A Coincidence Filtering-based Approach for CNNs in EEG-based Recognition

Zhongke Gao, *Senior Member, IEEE,* Yanli Li, Yuxuan Yang, Na Dong, Xiong Yang, and Celso Grebogi

***Abstract*—Electroencephalogram (EEG), obtained by wearable devices, can realize effective human health monitoring. Traditional methods based on artificially-designed features have achieved valid results in EEG-based recognition, and numerous studies start to apply deep learning techniques in this area. In this paper, we propose a coincidence filtering-based method to build a connection between artificial features-based methods and convolutional neural networks (CNNs), and design CNNs through simulating the information extraction pattern of artificial features-based methods. Based on this method, we propose a novel, simple, and effective CNNs structure for EEG-based classification. We implement two experiments to obtain EEG data, and perform experiments based on the two health monitoring tasks. The results illustrate that the proposed network can achieve a prominent average accuracy on the emotion recognition and fatigue driving detection task. Due to its generality, the proposed framework design of CNNs is expected to be useful for broader applications in health monitoring areas.**

***Index Terms*—Convolutional neural networks, Electroencephalogram (EEG), emotion recognition, fatigue driving detection**

I. Introduction

HUMAN psychological states, which are usually evoked by external stimuli, have a great impact on human daily activities. Positive emotion or negative emotion, alert and fatigue, these psychological states could greatly affect a person's physiological state [1]. These have triggered the emergence of affective computing (AC), fatigue driving detection, and other important research directions of human-computer interaction systems. Researchers from different fields have used different signals or clues to conduct the analysis of psychological state, such as facial expressions, speech signals, text messages, electrocardiogram (ECG) data [2], and EEG data [3], [4]. Among these clues, EEG signals are less likely to be compromised by individual human characteristics, such as economy, lifestyle, cultural background, and even the intent to defraud the sensors. Furthermore, EEG signals are directly related with the most concentrated organ of the human body information system: the brain. These advantages have made EEG signals attractive to researchers. Unfortunately, limited by EEG signals acquisition technology, the noise of EEG signals from the wearable brain caps is relatively high. EEG signals have low signal-to-noise ratio (SNR) due to the noise presence in the channels, which rather restricts the analysis and processing methods.

Many studies have analyzed, processed and classified EEG-based tasks through artificially-designed features, including time-frequency analysis [5] , Bayesian methods [6], complex network methods [7], and other nonlinear methods. Some artificially-designed features become the basis for many subsequent studies, and have been expanded to other EEG-based areas. Traditional features, such as differential entropy (DE) and power spectral density (PSD) features, have shown valid capabilities in emotion recognition and fatigue driving detection area [8]. However, these artificial feature methods are time-consuming to design proper feature for new kinds of signals, and subject to some limitations due to the nature of the data.

In recent years, convolutional neural networks (CNNs) have demonstrated impressive capabilities for data classification and feature extraction. For example, CNNs have been implemented for epilepsy prediction and monitoring [9], for detection of visual-evoked responses [10], for motor imagery classification [11], fatigue driving evaluation [12], [13], and emotion recognition [14], [15]. These studies have inspired many attempts to further improve the outcomes of EEG-based tasks through CNNs. However, finding a CNN structure that performs better in the new field is a challenging task. Common ideas for processing EEG-based signals with CNNs are: (1)directly use the CNN structures that are applied image and signals processing; (2)take artificially-designed features as the input of CNNs. These two common ideas are simple and effective in building a network model for EEG-based recognition, but still face some problems. First, among the CNNs architectures apply to image processing [16], the common features are the depth of the structure and the stereoscopic nature of the feature receptive field. These features are mainly aimed at the complex and multi-level features of the images. For EEG-based signal processing, the structure depth above the demand may lead to the reduction of model training efficiency. Second, for architectures from other signal-related works, which have achieved great performance in their own fields, the generalization on a new kind of EEG signals still needs to be further researched [17]. Third, some works have achieved results through building CNNs-based on artificial features instead of original dataset [18], but these approaches abandon the most important advantage of a CNN

This work was supported by National Natural Science Foundation of China under Grant Nos. 61873181, 61922062 and 61773282.

Z. Gao is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zhongkegao@tju.edu.cn).

N. Dong is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: dongna@tju.edu.cn).

Y. Li, Y. Yang, and X. Yang are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China(e-mail: casareli@tju.edu.cn).

C. Grebogi is with Institute for Complex Systems and Mathematical Biology, Kings College, University of Aberdeen, Old Aberdeen AB24 3UE, UK (e-mail: grebogi@abdn.ac.uk).

model: the feature extraction capability.

Artificial features usually revolves around the experts cognition of the data, with significant human components. These can enhance targeted information, leading to ignore other information. Traditional artificial features effectively filter out irrelevant information, but lose other relevant information, while neural networks retain most of the unnecessary information, both have their merits and weaknesses. Therefore, one idea is to use the artificial features to optimize the feature extraction capability of neural networks, thus retaining the advantages of neural networks and reducing noise emergency interference. For example, some DE-based methods have achieved state of art results in emotion recognition tasks and PSD-based methods are widely used in signal processing areas, which make these two features representative [19], [20]. These results illustrate the effectiveness and potential of artificial features. The remaining problem is how to reproduce and extend these feature extraction methods within neural networks, instead of directly input them. That means, we can retain the thoughts of artificial features extraction and use convolutional neural networks to simulate the thoughts instead of designing new artificial features.

