IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00,0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2021.Doi Number

JCS: an Explainable Surface Defects Detection Method for Steel Sheet by Joint Classification and Segmentation

SHIYANG ZHOU[1, 2, 3], HUAIGUANG LIU[1, 3], KETAO CUI[1, 3], and ZHIQIANG HAO[1, 3, *]

[1]Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
[2]Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, Wuhan 430074, China
[3]Precision Manufacturing Institute, Wuhan University of Science and Technology, Wuhan 430081, China

Corresponding author: Zhiqiang Hao (e-mail: haozhiqiangwust@163.com).

This research was funded by National Natural Science Foundation of China, under grant number 51805386, open fund of Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, under grant number IPIC2019-03, open fund of Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering at Wuhan University of Science and Technology, under grant number 2018A04.

ABSTRACT For surface defect images that captured from a practical steel production line, different shape, size, location and texture of defect object may cause inter-class similarity and intra-class difference of defect images. Despite attractive results have been achieved in some surface methods for defect classification and segmentation, it is still far from meeting the needs of real-world applications due to lack of adaptiveness of these methods. Considering the surface defect image can be decomposed into defect foreground image and defect-free background image, the paper develops a novel joint classification and segmentation (JCS) approach to perform surface defects detection for steel sheet. It comprises of the classification method based on a class-specific and shared discriminative dictionary learning (CASDDL) and the segmentation method based on a double low-rank based matrix decomposition (DLMD), respectively. For the proposed CASDDL method, we learn a shared sub-dictionary as well as several class-specific sub-dictionaries to explicitly capture common information shared by all classes and class-specific information belonging to corresponding class. We adopt a mutual incoherence constrain for each sub-dictionary, a Fisher-like discriminative criterion and low-rank constrain on coding vector to improve the discriminative ability of learned dictionary. For the proposed DLMD method, we formulate the segmentation task as a double low-rank based matrix factorization problem, and the Laplacian and sparse regularization terms are introduced into the matrix decomposition framework. Experimental results demonstrate that our proposed JCS method achieve a comparable or better performance than the state-of-the-art methods in classifying and segmenting surface defects of steel sheet.

INDEX TERMS Joint classification and segmentation for image; class-specific and shared dictionary learning; double low-rank matrix decomposition; surface defects of steel sheet

I. INTRODUCTION

Automated surface defect classification and segmentation based on machine vision are two most essential and related tasks in quality management of industrial products. For the real-time surface defect detection system based on machine vision, the classification task is used to classify normal

images and abnormal images, which is highly beneficial for improving the efficiency and accuracy of defect segmentation, whereas the segmentation task is used to detect the locations and boundaries of defects, which highlighting the critical defect regions for high-level image understanding [1].

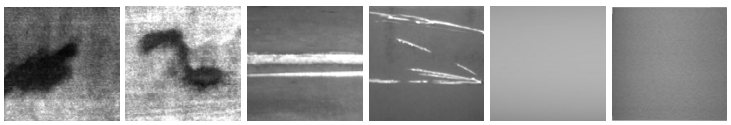


FIGURE 1. Examples of surface images of steel sheet: (a)-(b) Patch; (c)-(d) Scratch; (e)-(f) Non-defective.

As shown in Fig. 1, both classification and segmentation tasks are challenging due to the following reasons: heterogeneous and scattered defect: the number and type of defect are generally unknown in advance, and different surface images often have different imaging qualities, i.e., low contrast between each defect and its surrounding surface tissue results in fuzzy defect boundaries; cluttered and complicated background: non-defective background may also have great differences in different images; different types of defect might be contained in a single defect image, and they often exhibit substantial stochastic variability in terms of shape, size, gray, texture and location; the inter-type surface defects may share visual similarities, and the intra-type defects may have visual differences. In the past two decades, many efforts have been devoted for more efficient and accurate defect classification and segmentation methods [2-3]. These approaches focused on two aspects of feature extraction and classifier design, which are basically customized for a predefined or specific type of defect. Besides, the low computational speed of these methods is a limitation for real-time detection. These factors motivate researchers to develop some new methods for surface defect classification and segmentation.

Most recently, convolutional neural networks (CNN) and generative adversarial networks (GAN)-based deep learning methods have been achieving remarkable performance in image classification and segmentation. Therefore, some studies have attempted to adopt deep learning methods for defect detection [4-5]. As mentioned in [6-7], these deep learning models are complex with many parameters, and training them require a huge number of expert-labelled training samples, complex optimization algorithm, consume a significant amount of computing resources to keep running as its complex network structure, which are the significant challenging problem in industrial environments. Moreover, defective samples are difficult to obtain because of the probability of defect occurrence is very low in industrial manufacturing. In particular, these deep learning models lack of sufficient theoretical support and mostly rely on the human experiences, which limit the practical use.

Lately, dictionary learning has been successfully applied to many machine vision problems [8-9], such as surface defect classification of industrial products [10]. Sparse representation-based classification (SRC) [11] used original training data as a dictionary directly, and Aharon et al. [12] proposed K-SVD method to learn an over-complete dictionary from original training data. Ramirez et al. [13] developed a structured incoherence regularization term for dictionary learning (DLSI) to promote the independence between different sub-dictionaries. Ling et al. [14] developed a class-oriented discriminative dictionary learning (CODDL) method to emphasis class discrimination of dictionary atoms and representation coefficients. Fan et al. [15] exploited discriminative Fisher embedding dictionary transfer learning (DFEDTL) to preserve the interclass differences and intraclass similarities of training samples. As shown in Fig. 1, defect object in the surface image can be regarded as local anomaly against relatively homogeneous background. The background texture is useful for reconstruction rather than discrimination. For the aforementioned dictionary learning methods, most of atoms are used to represent non-defective background, causing only small part of atoms represent class-specific defect. Therefore, the discrimination of class-specific sub-dictionaries between different defect object will diminish, greatly degrading the classification performance. An intuitive way to capture and separate those shared components from training samples. Recent researches have yielded more promising results by using the idea of shared dictionary, which different classes not only have class-particular parts but also share commonality [16-17]. Gao et al. [18] constructed a joint dictionary learning algorithm to learned some category-specific sub-dictionaries and a shared sub-dictionary by imposing cross-incoherence constraint between different sub-dictionaries and self-incoherence constraint in each sub-dictionary. Wang et al. [19] established a category-specific and shared dictionary learning (COPAR) by exploiting the information of particularity and commonality across all classes. Lin et al. [20] constructed a class-shared, class-specific and disturbance dictionary by introducing a robust, discriminative and comprehensive dictionary learning (RDCDL). However, these methods overlook the low-rank ability of sub-dictionaries or coding vector over the shared sub-dictionary. Therefore, Jiang et al. [21], Rong et al. [22], Wen et al. [23] introduced a low-rank constraint on dictionary decomposition. Furthermore, Vu et al. [24] proposed a low-rank constraint on the shared dictionary (LRSDL) to encourage its subspace to be of low-dimensionality and its corresponding representations to be similar. Du et al. [25] presented a low-rank graph

preserving discriminative dictionary learning (LRGPDDL) by introducing a low-rank constraint on each sub-dictionary. Chen et al. [26] introduced an adaptive dictionary learning strategy combined with an adaptive low-rank representation (ALRR) method for classification. These methods show that incorporating low-rank regularization term into dictionary learning framework can enhance robustness of the learned dictionary and achieved impressive classification results.

Inspired by the idea of shared sub-dictionary and low-rank constrain, we develop a class-specific and shared discriminative dictionary learning (CASDDL) model for surface defect classification of steel sheet. Based on different classes of defect image share similar background, CASDDL-based classification method constructs c class-specific sub-dictionaries associated with corresponding classes and one shared sub-dictionary for all the classes, respectively. With these sub-dictionaries, exclusive features and shared features of surface defect image can be explicitly separated. CASDDL specially introduces incoherence promoting constraints on all the sub-dictionaries and low-rank constraints on coding vector over shared sub-dictionary, to make the learned dictionary more compact, discriminative and robust. Also, a Fisher-like regularization term on coding vectors over class-specific sub-dictionaries ensures more coherence for within-class coding vectors and more disparity for between-class coding vectors.

When the surface image is classified as the defect image, the defect object in defect image should be located and segmented. Some studies based on robust principal component analysis (RPCA) [27] have shown that matrix decomposition techniques are excellent unsupervised method for separating and segmenting the region of interest (ROI) from the image. RPCA assumes that an image can be represented as a combination of a highly redundant part (i.e., background regions) and a sparse part (i.e., foreground object). Mathematically, the feature matrix of input image can be decomposed into a low-rank matrix corresponding to background and a sparse matrix corresponding to foreground object. Some prior knowledge and regularization are incorporate into original RPCA model, which can improve segmentation results in terms of speed and accuracy [28-29]. Cen et al. [30], Li et al. [31] designed a model of low-rank matrix reconstruction for defect inspection. Yan et al. [32] performed a smooth-sparse decomposition (SSD) with regularized high-dimensional regression to decompose a defect image and separate anomalous regions. Cao et al. [33] presented prior knowledge guided least squares regression (PG-LSR) based on low-rank representation to detect diverse defects. Huang et al. [34] applied a texture prior to construct a novel weighted low-rank reconstruction (W-LRR), which is only suitable for the defect images with regular or near-regular texture. Wang et al. [35] studied the entity sparsity pursuit (ESP) to identify surface defects. These methods don't consider the low-rank characteristic for the defect foreground and defect-free background simultaneously, and ignore the spatial and pattern relations of these regions, which may influence the final segmentation performance.

Motivated by the above analysis, a double low-rank decomposition (DLMD) model for surface defect segmentation of steel sheet is exploited in the paper. Based on the unified low-rank assumption to characterize defect foreground and defect-free background, DLMD-based segmentation approach can be divided into two steps: firstly, the defect foreground image and defect-free background image are separated from surface defect image; secondly, the optimization strategy is further applied to improve the accuracy of the defect foreground image, leading to a higher segmentation performance.

To sum up, we propose a joint classification and segmentation (JCS)-based defect detection approach to provide explainable classification and segmentation results for steel sheet. As illustrated in Fig. 2, the proposed JCS approach first identifies the surface defect by a classification branch via CASDDL model. It's then feasible to discover the locations and areas of surface defect by a segmentation branch via DLMD model. With the explainable classification results and corresponding defect segmentation, JCS largely simplifies and accelerates the detection process for quality experts. This paper is an extension of our previous works of [36-37] with significant new proposals and more experiments. Our main contributions are summarized as follows:

- We propose a CASDDL approach to train discriminative dictionary for surface defect classification of steel sheet. It not only encourages intra-class samples to deliver the similar feature representation, but also minimizes the inter-class samples correlations.
- We develop a DLMD approach to segment various types of defects from surface defect images of steel sheet. It doesn't need training process by directly decomposing the surface defect image into the defect foreground image and defect-free background image.
- The feasibility and advantages of the proposed JCS method combined CASDDL and DLMD is evaluated by extensive experiments and comparisons with the other state-of-the-art methods, which show that it clearly improves both subjective and objective quality of surface defect detection for steel sheet.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce some related works about surface defects classification and segmentation, dictionary learning, and RPCA, respectively. Section 3 presents our proposed JCS detection approach, including CASDDL-based defect classification model, and DLMD-based defect segmentation model. In Section 4, we validate proposed JCS approach in extensive experiments and compare it with the other state-of-the-art methods. Some conclusions and future works are finally provided in Section 5.

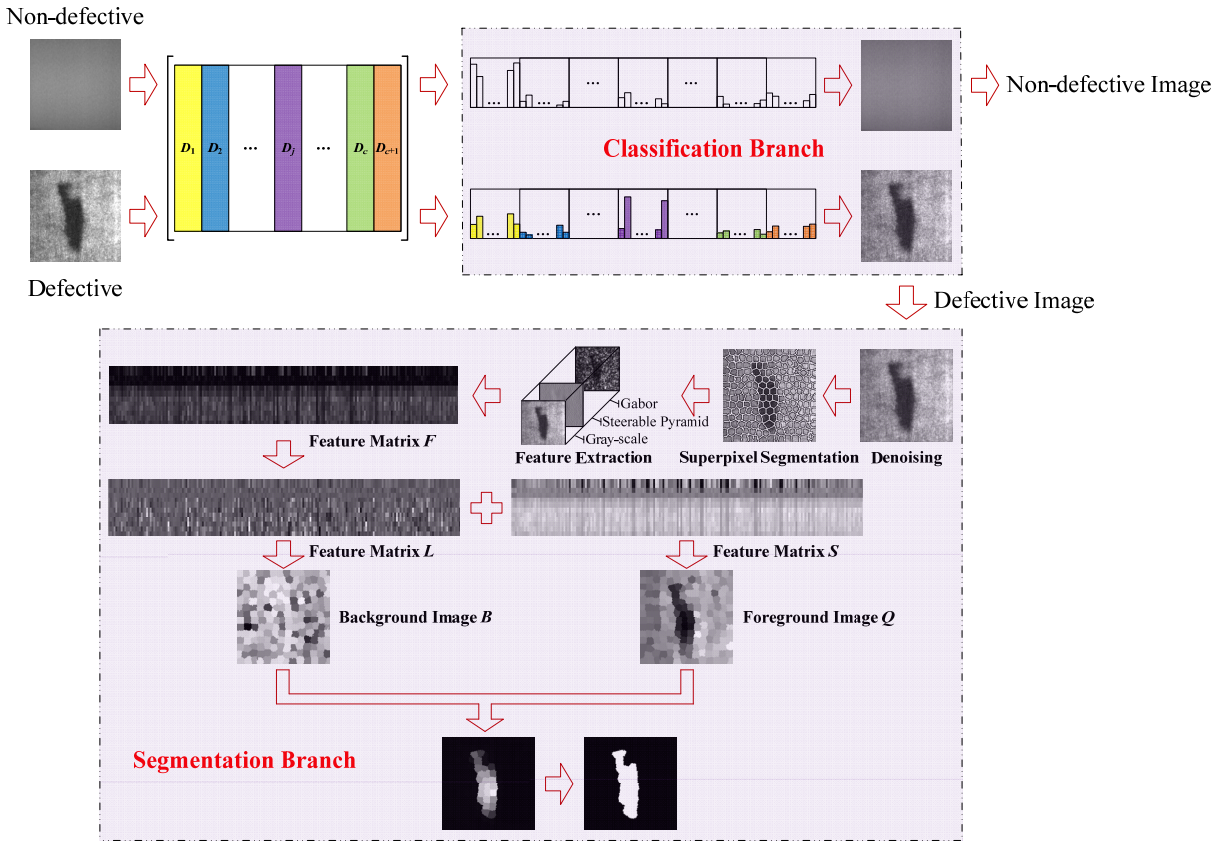


FIGURE 2. Diagram of the proposed JCS approach for surface defect detection of steel sheet.

