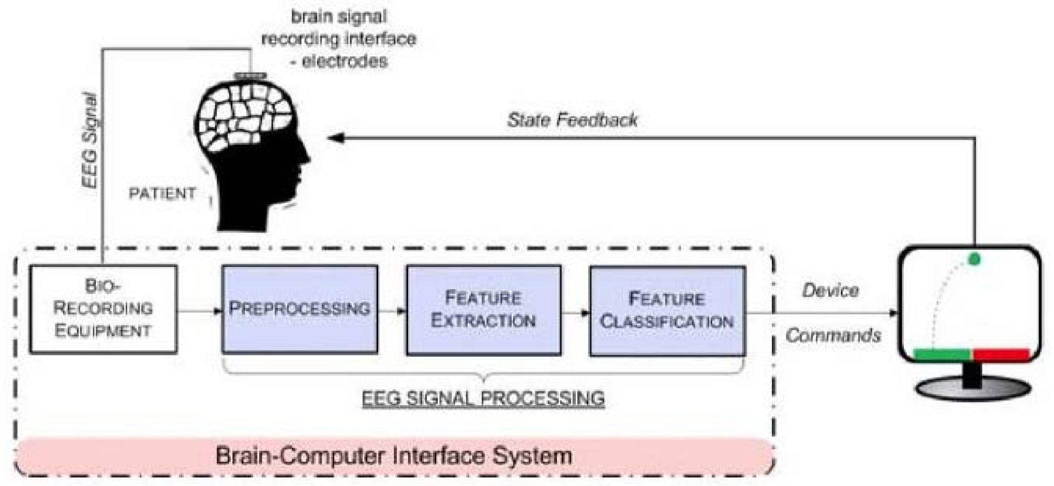
IEEE ICASSP 2018
Calgary, Alberta, Canada
15–20 April 2018

DEEP TRANSFER LEARNING FOR EEG-BASED BRAIN COMPUTER INTERFACE

Chuanqi Tan, Fuchun Sun, and Wenchang Zhang

State Key Laboratory of Intelligent Technology and Systems. Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Computer Science and Technology, Tsinghua University, 10084, China
fcsun@tsinghua.edu.cn, {tcq15, zhangwc14}@mails.tsinghua.edu.cn



INTRODUCTION

•The biggest challenge and most important component in a BCI system is the EEG classifier, which translates a raw EEG signal into the commands of the human brain.

- First, traditional EEG classification methods focus on frequency-domain information and cannot fully exploit multimodal information.
- Second, high-quality, large-scale annotated EEG datasets are extremely difficult to construct because biological data acquisition is challenging and quality annotation is costly.

•We solved these problems in the following ways.

- First, we modeled cognitive events based on EEG data by characterizing the data using EEG optical flow, which is designed to preserve multimodal EEG information.
- Second, we designed a deep transfer learning framework suitable for transferring knowledge by joint training, which contains an adversarial network and a special loss function.

REPRESENTATION

•Traditional methods **do not fully exploit multimodal information**. For example, they ignore the locations of the electrodes and the inherent information in the spatial dimension.

•In our approach, we convert raw EEG signals into EEG optical flow to represent the multimodal information of EEG.

- **3D locations -> 2D projected points -> EEG video -> EEG optical flow**
- We project the 3D locations of the electrodes to 2D points via azimuthal equidistant projection (AEP).
- Interpolate them to gray image by clough-tocher algorithm.
- Optical flow is introduced in our approach to describe the variant information of the EEG signal.

Consider $f(x, y, t)$ as a pixel at position (x, y) at time t, which moves $(\Delta x, \Delta y)$ in the next frame

$$f(x, y, t) = f(x + \Delta x, y + \Delta y, t + \Delta t)$$

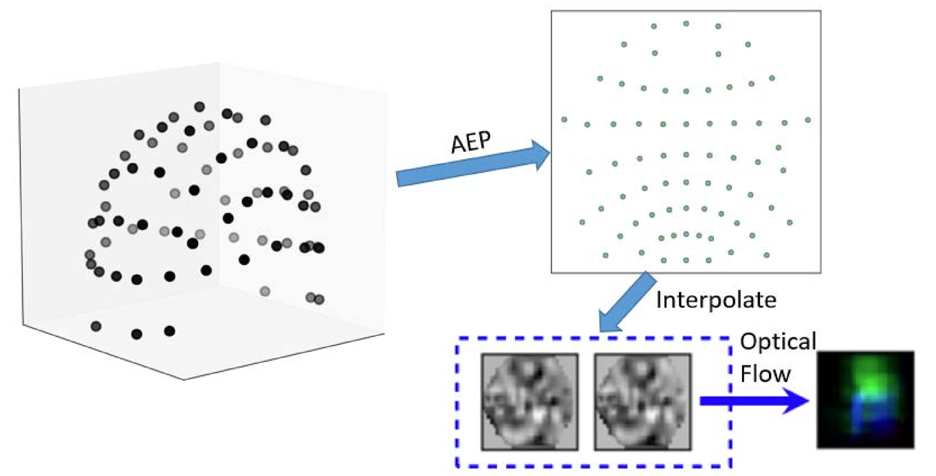
Performing Taylor series approximation on the right-hand side and ignoring the higher-order terms in the Taylor series, we can obtain the following equation:

$$f(x+\Delta x, y+\Delta y, t+\Delta t) = f(x,y,t) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \frac{\partial f}{\partial t}\Delta t + \dots$$