In this paper we propose a method based on coincidence filtering for designing and optimizing the CNNs of EEG-based signals. This method presents a novel CNNs model which has a simple structure, with fewer parameters to tune, higher training efficiency, and faster convergence than deep CNNs. Compared to artificial features, this CNNs model retains the feature extraction capabilities of convolutional layers. Besides, compared to existing CNNs methods, this CNNs model with channel average pooling layers can provide higher training efficiency and accuracy on certain EEG-based health monitoring tasks, and the construction process of this model provides a way of thinking to design the network structure.

To validate the proposed CNN structure, we apply the same CNN structure with little adjustment into two different EEG-based tasks: emotion classification task and fatigue driving detection task. Emotion recognition tasks focus on the recognition of human emotions, which are an important part of the human-computer interaction system and part of the field of affective computing. Many researchers have made contributions to the task of emotion recognition based on EEG signals, such as group sparse canonical correlation analysis [21], deep belief network based on differential entropy [20], bimodal deep auto-encoder [22]. The mental fatigue refers to a physiological and psychological condition accompanied with lessened alertness and decremental integrated performance. Fatigue driving devotes to a conservative estimate of above 10000 deaths, which is a great threat to human safety. The detection of the fatigue driving state in driving is beneficial to reduce the traffic accident and has attracted considerable attention [12], [23], [24]. These two tasks are important application of EEG-based recognition, which can prove the the effectiveness and generality of our method in EEG signals.

The layout of the article is organized as follows: in section II, we explain the effective artificial features, the way of combining CNNs with artificial features through coincidence filtering, and introduce our architecture. Then, the section III briefly describes our experiment design and data acquisition. Section IV provides the results obtained by our method on these datasets, and the advantages over other works. We also make some comparisons to validate our method in section IV, and make some brief discussions in section V to conclude the paper.

II. Methods

A. Artificial features and its inspiration

DE and PSD are commonly used artificial features for processing EEG-based recognition before deep learning technology. Of the two features, DE-based methods have achieved valid results in emotion recognition tasks, proved to be the most effective artificial feature in this area, and also achieved good performance in fatigue driving detection tasks. PSD is widely applied in most of EEG-based tasks, considered to be one of the most commonly used features. Some other artificial features, which are applied to these two areas are relatively deficient in accuracy or versatility compared to these two features, and some features require unique pre- or post-processing with complex methods combination. Therefore, we take these two representative features for analysis.

DE is referred to as the extension of the Shannon entropy, whose calculation formula can be expressed as:

$$h(x) = -\sum_{i=1}^{N} p(x_i) log(p(x_i)) \tag{1}$$

where x is a series of random variable that contains event x_i, $h(x)$ is the Shannon entropy, and the $p(x_i)$ the probability of x_i.

DE is the continuous version of Shannon entropy so the original computing formulation can be expressed as:

$$h(X) = -\int_X f(x)\log(f(x))dx \tag{2}$$

where X is a random variable and $f(x)$ is the probability density function of X.

For a time series X obeying the Gauss distribution $N(\mu, \sigma^2)$, the calculation formula can be simplified as [20]:

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}})dx \\ &= \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned} \tag{3}$$

The original EEG signals do not follow a certain fixed distribution, but are subject to a distribution when they are divided into a series of sub-bands after band-pass filtering. The PSD calculates the mean square value in the unit frequency band, usually uses the Fourier transform to convert the original signal into a frequency domain.

We believe that DE-based and PSD-based methods can provide an ideological guidance for the construction of CNNs, which can help focusing on the direction that counts, and giving up noise and irrelevant information. What we can learn from the features are as follows: (1)

1) DE and PSD adopt the method of extracting features in different frequency bands.
2) Within the channel, changes in the temporal direction carry the main information. In the past, the feature extraction is mostly performed in the single-channel temporal direction.
3) DE and PSD usually only obtain a single eigenvalue on the whole channel, so it is very likely that there is no internal difference in the information on the entire channel. The length of time mainly affects the amount of information and the resulting robustness against noise and disturbances.

B. Coincidence Filtering

Coincidence filtering, derived from a simple statistical classification idea, is a method of quantifying the similarity between the target data and the reference object through randomly generated filter clusters. Compared to other simple filtering mechanisms, coincidence filtering focuses manly on the number of individuals who have achieved high evaluation.

