Multi-Task CNN model for emotion recognition from EEG Brain maps

Conference Paper · December 2021

DOI: 10.1109/BioSMART54244.2021.9677807

8 authors, including:

Lou Laurent

Université Paris-Est Créteil Val de Marne - Université Paris 12

Ahmed Roshdy

The American University of the Middle East

Régis Fournier

Université Paris-Est Créteil Val de Marne - Université Paris 12

Taha Beyrouthy

The American University of the Middle East

Multi-Task CNN model for emotion recognition from EEG Brain maps

Evgenii Rudakov, Lou Laurent, Valentin Cousin
Faculty of Sciences and Technology
Univ Paris Est Creteil
Créteil, France
{evgenii.rudakov, lou.laurent, valentin.cousin}@etu.u-pec.fr

Ahmed Roshdi, Régis Fournier, Amine Nait-ali
LISSI
Université Paris-Est (UPE)
Vitry, France
ahmad.roshdi@aum.edu.kw, {rfournier, naitali}@u-pec.fr

Taha Beyrouthy, Samer Al Kork
American University of the Middle East
Kuwait, Lebanon
{Taha.Beyrouthy, Samer.AlKork}@aum.edu.kw

***Abstract*—Emotion identification plays a vital role in human interactions. For this purpose, Computer-vision methods for automatic emotion recognition is nowadays a widely studied topic. One of the most studied approaches for automatic emotion recognition is processing multi-channel Electroencephalogram signals (EEG). This paper presents a new model for emotion recognition using brain maps as input and providing emotion states in terms of arousal and valence as output. Brain maps are a spatial representation of features extracted from EEG signals. The proposed model, called Multi-Task Convolutional Neural Network (MT-CNN), is fed with stacked brain maps of four different waves of different frequency bands: alpha, beta, gamma and theta, using differential entropy and power spectra density and considering observation windows of 0.5s. This model is trained and tested on the DEAP dataset, a well-known dataset for comparison purposes. This work shows that the MT-CNN performs better than other methods.**

***Index Terms*—Convolutional Neural Network, EEG, Brain map, Emotion recognition, Multi-Task learning**

I. Introduction

Emotions can be defined as the reaction of humans to daily life interactions. They can manifest through various modes of human expression such as psychophysiology, face expression, gesture, or biological reactions. Increasing the knowledge in this field can help researchers to understand individuals better. Therefore, analyzing emotion using computer vision is considered a fundamental challenge. In order to understand and analyze emotions properly, one needs to detect the appropriate signals.

From literature, numerous techniques are used to detect emotion using non-physiological signals such as speech [1], posture, or facial expression [2]. They can also be detected through physiological signals like heart rate, respiration, or brain signals such as functional Magnetic Resonance Imaging (fMRI) [3] or electroencephalography (EEG) [4]. Commonly, emotion recognition in EEG-based emotion techniques requires several phases, including: (1) Placement of the electrodes, (2) Collecting data, (3) Pre-processing, (4) Feature extraction and (5) Emotion classification [5]. The feature's input of the emotion classification can be EEG signals [4] or their image representation called brain maps [6].

Nevertheless, creating an emotion classification model based on brain maps represents a challenging task due to the difficulties of selecting high-quality data, pre-processing methods and classification models. This article's primary purpose is to construct an evolutive model which could help scientist for future research. From brain maps as input, one can classify the emotion both in the arousal spectrum and in the valence spectrum. For this purpose, the first step is the creation of brain maps through differential entropy (DE) and power spectral density (PSD). Their combination with the spatial information of electrodes location allows the creation of multiple 2D array as input of the proposed approach. Based on previous work in this field [8], the inputs are passed through a CNN model, which was chosen for its ability to analyze spatially distributed data representations and its high computational efficiency.

This paper is organized as follows: in section II, we will discuss about the DEAP dataset, brain maps construction and the proposed model used in our work. Section III is dedicated to the presentation of the obtained results. After an objective discussion and comparison this work is concluded in section IV.

II. Methods and materials

A. DEAP dataset

As described [7], the DEAP dataset contains EEG signals from 32 participants. During the experiment the participants were asked to watch 40 different stimuli from music video clips. Each clip has a duration of 1 minute and is labelled in valence and arousal from 1 to 9 in both as reference of emotion related to the stimuli. EEG signals were collected using a 32-channel's Biosemi ActiveTwo device with the 10-20 international system. Signals were pre-processed by denoising signals using EMG and EOG recording and a band-pass filter between 4 Hz and 45 Hz. Then downsampling to 128 Hz

978-1-6654-0810-3/21/$31.00©2021 IEEE

2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART) | 978-1-6654-0810-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIOSMART54244.2021.9677807

was performed. After each trial, participants where asked to rank their emotions in valence and arousal. The valence range relates to unpleasant (1 in valence) to pleasant (9 in valence) emotions such as sad, stressed or happy where arousal scales from inactive (1 in arousal) to active (9 in arousal) emotions like bored, excited or alert.