II. RELATED WORK

A. SURFACE DEFECT CLASSIFICATION AND SEGMENTATION

For classifying surface defects, different customized feature extraction methods for a variety of problems have been developed. The representative feature extraction methods mainly include grayscale, shape, texture, morphological operator, Fourier, Gabor and wavelet transform. Then, these features are combined with powerful classifiers, such as artificial neural networks, support vector machines. Borwankar et al. [38] used the discrete wavelet transform and rotated wavelet transform for feature extraction, while a KNN classifier for classification. Luo et al. [39] exploited a generalized completed local binary patterns framework and simple nearest-neighbor classifier for steel surface defect classification. Ashour et al. [40] developed a method combining the use of discrete shearlet transform and gray-level co-occurrence matrix to classify surface defects of hot-rolled steel strips.

Traditional segmentation methods of surface defect can be mainly divided into three categories: statistical-based methods, filter-based methods and model-based methods. For the statistical-based methods, such as statistical moments, mathematical morphology, maximum entropy, are used to evaluate the spatial distribution characteristic of pixel intensities. These methods are sensitive to lighting, noise or outliers. In contrast, the filter-based methods, such as discrete Fourier transform, discrete Gabor transform and discrete wavelet transform, the energies of the filters response are utilized as features to segment the defects. These methods require the periodicity of texture structures, which may not suitable to random texture. Furthermore, it's not suitable for localizing the defect regions in the spatial domain. The model-based methods, such as level set, Markov random field, fractal model, and partial differential equation, construct the specific models with image feature distributions, which have a high computational complexity.

B. DICTIONARY LEARNING

Mathematically, dictionary learning can be formulated as follows:

$$\min_{D,\boldsymbol{x}} \|\boldsymbol{y} - D\boldsymbol{x}\|_2^2 + \lambda\theta(D,\boldsymbol{x}) \tag{1}$$

where, $\|\cdot\|_2$ denotes l_2 norm, $\boldsymbol{y} \in \mathbb{R}^d$ denotes a given d-dimensional feature vector of training sample, $\boldsymbol{x} \in \mathbb{R}^K$ denotes coding vector of $\boldsymbol{y}$ over dictionary $D = (\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_j, \dots, \boldsymbol{d}_K) \in \mathbb{R}^{d\times K}$, $\boldsymbol{d}_j \in \mathbb{R}^d$ denotes the k-th atom of D, $\theta(D,\boldsymbol{x})$ denotes a regularization term to constrain D or $\boldsymbol{x}$, λ is a positive parameter that balances the tradeoff between reconstructive error $\|\boldsymbol{y} - D\boldsymbol{x}\|_2^2$ and $\theta(D,\boldsymbol{x})$.

For the classification task, discriminative dictionary learning has demonstrated that a well-learned dictionary D

will greatly boost classification performance. The discrimination could be developed from the dictionary, coding vectors, or both. Several regularization terms, such as sparsity, low-rank, neighborhood preservation of graph, entropy, incoherence constraint on sub-dictionaries, have been introduced into the learning process to promote the discriminative power of learned dictionary.

Optimizing Eq. (1) can be carried out by an iterative method composing two steps: (a) fixing D to update x; (b) fixing x to update D, which can be solved efficiently by lots of algorithms [41]. According to the learned dictionary D, test sample $\hat{\boldsymbol{y}}$ is classified as class k^* if it satisfies: $k^* = \arg\min_k \|\hat{\boldsymbol{y}} - D l_k(\boldsymbol{x})\|_2^2$, where, $\boldsymbol{x}$ is coding vector, $l_k(\boldsymbol{x})$ denotes a vector only keeping the entries of $\boldsymbol{x}$ associated with the k-th class and changing others into zeros. As a result, $\hat{\boldsymbol{y}}$ is assigned to the class k^* corresponding to the minimum reconstruction error $\|\hat{\boldsymbol{y}} - D l_{k^*}(\boldsymbol{x})\|_2^2$.

C. ROBUST PRINCIPAL COMPONENT ANALYSIS

RPCA shows the low-rank representation has a better performance in discovering global structures of data, which can reveal the relationships of the samples: the within-class affinities are dense while the between-class affinities are all zeros [42]. RPCA can be formulated as follows:

$$\begin{aligned} &\min_{L,\,S}(\mathrm{rank}(L) + \lambda\|S\|_0) \\ &\text{s.t. } F = L + S \end{aligned} \tag{2}$$

where, $F \in \mathbb{R}^{m\times n}$ is the input matrix, $L \in \mathbb{R}^{m\times n}$ and $S \in \mathbb{R}^{m\times n}$ are two decomposed matrices; $\mathrm{rank}(\cdot)$ denotes the rank of matrix; $\|\cdot\|_0$ denotes l_0 norm of matrix, which equals the number of non-zero element of matrix; $\lambda > 0$ is a trades-off parameter between L and S.

As Eq. (2) is NP-hard problem, $\mathrm{rank}(L)$ can be replaced by nuclear norm $\|L\|_*$, and $\|S\|_0$ can be replaced by l_1 norm $\|S\|_1$ or $l_{2,1}$ norm $\|S\|_{2,1}$, where, $\|\cdot\|_*$ equals the sum of singular values of a matrix; $\|\cdot\|_1$ equals the sum of the absolute values of all entries in a matrix, $\|\cdot\|_{2,1}$ equals the sum of l_2 norms of the columns of a matrix, $\|S\|_{2,1} = \sum_{j=1}^{n}\|\boldsymbol{s}_j\|_2$ with $S = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n)$ with $\boldsymbol{s}_j \in \mathbb{R}^m$.

Several optimization algorithms have been proposed to solve RPCA [43], such as alternating direction method of multipliers, inexact augmented Lagrangian multipliers (inexact ALM) method. Supposing that $L \in \mathbb{R}^{m\times n}$ is a matrix with rank r, its singular value decomposition (SVD) operation is denoted as $\mathrm{svd}(L) = U\Sigma V^T$, where, $\Sigma = \mathrm{diag}(\{\sigma_i\}_{1\le i\le r})$ is the diagonal matrix with σ_1, σ_2,…, σ_r on the diagonal and zeros elsewhere, σ_i is the i-th singular value of L, $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{N\times r}$ are left, right singular matrices, respectively. For the traditional soft-thresholding shrinkage operator $\Psi_\lambda\{\Sigma_{ij}\} = \begin{cases}\Sigma_{ij} - \lambda & \Sigma_{ij} > \lambda \\ 0 & \Sigma_{ij} \le \lambda\end{cases}$, where, Σ_{ij} stands for the (i,j)-th element of Σ. Each singular value equally shrinks by subtracting the same constant λ, which means that all singular values have equal contributions. Given the weights vector $\boldsymbol{w} \in \mathbb{R}^r$, the non-uniform singular value thresholding operator can be defined as follows [44]: $\Psi_{\lambda\boldsymbol{w}}\{\Sigma_{ij}\} = \begin{cases}\Sigma_{ij} - \lambda w_i & \Sigma_{ij} > \lambda \\ 0 & \Sigma_{ij} \le \lambda\end{cases}$, where, $w_i = \frac{\sum_{j=1}^{r}\sigma_j}{\sigma_i}$. For the larger singular values which quantify the principal information of image, they should be reduced a little as much as possible, i.e., the larger the singular value is, the more contribution it makes to the major information. Different singular values are treated differently by assigning different weights and can adaptively shrink according to the specific information of image. For the surface defect image, matrix singular values have clear physical meanings, larger singular values corresponding to major projection directions are supposed to be less shrunk to preserve the major components, which can improve the accuracy of low-rank reconstruction and enhance the adaptivity of defect segmentation.

III. OUR SURFACE DEFECT DETECTION APPROACH

Our JCS detection approach consists of an explainable classification branch to identify the defect and a segmentation branch to discover the defect areas. The proposed CASDDL classification model identifies whether the surface image is defect or not, along with convincing visual explanations. To provide complementary pixel-level prediction, the proposed DLMD segmentation model recognizes fine-grained defect areas in the surface defect image. By combining these two models together for better performance, JCS provides informative detection results for surface defect of steel sheet.

A. EXPLAINABLE CLASSIFICATION

The proposed CASDDL-based classification method mainly comprises of two stages, including discriminative dictionary learning, and defect classification.

1) DISCRIMINATIVE DICTIONARY LEARNING

a: FORMULATION OF CASDDL

Supposing $Y = [Y_1, Y_2, \ldots, Y_i, \ldots, Y_c] \in \mathbb{R}^{d\times N}$ denotes whole training samples of c classes, each column denotes one sample, where, $Y_i \in \mathbb{R}^{d\times n_i}$ denotes the i-th class training samples, d is dimension of one sample, n_i denotes number of sample from class i, $\sum_{i=1}^{c} n_i = N$, where, N is

total number of training samples. Let $D = [D_1, D_2, \dots, D_j, \dots, D_c, D_{c+1}] = [D_{class}, D_{c+1}] \in \mathbb{R}^{d\times K}$ denotes learned dictionary of K atoms, $\{D_j\}_{j=1,2,\dots,c} \in \mathbb{R}^{d\times k_j}$ denotes the j-th class-specific sub-dictionary that trained from a corresponding training samples Y_i, $D_{c+1} \in \mathbb{R}^{d\times k_{c+1}}$ denotes a shared sub-dictionary that trained from the whole training samples Y, where, $K = \sum_{j=1}^{c+1} k_j$, k_j denotes number of atoms from the j-th sub-dictionary. Let $X = [X_1, X_2, \dots, X_i, \dots, X_c] \in \mathbb{R}^{d\times N}$ denotes coding matrix of Y over D, where, $X_i \in \mathbb{R}^{K\times n_i}$ denotes coding matrix of Y_i over D. Furthermore, X_i can be written as $X_i = [X_i^1; X_i^2; \dots; X_i^j; \dots; X_i^c; X_i^{c+1}] = [X_i^{class}; X_i^{c+1}]$, where, $X_i^j \in \mathbb{R}^{k_j\times n_i}$ denotes coding matrix of Y_i over sub-dictionary D_j, $X_i^{class} \in \mathbb{R}^{(K-k_{c+1})\times n_i}$ denotes coding matrix of Y_i over all class-specific sub-dictionaries, $X_i^{c+1} \in \mathbb{R}^{k_{c+1}\times n_i}$ denotes coding matrix of Y_i over the shared sub-dictionary D_{c+1}. To enhance the discriminative capability of dictionary, it's ideally desired that for each class, its samples have non-zero coding vectors intensively locating at the corresponding atoms, whereas the coding vectors at other atoms are zero. As shown in Fig. 3, a sample is supposed to be represented only by the corresponding class-specific sub-dictionary, while can't be represented by other class-specific sub-dictionaries at the same time. It can enhance the discriminative capability of learned dictionary by forcing that all other discriminative sub-dictionaries have poor representative capability of non-corresponding samples. Besides, different sub-dictionaries should be low coherence, which can guide the learned dictionary to be discriminative. What's more, in terms of intra-class compactness and inter-class separability, the coding vectors of same samples class should be similar, while the coding vectors of different samples class should be dissimilar. What's more, the coding vectors corresponding to the shared dictionary should be similar, the corresponding coding matrix should be low-rank, which well addresses the redundant information in the shared sub-dictionary and promotes coding vectors compact.

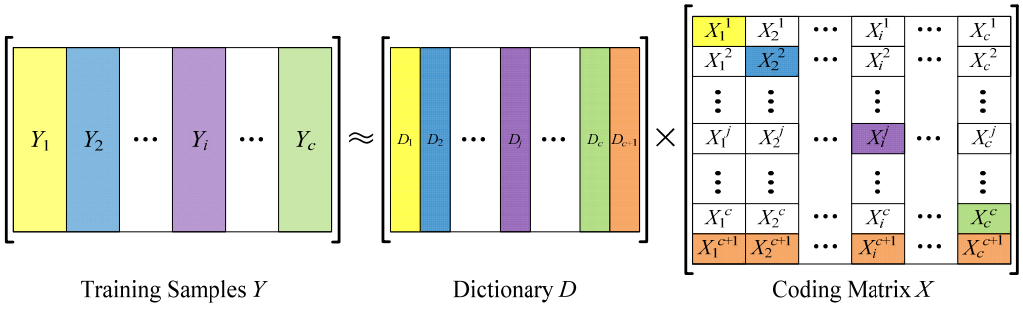


FIGURE 3. **Ideal structure of dictionary *D* and coding matrix *X* in the proposed CASDDL method.**

The block-diagonal constraints increase the discriminative capability of *D* and *X*.