In simple terms, when an individual in a region reaches a certain similarity threshold in relation to the reference, we declare that the individual exhibits characteristic A. However, no matter how high the similarity the individual achieves, we refuse to admit that the region reflects the characteristic A. Conversely, if the majority in the region reflects the characteristic A, even if the individuals have just reached the threshold, we admit that the region exhibits the characteristic A. The core framework of coincidence filtering is: (1) to establish a standard for quantifying the similarity between coincidence reference and original data through randomly generated filter clusters, and (2) to preset a default threshold for determining the tendency of an individual or a whole region. The former can be thought of as a filter whose output is positively correlated with the similarity of those two elements (original data and the coincidence reference), while the latter is a classifier that associates the output of the filter with the meaning of the representation.

C. The relationship between CNNs functional layers and artifically-designed features

For EEG signals, the signal distribution in the spatial direction should have a practical meaning, namely, the EEG signal acquisition channels, which except for the number of data dimension, are the main differences between EEG signals and images. The extraction of artificial features is also performed separately on a single channel, so it is standard to focus mainly on how to acquire and use filters in the time direction. On one hand, some studies have suggested that a particular convolution kernel with the form of $(1, n)$ has its corresponding frequency filtering bands [25], and this relationship can be expressed as:

$$|W_i(f)|^2 = \left| \sum_{k=1}^{N_t} w_{ik} e^{-j2\pi f(k-1)} \right|^2, \quad (4)$$

where $W_i(f)$ is the frequency, N_t is the length of the temporal filter and w_{ik} is the kth filter coefficient of the ith filter.

The two-dimensional $(1, n)$ convolution can be regarded as a form of filtering in the time dimension, which is exactly the extraction direction of artificial features. Therefore, a series of $2D$ $(1, n)$ convolution kernels can realize two important characteristics of the artificial features. Since the convolution kernel initialization is randomly generated, the enough number of convolution kernels is very important. In order to balance the calculation efficiency and calculation effect, we set a suitable number of convolution kernels via trial and error.

As aforementioned, DE and PSD usually yield a single eigenvalue on the whole channel, which indicates that there is no internal difference in the information on the entire channel. These inspire us to concentrate more on the generality, rather than on the maximum, of the filtered high values. In CNNs, this concentration is reflected in the abandonment of multiple convolutions on the local precise judgment, and we give up the adoption of information on each pixel of the feature map. Instead, we reduce the number of convolution layers and treat the entire channel as a basic evaluation unit.

By considering randomly generated convolution kernels as coincidence filters, we quantify the similarity to the unknown target. The remaining problem is to set the similarity threshold, and realize the frequency filters to ensure that the quantity of the pixels that exceed the threshold, determine the evaluation of the whole unit. In general, using the mean value of a region instead of a threshold for the frequency of good evaluations may result in classification errors due to the presence of extremely high-energy pixels. However, for EEG signals, the long total length of the data leads to the attenuation of the influence of extreme pixels after averaging. Therefore the mean value could represent the tendency or the level of the entire region in a sense. Besides, using the mean value instead, it is possible to circumvent the preset threshold and reduce the amount of parameters. Figure 1 provides the flow chart of proposed method applied in the network.

D. Coincidence filtering-based CNNs (CF-CNN)

Figure 2 shows the pipeline of the developed method for EEG-based recognition. Convolution is only operated on the temporal direction, the first layer for frequency division, and the second layer for coincidence filtering. According to previous inferences, in order to reduce noise interference, an average pooling layer is then used for each channel, called channel average pooling (CAP) layer, to obtain a convolution result for each coincidence. After that, two fully-connected layers are set as classifiers to classify the obtained feature values.

For more details, it is conducive to judge the situation of the entire sample by lowering the quantity of convolution layers, reducing the accuracy of every single pixel judgement, and by increasing the amount of pixel judgments. So, we only use two layers of convolution. CAP is used, instead of other forms