B. Features selection

Selected features are significant in emotion detection and creating brain maps. EEG signals are susceptible to noise, so the feature must give the best information about the signal. We follow the method described in previous work [8] where a 4D feature structure is built, including frequency, spatial and temporal features. Based on the work done [9], the original EEG signal is divided into Ts length segments, without overlapping to increase the amount of data. Then segments are decomposed into alpha, beta, gamma, and theta bands. There are five different waves known in brain activity in a range from 0 Hz to more than 30 Hz, Delta wave belongs to 0-4 Hz range and is often observed in deep sleep; it's not significant in emotion recognition of awake people [5]. Other waves which describe a state from relaxed to high brain activity or excited are often used in emotion recognition using EEG signals.

DE and PSD are extracted for each bands to describe these different waves. PSD describes the power present in the signal and it's one of the most used feature in frequency domain to achieve emotion recognition [10] and DE measures the complexity of the signal [6]. PSD and DE vectors extracted for each band from the original EEG signal are then transformed into brain maps to keep spatial information of electrodes locations. Brain maps, showed in Figure 1, are 2D representation of PSD and DE measured in all four frequency bands. For each window of 0.5s brain maps of all bands and the features are created and stacked in depth. Each brain map is an input of the model and has a height $h = 8$, a width $w = 9$ and a depth $d = 8$.

C. Model

This study utilises a fully convolutional architecture due to its ability to reason image-like data and the highly optimized GPU learning process for such architecture. Figure 2 shows the overall structure of the proposed Fully CNN Multi-Task Neural Network. Input to the model is a brain map which is a spatial-spectral representation of EEG signals. The model consists of four 2D convolution layers, a fully connected layer and Dropout and Batch Normalization layers after each of the ones mentioned above. At the end, the output goes to two streams: former is for classification the subject valence level, later is for arousal one. ReLU is used as a function of activation. The classification layers use a sigmoidal function to get probabilistic-like outputs. The model is trained to convergence.

D. Training process

In earlier state-of-the-art studies [8] [14], only the subject-specific subset of the data during the training process was used to maximize the performance of the emotion recognition. It is known that large training datasets are needed to increase the generalization of deep learning models. But due to the high cross-subject variability in EEG data, training with a full set of subjects led to the degradation of the performance model. For purposes of utilizing the broad range of subjects for the training process, methods of regularization of artificial neural networks are used in our study, including dropout, batch normalization, and learning rate exponential decay.

The performance definition of the final model is made by a cross-validation method with five folds. In each of the partitions, 122880 brain maps are used as a train set, and 30720 as a test set. The model is trained by the back-propagation method. Cross-entropy loss is used as a cost function. The final model is trained using the Multi-task learning principle. Multi-task learning is an approach in which a single model or part of it is used to solve more than one problem. Thus, solving the different tasks at once is beneficial for the model generalization. [15]. In addition, this approach allows to save computational resources during training and inference because in one pass the model is trained for several tasks at once. In this study, we are dealing with two tasks: classification of the valence and arousal levels. The final loss function is a weighted sum of the original loss functions, which can then be used to solve the problem using the back-propagation method. It can be represented as:

$$\mathcal{L}(\theta) = w_v \mathcal{L}_v(\theta) + w_a \mathcal{L}_a(\theta) \tag{1}$$

Where $\mathcal{L}_v$ and $\mathcal{L}_a$ are the loss functions for valence and arousal tasks respectively and w_v and w_a are weighting coefficients for them. And $\theta \in R^n$ are training model parameters.

The training hyper-parameters are presented in Table I.

TABLE I
HYPER-PARAMETERS USED FOR THE TRAINING OF THE MULTI-TASK CNN

Parameter	Value
Learning rate	0.001
Learning rate decay in plateau	1/2
Learning rate decay patience	5 epochs
Dropout rate	0.2
Batch size	64 samples
Valence loss weight coefficient	1
Arousal loss weight coefficient	1

III. RESULTS AND DISCUSSION

The model used in this experiment is inspired by the 4D-CRNN model [8]. Initially the input data for the model have 4D structure $X \in \mathbb{R}^{h \times w \times d \times T}$ with a depth $d = 4$ representing the four bands with DE as feature. In this paper, the changes presented previously are made to finally get 3D brain maps, representing PSD, DE and spatial locations of the electrodes for each 0.5s of the recorded signal $X \in \mathbb{R}^{h \times w \times d}$ with a depth $d = 8$ representing the four bands but including DE and PSD as features. The resulted structure are simplified for faster learning. Those inputs are used to feed a simple CNN structure

