LEVERAGING SPARSE CODING FOR EEG BASED EMOTION RECOGNITION IN SHOOTING

Yulu Wang[1], Yiwen Sun[1], Lei Fang[2] and Changshui Zhang[1],IEEE Fellow*

[1] Institute for Artificial Intelligence, Tsinghua University (THUAI), State Key Lab of Intelligent Technologies and Systems. Beijing National Research Center for Information Science and Technology (BNRist) Department of Automation, Tsinghua University, Beijing ,P.R.China
[2] DataCanvas Technology Co., Ltd.

ABSTRACT

Emotion recognition in shooting is of great importance for improving athletes' training methods. However, there is no open and high confident electroencephalography (EEG) dataset about shooting due to the difficulty of data acquisition, which made it a challenge for related studies. In this paper, we collected EEG of novice shooters and high-level shooters in different emotion states, and established two shooting datasets. Furthermore, instead of adopting the common convolutional neural network, we are the first to leverage sparse coding for EEG based emotion recognition in shooting process. Our proposed method can effectively solve the problem of low accuracy caused by data with low signal-noise ratio and small training set. The experimental results demonstrate that our method outperforms other representative deep learning based methods.

Index Terms— Sparse coding, EEG, Shooting, Emotion Recognition

1. INTRODUCTION

In international shooting competitions, psychological factors have become the key to success. Tension, depression and other psychological factors make shooters play poorly in competition, even for high-level athletes in the Olympic Games. Many famous athletes fail to win the championship because of psychological pressure. Therefore, it is vital for the relevant sports professionals to recognize their emotions in the shooting process. They believe that the recognition of Shooters' emotion in the shooting process can help to improve the training methods. EEG is an effective material for emotion recognition. Early studies have shown that EEG signals can reflect the information of shooting process [1–4]. Besides, EEG is an effective approach for emotion recognition [5]. Therefore, it is vital to recognize emotion in the shooting process through EEG signal.

In recent years, many researchers have adopted a variety of methods for emotion recognition based on EEG. At first, researchers use spatial filtering to extract EEG features, then use them for emotion recognition [6, 7], Lin et al. [8] use Support Vector Machines(SVM) to classify the four emotions states generated when listening to music. In [9], Bayes Network is used to extract features in EEG and ensemble classifiers such as Random Forest for the final emotion recognition. In [10], Simple Recurrent Units (SRU) network is used to grasp the temporal information of EEG for emotion recognition. Convolutional Neural Network (CNN) is introduced to EEG-based recognition in [5, 11] and it is considered to be an effective EEG classification method.

Most of the above methods only consider the temporal information in EEG without considering the spatial information of the EEG electrodes on the scalp. In the research of psychology and neuroscience, researchers believe that discrete emotions always correspond to different regions of the brain [12]. Therefore, the lack of spatial information may lose some key information in emotion recognition. Besides, these methods do not perform well on the processing and feature extracting of signal data. Sparse coding is a classical method in the field of signal processing [13–15], which aims to reconstruct sparse feature through the low-dimensional observations. The imperative for sparse codes to reconstruct an input signal helps improve the information degeneracy issues in deep architectures [16].

The main contributions in this work are as follows: First, we establish two data sets of novice shooters and high-level athletes under different emotion states, which can make up for the blank of this part of research. The recognition of shooter's emotions in shooting can help the coach to arrange a more appropriate training method for athletes. Then we summarize a complete set of data processing procedures to effectively remove the noise in EEG. Finally, we find a more effective method on small EEG data set and first adopt super-

⋆ Corresponding author. This work is funded by the Natural Science Fundation of China(NSFC) and the German Research Foundation(DFG) in Project Crossmodal Learning, NSFC 62061136001/ DFG TRR-169, and NSFC 62176132.Thanks associate professor Jia Cheng, Dr. Fangfang Wu of Department of Mechanical Engineering of Tsinghua University, associate professor Chunmei Cao of Division of Sports Science and Physical Education of Tsinghua University for their contributions in collecting the datasets.

vised sparse coding method to EEG-based emotion recognition. Our network consists of several bottleneck structures, each structure includes a compress sparse layer and an expand sparse layer. We found that the sparse coding structures can filter out the redundant information caused by other brain activities in EEG in the signal reconstruction process, so that the network can focus more on the information which is more relevant to the task. This is the key why our proposed method outperforms other state-of-the-art(SOTA) methods CNN.

