Reducing Calibration Efforts in RSVP Tasks With Multi-Source Adversarial Domain Adaptation

Wei Wei, Shuang Qiu, Xuelin Ma, Dan Li, Bo Wang, and Huiguang He, *Senior Member, IEEE*

***Abstract*— Rapid Serial Visual Presentation (RSVP)-based Brain-Computer Interface (BCI) is an efficient information detection technology by detecting event-related brain responses evoked by target visual stimuli. However, a time-consuming calibration procedure is needed before a new user can use this system. Thus, it is important to reduce calibration efforts for BCI applications. In this article, we propose a multi-source conditional adversarial domain adaptation with the correlation metric learning (mCADA-C) framework that utilizes data from other subjects to reduce the data requirement from the new subject for training the model. This model utilizes adversarial training to enable a CNN-based feature extraction network to extract common features from different domains. A correlation metric learning (CML) loss is proposed to constrain the correlation of features based on class and domain to maximize the intra-class similarity and minimize inter-class similarity. Also, a multi-source framework with a source selection strategy is adopted to integrate the results of multiple domain adaptation. We constructed an RSVP-based dataset that includes 11 subjects each performing three RSVP experiments on three different days. The experimental results demonstrate that our proposed method can achieve 87.72% cross-subject balanced-accuracy under one block calibration. The results indicate our method can realize a higher performance with less calibration efforts.**

***Index Terms*— EEG, RSVP-based BCI, calibration reduction, multi-source domain adaptation, correlation metric learning.**

Manuscript received May 22, 2020; revised August 5, 2020; accepted September 7, 2020. Date of publication September 14, 2020; date of current version November 6, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 81701785, and Grant 61906188; in part by the Strategic Priority Research Program of CAS under Grant XDB32040200; and in part by the CAS International Collaboration Key Project under Grant 173211KYSB20190024. *(Wei Wei and Shuang Qiu contributed equally to this work.) (Corresponding author: Huiguang He.)*

Wei Wei, Xuelin Ma, Dan Li, and Bo Wang are with the National Laboratory of Pattern Recognition, Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

Shuang Qiu is with the National Laboratory of Pattern Recognition, Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

Huiguang He is with the National Laboratory of Pattern Recognition, Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100864, China (e-mail: huiguang.he@ia.ac.cn).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.3023761

I. Introduction

THE brain-computer interfaces (BCIs) system provides a direct communication and control channel between a person's brain and external devices rather than depending on peripheral nerves and muscles [1]. BCIs have shown a variety of potential applications for communication, control, and rehabilitation [2]. Electroencephalogram (EEG) is one of the most common approach for building a BCI system because of its cost effectiveness, non-invasive implementation, and portability.

There are many traditional BCI paradigms such as motor imagery (MI) [3]–[5], steady-state visual evoked potentials (SSVEP) [6], [7], slow cortical potentials [8], and the P300 [9]–[11].

Rapid serial visual presentation (RSVP) is a process of sequentially displaying images at the same spatial location at high presentation rates. RSVP-based BCIs are a specific type of BCI that can be used for detecting target configurations in the stream through detecting event related brain responses. In particular, if targets are rare, the P300 component is likely to be produced in response to them a 300-500 millisecond delay after the onset of the target stimulus.

Compared with computer vision, human vision is more effective and robust [12]. RSVP-based BCI systems can deal with the problems that computer vision is difficult to solve. Christopher B. *et al.* developed an RSVP-based BCI system for finding the mine-like object from the sonar imagery, which takes advantage of the accuracy and rapidity of human vision [13]. Recently, various RSVP-based BCI applications have been developed such as speller [14], [15], image retrieval [12], image classification [16], [17], anomaly detection [13], anti-deception [18]. RSVP-based speller is gaze independent and not being easy to cause visual fatigue, thus it is a very effective way for patients and external communication, especially for those patients with severe oculomotor impairments [19].

Many different methods have been proposed to classify single-trial EEG in RSVP tasks, such as hierarchical discriminant component analysis (HDCA) [16], minimum distance to Riemannian mean (MDRM) [20], [21], MCNN [22], EEG-Net [23]. Although they showed their suitability to single-trial EEG classification in the RSVP paradigm, these methods were to train a new model for a new subject. Also, these models are hard to generalize across subjects due to the individual differences of EEG. Thus, for a new subject,

we need to collect a lot of EEG data for training models, which will be then used for the subsequent experiment. This means a time-consuming calibration procedure in the BCI experiment. Previous studies proved that the performance of the subject-dependent model decreased when there were few training data [21], [24]. Thus, it is still a challenge to train a reliable model with calibration reduction. In this article, we aim to present a transfer learning method that utilizes EEG data from different participants or different sessions to realize cross-subject and cross-session learning. This will reduce the demand of new participant for calibration data, so as to achieve calibration with less efforts for the new participant.

In the EEG transfer learning area, EEG data from different subjects or sessions are regarded as different domains. For a new user, a database of past users can be considered to initialize or train the model. This form of learning is referred to as cross-subject learning. Data from previous sessions of the user employed in training, this is referred to as cross-session learning. Data of the existing subject are called the source domain and the data of the new subject are called the target domain. Our proposed transfer learning method belongs to transductive transfer learning, also known as domain adaptation (DA), where the feature or distribution of source and target domain is distinct while the tasks are the same [25].

In this article, we propose a multi-source Conditional Adversarial Domain Adaptation network with Correlation metric learning (mCADA-C) framework to reduce the BCI calibration efforts and improve the classification performance in RSVP image retrieval. Our mCADA-C framework includes three parts: First, the adversarial training of feature extractor and conditional discriminator helps the feature extractor learn the common features of the target domain and the source domain. This uses a large amount of source domain data to reduce the calibration demand for the target. Secondly, correlation metric learning loss constrains the correlation of features depending on the class and domain, which maximizes the intra-class correlation and minimizes the inter-class correlation. Finally, a multi-source domain adaptation framework with a source selection strategy is brought up to effectively integrate multiple domain adaptation results to improve classification performance further.

The rest of the paper is structured as follows: First, we reviewed the related work in Section II. The RSVP experiment we conduct and the EEG data pre-process method will be illustrated in Section III. Section IV introduced the mCADA-C framework we developed, and the structure and function of each module in the network will be explained respectively. In section V, we apply our network to the image retrieval of RSVP. A variety of basic classification methods and transfer learning methods are compared. The validity of each part of the network is verified and analyzed. The results and analysis of each experiment are present in Section VI. The discussion on the research analyzes the problems and adds some details of solutions in the research, which is given in Section VII. Lastly, we draw a conclusion of this study in Section VIII.

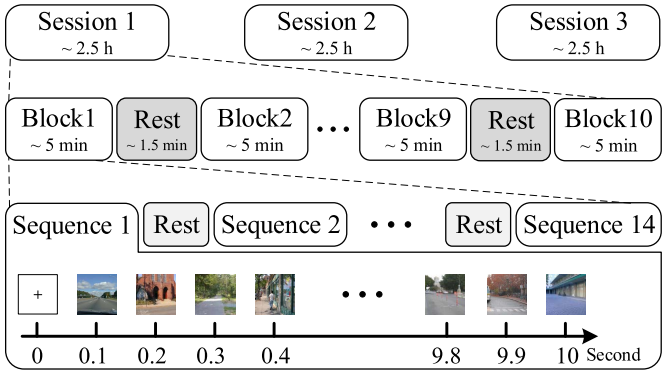
II. Related Work

Transfer learning can be used in the brain-computer interface for solving cross-session or cross-subject problems [26]. The BCI transfer learning method based on Riemannian geometry has been studied in the past few years. In 2018, Rodrigues *et al.* [27] proposed a Riemannian Procrustes Analysis (RPA) method based on the MDRM classifier, which matches the statistical distributions of symmetric positive definite (SPD) matrices from different domains. Compared with the basic MDRM classifier, RPA improved the classification accuracy of MI by 8%, but there was no significant improvement in P300 tasks. This might be due to that the temporal information in EEG could not be effectively utilized when the covariance matrix of EEG was constructed as the input. In [24], [28], and [29], the authors also used a similar Riemannian manifold alignment method for EEG transfer learning. He and Wu [30] promoted a data alignment method in Euclidean space.

