Robust Image Feature Extraction via Approximate Orthogonal Low-Rank Embedding

CONG FU1,2, ZHIGUI LIU2, AND LI LI2

1Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China
2College of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China

Corresponding author: Zhigui Liu (liuzhigui@swust.edu.cn)

This work was supported by the China Aerodynamic Research Center Support Program under Grant 18zh0270.

ABSTRACT Feature extraction (FE) plays an important role in machine learning. In order to handle the “dimensionality disaster” problem, the usual approach is to transform the original sample into a low-dimensional target space, in which the FE task is performed. However, the data in reality will always be corrupted by various noises or interfered by outliers, making the task of feature extraction extremely challenging. Thence, we propose a novel image FE method via approximate orthogonal low-rank embedding (AOLRE), which adopts an orthogonal matrix to reserve the major energy of the samples, and the introduction of the $\ell_2,1$-norm makes the features more compact, differentiated and interpretable. In addition, the weighted Schatten $p$-norm is adopted in this model for fully exploring the effects of different ranks while approaching the original hypothesis of low rank. Meanwhile, as a robust measure, the correntropy is applied in AOLRE. This can effectively suppress the adverse influences of contaminated data and enhance the robustness of the algorithm. Finally, the introduction of the classification loss item allows our model to effectively fit the supervised scene. Five common datasets are used to evaluate the performance of AOLRE. The results show that the recognition accuracy and robustness of AOLRE are significantly better than those of several advanced FE algorithms, and the improvement rate ranges from 2% to 15%.

INDEX TERMS Image feature extraction, machine learning, low-rank embedding, orthogonal matrix.

I. INTRODUCTION

As we all know, image processing technology has been applied to various areas of our lives, such as face recognition, image analysis, document clustering and medical diagnosis. By denoising, reducing noise, normalizing, enhancing, segmenting, extracting features, and registering the image, the image will be easier to observe and analyze, and the quality of the image can be significantly improved, which can facilitate the researcher to perform the image follow-up related operations. Feature extraction (FE), as the core link in image processing, plays a vital role in the final image processing result. Image feature extraction technology not only creates new scientific and technological achievements, but also greatly improves the accuracy and accuracy of digitization. The continuous optimization of data acquisition method makes the dimensionality of the collected samples rise sharply. While bringing more detailed feature information, irrelevant and redundant features also follow, which greatly weakens the inherent structural relationship between features, and ultimately reduces the algorithm performance. Therefore, how to find high-quality low dimensional representation features from massive high dimensional data has become a attention focus of researchers [1], [2]. In recent years, some scholars have carried out a lot of research on the application of FE. For example, there are many feature extraction algorithms used for their ability in terms of energy compaction property such as orthogonal polynomials [3]–[5]. This study focuses on the design of feature extraction algorithm, so the application of FE will not be further described here.

As a typical unsupervised FE method, the principal component analysis (PCA) [6] achieves the optimal data reconstruction by reducing the dimensionality of the data. Of course, another typical supervised FE method is linear discriminant analysis (LDA) [7], whose discriminative projection is obtained by setting different degrees of inter-class and intra-class dispersion. Unfortunately, both

The associate editor coordinating the review of this manuscript and approving it for publication was Paolo Remagnino.
methods assume that the data obeys a Gaussian distribution. Therefore, extended versions of these two methods have been proposed one after another [8]–[12].

Over the years, with the continuous development of manifold learning theory, many local-based FE methods have been proposed. For example, locality preserving projections (LPP) first finds an embedding to retain local structure, then get a subspace that can better detect the basic manifold structure [13]. Some of its improved versions have also been proposed [14]–[16]. In addition, local neighborhood reconstruction information is also used for feature extraction, such as neighborhood preserving embedding [17] and its improved versions [18], [19]. In general, if the target image satisfies the manifold constraint, these methods can achieve good performance. The core task of these methods is to model the structure of the data by constructing a graph. However, the actual data is often contaminated and corroded due to the interference of the external environment when acquiring, which makes the distance between samples deviate from the true level. At this time, the samples’ geometric connection reflected by the constructed graph has deviated from the real situation, so the effectiveness of the aforementioned methods may be extremely reduced. Therefore, improving the robustness of image FE is still a challenging task.

Lately, representation-based FE methods have received widespread attention [20]–[25], and the most typical ones are sparse representation (SR) and low-rank representation (LRR). As a general classification algorithm, SR classification (SRC) [20] utilizes the sparsity in the recognition problem to efficiently improve the effect of face-recognition. The key to this method is whether there are enough features and whether the sparse representation can be accurately obtained. Unfortunately, in the face of corrupt data, SRC often fails to achieve the desired results. Xu et al. proposed a two-phase test sample SR method, which makes coarse-to-fine classification decisions, and finally achieves good face recognition [22]. However, these methods are not expert in seizing the unique global information of samples. For this reason, many FE methods based on LRR have been proposed one after another [23], [24], [26]. These methods assume that samples of the same kind are gathered in the same subspace, so that LRR may easily capture the global features, thereby improving the robustness of FE. SR is expert in seizing the local information of the sample and LRR is expert in seizing the global information. Inspired by this, a method combining SR and LRR is proposed [27]. Although LRR has been widely used, it still has some limitations [28].

As mentioned earlier, traditional FE methods and representation-based FE methods have their own advantages and disadvantages. This makes many researchers aim at the fusion of the two types of methods to gain their respective advantages and overcome their respective disadvantages. Many advanced methods have been proposed one after another, which has promoted the continuous development of feature learning theory. For example, sparse PCA (SPCA) [29], robust PCA (RPCA) [30], latent LRR (LatLRR) [31], Collaboration-based method (CRC) [32], structured LRR for image classification (LRRC) [33], extended version of LatLRR (SALPL) [34]. Especially, in LatLRR and SALPL, the data is separated into two spaces from the perspective of rows and columns, and the data information is reconstructed by optimizing two low-rank matrices. As a recently proposed superior FE method, low-rank embedding (LRE) [35] can complete the latent embedding subspace search task while obtaining the optimal low-rank representation. However, this method still has some defects in data reconstruction and noise separation, and as an unsupervised method, its application range is also limited.