2. DATA SET

Sports professionals have an urgent need to recognize the emotions in the shooting process based on EEG. However, there is no reliable data set to support relevant research because it is hard to find suitable subjects and collect useful data in the shooting process. Therefore, we carry out research in related areas and established two EEG data sets for novice shooters and high-level athletes under different emotion states.

The EEG signal collection system used in this experiment is a 64-electrode EEG signal amplification and collection system. Considering the actual conditions and the accuracy requirements of the experiment, we selected 32 electrodes for EEG collection in shooting process referring to the specifications of $10-20$ electrode system [17]. After removing the reference electrode, the collected EEG data consists of 31 channels, of which is collected by corresponding electrode. We collected the emotions of novice shooters and high-level athletes in the shooting process, then named the two data sets as novice shooter and high-level shooter. Due to different shooting levels, novice shooters are asked to shoot in Virtual Reality scenes, while high-level athletes shoot in live-fire exercise. Table 1 shows details of the subjects of the two data sets.

Table 1. Details of two data sets.

Name	Age	Gender	Size
Novice shooter	18-35	male	35
High-level shooter	19-27	male/female	10

Novice shooter. We use virtual reality technology to induce positive and negative emotions in the virtual scene. The positive and negative emotions are evoked by watching videos, and the data of neutral emotion is EEG data of subjects in resting state. Before and after the emotion evoking, the subjects must fill in self assessment manikin (SAM) [18] and discrete emotion assessment table which is to estimate whether the subjects' emotion evoking is successful. After data collection, each subject must fill in a immersion table to estimate their degree of participation. Then we eliminate the samples failed to evoke emotion according to the above tables. As shown in table 1, the novice shooters are all male, aged from 18 to 35, right-handed. They are asked to shoot 20 shots under each induced emotion states, with a shooting interval of 10s. We collected data from 35 people.

High-level shooter. We collected EEG data from ten 10-m air rifle athletes (5 males and 5 females) from shooting team of Tsinghua University. They are also members of sports teams in some provinces of China. All of them are right-handed. The shooting is live-fire exercise, and the experimental site is the Shooting Hall of Tsinghua University. All of the athletes shoot 60 shots under pressure and relaxation state. Pressure is evoked by competition. Two athletes shoot at the same time. After each shot, each athlete will be informed of the ring number. All the subjects are informed that the winner will be rewarded and the loser will be punished. During the experiment, the athletes need to fill in SAM to judge whether the they are under pressure.

3. PROPOSED METHOD

3.1. EEG Signal Representation

EEG data sets generally have the characteristics of small amount of data and low signal-to-noise ratio data. Therefore, we should filter noise information and extract useful information from EEG to the greatest extent. Referring to previous studies [1, 4] on EEG of shooters, we mainly focus on alpha and beta EEG bands. In order to remove the environmental noise, we use the Finite Impulse Response (FIR) filter to band-pass filter the data and the retained frequency component is 0.5-30 Hz. Then we calibrate the trace of EEG data of each lead to reduce data drift.

In the EEG signal of shooters, the main noise comes from electrooculogram (EOG). Independent component analysis (ICA) is a general approach to separate independent components in EEG. We use FastICA [19] to separate the independent components of EEG and remove the EOG components. The result of EOG removal is shown in figure 1.

Then, the EEG time series is transformed into a two-dimensional matrix of time-frequency power by short-time Fourier transform (STFT). In addition, we used data from multiple electrodes. Each electrode reflect EEG information at a specific location of the brain. After the above processing, we get 3-D EEG data including time domain, frequency domain and space information simultaneously. In this way, we get a better representation of EEG signals.