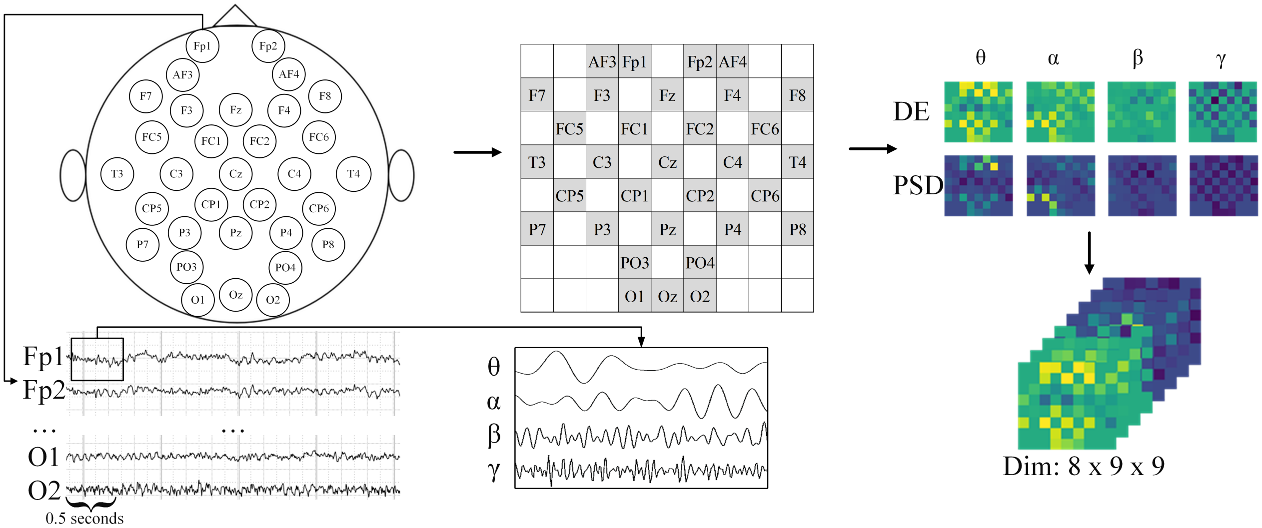

Fig. 1. Flowchart of 2D brain maps creation with electrodes location on the scalp, electrodes location on brain maps, extraction of frequency bands, representation of DE and PSD brain maps for each bands and representation of stacked 2D brain maps

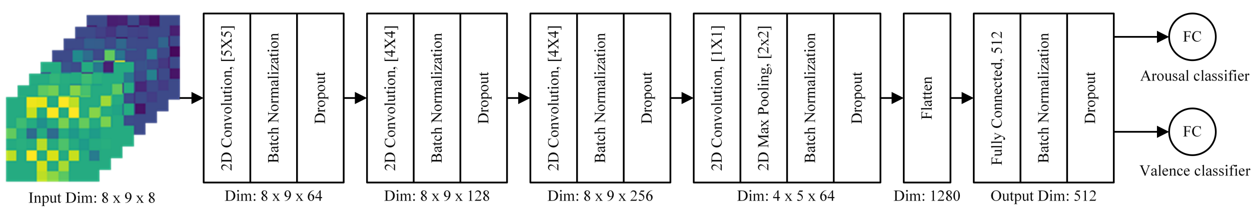

Fig. 2. Detail of layers composing the Multi-task CNN with input dimension, tensor dimension after each block and output dimension before classification

by removing RNN layers. Figure 3 presents the impact of each modification on the performances of the model.

A comparison of the proposed method with others is presented below. Then the proposed model will be discussed.

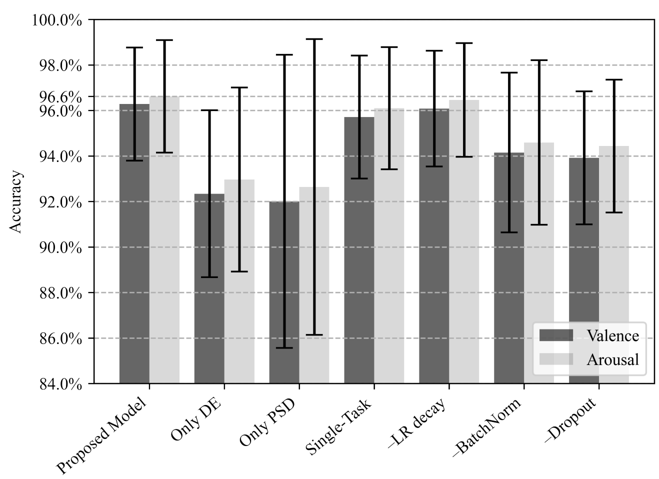

Fig. 3. Ablation studies on different input features, regularization methods and the use of the multi-task learning principle. "–" denotes the ablation on specific regularization methods.

