2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Insights of 3D Input CNN in EEG-based Emotion Recognition

Kris van Noord, Wenjin Wang*, and Hailong Jiao

Abstract— **Electroencephalogram (EEG) signals have shown to be a good source of information for emotion recognition algorithms in Human-Brain interaction applications. In this paper, a reproducible framework is proposed for classifying human emotions based on EEG signals. The framework consists of extracting frequency-dependent features from raw EEG signals to form a three-dimensional EEG image which is classified by a convolutional neural network (CNN). The framework is used to show that the 3D input CNN outperforms conventional methods with two-dimensional input, using a public dataset. The implementation of the framework is publicly available to facilitate further work on this topic: https://github.com/KvanNoord/3D-CNN-EEG-Emotion-Classification.**

I. Introduction

Recognition of human emotion plays an important role in intelligent Human-Brain interaction applications, such as virtual reality, autonomous driving, educational systems, and health care [1], [2]. For human interpretation, the main source of information for emotion estimation is the facial expression. However, it is known that some people can spoof their facial expression at a certain emotion [2]. Therefore this method is susceptible to fraud. More reliable sources of information are needed for human emotion recognition. One of these sources is the electroencephalogram (EEG) signal, i.e. a recording of the electrical field of a brain that represents different brain activities. The information from EEG is favoured over other physiological signals for emotion recognition because it comes directly from the brain [1], [3].

EEG signals have a complex nature and are easily distorted. Therefore, the EEG-based emotion recognition is a complex task. However, with the rise of machine learning (especially deep learning), the performance of emotion recognition models is rapidly increasing. Multiple strategies for emotion recognition have been proposed over the years, where the successful models often use Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).

In this paper, we propose a light-weight yet highly reproducible EEG-based emotion recognition framework that is made publicly available to facilitate the study in this field. We evaluated the differences between the conventional two-dimensional input representation and the new 3D input representation as proposed by Yang *et al.* [3] on a public dataset. Furthermore, we investigated various processing techniques within the proposed framework (e.g. frequency feature extraction methods, normalization techniques, CNN structures) and reported the insights that can further optimize EEG-based emotion recognition.

Kris van Noord and Wenjin Wang are with Eindhoven University of Technology, The Netherlands.

Hailong Jiao is with Peking University, China and Eindhoven University of Technology, The Netherlands.

*Correspondence: Wenjin Wang (wwang@tue.nl).

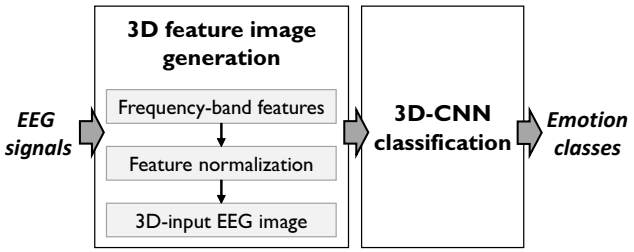

Fig. 1: Processing pipeline of the proposed framework for EEG-based emotion recognition.

II. Materials and Method

The processing pipeline of the proposed framework is shown in Figure 1. The EEG-based emotion recognition is separated into two parts. First, the emotion-related band features are extracted from the EEG signals and formed in a 3D representation. Secondly, a 3D-CNN classification algorithm is developed to predict the emotion state (Low/High) based on the input of 3D EEG-feature image.

The proposed framework is validated on the Database for Emotion Analysis using Physiological Signals (DEAP) [4]. In the experiment of DEAP, physiological and multi-channel EEG signals of 32 participants were recorded while watching 40 different music videos for 1 minute. During the video watching, the participants were asked to rate each video (on continuous 1-9 scale) in terms of 4 emotions: valence, arousal, dominance, and liking. Especially the first two are important for emotion classification, and therefore are focused emotion types in this work. Moreover, a 3-second pre-trial was recorded for each trial, in which the subjects were assumed to be in a low-emotional state. In DEAP, the EEG signals of 32 different channels were recorded. The channels were placed and labelled according to the International 10-20 System, a recognized method for describing the location of scalp electrodes in an EEG experiment [5]. For DEAP, a pre-processed dataset is also provided, of which all steps are explained in [4]. In the following subsections, we introduce the proposed framework step-by-step.

