A Transfer Learning Framework for RSVP-based Brain Computer Interface*

Wei Wei[1,2] Shuang Qiu[1] Xuelin Ma[1,2] Dan Li[1,2] Chuncheng Zhang[1] and Huiguang He[1,2,3,*]

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 paper, we collect an RSVP-based electroencephalogram (EEG) dataset, which includes 11 subjects. The experimental task is image retrieval. Also, we propose a multi-source transfer learning framework by utilizing data from other subjects to reduce the data requirement on the new subject for training the model. A source-selection strategy is firstly adopted to avoid negative transfer. And then, we propose a transfer learning network based on domain adversarial training. The convolutional neural network (CNN)-based network is designed to extract common features of EEG data from different subjects, while the discriminator tries to distinguish features from different subjects. In addition, a classifier is added for learning semantic information. Also, conditional information and gradient penalty are added to enable stable training of the adversarial network and improve performance. The experimental results demonstrate that our proposed method outperforms a series of state-of-the-art and baseline approaches.**

I. INTRODUCTION

Brain-Computer Interface (BCI) provides direct communication between a person's brain and an electronic device without the need for any muscle control [1]. EEG-based BCI is more widely used in research and application since EEG is measured non-invasively, more convenient and safe and has a high temporal resolution [2]. Rapid Serial Visual Presentation (RSVP) is a visual evoked paradigm based on the oddball paradigm and is one of many BCI paradigms. RSVP paradigm uses a high speed presented picture sequence containing a small number of target pictures to induce specific event-related potential (ERP) components, for example, the P300 components which have a positive potential that appears with a 300∼500 ms delay after the onset of the target stimulus. RSVP-based BCI has been used for many applications, for example, speller[3][4], image retrieval [5][6], anomaly detection [7], anti-deception [8], etc.

There are several classic and state-of-the-art methods in EEG analysis of RSVP tasks, including hierarchical discriminant component analysis (HDCA) [6], minimum distance to Riemannian mean (MDRM) [9], EEG-Net [10], etc. These methods can be used to train models and test within a subject. However, due to the individual differences of EEG, models are hard to generalize across subjects. Thus, RSVP-BCI still suffer from the requirement of time-consuming training and frequent calibration sessions. Meanwhile, the acquisition of EEG is very expensive, and a large number of labeled samples are scarce, which makes reducing calibration through transfer learning an area of increasing interest for the development of practical BCI systems.

In transfer learning, EEG from different subjects can be considered as different domains. In this study, we aim to develop a novel method of transfer learning, that can utilize the existing subject's data to reduce the calibration effort and improve the classification performance. We regard the new subject's data as the target domain, and the existing subject's data as the source domain. This form of transfer learning is referred to as cross-subject learning.

In the study of EEG transfer learning, based on the MDRM classifier, Zanini [11] promoted a Riemannian geometry transfer framework that works well on motor imagery (MI). Rodrigues [12] proposed a Riemannian Procrustes Analysis (RPA) method for improving this method. RPA can improve the classification of MI and steady state visual evoked potential (SSVEP) by 8% and 9% respectively, but there is no significant improvement in P300 task.

Recently, researchers used the idea of the generative adversarial network (GAN) [13] to conduct domain adaptive in transfer learning study [14][15]. In the EEG emotion recognition task, Li [16] promoted a domain adaptation model based on latent representation similarity. The model is optimized by minimizing the classification error of the source domain and improving the similarity of source and target domain latent representations at the same time. The model outperforms the state-of-the-art on unsupervised EEG emotion recognition. This model is based on the differential entropy feature of EEG, while the classical feature of EEG data in RSVP tasks is temporal. Thus, it is hard to employ this model directly in RSVP EEG transfer learning problem.

With the similar idea of GAN, the subject adaptation network (SAN) [17] learns features from several domains at the same time. The adversarial network is adopted to constrain the distribution of features to an artificial distri-

*This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 81701785, and Grant 61906188, and in part by the Strategic Priority Research Program of CAS under Grant XDB32040200, and in part by the China Postdoctoral Science Foundation 2019M650893.

