

Cross-Task Mental Workload Recognition Based on EEG Tensor Representation and Transfer Learning

Kai Guan[✉], Zhimin Zhang, Tao Liu[✉], and Haijun Niu[✉], *Member, IEEE*

Abstract—The accurate evaluation of mental workload of operators in human machine systems is of great significance in ensuring the safety of operators and the correct execution of tasks. However, the effectiveness of EEG based cross-task mental workload evaluation are still unsatisfactory because of the different EEG response patterns in different tasks, which hindered its generalization in real scenario severely. To solve this problem, this paper proposed a feature construction method based on EEG tensor representation and transfer learning, which was verified in various task conditions. Specifically, four working memory load tasks with different types of information were designed firstly. The EEG signals of participants were collected synchronously during task execution. Then, the wavelet transform method was used to perform time-frequency analysis of multi-channel EEG signals, and three-way EEG tensor (time-frequency-channel) features were constructed. EEG tensor features from different tasks were transferred based on the criteria of feature distribution alignment and class-wise discrimination criteria. Finally, the support vector machine was used to construct a 3-class mental workload recognition model. Results showed that compared with the classical feature extraction methods, the proposed method can achieve higher accuracy in both within-task and cross-task mental workload evaluation (91.1% for within-task and 81.3% for cross-task). These results demonstrated that the EEG tensor representation and transfer learning method is feasible and effective for cross-task mental workload evaluation, which can provide theoretical basis and application reference for future researches.

Index Terms—Mental workload, cross-task, EEG, tensor representation, transfer learning.

I. INTRODUCTION

THE mental workload (MWL) state of operators would directly influence the task performance in human-machine systems. Excessive workload might induce the decline of the operators' decision-making ability and therefore bring threat to its safety [1], [2]. It is of great significance to evaluate the mental workload accurately in ensuring the correct execution of tasks and the operators' safety. At present, the commonly used mental workload evaluation methods include subjective measurement (e.g., NASA-TLX, SWAT etc. [3]), task performance measurement (e.g., response time, accuracy, etc.) and physiological measurement (e.g., EEG, ECG and so on) [2], [4]. Of all these methods, the physiological measurement has attracted more attention from researchers since it can assess mental workload objectively and can be applied in real-time scenario.

Among all the physiological signals, EEG is considered to be the most sensitive physiological indicator to mental workload since it can directly reflect the cognitive process of human brain. Various studies have confirmed that the linear features (e.g., Power Spectral Density (PSD) [5], [6]), nonlinear features (e.g., Sample Entropy [7]) and functional connectivity features (nodal and global properties) [8], [9], [10] of EEG signals can effectively distinguish the mental workload level of operators in specific tasks. However, when the same features were applied to evaluate mental workload in different tasks (i.e., cross-task mental workload evaluation), the evaluation results usually tend to be corrupted and the recognition accuracy usually decreases to random levels [11], [12], [13]. Many studies attributed the cross-task problem of mental workload assessment to the differences of brain information processing mechanisms when operators performing different tasks. Specifically, different brain regions were involved in the process of different information, the response regularities of these brain regions differ from each other as well, which lead to the changes of EEG characters with mental workload under different tasks also varied [12], [14], [15]. From the perspective of the statistical characteristics of data, EEG features under different tasks usually appeared to have different probability distribution. However, the classical

Manuscript received 15 September 2022; revised 7 May 2023; accepted 15 May 2023. Date of publication 19 May 2023; date of current version 14 June 2023. This work was supported in part by the State Key Laboratory of Software Development Environment under Grant SKLSDE-2021ZX-08 and in part by the National Natural Science Foundation of China (NSFC) under Grant 32071315. (*Corresponding author: Haijun Niu.*)

Kai Guan is with the State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China, also with the Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China, and also with the Information Science Academy, China Electronics Technology Group Corporation, Beijing 100086, China (e-mail: guankaer@buaa.edu.cn).

Zhimin Zhang, Tao Liu, and Haijun Niu are with the State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China, and also with the Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China (e-mail: zhiminzhang@buaa.edu.cn; tao.liu@buaa.edu.cn; hjniu@buaa.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3277867

machine learning methods hypothesize that all the data were in an independent and identically distribution, these methods would be unsuitable in the cross-task scenarios therefore.

Researches of EEG-based cross-task mental workload evaluation methods can be divided into two categories, namely, common feature selection-based methods [8], [9], [16] and transfer learning-based methods [17]. The common feature selection method obtains the “common” indexes that suitable for various tasks by selecting physiological indexes that change significantly with mental workload in all different tasks, which can be realized by statistical test analysis [18], feature ranking strategies [9], [10] and so on. Ke et al. [18] explored the consistency of EEG power spectrum and task-independent auditory event-related potential with mental workload in two types of cognitive tasks (Verbal N-Back and MATB II tasks) by nonparametric statistical test method. Kakko [9] and Dimitrakopoulos [10] proposed EEG feature selecting methods based on sequence forward search algorithm and model-driven strategy respectively. By using the selected features, satisfactory recognition accuracy was achieved in the cross-task mental workload recognition (N-Back and mental arithmetic tasks).

Transfer learning-based methods aim to find a feature transformation function/matrix that EEG features under different tasks would have same/similar probability distribution after transformation. By using the transferred features, the recognition accuracy of cross-task mental workload evaluation was supposed to be improved [19], [20]. The feature transfer learning method has been widely used in the field of EEG-based Motor Imagery Brain-Computer Interface [19], emotion recognition [21], cognitive state detection [22], [23], [24] and other tasks. In terms of cross-task mental workload evaluation, Zhou et al. [17] first tried to use classical transfer learning models such as Transfer Component Analysis (TCA) and Joint Distribution Analysis (JDA) to evaluate the cross-task mental workload in working memory and mathematics addition tasks, and a higher recognition accuracy than non-transfer learning methods was achieved.