3.2. Supervised Sparse Coding Network For EEG

Sparse coding is a problem of reconstructing sparse features x_s from noisy linear measurement y:

$$y = Ax_s + \varepsilon, \tag{1}$$

where $y \in \mathbb{R}^M$, $x_s \in \mathbb{R}^N$, $A \in \mathbb{R}^{M \times N}$, and $\varepsilon \in \mathbb{R}^M$ is additive white Gaussian noise. To solve this sparse linear problem,

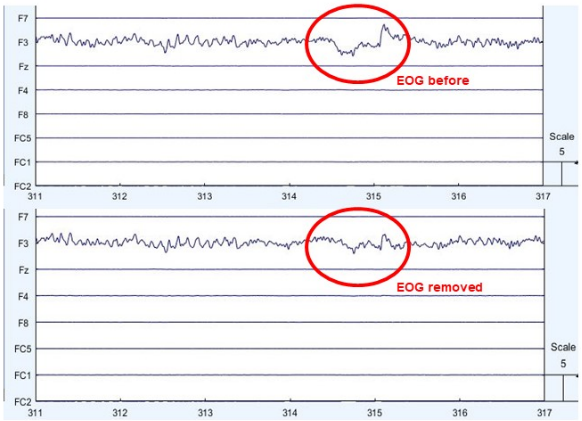

Fig. 1. The effect of EOG removal. The above shows EEG signal before EOG removal, and the red box shows an EOG signal. The below figure shows EEG signal after EOG removal, and the red box shows the signal corresponding to the red box in the above figure.

a popular method is to solve the following convex optimization problem [20]:

$$\min_{x} \ \frac{1}{2}\left\|y - Ax\right\|_2^2 + \lambda \left\|x\right\|_1 . \quad (2)$$

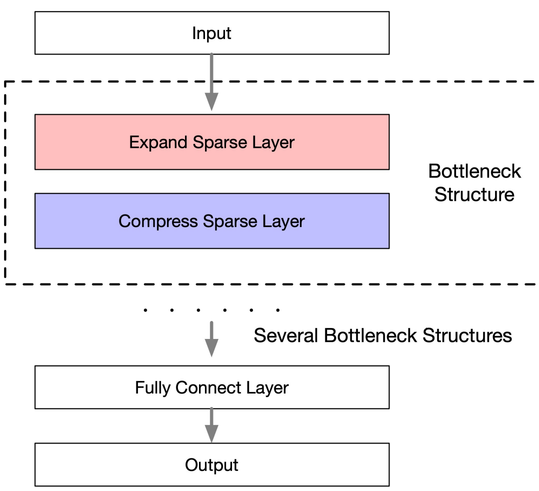

Fig. 2. Supervised sparse coding network. It consists of several bottleneck structures. Each bottleneck structure consists of an expand sparse layer and a compress sparse layer.

Sparse representation of data can be obtained by sparse coding. Sun et al. [16] extend conventional sparse coding to multi-layer architectures in image classification. They expand the learning capacity of sparse coding network by bottleneck structures which is used to reconstruct sparse features of network input and reduce the amount of parameters. In order to solve the problem of low signal-noise ratio and small data sets, we adopt sparse coding structure into emotion recognition to utilize the advantages of sparse coding in signal reconstruction and feature extraction. The structure of supervised deep sparse coding used in this paper is shown in figure 2.

As shown in figure 3.2, the bottleneck structure is consist of an expand sparse layer and a compress sparse layer. The input EEG representations are 3-D vectors. The length represents time, the width represents frequency, and the height represents space. First, the 3-D input pass through a sparse expansion layer where it is converted into high-dimensional sparse feature. Like the convolution process in CNN, we calculate sparse feature in the neighborhood of each position. The vectors in the neighborhood is concatenated together first, then the sparse feature is calculated by a large dictionary matrix. The process is described in formula (2). Unlike [16], here we didn't use l_2 norm in formula (2) to get better sparse features. The traditional sparse recovery algorithm FISTA [14] is used to solve the sparse coding problem. Like the expand sparse layer, the compress sparse layer repeat the above process with a thin dictionary, which is to compress the high-dimensional sparse representation. This can reduce the amount of calculation and avoid the network being too large. Here, the sparse feature is assumed to be non-negative to correspond to the nonlinear operation in deep learning. Each sparse coding layer have a batch normalization layer [21] after to enable the convergence. After several stacked bottleneck structures, the network passes through a global average pooling layer and a fully connected layer.