Then, removing the common terms and dividing by Δt, we can obtain:

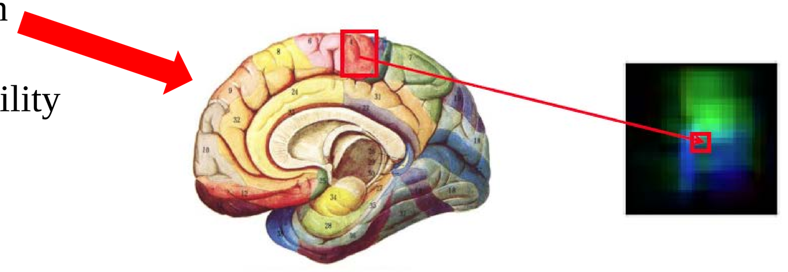
$$\frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial f}{\partial t}\frac{\Delta t}{\Delta t} = \frac{\partial f}{\partial x}u + \frac{\partial f}{\partial y}v + \frac{\partial f}{\partial t} = 0$$

In this equation, (u, v) is the value of the optical flow at $f(x, y, t)$



•Many benefits can be gained from using the EEG optical flow:

- Uniform representation of multimodal information
- More suitable for CNN
- Locality information
- Transfer learning ability



DEEP TRANSFER LEARNING

•**Insufficient training data** is a serious problem in all domains related to bioinformatics. Our answer to this problem is transfer learning, which transfers knowledge from computer vision.

•We construct a **deep transfer learning framework** that contains two steps to obtain the final EEG category labels:

- Joint Training
- EEG Classification

Joint training aimed to learn a better representation for natural images and EEG optical flow.

- Many studies have demonstrated that front layers in an CNN network can extract the general features of images, such as edges and corners.
- Parameters of the transfer network were pretrained on large-scale dataset like ImageNet and transfer to EEG classification network in directly **(shown in the red box)**.
- But the general feature extractor trained by natural images does not fully match the EEG optical flow.

Adversarial network: Inspired by generative adversarial nets (GAN), we apply an adversarial network **(shown in the green box)** to train a better general feature extractor.

- In order to achieve a better effect of transfer, the edge distribution of features from the source domain and target domain should be as similar as possible.
- We use features extracted from natural images and EEG optical flow as the inputs for the adversarial network and train it to identify their origins. If the adversarial network achieves worse performance, it means a small difference between the two types of feature and better transferability, and vice versa.
- We achieved this goal by a special loss function and a iteratively optimizing algorithm.
  - loss function

$$\mathcal{L}_{img} = -\sum_k \mathbb{I}[y=k]\log p_k + \alpha\mathcal{L}_{adver}$$

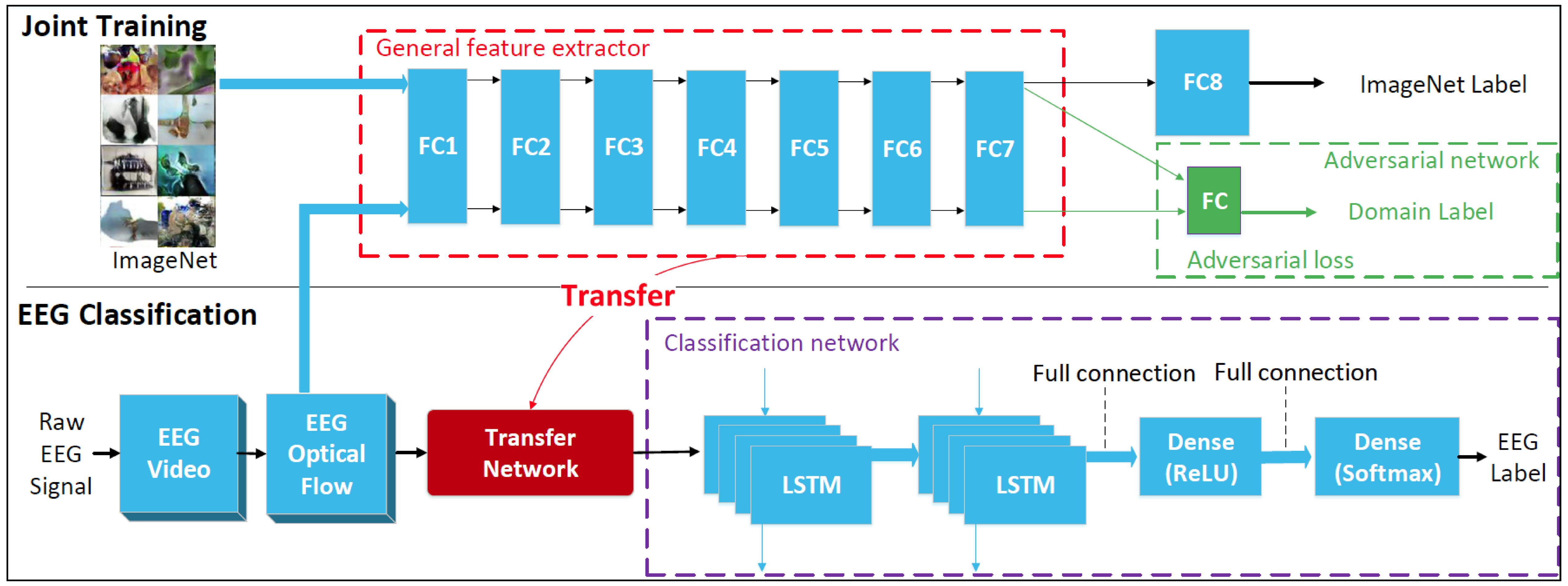
$$\mathcal{L}_{adver} = -\sum_d \frac{1}{D}\log p_d$$