In addition, it should also be noticed that the generally used feature analysis methods usually only consider the single-dimension information of EEG (such as time dimension, frequency dimension, etc.), and these features were processed in the form of vectors in the subsequent processing procedure. Since multiple brain regions were involved and these regions worked collaboratively in a dynamic way when the brain is processing information, the structural characteristics of the EEG signals between different dimensions (such as time, frequency, channel, etc.) would be ignored, which result in the loss of useful information in distinguishing mental workload levels [25], [26], [27]. Various studies demonstrated that the multi-dimension information of EEG signals is sensitive to the level of mental workload [6], [16], [28]. Therefore, it is necessary to explore a feature construction method that can simultaneously characterize multi-mode information of EEG for the accurate mental workload evaluation. Tensor is a representation of high-dimensional information of data, which has been widely used in the fields of gait recognition [29], image recognition [30] and so on. The intrinsic attributes of EEG data can be effectively retained by the tensorized form of representation [25], [26], [27]. In the field of brain-

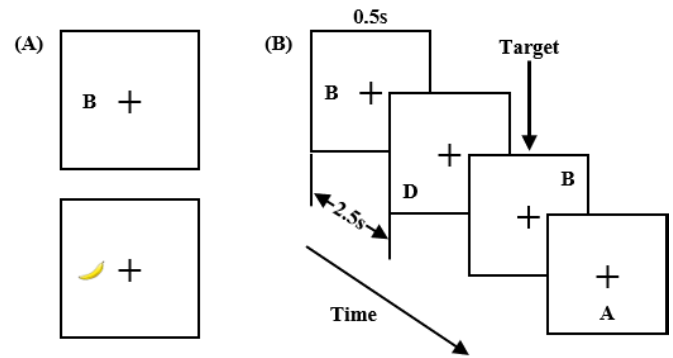


Fig. 1. N-back paradigm.

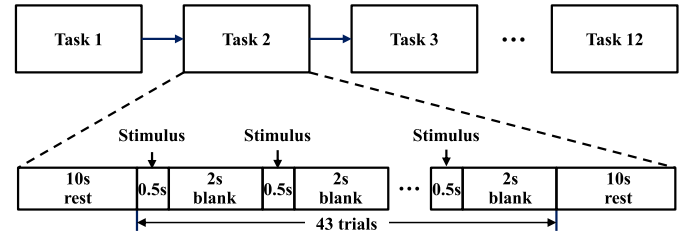


Fig. 2. Experiment protocol.

computer interface, Li [31] and Zhang [32] proposed EEG tensor features construction methods in decoding subjects’ motion intentions in motor imagery tasks. Dao [33] realized the detection of epileptic spikes based on EEG tensor features and tensor decomposition methods. These studies implied that EEG feature construction method based on tensor representation might be feasible for mental workload evaluation.

Considering all the above, it is believed that the tensor form of EEG features can represent the dynamic response characteristics of brain in task-state more comprehensively. Meanwhile, by applying transfer learning procedure on the EEG tensor data, the distance between feature distributions under different tasks can be effectively reduced, thus the accuracy of cross-task mental workload evaluation might be improved. Therefore, this study proposed a cross-task mental workload evaluation method based on the EEG tensor representation and feature transfer learning methods. Specifically, the time-frequency characteristics of multi-channel EEG under different information types and different load levels were extracted to construct EEG feature tensors. Tensorized feature transfer learning method and the corresponding optimization method were then introduced and applied. Finally, based on the classical support vector machine (SVM), mental workload recognition models for both within-task and cross-task scenario were constructed to verify the effectiveness of the proposed method.

II. METHODS

A. Subjects and Experiment Protocol

Sixteen participants were recruited in the experiment (8 males and 8 females, with an average age of 25.6 ± 2.4). All subjects signed the informed consent before the experiment, and this study was approved by the Ethics Committee of Beihang University.

The N-Back paradigm was used to carry out the experiment. Four different information types (verbal, object, space (verbal),