[1]Research Center for Brain-inspired Intelligence, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing, China

[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

[3]Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Science, Beijing, China

*Corresponding author e-mail: huiguang.he@ia.ac.cn

bution which includes prior information on the proportion of classes. This method provides a new idea of distributed constraint based on the combination of adversarial network and prior knowledge. However, the negative transfer and the common strong constraint on multiple subjects would make the model difficult to train in the case of small samples.

At present, there are few researches on EEG transfer learning for RSVP paradigm. In this study, we proposed a multi-source transfer learning framework to reduce the calibration efforts and improve the classification performance in the RSVP task.

To deal with the negative transfer in a multi-source transfer learning problem, a source selection strategy was firstly proposed and adopted before transfer learning. A convolutional neural network (CNN)-based network was designed to extract features from EEG data that collectively leverages the spatial and temporal information of RSVP EEG signals. Then, a domain adversarial training network was proposed in our model, the discriminator was used to constrain the feature extractor to learn common features between different domains and adapt the marginal distributions of features. Also, we added the category information to the adversarial network to form a conditional adversarial network. Besides, a classifier was added to the network for feature extractor learning semantic information of EEG. Furthermore, a gradient penalty was induced for stability and convergence in training our model. Experimental results showed our proposed framework outperformed other compared methods and can achieve the goal of model training with less calibration.

The rest of the paper is structured as follows: First, the details of the BCI experiment we conducted will be illustrated in Section II. Section III introduces the transfer learning framework we developed. Section IV illustrates the details of the experiment's setup. Section V presents the results, analysis, and discussion. Finally, we conclude our work in Section VI.

II. MATERIALS

A. Subjects

The experiment included 11 participants (6 males and 5 females; aged 22.8 ± 2.1, 9 right-handed). All participants had corrected-to-normal or normal vision, and were naïve in respect RSVP-based BCI experiments.

All participants was voluntary, fully informed, and have written consent. This study was ethically approved by the Institutional Review Board of the Institute of Automation, Chinese Academy of Science, before the research.

B. RSVP Paradigm

Visual stimuli images are sized 500×500 pixels RGB images. All the photos are street scenes, the target images include pedestrians, shown in Fig. 1 (a), (b). The pictures are from the scenes and objects database [18]. The probability of target images is 4%. All images were randomly presented at a rate of 10Hz. Subjects need to find target images in the image sequence and they were asked to press the space bar on the keyboard as quickly as possible when they found the target. As shown in Fig. 1 (c) each experimental session contains 10 blocks, and each block includes 1400 pictures.

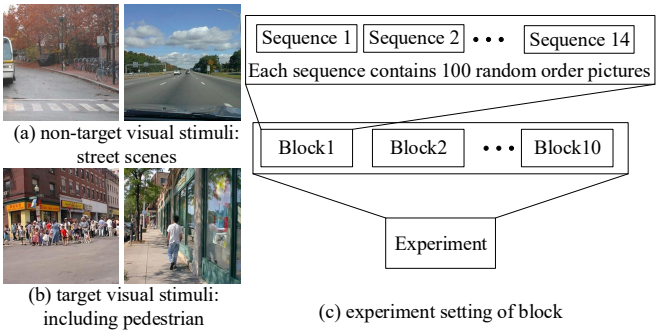


Fig. 1. (a) Examples of target pictures, (b) examples of non-target pictures and (c) settings of our RSVP experiment.

C. Data Acquisition and pre-Process

The EEG data were recorded using NeuroScan SynAmp2 system at a sampling rate of 1000 Hz from 64-channel Ag/AgCl electrodes cap according to the 10/20 international system. All 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, the signals were down-sampled to 250 Hz. Then, the signals were bandpass filtered between 0.1 and 15 Hz. Afterward, the EEG signals were divided into epochs. Each epoch contains EEG data from 0 to 1000 milliseconds after the stimulation onset. After epoch, there are 14,000 trials of EEG sample for each subject, 560 of them is target class the rest are non-target. The subsequent analysis and classification of EEG were all based on single-trial EEG data.