  - iteratively optimizing

$$\operatorname*{argmin}_{\theta_{extr}\theta_{img}} \mathcal{L}_{img}(X_{img}, X_{of}, \theta_{adver}; \theta_{extr}, \theta_{img})$$

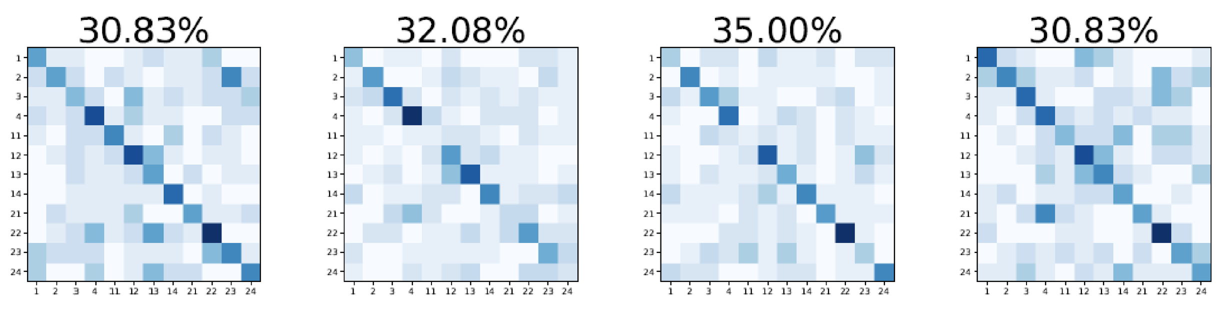
$$\operatorname*{argmin}_{\theta_{extr}\theta_{adver}} \mathcal{L}_{adver}(X_{img}, X_{of}, \theta_{img}; \theta_{extr}, \theta_{adver})$$

EEG classification aimed to obtain the final EEG label. General features are fed to a classification network (shown in the purple box) with two RNN layers and two fully connected layers.



EXPERIMENTS

We apply our approach to a famous public dataset in the BCI field, named Open Music Imagery Information Retrieval.



(a) AlexNet (b) VGG16 (c) VGG19 (d) ResNet

Classification accuracy while further reducing the size of the training dataset. N% in the table head indicate the percentage of training dataset were used during training. The values in parentheses indicate the classification results without joint training.

	100%	50%	25%
SVC	23.1	16.69	9.83
DNN	27.22	20.83	12.47
CNN	27.8	19.2	8.55
AlexNet	30.83(27.92)	24.58(21.72)	**15.42**(15.27)
VGG16	32.08(31.67)	**28.33**(26.2)	**15.42**(14.7)
VGG19	**35.0**(32.92)	25.83(23.14)	14.58(13.93)
ResNet	30.83(27.08)	22.5(22.67)	11.25(10.42)

DISSCUSSION

We can draw the following conclusions from the experimental results:

- Our proposed approach achieves accuracy that is obviously superior to that of the traditional methods;
- VGG16 and VGG19 are good choices of transfer network;
- Our approach can achieve acceptable results while further reducing the size of the training set;
- Joint training play a important and positive role in the final results.

CONCLUSIONS

- In this paper, we propose a novel EEG signal classification approach with EEG optical flow and deep transfer learning in response to two major problems in EEG classification: (1) the inability of traditional methods to fully exploit multimodal information and (2) insufficient training data.
- Our approach is superior to other state-of-the-art methods, which is important for building better BCI systems, and provides a new perspective for solving the problem of EEG classification.
- In addition, our approach can be viewed as a general bioelectrical signal classification framework that is suitable for other bioelectrical signals, such as functional magnetic resonance imaging (fMRI).