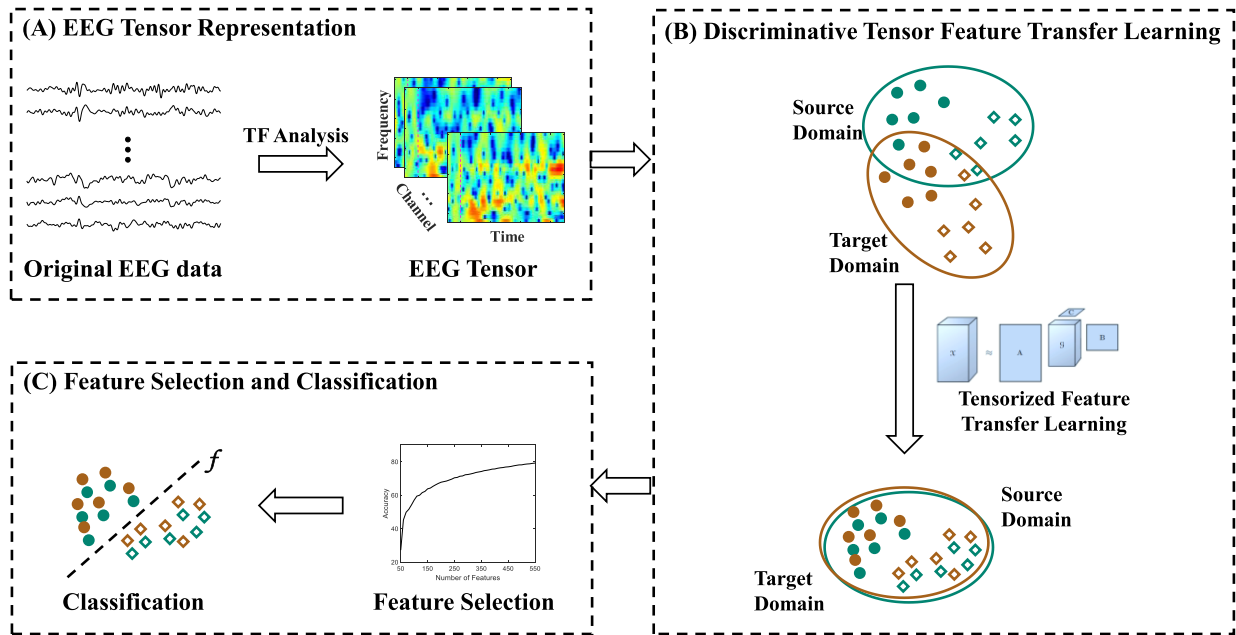


Fig. 3. Flowchart of the EEG Tensor Representation and Discriminative Feature Transfer Learning. (A) The original EEG data samples were represented as 3-ways tensors by using Time-Frequency Analysis. (B) The 3-ways tensors from Source (tensor samples from one task) and Target (tensor samples from another task) Domain were transferred under the constraint of distribution alignment and class-wise discrimination.

space (object)) were used in this study (Fig. 1). Three task loads were set by manipulating the number of items to be remembered in the N-Back tasks (N value from 1 to 3). That is, all subjects need to complete a total of 12 mental workload tasks (4 task types \times 3 task difficulty), which can be seen in Fig. 2. Specifically, under each task condition, the subject needs to judge whether the current presented information is consistent with the previous Nth information in a limited time (2s in this study). Each task condition consisted of 43 trials (of which the first three trials were not used for subsequent analysis). During the tasks, the response time was recorded. After the tasks, the response accuracy was calculated and the participants were asked to finish the NASA-TLX scale. Detailed experimental procedure can also be found in our previous work [8].

B. Data Acquisition and Pre-Processing

The entire experiment was carried out in a sound-shielded room. The NeuroScan system (SynAmps2, USA) was used to record 60-channel EEG signals of the participants in real time, and the signal sampling rate was 1000 Hz. During the acquisition procedure, the electrode impedance is kept below 10k Ω .

The effective components in the original EEG signal are often submerged in noise and artifacts, which would seriously affect the analysis results. Therefore, a band-pass filtering (0.5-40Hz) was used to eliminate the DC component and high-frequency noise of the signal, and the artifact such as blink, vertical/horizontal electrooculogram, and electromyography were removed by using the SASICA algorithm [34]. Signal from 0.5s before stimulation to 2s after stimulation was taken as a trial for subsequent feature analysis.

C. EEG Tensor Representation

Considering the non-stationary and dynamic characteristics of EEG signals, and the close relationship of its temporal,

frequency, and spatial dimensional information to mental workload, the time-frequency-space dimension was selected to construct the EEG tensor feature.

The whole process of EEG feature tensor representation and mental workload recognition is shown in Fig. 3. Considering that the most sensitive components to mental workload of EEG is mainly concentrated in the low frequency band, a band-pass filter (1-20Hz) is first applied, then the time-frequency analysis was carried out for each EEG signal channel from all trials based on the wavelet transform method. The time-frequency data of each channel were combined to construct 3-way tensors (time-frequency-channel) of the task-state EEG signal feature.

D. Tensorized Transfer Learning

The classical transfer learning methods mainly consider the distribution difference of features. Although these methods can align the source domain and target domain samples in the perspective of probability distribution, the effectiveness of classification tasks still needs to be considered. That is to say, it is necessary to consider both the consistency of feature probability distribution and classification discrimination ability when conducting feature learning.

To fill this gap, the proposed tensorized feature transfer learning methods considered the distribution alignment and class-wise discrimination learning in the same time, and the cost function can be expressed as $L = L_{distribution} + \theta L_{discrimination}$, in which the $L_{distribution}$ represents the distribution alignment and the $L_{discrimination}$ names the class-wise discrimination loss. θ is the tuning factors to adjust the weight of the two parts, when $\theta = 0$, the whole cost function was degenerated to the tensorized joint domain distribution alignment method [35].

1) *Distribution Alignment*: The distribution alignment assumes that the feature from source domain and target domain appears to have same or similar distribution in the

subspace with projection of matrix $U = \{U^{(k)}\}_{k=1,\dots,K}$. Considering about the large margin of feature distribution among different tasks, a source domain transfer matrix $P = \{P^{(k)}\}_{k=1,\dots,K}$ was also introduced in this study [36]. According to the classical joint domain distribution alignment method, the proposed distribution alignment method also composed of the marginal and conditional distribution alignment parts, which can be represented as $L_{distribution} = (1 - \mu)L_{marginal} + \mu L_{conditional}$, in which the μ is a tuning factor to regular the weight of the two parts and $\mu \in [0, 1]$. The distance of feature distribution was measured by the Maximum Mean Discrepancy (MMD) criteria. The detailed calculation of the distribution alignment loss was shown as below:

$$\begin{aligned}
 L_{distribution} &= (1 - \mu)L_{marginal} + \mu L_{conditional} \\
 &= (1 - \mu) \left\| \frac{1}{N_S} \sum_{i=1}^{N_S} \left[\left[\llbracket X_S; P \rrbracket; U^T \right] \right. \right. \\
 &\quad \left. \left. - \frac{1}{N_T} \sum_{j=1}^{N_T} \left[\llbracket X_T; U^T \right] \right] \right\|_F^2 \\
 &\quad + \mu \sum_{i=1}^C \left\| \frac{1}{N_S^{(C)}} \sum_{x_m \in X_S^{(C)}} \left[\left[\llbracket X_S^{(C)}; P \rrbracket; U^T \right] \right. \right. \\
 &\quad \left. \left. - \frac{1}{N_T^{(C)}} \sum_{x_n \in X_T^{(C)}} \left[\llbracket X_T^{(C)}; U^T \right] \right] \right\|_F^2 \quad (1)
 \end{aligned}$$