In recent years, the research of EEG decoding based on neural networks has been developed. There are different EEG decoding methods based on neural networks, such as [22], [23], and [31]. The transfer learning method based on the neural network has also attracted attention in the research of BCI [32]. The generative adversarial network (GAN) [33] has been successfully applied in domain adaptation areas, for example, DANN [34], ADDA [35]. Li *et al.* [36] promoted an unsupervised domain adaptation model to learn a similar latent representation shared in different domains by the gradient reversal in EEG emotion recognition. There was a 7.8% increase in classification accuracy compared to DANN. This method is based on the differential entropy feature of EEG, which is widely used in emotion tasks. But this feature or this model based on the feature may be not suitable for RSVP tasks because of the temporal characteristic of RSVP EEG. A multi-path network structure based on DANN has been used in a MI study by Jeon *et al.* [37], the average accuracy increased by 4% after domain adaptation. Tang and Zhang [38] promoted a conditional adversarial domain adaptation neural network for MI EEG Decoding. Jimnez-Guarneros and Gomez-Gil [39] promoted a domain adaptation for cross-subject cognitive load recognition. Manifold Embedded Knowledge Transfer (MEKT) [40] uses the domain adaptation on the tangent space of Riemannian manifold, to conduct offline unsupervised cross-subject EEG classification.

Ming proposed a multi-source DA method, subject adaptation network (SAN) [41]. The adversarial network in SAN is adopted to constrain the distribution of features to an artificial distribution which includes prior information on the proportion of classes. The classification accuracy of SAN outperformed EEG-Net. SAN needs to learn features from several different domains at the same time with an indirect constraint on feature space, which may be insufficient to learn a common feature. Compared to single-source domain adaptation, a multi-source domain adaptation can introduce more task-related knowledge which usually can force a better task performance on the target domain. In applications, there are EEG collected from different subjects which means there are multiple source domains, so multi-source DA needs more research and attention.

(a) non-target: street scenes (b) target: including pedestrian



(c) experiment setting of block

Fig. 1. (a) Examples of non-target pictures, (b) examples of target pictures and (c) settings of our RSVP experiment. The rest time between blocks were around 1-3 mins. The rest time between sequences were controlled by subjects, around 4-6 s.

III. Materials

A. Subjects

In the experiment, 11 participants (9 right-handed; 6 males and 5 females; aged 22.8 $\pm$ 2.1) were recruited. Participants had corrected-to-normal or normal vision. In addition, no participant had any previous history of visual disorders and neurological disease or injury. All participants were naïve in respect to RSVP-based BCI experiments. Experimental procedures were approved by the Institutional Review Board of the Institute of Automation, Chinese Academy of Science. All subjects gave their written informed consent before testing.

B. RSVP Paradigm

The visual stimuli of our experiment consisted of 1400 images (500 $\times$ 500 pixels) from the scenes and objects database [42] released by MIT CSAIL. These images were separated into target scenes that included pedestrians and non-target scenes that did not include pedestrians (See Fig. 1, (a), (b)).

The images were randomly presented at a rate of 10Hz. And the probability of target images was 4%. Participants were seated 1 meter from a monitor with 1920*1080 resolution. Subjects were instructed to press the space bar on the keyboard as quickly as possible when they found the target.

There were 10 blocks for each experimental session, and each block included 1400 pictures which were divided into 14 sequences, each consisting of 100 images. The interval between two sequences was controlled by the subject. Each block was completed in about 5 minutes, the rest time between blocks were around 2-3 mins. Each subject needed to carry out three session experiments, and the interval between the two adjacent sessions was 5-10 days.

TABLE I
Notation Summary

Notation	Meaning
x, y	EEG sample with m channels and n time points, $x \in R^{m \times n}$, and corresponding label $y \in \{0, 1\}$, represent the non-target and target, respectively.
X^S, Y^S	Data and labels of source domain, with subscript on domain X^{S_i} and Y^{S_i}, denote the data and labels of the i-th source.
X^T, Y^T	Data of Target domain, training data and label of target subject respectively.
X, Y	Training data in a CADA network, $X = X^T \cup X^S, Y = Y^T \cup Y^S$.
X_U^T	Test data of target domain, test data of target subject without label is not available during the training.
F, D, C	Feature extractor, Discriminator, Classifier.
$\theta_F, \theta_D, \theta_C$	Parameter of feature extractor, discriminator, and classifier, respectively.

C. Data Acquisition and Preprocessing

The EEG data were recorded by a SynAmp2 system (NeuroScan, Australia) using 64-channel Ag/AgCl electrodes placed according to the International 10/20 system, with a sampling rate of 1000 Hz. The electrodes were referenced to the vertex and grounded to the forehead. The impedance of each electrode was kept below 10 kΩ.

In the preprocessing stage, for each block of EEG data, the EEG data were down-sampled to 250 Hz. Then, the signals were bandpass filtered between 0.1 and 15 Hz with a 3-order ButterWorth filter, which was implemented as linear phase, to remove slow drift and high-frequency noise and prevent delay distortions. Afterward, the preprocessed data of each block were segmented into EEG trials. Each trial including 1 second EEG data, starting from their stimulation onset to 1000 milliseconds after the stimulation onset (0s to 1s). For each trial, data was normalized to zero mean and variance one. The subsequent analysis and classification of EEG were based on these segmented EEG trials (samples). According to our experiment paradigm, each subject has 10 block $\times$ 1,400 trial EEG samples per session, where 560 were target samples, and others were non-target samples.

IV. Method

In this section, we present our proposed multi-source conditional adversarial domain adaptation (mCADA) framework. We first introduce the source selection strategy and the multi-source DA framework. Then the basic EEG feature extraction and classification network in our method will be introduced. Next, the detail of our proposed CADA network will be elaborated. Finally, a correlation metric learning (CML) loss function is proposed for CADA.

We summarize the notations in Table I. for a clear narrative. Our application scenario on cross-session/subject learning is supervised domain adaptation. Using $\{(X^T, Y^T) \cup (X^{S_1}, Y^{S_1}) \cup \cdots \cup (X^{S_M}, Y^{S_M})\}$ to construct a model and predict the label of X_U^T.

A. Source Selection Strategy

There are 11 subjects in our dataset. If one subject was set as a target domain, thus, the other 10 subjects could be set as the

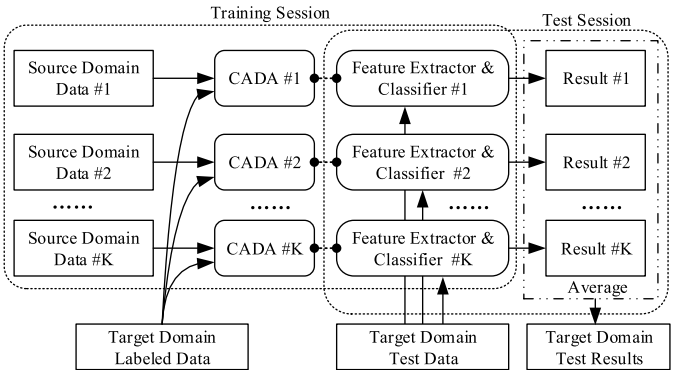


Fig. 2. The multi-source domain adaptation framework.

source domain ($M = 10$). Given the possible negative transfer and computation cost, we perform a source selection procedure for our multi-source DA framework. The more similar between the distribution of source domain and that of target domain is better for transfer learning. Here, we used cross-domain classification performance to measure the similarity between different domain data. Firstly, the labeled data of M subjects are used to train M classifiers. Secondly, for a new subject (X^T, Y^T), use the M classifiers to classify X^T and get M classification results. Thirdly, arrange M classification accuracy in a descending order. The data of subjects corresponding to the top K accuracy are selected as the source for subsequent domain adaptation. The K is an important parameter in our model, the choice of K will be present in Section VI.

