# Robust Transfer Subspace Learning: A Novel Data-Driven Scheme for Fault Diagnosis

1st Fuchao Yu
*School of Mechatronic Engineering and Automation*
*Shanghai University*
Shanghai, China
yufuchao19981025@163.com

2nd Xianchao Xiu
*School of Mechatronic Engineering and Automation*
*Shanghai University*
Shanghai, China
xcxiu@shu.edu.cn

***Abstract*—Transfer Subspace Learning (TSL) offers a promising solution for cross-conditional Fault Diagnosis (FD). However, when the data is corrupted by noise or the labels are inaccurate, the FD performance may severely degrade. In this paper, we propose a robust TSL-based scheme called RTSL for FD. By introducing a matrix with appropriate regularization terms, the proposed RTSL is robust to both Gaussian noise and non-Gaussian noise. By adaptively learning labels from inputs, the proposed RTSL is robust to inaccurate labels. In algorithms, we develop an efficient optimization algorithm based on the Alternating Direction Method of Multipliers (ADMM). Finally, we verify the efficiency by numerical studies on the Case Western Reserve University (CWRU) dataset. The results show that the proposed method is robust for FD.**

***Index Terms*—Fault diagnosis, domain adaptation, robust transfer subspace learning, optimization algorithm.**

## I. Introduction

Fault Diagnosis (FD) plays an very important role in modern industrial systems because the early detection of faults can prevent safety accidents and save valuable cost. Generally speaking, FD approaches can be divided into model-based and data-driven categories. Model-based FD approaches heavily rely on the prior knowledge, and thus the FD performance is determined by the accuracy of the system model. In contrast, data-driven FD approaches only require data collected by sensors, and they are more suitable for complex industrial systems, thus attracting widespread attention [1]–[3]. In the past decades, both traditional machine learning approaches [4], [5] and deep learning approaches [6], [7] have been successfully applied in the field of FD. However, when there exist few or no labels, it will pose a great challenge to most data-driven FD approaches.

In practical industrial systems, the collected data usually comes from different working conditions, which means the distribution of samples from virtual domains differs. The existence of similarities in different working conditions makes domain adaptation possible. Transfer Learning (TL) uses the learned knowledge on source domains to achieve good performance on different but related target domains [8]. According to [9], [10], TL approaches can be categorized into four types, including instance-based TL [11], [12], model-based TL [13]–[15], feature-based TL [16]–[18], and relational-based TL [19]. Feature-based TL, including manifold learning [20], feature extraction [21], and Transfer Subspace Learning (TSL) [22], [23], plays an significant role in TL. Among them, TSL exploits a common subspace where the disparity of the source and target domains is minimal. Pan *et al.* [24] transformed samples into a Reproducing Kernel Hilbert Space (RKHS) to reduce the difference in the marginal distribution. Long *et al.* [25] proposed an efficient approach, called Joint Distribution Adaptation (JDA). Xu *et al.* [23] introduced low-rank and sparse representation into transformation matrix to seek the optimum subspace. Wang *et al.* [26] balanced the importance of the marginal and conditional distributions in the subspace by introducing dynamic factors. Lu *et al.* [27] introduced between- and within-class graphs to preserve the geometric structure in the common subspace. Shi *et al.* [28] combined sparse representation, manifold learning, and low-rank representation to learn a transformation matrix.

Although TSL has been successfully applied in many fields, to our surprise, it has not been considered in the field of FD. In fact, there exist the following three challenges. The first one is that the data collected under different working conditions may be disorganized. In particular, some features are harmful to transfer. The second one is that the noise in data is the leading cause of negative transfer. Most of the existing approaches don't consider the effect of different types of noise. The third one is that some recent work has leveraged $\epsilon$-dragging technique to provide more degrees of freedom for the transformation matrix [23], [27], however, their flexibility needs to be further improved. Therefore, in this paper, we propose a novel Robust Transfer Subspace Learning (RTSL) approach to address all above challenges. Fig.1 shows the framework of the proposed RTSL.

Specially, the contribution of the current work can be summarized as following three parts.

- We apply neural networks for extracting features in order to overcome the disorganization of the data.
- We introduce an auxiliary matrix to filter out Gaussian noise in order to enhance the robustness.
- We learn the labels from the inputs in order to learn a flexible transformation matrix.

This work was supported in part by the National Natural Science Foundation of China under Grant 12001019.

