EasyChair Preprint

№ 6952

Subject Adaptive EEG-Based Visual Recognition

Pilhyeon Lee, Sunhee Hwang, Seogkyu Jeon and Hyeran Byun

October 30, 2021

Subject Adaptive EEG-based Visual Recognition

Pilhyeon Lee[1], Sunhee Hwang[4], Seogkyu Jeon[1], and Hyeran Byun[1,2,3⋆]

[1] Department of Computer Science, Yonsei University
[2] Graduate School of Artificial Intelligence, Yonsei University
[3] Graduate Program of Cognitive Science, Yonsei University
[4] AI Imaging Tech. Team, LG Uplus
{lph1114, jone9312, hrbyun}@yonsei.ac.kr, sunheehwang@lguplus.co.kr

Abstract. This paper focuses on EEG-based visual recognition, aiming to predict the visual object class observed by a subject based on his/her EEG signals. One of the main challenges is the large variation between signals from different subjects. It limits recognition systems to work only for the subjects involved in model training, which is undesirable for real-world scenarios where new subjects are frequently added. This limitation can be alleviated by collecting a large amount of data for each new user, yet it is costly and sometimes infeasible. To make the task more practical, we introduce a novel problem setting, namely *subject adaptive EEG-based visual recognition*. In this setting, a bunch of pre-recorded data of existing users (source) is available, while only a little training data from a new user (target) are provided. At inference time, the model is evaluated solely on the signals from the target user. This setting is challenging, especially because training samples from source subjects may not be helpful when evaluating the model on the data from the target subject. To tackle the new problem, we design a simple yet effective baseline that minimizes the discrepancy between feature distributions from different subjects, which allows the model to extract subject-independent features. Consequently, our model can learn the common knowledge shared among subjects, thereby significantly improving the recognition performance for the target subject. In the experiments, we demonstrate the effectiveness of our method under various settings. Our code is available at here[1].

Keywords: Brain-computer interface · Electroncephalography · Visual recognition · Subject adaptation · Deep Learning.

1 Introduction

Brain-computer interface (BCI) has been a long-standing research topic for decoding human brain activities, playing an important role in reading the human mind with various applications [44,32,40,21]. For instance, BCI systems enable a

⋆ Corresponding author
[1] https://github.com/DeepBCI/Deep-BCI

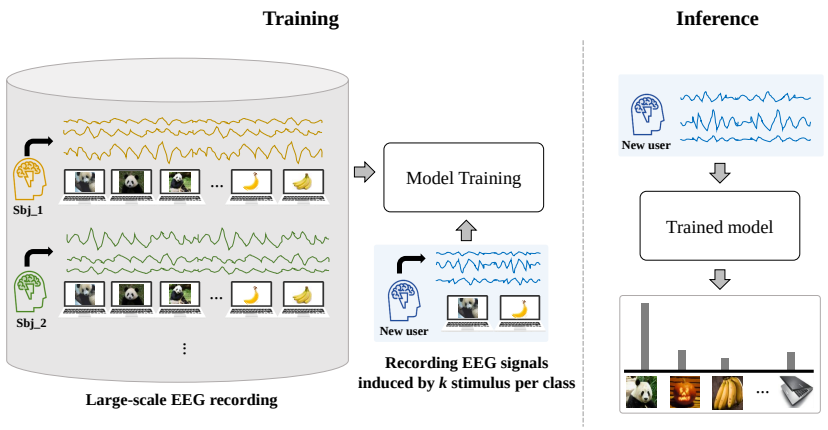

Fig. 1. An illustration of *Subject Adaptive EEG-based Visual Recognition*. During the large-scale EEG recording step, abundant sample images are observed by various subjects (source) and we collect their EEG signals. Afterwards, we record EEG signals from a new user (target) induced by only k stimuli per class. We train the model on the EEG signals from the source and the target subject and expect the trained model to correctly predict the visual classes given unseen EEG signals from the target subject.

user to comfortably control machines without requiring any peripheral muscular activities [3,27]. In addition, BCI is especially helpful for people suffering from speech or movement disorders, allowing them to freely communicate and express their feelings by thinking [4,12,7,24]. It also can be utilized to identify abnormal states of brains, such as seizure state, sleep disorder, and dementia [43,41,34,33]. Recently, taking it to the next level, numerous works attempt to decode brain signals for figuring out what audiovisual stimulus is being taken by a person, providing deeper insight for analyzing human perception [37,26,1,13].

There are different ways to collect brain signals, *e.g.*, electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Among them, EEG is considered the most favorable one to analyze human brain activities since it is non-invasive and promptly acquirable. With its numerous advantages, EEG-based models have been largely explored by researchers and developed for various research fields such as disorder detection [2,29], drowsy detection [17,23], emotion recognition [15,14,30], *etc.*

In this paper, we tackle the task of visual recognition based on EEG signals, whose goal is to classify visual stimuli taken by subjects. Recently, thanks to the effectiveness of deep neural networks (DNNs), existing models have shown impressive recognition performances [15,23,37,36]. However, they suffer from the large inter-subject variability of EEG signals, which greatly restricts their scalability. Suppose that a model faces a new user not included in the training set – note that this is a common scenario in the real world. Since the EEG signals from the user are likely to largely differ from those used for training, the model would fail to recognize the classes. Therefore, in order to retain the performance,

it is inevitable to collect EEG signals for training from the new subject, which requires additional costs proportional to the number of the samples. If we have sufficient training samples for the new subject, the model would show great performance, but it is not the case for the real-world scenario.