Based on above discussion, the proposed CASDDL can be modelled as the following optimization problem:

$$\min_{D,X} Z_{reconstruction}(Y, D, X) + Z_{incoherence}(D_i, D_j) + Z_{exclusiveness}(X_i^{class}) + Z_{lowrank}(X_i^{c+1}) \quad (3)$$

where, $Z_1 = Z_{reconstruction}(Y, D, X)$ denotes the reconstruction error term; $Z_2 = Z_{incoherence}(D_i, D_j)$ denotes the sub-dictionary incoherence term; $Z_3 = Z_{exclusiveness}(X_i^{class})$ denotes the discriminative promotion term for coding vectors over all the class-specific sub-dictionaries; $Z_4 = Z_{lowrank}(X_i^{c+1})$ denotes the low-rank preserving term for coding vectors over the shared sub-dictionary.

(1) RECONSTRUCTION ERROR TERM Z_1

To learn a representative and discriminative structured dictionary D, each class-specific sub-dictionary should be supposed to well represent samples from the i-th class, but not other classes. The most important property of the shared dictionary is to represent samples from all the classes. According to $Y_i \approx DX_i = D_1X_i^1 + D_2X_i^2 + \dots + D_jX_i^j + \dots + D_cX_i^c + D_{c+1}X_i^{c+1} = D_{class}X_i^{class} + D_{c+1}X_i^{c+1}$, small value of $\|Y_i - D_{class}X_i^{class}\|_F^2$ ensures that the dictionary D can represent Y_i well, where, $\|\cdot\|_F$ denotes Frobenius-norm. Besides, small value of $\|Y_i - D_jX_i^j\|_F^2 = \|Y_i - (D_{class}V_j)(V_j^TX_i^{class})\|_F^2$ $(j = 1, 2, \dots, c)$ makes sure that each class Y_i has a good representation over corresponding class-specific sub-dictionary D_j, where, $V_j \in \mathbb{R}^{K\times k_j}$ is the selection operator that selects the j-th class-specific sub-dictionary D_j from D, each column of V_j has only one nonzero element 1, which the location is column index of corresponding class-specific sub-dictionary atom in D, $V_{-j} = [V_1, V_2, \dots, V_{j-1}, V_{j+1}, \dots, V_c] \in \mathbb{R}^{K\times(K-k_{c+1}-k_j)}$. Meanwhile, the small value of $\|Y_i - D_{c+1}X_i^{c+1}\|_F^2$ ensures that the shared sub-dictionary D_{c+1} make contribution to represent Y_i. Hence, the reconstruction error term Z_1 can be defined as follows:

$$Z_1(Y, D, X) = \sum_{i=1}^{c}\left[\|Y_i - D_{class}X_i^{class}\|_F^2 + \|Y_i - (D_{class}V_i)(V_i^TX_i^{class})\|_F^2 + \|Y_i - D_{c+1}X_i^{c+1}\|_F^2\right] \quad (4)$$

(2) SUB-DICTIONARY INCOHERENCE TERM Z_2

To exploit desirable discriminative capability of learned dictionary D, different sub-dictionaries should be as orthogonal as possible, which ensures that each class-specific sub-dictionary is exclusive to represent

corresponding samples well. Therefore, the value of structural incoherence constraint $\|D_j^T D_{-j}\|_F^2$, $\|D_j^T D_j - I_{k_j}\|_F^2$ are supposed to be small, where, $D_{-j} \in \mathbb{R}^{K\times(K-k_j)}$ is the sub-matrix by removing D_j from D, I_{k_j} is an identity matrix. By adding these two terms, the redundancy among sub-dictionaries would be reduced effectively, which has a direct impact on the speed of computation. Hence, the sub-dictionary incoherence term Z_2 can be defined as follows:

$$Z_2 = \alpha \sum_{j=1}^{c+1} \left[\frac{n_j}{k_j(K-k_j)} \|D_j^T D_{-j}\|_F^2 + \frac{n_j}{k_j^2} \|D_j^T D_j - I_{k_j}\|_F^2 \right] \quad (5)$$

where, $n_{c+1} = N$, $\frac{n_j}{k_j(K-k_j)}$ and $\frac{n_j}{k_j^2}$ can alleviate the effect of imbalance between the number of samples and atoms of sub-dictionaries.

(3) DISCRIMINATIVE PROMOTION TERM Z_3

By directly constraining coding vectors, the separability and discriminability of coding vectors from different classes is increased and further enhanced. Based on Fisher's linear discriminant, which maximizes the ratio of between-class scatter matrix to within-class scatter matrix, we can minimize the within-class scatter matrix $S_W(X)$ and maximum the between-class scatter matrix $S_B(X)$. Denote $S_W(X) = \sum_{i=1}^{c} \sum_{l=1}^{n_i} (\boldsymbol{x}_i^l - \boldsymbol{u}_i)(\boldsymbol{x}_i^l - \boldsymbol{u}_i)^T$ $S_B(X) = \sum_{i=1}^{c} n_i (\boldsymbol{u}_i - \boldsymbol{u})(\boldsymbol{u}_i - \boldsymbol{u})^T$, where, $\boldsymbol{x}_i^l$ denotes the coding vector of the l-th training sample over the i-th class-specific sub-dictionary, $\boldsymbol{u}_i = \frac{1}{n_i}\sum_{l=1}^{n_i} \boldsymbol{x}_i^l$ and $\boldsymbol{u} = \frac{1}{N}\sum_{i=1}^{c}\sum_{l=1}^{n_i} \boldsymbol{x}_i^l$ are mean vector of X_i and X, respectively. Thus, $\mathrm{tr}(S_W) = \sum_{i=1}^{c} \|X_i^{class} - U_i\|_F^2$, $\mathrm{tr}(S_B) = \sum_{i=1}^{c} \|U_i - U\|_F^2$, where, $U_i \in \mathbb{R}^{K\times n_i}$, each column equals to $\boldsymbol{u}_i$, $U \in \mathbb{R}^{K\times n_i}$, each column equals $\boldsymbol{u}$. Hence, the discriminative coding vector term Z_3 can be defined as follows:

$$Z_3 = \beta \sum_{i=1}^{c} \left(\|X_i^{class} - U_i\|_F^2 + \|U_i - U\|_F^2 + \|X_i^{class}\|_1 \right) \quad (6)$$

(4) LOW-RANK PRESERVING TERM Z_4

As nuclear norm $\|\cdot\|_*$ is the convex relaxation of rank($\cdot$), the low-rank preserving term Z_4 can be defined as follows:

$$Z_4 = \gamma \sum_{i=1}^{c} \|X_i^{c+1}\|_* \quad (7)$$

Taking all mentioned above into consideration, we have the following CASDDL model:

$$\begin{aligned} \min_{D,X} & \sum_{i=1}^{c} \left[\|Y_i - D_{class}X_i^{class}\|_F^2 + \|Y_i - (D_{class}V_i)(V_i^T X_i^{class})\|_F^2 + \|Y_i - D_{c+1}X_i^{c+1}\|_F^2 \right] + \\ & \alpha \sum_{j=1}^{c+1} \left[\frac{n_j}{k_j(K-k_j)} \|D_j^T D_{-j}\|_F^2 + \frac{n_j}{k_j^2} \|D_j^T D_j - I_{k_j}\|_F^2 \right] + \\ & \beta \sum_{i=1}^{c} \left(\|X_i^{class} - U_i\|_F^2 + \|U_i - U\|_F^2 + \|X_i^{class}\|_1 \right) + \\ & \gamma \sum_{i=1}^{c} \|X_i^{c+1}\|_* \end{aligned} \quad (8)$$

b: OPTIMIZATION OF CASDDL

Eq. (8) can be divided into two sub-problems: updating X with fixed D; updating D with fixed X. In order to learn a discriminative dictionary better, K-means algorithm is chosen to initialize the dictionary at first: each class-specific sub-dictionary is initialized as cluster centers of corresponding training samples, a shared sub-dictionary is initialized as cluster center of whole training samples. As the dissimilarity between different cluster centers is high, the initial atoms in class-specific sub-dictionaries obtain approximately discriminative ability. We summarize CASDDL in Algorithm 1.

Algorithm 1: Class-specific and Shared Discriminative Dictionary Learning

Input: Training samples $Y = \{Y_i\}_{i=1,2,\ldots,c}$; number of atoms k_j in class-specific dictionary $\{D_i\}_{i=1,2,\ldots,c}$; number of atoms k_{c+1} in shared sub-dictionary D_{c+1}; parameters α, β, and γ.

Initialize: The class-specific sub-dictionary $\{D_j\}_{j=1,2,\ldots,c}$ is initialized by K-means in Y_i, the shared sub-dictionary D_{c+1} is initialized by K-means in Y.

While not converged **do**

step 1: Update X_i^{class} by Eq. (13);

step 2: Update X_i^{c+1} by **Algorithm 2**;

step 3: Update $\{D_i\}_{i=1,2,\ldots,c}$ by Eq. (29);

step 4: Update D_{c+1} by Eq. (37);

End While

Output: The learned dictionary $D = \{D_j\}_{j=1,2,\ldots,c,c+1}$.

(1) UPDATE CODING MATRIX X

When D is fixed, Eq. (8) becomes a coding problem of computing $X = [X_1, X_2, \ldots, X_i, \ldots, X_c]$. When computing X_i, all X_j ($j \neq i$), are fixed, Eq. (8) can be simplified as follows:

$$\begin{aligned} \min_X & \sum_{i=1}^{c} \left(\|Y_i - D_{class}X_i^{class}\|_F^2 + \|Y_i - (D_{class}V_i)(V_i^T X_i^{class})\|_F^2 \right) + \beta \sum_{i=1}^{c} \left(\|X_i^{class} - U_i\|_F^2 + \|U_i - U\|_F^2 + \|X_i^{class}\|_1 \right) + \sum_{i=1}^{c} \|Y_i - D_{c+1}X_i^{c+1}\|_F^2 + \\ & \gamma \sum_{i=1}^{c} \|X_i^{c+1}\|_* \end{aligned} \quad (9)$$

① *Update* X_i^{class}

With fixed D and X_i^{c+1}, Eq. (9) can be rewritten as follows:

$$\begin{aligned} \min_{X_i^{class}} & \|Y_i - D_{class}X_i^{class}\|_F^2 + \|Y_i - (D_{class}V_i)(V_i^T X_i^{class})\|_F^2 + \beta \left(\|X_i^{class} - U_i\|_F^2 + \sum_{i=1}^{c} \|U_i - U\|_F^2 \right) + \beta \|X_i^{class}\|_1 \end{aligned} \quad (10)$$

It can be rewritten as follows:

$$\min_{X_i^{class}} R(X_i^{class}) + 2\omega \|X_i^{class}\|_1 \quad (11)$$

where, $R\left(X_i^{class}\right) = \left\|Y_i - D_{class}X_i^{class}\right\|_F^2 + \left\|Y_i - (D_{class}V_i)\left(V_i^T X_i^{class}\right)\right\|_F^2 + \beta\left(\left\|X_i^{class} - U_i\right\|_F^2 + \sum_{i=1}^{c}\|U_i - U\|_F^2\right)$, $\omega = \frac{\beta}{2}$.

According to [45], a two-step iterative shrinkage/ thresholding (TwIST) algorithm can be adopted to solve Eq. (11). After first derivative of $R\left(X_i^{class}\right)$ with respect to X_i^{class} is calculated (**Appendix 1**), we have

$$\begin{aligned}\nabla_{X_i^{class}} R\left(X_i^{class}\right) = {} & 2D_{class}^T\left(D_{class}X_i^{class} - Y_i\right) + \\ & 2V_iV_i^T D_{class}^T\left(D_{class}V_iV_i^T X_i^{class} - Y_i\right) + \\ & 2\beta\left\{X_i^{class}O_iO_i^T + X_i^{class}P_iP_i^T - RP_i^T + \right. \\ & \left.\sum_{\substack{j=1 \\ j\neq i}}^{c}\left[X_i^{class}Q_i^j\left(Q_i^j\right)^T - T_j\left(Q_i^j\right)^T\right]\right\}\end{aligned} \tag{12}$$

where, $E_i^j = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}_{n_i \times n_j}$, $O_i = I_{n_i \times n_i} - \frac{E_i^i}{n_i}$, $P_i = \frac{E_i^i}{n_i} - \frac{E_i^i}{N} = \frac{E_i^i}{n_i} - Q_i^i$, $R = \sum_{\substack{j=1 \\ j\neq i}}^{c} X_jQ_j^i$, $T_j = X_j\frac{E_j^j}{n_j} - \sum_{\substack{l=1 \\ l\neq i}}^{c} X_lQ_l^j$.

Then, we have

$$\begin{aligned}\left(X_i^{class}\right)^{(t+1)} = {} & (1-\xi)\left(X_i^{class}\right)^{(t-1)} + (\xi - \nu)\left(X_i^{class}\right)^{(t)} + \\ & \nu\Psi_{\frac{\tau}{\sigma}}\left[\left(X_i^{class}\right)^{(t)} - \frac{1}{2\sigma}\nabla_{X_i^{class}}F\left(X_i^{class}\right)\right]\end{aligned} \tag{13}$$

where, $\Psi_{\frac{\tau}{\sigma}}(\Sigma)$ denotes soft-thresholding shrinkage operator, $\Psi_{\frac{\tau}{\sigma}}(\Sigma) = \begin{cases}\Sigma_{ij} - \frac{\tau}{\sigma} & \Sigma_{ii} \geq \frac{\tau}{\sigma} \\ 0 & \Sigma_{ii} < \frac{\tau}{\sigma}\end{cases}$, Σ_{ij} stands for the (i,j)-th element of matrix Σ; $\left(X_i^{class}\right)^{(t-1)}$ is the previous value of X_i^{class}, $\left(X_i^{class}\right)^{(t)}$ is the current value of X_i^{class}, $\left(X_i^{class}\right)^{(t+1)}$ is the next value of X_i^{class}; $\xi > 0$, $\nu > 0$, $\sigma > 0$.