In addition, the aforementioned LRR-based methods deal with low-rank problems by using nuclear norm for convex relaxation of optimization targets, and it’s almost impossible to get the optimal solution [36]. Hence, the Schatten norm (SN) is applied to enforce low-rank regularization [37], [38]. To better approach the optimal solution, Brbic and Kopriva [39] introduced the Schatten-0 ($S_0$) and $\ell_0$-regularized objective. However, these models treat all singular values the same, ignoring the difference in rank contribution. Recently, to treat each rank component differently, a weighted method is proposed to deal with low-rank problems (WNMM) [40]. On this basis, to make better use of the low-rank property, a weighted SN method (WSNM) is proposed [38], [41], [42].

Encouraged by LRE, to ameliorate the shortcomings of existing FE methods to get more excellent features, we then develop a novel FE algorithm via approximate orthogonal low-rank embedding (AOLRE). Summarized as follows:

- In our AOLRE model, the data reconstruction term and the regular term are cleverly combined. Specifically, by designing an orthogonal matrix firstly to ensure that a good principal component is preserved in the embedding space. In addition, the $\| \cdot \|_{2,1}$ is applied to constrain the embedded projection matrix $F$ to guarantee that the differences of obtained features are significant. Third, the introduction of low-rank constraint $\| \cdot \|_{p,Sp}$ enables AOLRE to superiorly seize the unique internal global information of samples, so as to recover the damaged data as much as possible. Finally, as a robust measure, the correntropy is applied in this model to better handle non-Gaussian noise.
- Our model introduces a classification loss item. In the optimization process, it can masterly use the label information to continuously obtain the most favorable information to classify.
- To enhance the solution speed of AOLRE, a specialized algorithm (HQ-ADMM) is designed by coalescing the alternating direction method of multipliers (ADMM) [43] and half-quadratic (HQ) techniques [44].
- Sufficient experimental results confirm that the effectiveness, robustness and reconstruction ability of this method are superior to some existing methods, especially in the case of noise and damaged observations, the advantages of our model are more obvious.
The remaining article content is divided into four Sections. Some related works are briefly reviewed in Section II. Section III presents our model and gives the optimization process. Sufficient experimental results and analysis are given in Section IV. Section V gives the conclusion of this work.

II. RELATED WORKS

Here, some related works are briefly reviewed, i.e., LRR-based FE, weighted Schatten p-norm and correntropy. Some notations used in our work are listed in Tab. 1.

| Notation | Definition                  |
|----------|----------------------------|
| 1        | All-one column vector      |
| I        | Identity matrix            |
| X ∈ R^{m×n} | Data matrix              |
| E ∈ R^{m×n} | Error matrix             |
| F ∈ R^{m×d} | Projection matrix        |
| J ∈ R^{n×n} | Auxiliary matrix         |
| L ∈ R^{n×n} | Low-rank matrix           |
| P ∈ R^{m×d} | Reconstruction matrix    |
| Y_1, Y_2 | Lagrange multipliers      |
| α, β, γ  | Balance parameters        |

A. LRR-BASED FE

First we review LatLRR [31], and then introduce LRE [35].

As a recently proposed LRR-Based FE method, LatLRR can effectively handle the situation where the quantity of samples is insufficient or severely damaged. In LatLRR, the data information is reconstructed by optimizing two low-rank matrices \( L_c \) and \( L_r \). For data \( X \in \mathbb{R}^{m\times n} \), LatLRR’s mathematical expression is

\[
\min_{L_c, L_r, E} \| L_c \|_* + \| L_r \|_* + \lambda \| E \|_1 \quad \text{s.t.} \quad X = XL_c + L_r X + E
\]  

(1)

where \( E \in \mathbb{R}^{m\times n} \) denotes the sparse error.

Obviously, there is no response between \( L_c \) and \( L_r \) in (1), so the obtained result may not be optimal, which will make its recognition performance unpredictable.

In LRE [35], the optimal LRR and the optimal subspace can be sought at the same time. In particular, first an orthogonal matrix \( F \) is introduced, and then \( L \) is obtained on the orthogonal subspace, and then the data is reconstructed. Its mathematical expression is,

\[
\min_{L, F} \| L \|_* + \lambda \| F^T X - F^T X L \|_{2,1} \quad \text{s.t.} \quad F^T F = I
\]  

(2)

In fact, when data is converted from one space to another, there must be a loss of information. As shown in (2), only measuring the data reconstruction ability in embedding space is obviously not perfect. In other words, LRE neglects the loss of information during space conversion.

B. THE WEIGHTED SCHATTEN P-NORM

Given \( X \in \mathbb{R}^{m\times n} \), its weighted Schatten p-norm is,

\[
\| X \|_w^p = \left( \sum_{i=1}^{\min(m, n)} w_i \delta_i^p \right)^{1/p}
\]  

(3)

where \( 0 < p \leq 1 \), \( w = [w_1, ..., w_{\min(m, n)}] > 0 \), and \( \delta_i \) is the \( i \)-th largest singular value. Usually we will use the following form,

\[
\| X \|_w^p = \sum_{i=1}^{\min(m, n)} w_i \delta_i^p = tr(W \Delta^p)
\]  

(4)

where both \( W \) and \( \Delta \) are diagonal matrices composed of \( w_i \) and \( \delta_i \), respectively.

C. CORRENTROPY

In [44], for two vectors \( a \) and \( b \), their correntropy is,

\[
V(a, b) = E[\exp(-\frac{|a - b|^2}{2\sigma^2})]
\]  

(5)

where \( E[\cdot] \) denotes the expectation. Based on this, the Correntropy Induced Metric (CIM) [44] is,

\[
CIM(a, b) = (1 - V(a, b))^\frac{1}{2} = \left( 1 - \frac{1}{N} \sum_{i=1}^{N} k_\sigma(a_i, b_i) \right) \frac{1}{2}
\]  

(6)

Thus, if we assume that \( c = a - b \), the similarity between \( a \) and \( b \) can be measured by the correntropy in the form of \( \sum_i k_\sigma(c_i) \), where \( c_i \) denotes the \( i \)-th element of variable \( c \).