In which, the X_S and X_T represents the tensor sample from source and target domain respectively and the number of samples are N_S and N_T , which was corresponding to the tensor features from different task conditions in this study. P and U denote the source domain feature transfer matrix and the subspace projection matrix respectively, and $\llbracket X_S; P \rrbracket$ means the mode-product of tensor X_S and matrix P ($\llbracket X_S; P \rrbracket = X_S \times_1 P^{(1)} \times_2 P^{(2)} \dots \times_k P^{(k)}$). C is the number of class/workload level, $N_S^{(C)}$ and $N_T^{(C)}$ are the sample number in source or target domain that belong to the class C . $tr(\cdot)$ is the trace of matrix, and k represents the mode of the feature tensor. By unfolding (1) with mode- k , the $L_{distribution}$ can be calculated as:

$$\begin{aligned}
 L_{distribution} &= (1 - \mu) tr \left(U^{(k)} X^{(k)} Q_0 X^{(k)T} U^{(k)T} \right) \\
 &\quad + \mu tr \left(U^{(k)} X^{(k)} Q_1 X^{(k)T} U^{(k)T} \right) \quad (2)
 \end{aligned}$$

in which, $X^{(k)} = \left[\llbracket X_S; P \rrbracket^{(k)}, X_T^{(k)} \right] = \left[Z_S^{(k)}, X_T^{(k)} \right]$, Q_0 and Q_1 are the MMD matrix of marginal and conditional distribution alignment, which can be calculated by:

$$Q_0 = \begin{cases} 1/(N_S \cdot N_S), & x_i, x_j \in X_S \\ 1/(N_T \cdot N_T), & x_i, x_j \in X_T \\ -1/(N_S \cdot N_T), & otherwise \end{cases} \quad (3)$$

$$Q_1 = \begin{cases} 1/(N_S^{(C)} \cdot N_S^{(C)}), & x_i, x_j \in X_S^{(C)} \\ 1/(N_T^{(C)} \cdot N_T^{(C)}), & x_i, x_j \in X_T^{(C)} \\ -1/(N_S^{(C)} \cdot N_T^{(C)}), & \begin{cases} x_i \in X_S^{(C)}, & x_j \in X_T^{(C)} \\ x_i \in X_T^{(C)}, & x_j \in X_S^{(C)} \end{cases} \\ 0, & otherwise \end{cases} \quad (4)$$

2) *Class-Wise Discrimination Learning*: Although minimizing the distribution distance of tensor features can reduce the distribution diversity, it cannot guarantee the learned features have enough ability in distinguishing workload levels. Therefore, the General Tensor Discriminative Analysis (GTDA) [29], [31] methods were introduced to construct the objective function of class-wise discrimination learning procedure after the distribution alignment, which can be seen as:

$$\begin{aligned}
 L_{discrimination} &= L_{between} - \zeta L_{within} \\
 &= \sum_{i=1}^C N_S^{(C)} \left\| M_S^{(C)} - M_S \right\|_F^2 + \sum_{j=1}^C N_T^{(C)} \left\| M_T^{(C)} - M_T \right\|_F^2 \\
 &\quad - \zeta \sum_{i=1}^C \left(\sum_{j=1}^{N_S^{(C)}} \left\| G_{S,j}^{(C)} - M_S^{(C)} \right\|_F^2 + \sum_{k=1}^{N_T^{(C)}} \left\| G_{T,k}^{(C)} - M_T^{(C)} \right\|_F^2 \right) \quad (5)
 \end{aligned}$$

In which, $L_{between}$ and L_{within} represents the margin of intra- and inter-class, M_S and M_T are the mean value of the data from source and target domain, $M_S^{(C)}$ and $M_T^{(C)}$ are the mean value of the data of class C in source and target domain, G_S and G_T denote the tensor feature after the projection and $G_S = \llbracket \llbracket X_S; P \rrbracket; U \rrbracket$, $G_T = \llbracket X_T; U \rrbracket$. By unfolding (5), the class-wise discrimination learning cost function can be represented as:

$$L_{discrimination} = tr \left(U^{(k)} \left(B^{(k)} - \zeta W^{(k)} \right) U^{(k)T} \right) \quad (6)$$

$B^{(k)}$ and $W^{(k)}$ are the intra- and inter-class matrix of the k -mode unfolded matrix, which can be calculated as:

$$\begin{aligned}
 B^{(k)} &= \sum_{i=1}^C \left[N_i mat^{(k)} \left((M_i - M) \times_k U^{(k)T} \right) \right. \\
 &\quad \left. \times mat^{(k)T} \left((M_i - M) \times_k U^{(k)T} \right) \right] \\
 W^{(k)} &= \sum_{i=1}^C \sum_{j=1}^{N_i} \left[mat^{(k)} \left((X_{i,j} - M_i) \times_k U^{(k)T} \right) \right. \\
 &\quad \left. \times mat^{(k)T} \left((X_{i,j} - M_i) \times_k U^{(k)T} \right) \right] \quad (7)
 \end{aligned}$$

In which, $mat^{(k)}(\cdot)$ is the mode- k matrix. The factor ζ is set as the maximal eigenvalue of $(W^{(k)})^{-1} \cdot B^{(k)}$.

3) *Optimization*: The objective of the proposed method is to learn the transfer matrix U and P . By using these two transfer matrixes, the EEG features from different tasks would have similar distribution and class-wise discrimination property. According to the cost function of distribution alignment and

the class-wise discrimination learning depicted in Eq. (2) and Eq. (6), the whole optimization object can be arranged as:

$$L = \arg \min_{U, P} (1 - \mu) \text{tr} \left(U^{(k)} X^{(k)} Q_0 X^{(k)T} U^{(k)T} \right) + \mu \cdot \text{tr} \left(U^{(k)} X^{(k)} Q_1 X^{(k)T} U^{(k)T} \right) + \theta \cdot \text{tr} \left(U^{(k)} \left(B^{(k)} - \zeta W^{(k)} \right) U^{(k)T} \right) \quad (8)$$

s.t. $\forall k, \quad U^{(k)T} U^{(k)} = I, \quad P^{(k)T} P^{(k)} = I$

Since there are several parameters to be solved (U and P), the Alternating Iteration Strategy was applied. In detail, the problem was decomposed into subproblems with one parameter to be solved by fixing the other. These subproblems were optimized iteratively until convergence to obtain an approximate solution [37]. The detailed calculation procedure are as follows:

Step 1 (Optimize U Given P): The subspace matrix U can be solved by High Order Singular Value Decomposition (HOSVD) of tensors. To solve this problem, an auxiliary variable $Z_S = \llbracket X_S; P \rrbracket$ was introduced and the optimizing problem (8) can be converted to the SVD of $(1 - \mu) Q_0^{(k)} + \mu Q_C^{(k)} + \theta (B^{(k)} - \zeta W^{(k)})$.