A. 3D EEG-image generation

1.1 Frequency band features

For EEG processing, a meaningful feature extraction method that extracts emotion-related information is desired. The frequency spectrum of EEG signals can be divided in separate bands, where frequencies in those bands are

Band	Frequency	Brain state
θ	4 - 8 Hz	Light sleep pattern
α	8 - 12 Hz	Relaxing state
β	12 - 30 Hz	Active thinking, focus, alert
γ	$>$ 30 Hz	Cross-modal sensory processing

TABLE I: Bands in EEG frequency spectrum defined for different emotions.

observed for different brain activities. The relevant bands are shown in Table I. Based on these brain state descriptions, emotions are observed mostly in the high frequency ranges, which is also shown in previous research [3]. These four bands are the basis of the two feature extraction methods that are investigated in this research: power spectral density and differential entropy. The first is a new simple approach with some intuitive meaning, the second is giving best results in earlier work [3].

• **Power spectral density** It involves the power of different frequencies. First, the power spectral density (PSD) is estimated for the EEG-signal. Then the mean of the PSD in the EEG frequency bands (as given in Table I) is calculated, giving four feature values per EEG channel per signal.

• **Differential entropy** According to [3], [6], differential entropy (DE) is an effective representation for EEG signals. The DE $h(X)$ of a random variable X is given as

$$h(X) = \int f(X)\log(f(x))dx, \tag{1}$$

where $f(X)$ is the probability density function of the random variable X. If EEG-signals are regarded as a random variable over time, and the distribution is close to Gaussian due to band-pass filtering [6], this simplifies to:

$$h(X) = \frac{1}{2}\log(2\pi e\sigma^2), \tag{2}$$

where σ^2 is the signal variance of a band-pass filtered signal. First the signal is filtered for each EEG frequency band, giving four different signals. Then, DE is calculated for each signal, again giving four values per EEG channel per signal.

1.2 Feature normalization

To reduce inter-subject differences and improve generalization of the representation, we apply normalization to the feature values. Two normalization methods are investigated, the first one showed good results in earlier work [3] and the second one is a new simple normalization technique.

• **Pre-trial normalization** Using the pre-trial signals provided in the DEAP dataset, pre-trial features are subtracted from the trial features. This gives a normalized feature which represents the difference between EEG-signals with and without experiencing emotions. This type of normalization was considered to improve the performance significantly [3].

• **Self-normalization** Since the neutral pre-trial may not be always available (e.g. when there is no resting period between trials), we investigate a second normalization technique: self-normalization. In this approach, we subtract the feature vector (with 4 feature values estimated by PSD or DE) by its average to make the feature vector zero-mean, thus a relative power (of PSD) or entropy (of DE) is calculated.

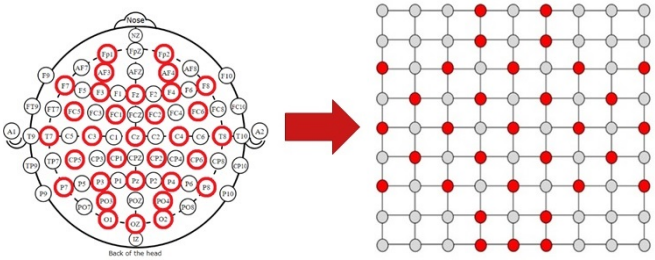

Fig. 2: With the locations specified by the 10-20 International EEG system, a sparse 9×9 matrix can be constructed.