To handle this limitation and bypass the expensive cost, we introduce a new practical problem setting, namely *subject adaptive EEG-based visual recognition*. In this setting, we have access to abundant EEG signals from various source subjects, whereas the signals from a new user (target subject) are scarce, *i.e.*, only a few samples (k-shot) are allowed for each visual category. At inference, the model should correctly classify the EEG signals from the target subject. Fig. 1 provides a graphical illustration of the proposed problem setting.

Naturally, involving the copious samples from source subjects in the model training would bring about performance gains compared to the baseline using only signals from the target subject. However, as aforementioned, the signals obtained from the source and the target subjects are different from each other, and thus the performance improvements are limited. To maximize the benefits of pre-acquired data from source subjects, we here provide a simple yet effective baseline method. Our key idea is to allow the model to learn subject-agnostic representations for EEG-based visual recognition. Technically, together with the conventional classification loss, we design a loss to minimize maximum mean discrepancy (MMD) between feature distributions of EEG signals from different subjects. On the experiments under a variety of circumstances, our method shows consistent performance improvements over the vanilla method.

Our contributions can be summarized in three-fold.

- We introduce a new realistic problem setting, namely subject-adaptive EEG-based visual recognition. Its goal is to improve the recognition performance for the target subject whose training samples are limited.
- We design a simple baseline method for the proposed problem setting. It encourages the feature distributions between different subjects to be close so that the model learns subject-independent representations.
- Through the experiments on the public benchmark, we validate the effectiveness of our model. Specifically, in the extreme 1-shot setting, it achieves the performance gain of 6.4% upon the vanilla model.

2 Related work

2.1 Brain activity underlying visual perception

Over recent decades, research on visual perception has actively investigated to reveal the correlation between brain activity and visual stimuli [35,31,9]. Brain responses induced by visual stimuli come from the occipital cortex that is a brain region for receiving and interpreting visual signals. In addition, visual information obtained by the occipital lobe is transmitted to nearby parietal and temporal lobes to perceive higher-level information. Based on this prior knowledge, researchers have tried to analyze brain activities induced by visual stimuli.

Eroğlu *et al.* [8] examine the effect of emotional images with different luminance levels on EEG signals. They also find that the brightness of visual stimuli can be represented by the activity power of the brain cortex. Stewart *et al.* [38] attempt to distinguish the presence of visual stimuli within a single trial in EEG recordings. It is revealed in their analyses that the individual components of EEG signals are spatially located in the visual cortex and are effective in classifying visual states. More recently, Spampinato *et al.* [37] tackle the problem of EEG-based visual recognition by learning a discriminative manifold of brain activities on diverse visual categories. Besides, they build a large-scale EEG dataset for training deep networks and demonstrate that human visual perception abilities can be transferred to deep networks. Kavasidis *et al.* [20] propose to reconstruct the observed images by decoding EEG signals. They find that EEG contains some patterns related to visual contents, which can be used to effectively generate images that are semantically coherent to the visual stimuli.

In line with these works, we build a visual recognition model to decode EEG signals induced by visual stimuli. In addition, we design and tackle a new practical problem setting where a limited amount of data is allowed for new users.

2.2 Subject-independent EEG-based classification

Subject-dependent EEG-based classification models have widely been studied, achieving the noticeable performances [5,19,14,30,16]. However, EEG signal patterns greatly vary among individuals, building a subject-independent model remains an important research topic to be solved. Hwang *et al.* [15] train a subject-independent EEG-based emotion recognition model by utilizing an adversarial learning approach to make the model not able to predict the subject labels. Zhang *et al.* [42] propose a convolutional recurrent attention model to classify movement intentions by focusing on the most discriminative temporal periods from EEG signals. In [17], an EEG-based drowsy driving detection model is introduced, which is trained in an adversarial manner with gradient reversal layers in order to encourage feature distribution to be close between subjects.

Besides, to eliminate the expensive calibration process for new users, zero-training BCI techniques are introduced which does not require the re-training. Lee *et al.* [25] try to find the network parameters that generalize well on common features across subjects. Meanwhile, Grizou *et al.* [11] propose a zero-training BCI method that controls virtual and robotic agents in sequential tasks without requiring calibration steps for new users.

Different from the works above, we tackle the problem of EEG-based visual recognition. Moreover, we propose a new problem setting to reduce the cost of acquiring labeled data for new users, as well as introduce a strong baseline.

3 Dataset

Before introducing the proposed method, we first present the dataset details for experiments. We use the publicly available large-scale EEG dataset collected

Table 1. The list of object classes utilized for collecting EEG signals with ImageNet [6] class indices.