② *Update* X_i^{c+1}

With fixed D and X_i^{class}, Eq. (9) is further reduced to:

$$\min_{X_i^{c+1}}\|Y_i - D_{c+1}X_i^{c+1}\|_F^2 + \gamma\|X_i^{c+1}\|_* \tag{14}$$

According to inexact ALM algorithm, introducing the auxiliary variable $H = X_i^{c+1}$, Eq. (14) can be defined as follows:

$$\begin{aligned}\mathcal{O}(X_i^{c+1}, H, P, \mu) = {} & \|Y_i - D_{c+1}H\|_F^2 + \gamma\|X_i^{c+1}\|_* + \\ & \langle P, H - X_i^{c+1}\rangle + \frac{\mu}{2}\|H - X_i^{c+1}\|_F^2\end{aligned} \tag{15}$$

where, $\langle\cdot,\cdot\rangle$ means the inner product operator for two matrices; $\|\cdot\|_F^2$ denotes the Frobenius norm, which equals the sum of squares of each element of matrix; P is a Lagrange multiplier; $\mu > 0$ is a penalty parameter.

Furthermore, we have

$$\begin{aligned}\mathcal{O}(X_i^{c+1}, H, P, \mu) = {} & \frac{1}{2}\left\|H + \frac{P}{\mu} - X_i^{c+1}\right\|_F^2 + \\ & \frac{1}{\mu}\|Y_i - D_{c+1}H\|_F^2 + \frac{\gamma}{\mu}\|X_i^{c+1}\|_*\end{aligned} \tag{16}$$

The detailed procedure of solving Eq. (16) is presented in Algorithm 2.

Algorithm 2: Solving Eq. (16) via inexact ALM

Input: Training samples Y_i, shared sub-dictionary D_{c+1}, parameter $\gamma > 0$

Initialize: $H^{(0)} = (X_i^{c+1})^{(0)} = 0$, $P^{(0)} = 0$, $\mu^{(0)} = 0.1$, $\mu_{\max} = 10^5$, $\rho = 1.1$, $k = 0$, $k_{\max} = 10$

While not converged **do**

step 1: Update $H^{(k+1)}$ by Eq. (19);

step 2: Update $(X_i^{c+1})^{(k+1)}$ by Eq. (21);

step 3: Update $P^{(k+1)}$ by Eq. (22);

step 4: Update $\mu^{(k+1)}$ by Eq. (23);

step 5: Check the convergence condition $k < k_{\max}$;

step 6: Update k by $k = k + 1$;

End While

Output: The optimal solution X_i^{c+1}.

① *Update* H

$$\frac{1}{2}\left\|H + \frac{P}{\mu} - X_i^{c+1}\right\|_F^2 + \frac{1}{\mu}\|Y_i - D_{c+1}H\|_F^2 \tag{17}$$

Differentiating it with respect to H, and let it to be zero:

$$H + \frac{P}{\mu} - X_i^{c+1} + \frac{2}{\mu}D_{c+1}^T(Y_i - D_{c+1}H) = 0 \tag{18}$$

Then, we have

$$\begin{aligned}H^{(k+1)} = {} & \\ & \left(I - \frac{2}{\mu^{(k)}}D_{c+1}^T D_{c+1}\right)^{-1}\left[(X_i^{c+1})^{(k)} - \frac{2}{\mu^{(k)}}D_{c+1}^T Y_i - \frac{P^{(k)}}{\mu^{(k)}}\right]\end{aligned} \tag{19}$$

② *Update* X_i^{c+1}

$$\frac{1}{2}\left\|H + \frac{P}{\mu} - X_i^{c+1}\right\|_F^2 + \frac{\gamma}{\mu}\|X_i^{c+1}\|_* \tag{20}$$

Then, we have

$$(X_i^{c+1})^{(k+1)} = U\Psi_{\frac{\gamma}{\mu}}\{\Sigma\}V^T \tag{21}$$

where, $(U, \Sigma, V) = \mathrm{svd}\left(H^{(k+1)} + \frac{P^{(k)}}{\mu^{(k)}}\right)$, $\mathrm{svd}(\cdot)$ denotes SVD operation, $\Sigma = \mathrm{diag}(\{\sigma_i\}_{1\leq i\leq r})$ is the diagonal matrix with σ_1, σ_2,…, σ_r on the diagonal and zeros elsewhere, σ_i is the i-th singular value of $H^{(k+1)} + \frac{P^{(k)}}{\mu^{(k)}}$, $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{N\times r}$ are left, right singular matrices, respectively.

③ *Update* P

$$P^{(k+1)} = P^{(k)} + \mu^{(k)}\left(H^{(k+1)} - (X_i^{c+1})^{(k+1)}\right) \tag{22}$$

④ *Update* μ

$$\mu^{(k+1)} = \min\left(\rho\mu^{(k)}, \mu_{\max}\right) \tag{23}$$

where, $\rho = 1.1$, $\mu_{\max} = 10^5$.

(2) UPDATE DICTIONARY D

When X is fixed, $\{D_j\}_{j=1, 2, \ldots, c, c+1}$ can be updated one by one. Eq. (8) can be simplified as follows:

$$\begin{aligned}\min_D \sum_{i=1}^{c}\Big(\|Y_i - D_{class}X_i^{class}\|_F^2 + {} & \\ \left\|Y_i - (D_{class}V_i)\left(V_i^T X_i^{class}\right)\right\|_F^2 + \|Y_i - D_{c+1}X_i^{c+1}\|_F^2\Big) + & \\ \alpha\sum_{j=1}^{c+1}\left[\frac{n_j}{k_j(K-k_j)}\left\|D_j^T D_{-j}\right\|_F^2 + \frac{n_j}{k_j^2}\left\|D_j^T D_j - I_{k_j}\right\|_F^2\right] & \end{aligned} \tag{24}$$

① *Update* $\{D_i\}_{i=1,2,\dots,c}$

With fixed X and other sub-dictionaries, Eq. (24) can be rewritten as follows:

$$\min_{D_i}\sum_{i=1}^{c}\Big(\big\|Y_i - D_{class}X_i^{class}\big\|_F^2 + \big\|Y_i - (D_{class}V_i)\big(V_i^T X_i^{class}\big)\big\|_F^2\Big) + \alpha\frac{n_i}{k_i(K-k_i)}\|D_i^T D_{-i}\|_F^2 + \alpha\frac{n_i}{k_i^2}\big\|D_i^T D_i - I_{k_i}\big\|_F^2 \quad (25)$$

We optimize $\{D_i\}_{i=1,2,\dots,c}$ class-by-class and meanwhile, make all other D_j ($j\neq i$) fixed. Then, we have

$$\min_{D_i}\big\|Y - D_{class}(X_1^{class},\cdots,X_c^{class})\big\|_F^2 + \big\|Y - D_{class}(V_1V_1^TX_1^{class},\cdots,V_cV_c^TX_c^{class})\big\|_F^2 + \alpha\frac{n_i}{k_i(K-k_i)}\|D_i^T D_{-i}\|_F^2 + \alpha\frac{n_i}{k_i^2}\big\|D_i^T D_i - I_{k_i}\big\|_F^2 \quad (26)$$

Denote $A = [Y, Y] \in \mathbb{R}^{d\times 2N}$, $C = [X_1^{class},\cdots,X_c^{class},V_1V_1^TX_1^{class},\cdots,V_cV_c^TX_c^{class}] \in \mathbb{R}^{(K-k_{c+1})\times 2N}$, we have

$$\min_{D_i}\|A - D_{class}C\|_F^2 + \alpha\frac{n_i}{k_i(K-k_i)}\|D_i^T D_{-i}\|_F^2 + \alpha\frac{n_i}{k_i^2}\big\|D_i^T D_i - I_{k_i}\big\|_F^2 \quad (27)$$

Therefore

$$\min_{D_i}\left\|A - \sum_{\substack{j=1\\ j\neq i}}^{c} D_jC^j - D_iC^i\right\|_F^2 + \alpha\frac{n_i}{k_i(K-k_i)}\|D_i^T D_{-i}\|_F^2 + \alpha\frac{n_i}{k_i^2}\big\|D_i^T D_i - I_{k_i}\big\|_F^2 \quad (28)$$

Denote $\tilde{A} = A - \sum_{\substack{j=1\\ j\neq i}}^{c} D_jC^j \in \mathbb{R}^{d\times 2N}$, we have

$$\min_{D_i}\big\|\tilde{A} - D_iC^i\big\|_F^2 + \alpha\frac{n_i}{k_i(K-k_i)}\|D_i^T D_{-i}\|_F^2 + \alpha\frac{n_i}{k_i^2}\big\|D_i^T D_i - I_{k_i}\big\|_F^2 \quad (29)$$

where, $B^i \in \mathbb{R}^{k_i\times 2N}$.

Eq. (29) can be solved by a coherence regularized (CORE) algorithm [46].

② *Update* D_{c+1}

With fixed X and all the class-specific sub-dictionaries, Eq. (24) can be rewritten as follows:

$$\min_{D_{c+1}}\sum_{i=1}^{c}\|Y_i - D_{c+1}X_i^{c+1}\|_F^2 + \alpha\frac{n_{c+1}}{k_{c+1}(K-k_{c+1})}\big\|D_{c+1}^T D_{-(c+1)}\big\|_F^2 + \alpha\frac{n_{c+1}}{k_{c+1}^2}\big\|D_{c+1}^T D_{c+1} - I_{k_{c+1}}\big\|_F^2 \quad (30)$$

Denote $X^{c+1} = [X_1^{c+1}, X_2^{c+1}, \dots, X_c^{c+1}] \in \mathbb{R}^{k_{c+1}\times N}$ is the coding matrix of Y over shared sub-dictionary D_{c+1}, then we have

$$\min_{D_{c+1}}\|Y - D_{c+1}X^{c+1}\|_F^2 + \alpha\frac{n_{c+1}}{k_{c+1}(K-k_{c+1})}\|D_{c+1}^T D_{class}\|_F^2 + \alpha\frac{n_{c+1}}{k_{c+1}^2}\big\|D_{c+1}^T D_{c+1} - I_{k_{c+1}}\big\|_F^2 \quad (31)$$

Similar to Eq. (29), Eq. (31) can be solved by CORE algorithm.

2) DEFECT CLASSIFICATION

The proposed CASDDL especially emphasizes class discrimination of both dictionary atoms and coding vector, which not only contributes for learning class-oriented discriminative dictionary, but also results in discriminative coding vector. Different from traditional classification method that treat the coding vector just as input to sophisticated classifiers, we can directly make full use of the discriminative capability of coding vector for a simple and efficient classification scheme, without adding any parameters to be learned.

For a test sample $\hat{\boldsymbol{y}}$, we use the obtained dictionary D to compute its coding vector $\hat{\boldsymbol{x}} = [\boldsymbol{x}_1; \boldsymbol{x}_2; \dots \boldsymbol{x}_i; \dots; \boldsymbol{x}_c]$, where, $\boldsymbol{x}_i$ is the coding sub-vector associated with class-specific sub-dictionary $\{D_i\}_{i=1,2,\dots,c}$. Considering the discrimination of $\hat{\boldsymbol{x}}$, if $\hat{\boldsymbol{y}}$ is from class i, $\boldsymbol{x}_i$ will be large than other part. Therefore, the class of $\hat{\boldsymbol{y}}$ is determined by $\arg\max_i \|\boldsymbol{x}_i\|_2^2$.

B. ACCURATE SEGMENTATION

The proposed DLMD-based segmentation method mainly comprises of four stages, including superpixel over-segmentation, feature extraction, feature matrix decomposition, and defect segmentation.

1) SUPERPIXEL OVER-SEGMENTATION

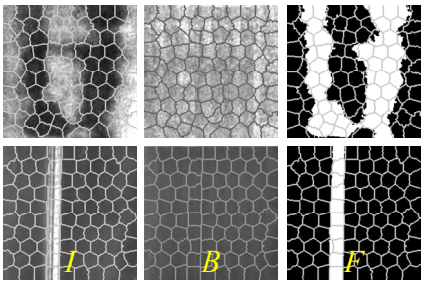
In order to capture structural information of defect, we adopt the superpixel-algorithm of adaptive simple linear iterative clustering (ASLIC) [47] to partition the surface defect image into several non-overlap sub-regions. It can generate regular shaped superpixels in both textured and non-textured regions alike. Only the number of superpixel sub-regions K should be specified. The bigger K should be chosen if the potential defect object is small and morphological complex, which can produce more deformable shape to enclose the region containing potential defect object, vice versa. As the number of superpixel sub-regions is far less than the pixel of image, which can ease the computational burden and improve the computation efficiency.

2) FEATURE EXTRACTION

The feature of gray-scale, Gabor filters with eight directions on two different scales, steerable pyramid filters with four directions on two different scales are computed and then stacked vertically to construct a 25-dimensional feature vector for each pixel. For each superpixel sub-region, its feature vector is calculated by taking mean of all the feature vectors of pixels contained in it, which is robust to noise. All the feature vectors of sub-regions are normalized into unit column vectors, and are stacked together to construct a feature matrix $D \in \mathbb{R}^{d\times K}$, where, d is the dimension of feature vector, K is the number of superpixels sub-regions.

3) FEATURE MATRIX DECOMPOSITION

a: Formulation of DLMD



(a) (b) (c)

FIGURE 4. Illustration of surface defect image decomposition with double low-rank assumption: (a) original surface defect image I by superpixel over-segmentation; (b) defect-free background image B by superpixel over-segmentation; (c) defect foreground image E by superpixel over-segmentation.

As shown in Fig. 4, we try to decompose surface defect image I into defect-free background image B and defect foreground image E. According to the ASLIC algorithm and stack all feature vector of superpixel sub-regions together to form feature matrix F constructed from the original defect image I, feature matrix L represents a background image B, and a feature matrix S represents a defect foreground image E in a certain feature space, respectively. Therefore, $F = L + S$, where, each column of these matrices stand for the feature vector of individual superpixel sub-regions. Both the background image B and the defect foreground image E contain multiple homogeneous and highly similar sub-regions. These two feature matrices L and S have redundant information and can be assumed to have low-rank due to the similarity among different sub-regions, which form a low-dimensional feature subspace. What's more, in order to reduce the influence of noises and improve the robustness to uneven illumination simultaneously, we assume that the background has the sparse property and lies in a sparse feature subspace.

