 (1.70)	75.88 (2.00)	75.42 (2.64)	75.08 (2.87)	74.63 (2.18)	74.88 (2.47)
RDNBMF	76.75 (1.50)	77.38 (1.50)	83.46 (1.13)	81.04 (1.29)	83.92 (0.77)	86.96 (0.97)	86.38 (1.22)
RDNNBMF	77.58 (0.71)	80.71 (0.96)	81.21 (1.02)	84.21 (0.73)	82.71 (0.79)	**88.17**	86.00 (1.10) (0.77)

TABLE 10
Mean ACC (%) Versus the Number of Underlying Basis Images on Stirling Face Dataset (TN=5)

r	40	60	80	100	120	140	160
DNBMF	46.21 (3.35)	53.18 (1.82)	51.29 (3.59)	51.44 (4.75)	50.76 (5.70)	50.00 (4.05)	50.08 (2.84)
RDNBMF	56.52 (1.63)	57.05 (2.94)	63.26 (1.86)	64.55 (2.11)	65.15 (2.85)	69.62 (3.02)	64.77 (3.07)
RDNNBMF	59.62 (2.01)	67.42 (3.12)	68.11 (3.66)	67.65 (1.44)	72.58 (1.62)	**72.80** (2.26)	71.52 (1.70)

(a) NMF

(b) DSNMF

(c) The proposed DNBMF

(d) The proposed RDNBMF

(e) The proposed RDNNBMF (the first layer)

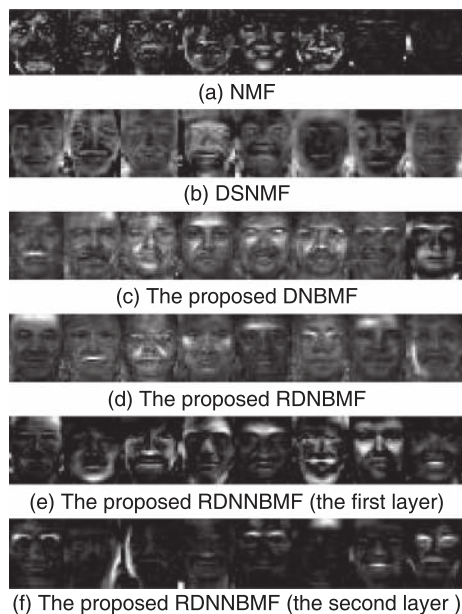(f) The proposed RDNNBMF (the second layer)

Fig. 8. The basis images learned by NMF, DSNMF, DNBMF, RDNBMF, and RDNNBMF.

a small range. The performance of the proposed algorithm is relatively stable, and the proposed algorithm has better stability when the parameter is within a certain range of variation.

6.4 Analysis on the Basis Images

In order to observe the basis images obtained by the proposed algorithms based on the UBIL intuitively, we present the basis images learned by NMF, DSNMF and DNBMF, RDNBMF and RDNNBMF, and make an intuitive comparison. The basis images learned by NMF, DSNMF and DSNMF, DNBMF, and RDNBMF are shown in Figs. 8a, 8b, 8c, and 8d, and the basis images learned by the first layer of RDNNBMF are shown in Fig. 8e. Fig. 8f shows the basis images learned by the second layer of RDNNBMF.

It can be seen from Fig. 8a that the basis images obtained by the NMF algorithm cannot clearly show the contour of the face, and there is no obvious local feature. This is due to the fact that only the shallow factorization is done in the NMF algorithm and there is no special regularization constraint on the basis images matrix or the coefficient matrix. Therefore, the obtained basis images are difficult to reflect the local characteristics of the samples as a set of bases representing the samples.

Although the base image extracted by the DSNMF algorithm shown in Fig. 8b can clearly display the contour of the face, the extracted local features are still not obvious. The underlying basis images obtained by the proposed linear algorithm DNBMF show a clearer face contour than NMF and DSNMF. The more obvious local features of faces (such as teeth, forehead, eyebrows, etc.) in Fig. 8c. Fig. 8d illustrates that the basis images extracted by the proposed linear algorithm RDNBMF with the regularization constraint on basis images has more prominent local characteristics than the basis images extracted by DNBMF. Fig. 8e shows that the characteristics of the basis images that has been factorized by shallow factorization are not localized enough. The

basis images deep extracted by the proposed nonlinear algorithm RDNNBMF shown in Fig. 8f have more accurate face contours and can display more prominent features of various parts of the face.

This experiment shows that the proposed deep NMF algorithm architecture based on the UBIL can obtain the basis images with more prominent local characteristics with the help of the regularization constraints.