In practice, we usually use the Maximum Correntropy Criterion (MCC) [44] to equivalent minimizing CIM. In [45], the performance of MCC is obviously better than Minimum Mean Square Error (MMSE), which confirms the robustness of MCC. In addition, we compared the effects of several loss functions, including absolute error (\( \ell_1 \)-norm), MSE (\( \ell_2 \)-norm) and CIM. Fig. 1 shows that CIM can handle large errors better. Hence, the correntropy will be used in the proposed model.

III. APPROXIMATE ORTHOGONAL LOW-RANK EMBEDDING

Next, we first give the derivation process of model AOLRE, and then develop an efficient optimization algorithm for it.

A. PROBLEM FORMULATION

As mentioned earlier, as an embedding-based method, LRE shows good performance in FE tasks, but it also has some shortcomings. In this regard, the goal of this work is to propose a FE method that can maintain the advantages of LRE while overcoming its shortcomings. As we all know, in the process of obtaining real data, it is always unavoidable to avoid noise pollution caused by equipment or the environment. This pushes us to introduce a noise matrix \( E \in \mathbb{R}^{m\times n} \) in the model. By comparing (2), we can get

\[
F^T X = F^T X L + E
\]  

As a robust measure, the correntropy is applied into our model, i.e., the term \( \sum_{i,j} \psi(E_{ij}) \), where
the principal components in the data are easier to obtain by the reconstruction, then introduce an adscititious matrix $\Phi$ into the target space. This deviates from the original intention of C. Fu et al. [193229] VOLUME 8, 2020

In (7) clearly shows that $F^\top X = F^\top XL + E$, $F^\top F = I$

where $F \in \mathbb{R}^{m \times d}$ is an orthogonal transformation matrix (or projection matrix).

Since the purpose of introducing $E$ is to model the noise in original space, but the constraint $F^\top X = F^\top XL + E$ in (7) clearly shows that $E$ is used to model the data noise in the target space. This deviates from the original intention of separating noise. To prevent this circumstance, we specially introduce an adscititious matrix $F \in \mathbb{R}^{m \times d}$ to perform data reconstruction, then $F$ can learn features more flexibly. So, the principal components in the data are easier to obtain by the embedding representation (i.e. $F^\top X$). Therefore, the target formula can be re-expressed as,

$$\min_{L.E.F.P} \|L\|_{w,S_p}^p + \alpha \sum_{i,j} \varphi(E_{i,j})$$

s.t. $X = PF^\top X + E$, $F^\top X = F^\top XL$, $F^\top F = I$ (8)

Analyzing problem (8), we can see that the term $F^\top X = F^\top XL$ in its constraint condition is exactly equivalent to the latent LRE. Then, mathematically, we can simplify (8) to

$$\min_{L.E.F.P} \|L\|_{w,S_p}^p + \alpha \sum_{i,j} \varphi(E_{i,j})$$

s.t. $X = PF^\top XL + E$, $F^\top F = I$ (9)

The orthogonal constraint $F^\top F = I$ is introduced in the model to avoid trivial solutions. Through analysis, substituting $P^\top P = I$ for $F^\top F = I$ can have the same effect, and the obtained $F$ and $P$ are more accurate.

Furthermore, given any sample $x_i$, its projected result $y_i = F^\top x_i$. Exactly,

$$(y_i | F^\top x_i) = \begin{bmatrix} y_{i,1} & f_{1,1} x_{i,1} & \cdots & f_{m,1} x_{i,m} \\ y_{i,2} & f_{1,2} x_{i,1} & \cdots & f_{m,2} x_{i,m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,d} & f_{1,d} x_{i,1} & \cdots & f_{m,d} x_{i,m} \end{bmatrix}$$ (10)

where $(y_i | F^\top x_i)$ is the augmented matrix of $y_i$ and $F^\top x_i$.

The analysis shows that encouraging the sparsity of the rows of $F$ can help to obtain a good embedding representation $F^\top X$. Considering the characteristics of $\ell_2$-norm [46], we introduce the $\|F\|_{2,1}$ into the model, and the optimization problem is transformed into

$$\min_{L.E.F.P} \alpha \sum_{i,j} \varphi(E_{i,j}) + \beta \|F\|_{2,1}$$

s.t. $X = PF^\top XL + E$, $P^\top P = I$ (11)

where $\alpha > 0$ and $\beta > 0$ are balance parameters, and $\|F\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{d} F_{i,j}}$.

Furthermore, the known label information can obviously help extract more effective features. To this end, we expect to introduce a special discriminant function $\Phi(X, F)$ into the model.

$$\min_{L.E.F.P} \alpha \sum_{i,j} \varphi(E_{i,j}) + \beta \|F\|_{2,1} + \gamma \|L\|_{w,S_p}^p + \Phi(X, F)$$

s.t. $X = PF^\top XL + E$, $P^\top P = I$ (12)

where $\gamma > 0$ is a parameter to balance term $L$.

Generally, in the learning process, the stronger the resolving power of the projection matrix, the better the classification effect. Specifically, for $n$ training samples $X = [x_1, x_2, \cdots, x_n] \in \mathbb{R}^{m \times n}$ from $c$ classes, we will learn its zero-one label matrix $B = [b_1, b_2, \cdots, b_n] \in \mathbb{R}^{c \times n}$. In the training phase of supervised FE, linear classifier is usually used to process sample features. In this paper, a linear classifier is adopted for $F^\top X$. Consequently, using multivariate rigid regression, we can get $\Phi(X, F) = \frac{1}{2} \|B - F^\top X\|_F^2$. Then, the final expression of AOLRE is,

$$\min_{L.E.F.P} \alpha \sum_{i,j} \varphi(E_{i,j}) + \beta \|F\|_{2,1} + \gamma \|L\|_{w,S_p}^p + \frac{1}{2} \|B - F^\top X\|_F^2$$

s.t. $X = PF^\top XL + E$, $P^\top P = I$ (13)