As shown in Fig. 2, we propose the multi-source domain adaptation framework to integrate the results of multi-source domain adaptation. In our framework, each source domain data $X^{S_{K_i}}, Y^{S_{K_i}}$ are feed in a CADA network with target domain labeled data X^T, Y^T in the training session, the data in target domain shares network parameters with the source domain. In the test session, the unlabeled test data X_U^T will be used as the input of K feature extractors and classifiers (part of the DA network, trained in the training session), and obtain K classification results. Finally, the average combination rule is used to integrate output of K classifiers.

B. EEG Feature Extraction and Classification Network

We design a concise CNN network to extract features from raw EEG data. The feature exactor (denoted as F) contains one convolution layer, followed by a batch normalization layer and a dropout layer. The convolution layer with kernel $(m, n/p)$ and stride $(1, n/p)$ is equivalent to p no overlap time windows and can extract features of each time window from spatial and time dimension simultaneously. The architecture of the feature extractor is shown in Fig. 3.

A classifier is used to classify the features learned by the feature extractor. The classifier (denoted as C) is made up of two fully connected layers and the activation functions are ReLU. The output layer contains one neuron with sigmoid function. The loss function we used for classifier is binary cross entropy, and the loss function for basic classification task is as follow:

$$L_{task} = -E_{x,y\sim X,Y} \sum_{y\in Y} y * log(\sigma(C(F(x)))) \tag{1}$$

C. Conditional Adversarial Domain Adaptation

We bring up a conditional adversarial domain adaptation (CADA) network to narrow the difference of features distribution between different domains, and thus the target domain can utilize the knowledge of the source domain. A discriminator (denoted as D, as shown in Fig. 3) consists of two fully connected layers (512 and 256 neurons) with batch normalization and one output layer. The adversarial training is conduct between the feature extractor and the discriminator, which is a minimax optimization problem. The feature extractor learns common features of the source and target domain to fool the discriminator, while the discriminator tries to distinguish whether the feature belongs to the source or target domain.

Firstly, we use Earth-Mover distance (EMD) in Wasserstein GANs [43] to measure the divergence between features of the source and target domain. The EMD can provide meaningful gradients for optimizing the feature extractor, which ensures the convergence of the adversarial network. Then, to improve the network performance, the class information is used to guide the training of feature extractor and discriminator. The label of each sample y is element-wise added to the feature extracted by feature extractor $F(x)$, and constructs a new conditional feature as the input of discriminator.

To further improve the stability and ensure fast convergence in the process of network training, the gradient penalty proposed in [44] is adopted in our network. The loss function of conditional adversarial network:

$$\begin{aligned} L_{feature} = {} & E_{x^S,y^S\sim X^S,Y^S}(D(F(x^S) \oplus y^S)) \\ & - E_{x^T,y^T\sim X^T,Y^T}(D(F(x^T) \oplus y^T)) \\ & + \lambda_{gp} E((\left\| \nabla D(\hat{X}) \right\|_2 - 1)^2) \end{aligned} \tag{2}$$

The λ_{gp} is a hyperparameter that controls the trade-off of the original objective and gradient penalty, the $\hat{x}$ is data randomly sampled from source feature and target feature:

$$\begin{aligned} & \hat{x} = \delta(F(x^S) \oplus y^S) + (1-\delta)(F(x^T) \oplus y^T) \\ & \delta \sim U[0,1]; x^S, y^S, x^T, y^T \sim X^S, Y^S, X^T, Y^T \end{aligned} \tag{3}$$

D. Correlation Metric Learning

We further improve the domain adaptation by reducing the inter-class correlation and improving the intra-class correlation of features. With the idea of marginal fisher analysis criterion [45], we propose a metric learning loss function based on the correlation coefficient.

The overall CML loss function for the proposed CADA network can be stated by the following formula:

$$L_{CML} = \alpha * L_{intra} + \beta * L_{inter} \tag{4}$$

where α, β are parameters which weights the importance of intra-class compactness and inter-class separability.

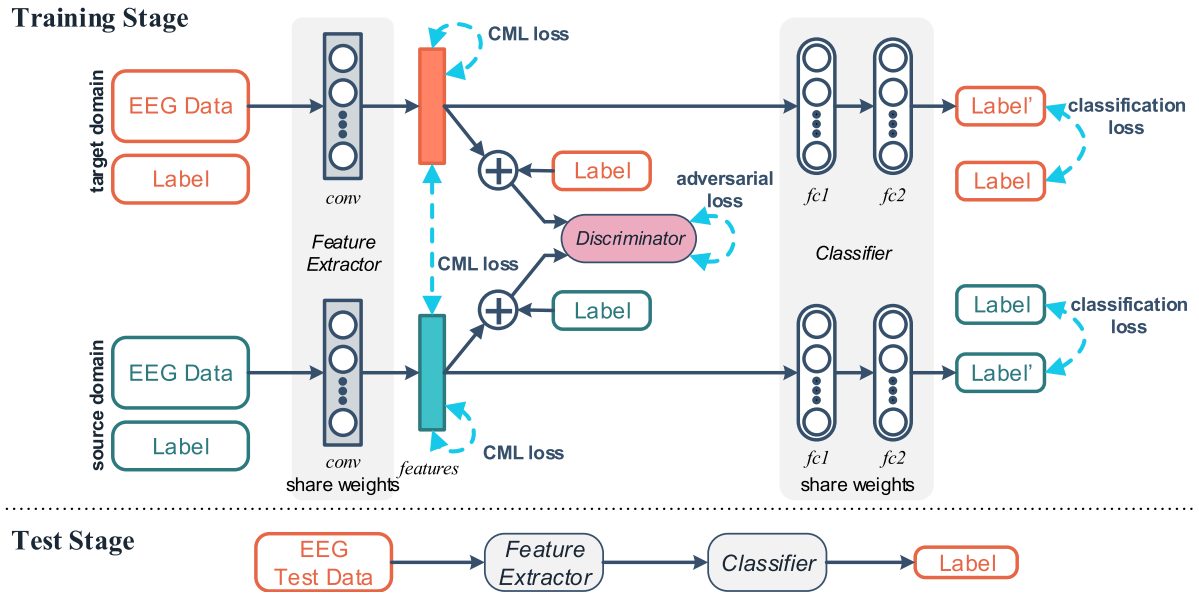


Fig. 3. The structure of the proposed CADA network. The network is made up of a feature extractor F, a classifier C, and a discriminator D. In the training session, the target domain X^T and source domain X^S are feed into F together. The C is training by the $F(X)$ which are features extracted by the F to predict labels. In addition, by adding labels (Y^T, Y^S) to $F(X)$ form the conditional features $F(X) \mid Y$, the D classifies the domain of $F(X) \mid Y$. The adversarial training of the F and the D, force F to extract common features of both domains to fool D. The CML loss further constrains the $F(X)$ depending on the class and domain for maximizes the intra-class correlation and minimizes the inter-class correlation. In the test session, the F extract the features of test data X_U^T, and the C predicts the label of X_U^T.

By minimizing the first term L_{intra}, we can achieve the maximum correlation of centralized features of same class in the source and target domain and increase the similarity between domains. The intra-loss is calculated as follows:

$$L_{intra} = \sum_{y^S=y^T x^S,x^T\sim X^S,X^T} [E \mid \rho(F(x^S), F(x^T)) - 1 \mid + \mid \rho(\overline{E(F(x^S))}, \overline{E(F(x^T))}) - 1 \mid] \quad (5)$$

The $\rho(\cdot,\cdot)$ in (5) is the correlation coefficient operation, the correlation coefficient is range from -1 to +1. $\overline{E(F(x^S))}$ and $\overline{E(F(x^T))}$ are centralized domain mean, calculated by minus the mean of all features in the source or target domain in the last training epoch, which is global structure information.

$$\begin{cases} \overline{E(F(x^S))} = E(F(x^S)) - E(F(X^S)) \\ \overline{E(F(x^T))} = E(F(x^T)) - E(F(X^T)) \end{cases} \quad (6)$$

In (5), the first item deals with sample level correlation, and the second introduces global information to calculate the correlation of category centers.