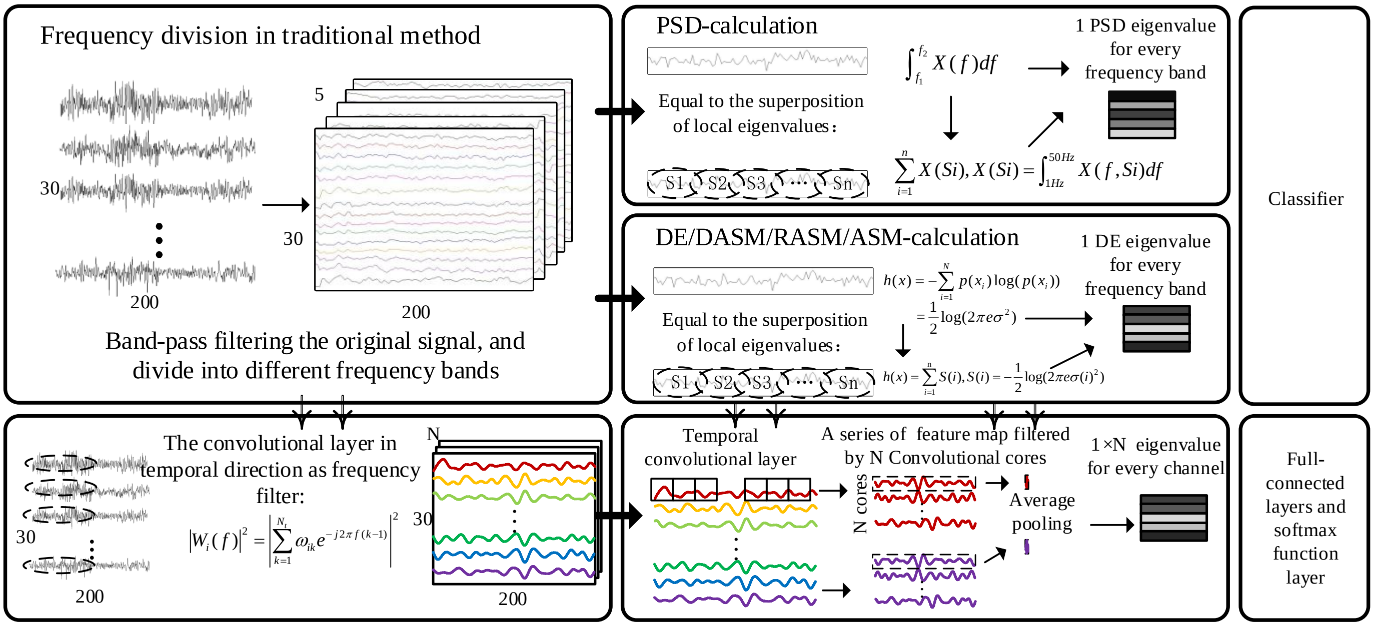


Fig. 1. The flow chart of proposed method applied in the network.

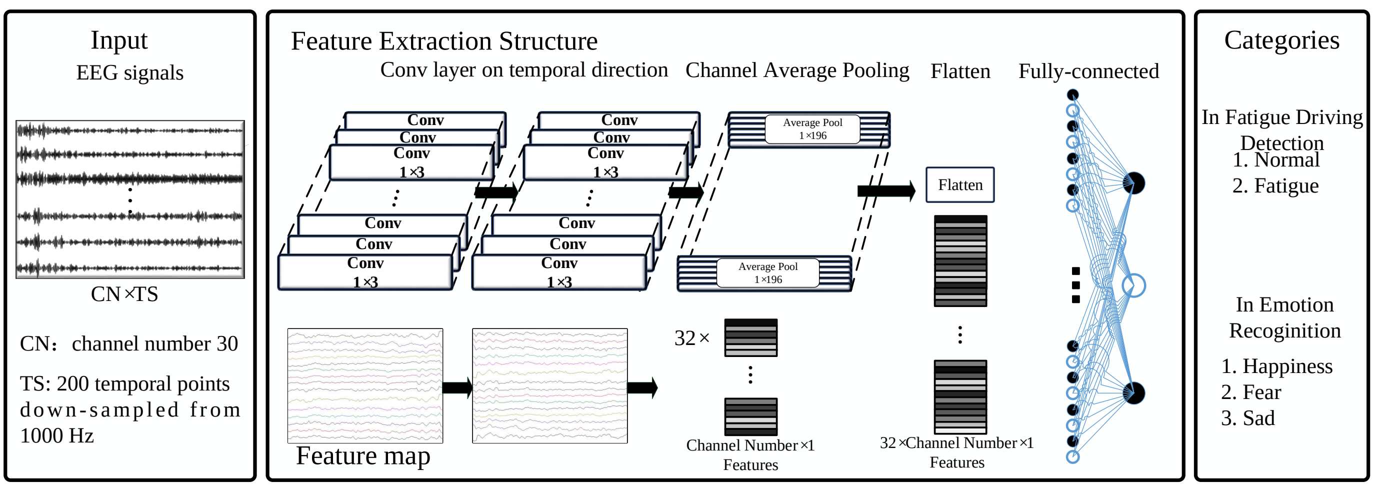


Fig. 2. The flow chart of the proposed model including the structure of CF-CNN.

of pooling, to simulate the information extraction pattern of the artificial feature method, using the entire channel as an evaluation region. Numerous studies have revealed that mental states such as emotion or fatigue produce different results in different regions of the brain, leading to significant differences among EEG channels. Because of the low signal-to-noise ratio, the differences among channels can greatly influence the process of network learning. This mechanism inspires us to preserve the uniqueness of the channel rather than simply using global averaging pooling, and the channel electrodes are arranged according to international 10-20 electrode placement system. In addition, we put a CAP layer between convolution layers and a fully-connected layer, rather than just use a fully-connected layer directly behind the convolution layers. Our decision is based on the following two reasons. (1) Through the CAP layer, the whole sample can be divided into individual channel regions, and the evaluation value of each region can be averaged for reducing the influence of extreme pixels and noise. (2) CAP layer can reduce the amount of parameters as well as the computational burden, and further accelerate the model convergence.

Finally, we can obtain a structure composed of one convolution layer with 32 kernels of form $(1, 3)$, stride size $(1, 1)$, another convolution layer with 32 kernels of form $(1, 3)$, stride size $(1, 3)$, then one CAP layer, and two fully-connected layers with 64 and 3/2 nodes, respectively. Two layers of convolutional layers (one for frequency division and one for feature extraction) with channel pooling layer are enough to form a coincidence filtering function unit. From the perspective of computational complexity, when we input 30*200 samples, the total number of parameters of the network structure with the CAP layer is two orders of magnitude smaller than that of the usual network structures. Through this structure, we can greatly reduce the

Fig. 3. The experimental scene of emotion recognition.