n02106662	German shepherd	n02951358	Canoe	n03445777	Golf ball	n03888257	Parachute
n02124075	Egyptian cat	n02992529	Cellular telephone	n03452741	Grand piano	n03982430	Pool table
n02281787	Lycaenid	n03063599	Coffee mug	n03584829	Iron	n04044716	Radio telescope
n02389026	Sorrel	n03100240	Convertible	n03590841	Jack-o'-lantern	n04069434	Reflex camera
n02492035	Capuchin	n03180011	Desktop computer	n03709823	Mailbag	n04086273	Revolver
n02504458	African elephant	n03197337	Digital watch	n03773504	Missile	n04120489	Running shoe
n02510455	Giant panda	n03272010	Electric guitar	n03775071	Mitten	n07753592	Banana
n02607072	Anemone fish	n03272562	Electric locomotive	n03792782	Mountain bike	n07873807	Pizza
n02690373	Airliner	n03297495	Espresso maker	n03792972	Mountain tent	n11939491	Daisy
n02906734	Broom	n03376595	Folding chair	n03877472	Pajama	n13054560	Bolete

by [37] that consists of 128-channel EEG sequences lasting for 440 ms from six different subjects (five male and one female). The EEG signals are filtered using a notch filter (49-51 Hz) and a band-pass filter (14-72 Hz) to include two frequency bands, *i.e.*, Beta and Gamma. The dataset contains 40 easily distinguishable object categories from ImageNet [6], which are listed in Table 1. The number of image samples looked at by subjects is 50 for each class, constituting a total of 2,000 samples. We use the official splits, keeping the ratio of training, validation, and test sets as 4:1:1. The dataset contains a total of 6 splits and we measure the mean and the standard deviation of performance of 6 runs in the experiments. We refer readers to the original paper [37] for further details about the dataset.

4 Method

In this section, we first define the proposed problem setting (Sec. 4.1). Then, we introduce a baseline method with subject-independent learning to tackle the problem. Its network architecture is illustrated in Sec. 4.2, followed by the detailed subject-independent learning scheme (Sec. 4.3). An overview of our method is depicted in Fig. 2.

4.1 Subject Adaptive EEG-based Visual Recognition

We start by providing the formulation of the conventional EEG-based visual recognition task. Let $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$ denote the dataset collected from the s-th subject. Here, $x_i^s \in \mathbb{R}^{D \times T}$ denotes the i-th EEG sample of subject s with its channel dimension D and the duration T, while $y_i^s \in \mathbb{R}^K$ is the corresponding ground-truth visual category observed by the subject and N^s is the number of the samples for subject s. In general, the EEG samples are abundant for each subject, *i.e.*, $N^s \gg 0$. To train a deep model, multiple datasets from different subjects are assembled to build a single training set $\mathcal{D} = \{\mathcal{D}^1, \mathcal{D}^2, ..., \mathcal{D}^S\}$, where S is the total number of subjects. At inference, given an EEG sample $x^{s'}$, the model should predict its category. Here, it is assumed that the input signal at test time is obtained by one of the subjects whose samples are used during the training stage, *i.e.*, $s' \in [1, S]$. However, this conventional setting is impractical especially for the case where EEG data from new subjects are scarce.

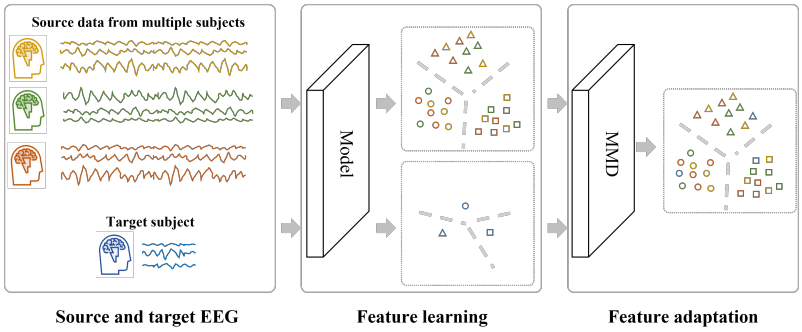

Fig. 2. An overview of the proposed method. Colors and shapes respectively represent subject identities and classes. During feature learning, we train the model to accurately predict the class from the EEG signals. To alleviate the feature discrepancy of source and target signals, we propose a feature adaptation stage which minimizes the maximum mean discrepancy. Consequently, both source and target features are projected on the same manifold, enabling accurate predictions on target signals during inference.

Instead, we propose a more realistic problem setting, named *Subject Adaptive EEG-based Visual Recognition*. In this setting, we aim to utilize the knowledge learned from abundant data of source subjects to classify signals from a target subject whose samples are rarely accessible. For that purpose, we first divide the training set into source and target sets, *i.e.*, $\mathcal{D}_{src}$ and $\mathcal{D}_{trg}$. We choose a subject and set it to be the target while the rest become the sources. For example, letting subject S be the target, $\mathcal{D}_{src} = \{\mathcal{D}^1, \mathcal{D}^2, ..., \mathcal{D}^{S-1}\}$ and $\mathcal{D}_{trg} = \hat{\mathcal{D}}^S \subset \mathcal{D}^S$. Based on the sparsity constraint, the target dataset contains only a few examples, *i.e.*, $\hat{\mathcal{D}}^S = \{(x_j^S, y_j^S)\}_{j=1}^{\hat{N}^S}$, where $\hat{N}^S \ll N^S$. In practice, we make the target set have only k samples with their labels per class (k-shot). Note that we here use the S-th subject as the target, but any subject can be the target without loss of generality. After trained on $\mathcal{D}_{src}$ and $\mathcal{D}_{trg}$, the model is supposed to predict the class of an unseen input signal x^S which is obtained from the target subject S.