III. METHOD

We consider the application scenario of cross-subject learning is supervised domain adaptation and treat the data of each subject as a domain. The new subject is the target subject (domain) and existing subjects are source subjets (domains). The source data X_s and label Y_s from other Q subjects and we trained Q corresponding classifiers C_s. The target subject has training data X_t, label Y_t, and test data $X_t^{'}$. Our purpose is to construct a model using $\{(X_s^1, Y_s^1) \cup \cdots \cup (X_s^Q, Y_s^Q) \cup (X_t, Y_t)\}$ data to predict the label of $X_t^{'}$.

A. Multi-Source Transfer Learning Framework

Existing studies indicate that brute force leveraging of the sources poorly related to the target may decrease the performance [19]. To avoid negative transfer and reduce computation in multi-source transfer learning, we select appropriate sources first.

We use cross-domain classification performance to measure the similarity between different domain data. For the target subject, we use $C_s^i, i \in (1, \ldots, Q)$ to classify X_t and acquire Q classification results. Q subjects were ranked from high to low according to the classification accuracy, and the top K subjects were selected as the source domain

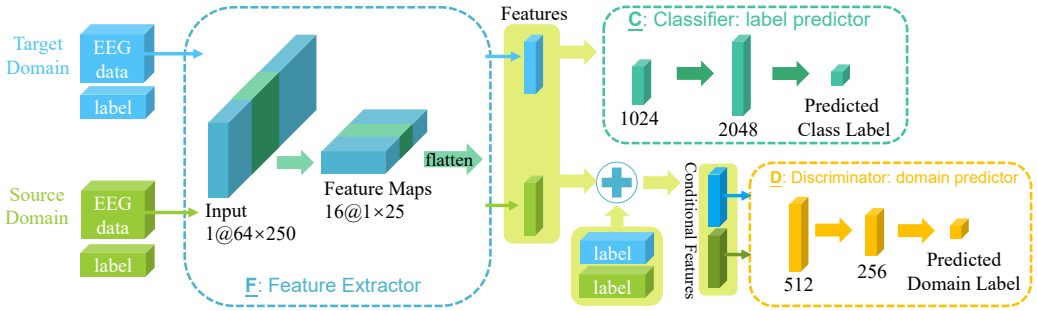


Fig. 2. Architecture of our transfer learning network: the network model is made up of a feature extractor F, a classifier C, and a discriminator D. The X_t and X_s^i are feed into one feature extrator F_i together. The extracted features $F_i(X)$ train the label predictor C_i. In addition, by adding labels (Y_t, Y_s^i) to $F_i(X)$ form the conditional features $F_i(X)|Y$, the domain of $F_i(X)|Y$ is classified by the doamin predictor D_i.

for subsequent transfer learning. The K is an important parameter in our model, the choice of K will be present in Section V.

After source selection, during the training session, each source domain $X_s^{K(i)}$ $(i \in 1,\ldots,K)$ is fed in a transfer learning network with target domain labeled data X_t. Our proposed transfer learning network shows in Fig. 2. Each transfer learning network learns the common feature representation of two domains and train the classifiation network. In the test session, $X_t^{'}$ will be classified by the K transfer learning network, then obtain K classification results. Finally, the average voting method is used to integrate the K results.

B. Basic Classification Network

The basic classification network, which consists of a feature extractor F and a classifier C. The feature extractor F projects EEG data from the original EEG space into an embedding feature space. The extracted feature can be used to classify through classifier C.

The features of RSVP-based EEG are temporal, while there are several kinds of neural network architecture designed for EEG analysis, usually involved with more than three convolution layers, max-pooling, depth-wise convolution, etc. [10][20] However, the temporal information is likely to be missed due to the strike and pooling operation. Besides, while the labeled data in the target domain is few, which is easily over-fitting for a complex model.

To avoid this problem, we designed a concise CNN-based feature extraction model for RSVP-based EEG, which is inspired by the HDCA method [6]. The HDCA uses a fisher linear discriminator to learn weights for each channel and time window. With back propagation, the nonlinear CNN network can learn a better kernel parameter for EEG classification.

The structure of network F is shown in Fig. 2. Network F principal makes up with only one convolution layer to extract features from spatial and time simultaneously. The network F's architecture is similar with [21] while our kernel size is (64, 10), batch normalization and dropout is used. The output of this layer has a shape of 16×25 and flatten to a 400×1 vector as output.