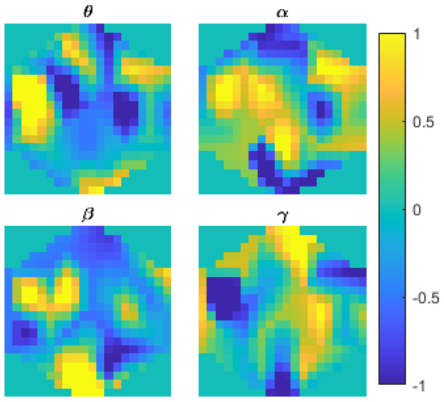

Fig. 3: Example of four EEG images (corresponding to four bands) where differential entropy is calculated from the EEG.

1.3 3D-input EEG image

For each EEG channel four feature values are obtained, giving in total 32×4 feature values. If using 32×4 matrix as the input of the algorithm for classification, the spatial information would be neglected. Therefore the channel feature values are placed in a matrix according to the place of each channel in the 10-20 International EEG system, as shown in Figure 2. The EEG channel values in DEAP are represented by red dots. With the use of spatial interpolation, the sparse EEG images are converted to 20×20 non-sparse images. Interpolation makes the model more robust: the images are less susceptible to small deviations in the positions of the electrodes. As there are four frequency bands, the input is a 2D matrix with four channels, thus a 3D cube.

With this 3D representation, EEG processing is similar to image processing. In image processing, three color channels (R, G, B) are used, wherein EEG processing four EEG-band channels (θ, α, β, γ) are used. Another difference is that EEG-images have a much lower resolution. For the upcoming processing steps, EEG processing is considered as a low-resolution image processing.

B. 3D-CNN based classification

Based on the aforementioned similarity to regular image processing, a CNN with 3D-input EEG image is proposed for classification. As a basic model, the CNN structure from Yang *et al.* [3] is used. This CNN structure, consisting of 4 hidden convolutional layers and 2 fully connected layers, is shown in Table II. The structure does not include pooling layers, as pooling is used to reduce the spatial dimension, which

Layer	Size	Output
Input	(20×20)×4	-
2D Conv1	(4×4)×64	20×20×64
2D Conv2	(4×4)×128	20×20×128
2D Conv3	(4×4)×256	20×20×256
2D Conv4	(1×1)×64	20×20×64
Flatten	-	25600
Fully connected	1024	1024
Fully connected	2	2
Softmax	-	2

TABLE II: Structure of the used CNN model.

is not necessary here because the resolution of the EEG-image is already low. All convolutional and fully connected layers are followed by ReLU activation for nonlinearity and Dropout layers to avoid overfitting. The goal of this work is to verify if the approach with 3D EEG input outperforms other 2D-input (channels, frequency bands) based methods. This is achieved by changing all kernel sizes in the model of Table II to 1×1. This leaves out the local spatial information (neighbouring pixels) in the convolution and is therefore similar to a conventional Neural Network (NN).

III. Experimental setup

All processing in this research is performed in Matlab, with the use of two toolboxes: EEGLab [7], Deep Learning and Parallel Computing [8]. A sliding window is used to segment the 60 s signal (in one trial) into short intervals. The sliding window approach involves two parameters: window length and stride. An appropriate window length is determined by experiments, the stride was set to 0.25 s.

For training a model, all available data is separated into 5 chunks, and 5-fold cross validation is applied. The CNN is trained using the Adam optimizer and cross-entropy loss. Training is performed in mini-batches of 64 training samples and is stopped if the validation accuracy stopped increasing for 5 epochs. The CNN model is trained per subject independently, as the EEG-based emotion recognition is too complex and the differences between subjects are too large to train a generalized model for all subjects. The performance per subject is given as the mean of all 5 validation accuracies. To emphasize on the difference between subjects, the performance of models is presented as median with interquartile range for all subjects.

IV. Results

In this section, all the results are shown and discussed for the valence and arousal class.