B. SOLUTION FOR AOLRE

Considering the particularity of the model, we design an effective algorithm HQ-ADMM to improve the efficiency of solving our model by coalescing ADMM [43] and HQ [44]. In [47], the convergence of ADMM for non-convex problems is analysed. For convenience, we introduce an adjuvant $J$ to get the following mathematical formula,

$$\min_{L.E.F.I.P} \alpha \sum_{i,j} \varphi(E_{i,j}) + \beta \|F\|_{2,1} + \gamma \|J\|_{w,S_p}^p$$
+ \frac{1}{2} \| B - F^T X \|_F^2 \\
 s.t. \ X = PF^T XL + E, \ P^T P = I, \ L = J 

(14)

Its augmented Lagrangian form is,

\[ \mathcal{L}(E, F, J, L, P) = \alpha \sum_{i,j} \phi(E_{i,j}) + \beta \| F \|_{2,1} + \gamma \| J \|_W S_p \\
= \frac{1}{2} \| B - F^T X \|_F^2 + \lambda \left( \| L - J \| + \frac{1}{\lambda} \| Y \|_F^2 \right) \\
+ \| X - PF^T XL - E + \frac{Y_2}{\lambda} \|_F^2 
\]

s.t. \ P^T P = I 

(15)

where \( \lambda \) denotes the penalty parameter.

1) UPDATE \ E

It can be optimized by fixing \( (F, J, L, P) \) as,

\[ E^* = \arg \min_E \frac{1}{2} \| X - PF^T XL - E + \frac{Y_2}{\lambda} \|_F^2 \\
+ \frac{\alpha}{\lambda} \sum_{i,j} \phi(E_{i,j}) 
\]

(16)

However, the correntropy measurement term \( \phi(E_{i,j}) \) in (16) is difficult to optimize directly. According to [48], the HQ technique can be applied to optimize \( E \). So, we introduce an adjuvant \( S \) to get the following formula:

\[ \phi(E_{i,j}) = \min_{S_{i,j} \in \mathbb{R}} \frac{1}{2} \| S_{i,j} E_{i,j} \|^2 + \psi(S_{i,j}) 
\]

(17)

where \( \psi(\cdot) \) denotes the dual function of \( \phi(\cdot) \).

Substituting (17) into (16) and transforming the optimization problem as,

\[ \min_{S} \frac{1}{2} \left\| X - PF^T XL - E + \frac{Y_2}{\lambda} \right\|_F^2 \\
+ \frac{\alpha}{\lambda} \| S \otimes E \|_F^2 + \sum_{i,j} \psi(S_{i,j}) 
\]

(18)

where \( \otimes \) denotes the dot product.

This problem can be obtained by the optimal solution by alternate updating \( E \) and \( S \). In detail, when \( E_{i,j} \) is fixed, we can obtain,

\[ S_{i,j}^* = \exp(-\frac{E_{i,j}^2}{2\sigma^2}) 
\]

(19)

where \( \sigma^2 = \frac{1}{2m \times n} \| X - PF^T XL \|_F^2 \). Then, fixing \( S_{i,j} \), \( E_{i,j} \) can be optimized by,

\[ E^* = \arg \min_E \frac{1}{2} \| X - PF^T XL - E + \frac{Y_2}{\lambda} \|_F^2 \\
+ \frac{\alpha}{\lambda} \| S \otimes E \|_F^2 \\
= \left( X - PF^T XL + \frac{Y_2}{\lambda} \right) \cdot \left( \frac{\alpha}{\lambda} S^* + 1 \right) 
\]

(20)

where \( \cdot / \) denotes the division of element-by-element.

2) UPDATE \ F

It can be optimized by fixing \( (E, J, L, P) \) as,

\[ \mathcal{L}(F) = \beta \| F \|_{2,1} + \frac{1}{2} \| B - F^T X \|_F^2 \\
+ \frac{\lambda}{2} \| X - PF^T XL - E + \frac{Y_2}{\lambda} \|_F^2 
\]

(21)

By setting \( \frac{\partial \mathcal{L}(F)}{\partial F} = 0 \) to get the analytical solution of (21):

\[ F^* = (\beta A + XX^T + \lambda DD^T)^{-1}(XB^T + \lambda DC^T P) \]

(22)

where \( D = XL, C = X - E + \frac{Y_2}{\lambda} \). \( A \in \mathbb{R}^{m \times m} \) is a diagonal matrix and meets the following conditions:

\[ A_{ij} = \begin{cases} 
\frac{1}{2} \| F_{i..} \|_2^2, & i = j \text{ and } \| F_{i..} \|_2 \neq 0, \\
0, & i \neq j \text{ or } \| F_{i..} \|_2 = 0.
\end{cases} \]

where \( F_{i..} \) is the \( i \)-th row of \( F \).

3) UPDATE \ J

It can be optimized by fixing \( (E, F, L, P) \) as,

\[ J^* = \arg \min_J \psi(J)^p + \frac{\lambda}{2} \| J - Q \|_F^2 
\]

(23)

where \( Q = L + \frac{Y_2}{\lambda} \).

This is a typical WSNM problem. In light of [41], the general solution of (23) can be calculated by \textbf{Lemma 1}. The detailed solution process is given in [41].

\textbf{Lemma 1}: Let the SVD of \( Q \in \mathbb{R}^{m \times n} \) be \( Q = U \Theta V^T \) with \( \Theta = (\theta_1, \cdots, \theta_k), \) where \( k = \min(m, n), \) and \( \theta_1 \geq \cdots \geq \theta_m. \) If \( 0 \leq w_1 \leq \cdots \leq w_k, \) then the optimal solution of (23) will be \( J = U \Delta V^T \) with \( \Delta = \text{diag}(\delta_1, \cdots, \delta_k), \) where \( \delta_i \) can be solved by:

\[ \min_{\delta_1, \cdots, \delta_k} \sum_{i=1}^{k} \left( \frac{\lambda}{2} (\delta_i - \theta_i)^2 + \gamma w_i \delta_i^p \right), \ i = 1, \cdots, k \]

(24)

\[ s.t. \ \delta_i \geq 0, \ and \ \delta_i \geq \delta_j, \ for \ i \leq j. \]

where \( w_i = 10^{1/p} \sqrt{m \times n} / (\delta_i + 10^{-16}) \).