7 CONCLUSION

This paper proposes a novel deep NMF architecture based on the UBIL and applies it to face recognition tasks. Because the NMF-based algorithm uses the Moore-Penrose pseudo-inverse of the basis images matrix to map the original samples and obtain new features of the sample, the basis images matrix is very important in the NMF-based algorithm. The proposed deep NMF architecture performs deep factorization on the basis images matrix and obtains the underlying basis images matrix which can reflect the local characteristics of the samples. It can thusly obtain a more accurate sample representation. There is a strong interpretability that contains the actual meaning in the factorization process of each layer. To implement the proposed deep NMF algorithm based on the UBIL, this paper proposes three algorithms. The linear algorithm DNBMF continuously factorizes the basis images matrix until the underlying basis images matrix is obtained. To make a large difference between the factorized basis images in each layer, this paper proposes the RDNBMF which adds a regularization term to the objective function and directly constrains the basis images matrix. Then a nonlinear RDNNBMF is proposed. The RDNNBMF can adapt to the recognition problem with more complex data and at the same time make the algorithm conform to the idea that the human visual system adopts a hierarchical and non-linear method when analyzing images. In this paper, the convergence of the algorithm is proved theoretically. Finally, the paper experiments on 6 public face data sets, namely FERET, ORL, AR, CMU, Stirling and Yale databases. The three proposed algorithms are compared with the other 10 currently popular algorithms. The ACC, ROC curve and AUC values are used to verify the performance of the three proposed algorithms based on the proposed deep NMF architecture. Experimental results show that the deep NMF algorithm performs better than the shallow NMF algorithm. The deep NMF architecture based on the UBIL proposed in this paper has better recognition performance than the existing deep factorization architecture based on the coefficient matrix. In addition, the underlying basis images obtained by the algorithm are analyzed. It is observed that the deep NMF algorithm architecture proposed in this paper can obtain the basis images with more local features. In general, the algorithm based on the proposed deep NMF architecture can obtain better recognition performance.

Due to the important role of the basis images in the NMF-based algorithm, our future work will continue to focus on the optimization constraints of the basis images matrix. In addition, because the basis images of each class of samples may be different, we will also continue our research on supervised deep NMF algorithms based on UBIL.

ACKNOWLEDGMENTS

REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. 13th Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[3] T. Shi, K. Kang, J. Choo, and C. K. Reddy, "Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations," in *Proc. World Wide Web Conf.*, 2018, pp. 1105–1114.

[4] V. Gligorijević, Y. Panagakis, and S. Zafeiriou, "Non-negative matrix factorizations for multiplex network analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 928–940, Apr. 2018.

[5] B. Sabzalian and V. Abolghasemi, "Iterative weighted non-smooth non-negative matrix factorization for face recognition," *Int. J. Eng.*, vol. 31, no. 10, pp. 1698–1707, 2018.

[6] D. P. Varikuti *et al.*, "Evaluation of non-negative matrix factorization of grey matter in age prediction," *NeuroImage*, vol. 173, pp. 394–410, 2018.

[7] Z. Chen, Y. Shi, and Z. Qi, "Constrained matrix factorization for semi-weakly learning with label proportions," *Pattern Recognit.*, vol. 91, pp. 13–24, 2019.

[8] H. Kim, J. Piven, and G. Gerig, "Longitudinal structural connectivity in the developing brain with projective non-negative matrix factorization," in *Proc. Med. Imag.: Image Process.*, 2019, Art. no. 109490Q.

[9] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Non-negative matrix factorization with locality constrained adaptive graph," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2019.2892971.

[10] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, no. Nov., pp. 1457–1469, 2004.

[11] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 403–415, Mar. 2006.

[12] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2010.

[13] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

[14] F. Rousset, F. Peyrin, and N. Ducros, "A semi nonnegative matrix factorization technique for pattern generalization in single-pixel imaging," *IEEE Trans. Comput. Imag.*, vol. 4, no. 2, pp. 284–294, Jun. 2018.

[15] J. Shawe-Taylor *et al.*, *Kernel Methods for Pattern Analysis*. Cambridge, UK: Cambridge Univ. Press, 2004.

[16] I. Buciu, N. Nikolaidis, and I. Pitas, "Nonnegative matrix factorization in polynomial feature space," *IEEE Trans. Neural Netw.*, vol. 19, no. 6, pp. 1090–1100, Jun. 2008.

[17] S. Zafeiriou and M. Petrou, "Nonlinear non-negative component analysis algorithms," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1050–1066, Apr. 2010.

[18] W.-S. Chen, Y. Zhao, B. Pan, and B. Chen, "Supervised kernel non-negative matrix factorization for face recognition," *Neurocomputing*, vol. 205, pp. 165–181, 2016.

[19] D. Tolić, N. Antulov-Fantulin, and I. Kopriva, "A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering," *Pattern Recognit.*, vol. 82, pp. 40–55, 2018.

[20] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neurosci.*, vol. 2, no. 11, pp. 1019–1025, 1999.

[21] E. Goceri, "Skin disease diagnosis from photographs using deep learning," in *Proc. Thematic Conf. Comput. Vision Med. Image Process.*, 2019, pp. 239–246.

[22] E. Goceri, "Analysis of deep networks with residual blocks and different activation functions: Classification of skin diseases," in *Proc. IEEE 9th Int. Conf. Image Process. Theory Tools Appl.*, 2019, pp. 1–6.

[23] S. Wisdom, T. Powers, J. Pitton, and L. Atlas, "Deep recurrent NMF for speech separation by unfolding iterative thresholding," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2017, pp. 254–258.