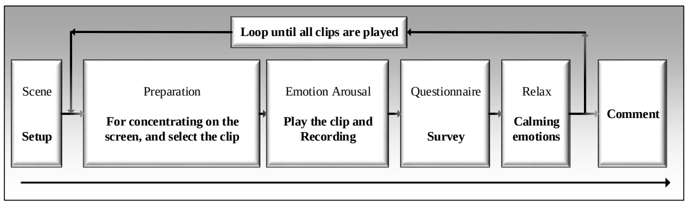


Fig. 4. The flow chart of emotion arousal experiments.

computational complexity of back-propagation, improve the computational efficiency and reduce the risk of over-fitting to some extent. Besides, after experiments for searching suitable number of convolutional layers, we find that more convolutional layers will not be helpful for improving the accuracy significantly, and will lead to a decrease in network operation efficiency. Therefore, we employ basic coincidence filtering functional unit. The input of the entire network structure is obtained by cutting the complete EEG data into 1s segments and down-sampling them to 200 Hz from 1000 Hz. The training iteration is set at 300,and we employ early-stop function to reduce the time cost of training process. The cross-entropy objective function is employed in the softmax layer to generate a distribution for each category. Input matrix is $X \in R^{T \times E}$, and matching labels is $Y \in R^N$, where T means the time direction or signal sampling points, E denotes the channels, and N is the categories of emotion states.

III. Experiment design and data acquisition

We design two different experiments to obtain EEG signals, the tasks are the emotion recognition task and fatigue driving detection task, including 15 and 10 volunteers, respectively. In this section we briefly introduce the two datasets.

A. Self-conducted emotion recognition task

Nine emotion-evoking film clips with audios and scenes are chosen as stimulus sources to evoke three specific emotions (fear, happiness and sadness). Each film clip lasts for 3 to 8 minutes, and each emotion corresponded to a certain number of clips to ensure that the total number of emotion clips is similar. Sixteen volunteers (11 male and 5 female), aged from 20 to 24 (mean: 22.14), participate in the experiments. All the subjects sign written consent prior to the recording, containing information about the design and purpose of the experiment. Besides, before the experiment, all of them are advised to follow the procedure and to refrain from unnecessary body movements while watching the movie clips. The EEG signals are collected by a 30-channel recording cap (ESI Neuroscan), and the sampling rate is 1000 Hz, but we downsample the signals to 200Hz in the preprocessing. All electrodes are arranged with standard 10-20 system. EEG signals are processed with a band pass filter of 1-50 Hz and with an Independent Component Analysis (ICA) for removing noise and artifacts. Besides, we collect face videos of subjects with a frontal camera.

Figure 4 provides the flow chart of emotion arousal experiments. For each subject, the clip projection durations of the experiments are equal and fixed. Figure 3 provides the experimental scene of the experiment.

EEG signals of each subject are cut into samples of 1 second without overlapping and the corresponding number of samples for each emotion are 855, 1110 and 945, respectively. To make the number of each category balanced, we randomly select 850 samples out of each category (total sample number 2550). The source, emotion type, and duration of each clip are shown in Table 1.

TABLE I
The sources of the selected movie clips and their corresponding emotional labels

Number	Film clip sources	Emotional label	Duration
1	Find miracle in cell No.7	Sadness	3 min 25 sec
2	Dearest	Sadness	4 min 30 sec
3	A hero or not	Happiness	5 min 30 sec
4	Mr. Bean	Happiness	7 min 20 sec
5	Mr. Bean	Happiness	5 min 52 sec
6	Dead silence	Fear	8 min 24 sec
7	The Conjuring 2	Fear	4 min
8	Lights out	Fear	3 min 25 sec

B. Self-conducted fatigue driving task

For further testing our method and verifying its generalization ability, we also built a dataset 2 based on our self-conducted fatigue driving experiment.

In the study, ten right-handed subjects(7 males and 3 female) without any disorders related to psychiatric aged from 22 to 26 (mean 23.2, std. 1.32) voluntarily join the experiment. The subjects are required to refrain from anti-fatigue drinks or medications before the experiment. Driving on a real highway while conducting another task is highly perilous for subjects and other drivers. Therefore, the study is performed in an indoor environment with a simple driving simulator PGFD001 equipped with a pedal, a steering wheel and a clutch. In the relevant driving software 3DInstructor2, a regular car Phaeton2.0L with automatic shifting is occupied, which forms a quite realistic driving conditions for the study. Trials

Fig. 5. The experimental scene of fatigue driving detection.