4) UPDATE \ L

It can be optimized by fixing \( (E, F, J, P) \) as,

\[ L^* = \arg \min_L \mathcal{L}(L) \]

\[ = \arg \min_L \frac{\lambda}{2} \left( \| L - G_1 \|_F^2 + \| PF^T XL - G_2 \|_F^2 \right) \]

(25)

where \( G_1 = J - \frac{Y_2}{\lambda} \) and \( G_2 = X - E + \frac{Y_2}{\lambda} \).

By setting \( \frac{\partial \mathcal{L}(L)}{\partial L} = 0 \), we have:

\[ (X^T FF^T X + I) L - (X^T FP^T G_2 + G_1) = 0 \]

(26)

Obviously,

\[ L^* = (X^T FF^T X + I)^{-1} (X^T FP^T G_2 + G_1) \]

(27)
5) UPDATE $P$

It can be optimized by fixing $(E,F,J,L)$ as,

$$P^* = \arg \min_{P:F^{T}F \leq I} \frac{\lambda}{2} ||X - PF^{T}XL - E + \frac{Y_2}{\lambda}||_F^2$$

$$= \arg \min_{P:F^{T}F \leq I} \frac{\lambda}{2} ||G_3^{T} - L^{T}X^{T}FP^{T}||_F^2$$

(28)

where $G_3 = X - E + \frac{Y_2}{\lambda}$. Obviously, Eq.(28) is an orthogonal procrustes problem (OPP) [49]. So, we can get the solution directly:

$$P^* = UV^T$$

(29)

where $U$ and $V$ can be solved by $UAV^T = G_3L^TX^TF$.

6) UPDATE $Y_1$ AND $Y_2$

By fixing other parameters, they can be updated by,

$$Y_1 := Y_1 + \lambda(L - J)$$

$$Y_2 := Y_2 + \lambda(X - PF^{T}XL - E)$$

(30)

To clearly explain the optimization process, we assume that $(E^t,F^t,J^t,L^t,P^t)$ and $(E^{t+1},F^{t+1},J^{t+1},L^{t+1},P^{t+1})$ represent the results of the $t$-th iteration and $(t+1)$-th iteration, respectively. Algorithm 1 detailed reveals main steps of AOLRE.

Algorithm 1 AOLRE

Require: Data matrix $X$, balance parameters $\alpha, \beta, \gamma$.

1: Initialize: $[E^0, F^0, J^0, L^0, P^0, Y_1^0, Y_2^0] = 0$; Set $t = 0$, $\lambda = 0.1$, $\lambda_m = 10^5$, $\maxIter = 50$, $\eta = 2.1$, $\epsilon_1 = 10^{-5}$, $\epsilon_2 = 10^{-6}$.

2: while not converged and $t < \maxIter$ do

3: Given $F^t, J^t, L^t, P^t$, update $E^{t+1}$ and $S$ cyclically according to (19) and (20), respectively, until: $\|X - PF^{T}XL - E\|_F/\|X\|_F < \epsilon_1$;

4: Given $E^{t+1}, J^t, L^t, P^t$, update $F^{t+1}$ according to (22);

5: Given $E^{t+1}, F^{t+1}, J^t, P^t$, update $J^{t+1}$ according to (24);

6: Given $E^{t+1}, F^{t+1}, J^{t+1}, P^t$, update $L^{t+1}$ according to (27);

7: Given $E^{t+1}, F^{t+1}, J^{t+1}, L^{t+1}$, update $P^{t+1}$ according to (29);

8: Update $Y$ according to (30);

9: Update the penalty variable $\lambda = \min(\eta \lambda, \lambda_m)$;

10: Update $t = t + 1$;

end while

Ensure: $E, F, L$ and $P$.

C. COMPLEXITY ANALYSIS

It can be known by analyzing Algorithm 1, the computational complexities of optimizing $E, F, J, L$ and $P$ are $O(n^3), O(m^3), O(n^3), O(n^3)$ and $O(d^3)$, respectively. Optimizing $Y_1$ and $Y_2$ requires $O(2n)$. Hence, the computational complexity of AOLRE is approximately $O(n^3 + m^3 + n^3 + n^3 + d^3 + 2n)$ for each iteration. Actually, $d$ is small and $m \ll n$, so the total computational complexity is approximately $O(n^3)$, where $t$ denotes the times of iteration.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. SETTINGS

In the following, extensive experimental verification of the proposed AOLRE method is performed. We assess the capability of AOLRE from multiple aspects and compare it with several advanced methods, including SRC [20], LRC [21], LatLRR [31], CRC [32], LRRC [33], LRE [35] and SALPL [34].

In our experiments, five general datasets are used to evaluate the performance of AOLRE, which are taken from different application scenarios, such as face images (i.e., Extended Yale B [50], AR [51], ORL [52]), digital images (i.e., USPS [53]), and object images (i.e., COIL-20 [54]). In particular, the Extended Yale B (YaleB) dataset is used for three types of performance verification, namely different illumination, random pixel corruption and block occlusion corruption. The detailed information of these datasets used in this work is described in Tab. 2. Similar to the strategy used by several comparison algorithms, all samples are separated two parts, which are used for feature extraction and feature extraction capability evaluation, respectively.

| DID | Dataset | Samples | Pixels | Classes | Image types |
|-----|---------|---------|--------|---------|-------------|
| 1   | YaleB   | 2414    | 32 x 32| 38      | Face        |
| 2   | AR      | 2600    | 60 x 44| 10      | Face        |
| 3   | ORL     | 408     | 32 x 32| 40      | Face        |
| 4   | USPS    | 1000    | 16 x 16| 10      | Digit       |
| 5   | COIL-20 | 1440    | 32 x 32| 20      | Object      |

The parameters in our model are set as $\alpha, \gamma = \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, $\beta = \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$ and $p = \{1.0, 0.8, 0.5, 0.1, 0.05\}$. The parameters used in the comparison algorithm come from its original article. In particular, in order to increase comparability, the unsupervised LRE is extended to the supervised scenario by adding a classification loss item $\|B - F^TX\|_F$. For this reason, the dimension of $F$ needs to be reduced to the same number of training classes, that is, $c$. In addition, to ensure the reliability of the results, each experiment is repeated 20 times, and the final results are listed in the form of average and standard deviation.