Step 2 (Optimize P Given U): The optimization object was degenerated to $L_{distribution}$ when U was fixed, that is:

$$\min_P (1 - \mu) \times \left\| \frac{1}{N_S} \sum_{i=1}^{N_S} \left[\llbracket X_S; P \rrbracket; U^T \right] - \frac{1}{N_T} \sum_{j=1}^{N_T} \left[X_T; U^T \right] \right\|_F^2 + \mu \sum_{i=1}^C \left\| \frac{1}{N_S^{(C)}} \sum_{x_m \in X_S^{(C)}} \left[\llbracket X_S^{(C)}; P \rrbracket; U^T \right] - \frac{1}{N_T^{(C)}} \sum_{x_n \in X_T^{(C)}} \left[X_T^{(C)}; U^T \right] \right\|_F^2 \quad (9)$$

By unfolding (9), the problem can be converted to (10), which can be solved by using the classical optimizing calculating procedure [38].

$$\min_P \left[\begin{array}{l} \text{tr} \left(P^{(k)} \frac{1}{N_S^2} U^{T(k)} X_S^{(k)} I_S I_S^T X_S^{(k)T} U^{(k)} P^{(k)T} \right) \\ - \text{tr} \left(P^{(k)} \frac{2}{N_S} U^{T(k)} X_S^{(k)} I_S \right) + \text{tr} \left(Z_T^{(k)} Z_T^{T(k)} \right) \end{array} \right] = \min_P \left[\text{tr} \left(P^{(k)} Q_{QT} P^{(k)T} \right) - \text{tr} \left(P^{(k)} Q_{YT} \right) + \text{tr} \left(Y_{YT} \right) \right] \quad (10)$$

At last, the tensor feature from source and target domain can be represented as X_S^* and X_T^* by using the projection matrix U and P , which can be calculated by (11).

$$X_S^* = \left[\llbracket X_S; P \rrbracket; \quad U^T \right] = X_S \times_1 \left(U^{(1)T} P^{(1)} \right) \times_2 \cdots \times_k \left(U^{(k)T} P^{(k)} \right) X_T^* = \left[X_T; U^T \right] = X_T \times_1 U^{(1)T} \times_2 \cdots \times_k U^{(k)T} \quad (11)$$

Algorithm Discriminative Tensor Feature Transfer Learning

Input: Source and target domain data sets: $\{X_S^i, Y_S^i\}_{i=1}^{N_S}, \{X_T^i, Y_T^i\}_{i=1}^{N_T}$; Tuning parameters: μ, θ .

Output: Subspace projection matrix U ; Source domain transformation matrix P ; Transferred data: X_S^*, X_T^* .

- 1: Initialize U, P
 - 2: **FOR** iter = 1 to T
 - 3: **FOR** k = 1 to nDim
 - 4: Update U by SVD on $(1 - \mu) Q_0^{(k)} + \mu Q_C^{(k)} + \theta (B^{(k)} - \zeta W^{(k)})$
 - 5: Update P by solving Eq. (10)
 - 6: **END FOR**
 - 7: Check for convergence
 - 8: **END FOR**
 - 9: Calculate X_S^*, X_T^* by Eq. (11)
 - 10: Return X_S^*, X_T^*, U, P
-

In total, the whole procedure of the proposed algorithm is as follows:

E. Feature Selection and Classification

In order to remove the redundant information in the feature and reduce the complexity of the classification model, Fisher Score is used to select features that give most contributions in distinguishing mental workload for subsequent classification steps [39]. The central idea of the Fisher Score is to find out a subset of features which have the largest inter-class margin and the smallest intra-class margin. Specifically, for a certain feature X^i , let μ_c^i and σ_c^i be the mean and standard deviation of the features in category C , and μ_c^i, σ_c^i be the mean and standard deviation of all samples, respectively. The Fisher Score of this feature can be calculated by Equation (12). After sorting the Fisher Score of all the features in a descending order, the first N features were taken as the input of the classification model.

$$FisherScore \left(X^i \right) = \frac{\sum_{c=1}^C N_c \left\| \mu_c^i - \mu^i \right\|^2}{\sum_{c=1}^C N_c \left(\sigma_c^i \right)^2} \quad (12)$$

In order to verify the feasibility of the proposed feature construction method for mental workload recognition, a SVM classifier was used to construct a 3-class recognition model for both within-task and cross-task conditions. For within-task classification, a 5-fold cross-validation procedure for each of the four tasks was applied. For cross-task mental workload identification, one of the four tasks was selected as the training task, and the other three were used as the validation set, which can also be called leave-one-task-out approach. In order to evaluate the recognition effect, the accuracy, sensitivity and specificity were calculated respectively.

III. RESULTS

A. Classification Performance

Table I shows the classification results of the proposed method in within-task and cross-task workload evaluation.