By minimizing the second term L_{inter}, we can reduce the correlation between classes and increase the differentiation of categories. The inter-loss is calculated as follows:

$$L_{inter} = \sum_{y_i\neq y_j (x_i,x_j)\sim X} [E \mid \rho(F(x_i), F(x_j)) \mid + \mid \rho(\overline{E(F(x_i))}, \overline{E(F(x_j))}) \mid] \quad (7)$$

$\overline{E(F(x_i))}$ and $\overline{E(F(x_j))}$ are centralized class mean, calculated by minus the corresponding domain mean of all features in the last training epoch, which is global structure information. The reduction of inter-class correlation loss also proceeds on sample level and class mean level, by minimizing the absolute value of the correlation coefficient of different classes.

In order to reduce over-fitting, L-2 regularization is added to the objective function of network optimization. Taken together, the overall CADA model train then corresponds to solve the optimization problem:

$$\min_{\theta_F,\theta_C} \max_{\theta_D} L(X,Y) = L_{feature} + L_{CML} + L_{task} + \lambda_1 \|\theta_F + \theta_C\|_2^2 + \lambda_2 \|\theta_D\|_2^2 \quad (8)$$

V. Experiments

A. Dataset Setup and Parameters

In our dataset, each session of each subject has 10 blocks of data. Block was the smallest unit of data set partition. For the subjects selected as sources, all 10 blocks of data were used. For the target subjects, when b blocks were selected as calibration (training) data and the remaining $10-b$ blocks were used as the test data. To analyze the effect of the amount of calibration data on classification performance, we chose different numbers of calibration blocks from 1 to 5 for further research. Five-fold cross-validation was adopted to divide the data set. In the training, we down-sampled the non-target samples on each subject to balance the two classes of samples, to reduce the influence induced by the extreme imbalance of two classes. The test set in the target domain maintains the original proportion of two classes.

The time window number p in the feature extractor was 25. During the training session, the network was optimized with Adam optimizer [46] with a learning rate of 0.0002, and

beta1 and beta2 of Adam were set as 0.5 and 0.999 respectively. Weight decay $\lambda_1 = 0.01, \lambda_2 = 0.001$. The feature extractor and classifier were optimized together, and the discriminator was optimized separately. The batch size was set to 10. The hyper-parameters in the loss function, λ_{gp}, α, and β were set as 1.0, 10.0, and 1.0, respectively.

B. Performance Evaluation

We evaluated our framework from two perspectives: comparison with other methods and ablation study.

We compared our method with the most recently published approaches, which include subject-dependent methods and transfer learning methods. For subject-dependent methods that train models based on data from the target domain only, HDCA [16], MDRM [21], MCNN [22], OLCNN [31], and EEG-Net [23] were re-implemented to compare with our approach. At the same time, we compared our method with transfer learning methods: RPA [27], SAN [41], and Li *et al.* [36]. For the single-source transfer learning methods RPA and Li, we applied the same multi-source vote framework to make a fair comparison. SAN is a multi-source transfer learning method, thus, we re-implemented it. Parameters of all compared methods were set according to their published paper. In addition, for our feature extraction and classification network, we designed three baseline contrast methods: Baseline of Direct Cross-domain (DC), Baseline of Calibration, and Baseline without Domain Adaptation.

- HDCA: a linear discrimination method that learns weights on the channel and time window of EEG.
- MDRM: a Riemannian geometry classifier, classify samples according to the geodesic distance to the center of the category. The method is re-implemented based on the Python package pyRiemann.
- MCNN: a three layers CNN proposed by Manor, we re-implemented the network under the description in [22].
- OLCNN: a one-layer CNN, we re-implemented the network under the description in [31].
- EEG-Net: a multi-layer CNN with depth-wise and separable convolution layers. To fit the model designed by the author, we resample our data to 128 Hz. The method was implemented according to https://github.com/vlawhern/arl-eegmodels.
- RPA: a transfer learning method based on the MDRM classifier. This method was implemented according to https://github.com/plcrodrigues/RPA.
- SAN: an adversarial transfer learning network. This method was implemented according to https://github.com/mingyr/san.
- Li: an unsupervised domain adaptation network. This method was re-implemented according to the code provided by author Li. We down-sampled EEG to match input and train the model in a supervised way.
- Baseline DC: The basic network of our framework is made up with the feature extractor and classifier. Our basic network is trained by the source domain data only.
- Baseline Calibration: Our basic network is trained by the calibration data in the target domain only.
- Baseline w/o DA: our basic network is trained by the calibration data in the target domain and one source domain data without DA, and apply the same multi-source vote framework.

To investigate the effectiveness of our proposed mCADA-C framework, and study the improvement of mCADA and the CML loss function, we conducted an ablation study on three models: Baseline w/o DA, mCADA, and mCADA-C.

RSVP image retrieval is a binary classification problem, with an extreme class imbalance in the test set. To reduce the impact of class imbalance when evaluating classification performance, we mainly use and analyze the balanced-accuracy (BA) to get a more comprehensive evaluation. We also provide other metrics for discussion, true positive rate (TPR), false positive rate (FPR).

$$\begin{cases} BA = \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \Big/ 2 \\ TPR = \frac{TP}{TP+FN} \\ FPR = \frac{FP}{TN+FP} \end{cases} \tag{9}$$

For target and non-target samples that are correctly classified is true positive (TP) and true negative (TN), otherwise are false negative (FN) and false positive (FP).

Two-way repeated measures ANOVA was employed to analyze the influence of different factors on the RSVP classification performance. The Greenhouse-Geisser correction was used when the Sphericity assumptions were not met. The Holm-Bonferroni correction was used to correct the p-value in pairwise comparison. Bland Altman analysis was employed to perform agreement analysis. The significant level was set at 0.05. All results were presented as mean ± standard deviation.

VI. Results

A. Analysis of ERPs

Fig. 4 shows the grand-average of target and non-target ERP responses, from Fz, Cz, Pz, and Oz channels obtained in our RSVP experiments. In Fig. 4, all channels were referenced to the average of the left and right mastoids. For the target ERP, we can see the components of N200 and P300 clearly, while it is mainly the harmonic response caused by the 10Hz picture flicker for the non-target ERP.

B. Source Selection Strategy

Fig. 5 shows the balanced-accuracy changes with different source numbers under different amount of calibration data. The two-way repeated measures ANOVA revealed that a significant effect of source number as well as a significant effect of calibration data amount (source number: $F(1.579, 15.789) = 69.045$, $p < 1e\text{-}7$; calibration data amount: $F(1.633, 16.326) = 20.686$, $p = 7e\text{-}6$). There was no interaction between these two factors ($F(3.33, 33.30) = 2.392$, $p = 0.08$). For different amounts of calibration block, the BA values under the data of 1 are 2 sources were significantly lower than those under

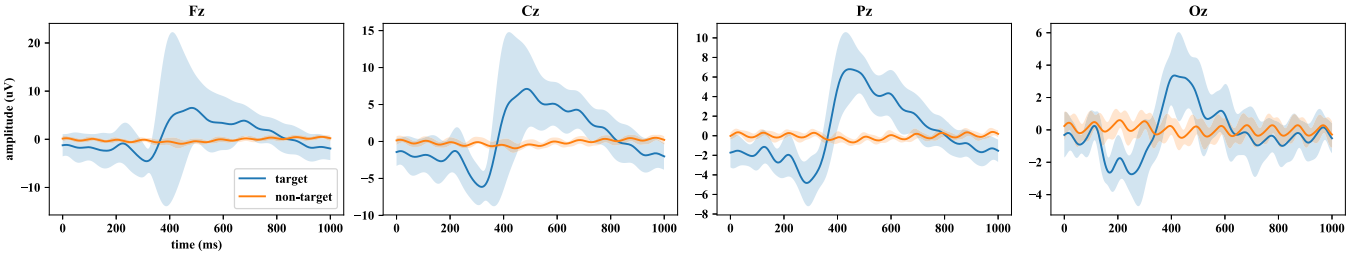


Fig. 4. Grand-Average target and non-target ERP responses. The responses from channel Fz, Cz, Pz, and Oz are shown from left to right. Shaded regions represent standard deviation. X-axis shows time in ms, and y-axis represents signal amplitude in μV.