B. FACE RECOGNITION USING THE YaleB DATASET

This dataset collects 2414 samples of 38 categories with various illuminations, and each category contains approximately 64 photos. As the light intensity and angle change, all images are damaged to varying degrees. In this work, all samples are resized to $32 \times 32$ pixels.
1) THE RECOGNITION PERFORMANCE UNDER DIFFERENT ILLUMINATIONS
As shown in Fig. 2 (a), most images are damaged by shadows or noise, which makes the FE task very challenging. Here, the training set is composed of Tr = \{10, 15, 20, 25\} random samples of each category (“Tr” represents the number of training samples), and all the other samples constitute a test set. The test results of all methods are revealed in Fig.3. Obviously, AOLRE gets the best performance under different number of training samples. Taking the Avg value as an example, the recognition rate of AOLRE is 2.23%, 3.78%, 6.70%, 4.71%, 5.02%, 3.35% and 2.22% higher than that of SRC, LRC, LatLRR, CRC, LRRC, LRE and SALPL, respectively.

2) THE RECOGNITION PERFORMANCE WITH BLOCK OCCLUSION
In this sub-experiment, for training set, we set Tr = 20, and all the other samples constitute test set. In particular, 5 of the 20 images have randomly added occlusion block corruption. The position of these occlusion blocks in each sample is random, and the ratio is from 5% to 30%. Some examples are revealed in Fig. 2 (b), and the test results are revealed in Tab. 3. Taking the Avg value as an example, the recognition rate of AOLRE is 7.78%, 7.97%, 14.24%, 8.22%, 10.97%, 3.38% and 3.63% higher than that of SRC, LRC, LatLRR, CRC, LRRC, LRE and SALPL, respectively.

3) THE RECOGNITION PERFORMANCE WITH PIXEL CORRUPTION
Similar to the previous sub-experiment, Tr = 20 for training set, and all the other samples constitute test set. In particular, the pixels in 5 out of 20 pictures are damaged. These pixels are randomly selected and replaced with uniformly distributed values (the values is [0, 255]), and the ratio is 10% to 40%. Some examples are revealed in Fig. 2 (c), and the recognition accuracies are revealed in Tab. 4. Obviously, AOLRE has achieved the highest recognition rate, and it is 4.87% higher than the second best method SALPL.

C. FACE RECOGNITION USING THE AR DATASET
The images in this dataset come from 70 male subjects and 56 female subjects, a total of more than 4000 facial images. The 26 samples of each category are gathered during two sessions, 13 samples are gathered each session. Of the 13 images, 6 were covered by sunglasses/scarves and 7 were uncovered. Fig. 4 shows some examples of the AR dataset. In this work, we use a total of 2600 image samples of 50 female and 50 male subjects to form a sub-dataset, and all samples are resized to 60 × 44 pixels. The 1200 samples in the test set come from 6 samples randomly selected from each person in each session (including 1 sunglasses, 1 scarf, and 4 neutrals), and the rest are used for testing. In Fig. 5, the test result of AOLRE is 4.01% higher than the next best model SALPL.

D. FACE RECOGNITION USING THE ORL DATASET
This dataset consists of 40 individual image samples, each of which provides 10 different images. When these pictures are collected, the angle and expression of the face are changed. In addition, the facial details of these subjects are also different. For example, some people wear glasses and some don’t. Some examples are revealed in Fig.6. Similar to the YaleB dataset, the pixels of all samples are reduced to 32 × 32. the training set is composed of Tr = 5 random samples of each category, and all the other samples constitute a test set. The recognition accuracies of all algorithms are shown in Fig.7. As we can see, the recognition accuracy of AOLRE is the best among all methods, 2.01% higher than the next best model SALPL.

E. DIGIT RECOGNITION USING THE USPS DATASET
Next, we evaluate the recognition level of model AOLRE by using the USPS [53] handwritten digit database for digit recognition. This dataset is composed of handwritten images of 10 numbers from ‘0’ to ‘9’, and each subject has 1100 samples. Some examples are revealed in Fig.8. Here, we select 1000 samples from the original dataset to form a sub-dataset, with 100 pictures per category. For this sub-data set, there are 40 random samples of each category for training, and others for testing. As we can see from Fig.9, the recognition accuracy of AOLRE is the best among all methods, 2.54% higher than the second best method SALPL.

F. OBJECT RECOGNITION USING THE COIL-20 DATASET
This dataset is composed of 20 different categories, and each category has 72 black background images taken from different views. Here, there are 10 random samples of each category for training, and others for testing. Images of 10 categories in this dataset are displayed in Fig.10. The recognition accuracies are revealed in Fig.11. The analysis reveals that
G. EXPERIMENTAL ANALYSIS

Based on the results of extensive experiments on five datasets, we can summarize the following points:

1. Compared with other advanced representation based feature learning methods, AOLRE achieves better results in face recognition, object recognition and so on. There are the

AOLRE obtains the best result in all methods, which proves that model AOLRE can effectively extract the features of object images.
following reasons: (1) In the process of dimensionality reduction, \( \|F\|_{2,1} \) plays a key role in the retention of key features, and can effectively remove redundant information; (2) The introduction of data reconstruction enables AOLRE to retain the main energy of the original data. (3) Classification loss term \( \|B - F^TX\|^2_F \) in AOLRE can clearly guide the FE and make the learned features more distinctive. These characteristics enable AOLRE to effectively complete the classification task. The specific experimental results are revealed in Fig.3, Fig.5, Fig.7, Fig.9 and Fig.11.

2. As the robustness verification experiments, it can be clearly seen from Tab.3 and Tab.4 that the recognition accuracies of LRE, SALPL, AOLRE are evidently exceed other comparison models, and the performance of AOLRE is the best. The main reasons are as follows: (1) LRR can mine the global information of data, which is very helpful to clear away the corrupt part. The introduction of \( \|L\|_{p,w}^{Sp} \) enables AOLRE get better low-rank performance. (2) The correntropy is applied in AOLRE, which is better for handling non-Gaussian noise. (3) As an extended version of LRE, AOLRE seeks important effective features by learning a robust embedding subspace. All in all, the above three aspects respectively focus on damaged blocks, outliers samples, and damaged characteristics. The integration of these advantages significantly improves the robustness of AOLRE. In the face of actual data with complex noise, our model can also be competent for FE tasks.