[24] Y. Liu, N. Guan, and J. Liu, "Deep transductive nonnegative matrix factorization for speech separation," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl.*, 2017, pp. 249–254.

[25] F. Ye, C. Chen, and Z. Zheng, "Deep autoencoder-like nonnegative matrix factorization for community detection," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1393–1402. [Online]. Available: http://doi.acm.org/10.1145/3269206.3271697

[26] H. Fang, A. Li, T. Wang, H. Chang, and H. Xu, "Hyperspectral unmixing using graph-regularized and sparsity-constrained deep NMF," in *Proc. Opt. Sens. Imag. Technol. Appl.*, 2018, Art. no. 108462J.

[27] A. Imakura, Y. Inoue, T. Sakurai, and Y. Futamura, "Parallel implementation of the nonlinear Semi-NMF based alternating optimization method for deep neural networks," *Neural Process. Lett.*, vol. 47, no. 3, pp. 815–827, 2018.

[28] Z. Li, J. Tang, and T. Mei, "Deep collaborative embedding for social image understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2070–2083, Sep. 2019.

[29] J.-H. Ahn, S. Kim, J.-H. Oh, and S. Choi, "Multiple nonnegative-matrix factorization of dynamic pet images," in *Proc. Asian Conf. Comput. Vis.*, 2004, pp. 1009–1013.

[30] A. Cichocki and R. Zdunek, "Multilayer nonnegative matrix factorisation," *Electron. Lett.*, vol. 42, no. 16, pp. 947–948, 2006.

[31] A. Cichocki and R. Zdunek, "Multilayer nonnegative matrix factorization using projected gradient approaches," *Int. J. Neural Syst.*, vol. 17, no. 6, pp. 431–446, 2007.

[32] H. A. Song, B.-K. Kim, T. L. Xuan, and S.-Y. Lee, "Hierarchical feature extraction by multi-layer non-negative matrix factorization network for classification task," *Neurocomputing*, vol. 165, pp. 63–74, 2015.

[33] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[34] M. Tong, Y. Chen, M. Zhao, H. Bu, and S. Xi, "A deep discriminative and robust nonnegative matrix factorization network method with soft label constraint," *Neural Comput. Appl.*, vol. 31, pp. 7447–7475, 2018.

[35] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3840–3853, Nov. 2017.

[36] L. Zong, X. Zhang, L. Zhao, H. Yu, and Q. Zhao, "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," *Neural Netw.*, vol. 88, pp. 74–89, 2017.

[37] Y. X. Wang and Y. J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1336–1353, Jun. 2013.

[38] E. Goceri and N. Goceri, "Deep learning in medical image analysis: Recent advances and future trends," in *Proc. 11th Int. Conf. Comput. Graphics Vis. Comput. Vis. Image Process.*, 2017, pp. 305–311.

[39] E. Goceri, "Formulas behind deep learning success," in *Proc. Int. Conf. Appl. Anal. Math. Model.*, 2018, Art. no. 156.

[40] E. Goceri, "Challenges and recent solutions for image segmentation in the era of deep learning," in *Proc. 9th Int. Conf. Image Process. Theory Tools Appl.*, 2019, pp. 1–6.

[41] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[42] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.

[43] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[44] A. Martínez and R. Benavente, "The AR face database," Computer Vision Center, no. 24, Jun. 1998. [Online]. Available: http://www.cat.uab.cat/Public/Publications/1998/MaB1998

[45] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[46] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[47] H. Yu and J. Yang, "A direct lda algorithm for high-dimensional data with application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.

[48] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[49] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, 2018. [Online]. Available: https://doi.org/10.1016/j.aci.2018.08.003

[50] B. Pan, J. Lai, and W. S. Chen, "Nonlinear nonnegative matrix factorization based on mercer kernel construction," *Pattern Recognit.*, vol. 44, no. 10, pp. 2800–2810, 2011.
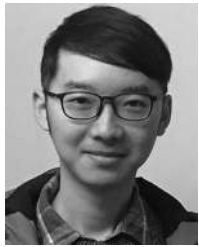
Yang Zhao received the BSc degree in applied mathematics from Huanggang Normal University, China, in 2012. He received the MSc degree in applied mathematics, Shenzhen University, China, in 2016. He is currently working toward the PhD degree in the College of Electronics and Information Engineering, Shenzhen University, China. His research interests include pattern recognition and image processing.

Huiyang Wang received the BSc degree in mechanical design manufacture and automation from Jiangxi Agricultural University, China, in 2017. He is currently working toward the MSc degree in the College of Mechatronics and Control Engineering, Shenzhen University, China. His research interests include artificial intelligence and image processing.

Jihong Pei received the BS degree in electronics engineering from Beihang University, China, in 1989, the MS degree in physical electronics and optoelectronics, in 1994, and the PhD degree in signal and information processing, in 1998, both from, Xidian University. He is currently a professor with Shenzhen University and the director of the Electronics Engineering Department at CIE, Shenzhen University. His research interests include intelligent information processing and pattern recognition. He is a member of the IEEE.