Based on above analysis, the proposed DLMD can be modelled as the following optimization problem:

$$\begin{aligned} &\min_{L,\,S}\left(\operatorname{rank}(L)+\operatorname{rank}(S)+\eta\Theta(L,S)+\tau\|L\|_0\right)\\ &\text{s.t. } F=L+S \end{aligned} \tag{32}$$

where, $\Theta(L,S)$ denotes the regularization term to enlarge the margin and reduce the coherence between the feature subspaces induced by L and S; $\eta>0$, $\tau>0$ are regularization parameters.

The local invariance assumption based Laplacian regularization term $\Theta(L,S)$ can be defined as follows:

$$\Theta(L,S)=\frac{1}{2}\sum_{i,j=1}^{K}\left\|\boldsymbol{s}_i-\boldsymbol{s}_j\right\|_2^2 w_{ij}=\operatorname{tr}(SMS^T) \tag{33}$$

where, $M\in\mathbb{R}^{K\times K}$ is a Laplacian matrix; $\operatorname{tr}(\cdot)$ denotes the trace of a matrix; $\boldsymbol{s}_i$, $\boldsymbol{s}_j$denotes the i-th and j-th column of S; w_{ij} of affinity matrix $W\in\mathbb{R}^{K\times K}$ denotes the weight that represents the feature similarity between sub-regions R_i and R_j.

Supposing that each sub-region of surface defect image is represented by a node, the Laplacian matrix M is defined:

$M_{ij}=\begin{cases}-w_{ij} & i\neq j\\ \sum_{i\neq j}w_{ij} & \text{otherwise}\end{cases}$, where, W is an affinity matrix, when R_i and R_j are directly adjacent, $w_{ij}=\exp\left(\frac{-\|\boldsymbol{p}_i-\boldsymbol{p}_j\|_2^2}{2\sigma_{\boldsymbol{p}}^2}\right)\exp\left(\frac{-\|\bar{\boldsymbol{f}}_i-\bar{\boldsymbol{f}}_j\|_2^2}{2\sigma_{\boldsymbol{f}}^2}\right)$, otherwise, $w_{ij}=0$; $\boldsymbol{p}_i\in\mathbb{R}^2$ and $\boldsymbol{p}_j\in\mathbb{R}^2$ denote the central coordinate of R_i and R_j; $\bar{\boldsymbol{f}}_i\in\mathbb{R}^d$ and $\bar{\boldsymbol{f}}_j\in\mathbb{R}^d$ denote the feature vector of R_i and R_j; $\exp\left(\frac{-\|\boldsymbol{p}_i-\boldsymbol{p}_j\|_2^2}{2\sigma_{\boldsymbol{p}}^2}\right)$ represents the spatial contiguity between R_i and R_j; $\exp\left(\frac{-\|\bar{\boldsymbol{f}}_i-\bar{\boldsymbol{f}}_j\|_2^2}{2\sigma_{\boldsymbol{f}}^2}\right)$ gives the feature similarity between R_i and R_j; $\sigma_{\boldsymbol{p}}$ and $\sigma_{\boldsymbol{f}}$ are two scalars.

b: Optimization of DLMD

Eq. (32) can be converted into the following optimization problem:

$$\begin{aligned} &\min_{L,\,S}\left(\|L\|_*+\|S\|_*+\eta\operatorname{tr}(SMS^T)+\tau\|L\|_{2,1}\right)\\ &\text{s.t. } F=L+S \end{aligned} \tag{34}$$

where, $l_{2,1}$ norm-based penalty term $\|L\|_{2,1}$ aims to characterize the noise or illumination interference of surface defect image.

According to inexact ALM algorithm, introducing the auxiliary variables $H=L$, $J=S$, Eq. (34) can be defined as follows:

$$\begin{aligned} &\mathcal{O}(L,S,H,J,Y_1,Y_2,Y_3,\mu)=\\ &\|L\|_*+\|S\|_*+\eta\operatorname{tr}(JMJ^T)+\tau\|H\|_{2,1}+\langle P_1,F-L-S\rangle+\\ &\frac{\mu}{2}\|D-L-S\|_F^2+\langle P_2,H-L\rangle+\frac{\mu}{2}\|H-L\|_F^2+\\ &\langle P_3,J-S\rangle+\frac{\mu}{2}\|J-S\|_F^2 \end{aligned} \tag{35}$$

where, P_1, P_2 and P_3 are Lagrange multipliers; $\mu>0$ is a penalty parameter.

The detailed procedure of solving Eq. (35) is presented in Algorithm 3.

① *Update* H

In order to solve H, we can further simplify Eq. (35) as follows:

$$\min_H\left(\frac{1}{2}\left\|L-\frac{P_2}{\mu}-H\right\|_F^2+\frac{\tau}{\mu}\|H\|_{2,1}\right) \tag{36}$$

The optimal solution can be obtained as follows:

$$H^{(k+1)}(:,j)=\begin{cases}\frac{\|Z^{(k)}(:,j)\|_2-\frac{\tau}{\mu^{(k)}}}{\|Z^{(k)}(:,j)\|_2}Z^{(k)}(:,j) & \left\|Z^{(k)}(:,j)\right\|_2>\frac{\tau}{\mu^{(k)}}\\ 0 & \text{otherwise}\end{cases} \tag{37}$$

where $Z^{(k)}=L^{(k)}-\frac{P_2^{(k)}}{\mu^{(k)}}$, $Z(:,j)$ denotes the j-th column of matrix Z.

② *Update* J

In order to solve J, the optimal solution can be obtained as follows:

$$\min_J \left(\frac{1}{2} \left\| J - S + \frac{P_3}{\mu} \right\|_F^2 + \frac{\eta}{\mu} \mathrm{tr}(JMJ^T) \right) \tag{38}$$

Differentiating it with respect to J, and let it to be zero:

$$J - S + \frac{P_3}{\mu} + \frac{2\eta}{\mu} JM = 0 \tag{39}$$

The close-form solution can be obtained as follows:

$$J^{(k+1)} = \left(S^{(k)} - \frac{P_3^{(k)}}{\mu^{(k)}} \right) \left(I + \frac{2\eta}{\mu^{(k)}} M \right)^{-1} \tag{40}$$

③ *Update* L

To solve L, Eq. (12) can be transformed to Eq. (22):

$$\min_L \left(\frac{1}{2} \left\| F - S + \frac{P_1}{\mu} - L \right\|_F^2 + \frac{1}{2} \left\| H + \frac{P_2}{\mu} - L \right\|_F^2 + \frac{1}{\mu} \|L\|_* \right) \tag{41}$$

It can be rewritten as follows:

$$\min_L \left(\frac{1}{2} \left\| \frac{1}{2} \left(F - S + H + \frac{P_1 + P_2}{\mu} \right) - L \right\|_F^2 + \frac{1}{4\mu} \|L\|_* \right) \tag{42}$$

The optimal solution can be obtained by Eq. (21):

$$L^{(k+1)} = U \Psi_{\frac{w}{4\mu^{(k)}}}(\Sigma) V^T \tag{43}$$

where, $(U, \Sigma, V) = \mathrm{svd}\left[\frac{1}{2} \left(F - S^{(k)} + H^{(k+1)} + \frac{P_1^{(k)} + P_2^{(k)}}{\mu^{(k)}} \right) \right]$; $\Psi_{\frac{w}{4\mu}}(\cdot)$ denotes non-uniform singular value thresholding operator, $\{\sigma_i\}_{i=1,2,\ldots,r}$ is the singular value of $\frac{1}{2} \left(F - S^{(k)} + H^{(k+1)} + \frac{P_1^{(k)} + P_2^{(k)}}{\mu^{(k)}} \right)$, $w_i = \frac{\sum_{j=1}^r \sigma_j}{\sigma_i}$.

④ *Update* S

In order to solve S, Eq. (35) can be transformed as follows:

$$\min_S \left(\frac{1}{2} \left\| F - L + \frac{P_1}{\mu} - S \right\|_F^2 + \frac{1}{2} \left\| J + \frac{P_3}{\mu} - S \right\|_F^2 + \frac{1}{\mu} \|S\|_* \right) \tag{44}$$

It can be rewritten as follows:

$$\min_S \left(\frac{1}{2} \left\| \frac{1}{2} \left(F - L + J + \frac{P_1 + P_3}{\mu} \right) - S \right\|_F^2 + \frac{1}{4\mu} \|S\|_* \right) \tag{45}$$

Its solution is

$$S^{(k+1)} = U \Psi_{\frac{w}{4\mu}}(\Sigma) V^T \tag{46}$$

where, $(U, \Sigma, V) = \mathrm{svd}\left[\frac{1}{2} \left(F - L^{(k+1)} + J^{(k+1)} + \frac{P_1^{(k)} + P_3^{(k)}}{\mu^{(k)}} \right) \right]$; $\{\sigma_i\}_{i=1,2,\ldots,r}$ is the singular value of $\frac{1}{2} \left(F - L^{(k+1)} + J^{(k+1)} + \frac{P_1^{(k)} + P_3^{(k)}}{\mu^{(k)}} \right)$, $w_i = \frac{\sum_{j=1}^r \sigma_j}{\sigma_i}$.

⑤ *Update* P_1, P_2 *and* P_3

$$\begin{aligned} P_1^{(k+1)} &= P_1^{(k)} + \mu^{(k)} \left(F - L^{(k+1)} - S^{(k+1)} \right) \\ P_2^{(k+1)} &= P_2^{(k)} + \mu^{(k)} \left(H^{(k+1)} - L^{(k+1)} \right) \\ P_3^{(k+1)} &= P_3^{(k)} + \mu^{(k)} \left(J^{(k+1)} - S^{(k+1)} \right) \end{aligned} \tag{47}$$

⑥ *Update* μ

$$\mu^{(k+1)} = \min\left(\rho \mu^{(k)}, \mu_{\max} \right) \tag{48}$$

where, $\rho = 1.1$, $\mu_{\max} = 10^5$.

4) DEFECT SEGMENTATION

Each column of $L = (\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_K)$ and $S = (\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K)$ are the feature vector of corresponding superpixel sub-region of decomposed background image B and defect foreground image E, respectively. Then, we transfer L and S from feature domain to spatial domain for visualizing. The gray-value of each superpixel sub-region is maximum value of corresponding feature vector, then allocating it to corresponding pixels to visualize background image B and defect foreground image E, as shown in Fig. 2.

To enhance the completeness of defect objects and suppress the background noise in defect foreground image E, the regression optimization algorithm is adopted as follows:

$$\min_{s_i} \left(\sum_{i=1}^K w_i^f (s_i - 1)^2 + \sum_{i=1}^K w_i^b s_i^2 + \sum_{i,j=1}^K w_{ij} \left(s_i - s_j \right)^2 \right) \tag{49}$$

where, w_i^f and w_i^b denotes gray-value of sub-region in defect foreground image E and background image B, respectively; $s_i \in \boldsymbol{s} = (s_1, s_2, \cdots, s_K)^T$ denotes the enhanced gray-value of i-th sub-region of defect foreground image E.

Following $W^b = \mathrm{diag}[(w_1^b, w_2^b, \cdots, w_K^b)^T] \in \mathbb{R}^{K \times K}$, $W^f = \mathrm{diag}\left[\left(w_1^f, w_2^f, \cdots, w_K^f \right)^T \right] \in \mathbb{R}^{K \times K}$, Eq. (49) can be reformulated as follows:

$$\min_{\boldsymbol{s}} (\boldsymbol{s}^T W^b \boldsymbol{s} + \boldsymbol{s}^T W^f \boldsymbol{s} - 2W^f \boldsymbol{s} + W^f \mathbf{1} + 2\boldsymbol{s}^T M \boldsymbol{s}) \tag{50}$$

where, $\mathbf{1} \in \mathbb{R}^{K \times 1}$ denotes the unit vector; $M \in \mathbb{R}^{K \times K}$ denotes the same Laplacian matrix in Eq. (33).

Differentiating Eq. (50) with respect to $\boldsymbol{s}$, and let it to be zero, we have

$$2W^b \boldsymbol{s} + 2W^f \boldsymbol{s} - 2W^f \mathbf{1} + 4M\boldsymbol{s} = 0 \tag{51}$$

Its solution is

$$\boldsymbol{s} = (W^f + W^b + 2M)^{-1} W^f \mathbf{1} \tag{52}$$

Through Eq. (49), the gray-value of defect sub-region in defect foreground image E will become bigger, so the defect object can be highlighted further. Finally, the shape, location and size of surface defect can be easily localized and segmented through a simple thresholding operation.

IV. EXPERIMENT

In this section, various experiments, such as parameters analysis, convergence analysis, robustness to noise, comparisons between our method and some state-of-the-art methods, are conducted to verify the proposed JCS method.