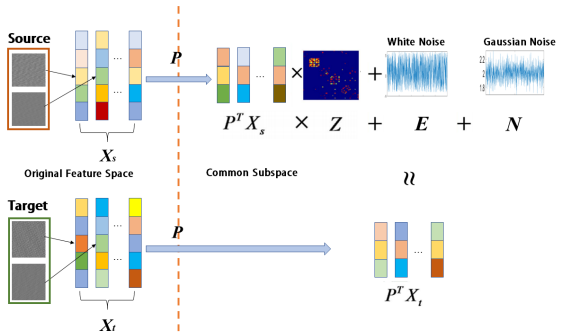


Fig. 1. The framework of the proposed approach.

## II. Robust Transfer Subspace Learning

### A. Problem Formulation

Given the source domain data $X_s$ and the target domain data $X_t$. TSL aims to find a transformation matrix $P$ and a reconstruction matrix $Z$ such that the target domain data $X_t$ can be linearly represented by the source domain data $X_s$ in a common subspace, i.e.,

$$P^\top X_t = P^\top X_s Z. \tag{1}$$

Let $Y$ be the label matrix. Then, TSL can be described by the following minimization problem

$$\begin{aligned} \min_{Z,P} \ & \frac{1}{2}\phi_1(P,Y,X_s) + \|Z\|_* + \alpha\|Z\|_1 \\ \text{s.t.} \ & P^\top X_t = P^\top X_s Z, \end{aligned} \tag{2}$$

where $\phi_1(P,Y,X_s) = \|P^\top X_s - Y\|_{\mathrm{F}}^2$ is a discriminant subspace learning function, $\|Z\|_*$ is the nuclear norm, defined as the sum of its singular values, which can capture the global structure, $\|Z\|_1$ is the $\ell_1$-norm of $Z$, which can capture the local structure, and $\alpha > 0$ is a trade-off parameter to balance $\|Z\|_*$ and $\|Z\|_1$.

Recently, Xu *et al.* [23] introduced a matrix $E$ with $\ell_1$-norm regularization to filter out random noise, and enlarged the margins between different classes. The optimization model is characterized by the form of

$$\begin{aligned} \min_{Z,E,P,M} \ & \frac{1}{2}\phi_2(P,Y,X_s) + \|Z\|_* + \alpha\|Z\|_1 + \beta\|E\|_1 \\ \text{s.t.} \ & P^\top X_t = P^\top X_s Z + E. \end{aligned} \tag{3}$$

Note that $\phi_2(P,Y,X_s) = \|P^\top X_s - (Y + B \circ M)\|_{\mathrm{F}}^2$, where $B$ is called a luxury matrix, defined as

$$B_{ij} = \begin{cases} +1, & \text{if } Y_{ij} = 1, \\ -1, & \text{if } Y_{ij} = 0, \end{cases} \tag{4}$$

$M$ is a label relaxation matrix with each element being non-negative, and $\circ$ is the Hadamard product.

To improve the robustness to noise and labels, we proposes a novel scheme called RTSL, which can be formulated as follows

$$\begin{aligned} \min_{Z,E,N,R,P} \ & \frac{1}{2}\phi_3(P,R,X_s) + \|Z\|_* + \alpha\|Z\|_1 + \beta\|E\|_1 \\ & + \gamma\|N\|_{\mathrm{F}}^2 \\ \text{s.t.} \ & P^\top X_t = P^\top X_s Z + E + N, \\ & R_{iy_i} - \max_{j \neq y_i} R_{ij} \geq 1, \ i = 1, \ldots, n, \end{aligned} \tag{5}$$

where $R$ is the learned label matrix, and $y_i$ is the index of the true class of the $i$th sample in the source domain.

Obviously, compared with [23], [27], [28], the main differences lie in two aspects.

- It embeds a matrix $N$ with F-norm regularization to represent the Gaussian noise. Thus, RTSL can detect both random noise and Gaussian noise.
- It considers $\phi_3(P,R,X_s) = \|P^\top X_s - R\|_{\mathrm{F}}^2$, where $R$ provides more freedom than $B \circ M$ [29]. Thus, RTSL improves the discriminative ability and flexibility.