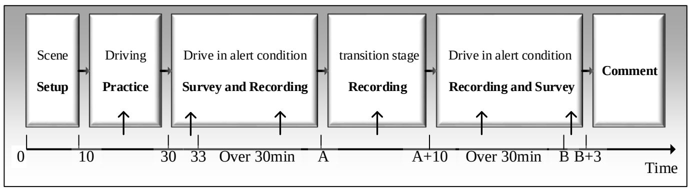


Fig. 6. The timeline of the fatigue driving experiment.

begin during 14:00 to 15:30 which is proved as an easy-trapped period for fatigue and one trial last approximately 80 minute. According to the 9-point Karolinska Sleepiness Scale (KSS) that assesses from 1(extremely alert) to 9 (very sleepy), drivers' fatigue state is divided into alert, mild fatigue and fatigue for assessing subjective fatigue. Figure 5 provides the experimental scene of the experiment, and the timeline of the experiment is shown in Figure 6.

In order to avoid interference caused by unclear fatigue conditions, and to make the number of each category balanced, we select 20 minutes of each subject's driving process as clarity and fatigue, respectively, based on the moment that subjects indicate suffering mild fatigue. The total sample number is 2400, and 1200 for each category.

IV. Results and analysis

We test our architecture on both emotion recognition task and fatigue driving. Although these tasks are distinct among objects and data materials, we apply the same architecture on both of them.

A. Overall performance

Both of the tasks have the same input format. Input samples are 30-channel data from 1s segments (200 Hz) and the label prediction is exported for the evaluation of each model. Figure 7 and 8 present the overall results on emotion recognition task, and Figure 9 and 10 presents the overall performance on fatigue driving detection task.

From Figure 7 and 9, we conclude that our method is stably effective on each task in subject-dependent test. We obtain the average result of 97.1% on emotion recognition task

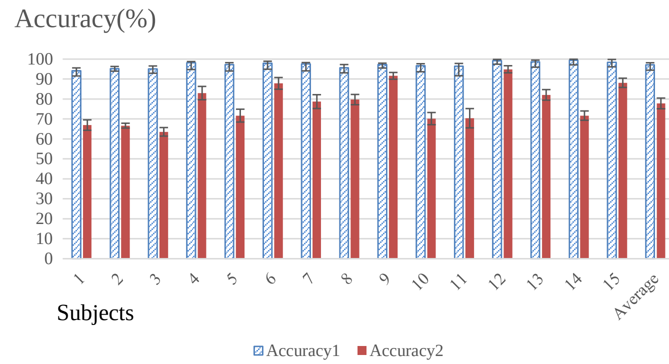


Fig. 7. The overall performance on emotion recognition task.

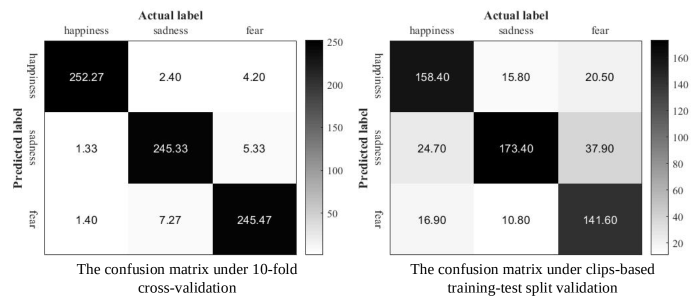


Fig. 8. The confusion matrix on emotion recognition task.

and 98.25% on fatigue driving under 10-fold cross validation evaluation (accuracy1). In addition, we derive the average result of 77.81% and 96.89% under the evaluation that the test set and training set are selected from totally different time periods (accuracy2). There exist differences among individuals, which may stem from subjective personal wills, environment, and backgrounds, or from objective experimental environments and physiological states. In addition, there are significant differences in accuracy of emotion recognition task between the two different evaluation methods. This might be because there is the significant difference of static EEG signals among different time periods, which has a great interference to classification.

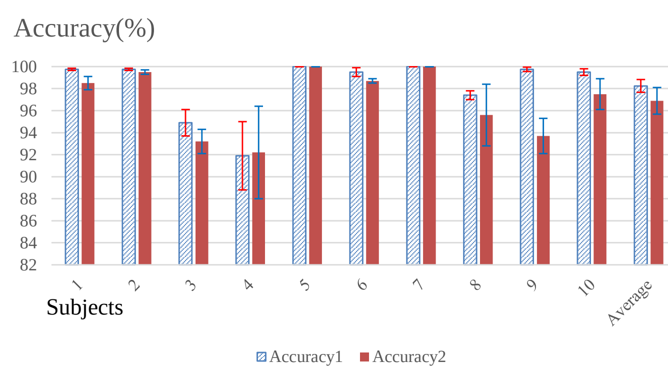
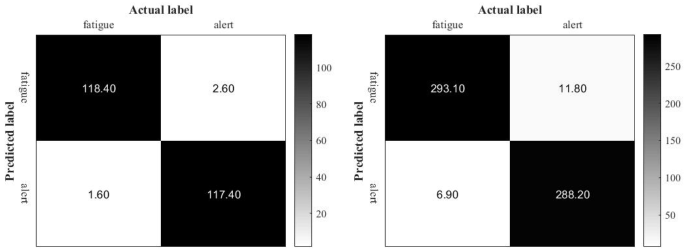


Fig. 9. The overall performance on fatigue driving detection task.