A. Sliding window length

The model performance is compared for different window lengths for both the DE and PSD methods. The results for the valence and arousal class are shown in Figure 4. For both methods and both classes, the dependence is similar. For small window lengths the amount of information that is extracted is limited. By increasing the window length, the performance increases up to a maximum around 3 - 7 s and then decreases again for longer window lengths. Long windows will include more distortions and are less responsive

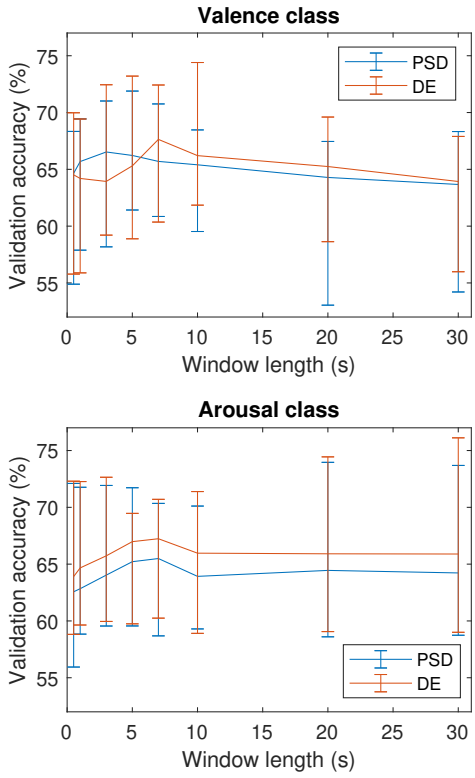

Fig. 4: Validation accuracies for PSD and DE features with different sliding window lengths.

to spontaneous emotion-related frequency fluctuations. Based on these findings, a window length in the range of 3 - 7 s seems appropriate.

B. Feature extraction and normalization

The feature extraction methods and normalization techniques as discussed in Section II are compared. Figure 5 shows that the normalization techniques do not improve performance, contradicting to the results shown by Yang *et al.* [3]. If no normalization method is used, both PSD and DE seem suitable features for emotion recognition, but the performance of the DE method is consistently higher than the PSD method. Regarding the reproducibility, our implementation is released for replicating this experiment and comparison.

C. Kernel size

As the third experiment, the model is trained for different sizes of the convolutional kernel applied in the convolutional layers. Here the 1×1 kernel represent the conventional NN without local spatial information. To compare our results with a baseline method (i.e. not CNN-based), the results of the Multinomial Logistic Regression (MLR) on this dataset are also added. Figure 6 shows that the performance of our proposed neural network structure is clearly better than a basic classifier. The poor performance of the basic classifier in the arousal class also indicates the complexity of EEG-based emotion recognition. Furthermore, the validation accuracy

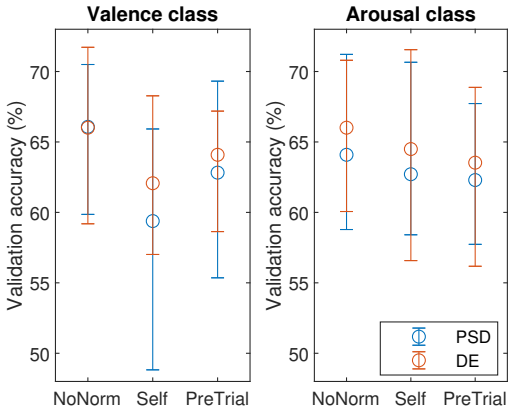

Fig. 5: Validation accuracies for PSD/DE feature for no normalization, self-normalization and pre-trial normalization.