A. EXPERIMENTAL SETUP

Two typical surface defects images (Patch, Scratch) and defect-free image are selected in the following experiments. There are 300 grayscale images (200×200 pixels) per class, and the pixel-level ground truth of defect image is manually marked by using white to denote defective pixels and black to denote defect-free pixels. We evaluated classification results using classification accuracy N_R/N, where, N_R is the number of test samples that are correctly classified, N is the total number of test samples. All the surface images are normalized and resized to 40×40 pixels, then randomly divide into training samples and test samples in 1:1 ratio. We repeated each experiment ten times, and the average

values and standard deviations of the classification results are given. We evaluated segmentation results using qualitative and quantitative metrics: the qualitative metrics refer to human subjective feeling for segmentation performance (i.e., boundary of defect object is clear, contrast between defect and background is obvious); the quantitative metrics refer to precision-recall (P-R) curve, receiver operating characteristic (ROC) curve, average F-Measure (F_β) curve, area under ROC curve (AUC) and mean square error (MAE). Supposing that the pixel belonging to defect is defined as a positive example, and the pixel belonging to background is defined as a negative example. The symbols *TP* (True Positive), *TN* (True Negative), *FP* (False Positive), and *FN* (False Negative) correspond to the number of defect pixel correctly recognized as defect object, the number of background pixel correctly recognized as background, the number of background pixel mistakenly recognized as defect object, and the number of defect pixel mistakenly recognized as background, respectively. Then, Precision, Recall, TPR (True Positive Rate), FPR (False Positive Rate), F_1, and MAE are computed as follows: $\text{Precision} = \frac{TP}{TP+FP}$, $\text{Recall} = \frac{TP}{TP+FN}$, $\text{TPR} = \frac{TP}{TP+FN}$, $\text{FPR} = \frac{FP}{FP+TN}$, $F_1 = \frac{2}{N}\sum_{i=1}^{N}\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$, $\text{MAE} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}|BW(i,\ j)-G(i,\ j)|}{H\times W}$, where, N, H and W denotes the number, height and width of surface defect image.

B. CLASSIFICATION RESULTS ANALYSIS

1) PARAMETERS ANALYSIS

Table 1. Classification accuracy of CASDDL with different parameters α, β, and γ, fix α = 0.1 to tune β and γ.

γ \ β	0	0.6	0.8	1	
0	0.8711	0.9105	0.9214	0.9105	0.9034
0.5	0.8867	0.9134	0.9256	0.9096	0.9088
0.7	0.8892	0.9141	**0.9287**	0.9136	0.9114
0.9	0.8873	0.9125	0.9222	0.9151	0.9093
	0.8836	0.9126	0.9245	0.9122	

The tuning three regularization parameters α, β, γ in Eq. (8) are chosen by 5-fold cross validation. α controls mutual incoherence between each sub-dictionary, β controls discrimination of coding vectors over all the class-specific sub-dictionaries, γ controls the low-rank ability of coding vector over the shared sub-dictionary. Following the work in [18], we set $\alpha = 0.1$, then search for their best values in a small set {0, 0.6, 0.8, 1.0}, {0, 0.5, 0.7, 0.9}, respectively.

Let k_c, k_s denotes number of atoms of class-specific sub-dictionary and shared sub-dictionary, respectively. We vary k_c from 10 to 45 with five interval, k_s from 2 to 30 with four interval. For each parameter combination, we compute the classification accuracy of all the sub-dictionary combinations in terms of mean value, and illustrate the classification accuracy in Table 1. The bottom row of Table 1 denotes the mean value of classification accuracy with one β corresponding to different γ, the right column of Table 1 denotes the mean value of classification accuracy with one γ corresponding to different β. As shown in Table 1, classification accuracy rises as the increase of β at first, but a further increase of β over a proper value will decrease the classification performance. The classification accuracy will degrade with a small value of β, which shows that the discriminative coding vector term is useful in learning class-oriented dictionary. Comparably, a larger value of γ will capture the inter-class similarity, and the shared sub-dictionary is more readily to capture the commonality features. However, too large value of γ will decrease the representation ability of shared sub-dictionary, the classification performance will be degraded. For γ, we empirically observe that a value lying in the range [0.5, 0.9] can always achieve an acceptable result. Furthermore, the classification accuracy with $\gamma = 0$ are lower than that with $\gamma = 0.7$, which illustrates the importance of low-rank term.

From Table 1, the highest classification accuracy is achieved when $\alpha = 0.1$, $\beta = 0.8$, and $\gamma = 0.7$, and this parameter combination will be adopted in the following experiments. Besides, we observe that the classification accuracy is robust to different parameter combinations being greater than 89% in most cases.

2) CONVERGENCE ANALYSIS

Although Eq. (8) is non-convex, the optimization algorithm actually adopts an alternatively updating fashion, and the convergence of each sub-problem can be guaranteed. On the one hand, for updating X with D fixed, the optimal solution is gained by TwIST and ALM algorithms. On the other hand, in the process of updating D with X fixed, each atom is optimally renewed for the sub-problem, and the optimal solution is gained by CORE algorithm. As a consequence, the objective function is non-increasing during the whole process of alternatively updating X and D. In addition, we provide the empirical evidence to illustrate the good convergence behavior of CASDDL in Fig. 5. With the increase of iteration numbers, the curve of error gradually decreases and eventually becomes stable, and the curve of accuracy increases for different combination of sub-dictionaries. It shows the proposed CASDDL enjoys a good convergence performance.

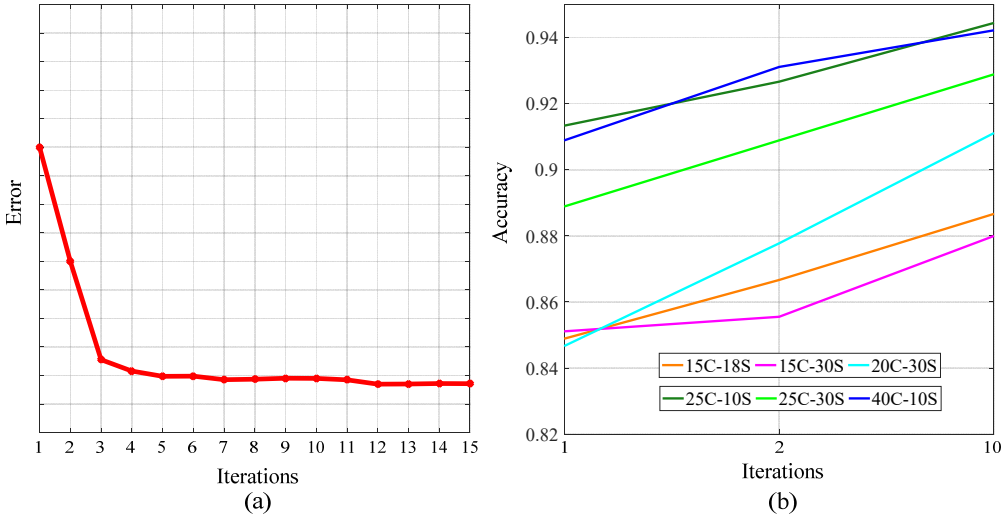


FIGURE 5. **Convergence ability of CASDDL.**

3) COMPUTATIONAL COMPLEXITY

The drawback of CASDDL is that it is computationally more complex. Although dictionary learning can be done in parallel and off-line, it is still important to see how long the dictionary learning process would take. A number of experimental parameters can affect the run time of CASDDL, including the number of classes, number of training samples, dictionary size and dimension of feature vectors.

4) ROBUSTNESS TO NOISE

We evaluate the robustness of the proposed CASDDL by corrupting original surface images with additive Gaussian noise in different signal to noise ratio (SNR), including 24dB, 20dB, and 16dB. As shown in Table 2, the classification accuracy is decreased slower when the noise level is increased; CASDDL can achieve 80.81% classification accuracy even at 20 dB noise, which is considered as less sensitive to noise.

Table 2. **Classification accuracy of CASDDL with different noise level.**

SNR (dB)	Classification Accuracy (%)
24	85.70±0.68
20	80.81±1.09
16	72.22±1.34

5) NUMBER OF ATOMS IN SUB-DICTIONARY

Supposing k_c, k_s denotes number of atoms of class-specific sub-dictionary and shared sub-dictionary, respectively. As shown in Table 3, we can observe that increasing k_c will lead to a higher classification performance. The possible reason is that more discriminative information can be captured by a larger class-specific sub-dictionary. When k_s is fixed, the classification accuracy is dropped as the increase of k_c. The possible reason is that smaller shared sub-dictionary is enough to capture the shared features of defect images, and larger shared sub-dictionary tends to absorb class-specific features into the shared sub-dictionary, causing some discriminative information lost. The proposed CASDDL always achieves higher classification accuracy despite different number of atoms, which indicates that it has a better ability to reconstruct defect images, even if learned dictionary has a small size. In fact, larger size of dictionary may have stronger representative ability and achieve better classification performance at the expense of increasing computational load. Therefore, we should make a tradeoff between classification performance and computational efficiency. When k_c = 30, the classification accuracy gain is merely promoted very little (~1%) as the increase of k_c. When k_s = 2 and k_c = 30, CASDDL can still have higher classification accuracy 94.89%, and this parameter combination will be used in the following experiments.

6) VISUALIZATION OF CODING VECTORS

The proposed CASDDL aims to get highly-discriminative coding vectors, through the learned discriminative dictionary, to achieve surface defect classification. Fig. 6 illustrates the coding vectors of training and testing samples are approximately block-diagonal, which further shows the class-label discriminative information in coding vectors.

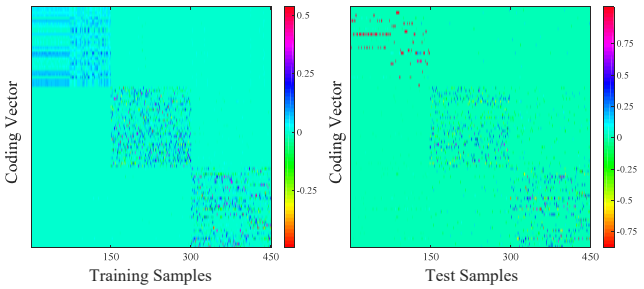


FIGURE 6. **Visualization of coding vectors of CASDDL in training and testing process.**

Table 3. Classification accuracy of CASDDL with different number of atoms k_c and k_s

k_s \ k_c	10	15	20	25	30	35	40	45
2	0.8407	0.8993	0.9282	0.9438	0.9489	0.9516	0.9544	0.9560
6	0.8256	0.8842	0.9144	0.9284	0.9404	0.9456	0.9458	0.9436
10	0.8158	0.8813	0.9173	0.9324	0.9402	0.9349	0.9402	0.9447
14	0.8298	0.8904	0.9158	0.9307	0.9296	0.9431	0.9404	0.9471
18	0.8209	0.8762	0.9109	0.9273	09404	0.9411	0.9396	0.9460
22	0.8222	0.8733	0.9067	0.9249	0.9364	0.9420	0.9444	0.9427
26	0.8038	0.8667	0.9000	0.9238	0.9324	0.9458	0.9449	0.9458
30	0.7949	0.8638	0.9073	0.9218	0.9349	0.9400	0.9413	0.9456

7) CLASSIFICATION RESULTS COMPARISON

We compare the proposed CASDDL with other well-known dictionary learning methods, including SRC [11], DLSI [13], CODDL [14], DFEDTL [15], COPAR [19], RDCDL [20], LRSDL [24], LRGPDDL [25], ALRR [26].

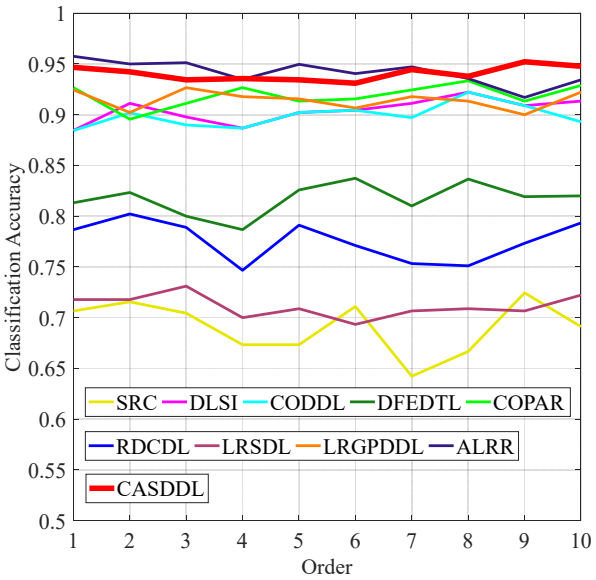


FIGURE 7. Visualization of classification accuracy between CASDDL and other approaches.

As shown in Fig. 7 and Table 4, CASDDL achieves 94.07% classification accuracy, compared to 91.89% for COPAR, 90.42% for DLSI, 89.90% for CODDL and 81.72 for DFEDTL. Compared to SRC, which is the baseline method in the experiment, CASDDL improves the classification accuracy with a margin of more than 24%. Among above approaches, ALRR performs the best, which is superior to ours by 0.11% for accuracy, and is inferior ours by stability. Besides, CASDDL outdoes LRGPDDL by a significant improvement of above 2.5%.

Table 4. Performance comparison between CASDDL and other approaches.

Method	Classification Accuracy (%)
SRC [11]	69.09±2.6
DLSI [13]	90.42±1.2
CODDL [14]	89.90±1.1
DFEDTL [15]	81.72±1.6
COPAR [19]	91.89±1.1
RDCDL [20]	77.58±2.0
LRSDL [24]	71.13±1.1
LRGPDDL [25]	91.47±0.9
ALRR [26]	94.18±1.6
CASDDL	**94.07±0.7**

C. SEGMENTATION RESULTS ANALYSIS

1) PARAMETERS ANALYSIS

The tuning two regularization parameters η, τ in Eq. (34) are chosen by 5-fold cross validation, and the experimental results measured by AUC metric are shown in Table 5. Its show that when the values of η and τ are set properly, the proposed DLMD can achieve better segmentation performance. When η is small, the performance is very sensitive to the changes of τ; while η is big, τ is insensitivity. Especially, it would be better to set the values of η much larger than that of τ in order to penalize the feature matrix of defect-free background image to be sparse. The segmentation performance reaches a high level when $\eta = 1.25$ and $\tau = 0.25$, and this parameter combination will be used in the following experiments.

Table 5. Experimental results of DLMD with different parameters η and τ.