The classifier C is made up with fully connected layer. The two hidden layers contain 1024 and 2048 neurons respectively, and using ReLU as activation function, and dropout parameter for two hidden layer is 0.5, 0.2. The output layer contains one neuron with sigmoid function.

C. Transfer Learning Network

In our application scenario, there are similar visual evoked P300 characteristics between EEG of different domains while the distribution is different. With the adversarial network idea, and under transfer learning task, the network model we proposed is made up of a feature extractor F, a classifier C, and a discriminator D, the structure is shown in Fig. 2.

In the training session, the adversarial training of feature extractor F and discriminator D force the F extract common features to confuse the discriminator D, while the D tries to distinguish whether a sample belong to source or target domain. The classifier C forces feature extractor F to learn semantic information of EEG. In the test session, the trained F and C are used to classify $X_t^{'}$.

The adversarial in training is a minimax optimization problem. We adopted the Earth-Mover distance to measure the divergence between features of source and target domain, which overcome the problems of difficulty and instability of adversarial training [22].

Previous studies have shown that adding task-related additional information to network training can improve the network performance, such as cGAN [23], the additional information can be category information, other modal data and so on. Here we use the class information to guide the training of feature extractor F and discriminator D. The label of each sample is added to the feature extracted by feature extractor F to construct a new conditional feature as the input of discriminator D.

Instead of using weight clipping to constrain the gradient norm, the gradient penalty proposed in [24] is adopted in our network. The loss function of adversarial network:

$$\begin{aligned} L_{feature}(F,D,X_s,X_t) = {} & \mathbb{E}_{x_s \sim X_s, y_s \sim Y_s}[D(F(x_s)|y_s)] \\ & -\lambda_g \mathbb{E}_{x_t \sim X_t, y_t \sim Y_t}[D(F(x_t)|y_t)] \\ & +\lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (1)$$

λ_g is a hyper-parameter to balance the unequal information of the source and target domain, $\lambda_g = N_s/N_t$, where N_s, N_t

is the number of samples in the source and target domain. The λ 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:

$$\begin{aligned} \hat{x} &= \alpha(F(x_s)|y_s) + (1-\alpha)(F(x_t)|y_t) \\ \alpha &\sim U[0,1]; x_s, y_s, x_t, y_t \sim X_s, Y_s, X_t, Y_t \end{aligned} \tag{2}$$

During the training iteration, the classifier is added to the training, the loss function we used for the classification task is cross-entropy:

$$L_{task} = -\mathbb{E}_{(x,y)\sim(X,Y)} \sum_{k \in Y} 1_{[k=y]} log(\sigma(C(x))) \tag{3}$$

L_2 regularization was added to the objective function of network optimization to reduce over-fit. Finally, the model training corresponds to solve the optimization problem:

$$\begin{aligned} \min_{\theta_F, \theta_C} \max_{\theta_D} L(X_s, Y_s, X_t, Y_t) &= L_{feature} + L_{task} \\ &+ \lambda_1 \|\theta_F + \theta_C\|^2 + \lambda_2 \|\theta_D\|^2 \end{aligned} \tag{4}$$

IV. EXPERIMENTS

A. Dataset Setup

In the experiments, 5-fold cross-validation was adopted to divide the target domain data into the training set and test set. The unit of different amounts of data set is block, which is defined in Section II. B. Each subject has 10 blocks of data, we analyze the results under different numbers of calibration data, for target subject we use the calibration data of 1 to 8 blocks as the training set and the rest as the test set.

To overcome the influence made by the extreme imbalance of two classes, we adopt resampling. Down-sampling the non-target class to the same number as the target class, this operation proceeded only in the source domain and target domain training set.

B. Network Parameters

The network is trained by minimizing the objective function (4). During the training session, the network F and network C were optimized together with Adam, with learning rate of 0.0001 and 0.001 weight decay. Besides, the discriminator was optimized separately with Adam optimizer, with learning rate of 0.0002 and 0.001 weight decay, the gradient penalty coefficient $\lambda = 10$.