### B. Optimization Algorithm

By introducing two variables $Z_1, Z_2$, problem (5) can be rewritten as

$$\begin{aligned} \min_{Z,Z_1,Z_2,E,N,R,P} \ & \frac{1}{2}\|P^\top X_s - R\|_{\mathrm{F}}^2 + \|Z_1\|_* + \alpha\|Z_2\|_1 \\ & + \beta\|E\|_1 + \gamma\|N\|_{\mathrm{F}}^2 \\ \text{s.t.} \ & P^\top X_t = P^\top X_s Z + E + N, \\ & R_{iy_i} - \max_{j \neq y_i} R_{ij} \geq 1, \ i = 1, ..., n, \\ & Z_1 = Z, \ Z_2 = Z. \end{aligned} \tag{6}$$

Next, we will show how to solve problem (6) by Alternating Direction Method of Multipliers (ADMM). The augmented Lagrangian function of problem (6) is

$$\begin{aligned} & \mathcal{L}_\mu(Z, Z_1, Z_2, E, N, R, P, \Lambda_1, \Lambda_2, \Lambda_3) \\ & = \frac{1}{2}\|P^\top X_s - R\|_{\mathrm{F}}^2 + \|Z_1\|_* + \alpha\|Z_2\|_1 \\ & \quad + \beta\|E\|_1 + \gamma\|N\|_{\mathrm{F}}^2 \\ & \quad + \langle \Lambda_1, P^\top X_t - P^\top X_s Z - E - N \rangle \\ & \quad + \frac{\mu}{2}\|P^\top X_t - P^\top X_s Z - E - N\|_{\mathrm{F}}^2 \\ & \quad + \langle \Lambda_2, Z - Z_1 \rangle + \frac{\mu}{2}\|Z - Z_1\|_{\mathrm{F}}^2 \\ & \quad + \langle \Lambda_3, Z - Z_2 \rangle + \frac{\mu}{2}\|Z - Z_2\|_{\mathrm{F}}^2, \end{aligned} \tag{7}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are the Lagrangian multipliers.

*1) Update $Z^{k+1}$:* When other variables are fixed, $Z^{k+1}$ can be updated by minimizing

$$\begin{aligned} \min_{Z} \ & \frac{\mu}{2}\|P^{k\top} X_t - P^{k\top} X_s Z - E^k - N^k + \Lambda_1^k/\mu\|_{\mathrm{F}}^2 \\ & + \frac{\mu}{2}\|Z - Z_1^k + \Lambda_2^k/\mu\|_{\mathrm{F}}^2 + \frac{\mu}{2}\|Z - Z_2^k + \Lambda_3^k/\mu\|_{\mathrm{F}}^2, \end{aligned} \tag{8}$$

which admits the following closed-form solution

$$Z^{k+1} = \left(\mu X_s^\top P^k P^{k\top} X_s + 2\mu I\right)^{-1} (T_1^k - X_s^\top P^k T_2^k), \tag{9}$$

**Algorithm 1** The solution of problem (18)

**Input:** $P_i$, $y_i$.
**Initialize:** $k_j = Q_{ij} + 1 - Q_{iy_i}$, $\forall i = 1, 2, 3, ..., n$, $\eta = 0$, $iter = 0$.
**for** $j = 1$ to $c (j \neq y_i)$ **do**
  **if** $f'(k_j) > 0$ **then**
    Calculate $\eta = \eta + k_j$, $iter = iter + 1$
Calculate $\eta = \eta / (iter + 1)$
Update $R_i = [R_{i1}, R_{i2}, ..., R_{ij}]$ by (21)
**Output:** R.

where $T_1^k = Z_1^k - \Lambda_2^k/\mu + Z_2^k - \Lambda_3^k/\mu$ and $T_2^k = P^{k\top} X_t - E^k - N^k - \Lambda_1^k/\mu$.

*2) Update $Z_1^{k+1}$:* The minimization problem in terms of $Z_1^{k+1}$ can be reformulated as

$$\min_{Z_1} \ \|Z_1\|_* + \frac{\mu}{2} \|Z^{k+1} - Z_1 + \Lambda_2^k/\mu\|_{\mathrm{F}}^2. \tag{10}$$

According to [30], it has the following closed-form solution

$$Z_1^{k+1} = \mathcal{D}_{1/\mu}(Z^{k+1} + \Lambda_2^k/\mu), \tag{11}$$

where $\mathcal{D}_\tau(T)$ is the singular value thresholding, defined as

$$\mathcal{D}_\tau(T) = U \mathcal{S}_\tau(\Sigma) V^\top, \tag{12}$$

of which $\mathcal{S}_\tau(\Sigma_{ii}) = \mathrm{sign} \cdot (\Sigma_{ii}) \max(0, |\Sigma_{ii} - \tau|)$ is the soft-thresholding operator [31].