**H. PARAMETERS SENSITIVITY AND CONVERGENCE OF AOLRE**

Next, we analyze the parameter sensitivity and convergence of AOLRE on the YaleB dataset (textTr = 10). In our model, there are three balance parameters (\( \alpha, \beta \) and \( \gamma \)) and one
low-rank parameter $p$. The parameter settings are as described in section IV-A. For the balance parameters, we fix $\alpha$ first, and find a better combination by varying $\beta$ and $\gamma$ synchronously. Then replace $\alpha$ and repeat the above process until the optimal solution is obtained. Fig.12 and Fig.13 show the recognition result of AOLRE on the YaleB dataset when varying the parameters. Obviously, the value of the parameters has different influence on the recognition performance, and overall, they are relatively stable.

Furthermore, the convergence of AOLRE is verified on the YaleB dataset ($textTr = 10$). In Fig. 14, we plot the curves of primal residual (computed as $\max(\|L - J\|_F, \|X - PF^T XL - E\|_2^2)$) and the objective function value of Eq. (13). Obviously, our model converges in 20 iterations. Similar outcomes can also be obtained on other datasets. Due to space constraints, we don’t show experimental results on other datasets.

FIGURE 12. Recognition accuracies (%) of AOLRE versus the balance parameters on the YaleB dataset.

FIGURE 13. Recognition accuracies (%) of AOLRE versus $p$ on the YaleB dataset.

FIGURE 14. Convergence curves for AOLRE on the YaleB dataset ($textTr = 10$).

I. COMPUTATIONAL PERFORMANCE

In addition, the computation time is also a key criterion for evaluating algorithms, so we test all algorithms on the YaleB dataset ($textTr = 10$), and the results are shown in Fig. 15. Obviously, compared with LatLRR, LRRC, and SALPL, AOLRE has a lower computational cost. For example, SRC, LRC and CRC, although their computational cost is lower than AOLRE, their recognition rate is low, especially for noisy data, the recognition accuracy of AOLRE is significantly higher than the other three algorithms.

V. CONCLUSION

In this work, a novel robust FE model (AOLRE) is designed, which can be thought of as an evolutionary version of LRE. Especially, this method uses an orthogonal matrix to retain the primary components of the samples. In addition, the introduction of $\|L\|_{w, Sp}^p$ enables AOLRE get better low rank performance. Meanwhile, the correntropy is applied in AOLRE, which successfully improves the model’s resistance to complex noise, thereby enhance the robustness of the algorithm. Finally, the introduction of the classification loss item allows our model to effectively fit the supervised scene. Adequate experimental results certify that AOLRE is an advanced...
feature extraction method and can be competent for feature extraction of data with complex noise.

Certainly, there are still rooms for further improvement in AORLE: 1) The solution of the model (especially the relationship between the balance parameters) needs to be further improved and optimized to be suitable for processing large-scale data; 2) It is worth trying to integrate the “kernel trick” into the model to deal with the nonlinearity of data; 3) Similar to [55], it may be an interesting attempt to introduce structured auto-encoder into the model.

REFERENCES

[1] B. Alwaely and C. Abhayaratne, “Graph spectral domain feature learning with application to in-air hand-drawn number and shape recognition,” IEEE Access, vol. 7, pp. 13135–13148, 2019.

[2] L. Gao, X. Zhang, J. Gao, and S. You, “Fusion image based radar signal feature extraction and modulation recognition,” IEEE Access, vol. 7, pp. 159661–159673, 2019.

[3] S. H. Abdullhusain, A. R. Ramli, B. M. Mahmood, M. I. Saripan, S. A. R. Al-Haddad, and W. A. Jassim, “A new hybrid form of Krawtchouk and Tchebycheff polynomials: Design and application,” J. Math. Imag. Vis., vol. 40, no. 4, pp. 550–570, May 2010.

[4] A. M. Abdul-Hadi, S. H. Abdullhusain, and B. M. Mahmood, “On the computational aspects of Charlier polynomials,” Cogent Eng., vol. 7, no. 1, Jan. 2020, Art. no. 1763553.

[5] B. M. Mahmood, A. M. Abdul-Hadi, S. H. Abdullhusain, and A. Hussien, “On the computational aspects of Krawtchouk polynomials for high orders,” J. Imag., vol. 6, no. 8, p. 81, Aug. 2020.

[6] M. Turk and A. Pentland, “Eigenfaces for recognition,” J. Cognit. Neurosci., vol. 3, no. 1, pp. 71–86, 1991.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, Jul. 1997.

[8] F. Song, D. Zhang, D. Mei, and Z. Guo, “A multiple maximum scatter difference discriminant criterion for facial feature extraction,” IEEE Trans. Syst. Man, Cybern. B, Cybern., vol. 37, no. 6, pp. 1599–1606, Dec. 2007.

[9] W. Zuo, D. Zhang, J. Yang, and K. Wang, “BDPCA plus LDA: A novel fast feature extraction technique for face recognition,” IEEE Trans. Syst. Man, Cybern. B, Cybern., vol. 36, no. 4, pp. 946–953, Aug. 2006.

[10] C. Fraley and A. E. Raftery, “Model-based clustering, discriminant analysis, and density estimation,” J. Amer. Stat. Assoc., vol. 97, no. 458, pp. 611–631, Jun. 2002.

[11] Q. Liu, H. Lu, and S. Ma, “Improving kernel Fisher discriminant analysis and density estimation,” J. Man, Cybern. B, Cybern., vol. 36, no. 4, pp. 946–953, Aug. 2006.

[12] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, “A two-phase test sample sparse representation method for use with face recognition,” IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 9, pp. 1255–1262, Sep. 2011.

[13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, “Robust recovery of subspace structures by low-rank representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 171–184, Jan. 2013.

[14] S. Li and Y. Fu, “Learning balanced and unbalanced graphs via low-rank coding,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 5, pp. 1274–1287, May 2015.