C. Compared Methods

We compared our framework to following methods:

- Baseline: HDCA [6], MDRM [9], MCNN [20], and baseline (target only).
- State-of-the-art: OLCNN [21], EEG-Net [10].
- Transfer learning methods: RPA [12], SAN [17], and Li's method [14].

The baseline (target only) training our basic classification network with target domain data only. EEG-Net [10], RPA [12], and SAN [17] was implemented according to follows:

- https://github.com/vlawhern/arl-eegmodels
- https://github.com/plcrodrigues/RPA

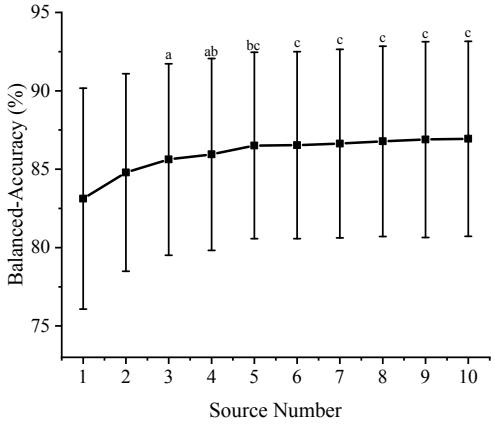


Fig. 3. The BA curve of different source number. Factors marked with the same lowercase letter indicate no significance between each other.

- https://github.com/mingyr/san

Li's method [14] was implemented according to the code provided by author Li. We down-sampled EEG as the input of the model instead of using the DE feature. To perform a fair comparison, we added the label information of target data in training accordingly.

D. Evaluation Metrics

The RSVP image retrieval classification is a binary classification problem that the target samples and non-target samples that are correctly classified are called true positive (TP) and true negative (TN) respectively, otherwise, they are called false negative (FN) and false positive (FP).

Given the class imbalance in the test set, we adopted a more comprehensive metrics balanced-accuracy (BA) other than the accuracy. BA is the average accuracy of all categories, which is calculated in (5). BA can give a better view of classification performance both classes.

$$BA = \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \Big/ 2 \tag{5}$$

V. RESULTS AND DISCUSSION

A. Source Selection Parameter

To find out the optimal source number for our transfer learning framework, we conducted experiments with different numbers of sources under one block calibration data situation. Fig. 3 shows the BA at different numbers of sources. As shown in Fig. 3, the BA showed a trend of increasing from the condition with 1 source to the condition with 5 sources and reached a plateau in the condition with 5 sources. A Friedman test showed the source number have a significant influence on the performance for single-trial detection ($\chi^2(9) = 49.493,\ p < 0.001$). And the post-hoc analysis showed the results at 1 source was significantly lower than that at $2 \sim 10$ ($all : p < 0.05$), and the results at 5 sources was significantly higher than that at $1 \sim 3$ ($1 : p = 0.016,\ 2 : p = 0.006,\ 3 : p = 0.033$). After the number of sources increased over 5, the BA no longer showed any