*3) Update $Z_2^{k+1}$:* When other variables are fixed, $Z_2^{k+1}$ can be presented as the form of

$$\min_{Z_2} \ \alpha \|Z_2\|_1 + \frac{\mu}{2} \|Z^{k+1} - Z_2 + \Lambda_3^k/\mu\|_{\mathrm{F}}^2, \tag{13}$$

which has the following closed-form solution

$$Z_2^{k+1} = \mathcal{S}_{\alpha/\mu}(Z^{k+1} + \Lambda_3^k/\mu). \tag{14}$$

*4) Update $E^{k+1}$:* Similarly, $E^{k+1}$ can be solved by

$$E^{k+1} = \mathcal{S}_{\beta/\mu}(P^{k\top} X_t - P^{k\top} X_s Z^{k+1} - N^k + \Lambda_1^k/\mu). \tag{15}$$

*5) Update $N^{k+1}$:* After simple linear transformations, $N^{k+1}$ can be achieved by

$$N^{k+1} = \frac{\mu}{2\gamma + \mu}(P^{k\top} X_t - P^{k\top} X_s Z^{k+1} - E^{k+1} + \Lambda_1^k/\mu). \tag{16}$$

*6) Update $R^{k+1}$:* When other variables are fixed, $R^{k+1}$ can be optimized by

$$\begin{aligned} \min_{R} \ & \frac{1}{2} \|P^{k\top} X_s - R\|_{\mathrm{F}}^2 \\ \text{s.t.} \ & R_{iy_i} - \max_{j \neq y_i} R_{ij} \geq 1, \ i = 1, ..., n. \end{aligned} \tag{17}$$

Denote $Q = P^{k\top} X_s$. Then, problem (17) can be handled as

$$\begin{aligned} \min_{R_i} \ & \frac{1}{2} \|Q_i - R_i\|_2^2 \\ \text{s.t.} \ & R_{iy_i} - \max_{j \neq y_i} R_{ij} \geq 1, \end{aligned} \tag{18}$$

where $Q_i$ and $R_i$ are the $i$th row of $Q$ and $R$, respectively. Further, denote $R_{iy_i} = Q_{iy_i} + \eta$, where $\eta$ represents the parameter that can be optimized. Thus, the above objective can be transformed to $\sum_{j=1}^{c} (Q_{ij} - R_{ij})^2$, where $c$ is the class number of samples. Therefore, problem (18) can be decomposed as

$$\begin{aligned} \min_{R_{ij}} \ & \frac{1}{2} (Q_{ij} - R_{ij})_2^2 \\ \text{s.t.} \ & Q_{iy_i} + \eta - R_{ij} \geq 1, \ \forall j \neq y_i. \end{aligned} \tag{19}$$

For a new variable $d$, denote the $j$th element as

$$d_j = Q_{ij} + 1 - Q_{iy_i}, \ \forall j \neq y_i, \tag{20}$$

where $d_j \leq 0$ and $d_j > 0$ represent the marginal constraint between the $i$th sample and the $y_i$th class is violated and satisfied, respectively. Then, $R_{ij}$ in (19) can be calculated by

$$R_{ij} = \begin{cases} Q_{ij} + \eta, & \text{if } j = y_i, \\ Q_{ij} + \min(\eta - d_j, 0), & \text{otherwise.} \end{cases} \tag{21}$$

Therefore, vector $R_i$ in (18) can be replaced by parameter $\eta$, and problem (18) can be rewritten as

$$\min_{\eta} f(\eta) = \eta^2 + \sum_{j \neq y_i} (\min(\eta - k_j, 0))^2. \tag{22}$$

The optimal solution of $\eta$ is

$$\eta = \frac{\sum_{j \neq y_i} k_j \mathrm{g}(f'(k_j) > 0)}{1 + \sum_{j \neq y_i} \mathrm{g}(f'(k_j) > 0)}, \tag{23}$$

where $\mathrm{g}(\cdot)$ denotes the indicator operator. If $f'(k_j) > 0$ is satisfied, $\mathrm{g}(\cdot) = 1$, or else $\mathrm{g}(\cdot) = 0$. Algorithm 1 provides the detailed steps of getting the optimal row vector $R_i$, then it is possible for us to get the matrix $R^{k+1}$.