TABLE II
COMPARISON OF BALANCED-ACCURACY (%) OF DIFFERENT METHODS

Methods	Different calibration data amounts (blocks)				
	1	2	3	4	5
HDCA[16]	67.41±9.38	73.53±10.30	76.22±9.92	77.62±9.81	78.55±9.72
MDRM[21]	78.46±9.82	81.15±8.67	82.16±7.88	82.54±7.46	82.90±6.71
MCNN[22]	76.71±8.72	80.85±7.54	82.64±6.92	83.83±6.91	84.77±6.54
OLCNN[31]	73.44±4.25	77.16±3.52	79.98±3.36	81.84±3.28	83.12±2.85
EEG-Net[23]	78.89±5.80	83.96±5.86	86.01±5.07	87.27±4.93	88.27±4.72
SAN[41]	65.50±9.71	69.28±8.02	69.48±7.91	69.44±7.83	69.23±7.85
Li[36]	71.78±7.51	75.41±7.78	77.48±7.63	79.06±7.40	80.01±7.01
RPA[27]	81.75±8.40	83.70±6.83	84.81±6.14	84.98±5.88	85.11±5.55
Baseline DC (source only)	66.94±6.72	66.92±6.81	67.02±6.83	66.93±6.80	67.00±6.76
Baseline Calibration (target only)	79.13±7.76	82.44±6.61	84.50±6.17	85.78±5.56	86.23±5.07
Baseline w/o DA	83.66±7.10	85.73±5.62	87.02±5.03	87.83±4.54	87.99±4.40
mCADA-C	**87.72±5.58**	**89.31±4.91**	**90.18±4.22**	**90.74±4.00**	**91.28±3.51**

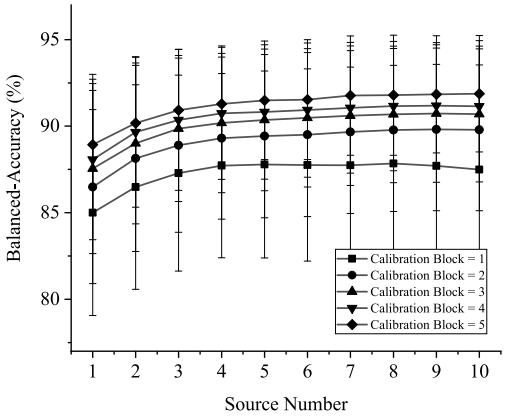


Fig. 5. The balanced-accuracy changes with the source number under different calibration data amounts.

more sources (all: $p < 0.02$). The BA value under 3 sources was significantly lower those with 5-10, and tends to be significantly lower than that under 4 sources ($p = 0.068$). The rising BA values reaches a plateau when the source number is 4. The BA values with 8 sources and 9 sources were significantly higher than those with 4 sources and 5 sources (Source numbers: 8 vs. 4, $p = 0.029$; 9 vs. 4, $p = 0.013$; 8 vs. 5, $p = 0.010$; 9 vs. 5, $p = 0.006$), but the difference values were less than 0.5%. Except that, the BA values have no significant difference between each other (all: $p > 0.1$). Therefore, we select 4 sources for transfer learning experiments based on our dataset.

C. Cross-Subject Domain Adaptation Under Different Amount Calibration Data

Table II summaries the comparison results of different methods. The two-way repeated measures ANOVA revealed there were the significant main effects of method and calibration data amount, as well as an interaction effect between these factors on classification accuracy, $F(4.593, 45.93) = 20.962$, $p < 2e\text{-}10$. And, under each amount of calibration data, the methods have significant influence on classification performance (all: $p < 6e\text{-}12$). Post-hoc tests showed that our method was significantly higher than all other compared methods (all: $p < 0.015$). Therefore, our method can improve the RSVP classification performance in cross-subject study.

Among the comparison results of different subject-dependent methods, EEG-Net outperformed other methods on RSVP dataset. Post-hoc tests showed that our baseline method was not different from the EEG-Net. Thus, for RSVP EEG, our proposed feature extractor can effectively extract features, and the classification performance is to that of EEG-Net.

Furthermore, we compare our mCADA-C method under 1 calibration block with that of EEG-Net and RPA methods to evaluate the reduced calibration data amount applied in our method. BA of our method at 1 calibration block was significantly higher than BA of EEG-Net at 1 to 3 blocks of calibration data ($p = 1e\text{-}5$, $p = 3e\text{-}4$, and $p = 0.01$, respectively). But it was not significantly different from that of EEG-net at 4 and 5 blocks. Moreover, this BA value was significantly higher than that of RPA at 1-5 blocks of

TABLE III
BALANCED-ACCURACY (%) OF ABLATION EXPERIMENTS

Methods	Different calibration data amounts (blocks)				
	1	2	3	4	5
Baseline w/o DA	83.66±7.10	85.73±5.62	87.02±5.03	87.83±4.54	87.99±4.40
mCADA	86.14±6.13	87.89±5.42	88.79±4.83	89.60±4.69	89.97±4.47
mCADA-C	**87.72±5.58**	**89.31±4.91**	**90.18±4.22**	**90.74±4.00**	**91.28±3.51**

calibration (all: $p < 0.005$). Moreover, Bland Altman analysis showed that our mCADA-C with 1 calibration block was agree with the EEG-Net with 4 and 5 calibration blocks ($p = 0.344$, $p = 0.340$). Therefore, our mCADA-C with 1 block calibration data can achieve similar classification performance with the EEG-Net and RPA with 4 or 5 calibration blocks.

For detailed performance (ACCuracy, TPR, FPR, BA, AUC, and ROC curve) of our method on each subject under 1 calibration block see Supplementary Material Fig. A. and Table A. Table II with extended metrics (ACC, TPR, FPR, and AUC) can be found in Supplementary Material Table B.

D. Ablation Experiments

To evaluate the effectiveness of the proposed CADA network and CML loss, we conducted ablation experiments. The comparison results are listed in Table III. The two-way repeated measures ANOVA revealed that the two factors (model and number of calibration data) showed significantly main effect on classification performance. The interaction between the two factors ($F(3.365, 33.65) = 1.802$, $p = 0.16$) was not significant. The main effect analysis showed that the estimated boundary mean BA of mCADA is significantly higher than baseline w/o DA by 1.8% ($p = 0.002$). The performance improvement indicates that the CADA is capable of matching the feature distribution of different subjects, thus increasing classification accuracy.

Moreover, compared with mCADA, mCADA-C can significantly improve the estimated boundary mean BA by 1.4% ($p = 0.003$). Thus, this proves the necessity of the CML. Therefore, CADA and CML can improve the classification performance of cross-subject learning.

E. Visualization

Fig. 6 shows waveforms of P300 waves of Pz and the topographies of EEG data before the feature extraction and waveforms and the topographies of features extracted by the feature extractor from one target subject and the corresponding source subject. All data in source domain and test data in target domain under 1 calibration block are used.