[15] A. Humeau-Heurtier, “Texture feature extraction methods: A survey,” IEEE Access, vol. 7, pp. 8975–9000, 2019.

[16] K. Tang, R. Liu, Z. Su, and J. Zhang, “Structure-constrained low-rank representation,” IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 12, pp. 2167–2179, Dec. 2014.

[17] L. Zhuang, S. Gao, J. Tang, J. Wang, Z. Lin, Y. Ma, and N. Yu, “Constructing a nonnegative low-rank and sparse graph with data-adaptive features,” IEEE Trans. Image Process., vol. 24, no. 11, pp. 3717–3728, Nov. 2015.

[18] J. Li, Y. Kong, H. Zhao, J. Yang, and Y. Fu, “Learning fast low-rank projection for image classification,” IEEE Trans. Image Process., vol. 25, no. 9, pp. 4803–4814, Oct. 2016.

[19] H. Zou, T. Hastie, and R. Tibshirani, “Sparse principal component analysis,” J. Comput. Graph. Statist., vol. 15, no. 2, pp. 265–286, Jun. 2006.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” J. ACM, vol. 58, no. 3, p. 11, 2011.

[21] G. Liu and S. Yan, “Latent low-rank representation for subspace segmentation and feature extraction,” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 1615–1622.

[22] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 471–478.

[23] Y. Zhang, Z. Jiang, and L. S. Davis, “Learning structured low-rank representations for image classification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 676–683.

[24] X. Fang, N. Han, J. Wu, Y. Xu, J. Yang, W. K. Wong, and X. Li, “Approximate low-rank projection learning for feature extraction,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 11, pp. 5228–5241, Nov. 2018.

[25] W. K. Wong, Z. Lai, J. Wu, X. Fang, and Y. Lu, “Low-rank embedding for robust image feature extraction,” IEEE Trans. Image Process., vol. 26, no. 6, pp. 2905–2917, Jun. 2017.

[26] X. Peng, C. Lu, Z. Su, and J. Zhang, “Connections between nuclear-norm and frobenius-norm-based representations,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 1, pp. 218–224, Jan. 2018.

[27] L. Liu, W. Huang, and D.-R. Chen, “Exact minimum rank approximation via schatten p-norm minimization,” J. Comput. Appl. Math., vol. 267, pp. 218–227, Sep. 2014.

[28] X. Zhang, H. Sun, Z. Liu, Z. Ren, Q. Cui, and Y. Li, “Robust low-rank kernel multi-view subspace clustering based on the schatten p-norm and corentropy,” Inf. Sci., vol. 477, pp. 430–447, Mar. 2019.

[29] M. Brbić and I. Kopriva, “ε₀-motivated low-rank sparse subspace clustering,” IEEE Trans. Cybern., vol. 50, no. 4, pp. 1711–1725, Apr. 2020.

[30] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2862–2869.

[31] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, “Weighted schatten p-norm minimization for image denoising and background subtraction,” IEEE Trans. Image Process., vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[32] X. Xue, X. Zhang, X. Feng, H. Sun, W. Chen, and Z. Liu, “Robust subspace clustering based on non-convex low-rank approximation and adaptive kernel,” Inf. Sci., vol. 513, pp. 190–205, Mar. 2020.

[33] S. Boyd, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

[34] W. Liu, P. P. Pokharel, and J. C. Principe, “Corentropy: Properties and applications in non-Gaussian signal processing,” IEEE Trans. Signal Process., vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[35] G. Xia, H. Sun, L. Feng, G. Zhang, and Y. Liu, “Human motion segmentation via robust kernel sparse subspace clustering,” IEEE Trans. Image Process., vol. 27, no. 1, pp. 135–150, Jan. 2018.
[46] X. Peng, Z. Yu, Z. Yi, and H. Tang, “Constructing the L2-graph for robust subspace learning and subspace clustering,” IEEE Trans. Cybern., vol. 47, no. 4, pp. 1053–1066, Apr. 2017.

[47] M. Hong, Z.-Q. Luo, and M. Razaviyayn, “Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems,” SIAM J. Optim., vol. 26, no. 1, pp. 337–364, 2016.

[48] M. Nikolova and M. K. Ng, “Analysis of half-quadratic minimization methods for signal and image recovery,” SIAM J. Sci. Comput., vol. 27, no. 3, pp. 937–966, 2005.

[49] H. Park, “A parallel algorithm for the unbalanced orthogonal procrustes problem,” Parallel Comput., vol. 17, no. 8, pp. 913–923, 1991.

[50] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643–660, Jun. 2001.

[51] A. Martinez and R. Benavente, “The AR face database,” CVC, Tech. Rep. 24, 1998, [Online]. Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[52] F. S. Samaria and A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in Proc. IEEE Workshop Appl. Comput. Vis., Dec. 1994, pp. 138–142.

[53] J. J. Hull, “A database for handwritten text recognition research,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 5, pp. 550–554, May 1994.

[54] S. A. Nene et al., “Columbia object image library (COIL-20),” Dept. Comp. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. 1996.

[55] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, “Structured AutoEncoders for subspace clustering,” IEEE Trans. Image Process., vol. 27, no. 10, pp. 5076–5086, Oct. 2018.

**CONG FU** received the B.S. degree in electronic information engineering from the Southwest University of Science and Technology, Mianyang, China, in 2003, and the M.S. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, China, in 2010. He is currently pursuing the Ph.D. degree in radio physics with the China Academy of Engineering Physics. His research interests include machine learning, deep learning, signal processing, and pattern recognition.

**ZHIGUI LIU** received Ph.D. degree in traffic information engineering and control from Southwest Jiaotong University, Chengdu, China, in 2006. He is currently a Professor with the Department of Information Engineering, Southwest University of Science and Technology. His current interests include machine vision and control theory.

**LI LI** was born in 1981. He received the Ph.D. degree in mechanical and electronic engineering from Sichuan University (SCU), China, in 2013, respectively. He is currently an Assistant Professor with the School of Computer Science and Technology, Southwest University of Science and Technology. He is also leading a research group in doing some edge computing and CPS designing. His research interests include edge computing, CNN, droid navigation, and machine vision.