TABLE I
CLASSIFICATION RESULTS FOR BOTH
WITHIN AND CROSS TASK SCENARIOS

	Train on	Test on	Accuracy	Sensitivity	Specificity
Within task	V	V	0.910±0.044	0.865±0.056	0.865±0.034
	O	O	0.902±0.008	0.854±0.015	0.853±0.007
	S(V)	S(V)	0.919±0.023	0.879±0.082	0.878±0.044
	S(O)	S(O)	0.914±0.091	0.872±0.054	0.871±0.026
	Average		0.911	0.867	0.867
Cross task		O	0.808±0.086	0.831±0.043	0.802±0.042
	V	S(V)	0.822±0.067	0.853±0.018	0.793±0.090
		S(O)	0.791±0.065	0.754±0.014	0.780±0.039
		V	0.830±0.056	0.793±0.055	0.878±0.013
	O	S(V)	0.785±0.043	0.795±0.087	0.768±0.011
		S(O)	0.798±0.069	0.794±0.035	0.782±0.006
		V	0.830±0.063	0.846±0.038	0.806±0.044
	S(V)	O	0.802±0.091	0.798±0.028	0.805±0.073
		S(O)	0.814±0.082	0.822±0.051	0.826±0.045
		V	0.816±0.067	0.819±0.062	0.827±0.051
		O	0.833±0.068	0.828±0.059	0.814±0.082
	S(V)	0.815±0.072	0.807±0.080	0.794±0.024	
	Average		0.813	0.812	0.806

Note: V: Verbal, O: Object, S(V): Spatial (Verbal), S(O): Spatial (Object)

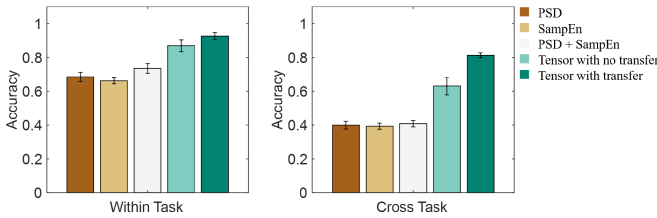


Fig. 4. Comparison with widely used no-transferred EEG features.

An average recognition accuracy of 91.1% for the within-task and 81.3% for cross-task mental workload recognition were achieved by using the EEG tensor feature and transfer learning procedure.

Previous researches have demonstrated that various EEG features can be used in distinguishing mental workload levels. In order to validate the effectiveness of the proposed tensor transfer learning feature and the classical feature extraction method for mental workload recognition, the classical single mode features and tensor features without transfer learning were also tested in both within and cross-task workload classification under the same N-Back workload task condition. Specifically, the single mode features include PSD, multi-scale sample entropy respectively. Fig. 4 illustrates the comparison between these classical feature extraction methods and the proposed method in this paper. It can be seen that although the feature extraction method based on time domain (multi-scale sample entropy) and frequency domain (PSD) can achieve satisfactory results in within-task mental workload recognition (Sample Entropy: 0.662 ± 0.016 , PSD: 0.685 ± 0.024 , Sample Entropy + PSD: 0.736 ± 0.025), the recognition effectiveness in cross-task scenario declined severely (Sample Entropy: 0.394 ± 0.018 , PSD: 0.399 ± 0.020 , Sample Entropy + PSD: 0.408 ± 0.025). Compared with these single-dimensional features, the recognition accuracy of tensor features without transfer learning was significantly improved in both within-task and cross-task scenarios (within-task: 0.871 ± 0.079 , cross-task: 0.625 ± 0.029). The feature extraction method based on tensor representation and transfer learning proposed in this paper achieved the highest recognition accuracy.

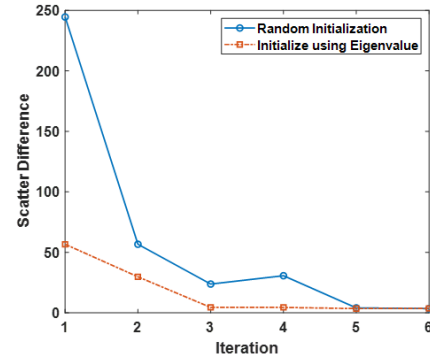


Fig. 5. Scatter difference with iteration by different initialization methods.

Furtherly, Table II illustrated the typical state-of-the-art studies concerning the cross-task mental workload evaluation. It should be noticed that there is no uniformed task paradigm in this field, which means the tasks used in the workload evaluation varied from each other. Therefore, the comparison is not rigorous. To be consistency, Table II mainly focus on researches that involving similar cognitive process (i.e., working memory). It can be seen from the table that researches using only PSD features could hardly obtain a satisfactory classification results, while the accuracies have been improved in distinguishing high and low workload (binary classification) when the connectivity relationship of different EEG channels (or different brain regions) were considered, which is consistency to the results depicted by Fig. 4. The proposed method of this study achieved comparable classification accuracy in a 3-class situation when taking the multi-mode information of EEG account.

B. Convergence and Parameter Sensitivity

In each iteration, the projection matrix U of each mode is determined by the $tr(U(1 - \mu)Q_0^{(k)} + \mu Q_C^{(k)} + \theta(B^{(k)} - \zeta W^{(k)})U')$. Given $U(k)$ and $U(k+1)$ as the projection matrix in k th and $k+1$ th iteration, the total scatter difference matrix of the two iterations is $\psi_{y(k)}$ and $\psi_{y(k+1)}$, respectively according to the calculation formula of intra-class and inter-class dispersion matrix. Since U is taken from the eigenvectors of $\psi_{y(k)}$ with the largest scatter difference matrix, it can be known that $\psi_{y(k)} \leq \psi_{y(k+1)}$. Therefore, it is reasonable that:

$$a = \psi_{y(1)} \leq \psi_{y(2)} \leq \dots \leq \psi_{y(K-1)} \leq \psi_{y(K)} = b \quad (13)$$

That means the total scatter difference is a monotonically changing process with an upper limit a and a lower limit b .