The confusion matrix under 10-fold cross-validation

The confusion matrix under clips-based training-test split validation

Fig. 10. The confusion matrix on fatigue driving detection task.

TABLE II
THE PERFORMANCES OF REPRESENTATIVE METHODS ON OUR DATASET

Method	Details	Acc.1(%)	Acc.2(%)
PSD+SVM	PSD feature value on each channel for SVM	70.38±10.62	63.67±11.12
DE+SVM	DE feature value of 5 frequency bands on each channel with SVM	87.39±6.69	67.97±9.84
DE+DBN	DE features with DBN as classifier	89.75±6.35	71.73±9.31
GCNN	GCNNs, DE features	93.62±2.61	77.46±6.42
CFCNN	Coincidence filtering	97.13±1.52	77.81±9.56

B. Comparisons with previous studies

1) Comparisons on emotion recognition task: To verify the effectiveness of the proposed method, we compare our method with other existing methods. The methods that performed well on emotion recognition mainly include two types: methods using DE features as input, and methods that directly input original signals. The former methods include: group sparse canonical correlation analysis (GSCCA) [21], deep belief network with DE features (DBN) [20], support vector machine using DE features (SVM) [20], minimalist neural network (MNN) [26], space-temporal recurrent neural network (STRNN) [14], dynamical graph convolutional neural networks (DGCNN) [18], Bimodal deep autoencoder (BDAE) [22], Graph regularized extreme learning machine (GELM) [27], and SyncNet [28]. We apply several representative methods on our dataset for comparison, the methods and models involved are as follows:

PSD+SVM: Classical emotion recognition method on EEG signals, where PSD features are extracted from each channel of EEG signals at five specific frequency bands (delta: 1-3 Hz, theta: 4-7 Hz, alpha: 8-13 Hz, beta: 14-30 Hz, gamma: 31-50 Hz), and are fed into the traditional SVM.

DE+SVM: DE is extracted from each channel of EEG signals at five specific frequency bands, and then is fed into SVM.

DE+GCNN [18]: DGCNN, which is composed of a graph convolution layer, a 1×1 convolution layer, a ReLU activation, fully-connected layers, and an output layer with Softmax function.

DE+DBN [20]: DE+DBN, DE features as input, and DBN as classifier.

CFCNN: Our method.

The accuracy1 is the performance under 10 fold cross validation, and the accuracy2 is the performance under the evaluation that the training set and test set from different time periods. Both of the result indicates our CNN can achieve a good adaptability and performance in EEG-based emotion recognition.

2) Comparisons on fatigue driving task: In order to further test the versatility of our approach, we refer to some traditional methods used in fatigue driving and EEG-based signals analysis, such as CSP+SVM [12] with 62.88%, auto regressive (AR) modeling and Bayesian neural network method [23] with 88.02%, neural networks method [24] with 87.4%, channel-wise CNNs [12] with 86.08%. We apply several competitive methods on our dataset for comparison, the methods and models involved are as follows:: The methods and models involved are as follows:

PSD-SVM: It extract the EEG power spectrum density features and used SVM classifier to determine the fatigue level.

CSP-SVM: It feed the relative energy of the filtered channels from CSPs methods into SVM classifiers, which is used to test the results in [12].

CNN-1: It develop a four-layer CNN for spatial feature fusion and temporal feature extraction on steady state visual evoked potential classification in [29].

CNN-2: A novel channel-wise CNN with raw EEG data on drivers cognitive performance prediction tasks in [12].

LSTM: The deep long short-term memory (LSTM) architecture for binary classification in [30], which consists of two LSTM layers and a sigmoid activation function.

CFCNN: Our method.

The detailed results are presented in Table 3. We can see that, traditional PSD+SVM and CSP+SVM get a result of $71.35\%\pm7.45\%$ and $74.56\%\pm6.88\%$, respectively. Compared to the traditional methods, our CNN structure, derived from them, achieves a result with significant improvements both in mean accuracy and standard deviation. Compared with CNN-2, which receives a result of $95.11\% \pm 2.24\%$, our structure still shows a superiority.

TABLE III
THE PERFORMANCES OF REPRESENTATIVE METHODS ON FATIGUE DRIVING TASK

Method	Details	Acc.1(%)	Acc.2(%)
PSD+SVM	PSD feature value on each channel for SVM	71.35±7.45	68.62±8.83
CSP+SVM	the relative energy from C-SPs methods into SVM	74.56±6.88	72.28±7.14
CNN-1	a four-layer CNNs	90.84±3.02	88.16±5.47
CNN-2	a channel-wise CNNs	95.11±2.24	92.78±3.27
LSTM	long short-term memory (LSTM) architecture	92.17±4.62	89.15±6.79
CFCNN	Coincidence filtering	98.25±0.58	96.89±1.21

C. Discussion

All these comparison results indicate that our architecture performs substantially better in EEG-based emotion recognition and fatigue driving detection. This is mainly because our structure attaches importance to information in the temporal dimension and to the meaning of the channel space information, which could be derived from the learning and imitation of artificial features through coincidence filtering. For the perspective of practical applications, we did not use cross-sample feature smoothing to correct the data, all methods use 1 second data samples only. Therefore, some DE feature-based methods perform worse than they did on their own datasets. Besides, since sadness and fear both belong to negative emotions, leading to the decrease in the accuracy of all compared methods.