*7) Update $P^{k+1}$:* When other variables are determined, $P^{k+1}$ can be obtained by optimizing

$$\begin{aligned} \min_{P} \ & \frac{1}{2} \|P^\top X_s - R^{k+1}\|_{\mathrm{F}}^2 \\ & + \frac{\mu}{2} \|P^\top X_t - P^\top X_s Z^{k+1} - E^{k+1} - N^{k+1} + \Lambda_1^k/\mu\|_{\mathrm{F}}^2. \end{aligned} \tag{24}$$

**Algorithm 2** The solution of problem (6)

**Input:** Data $X_s, X_t, Y$, parameters $\alpha, \beta, \gamma$.
**Initialize:** $(Z^0, Z_1^0, Z_2^0, E^0, N^0, R^0, P^0, \Lambda_1^0, \Lambda_2^0, \Lambda_3^0)$, $\mu = 0.1$, $\rho = 1.01$, $\mu_{\max} = 10^7$, $\epsilon = 10^{-7}$.
**While** not converged **do**
Update $Z^{k+1}$ by (9)
Update $Z_1^{k+1}$ by (11)
Update $Z_2^{k+1}$ by (14)
Update $E^{k+1}$ by (15)
Update $N^{k+1}$ by (16)
Update $R^{k+1}$ by Algorithm 1
Update $P^{k+1}$ by (26)
Update $\Lambda_1^{k+1}, \Lambda_2^{k+1}, \Lambda_3^{k+1}$ by (27)
**End while**
**Output:** $(Z, P)$

TABLE I
THE DESCRIPTION OF CLASS LABELS ON THE CWRU DATASET.

| Class Label | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Fault Location | NB | IF | BF | OF | IF |
| Fault Size (mils) | 0 | 7 | 7 | 7 | 14 |
| Class Label | 6 | 7 | 8 | 9 | 10 |
| Fault Location | BF | PF | IF | BF | OF |
| Fault Size (mils) | 14 | 14 | 21 | 21 | 21 |

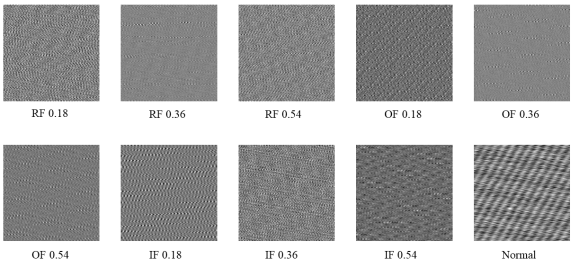


Fig. 2. The examples of tested images.

Let $T_3^{k+1} = X_s R^{k+1}$, $T_4^{k+1} = X_t - X_s Z^{k+1}$, and $T_5^{k+1} = E^{k+1} + N^{k+1} - \Lambda_1^k/\mu$. After simplification, it has the following closed-form solution

$$\begin{aligned} P^{k+1} =& (X_s X_s^\top + \mu T_4^{k+1} T_4^{(k+1)\top})^{-1} \\ & (T_3^{(k+1)\top} + \mu T_4^{k+1} T_5^{(k+1)\top}). \end{aligned} \tag{25}$$

To obtain a numerically stable solution, one can calculate

$$\begin{aligned} P^{k+1} =& (X_s X_s^\top + \mu T_4^{k+1} T_4^{(k+1)\top} + \lambda I)^{-1} \\ & (T_3^{(k+1)\top} + \mu T_4^{k+1} T_5^{(k+1)\top}), \end{aligned} \tag{26}$$

where $\lambda > 0$ is a small positive constant.

*8) Update $\Lambda_1^{k+1}, \Lambda_2^{k+1}, \Lambda_3^{k+1}$:* Once all the primal variables have been updated, the multipliers can be easily computed by

$$\begin{cases} \Lambda_1^{k+1} = \Lambda_1^k + \mu(P^{(k+1)\top} X_t - P^{(k+1)\top} X_s Z^{k+1} \\ \qquad - E^{k+1} - N^{k+1}), \\ \Lambda_2^{k+1} = \Lambda_2^k + \mu(Z^{k+1} - Z_1^{k+1}), \\ \Lambda_3^{k+1} = \Lambda_3^k + \mu(Z^{k+1} - Z_2^{k+1}). \end{cases} \tag{27}$$

In summary, the iterative framework of solving problem (6) can be given in Algorithm 2, where $\mu = \min(\rho\mu, \mu_{\max})$. Note that the algorithm is stopped by checking $\max\{\|P^\top X_t - P^\top X_s Z - E - N\|_\infty, \|Z - Z_1\|_\infty, \|Z - Z_2\|_\infty\} < \epsilon$.