Fig. 5 shows the changes of the total scatter difference obtained by the eigenvalue matrix of $(1 - \mu)Q_0^{(k)} + \mu Q_C^{(k)} + \theta(B^{(k)} - \zeta W^{(k)})$ and by the random initialization way. It can be seen that the scatter difference calculated by the two methods for the first time is quite different, but with the iteration proceeding, the scatter difference quickly converged. After 3~4 iterations, both initialization methods can reach the convergence state.

In the tensor feature representation and transfer learning procedure, the initialization parameters would directly affect the recognition accuracy of mental workload. In order to find out the influence of parameters on classification performance,

TABLE II
COMPARISON BETWEEN THE PROPOSED METHOD AND STATE-OF-THE-ART STUDIES

	Features	Methods	Tasks	Classes	Accuracy
Baldwin [11]	PSD	ANN	Reading span, Spatial N-Back, Sternberg tasks	Binary	0.450
Walter [12]	ERS/ERD	SVM	GO/NO-GO, Verbal N-Back, Reading span, Algebra task	Binary	0.540
Ke [13]	PSD	SVM	Verbal N-Back, Spatial N-Back	4-class	0.290
Zhang [28]	Spatial-spectral-temporal	R3DCNN	Spatial N-Back, Arithmetic tasks	Binary	0.889
Zhou[17]	PSD and connectivity	TCA+SVM	Working Memory, Mathematics Addition tasks	Binary	0.714
Kakkos[9]	PSD and connectivity	SVM	Verbal N-Back, Arithmetic tasks	Binary	0.940
The Proposed Method	Tensorized Transfer Learning	SVM	Verbal, Object, Spatial (Verbal), Spatial (Object) N-Back	3-class	0.813

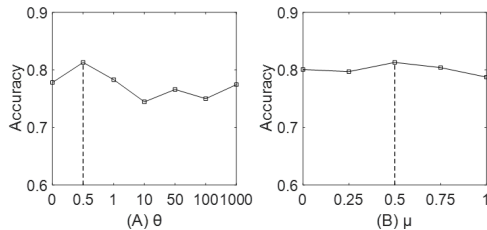


Fig. 6. Accuracy with different θ and μ .

different initialization parameter settings were compared and analyzed. Among these parameters, μ determines the importance of marginal and conditional probability distribution alignment. When it tends to 1, the cost function mainly considers the conditional distribution, and when it tends to 0, the marginal distribution was mainly considered. θ determines the weight of distribution alignment and the class-wise discrimination. Fig. 6 shows the cross-task recognition accuracy changing with θ and μ . It can be seen that when θ is set to 0.5, the recognition accuracy of mental workload is the highest, and when μ is set to 0.5, the recognition accuracy is slightly higher than other values.

C. Computational Complexity

The computation process of the proposed method includes tensor representation of EEG data and feature transfer learning. The tensor representation procedure based on time-frequency analysis of the N EEG channels with signal length L would take $O(L \cdot N)$. Considering about the feature transfer learning, the projection and transformation matrix U and P of each mode (k) were calculated in each iteration, so the number of calculations is $O(T \cdot K)$. Furtherly, the SVD of U and the optimizing of P (Eq. (10)) both cost $O(N^3)$, which leads to a total time complexity of $O(L \cdot N) + O(2 \cdot T \cdot K \cdot N^3)$ [38].

IV. DISCUSSION

The effectiveness of the cross-task mental workload evaluation based on physiological signals such as EEG still remains unsatisfactory. From the perspective of mining multi-dimensional character of EEG signals, a tensor representation and transfer learning-based EEG feature construction method was proposed. After that, recognition model based on the classical SVM classifier in both within-task and cross-task scenario was also realized. Results showed that by using the tensorized and transferred EEG feature, an averaged accuracy of 91.1% for within-task and 81.3% for cross-task mental workload evaluation were achieved, which is significantly higher than the recognition results by using classical features such as PSD, sample entropy and so on. These results demonstrated that it is feasible to use EEG tensor features

for mental workload evaluation, and the tensorized transfer learning method is effective for cross-task mental workload evaluation.

Many studies have demonstrated that the time, frequency, spatial information of EEG signals could reflect the mental workload state of operators [6], [16], [28]. The classical method based on linear and nonlinear characteristics of EEG usually ignored the time information of EEG signals as well as the interaction information between different brain areas (or EEG channels), which might not able to characterize the workload state comprehensively and thus lead to poor effects of mental workload evaluation [28]. The tensor representation of EEG data can fill the gap of the information losing in classical feature construction methods by considering multi-dimensional information of EEG, thus its recognition accuracy is significantly higher than that of classical method. Meanwhile, this result is consistent with the previous researches which believe that multi-dimensional information of EEG makes contribution to mental workload recognition. At the same time, it also confirms the necessity of applying multi-dimensional feature representation (i.e., tensor representation) to evaluate mental workload. Furtherly, compared with the EEG time-frequency-spatial tensor features without transfer learning, the EEG tensor features after transfer learning have higher recognition accuracy in both within-task and cross-task conditions, indicating that the feature distribution and class discriminative learning procedure both have good effect on distinguishing mental workload levels, which also verified the validity of the proposed method.

Considering the influence of parameter settings on recognition accuracy, it is found that the optimal classification results cannot be achieved in the cross-task mental workload evaluation when the distribution alignment of features (with θ tending to 0) or class discrimination information (with θ tending to 1000) was the only factor considered. Although the analysis of parameter sensitivity in this study relies on experience and attempts, these results can still demonstrate that it is necessary to consider both the distribution of features and the discriminative features in the transfer learning process. Optimization criteria-based methods for parameter selection can be explored in the future researches [40]. In addition, tensorized classifiers such as support tensor machine can be applied in the future to further verify the effectiveness of the proposed method instead of the classic SVM model [41].

In conclusion, this paper proposed a cross-task mental workload evaluation method based on EEG tensor representation and transfer learning. Satisfactory results in both within-task and cross-task conditions were achieved, which can provide theoretical basis and application reference value

for future research. However, the research of this paper is still limited to mental workload recognition in a pre-set task environment (i.e., N-Back tasks), the effectiveness of the proposed methods in real human-machine systems with more complex information and operations needs to be further verified.