η \ τ	0.05	0.15	0.25	0.35	0.45	0.55
0.25	0.809765	0.620011	0.617729	0.617857	0.617848	0.617807
0.5	0.828737	0.787541	0.716342	0.688683	0.683770	0.681919
0.75	0.820643	0.833686	0.804402	0.750935	0.713586	0.702458
1	0.817401	0.834645	0.842852	0.818865	0.778656	0.740081
1.25	0.813881	0.834150	**0.845304**	0.832574	0.821758	0.799830
1.5	0.810798	0.833805	0.834686	0.835088	0.833225	0.826053

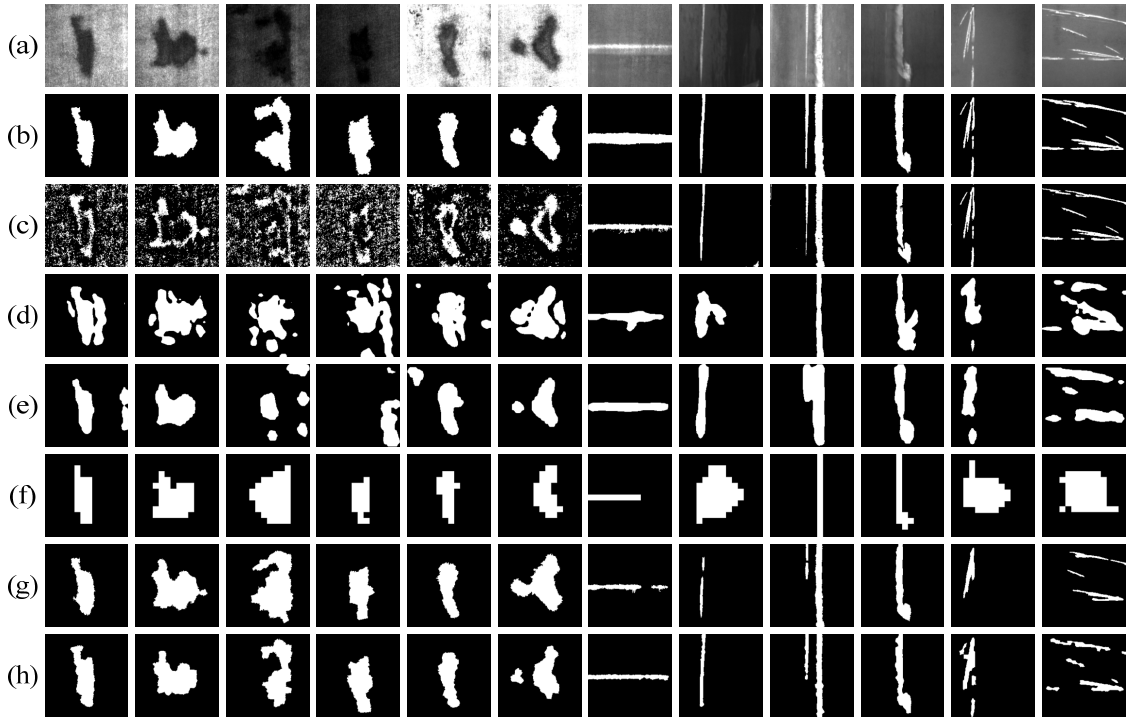


FIGURE 9. Qualitative comparison: (a) input image; (b) manual-labeled ground-truth image; (c) RPCA; (d) SSD; (e) PG-LSR; (f) W-LRR (g) ESP; (h) DLMD.

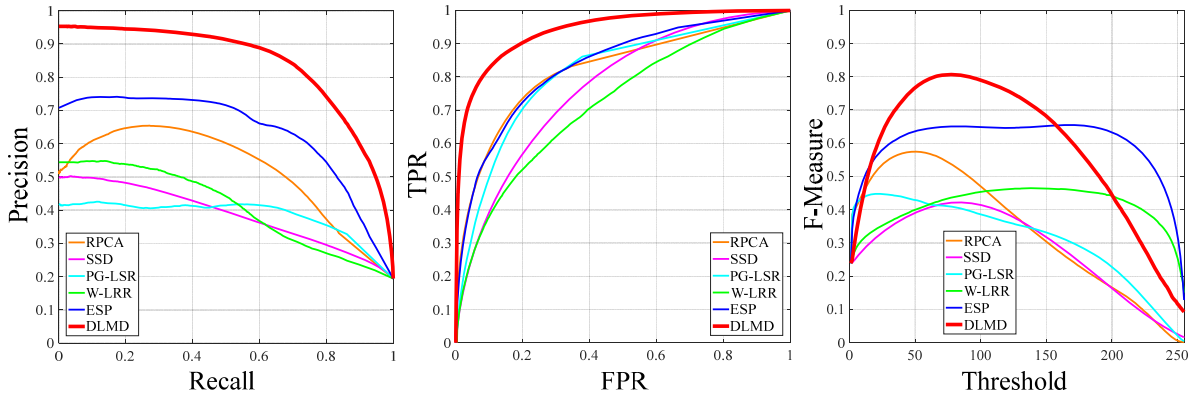


FIGURE 10. Quantitative comparison results with P-R curves, ROC curves and F-measure curves.

2) Convergence Analysis

We evaluate the convergence of the proposed DLMD to empirically show the convergence through experiments in different iterations, which is calculated via the relative error $\|D-L-S\|_F/\|D\|_F$. As shown in Fig. 8, the error converges very fast, usually within 20 iterations.

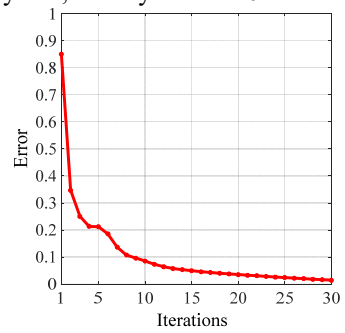


FIGURE 8. Convergence Curve of DLMD.

3) ROBUSTNESS TO NOISE

We evaluate the robustness of the proposed DLMD by corrupting original surface images with additive Gaussian noise in different SNR, including 24dB, 20dB, 16dB and 12dB. As shown in Table 6, when SNR decreases gradually, the AUC and MAE can remain a relative high level, especially when SNR = 16dB, AUC still remain around 0.8. In general, the proposed DLMD method is considered as robust to noise.

Table 6. Experimental results of DLMD with different noise level.

Index \ SNR(dB)	24	20	16	12
AUC	0.8298	0.8123	0.7759	0.7183
MAE	0.1610	0.1731	0.1939	0.2220

4) SEGMENTATION RESULTS COMPARISON

The proposed DLMD is compared with five representative segmentation methods quantitatively and qualitatively, including RPCA [27], SSD [32], PG-LSR [33], W-LRR [34], and ESP [35].

Table 7. Comparison of AUC and MAE of the proposed DLMD with other methods

Index \ Method	RPCA [27]	SSD [32]	PG-LSR [33]	W-LRR [34]	ESP [35]	**DLMD**
AUC	0.7636	0.7144	0.7133	0.6636	0.7500	**0.8453**
MAE	0.1860	0.2500	0.2010	0.2598	0.1937	**0.1593**

a: Qualitative Comparison

The qualitative comparison results between the proposed DLMD and other methods are shown in Fig. 9. It's shown that most of methods can handle simple defect images with relatively homogeneous background (i.e., column 5, and 10). For some complex defect images that containing multiple objects (i.e., column 6, 11 and 12), or having visually indistinguishable background (i.e., column 3, and 4), some parts of background being falsely classified as the defect. By contrast, the proposed DLMD separates the defect objects from the image background successfully and locates defects precisely, which has achieved the goal of "highlight the foreground and suppressing the background".

b: Quantitative Comparison

The six methods are evaluated by P-R curves, ROC curves, AUC values, F-measure curves and MAE values are illustrated in Fig. 10 and Table 7, respectively. They show that the proposed DLMD significantly outperforms the other five methods. Especially, Precision can remain above 90% within a large threshold range, which reflects a better segmentation performance. Most of AUC is higher than 70%, and DLMD achieves 84.53%, which is competitive with 9.53% improvement to 75.00% achieved by ESP. MAE of DLMD is typically the lowest among all the methods. Compared with ESP, it's increased by 9.53% and 3.44% in AUC and MAE, respectively. These experimental results illustrate the proposed DLMD is effective for segmenting a variety of defects from surface defect image, even if types and number of defects are unknown and exhibit diverse visual features of shapes, scales, directions and locations. Besides, double low-rank constrain of DLMD contributes to the good segmentation performance.

V. CONCLUSION

In this paper, we develop the JCS method including CASDDL and DLMD models to perform surface defects detection for steel sheet. Based on the anomaly characteristics of defect in the surface defect image of steel sheet, we propose a CASDDL method to learn a discriminative dictionary that consists of several class-specific sub-dictionaries associated with corresponding classes and a shared sub-dictionary shared by all the classes, in which class-specific sub-dictionaries are responsible for exploiting class-specific information, and the shared sub-dictionary is used for capturing and separating the common information. By introducing low-rank, mutual incoherence and Fisher-like discriminative constraints, it can effectively reduce redundancy in training samples. Moreover, we formulate a double low-rank decomposition model to obtain high-quality defect foreground image directly, which provides a robust way to segment the surface defect. Experimental results verify the effectiveness and robustness of JCS for detecting surface defects of steel sheet.

APPENDIX

Computing $\nabla_{X_i^{class}}\left[\left(\left\|X_i^{class}-U_i\right\|_F^2+\sum_{i=1}^{c}\|U_i-U\|_F^2\right)\right]$ in Eq. (12)

Denoting $E_i^j=\begin{pmatrix}1 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 1\end{pmatrix}_{n_i\times n_j}$, $I_{n_i\times n_i}=\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{pmatrix}_{n_i\times n_i}$, $O_i=I_{n_i\times n_i}-\frac{E_i^i}{n_i}$, then, we have

$$\begin{aligned}\left\|X_i^{class}-U_i\right\|_F^2&=\left\|X_i^{class}\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{pmatrix}_{n_i\times n_i}\right.\\&\quad\left.-X_i^{class}\begin{pmatrix}\frac{1}{n_i} & \cdots & \frac{1}{n_i}\\ \vdots & \ddots & \vdots\\ \frac{1}{n_i} & \cdots & \frac{1}{n_i}\end{pmatrix}_{n_i\times n_i}\right\|_F^2\\&=\left\|X_i^{class}I_{n_i\times n_i}-X_i^{class}\frac{E_i^i}{n_i}\right\|_F^2\\&=\left\|X_i^{class}O_i\right\|_F^2\end{aligned}$$

Denoting $Q_i^j=\frac{E_i^j}{N}$, $P_i=\frac{E_i^i}{n_i}-\frac{E_i^i}{N}=\frac{E_i^i}{n_i}-Q_i^i$, $R=\sum_{\substack{j=1\\j\neq i}}^{c}X_j^{class}Q_j^i$, then, we have

$$\begin{aligned}\sum_{i=1}^{c}\|U_i-U\|_F^2&=\left\|X_i^{class}\frac{E_i^i}{n_i}\right.\\&\quad\left.-\left(X_i^{class}\frac{E_i^i}{N}+\sum_{\substack{j=1\\j\neq i}}^{c}X_j^{class}\frac{E_j^i}{N}\right)\right\|_F^2\\&=\left\|X_i^{class}\left(\frac{E_i^i}{n_i}-Q_i^i\right)-\sum_{\substack{j=1\\j\neq i}}^{c}X_j^{class}Q_j^i\right\|_F^2\\&=\left\|X_i^{class}P_i-\sum_{\substack{j=1\\j\neq i}}^{c}X_j^{class}Q_j^i\right\|_F^2\end{aligned}$$

For the i-th class, we have $\left\|X_i^{class}P_i-\sum_{\substack{j=1\\j\neq i}}^{c}X_j^{class}Q_j^i\right\|_F^2$, for the non i-th class, we have $\sum_{\substack{j=1\\j\neq i}}^{c}\left\|X_j^{class}P_j-\sum_{\substack{l=1\\l\neq j}}^{c}X_l^{class}Q_l^j\right\|_F^2$.

Choosing the j-th part of the i-th class, we have $\left\|X_j^{class}P_j-\sum_{\substack{l=1\\l\neq j}}^{c}X_l^{class}Q_l^j\right\|_F^2=\left\|X_j^{class}\frac{E_j^j}{n_j}-\left(X_j^{class}Q_j^j+\sum_{\substack{l=1\\l\neq j}}^{c}X_l^{class}Q_l^j\right)\right\|_F^2$

Separating X_i^{class}, we have $\left\|X_j^{class}\frac{E_j^j}{n_j}-\left(X_i^{class}Q_i^j+\sum_{\substack{l=1\\l\neq i}}^{c}X_l^{class}Q_l^j\right)\right\|_F^2=\left\|X_j^{class}\frac{E_j^j}{n_j}-\sum_{\substack{l=1\\l\neq i}}^{c}X_l^{class}Q_l^j-X_i^{class}Q_i^j\right\|_F^2=\left\|T_j-X_i^{class}Q_i^j\right\|_F^2$

Expanding the anyone non i-th class, $\sum_{\substack{j=1\\j\neq i}}^{c}\left\|T_j-X_i^{class}Q_i^j\right\|_F^2$, then, we have

$$\begin{aligned}&\left\|X_i^{class}-U_i\right\|_F^2+\sum_{i=1}^{c}\|U_i-U\|_F^2\\&\quad=\left\|X_i^{class}O_i\right\|_F^2-\left\|X_i^{class}P_i-R\right\|_F^2\\&\qquad-\sum_{\substack{j=1\\j\neq i}}^{c}\left\|T_j-X_i^{class}Q_i^j\right\|_F^2\end{aligned}$$

Calculating the first derivative of $R\left(X_i^{class}\right)$ with respect to X_i^{class}, we have

$$\frac{\partial\left\|X_i^{class}O_i\right\|_F^2}{\partial X_i^{Class}}=2X_i^{class}O_iO_i^T$$

$$\frac{\partial\left\|X_i^{class}P_i-R\right\|_F^2}{\partial X_i^{Class}}=2\left(X_i^{class}P_iP_i^T-RP_i^T\right)$$

$$\frac{\partial\sum_{\substack{j=1\\j\neq i}}^{c}\left\|T_j-X_i^{class}Q_i^j\right\|_F^2}{\partial X_i^{class}}=2\sum_{\substack{j=1\\j\neq i}}^{c}\left[X_i^{class}Q_i^j\left(Q_i^j\right)^T-T_j\left(Q_i^j\right)^T\right]$$

ACKNOWLEDGMENTS

The authors would like to thank editors and anonymous reviewers for their constructive suggestions to improve the manuscript.