## III. CASE STUDIES

In this section, we compare with some state-of-the-art approaches, including Transfer Component Analysis (TCA) [32], Joint Distribution Adaptation (JDA) [25] and Low-Rank and Sparse Representation (LRSR) [23]. Moreover, 1-Nearest Neighbor (NN) classification is chosen as the baseline classifier. We would like to point out that all the parameters $\alpha$, $\beta$, $\gamma$ are chosen by grid-search methods from the candidate set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$.

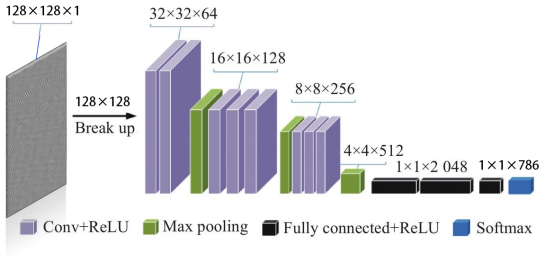


Fig. 3. The illustration of VGG used in this paper.

TABLE II
THE CLASSIFICATION ACCURACIES (%) OF COMPARED APPROACHES.

| **Transfer Task** | NN | TCA | JDA | LRSR | RTSL |
|---|---|---|---|---|---|
| SF0 → SF1 | **66.43** | 57.50 | 65.57 | 63.39 | 64.91 |
| SF0 → SF2 | 61.84 | 63.70 | 60.41 | 63.91 | **69.40** |
| SF0 → SF3 | 34.56 | 37.40 | 43.57 | 40.33 | **47.11** |
| SF1 → SF0 | 72.95 | 68.60 | 74.42 | **75.24** | 70.35 |
| SF1 → SF2 | 57.89 | 52.70 | 57.02 | 62.97 | **66.18** |
| SF1 → SF3 | 48.85 | 37.90 | 42.67 | 51.79 | **58.88** |
| SF2 → SF0 | 68.86 | 66.60 | 73.69 | 69.81 | **73.83** |
| SF2 → SF1 | 60.98 | 62.10 | 62.14 | 61.96 | **63.84** |
| SF2 → SF3 | 50.32 | 48.50 | 54.49 | 56.92 | **59.98** |
| SF3 → SF0 | 44.95 | 42.10 | 47.14 | 46.48 | **54.24** |
| SF3 → SF1 | 47.86 | 38.30 | 48.30 | 53.39 | **56.21** |
| SF3 → SF2 | 47.93 | 38.80 | 48.79 | 53.95 | **54.82** |
| Average | 55.29 | 51.18 | 56.52 | 58.35 | **61.65** |

### A. Data Preparation

*1) Dataset:* The Case Western Reserve University (CWRU) dataset is a benchmark open source dataset which has been widely used in FD. According to [33], the drive end bearing fault data is considered whose sampling frequency is 12kHz. In this sampling frequency, there are four operating speeds (four work conditions). Under each condition, there are four types, i.e., Normal Bearing (NB) and three fault types, including Internal Failure (IF), Ball Failure (BF) and External Fault (OF). For each fault type, there are three fault sizes. As seen in Table I, there are ten categories (one health state and nine fault states) for each condition.

*2) Preprocessing:* Inspired by [6], one-dimensional vibration signals are converted into two-dimensional images with 128*128 gray pixels. The example images of ten conditions are presented in Fig. 2. It can be observed that the images under different fault conditions look totally different, providing an intuitionistic way to classify them. Then a Convolution Neural Network (CNN) is used to extract features of images. Each image is quantized into 786-dimensional feature points by VGG; see Fig. 3. In order to better simulate the real situations, different numbers of images under the four working conditions are generated. After feature extraction, we obtain four TSL datasets: SF0, SF1, SF2, and SF3. Each dataset corresponds one work condition, and there are 12 TL tasks.

### B. Experimental Results

The results are listed in Table II. It can be concluded that

- The proposed RTSL can achieve the best performance in the five baseline approaches with statistical significance.