As shown in Fig. 6 (a), (c), the ERP waveform of the source domain and target domain are different. The P300 amplitude of the target subject is higher than that of the source subjects, and the latency is also higher. The pattern of the scalp map also varies from different domains and classes. For the target domain, the topography patterns of the two classes are similar. In Fig. 6 (b), (d), we can see similar ERP patterns and the topography patterns for the features from different domains with same classes. And, the topography

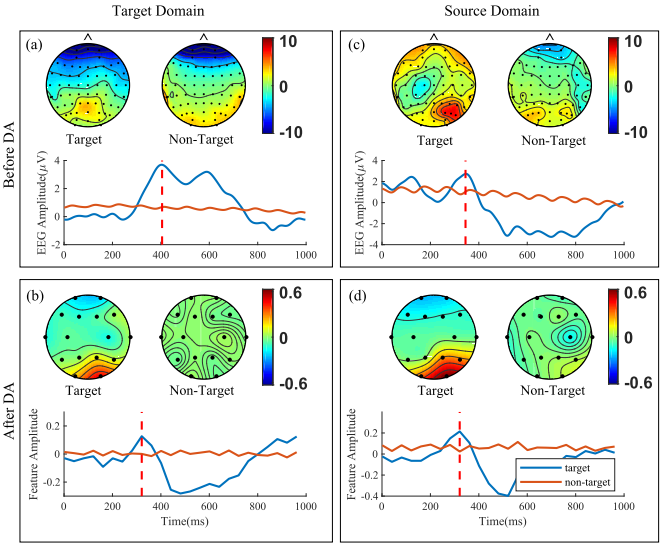


Fig. 6. Waveforms and topographies of P300 waves (Pz) for target subject (a) and source subject (c) before DA, and waveforms and topographies of feature extracted by the feature exactor for target subject (b) and source subject (d) after DA.

patterns for the features from target domain are obviously different between of the target samples and non-target samples. And, the topography patterns of target samples from different domains are much more similar. The result of Oz channel see Supplementary Material Fig. B. Therefore, the designed feature extractor learned the ERP patterns of EEG data, and the constrains provided by CML loss, discriminator, and classifier increase the inter-class differences and intra-class similarity of features.

We further applied t-distributed Stochastic Neighbor Embedding (t-SNE) to project the features in our CADA network into 2 dimensions. Fig. 7 shows an example of visualization under 1 calibration block situation. For one subject as a target domain and the corresponding first one of corresponding selected sources as source domain. The source domain data and class balanced target domain test data are employed as input of trained CADA model for visualization.

As shown in Fig. 7 (a), either source domain and target domain or target and non-target are maximum overlap for raw data. After extracting feature thought feature extractor, the distribution of features of source domain and target domain is more consistent. For both target domain and source domain, the target class and the non-target class more separated but still cannot be linearly separated, which can be seen in Fig. 7 (b). As the features are processed over the classifier, the features

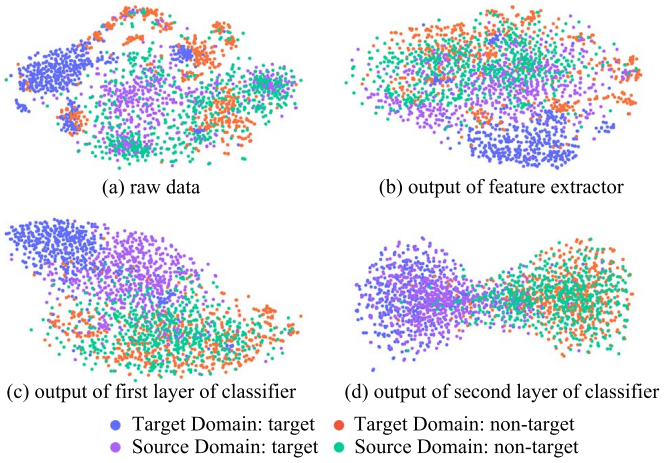


Fig. 7. An example of visualization: (a) raw data, (b) output by the feature extractor network F, (c) the output of first layer of classifier, (d) the output of second layer of classifier.

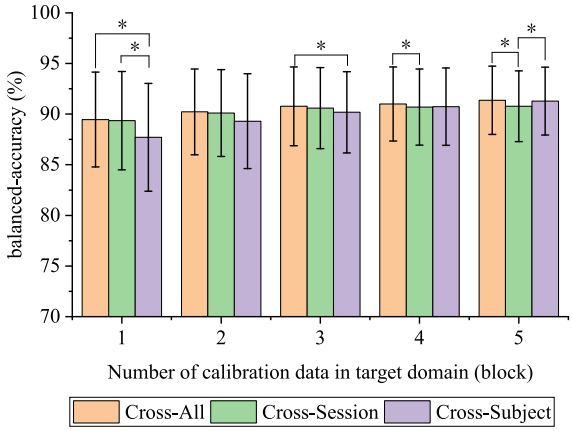


Fig. 8. BA of different cross-domain pattern, the * sign indicates that there is statistical significance between two groups ($p < 0.05$).

from different classes become more and more separable, and the samples of the same class samples from different domains is more similar. For the output of the second layer of the classifier, the target and non-target samples in both domains are linearly separable in Fig. 7 (d).

Thus, the visibility of separation of target and non-target increases from raw data to the output of the network, the feature distribution of samples in the same class is more similar and different classes are easier to classify.

F. Comparison of Different Cross-Domain Modes

Finally, we performed extensive research on the comparison of three different kinds of cross-domain mode: cross-subject, cross-session, and cross-all. The cross-subject and cross-session explained before. The cross-all mode in our experiments indicate for target subjects, their source selection range includes the labeled data from themselves in other sessions and other subjects at all experiment sessions.

Fig. 8 shows the comparison results of three conditions (cross-subject, cross-session and cross-all). The two-way repeated measures ANOVA showed a significant interaction between cross-domain mode and the amount of calibration data ($F(3.066, 30.659) = 9.387$, $p = 1e\text{-}4$). Post-hoc analysis revealed that the average BA have an increasing trend with the increase of calibration data in each cross mode ($p = 0.001$). Under 1 calibration block and 2 calibration blocks, the BA values of cross-subject was significantly lower than that of cross-all and cross-session ($p = 0.005$ and $p = 0.01$), but the difference between cross-all and cross-session was not significant. Under 3 calibration blocks, the BA of cross-subject was significantly lower than that of cross-all ($p = 0.0015$), and tends to be lower than that of cross-session ($p = 0.09$). It is noteworthy that cross-session tends to be lower than that of cross-all ($p = 0.1$). When the calibration block is 4 and 5, the cross-subject was not different with cross-all ($p = 0.65$).

The results indicated that when the calibration data is few, the data from other sessions contribute much more improvement on classification performance than data from other subjects. With the increase of calibration data, the data provides by other subjects can greatly compensate for the data defect of single target subject.

TABLE IV
VERIFICATION OF DIFFERENT SOURCE SELECTIONS

Different sources combination	BA (%)
Our source select strategy	87.72±5.32
Max performance	88.43±5.79
Min performance	84.76±6.69
Average performance	86.76±6.24
10 sources	87.49±6.05

VII. DISCUSSION

A. Source Selection Analysis

In our study, with the increase in the number of source domains, the performance of RSVP decoding increases, and then arrives a plateau in the condition of 4 sources. This phenomenon is in consistent with the previous study [47]. For our dataset, there are C_{10}^{4} combinations when we select 4 sources. We conduct a further study on source selection under 1 calibration block to test the validity of the source selection strategy. Table IV lists the performance of our source selection strategy, maximum, minimum and average performance in those C_{10}^{4} results and all 10 sources used performance. The one-way repeated measures ANOVA revealed the combination of sources can significantly influence the BA ($F(4, 40) = 25.806$, $p = 6e\text{-}6$). The performance of our source selection strategy is significantly higher than the minimum performance ($p = 0.002$) and not different from the maximum, average and 10 sources performance ($p = 0.297$, $p = 0.297$, and $p = 0.481$, respectively). The results indicate that our source selection strategy can reduce the impact of negative transfer and the computational complexity of using all sources.

It is worth noting that the mean value of our method is slightly lower than that of the optimal result. It may be due to that the distribution difference between labeled data and test data in the target domain result in the suboptimal performance in the testing with the optimal sources in the training.