V. Conclusion

We have proposed a simple and effective method for designing a targeted CNN structure for EEG-based tasks. Referring to the existing studies, we design two experiments, and obtain the EEG signals of two different tasks. The present structure reaches a good result on both of the tasks. Learning from artificial features, our method focuses on the temporal dependency information in EEG signals, and preserves the practical significance of the channel. Based on coincidence filtering, we concentrate on the evaluation of the whole region instead of partial pixels.

One major contribution of this research is that our framework is integrated with the prior knowledge, namely, artificial features, instead of simply using CNNs from image processing or other areas, which allows us to use previous researches and knowledge to improve the classification accuracy. Another one is that by observing the eigenvalues of the channel average pooling layer, we can explore the relationship between different psychological states and channels. Due to the generality of our method, we expect it to be useful for broader application in EEG-based health monitoring tasks.

References

[1] S. D. Pressman and S. Cohen, "Does positive affect influence health?" *Psychological bulletin*, vol. 131, no. 6, p. 925, 2005.

[2] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2012.

[3] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[4] N. Mammone, S. De Salvo, L. Bonanno, C. Ieracitano, S. Marino, A. Marra, A. Bramanti, and F. C. Morabito, "Brain network analysis of compressive sensed high-density EEG signals in ad and mci subjects," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 527–536, 2019.

[5] V. Vanitha and P. Krishnan, "Time-frequency analysis of EEG for improved classification of emotion," *International Journal of Biomedical Engineering and Technology*, vol. 23, no. 2/3/4, pp. 191–212, 01 2017.

[6] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Networks*, vol. 57, pp. 39 – 50, 2014.

[7] Z. K. Gao, K. L. Zhang, and a. Dang, W. D. et, "An adaptive optimal-kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowledge-Based Systems*, vol. 152, pp. 163–171, 2018.

[8] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, Nov. 2013, pp. 81–84.

[9] P. Mirowski, D. Madhavan, Y. Lecun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1927–1940, 2009.

[10] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2030–2042, 2014.

[11] R. T. Schirrmeister, J. T. Springenberg, F. Ldj, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[12] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Processing Image Communication*, vol. 47, pp. 549–555, 2016.

[13] Z. K. Gao, X. M. Wang, Y. X. Yang, C. X. Mu, Q. Cai, W. D. Dang, and S. Y. Zuo, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2755–2763, 2019.

[14] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatialtemporal recurrent neural network for emotion recognition," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 839–847, 2019.

[15] M. Soleymani, S. A. Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2016.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.

[17] V. J. Lawhern, A. J. Solonm, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, DOI:10.1088/1741-2552/aace8c.

[18] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, no. 99, DOI:10.1109/TAFFC.2018.2817622, 2018.

[19] M. H. Bhatti, J. Khan, U. G. Khan, R. Iqbal, M. Aloqaily, Y. Jararweh, and B. Gupta, "Soft computing based EEG classification by optimal feature selection and neural networks," *IEEE Transactions on Industrial Informatics*, DOI:10.1109/TII.2019.2925624, 2019.

[20] W. L. Zheng and B. L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[21] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 3, pp. 281–290, 2017.

[22] W. Liu, L. Z. Wei, and B. L. Lu, "Emotion recognition using multimodal deep learning," in *International Conference on Neural Information Processing*, vol. 1602.08225. Springer International Publishing, 2016, pp. 521–529.

[23] R. Chai, G. R. Naik, T. N. Nguyen, S. H. Ling, Y. Tran, A. Craig, and H. T. Nguyen, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 715–724, 2017.

[24] A. G. Correa, L. Orosco, and E. Laciar, "Automatic detection of drowsiness in EEG records based on multimodal analysis," *Medical Engineering and Physics*, vol. 36, no. 2, pp. 244–249, 2014.

[25] D. Hauke, S. M. Jakob, K. I. Helle, and P. Sadasivan, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Systems with Applications*, vol. 114, pp. 532–542, 2018.

[26] S. Keshmiri, H. Sumioka, J. Nakanishi, and H. Ishiguro, "Emotional state estimation using a modified gradient-based neural architecture with weighted estimates," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 4371–4378.

[27] W. Zheng, J. Zhu, and B. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 10, pp. 417 – 429, 2019.

[28] Y. Li, M. Murias, S. Major, G. Dawson, K. Dzirasa, L. Carin, and D. Carlson, "Targeting EEG/LFP synchrony with neural nets," in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 4620–4630.

[29] N. S. Kwak, K. R. Mller, and S. W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLOS ONE*, vol. 12, p. e0172578.

[30] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. S. Kabban, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognition Letters*, vol. 94, no. C, pp. 96–104, 2017.