The loss function is the cross entropy loss function:

$$Z = \sum_{i=1}^{n} p\left(x_i\right) \ln\left(q\left(x_i\right)\right) \quad (3)$$

where $p(x_i)$ is the label, $q(x_i)$ is the prediction.

3.3. The whole emotion recognition process.

Figure 3 demonstrates the complete process for EEG-based emotion recognition with sparse coding in the shooting process. The original data is 31-channel time series. EEG Data in each channel is obtained from the electrode at the corresponding positions in cerebral cortex. The time series of each channel is transformed into a time-frequency matrix and we get 3-D EEG representations. Then the data containing time, frequency and spatial information is input into several sparse coding layers for processing, and finally the prediction is obtained after the fully connected layer.

4. EXPERIMENTS AND DISSCUSSION

4.1. Compared Methods

We compare our SCN-based emotion recognition network with two representative networks on EEG emotion recognition on two real-world data sets of shooters. We select SVM as the representation of traditional method. It has been widely used in the classification of EEG signals. And CNN is the

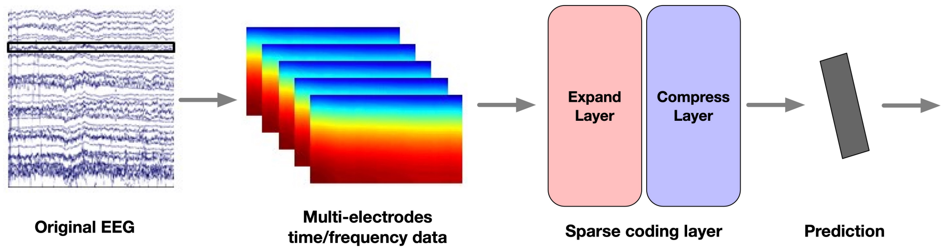

Fig. 3. The whole emotion recognition process. On the far left is the original EEG data, and the black box shows the data of one electrode. Then the data is transformed into a time-frequency matrix with 31 channels. After passing through several sparse coding layers, the prediction results are obtained after a fully connected layer.

SOTA method in shooting EEG recognition. We apply different structure of CNN include Resnet [22] to ensure the credibility of the experimental results.

4.2. Results Analysis

We cut the EEG data under each emotion state into 10s segments. Then we randomly select 80% of the data as the training set and the other as the testing set. Data in the training set does not appear in the testing set. After that, SCN network is used to classify them, and the learning rate is 10^{-5}.

4.2.1. Results on novice shooter dataset.

On novice shooter dataset with three categories (neutral, positive and negative), we compare the SCN network with CNN using one convolution layer, two convolution layers and Resnet. The results are shown in the table 2.

Table 2. Average accuracy on three-emotion recognition on novice shooter dataset.

Methods	Average Accuracy
SVM	71.3%
CNN(1-layer)	86.5%
CNN(2-layer)	87.1%
CNN(Resnet)	86.8%
Proposed Method(1-bottleneck)	87.5%
Proposed Method(2-bottleneck)	87.9%
Proposed Method(5-bottleneck)	**88.0%**

It can be seen from table 2 that the traditional SVM method performs poorly in processing time-frequency features. Though CNN performs good in processing time-frequency features, the proposed method is more accurate. The average accuracy of proposed method with one bottleneck structure is 87.5%, which is better than CNN with different depth. We also compared the impact of different numbers of bottleneck structures. The more the number of bottlenecks, the better the prediction results of the network. Considering the larger amount of parameters in 5 bottleneck network, network with 2 bottlenecks is suitable in practice.

Table 3. Average accuracy on two-emotion recognition on high-level shooter dataset.

Methods	Accuracy
SVM	69.5%
CNN(2-layer)	83.2%
Proposed Method(2-bottleneck)	**85.1%**

4.2.2. Results on high-level shooter dataset.

The average accuracy on two-emotion recognition on high-level shooter dataset is shown in table 3. We can see that our proposed methods outperforms CNN by 1.9%. This dataset collected shooting data of athletes with different genders, so there's more disturbance information in it. And the datasets is smaller than the novice shooters dataset, which makes it difficult for CNN to have good performance. Instead, sparse coding layer can filter out more useless information and retain more useful features in the process of reconstructing sparse features.