TABLE I

COMPARISON OF BALANCED-ACCURACY OF DIFFERENT METHODS

Methods	Balanced-Accuracy (%) under Different Blocks of Calibration Data							
	1	2	3	4	5	6	7	8
HDCA [6]	66.84 ± 8.7	72.94 ± 9.8	75.80 ± 9.6	77.53 ± 9.5	78.86 ± 9.2	79.82 ± 8.9	80.61 ± 8.8	81.40 ± 8.5
MDRM [9]	78.75 ± 8.7	80.82 ± 8.0	81.64 ± 7.7	82.38 ± 7.1	82.71 ± 7.0	83.18 ± 6.6	83.28 ± 6.6	83.66 ± 6.3
MCNN [20]	78.30 ± 8.0	82.02 ± 7.2	84.27 ± 6.3	85.58 ± 5.9	86.21 ± 5.9	86.94 ± 5.8	87.45 ± 5.4	88.23 ± 5.2
OLCNN [21]	74.49 ± 4.2	79.37 ± 4.3	82.79 ± 4.3	84.91 ± 4.3	86.26 ± 4.4	87.08 ± 4.3	87.87 ± 4.2	88.34 ± 4.2
EEG-Net [10]	79.16 ± 6.3	83.43 ± 5.4	86.26 ± 4.8	87.42 ± 4.5	87.93 ± 4.3	89.10 ± 3.9	89.24 ± 3.8	89.85 ± 3.5
SAN [17]	67.28 ± 6.1	67.75 ± 7.0	68.08 ± 6.9	68.40 ± 6.9	68.81 ± 7.2	69.03 ± 7.1	68.73 ± 7.7	68.79 ± 7.4
Li [14]	68.97 ± 6.5	72.63 ± 7.1	74.53 ± 7.1	75.61 ± 6.6	76.60 ± 6.9	77.38 ± 6.7	77.48 ± 7.0	78.13 ± 7.1
RPA [12]	80.58 ± 8.3	82.98 ± 7.5	83.86 ± 6.5	83.58 ± 6.6	84.04 ± 6.6	84.41 ± 6.4	83.72 ± 6.1	84.24 ± 6.1
Baseline (target only)	78.84 ± 9.2	83.53 ± 7.3	84.99 ± 6.8	86.92 ± 6.0	88.12 ± 5.7	88.24 ± 5.0	88.73 ± 4.6	89.73 ± 4.6
Our (transfer learning)	$\mathbf{86.71 \pm 5.6}$	$\mathbf{88.31 \pm 4.9}$	$\mathbf{89.13 \pm 4.4}$	$\mathbf{89.93 \pm 4.2}$	$\mathbf{90.31 \pm 4.1}$	$\mathbf{90.50 \pm 3.8}$	$\mathbf{91.10 \pm 3.6}$	$\mathbf{91.47 \pm 3.5}$

TABLE II

BALANCED-ACCURACY OF ABLATION EXPERIMENTS. M1: CALIBRATION BASELINE, M2: ADVERSARIAL TRANSFER, M3: CONDITIONAL ADVERSARIAL TRANSFER WITH GRADIENT PENALTY

Models	Balanced-Accuracy (%) under Different Blocks of Calibration Data							
	1	2	3	4	5	6	7	8
M1	82.25 ± 7.2	83.92 ± 6.5	85.49 ± 5.7	86.37 ± 5.1	86.52 ± 5.1	87.88 ± 4.7	87.71 ± 4.6	88.28 ± 5.0
M2	85.15 ± 6.6	87.06 ± 5.6	87.91 ± 5.1	88.82 ± 4.7	89.28 ± 4.6	89.80 ± 4.3	90.10 ± 4.5	90.37 ± 4.8
M3	$\mathbf{86.71 \pm 5.6}$	$\mathbf{88.31 \pm 4.9}$	$\mathbf{89.13 \pm 4.4}$	$\mathbf{89.93 \pm 4.2}$	$\mathbf{90.31 \pm 4.1}$	$\mathbf{90.50 \pm 3.8}$	$\mathbf{91.10 \pm 3.6}$	$\mathbf{91.47 \pm 3.5}$

significant difference between each two groups ($p > 0.05$). Therefore, we selected 5 sources for subsequent transfer learning experiments. The source number of 5 is a balance point of best classification performance and least calculation.

B. Classification Performance

In order to validate the efficiency of the method proposed in this study, we compared the performance of our method with our within-subject baseline and other compared methods with different calibration data.

The BA of classification is presented in Table I with mean $\pm$ standard deviation. The two-way repeated measures ANOVA revealed that there was an interaction effect between methods and the quantity of calibration data for the BA ($p < 0.001$ with Greenhouse & Geisser epsilon correction). The post-hoc analysis showed BA of our method was significantly higher than HDCA, OLCNN, MCNN, MDRM, Li, SAN, and RPA ($all : p < 0.002$). Compared with EEG-Net, the balanced-accuracy of our method was significantly higher at $1 \sim 3$ and 5, 8 block calibration data, and tends to be higher at 4, 6, 7 block calibration data ($p < 0.1$). Therefore, these results indicated that our method performed better than other methods under the same calibration data.