4.2 Network Architecture

In this section, we describe the architectural details of the proposed simple baseline method. Our network is composed of a sequence encoder f, an embedding layer g, and a classifier h. The sequence encoder $f(\cdot)$ is a single-layer gated recurrent unit (GRU), which takes as input an EEG sample and outputs the extracted feature representation $z = f(x) \in \mathbb{R}^{D_{seq}}$, where $\mathbb{R}^{D_{seq}}$ is the feature dimension. Although the encoder produces the hidden representation for every timestamp, we only use the last feature and discard the others since it encodes the information from all timestamps. Afterwards, the feature z is embedded to the semantic manifold by the embedding layer $g(\cdot)$, *i.e.*, $w = g(z) \in \mathbb{R}^{D_{emb}}$, where $\mathbb{R}^{D_{emb}}$ is the dimension of embedded features. The embedding layer $g(\cdot)$ is composed of a fully-connected (FC) layer with an activation function. As the final step, we

feed the embedded feature w to the classifier $h(\cdot)$ consisting of a FC layer with the softmax activation, producing the class probability $p(\mathbf{y}|x;\theta) = h(w) \in \mathbb{R}^K$. Here, θ is a set of the trainable parameters in the overall network. To train our network for the classification task, we minimize the cross-entropy loss as follows.

$$\mathcal{L}_{\text{cls}} = \frac{-1}{|\mathcal{D}_{src}| + |\mathcal{D}_{trg}|} \sum_{(x_i, y_i) \in \mathcal{D}_{src} \cup \mathcal{D}_{trg}} y_i \log p(y_i|x_i;\theta), \tag{1}$$

where $|\mathcal{D}_{src}|$ and $|\mathcal{D}_{trg}|$ indicate the number of samples in source and target sets.

4.3 Subject-independent Feature Learning

In spite of the learned class-discriminative knowledge, the model might not fully benefit from the data of source subjects due to the feature discrepancy from different subjects. To alleviate this issue and better exploit the source set, we propose a simple yet effective framework, where subject-independent features are learned by minimizing the divergence between feature distributions of source and target subjects. Concretely, for the divergence metric, we estimate the multi-kernel maximum mean discrepancy (MK-MMD) [28] between the feature distributions Z^{s_i} and Z^{s_j} from two subjects s_i and s_j as follows.

$$\text{MMD}(Z^{s_i}, Z^{s_j}) = \left\| \frac{1}{N^{s_i}} \sum_{n=1}^{N^{s_i}} \phi(z_n^{s_i}) - \frac{1}{N^{s_j}} \sum_{m=1}^{N^{s_j}} \phi(z_m^{s_j}) \right\|_F, \tag{2}$$

where $\phi(\cdot) : \mathcal{W} \rightarrow \mathcal{F}$ is the mapping function to the reproducing kernel Hilbert space, while $\|\cdot\|_F$ indicates the Frobenius norm. $z_n^{s_i}$ denotes the n-th feature from subject s_i encoded by the sequence encoder f, whereas N^{s_i} and N^{s_j} are the total numbers of samples from the s_i-th and the s_j-th subjects in the training set, respectively. In practice, we use the samples in an input batch rather than the whole training set due to the memory constraint. We note that the embedded feature w_n^i could also be utilized to compute the discrepancy, but we empirically find that it generally performs inferior to the case of using z_n^i (Sec. 5.3).

Reducing the feature discrepancy between different subjects allows the model to learn subject-independent features. To make feature distributions from all subjects close, we compute and minimize the MK-MMD of all possible pairs of the subjects. Specifically, we design the discrepancy loss that is formulated as:

$$\mathcal{L}_{\text{disc}} = \frac{2}{S(S-1)} \sum_{s_i=1}^{S} \sum_{\forall s_j \neq s_i} \text{MMD}(Z^{s_i}, Z^{s_j}), \tag{3}$$

where S is the number of the subjects in the training data including the target.

By minimizing the discrepancy loss, our model could learn subject-independent features and better utilize the source data to improve the recognition performance for the target subject. The overall training loss of our model is a weighted sum of the losses, which is computed as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{disc}}, \tag{4}$$

where λ is the weighting factor, which is empirically set to 1.

Table 2. Quantitative comparison of methods by changing the target subject. For evaluation, we select one subject as a target and set the rest as sources, then compute the top-k accuracy for the test set from the target subject. Note that only a single target sample for each class is included in training, *i.e.*, 1-shot setting. We measure the mean and the standard deviation of a total of 5 runs following the official splits.