The mCADA-C framework was implemented with the Pytorch framework and trained on one Nvidia 1080 GPU. The training stage contains source selection and 4 CADA network training. Under 1 calibration block, the average time consuming of the training stage is 3.2±1.7 minutes. On the one hand, the time consuming of the training stage ranged from 1.4 to 4.8 minutes, which is shorter than the calibration time of 1 block with rest (5 to 6 minutes). On the other hand, the performance of our method under 1 calibration block is significantly higher than EEG-Net and RPA under 2 calibration blocks. Hence, the time-consuming is acceptable and our framework is feasible. Furthermore, with more hardware support, the training of different CADA networks can be deployed to multiple GPUs to further reduce the training time. Based on the trained framework, we conduct a simulation online testing, the test time of 1block data is 0.5978±0.06 second, and that of 1 trial data is 0.1630±0.01 second.

B. Analysis of the Comparison Results

Comparison experiments showed that our proposed mCADA-C framework outperformed the other methods, and produced a decrease in the calibration block. Compared with EEG-Net, a deeper network, our framework utilized data of sources effectively and have a higher classification performance, this mainly depends on two aspects. Our framework enables feature extractor to learn common features of source and target by adversarial training. Also, CML loss function constricts the correlation of features based on class and domain and improves the consistency and separability of features in the source domain and target domain. In our framework, the feature extractor was designed to extract time information from EEG input, while previous comparable transfer learning methods (RPA, SAN, and Li) ignore the time characteristics of EEG. Therefore, our method outperformed previous methods.

In our framework, we built 4 one-to-one CADA-C models and integrate their classification results, which is different from the one many-to-one model in SAN. In one-to-one DA framework, the model needs to find common features from one source domain and the target domain, while the model needs to find common features from several different domains directly in many-to-one DA framework. Hence, the many-to-one framework more likely to cause the overfitting of training data and be difficult to training models. In addition, the integration of multiple one-to-one DA networks, can improve the classification of RSVP. Therefore, our multi-source strategy is valuable for decoding RSVP.

Generally, EEG preprocessing has an impact on the follow-up analysis. To evaluate this impact, we conduct experiments on three compared methods designed for RSVP task: HDCA, MCNN, and EEG-Net, to compare the performance under our preprocessing and their original preprocessing. Compared with our preprocessing, the original preprocessing of EEG-Net has no improvement. For HDCA and MCNN, the original preprocessing can improve the BA by 3.1% and 0.7% compared to the present preprocessing. However, the improved performance is still lower than our method, for more details of the results see Supplementary Material Table C.

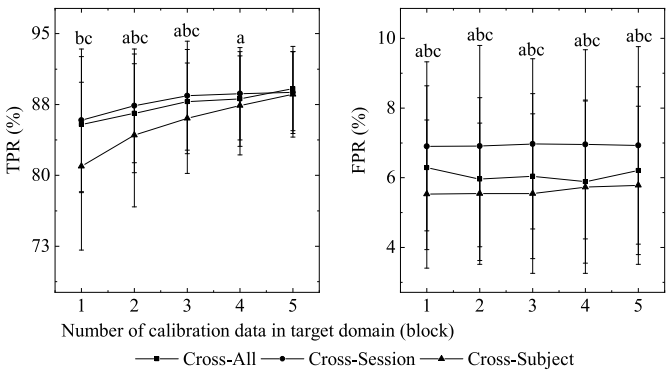


Fig. 9. TPR and FPR of different cross-domain mode. The a, b, c sign indicates that there is statistical significance between cross-all and cross-session, cross-session and cross-subject, cross-all and cross-subject.

TABLE V
THE RESULTS OF SOURCE SELECTION UNDER CROSS-ALL MODE

Target Sub/Sess*	Source 1 Sub/Sess	Source 2 Sub/Sess	Source 3 Sub/Sess	Source 4 Sub/Sess
1/1	**1**/2	**1**/3	5/1	11/1
2/1	**2**/2	**2**/3	10/3	8/3
3/1	**2**/2	**2**/3	11/3	11/2
4/1	**4**/3	**4**/2	5/1	2/3
5/1	**5**/3	**5**/2	7/3	9/2
6/1	**6**/3	**6**/2	4/1	4/3
7/1	**7**/3	**7**/2	9/2	8/1
8/1	**8**/3	**8**/2	2/2	11/1
9/1	**9**/2	**9**/3	5/2	7/2
10/1	**10**/2	5/1	**10**/3	7/2
11/1	**11**/3	**11**/2	8/1	10/3

* Sub and Sess are abbreviations of subject and session respectively.

C. Comparison Among Different Cross-Domain Modes

Fig. 9 shows TPR and FPR of different cross-domain modes. The two-way repeated measures ANOVA showed that there are significant main effects of cross-domain mode and calibration data amount and an interaction effect between these factors on classification TPR ($F(2.899, 28.994) = 8.476$, $p = 4e\text{-}4$). For FPR, these factors do not have an interaction effect and the calibration data amount has no significant main effects on FPR ($F(2.057, 20.568) = 0.457$, $p = 0.645$).

Under 1-5 calibration blocks, the FPR of cross-subject is significantly lower than the others (all: $p < 0.005$), and cross-all is significantly lower than cross-session ($p = 5e\text{-}5$). The BA results are combined to analyze the effect of mode. For 1 calibration block, TPR and BA of cross-all mode are similar with cross-session and higher than cross-subject, while lower FPR indicates that cross-all mode is best. For 2 and 3 calibration blocks, compared with cross-session, TPR of cross-all session shows a weak decrease. Considering that the non-target class samples are far more than the target class in this RSVP task, the cross-all with lower FPR is better for efficiency of the BCI system.

Due to individual differences in EEG, the cross-subject DA may learn fewer common features from source subject and target subject in training. Thus, during the testing of target subject, this may lead to identify target samples more conservatively, as well as reduce FPR greatly. Table V summaries

the selected sources in cross-all mode, the source selection strategy automatically includes the data of two other sessions of the same subject. Thus, besides cross-session, the cross-all consists of not only information from different subjects but also information from other session of the same subject. Therefore, to reduce calibration, the cross-all mode includes data of both other sessions and other subjects is a better choice for the application of RSVP-based BCI.

An individual's neural activity pattern is unique [49], which makes the individual difference of EEG. For one individual, the EEG pattern changes with physiological, environmental, and instrument (e.g., electrode positioning and impedance). While the RSVP paradigm can evoke a strong and stable individual response [18]. Hence, the difference of EEG between sessions within subjects was smaller than that between subjects. The results of our different cross-mode experiments are consistent with this characteristic. This also makes using different session data for zero-training possible.

D. Towards Zero-Training

Zero-training is an important challenge for the practical application of BCI. For the evoked BCI paradigm, the data from other users may be helpful for realizing zero-training BCI [50]. In our experiments, the 'Baseline DC (source only)' results in Table II. can be considered as zero-training for new subjects, in which the performance is poor due to the individual difference. Hence, for our proposed mCADA-C framework, we further extend its application towards zero-training for subjects who have data from previous sessions. For an 'experienced' subject, the data from the previous session can be considered as the calibration data in the target domain, and the existing data from the other subjects can be considered as source domains. We conducted the experiments according to this strategy.

In our experiments, for a new session of a target subject, data of one existing session from the target subject was used as calibration data to perform source selection and domain adaptation. Data from other subjects were used as source domains. The average BA of zero-training is 86.54 $\pm$ 5.36%. The BA of zero-training is significantly lower than cross-session and cross-all mode with 1 block of calibration data by 2.8% and 2.9% ($p = 8e\text{-}5$ and $p = 2e\text{-}5$) and tends to be lower than cross-subject with 1 block of calibration data by 1.2% ($p=0.06$). This result indicates that indeed, there is a similarity of EEG across sessions. However, a difference of time delay and amplitude of P300 and mismatch in data distribution remain. Still, the BA of zero-training is significantly higher than baseline calibration (target only) with 1 to 3 blocks of calibration data (all: $p < 0.01$). In addition, it is significantly higher than that of baseline w/o DA under 1 block of calibration ($p = 0.004$), and has no significant difference with performance under 2-3 blocks of calibration ($p = 0.112$ and $p = 0.284$). These results show the potential for our proposed method to be extended towards zero-training when the subject has data from other sessions. Towards zero-training application, P300 may be induced by false alarm [10], [11] rather than a target event, which should be considered in application and future work.