ACKNOWLEDGMENT

The authors would like to thank all of the participants who generously volunteered their time to participate in this study.

REFERENCES

- [1] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015.
- [2] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, "Cognitive workload recognition using EEG signals and machine learning: A review," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 799–818, Sep. 2022.
- [3] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988.
- [4] R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Appl. Ergonom.*, vol. 74, pp. 221–232, Jan. 2019.
- [5] A. Gevins, "High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice," *Cerebral Cortex*, vol. 7, no. 4, pp. 374–385, Jun. 1997.
- [6] E. Astrand, "A continuous time-resolved measure decoded from EEG oscillatory activity predicts working memory task performance," *J. Neural Eng.*, vol. 15, no. 3, Apr. 2018, Art. no. 036021.
- [7] P. Liu, W. Beh, C. Shih, Y. Chen, and A. A. Wu, "Entropy and complexity assisted EEG-based mental workload assessment system," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2019, pp. 1–4.
- [8] K. Guan, Z. Zhang, X. Chai, Z. Tian, T. Liu, and H. Niu, "EEG based dynamic functional connectivity analysis in mental workload tasks with different types of information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 632–642, 2022.
- [9] I. Kakkos et al., "EEG fingerprints of task-independent mental workload discrimination," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3824–3833, Oct. 2021.
- [10] G. N. Dimitrakopoulos et al., "Task-independent mental workload classification based upon common multiband EEG cortical connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017.
- [11] C. L. Baldwin and B. N. Penaranda, "Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48–56, Jan. 2012.
- [12] C. Walter, S. Schmidt, W. Rosenstiel, P. Gerjets, and M. Bogdan, "Using cross-task classification for classifying workload levels in complex learning tasks," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 876–881.
- [13] Y. Ke et al., "An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task," *Frontiers Hum. Neurosci.*, vol. 8, pp. 1–12, Sep. 2014.
- [14] T. B. Christophel, P. C. Klink, B. Spitzer, P. R. Roelfsema, and J.-D. Haynes, "The distributed nature of working memory," *Trends Cognit. Sci.*, vol. 21, no. 2, pp. 111–124, Feb. 2017.
- [15] H. Walter et al., "Evidence for quantitative domain dominance for verbal and spatial working memory in frontal and parietal cortex," *Cortex*, vol. 39, nos. 4–5, pp. 897–911, 2003.
- [16] M. J. Boring, K. Ridgeway, M. Shvartsman, and T. R. Jonker, "Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 056016.
- [17] Y. Zhou et al., "Cross-task cognitive workload recognition based on EEG and domain adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 50–60, 2022.
- [18] Y. Ke et al., "Cross-task consistency of electroencephalography-based mental workload indicators: Comparisons between power spectral density and task-irrelevant auditory event-related potentials," *Frontiers Neurosci.*, vol. 15, Nov. 2021, Art. no. 703139.
- [19] D. Wu, Y. Xu, and B. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2022.
- [20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.
- [21] W. Li, W. Huan, B. Hou, Y. Tian, Z. Zhang, and A. Song, "Can emotion be transferred?—A review on transfer learning for EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 833–846, Sep. 2022.
- [22] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C. Lin, "EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1752–1764, Mar. 2021.
- [23] Y. Cui, Y. Xu, and D. Wu, "EEG-based driver drowsiness estimation using feature weighted episodic training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2263–2273, Nov. 2019.
- [24] L.-L. Chen, A. Zhang, and X.-G. Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Syst. Appl.*, vol. 137, pp. 266–280, Dec. 2019.
- [25] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, Jun. 2015.
- [26] Y. Pei, "A tensor-based frequency features combination method for brain-computer interfaces," in *Proc. Int. Conf. Cognit. Syst. Signal Process.*, 2021, pp. 511–526.
- [27] S. Huang, Y. Chen, T. Wang, and T. Ma, "Spectrum-weighted tensor discriminant analysis for motor imagery-based BCI," *IEEE Access*, vol. 8, pp. 93749–93759, 2020.
- [28] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1149–1159, Jun. 2019.
- [29] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [30] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [31] J. Li, L. Zhang, D. Tao, H. Sun, and Q. Zhao, "A prior neurophysiologic knowledge free tensor-based scheme for single trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 2, pp. 107–115, Apr. 2009.
- [32] Q. Zhang, B. Guo, W. Kong, X. Xi, Y. Zhou, and F. Gao, "Tensor-based dynamic brain functional network for motor imagery classification," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102940.
- [33] L. T. Thanh, N. T. A. Dao, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Multi-channel EEG epileptic spike detection by a new method of tensor decomposition," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016023.
- [34] M. Chaumon, D. V. M. Bishop, and N. A. Busch, "A practical guide to the selection of independent components of the electroencephalogram for artifact correction," *J. Neurosci. Methods*, vol. 250, pp. 47–63, Jul. 2015.
- [35] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2013, pp. 2200–2207.
- [36] A. Braytee, M. Naji, and P. J. Kennedy, "Unsupervised domain-adaptation-based tensor feature learning with structure preservation," *IEEE Trans. Artif. Intell.*, vol. 3, no. 3, pp. 370–380, Jun. 2022.
- [37] H. Lu et al., "When unsupervised domain adaptation meets tensor representations," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 599–608.
- [38] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, nos. 1–2, pp. 397–434, Dec. 2013.
- [39] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," 2012, *arXiv:1202.3725*.
- [40] J. Wang, Y. Chen, W. Feng, H. Yu, M. Huang, and Q. Yang, "Transfer learning with dynamic distribution adaptation," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 1, pp. 1–25, Feb. 2020.
- [41] Z. Hao, L. He, B. Chen, and X. Yang, "A linear support higher-order tensor machine for classification," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2911–2920, Jul. 2013.