Validation set						
Subject	top-1 accuracy (%)			top-3 accuracy (%)		
	k-shot	Vanilla	Ours	k-shot	Vanilla	Ours
#0	$13.5_{\pm 2.1}$	$29.3_{\pm 1.9}$	$\mathbf{35.7}_{\pm 1.9}$	$22.6_{\pm 2.8}$	$51.6_{\pm 3.0}$	$\mathbf{58.1}_{\pm 2.9}$
#1	$12.6_{\pm 2.1}$	$21.8_{\pm 2.3}$	$\mathbf{29.0}_{\pm 3.6}$	$22.3_{\pm 2.5}$	$41.0_{\pm 5.1}$	$\mathbf{49.5}_{\pm 3.5}$
#2	$17.0_{\pm 1.6}$	$25.3_{\pm 0.9}$	$\mathbf{30.8}_{\pm 2.2}$	$29.8_{\pm 2.2}$	$44.4_{\pm 2.1}$	$\mathbf{53.1}_{\pm 2.6}$
#3	$27.8_{\pm 1.7}$	$28.8_{\pm 2.2}$	$\mathbf{31.9}_{\pm 3.9}$	$41.6_{\pm 2.1}$	$47.8_{\pm 4.1}$	$\mathbf{52.6}_{\pm 3.7}$
#4	$16.3_{\pm 2.8}$	$25.9_{\pm 1.9}$	$\mathbf{36.2}_{\pm 3.3}$	$25.9_{\pm 2.3}$	$44.4_{\pm 2.7}$	$\mathbf{61.0}_{\pm 4.7}$
#5	$9.2_{\pm 1.4}$	$20.7_{\pm 2.9}$	$\mathbf{25.8}_{\pm 1.7}$	$16.9_{\pm 2.5}$	$40.1_{\pm 3.9}$	$\mathbf{47.5}_{\pm 3.4}$
Test set						
Subject	top-1 accuracy (%)			top-3 accuracy (%)		
	k-shot	Vanilla	Ours	k-shot	Vanilla	Ours
#0	$12.2_{\pm 2.1}$	$24.3_{\pm 0.9}$	$\mathbf{29.6}_{\pm 4.9}$	$20.4_{\pm 2.5}$	$48.3_{\pm 2.3}$	$\mathbf{56.8}_{\pm 4.1}$
#1	$10.3_{\pm 2.2}$	$18.1_{\pm 2.7}$	$\mathbf{25.4}_{\pm 2.4}$	$20.8_{\pm 2.1}$	$39.0_{\pm 1.9}$	$\mathbf{49.0}_{\pm 2.4}$
#2	$15.5_{\pm 2.9}$	$23.9_{\pm 3.0}$	$\mathbf{29.2}_{\pm 3.7}$	$29.9_{\pm 3.4}$	$44.3_{\pm 4.3}$	$\mathbf{54.5}_{\pm 3.1}$
#3	$26.2_{\pm 3.2}$	$27.4_{\pm 3.2}$	$\mathbf{32.1}_{\pm 4.3}$	$41.7_{\pm 3.9}$	$47.9_{\pm 4.2}$	$\mathbf{53.6}_{\pm 4.0}$
#4	$15.2_{\pm 1.9}$	$22.7_{\pm 1.2}$	$\mathbf{35.3}_{\pm 3.6}$	$24.5_{\pm 2.0}$	$44.8_{\pm 3.5}$	$\mathbf{60.7}_{\pm 4.9}$
#5	$7.0_{\pm 1.0}$	$18.9_{\pm 2.9}$	$\mathbf{21.4}_{\pm 2.6}$	$15.3_{\pm 1.8}$	$38.4_{\pm 4.1}$	$\mathbf{45.0}_{\pm 4.1}$

5 Experiments

5.1 Implementation Details

The input signals for our method contain a total of 128 channels ($D = 128$) with a recording unit of 1 *ms*, each of which lasts for 440 *ms*. Following [37], we only use the signals within the interval of 320-480 *ms*, resulting in the temporal dimension $T = 160$. As described in Sec. 4.2, our model consists of a single-layer gated recurrent unit (GRU) followed by two fully-connected layers respectively for embedding and classification. For all layers but the classifier, we set their hidden dimensions to the same one with input signals to preserve the dimensionality, *i.e.*, $D_{seq} = D_{emb} = 128$. For non-linearity, we put the Leaky ReLU activation after the embedding layer g with $\alpha = 0.2$. To estimate multi-kernel maximum mean discrepancy, we use the radial basis function (RBF) kernel [39] as the mapping function. For effective learning, we make sure that all the subjects are included in a single batch. Technically, we randomly pick 200 examples from each source dataset and take all samples in the target dataset to configure a batch. Our model is trained in an end-to-end fashion from scratch without pre-training. For model training, we use the Adam [22] optimizer with a learning rate of 10^{-3}.

5.2 Quantitative Results

To validate the effectiveness of our method, we compare it with two different competitors: k-shot baseline and the vanilla model. First, the k-shot method is

Table 3. Quantitative comparison of methods by changing the number of target samples per class provided during training. The value of k means that only k samples of the target subject are used for training. We measure the mean and the standard deviation of a total of 5 runs for all subjects following the official splits.