REFERENCES

[1] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C. M. Oddo, and P. Dario. "Visual-based defect detection and classification approaches for industrial applications-a survey," *SENSORS-BASEL*, vol. 20, no. 5, p. 1459, 2020, doi: 10.3390/s20051459.

[2] T. Ehret, A. Davy, J. M. Morel, and M. Delbracio. "Image anomalies: A review and synthesis of detection methods," *J MATH IMAGING VIS*, vol. 61, no. 5, pp. 710-743, 2019, doi: 10.1007/s10851-019-00885-0.

[3] Q. W. Luo, X. X. Fang, L. Liu, C. H. Yang, and Y. C. Sun. "Automated visual defect detection for flat steel surface: a survey," *IEEE TRANS INSTRUM MEAS*, vol. 69, no. 3, pp. 626-644, 2020, doi: 10.1109/TIM.2019.2963555.

[4] J. H. Liu, C. Y. Wang, H. Su, B. Du, and D. C. Tao. "Multistage GAN for fabric defect detection," *IEEE TRANS IMAGE PROCESS*, vol. 29, pp. 3388-3400, 2020, doi: 10.1109/TIP.2019.2959741.

[5] J. B. Zhang, H. Su, W. Zou, X. Y. Gong, Z. T. Zhang, and F. Shen. "CADN: a weakly supervised learning-based category-aware object detection network for surface defect detection," *PATTERN RECOGN*, vol. 109, p. 107571, 2021, doi: 10.1016/j.patcog.2020.107571.

[6] Y. P. Gao, L. Gao, X. Y. Li, and X. G. Yan. "A semi-supervised convolutional neural network-based method for steel surface defect recognition," *ROBOT COMPUT INTEGR MANUF*, vol. 61, p. 101825, 2020, doi: 10.1016/j.rcim.2019.101825.

[7] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj. "Segmentation-based deep-learning approach for surface-defect detection," *J INTELL MANUF*, vol. 31, pp. 759-776, 2020, doi: 10.1007/s10845-019-01476-x.

[8] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. C. Yan. "Sparse representation for computer vision and pattern recognition," *P IEEE*, vol. 98, no. 6, pp. 1031-1044, 2015, doi: 10.1109/JPROC.2010.2044470.

[9] J. C. Yao，H. M. Yu, and R. Hu. "A new sparse representation-based object segmentation framework," *VISUAL COMPUT*, vol. 33, no. 2, pp. 179-192, 2017, doi: 10.1007/s00371-015-1171-2.

[10] J. Z. Wang, Q. Y. Li, J. R. Gan, H. M. Yu, and X. Yang. "Surface defect detection via entity sparsity pursuit with intrinsic priors," *IEEE TRANS IND INFORM*, vol. 16, no. 1, pp. 141-150, 2019, doi: 10.1109/TII.2019. 2917522.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. "Robust face recognition via sparse representation," *IEEE TRANS PATTERN ANAL MACH INTELL*, vol. 31, no. 2, pp. 210-227, 2008, doi: 10.1109/TPAMI.2008.79.

[12] M. Aharon, M. Elad, and A. K. Bruckstein. "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE TRANS SIGNAL PROCES*, vol. 54, no. 11, pp. 4311-4322, 2006, doi: 10.1109/TSP.2006.881199.

[13] I. Ramirez, P. Sprechmann, and G. Sapiro. "Classification and clustering via dictionary learning with structured incoherence and shared features," *IEEE CVPR*, vol. 23, no. 3, pp. 3501-3508, 2010, doi: 10.1109/CVPR.2010.5539964.

[14] J. Ling, Z. Z. Chen, and F. Wu. "Class-oriented discriminative dictionary learning for image classification," *IEEE TRANS CIRC SYST VID*, vol. 30, no. 7, pp. 2155-2166, 2020, doi: 10.1109/TCSVT.2019.2918852.

[15] Z. Z. Fan, L. R. Shi, Q. Liu, Z. M. Li, and Z. Zhang. "Discriminative Fisher embedding dictionary transfer learning for object recognition," *IEEE T NEUR NET LEAR*, vol. 99, pp. 1-15, 2021, doi: 10.1109/TNNLS. 2016.2545112.

[16] C. J. Zhang, C. Liang, L. Li, J. Liu, Q. M. Huang, and Q. Tian. "Fine-grained image classification via low-rank sparse coding with general and class-specific codebooks," *IEEE T NEUR NET LEAR*, vol. 28, no. 7, pp. 1550-1559, 2017, doi: 10.1109/TNNLS. 2016.2545112.

[17] W. H. Deng, J. N. Hu, and J. Guo. "Face recognition via collaborative representation: its discriminant nature and superposed representation," *IEEE TRANS PATTERN ANAL MACH INTELL*, vol. 40, no. 1, pp. 2513-2521, 2018, doi: 10.1109/TPAMI.2017.2757923.

[18] S. H. Gao, I. W. H. Tsang, and Y. Ma. "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE TRANS IMAGE PROCESS*, vol. 23, no. 2, pp. 623-634, 2014, doi: 10.1109/TIP.2013.2290593.

[19] D. H. Wang, and S. Kong. "A classification-oriented dictionary learning model: explicitly learning the particularity and commonality across categories," *PATTERN RECOGN*, vol. 47, no. 2, pp. 885-898, 2014, doi: 10.1016/j.patcog. 2013.08.004.

[20] G. J. Lin, M. Yang, J. Yang, L. L. Shen, and W. C. Xie. "Robust, discriminative and comprehensive dictionary learning for face recognition," *PATTERN RECOGN*, vol. 81, pp. 341-356, 2018, doi: 10.1016/j.patcog.2018.03.021.

[21] X. D. Jiang, and J. Lai. "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE TRANS PATTERN ANAL MACH INTELL*, vol. 37, no. 5, pp. 1067-1079, 2015, doi: 10.1109/TPAMI.2014.2359453.

[22] Y. Rong, S. W. Xiong, and Y. S. Gao. "Low-rank double dictionary learning from corrupted data for robust image classification," *PATTERN RECOGN*, vol. 72, pp. 419-432, 2017, doi: 10.1016/j.patcog.2017.06.038.

[23] Z. D. Wen, B. Hou, and L. C. Jiao. "Discriminative dictionary learning with two-level low rank and group sparse decomposition for image classification," *IEEE TRANS CYBERNETICS*, vol. 47, no. 1, pp. 3758-3771, 2017, doi: 10.1109/TCYB.2016.2581861.

[24] T. H. Vu, and V. Monga. "Fast low-rank shared dictionary learning for image classification," *IEEE TRANS IMAGE PROCESS*, vol. 26, no. 11, pp. 5160-5175, 2017, doi: 10.1109/TIP.2017.2729885.

[25] H. S. Du, L. G. Ma, G. D. Li, and S. Wang. "Low-rank graph preserving discriminative dictionary learning for image recognition," *KNOWL-BASED SYST*, vol. 187, p. 104823, 2020, doi: 10.1016/j.knosys.2019.06.031.

[26] J. Chen, H. Mao, Z. Wang, and X. P. Zhang. "Low-rank representation with adaptive dictionary learning for subspace clustering". *KNOWL-BASED SYST*, vol. 223, pp. 107053, 2021, doi: 10.1016/j.knosys.2021.107053.

[27] E. J. Candès, X. D. Li, Y. Ma, and J. Wright. "Robust principal component analysis," *J ACM*, vol. 58, no. 3, p. 11, 2011, doi: 10.1145/1970392.1970395.

[28] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E. H. Zahzah. "Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset," *COMPUT SCI REV*, vol. 23, pp. 1-71, 2017, doi: 10.1016/j.cosrev. 2016.11.001.

[29] H. W. Peng, B. Li, H. B. Ling, W. M. Hu, W. H. Xiong, and S. J. Maybank. "Salient object detection via structured matrix decomposition," *IEEE TRANS PATTERN ANAL MACH INTELL*, vol. 39, no. 4, pp. 818-832, 2017, doi: 10.1109/TPAMI. 2016.2562626.

[30] Y. G. Cen, R. Z. Zhao, L. H. Cen, L. H. Cui, Z. J. Miao, and Z. Wei. "Defect inspection for TFT-LCD images based on the low-rank matrix reconstruction," *NEUROCOMPUTING*, vol. 149, pp. 1206-1215, 2015, doi: 10.1016/j.neucom. 2014.09.007.

[31] C. L. Li, G. S. Gao, Z. F. Liu, D. Huang, and J. T. Xi. "Defect detection for patterned fabric images based on GHOG and low-rank decomposition," *IEEE ACCESS*, vol. 7, pp. 83962-83973, 2019, doi: 10.1109/ACCESS.2019.2925196.

[32] H. Yan, K. Paynabar, and J. J. Shi. "Anomaly detection in images with smooth background via smooth-sparse decomposition," *TECHNOMETRICS*, vol. 59, no. 1, pp. 102-114, 2017, doi: 10.1080/00401706.2015.1102764.

[33] J. J. Cao, J. Zhang, Z. J. Wen, N. N. Wang, and X. P. Liu. "Fabric defect inspection using prior knowledge guided least squares regression," *MULTIMED TOOLS APPL*, vol. 76, no. 3, pp. 4141-4157, 2017, doi: 10.1007/s11042-015-3041-3.

[34] Q. Z. Huang Peng, H. Zhang, X. R. Zeng, and W. Huang. "Automatic visual defect detection using texture prior and low-rank representation," *IEEE ACCESS*, vol. 6, pp. 37965-37976, 2018, doi: 10.1109/ACCESS.2018.2852663.

[35] J. Z. Wang, Q. Y. Li, J. R. Gan, H. M. Yu, and X. Yang. "Surface defect detection via entity sparsity pursuit with intrinsic priors," *IEEE TRANS IND INFORM*, vol. 16, no. 1, pp. 141-150, 2019, doi: 10.1109/TII. 2019.2917522.

[36] S. Y. Zhou, Y. P. Chen, D. L. Zhang, J. M. Xie, and Y. F. Zhou. "Learning a class-specific and shared dictionary for classifying surface defects of steel sheet," *ISIJ INT*, vol. 57, no. 1, pp. 123-130, 2017, doi: 10.2355/isijinternational.ISIJINT-2016-478.

[37] S. Y. Zhou, S. Q. Wu, K. T. Cui, and H. G. Liu. "Double low-rank based matrix decomposition for surface defect segmentation of steel sheet," *ISIJ INT*, vol. 61, no. 7, pp. 2111-2121, 2021, doi: 10.2355/isijinternational.ISIJINT-2021-024.

[38] R. Borwankar, and R. Ludwig. "An optical surface inspection and automatic classification technique using the rotated wavelet transform," *IEEE TRANS INSTRUM MEAS*, vol. 67, no. 3, pp. 690-697, 2018, doi: 10.1109/TIM.2017.2783098.

[39] Q. W. Luo, Y. C. Sun, P. C. Li, O. Simpson, L. Tian, and Y. G. He. "Generalized completed local binary patterns for time-efficient steel surface defect classification," *IEEE TRANS INSTRUM MEAS*, vol. 68, no. 3, pp. 667-679, 2019, doi: 10.1109/TIM.2018. 2852918.

[40] M. W. Ashour, F. Khalid, A. A. Halin, L. N. Abdullah, and S. H. Darwish. "Surface defects classification of hot-rolled steel strips using multi-directional shearlet features," *ARAB J SCI ENG*, vol. 44, no. 4, pp. 2925-2932, 2019, doi: 10.1007/s13369-018-3329-5.

[41] S. Ravishankar, J. C. Ye, and J. A. Fessler. "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *P IEEE*, vol. 108, no. 1, pp. 86-109, 2019, doi: 10.1109/JPROC.2019.2936204.

[42] T. Bouwmans, S. Javed，H. Y. Zhang, Z. C. Lin, and R. Otazo. "On the applications of robust PCA in image and video processing," *P IEEE*, vol. 106, no. 8, pp. 1427-1457, 2018, doi: 10.1109/JPROC.2018.2853589.

[43] S. Q. Ma, and N. S. Aybat. "Efficient optimization algorithms for robust principal component analysis and its variants," *P IEEE*, vol. 106, no. 8, pp. 1411-1426, 2018, doi: 10.1109/JPROC.2018.2846606.

[44] S. H. Gu, L. Zhang, W. M. Zuo, and X. C. Feng. "Weighted nuclear norm minimization with application to image denoising," *IEEE CVPR*, pp. 2862-2869, 2014, doi: 10.1109/CVPR.2014.366.

[45] J. M. Bioucas-Dias, and M. A. T. Figueiredo. "A new TwIST: two-step iterative shrinkage thresholding algorithms for image restoration," *IEEE TRANS IMAGE PROCESS*, vol. 16, no. 12, pp. 2992-3004, 2007, doi: 10.1109/TIP.2007.909319.

[46] M. Nejati, S. Samavi, S. M. R. Soroushmehr, and K. Najarian. "Coherence regularized dictionary learning," *IEEE ICASSP*, 2016, pp. 4717-4721, doi: 10.1109/ICASSP.2016.7472572.

[47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE TRANS PATTERN ANAL MACH INTELL*, vol. 34, no. 11, pp. 2274-2282, 2012, doi: 10.1109/TPAMI.2012.120.