Among other methods, RPA has the highest BA when the quantity of calibration data was 1 block, the classification performance of EEG-net wsa better than all other methods at $2 \sim 8$ blocks.

BA of our method at 1 block of calibration data (estimated boundary mean $\pm$ standard deviation $86.71 \pm 1.8\%$, Bonferroni correction of 9) was significantly higher than that of RPA at $1 \sim 8$ blocks of calibration data ($all : p < 0.01$). Hence, our method performs better with one block of calibration data than RPA with any number of calibration data. Compared with BA of EEG-Net at 2 blocks of calibration data, BA of our method at 1 block of calibration was significantly higher ($F(1,10) = 15.886,\ p = 0.003$). BA of our method at 1 block of calibration data was not significantly different from that of EEG-net at 3 blocks, though our average BA was 0.45% higher. Therefore, our method using only 1 block of calibration data can achieve the BA performance that the EEG-Net needs 3 blocks of calibration data.

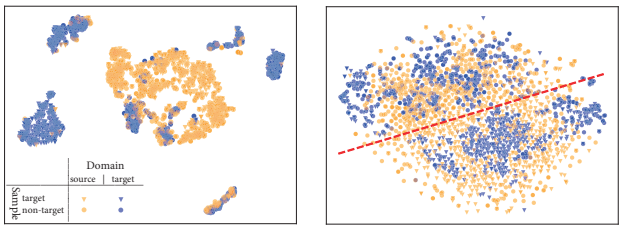
C. Ablation Experiments

To evaluate the effectiveness of each component of our transfer learning network, we decomposed the proposed network into three parts and marked as model M1 to M3.

M1: The model including the F and the C constitutes the basic calibration classification network. The same multi-source voting framework is employed in M1. **M2**: Based on M1, the D is added to perform adversarial domain adaptation; **M3**: Based on M2, the label information is used as a condition to guide the training, and the gradient penalty is introduced into the training. We performed experiments on these three architectures and results are shown in Table II with mean $\pm$ standard deviation.

The two-way repeated measures ANOVA revealed that the two factors (model and number of calibration data) have significantly influence on classification performance, and there was no interaction between the two factors ($F(4.099, 40.986) = 2.089,\ p = 0.098$).

The main effect analysis showed that the estimated boundary mean BA of M2 ($88.6 \pm 1.5\%$) was 3.1% higher than M1 ($86.1 \pm 1.6\%$) with significance ($p < 0.01$, Bonferroni correction of 3). This indicated the adversarial network architecture can improve the classification performance compared



(a) (b)

Fig. 4. An example of visualization, (a) visualization of rawData, (b) visualization of output features of feature extractor after transfer leaning. Triangle: target samples, Dots: non-target samples, Yellow: source domain data, Blue: target domain data.

to the basic model. Comparison of M3 and M2 shows the conditional constraint and gradient penalty can further improve the mean classification performance to $89.2 \pm 1.4\%$ (estimated boundary mean BA, Bonferroni correction of 3).

D. Visualization

We applied t-distributed Stochastic Neighbor Embedding (t-SNE) to project the features into 2 dimensions and plotted scatter. We took a subject as an example and visualized the test data and the corresponding first source under 1 block of training data. The feature visualization before and after transfer shows in Fig. 4.

In Fig. 4(a), we can see that the distribution of the source domain and the target domain is greatly different before the transfer, and it is difficult to distinguish between the target class and the non-target class. In Fig. 4(b), after training the extracted feature distribution of the target domain tends to be more consistent with that of the source domain. And different classes in both target domain and source domain are distinguishable, as shown in the Fig. 4(b), the two types of samples have relatively obvious boundaries (the red dotted line). Therefore, our transfer learning network reduces the difference of distribution of the source domain and the target domain and increases the discrimination of two kinds of samples.

VI. CONCLUSION

In this paper, we proposed a multi-source transfer learning framework to reduce the time that the calibration required for a RSVP-based BCI system. A source selection strategy was proposed and adopted to avoid negative transfer and reduce computation. We employed a domain-adversarial training network with conditional information to enable the CNN-based network learn common features from different domains. The experimental results indicated that our method significantly 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 application under the small calibration data.