Validation set						
k	top-1 accuracy (%)			top-3 accuracy (%)		
	k-shot	Vanilla	Ours	k-shot	Vanilla	Ours
1	$16.0_{\pm 0.6}$	$25.3_{\pm 1.0}$	$\mathbf{31.7}_{\pm 1.5}$	$26.5_{\pm 0.9}$	$44.9_{\pm 1.3}$	$\mathbf{53.6}_{\pm 1.9}$
2	$33.2_{\pm 1.2}$	$41.7_{\pm 1.9}$	$\mathbf{46.3}_{\pm 1.8}$	$50.1_{\pm 1.0}$	$65.2_{\pm 2.0}$	$\mathbf{70.2}_{\pm 1.6}$
3	$49.9_{\pm 0.4}$	$54.4_{\pm 1.0}$	$\mathbf{58.9}_{\pm 0.7}$	$68.5_{\pm 0.7}$	$77.6_{\pm 0.7}$	$\mathbf{80.8}_{\pm 1.2}$
4	$61.9_{\pm 2.0}$	$64.6_{\pm 1.5}$	$\mathbf{67.5}_{\pm 1.2}$	$79.6_{\pm 1.7}$	$85.1_{\pm 1.1}$	$\mathbf{86.8}_{\pm 1.2}$
5	$70.0_{\pm 1.6}$	$72.0_{\pm 1.3}$	$\mathbf{73.5}_{\pm 1.1}$	$85.6_{\pm 1.7}$	$89.6_{\pm 0.9}$	$\mathbf{90.0}_{\pm 1.0}$
Test set						
k	top-1 accuracy (%)			top-3 accuracy (%)		
	k-shot	Vanilla	Ours	k-shot	Vanilla	Ours
1	$14.4_{\pm 1.6}$	$22.5_{\pm 0.8}$	$\mathbf{28.8}_{\pm 1.2}$	$25.4_{\pm 1.8}$	$43.8_{\pm 1.6}$	$\mathbf{53.3}_{\pm 1.9}$
2	$31.2_{\pm 1.2}$	$39.9_{\pm 2.0}$	$\mathbf{43.8}_{\pm 1.4}$	$49.3_{\pm 2.0}$	$65.1_{\pm 2.1}$	$\mathbf{69.5}_{\pm 1.4}$
3	$48.2_{\pm 2.6}$	$52.6_{\pm 1.7}$	$\mathbf{56.4}_{\pm 1.7}$	$67.2_{\pm 1.7}$	$77.0_{\pm 1.5}$	$\mathbf{80.4}_{\pm 1.1}$
4	$60.4_{\pm 0.9}$	$62.4_{\pm 1.7}$	$\mathbf{64.7}_{\pm 1.6}$	$79.4_{\pm 1.1}$	$84.3_{\pm 0.9}$	$\mathbf{85.9}_{\pm 1.1}$
5	$68.1_{\pm 1.6}$	$69.5_{\pm 1.1}$	$\mathbf{70.1}_{\pm 1.0}$	$85.6_{\pm 1.3}$	$89.0_{\pm 0.5}$	$\mathbf{89.2}_{\pm 0.5}$

trained exclusively on the target dataset. As the amount of target data is limited, the model is expected to poorly perform and it would serve as the baseline for investigating the benefit of source datasets. Next, the vanilla model is a variant of our method that discards the discrepancy loss. Its training depends solely on the classification loss without considering subjects, and thus it can demonstrate the effect of abundant data from other unrelated subjects.

Comparison in the 1-shot setting. We first explore the most extreme scenario of our subject adaptive EEG-based visual classification, *i.e.*, the 1-shot setting. In this setting, only a single example for each visual category is provided for the target subject. The experimental results are summarized in Table 2. As expected, the k-shot baseline performs the worst due to the scarcity of training data. When including the data from source subjects, the vanilla setting improves the performance to an extent. However, we observe that the performance gain is limited due to the representation gap between subjects. On the other hand, our model manages to learn subject-independent information and brings a large performance boost upon the vanilla method without regard to the choice of the target subject. Specifically, the top-1 accuracy of subject #1 on the validation set is improved by 7.2% from the vanilla method. This clearly validates the effectiveness of our approach.

Comparison with varying k. To investigate the performance in diverse scenarios, we evaluate the models with varying k for the k-shot setting. Specifically, we change k from 1 to 5 and the results are provided in Table 3. Obviously, increasing

Table 4. Ablation on the location of feature adaptation. We compare two variants that minimize discrepancy after the sequence encoder f and the embedding layer g, respectively. We measure the mean and the standard deviation of a total of 5 runs for all subjects.

k	top-1 accuracy (%)		top-3 accuracy (%)	
	after f	after g	after f	after g
1	$31.7_{\pm 1.5}$	$\mathbf{32.4}_{\pm 0.7}$	$53.6_{\pm 1.9}$	$\mathbf{54.8}_{\pm 1.1}$
2	$\mathbf{46.3}_{\pm 1.8}$	$46.0_{\pm 1.8}$	$\mathbf{70.2}_{\pm 1.6}$	$69.6_{\pm 1.9}$
3	$\mathbf{58.9}_{\pm 0.7}$	$58.3_{\pm 1.3}$	$\mathbf{80.8}_{\pm 1.2}$	$80.4_{\pm 1.3}$
4	$\mathbf{67.5}_{\pm 1.2}$	$65.6_{\pm 1.5}$	$\mathbf{86.8}_{\pm 1.2}$	$86.0_{\pm 0.9}$
5	$\mathbf{73.5}_{\pm 1.1}$	$72.3_{\pm 1.3}$	$\mathbf{90.0}_{\pm 1.0}$	$89.7_{\pm 0.7}$

k leads to performance improvements for all the methods. On the other hand, it can be also noticed that regardless of the choice of k, our method consistently outperforms the competitors with non-trivial margins, indicating the efficacy and the generality of our method. Meanwhile, the performance gaps between the methods get smaller as k grows, since the benefit of source datasets vanishes as the volume of the target dataset increases. We note, however, that a large value of k is impractical and sometimes even unreachable in the real-world setting.

5.3 Analysis on the location of feature adaptation

Our feature adaptation with the discrepancy loss (Eq. 3) can be adopted into any layer of the model. To analyze the effect of its location, we compare two variants that minimize the distance of feature distributions after the sequence encoder f and the embedding layer g, respectively. The results are shown in Table 4, where the variant "after f" generally shows better performance compared to "after g" except for the case where k is set to 1. We conjecture that this is because it is incapable for a single GRU encoder (*i.e.*, f) to align feature distributions from different subjects well when the amount of the target dataset is too small. However, with a sufficiently large k, the variant "after f" consistently performs better with obvious margins. Based on these results, we compute the MK-MMD on the features after the sequential encoder f by default.