5. CONCLUSION

In this paper, we established two shooting datasets of novice shooters and high-level athletes under different emotion states. We use multi electrodes EEG data to retain spatial information and extract representations of EEG by denoising, frequency band selection, FastICA and STFT. In addition, we first adopt supervised sparse coding method to EEG-based emotion recognition. Our proposed method outperforms other classifiers in recognizing positive, negative and neutral emotions in EEG for novice shooters. On the dataset of high-level shooters with less data, our method has more obvious advantages than other representative machine learning methods in EEG-based emotion recognition.

6. REFERENCES

[1] N. Konttinen, H. Lyytinen, and R. Konttinen, "Brain slow potentials reflecting successful shooting performance," *Research Quarterly for Exercise and Sport*, vol. 66, no. 1, pp. 64–72, 1995.

[2] G. M. Loze, D. Collins, and P. S. Holmes, "Pre-shot eeg alpha-power reactivity during expert air-pistol shooting: a comparison of best and worst shots," *Journal of Sports Sciences*, vol. 19, no. 9, pp. 727–733, 2001.

[3] M. Doppelmayr, T. Finkenzeller, and P. Sauseng, "Frontal midline theta in the pre-shot phase of rifle shooting: Differences between experts and novices," *Neuropsychologia*, vol. 46, no. 5, pp. 1463–1467, 2008.

[4] C. D. Percio, C. Babiloni, M. Bertollo, N. Marzano, and Fabrizio Eusebi, "Visuo-attentional and sensorimotor alpha rhythms are related to visuo-motor performance in athletes," *Human Brain Mapping*, vol. 30, no. 11, pp. 3527–3540, 2009.

[5] S. E. Moon, S. Jang, and J. S. Lee, "Convolutional neural network approach for eeg-based emotion recognition using brain connectivity and its spatial information," 2018, pp. 2556–2560.

[6] D.J. Mcfarland, L.M. McCaNe, S.V. DaVid, and J.R. Wolpaw, "Spatial filter selection for eeg-based communication," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 386–94, 1997.

[7] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Trans Biomed Eng*, vol. 58, no. 2, pp. 355–362, 2011.

[8] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. Chen, "Eeg-based emotion recognition in music listening.," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[9] Jordan J. Bird, Aniko Ekart, Christopher D. Buckingham, and Diego R. Faria, "Mental emotional sentiment classification with an eeg-based brain-machine interface," in *The International Conference on Digital Image and Signal Processing (DISP'19)*, 2019.

[10] L.L. Wei, C.and Chen, Z.Z. Song, XG Lou, and D.D. Li, "Eeg-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical signal processing and control*, vol. 58, no. Apr., pp. 101756.1–101756.13, 2020.

[11] X. H. Wang, T. Zhang, X. M. Xu, L. Chen, X. F. Xing, and Clp Chen, "Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019.

[12] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review.," *Behavioral and Brain Sciences*, vol. 35, no. 3, pp. 121–143, 2012.

[13] David L Donoho and Michael Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[14] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Omnipress, 2010, pp. 399–406.

[15] Jialin Liu, Xiaohan Chen, Zhangyang Wang, and Wotao Yin, "Alista: Analytic weights are as good as learned weights in lista," 2018.

[16] Xiaoxia Sun, Nasser M Nasrabadi, and Trac D Tran, "Supervised deep sparse coding networks for image classification," *IEEE Transactions on Image Processing*, vol. 29, pp. 405–418, 2019.

[17] Marc R Nuwer, Giancarlo Comi, Ronald Emerson, Anders Fuglsang-Frederiksen, Jean-Michel Guérit, Hermann Hinrichs, Akio Ikeda, Fransisco Jose C Luccas, and Peter Rappelsburger, "Ifcn standards for digital recording of clinical eeg," *Electroencephalography and clinical Neurophysiology*, vol. 106, no. 3, pp. 259–261, 1998.

[18] Jon D Morris, "Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response," *Journal of advertising research*, vol. 35, no. 6, pp. 63–68, 1995.

[19] A. Hyvarinen, "Fast ica for noisy data using gaussian moments," in *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, 1999.

[20] Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[21] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.