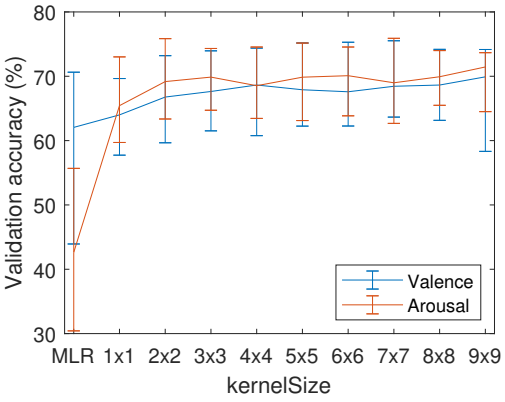

Fig. 6: Validation accuracies for MLR and CNN with different kernel sizes.

increases if the kernel size is increased from 1×1 to 4×4. This means that the 3D EEG representation outperforms a regular 2D representation (channel and band) that neglects local spatial information. Further increasing the kernel size increases the interquartile range (deviation) without significantly improving the median accuracy.

V. Discussion

The proposed method has shown to give decent results in EEG-based emotion classification. Nevertheless, all results in this work show that the EEG-based emotion recognition is a complex task and the performance is highly subject-dependent. Due to this issue, we consider the applicability in real-life applications to be limited in this stage. The model has to be trained on patient specific data, therefore a preliminary stage in which patient data will be gathered is necessary, which is time-consuming. The experiments show that both power spectral density and differential entropy are suitable for feature extraction, where the performance of the differential entropy method is higher. However, the normalization methods seem not improving the performance, which is different from claims made by the previous research [3]. The proposed 3D CNN structure outperforms simple classifier and conventional NN structure. An important limitation to take into account is the computational resource for embedded devices, which could limit the usability in real-time applications. For this reason, the proposed framework is kept light-weight.

The methodology of the current work is similar to the one of Yang *et al.* [3], but the reported high accuracies in their work are not achieved. In other literature, classification accuracies above 80% are found [9], but only if more complex models are used. We emphasize that the implementation of our study is made publically available to support the replication and comparison. For the future work, we consider to use a memory-based network (e.g. RNN) on top of CNN to improve the performance. As a second possible improvement, the subject normalization or calibration technique could be investigated and its understanding should be improved to eliminate the subject dependency of EEG representations.

VI. Conclusions

With the proposed light-weight 3D-CNN framework, classification accuracy of EEG-based emotion recognition is improved by using a 3D-input EEG image instead of regular signal inputs. The 3D EEG image is obtained by calculating features (e.g. power spectral density or differential entropy) in the EEG frequency bands and using the spatial location of the EEG electrodes. Due to the large inter-subject variation, the emphasis in EEG-based emotion recognition lies on generalization capability and the normalization/calibration techniques should be further investigated to achieve a more robust performance towards a subject-independent model.

References

[1] Y. Li, J. Huang, H. Zhou, and N. Zhong, "Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks," *Applied Sciences*, vol. 7, p. 1060, Oct. 2017.

[2] E. S. Salama, R. A.El-Khoribi, M. E.Shoman, and M. A.Wahby, "EEG-Based Emotion Recognition using 3D Convolutional Neural Networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018.

[3] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous Convolutional Neural Network with 3D Input for EEG-Based Emotion Recognition," in *Neural Information Processing* (L. Cheng, A. C. S. Leung, and S. Ozawa, eds.), vol. 11307, pp. 433–443, Cham: Springer International Publishing, 2018. Series Title: Lecture Notes in Computer Science.

[4] S. Koelstra et al., "DEAP: A Database for Emotion Analysis using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, Jan. 2012.

[5] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology," *Electroencephalography and Clinical Neurophysiology. Supplement*, vol. 52, pp. 3–6, 1999.

[6] 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)*, (San Diego, CA, USA), pp. 81–84, IEEE, Nov. 2013.

[7] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, Mar. 2004.

[8] G. Sharma and J. Martin, "MATLAB®: A Language for Parallel Computing," *International Journal of Parallel Programming*, vol. 37, pp. 3–36, Feb. 2009.

[9] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 93–110, Mar. 2017.