REFERENCES

[1] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M., 2002. Brain–computer interfaces for communication and control. Clinical neurophysiology, 113(6), pp.767-791.

[2] Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T.P. and Ming, D., 2018. A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. IEEE Transactions on Biomedical Engineering, 65(5), pp.1166-1175.

[3] Acqualagna, L. and Blankertz, B., 2013. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clinical Neurophysiology, 124(5), pp.901-908.

[4] Lin, Z., Zhang, C., Zeng, Y., Tong, L. and Yan, B., 2018. A novel P300 BCI speller based on the Triple RSVP paradigm. Scientific reports, 8(1), p.3350.

[5] Pohlmeyer, E.A., Wang, J., Jangraw, D.C., Lou, B., Chang, S.F. and Sajda, P., 2011. Closing the loop in cortically-coupled computer vision: a brain–computer interface for searching image databases. Journal of neural engineering, 8(3), p.036025.

[6] Gerson, A.D., Parra, L.C. and Sajda, P., 2006. Cortically coupled computer vision for rapid image search. IEEE Transactions on neural systems and rehabilitation engineering, 14(2), pp.174-179.

[7] Barngrover, C., Althoff, A., DeGuzman, P. and Kastner, R., 2015. A brain–computer interface (BCI) for the detection of mine-like objects in sidescan sonar imagery. IEEE Journal of Oceanic Engineering, 41(1), pp.123-138.

[8] Wu, Q., Yan, B., Zeng, Y., Zhang, C. and Tong, L., 2018. Anti-deception: Reliable EEG-based biometrics with real-time capability from the neural response of face rapid serial visual presentation. Biomedical engineering online, 17(1), p.55.

[9] Barachant, A. and Congedo, M., 2014. A plug&play P300 BCI using information geometry. arXiv preprint arXiv:1409.0107.

[10] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P. and Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering, 15(5), p.056013.

[11] Zanini, P., Congedo, M., Jutten, C., Said, S. and Berthoumieu, Y., 2017. Transfer learning: a Riemannian geometry framework with applications to brain–computer interfaces. IEEE Transactions on Biomedical Engineering, 65(5), pp.1107-1116.

[12] Rodrigues, P.L.C., Jutten, C. and Congedo, M., 2018. Riemannian procrustes analysis: Transfer learning for brain-computer interfaces. IEEE Transactions on Biomedical Engineering.

[13] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[14] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V., 2017. Domain-adversarial training of neural networks. In Domain Adaptation in Computer Vision Applications (pp. 189-209). Springer, Cham.

[15] Tzeng, E., Hoffman, J., Saenko, K. and Darrell, T., 2017. Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7167-7176).

[16] Li, J., Qiu, S., Du, C., Wang, Y. and He, H., 2019. Domain Adaptation for EEG Emotion Recognition Based on Latent Representation Similarity. IEEE Transactions on Cognitive and Developmental Systems.

[17] Ming, Y., Ding, W., Pelusi, D., Wu, D., Wang, Y.K., Prasad, M. and Lin, C.T., 2019. Subject adaptation network for EEG data analysis. Applied Soft Computing, 84, p.105689.

[18] A Torralba, K. P. Murphy and W. T. Freeman, The MIT-CSAIL Database of Objects and Scenes. available at http://web.mit.edu/torralba/www/database.html

[19] Yao, Y. and Doretto, G., 2010, June. Boosting for transfer learning with multiple sources. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1855-1862). IEEE.

[20] Manor, R. and Geva, A.B., 2015. Convolutional neural network for multi-category rapid serial visual presentation BCI. Frontiers in computational neuroscience, 9, p.146.

[21] Shan, H., Liu, Y. and Stefanov, T.P., 2018, July. A Simple Convolutional Neural Network for Accurate P300 Detection and Character Spelling in Brain Computer Interface. In IJCAI (pp. 1604-1610).

[22] Arjovsky, M., Chintala, S. and Bottou, L., 2017. Wasserstein GAN. arXiv preprint arXiv:1701.07875.

[23] Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[24] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A.C., 2017. Improved training of wasserstein GANs. In NIPS (pp. 5767-5777).