E. Limitation and Future Work

Our work reduced calibration efforts in RSVP tasks through multi-source conditional adversarial domain adaptation. There are still several limitations. The number of participants in our study is 11. More subjects will be included in the experiments in the future. Also, our experiment only applied the typical RSVP speed of 10Hz. Different RSVP speeds will be considered in the future.

VIII. Conclusion

In this article, we proposed a multi-source conditional adversarial domain adaptation with the correlation metric learning (mCADA-C) framework for reducing the calibration efforts of RSVP-based BCIs. The multi-source framework with source selection strategy was employed for avoiding negative transfer and reducing computation. The adversarial domain adaptation network enables the target to utilize the knowledge of the source domain. The proposed CML loss maximizes the intra-class correlation and minimizes the inter-class correlation. The experiment results of cross-subject learning proved that our mCADA-C framework outperformed the previous transfer learning methods for EEG analysis as well as conventional subject-dependent approaches. This work could be regarded as a powerful method to improve the performance of RSVP-based applications under small calibration data. We further studied different cross-domain modes and towards zero-training, and we drew the conclusion that combining cross-session and cross-subject data can further improve performance and our framework is feasible for the zero-training of 'experienced' subjects.

References

[1] J. R. Wolpaw, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791. 2002.
[2] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Inform. J.*, vol. 16, pp. 213–230, Jul. 2015.
[3] K. Choi and A. Cichocki, "Control of a wheelchair by motor imagery in real time," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, Berlin, Germany, 2008, pp. 330–337.
[4] C. Chen *et al.*, "G-causality brain connectivity differences of finger movements between motor execution and motor imagery," *J. Healthcare Eng.*, vol. 2019, pp. 1–12, Oct. 2019.
[5] A. N. Belkacem, S. Nishio, T. Suzuki, H. Ishiguro, and M. Hirata, "Neuromagnetic decoding of simultaneous bilateral hand movements for multidimensional brain–machine interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 6, pp. 1301–1310, Jun. 2018.
[6] O. Friman, I. Volosyak, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 742–750, Apr. 2007.
[7] L.-W. Ko, R. K. Chikara, Y.-C. Lee, and W.-C. Lin, "Exploration of User's mental state changes during performing Brain–Computer interface," *Sensors*, vol. 20, no. 11, p. 3169, Jun. 2020.
[8] N. J. Beuchat, R. Chavarriaga, S. Degallier, and J. R. del Millán, "Offline decoding of upper limb muscle synergies from EEG slow cortical potentials," in *Proc. EMBC*, Jul. 2013, pp. 3594–3597.
[9] M. Xu, X. Xiao, Y. Wang, H. Qi, T.-P. Jung, and D. Ming, "A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1166–1175, Jan. 2018.

[10] R. K. Chikara and L.-W. Ko, "Neural activities classification of human inhibitory control using hierarchical model," *Sensors*, vol. 19, no. 17, p. 3791, Sep. 2019.

[11] R. K. Chikara, O. Komarov, and L.-W. Ko, "Neural signature of event-related N200 and P300 modulation in parietal lobe during human response inhibition," *Int. J. Comput. Biol. Drug Des.*, vol. 11, nos. 1–2, pp. 171–182. 2018.

[12] E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S.-F. Chang, and P. Sajda, "Closing the loop in cortically-coupled computer vision: A brain–computer interface for searching image databases," *J. Neural Eng.*, vol. 8, no. 3, May 2011, Art. no. 036025.

[13] C. Barngrover, A. Althoff, P. DeGuzman, and R. Kastner, "A brain–computer interface (BCI) for the detection of mine-like objects in sidescan sonar imagery," *IEEE J. Ocean. Eng.*, vol. 41, no. 1, pp. 123–138, Jan. 2016.

[14] L. Acqualagna and B. Blankertz, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 901–908, May 2013.

[15] Z. Lin, C. Zhang, Y. Zeng, L. Tong, and B. Yan, "A novel P300 BCI speller based on the triple RSVP paradigm," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, Feb. 2018.

[16] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[17] A. R. Marathe, V. J. Lawhern, D. Wu, D. Slayback, and B. J. Lance, "Improved neural signal classification in a rapid serial visual presentation task using active learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 333–343, Mar. 2016.

[18] Q. Wu, B. Yan, Y. Zeng, C. Zhang, and L. Tong, "Anti-deception: Reliable EEG-based biometrics with real-time capability from the neural response of face rapid serial visual presentation," *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 55, Dec. 2018.

[19] D.-O. Won, H.-J. Hwang, D.-M. Kim, K.-R. Müller, and S.-W. Lee, "Motion-based rapid serial visual presentation for gaze-independent brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 334–343, Feb. 2018.

[20] A. Barachant, "Riemannian geometry applied to BCI classification," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.*, 2010, pp. 629–636.

[21] A. Barachant and M. Congedo, "A plug & play P300 BCI using information geometry," 2014, *arXiv:1409.0107*. [Online]. Available: http://arxiv.org/abs/1409.0107

[22] R. Manor and A. B. Geva, "Convolutional neural network for multi-category rapid serial visual presentation BCI," *Frontiers Comput. Neurosci.*, vol. 9, p. 146, Dec. 2015.

[23] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013.

[24] N. R. Waytowich, V. J. Lawhern, A. W. Bohannon, K. R. Ball, and B. J. Lance, "Spectral transfer learning using information geometry for a user-independent brain-computer interface," *Frontiers Neurosci.*, vol. 10, p. 430, Sep. 2016.

[25] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[26] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.

[27] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for Brain–Computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2390–2401, Aug. 2019.

[28] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: A Riemannian geometry framework with applications to Brain–Computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.

[29] O. Yair, M. Ben-Chen, and R. Talmon, "Parallel transport on the cone manifold of SPD matrices for domain adaptation," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1797–1811, Apr. 2019.

[30] H. He and D. Wu, "Transfer learning for Brain–Computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.

[31] H. Shan, Y. Liu, and T. Stefanov, "A simple convolutional neural network for accurate P300 detection and character spelling in brain computer interface," in *Proc. IJCAI*, Jul. 2018, pp. 1604–1610.

[32] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.

[33] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[34] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.

[35] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. CVPR*, Jul. 2017, pp. 7167–7176.

[36] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.

[37] E. Jeon, W. Ko, and H.-I. Suk, "Domain adaptation with source selection for motor-imagery based BCI," in *Proc. 7th Int. Winter Conf. BCI*, Feb. 2019, pp. 1–4.

[38] X. Tang and X. Zhang, "Conditional adversarial domain adaptation neural network for motor imagery EEG decoding," *Entropy*, vol. 22, no. 1, p. 96, Jan. 2020.

[39] M. Jimenez-Guarneros and P. Gomez-Gil, "Custom domain adaptation: A new method for cross-subject, EEG-based cognitive load recognition," *IEEE Signal Process. Lett.*, vol. 27, pp. 750–754, 2020.

[40] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, May 2020.

[41] Y. Ming *et al.*, "Subject adaptation network for EEG data analysis," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105689.

[42] A. Torralba, *The MIT-CSAIL Database of Objects and Scenes*. Accessed: Sep. 15, 2020. [Online]. Available: http://web.mit.edu/torralba/www/database.html

[43] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[44] I. Gulrajani, "Improved training of Wasserstein GANs," in *Proc. NIPS*, 2017, pp. 5767–5777.

[45] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[47] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.

[48] S. Lees, "A review of rapid serial visual presentation-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 2, 2018, Art. no. 021001.

[49] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 738–742, Apr. 2007.

[50] J. Lee, K. Won, M. Kwon, S. C. Jun, and M. Ahn, "CNN with large data achieves true zero-training in online P300 brain-computer interface," *IEEE Access*, vol. 8, pp. 74385–74400, Apr. 2020.