6 Concluding Remarks

In this paper, we introduce a new setting for EEG-based visual recognition, namely *subject adaptive EEG-based visual recognition*, where plentiful data from source subjects and sparse samples from a target subject are provided for training. This setting is cost-effective and practical in that it is often infeasible to acquire sufficient samples for a new user in the real-world scenario. Moreover, to better exploit the abundant source data, we introduce a strong baseline that minimizes the feature discrepancy between different subjects. In the experiments with various settings, we clearly verify the effectiveness of our method compared to the vanilla model. We hope this work would trigger further research under realistic scenarios with data scarcity, such as subject generalization [10,18].

Acknowledgments

This work was supported by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00451: Development of BCI based Brain and Cognitive Computing Technology for Recognizing Users Intentions using Deep Learning, No. 2020-0-01361: Artificial Intelligence Graduate School Program (YONSEI UNIVERSITY)).

References

1. An, W.W., Shinn-Cunningham, B., Gamper, H., Emmanouilidou, D., Johnston, D., Jalobeanu, M., Cutrell, E., Wilson, A., Chiang, K.J., Tashev, I.: Decoding music attention from "eeg headphones": A user-friendly auditory brain-computer interface. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 985–989. IEEE (2021)
2. Anuragi, A., Sisodia, D.S.: Alcohol use disorder detection using eeg signal features and flexible analytical wavelet transform. Biomedical Signal Processing and Control **52**, 384–393 (2019)
3. Bos, D.O., Reuderink, B.: Brainbasher: a bci game. In: Extended Abstracts of the International Conference on Fun and Games. pp. 36–39. Eindhoven University of Technology Eindhoven, The Netherlands (2008)
4. Chambayil, B., Singla, R., Jha, R.: Virtual keyboard bci using eye blinks in eeg. In: 2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications. pp. 466–470. IEEE (2010)
5. Dai, M., Zheng, D., Na, R., Wang, S., Zhang, S.: Eeg classification of motor imagery using a novel deep learning framework. Sensors **19**(3), 551 (2019)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
7. Eidel, M., Kübler, A.: Wheelchair control in a virtual environment by healthy participants using a p300-bci based on tactile stimulation: training effects and usability. Frontiers in Human Neuroscience **14** (2020)
8. Eroğlu, K., Kayıkçıoğlu, T., Osman, O.: Effect of brightness of visual stimuli on eeg signals. Behavioural Brain Research **382**, 112486 (2020). https://doi.org/https://doi.org/10.1016/j.bbr.2020.112486, `https://www.sciencedirect.com/science/article/pii/S0166432819313038`
9. Foxe, J.J., Simpson, G.V., Ahlfors, S.P.: Parieto-occipital 1 0hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport **9**(17), 3929–3933 (1998)
10. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D.: Domain generalization for object recognition with multi-task autoencoders. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2551–2559 (2015)
11. Grizou, J., Iturrate, I., Montesano, L., Oudeyer, P.Y., Lopes, M.: Calibration-free bci based control. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 28 (2014)

12. Huang, H., Xie, Q., Pan, J., He, Y., Wen, Z., Yu, R., Li, Y.: An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness. IEEE Transactions on Affective Computing (2019)
13. Hwang, S., Hong, K., Son, G., Byun, H.: Ezsl-gan: Eeg-based zero-shot learning approach using a generative adversarial network. In: 2019 7th International Winter Conference on Brain-Computer Interface (BCI). pp. 1–4 (2019). https://doi.org/10.1109/IWW-BCI.2019.8737322
14. Hwang, S., Hong, K., Son, G., Byun, H.: Learning cnn features from de features for eeg-based emotion recognition. Pattern Analysis and Applications **23**(3), 1323–1335 (2020)
15. Hwang, S., Ki, M., Hong, K., Byun, H.: Subject-independent eeg-based emotion recognition using adversarial learning. In: 2020 8th International Winter Conference on Brain-Computer Interface (BCI). pp. 1–4 (2020). https://doi.org/10.1109/BCI48061.2020.9061624
16. Hwang, S., Lee, P., Park, S., Byun, H.: Learning subject-independent representation for eeg-based drowsy driving detection. In: 2021 9th International Winter Conference on Brain-Computer Interface (BCI). pp. 1–3 (2021). https://doi.org/10.1109/BCI51272.2021.9385364
17. Hwang, S., Park, S., Kim, D., Lee, J., Byun, H.: Mitigating inter-subject brain signal variability for eeg-based driver fatigue state classification. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 990–994 (2021). https://doi.org/10.1109/ICASSP39728.2021.9414613
18. Jeon, S., Hong, K., Lee, P., Lee, J., Byun, H.: Feature stylization and domain-aware contrastive learning for domain generalization. In: Proceedings of the 29th ACM International Conference on Multimedia (2021)
19. Jin, Z., Zhou, G., Gao, D., Zhang, Y.: Eeg classification using sparse bayesian extreme learning machine for brain–computer interface. Neural Computing and Applications **32**(11), 6601–6609 (2020)
20. Kavasidis, I., Palazzo, S., Spampinato, C., Giordano, D., Shah, M.: Brain2image: Converting brain signals into images. In: Proceedings of the 25th ACM International Conference on Multimedia. pp. 1809–1817 (2017)
21. Khurana, V., Gahalawat, M., Kumar, P., Roy, P.P., Dogra, D.P., Scheme, E., Soleymani, M.: A survey on neuromarketing using eeg signals. IEEE Transactions on Cognitive and Developmental Systems (2021). https://doi.org/10.1109/TCDS.2021.3065200
22. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (2015)
23. Ko, W., Oh, K., Jeon, E., Suk, H.I.: Vignet: A deep convolutional neural network for eeg-based driver vigilance estimation. In: 2020 8th International Winter Conference on Brain-Computer Interface (BCI). pp. 1–3. IEEE (2020)
24. Kumar, P., Saini, R., Roy, P.P., Sahu, P.K., Dogra, D.P.: Envisioned speech recognition using eeg sensors. Personal and Ubiquitous Computing **22**(1), 185–199 (2018)
25. Lee, J., Won, K., Kwon, M., Jun, S.C., Ahn, M.: Cnn with large data achieves true zero-training in online p300 brain-computer interface. IEEE Access **8**, 74385–74400 (2020). https://doi.org/10.1109/ACCESS.2020.2988057
26. Lee, S.H., Lee, M., Lee, S.W.: Neural decoding of imagined speech and visual imagery as intuitive paradigms for bci communication. IEEE Transactions on Neural Systems and Rehabilitation Engineering **28**(12), 2647–2659 (2020)
27. Li, M., Li, F., Pan, J., Zhang, D., Zhao, S., Li, J., Wang, F.: The mindgomoku: An online p300 bci game based on bayesian deep learning. Sensors **21**(5), 1613 (2021)

28. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning. p. 97–105 (2015)
29. Mahato, S., Paul, S.: Detection of major depressive disorder using linear and non-linear features from eeg signals. Microsystem Technologies **25**(3), 1065–1076 (2019)
30. Placidi, G., Di Giamberardino, P., Petracca, A., Spezialetti, M., Iacoviello, D.: Classification of emotional signals from the deap dataset. In: International congress on neurotechnology, electronics and informatics. vol. 2, pp. 15–21. SCITEPRESS (2016)
31. Qin, W., Yu, C.: Neural pathways conveying novisual information to the visual cortex. Neural plasticity **2013** (2013)
32. Ramsey, N.F., Van De Heuvel, M.P., Kho, K.H., Leijten, F.S.: Towards human bci applications based on cognitive brain systems: an investigation of neural signals recorded from the dorsolateral prefrontal cortex. IEEE Transactions on Neural Systems and Rehabilitation Engineering **14**(2), 214–217 (2006)
33. Runnova, A., Selskii, A., Kiselev, A., Shamionov, R., Parsamyan, R., Zhuravlev, M.: Changes in eeg alpha activity during attention control in patients: Association with sleep disorders. Journal of Personalized Medicine **11**(7), 601 (2021)
34. Rutkowski, T.M., Koculak, M., Abe, M.S., Otake-Matsuura, M.: Brain correlates of task–load and dementia elucidation with tensor machine learning using oddball bci paradigm. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8578–8582. IEEE (2019)
35. Salenius, S., Kajola, M., Thompson, W., Kosslyn, S., Hari, R.: Reactivity of magnetic parieto-occipital alpha rhythm during visual imagery. Electroencephalography and Clinical Neurophysiology **95**(6), 453–462 (1995). https://doi.org/https://doi.org/10.1016/0013-4694(95)00155-7, `https://www.sciencedirect.com/science/article/pii/0013469495001557`
36. Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with convolutional neural networks for eeg decoding and visualization. Human brain mapping **38**(11), 5391–5420 (2017)
37. Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N., Shah, M.: Deep learning human mind for automated visual classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6809–6817 (2017)
38. Stewart, A.X., Nuthmann, A., Sanguinetti, G.: Single-trial classification of eeg in a visual object task using ica and machine learning. Journal of Neuroscience Methods **228**, 1–14 (2014). https://doi.org/https://doi.org/10.1016/j.jneumeth.2014.02.014, `https://www.sciencedirect.com/science/article/pii/S0165027014000752`
39. Vert, J.P., Tsuda, K., Schölkopf, B.: A primer on kernel methods. Kernel methods in computational biology **47**, 35–70 (2004)
40. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., Vaughan, T.M., et al.: Brain-computer interface technology: a review of the first international meeting. IEEE transactions on rehabilitation engineering **8**(2), 164–173 (2000)
41. Yuan, Y., Xun, G., Jia, K., Zhang, A.: A multi-view deep learning framework for eeg seizure detection. IEEE journal of biomedical and health informatics **23**(1), 83–94 (2018)

42. Zhang, D., Yao, L., Chen, K., Monaghan, J.: A convolutional recurrent attention model for subject-independent eeg signal analysis. IEEE Signal Processing Letters **26**(5), 715–719 (2019). https://doi.org/10.1109/LSP.2019.2906824
43. Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H., Xiang, J.: Epileptic seizure detection based on eeg signals and cnn. Frontiers in neuroinformatics **12**, 95 (2018)
44. Zickler, C., Di Donna, V., Kaiser, V., Al-Khodairy, A., Kleih, S., Kübler, A., Malavasi, M., Mattia, D., Mongardi, S., Neuper, C., et al.: Bci applications for people with disabilities: defining user needs and user requirements. Assistive technology from adapted equipment to inclusive environments, AAATE **25**, 185–189 (2009)