The average classification accuracy of RTSL on 12 tasks is 61.65%, and the performance improvement is 2.9% compared to the best baseline approach, i.e., LRSR.
- The proposed RTSL can construct more effective and robust representation for cross-domain FD tasks.
- The proposed RTSL can provide more freedom for the transformation matrix and the original data contains a large amount of Gaussian noise.
- The NN appraoch without transfer outperforms some methods in some tasks. The reason is that negative migration has occurred. We believe this is related to the features extracted by the trained CNN.

## IV. Conclusion

In this paper, we have proposed a novel robust TSL approach for data-driven fault diagnosis. First, for the processing of the data, we convert 1D data into 2D images and extract features by the VGG. Then, we alleviate the noise in data by introducing a sparse matrix with $\ell_1$-norm regularization and a matrix with F-norm regularization. Finally, we improve the least square regression by directly learning labels from the inputs to enhance the flexibility and discernment of the model. Extensive experiments show that the proposed approach is effective for cross-condition FD.

In the future, we are interested in extending the scheme to neural networks-based framework for complicated FD.

## References


[1] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1700–1716, 2022.

[2] Z. Gao, C. Cecati, S. X. Ding *et al.*, "A survey of fault diagnosis and fault-tolerant techniques—part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Transactions on Industrial Electronics*, 2015.

[3] X. Xiu, Z. Miao, and W. Liu, "A sparsity-aware fault diagnosis framework focusing on accurate isolation," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1356–1365, 2023.

[4] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mechanical Systems and Signal Processing*, vol. 70, pp. 87–103, 2016.

[5] Y. Li, X. Xiu, and W. Liu, "Towards efficient process monitoring using spatiotemporal pca," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, no. 2, pp. 551–555, 2023.

[6] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] F. Yu, X. Xiu, and Y. Li, "A survey on deep transfer learning and beyond," *Mathematics*, vol. 10, no. 19, p. 3619, 2022.

[11] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," vol. 227, 2007, pp. 193–200.

[12] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[14] M. Ghifary, W. B. Kleijn, and M. Zhang, "Domain adaptive neural networks for object recognition," in *Pacific Rim International Conference on Artificial Intelligence*. Springer, 2014, pp. 898–904.

[15] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *ArXiv:1412.3474*, 2014.

[16] E. V. Bonilla, K. Chai, and C. Williams, "Multi-task gaussian process prediction," *Advances in Neural Information Processing Systems*, vol. 20, 2007.

[17] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," in *Proceedings of the 15th ACM International Conference on Multimedia*, 2007, pp. 188–197.

[18] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[19] T. Isobe, X. Jia, S. Chen, J. He, Y. Shi, J. Liu, H. Lu, and S. Wang, "Multi-target domain adaptation with collaborative consistency learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8187–8196.

[20] S. Yao, Q. Kang, M. Zhou, M. J. Rawa, and A. Albeshri, "Discriminative manifold distribution alignment for domain adaptation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 2, pp. 1183–1197, 2023.

[21] J. Yoon, D. Kang, and M. Cho, "Semi-supervised domain adaptation via sample-to-sample self-distillation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2022, pp. 1978–1987.

[22] S. Ren and C. Q. Li, "Robustness of transfer learning to image degradation," *Expert Systems with Applications*, vol. 187, p. 115877, 2022.

[23] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 850–863, 2016.

[24] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[25] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2200–2207.

[26] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *Proceedings of the 26th ACM International Conference on Multimedia*, 2018, pp. 402–410.

[27] Y. Lu, W. Wang, C. Yuan, X. Li, and Z. Lai, "Manifold transfer learning via discriminant regression analysis," *IEEE Transactions on Multimedia*, vol. 23, pp. 2056–2070, 2021.

[28] K. Shi, Z. Liu, W. Lu, W. Ou, and C. Yang, "Unsupervised domain adaptation based on adaptive local manifold learning," *Computers and Electrical Engineering*, vol. 100, p. 107941, 2022.

[29] J. Jin, Z. Qin, D. Yu, T. Yang, C. P. Chen, and Y. Li, "Relaxed least square regression with $\ell_{2,1}$-norm for pattern classification," *International Journal of Wavelets, Multiresolution and Information Processing*, 2023.

[30] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[31] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[32] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[33] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–28, 2021.