A. Comparison with different methods

We compare the proposed model with the existing models for emotion recognition using EEG signals or brain maps. The comparison is made using only DEAP datasets to have a meaningful comparison. In Table II the comparison of accuracy in valence and arousal for each method is presented. The accuracy of the proposed model is 96.28% in valence, which beats all the compared methods by 6.27% the Multi-column CNN, 15.18% the SAE-LSTM, 2.06% the 4D-CRNN and 6.02% the FBCCNN in absolute values. In arousal, an accuracy of 96.62% is obtain, which outperforms compared methods by 5.97%, 22.24%, 2.04% and 7.72% in absolute values.

TABLE II
PERFORMANCES COMPARISON OF PROPOSED METHOD AND OTHER METHODS

Nos.	Method	DEAP-valence	DEAP-arousal
1	Multi-column CNN [11]	90.01%	90.65%
2	SAE-LSTM [12]	81.10%	74.38%
3	4D-CRNN [8]	94.22%	94.58%
4	FBCCNN [13]	90.26%	88.90%
5	MT-CNN (Proposed method)	96.28%	96.62%

B. Discussion

A new approach for emotion recognition using a brain map is presented in this paper. The comparison analysis presented above shows that our proposed method obtains better performances than other methods.

Firstly, the position of the electrodes contains important information for the analysis, as was done in [8]. The combination of the spatial features with others can be defined as brain maps. This way, passing brain maps in a model can get better results than just using EEG signals directly.

The model 4D-RCNN is also based on brain maps but got worse results than our model, 94.22% for valence and 94.58% for the arousal compared to 96.28% and 96.62%, respectively for the Multi-Task CNN. It can be explained by adding a feature, PSD, to DE instead of using only DE and exploiting regularizations during the training.

Generally, models for emotion recognition using EEG signals or brain maps use specific-subject methods. The proposed method shows that based on the whole dataset, all subjects at a time, using a simple structure, a CNN which performs well and fast on images, we get better results than other methods proposed from the literature.

Finally, in this paper, the model is a single for both valence and arousal classification tasks. The training of the model in multi-task manner allows to have only one model for prediction in valence and arousal instead of two ones, as it was in other methods.

IV. Conclusion

The main purpose of the article was the creation of the model for emotion recognition using brain maps. As said in the previous section, the model is based on different features extracted from EEG signals such as DE and PSD. Table II shows us that this method provides us better results than the previous methods. This method shows results on recorded high-quality EEG signals without considering EEG signals recorded in real-time with less precise devices. It gives results in valence and arousal not as an emotion, however, still, model improvement can be considered.

Indeed, we see that adding brain maps created from PSD features give us better results than just using brain maps created from DE features. So, we could add new features in addition to the current features, such as standard deviation. Another approach would be to train and test the model with a lot more data than we are using right now. The cross-dataset approach would permit to generalize the model to real-world problems. Last improvement which could increase drastically the performances for the recognition of the emotion would be using a multi-modal approach. For example, extracting videos of the face of the individuals at the same time than extracting features with EEG signals would allow combine different models and give us an efficient multi-modal approach for emotion recognition.

References

[1] Prasanth, S., M. Roshni Thanka, E. Bijolin Edwin, et V. Nagaraj. Speech Emotion Recognition Based on Machine Learning Tactics and Algorithms . Materials Today: Proceedings, février 2021, S2214785320398928. https://doi.org/10.1016/j.matpr.2020.12.207.

[2] Revina, I.Michael, et W.R. Sam Emmanuel. A Survey on Human Face Expression Recognition Techniques . Journal of King Saud University - Computer and Information Sciences 33, no 6 (juillet 2021): 619-28. https://doi.org/10.1016/j.jksuci.2018.09.002.

[3] Chen, Xiaoying, Weiming Zeng, Yuhu Shi, Jin Deng, et Yuan Ma. Intrinsic Prior Knowledge Driven CICA fMRI Data Analysis for Emotion Recognition Classification . IEEE Access 7 (2019): 59944-50. https://doi.org/10.1109/ACCESS.2019.2915291.

[4] Suhaimi, Nazmi Sofian, James Mountstephens, et Jason Teo. EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities . Computational Intelligence and Neuroscience 2020 (16 septembre 2020): 1-19. https://doi.org/10.1155/2020/8875426.

[5] Dadebayev, Didar, Wei Wei Goh, et Ee Xion Tan. EEG-Based Emotion Recognition: Review of Commercial EEG Devices and Machine Learning Techniques . Journal of King Saud University - Computer and Information Sciences, avril 2021, S1319157821000732. https://doi.org/10.1016/j.jksuci.2021.03.009.

[6] Topic, Ante, et Mladen Russo. Emotion Recognition Based on EEG Feature Maps through Deep Learning Network . Engineering Science and Technology, an International Journal 24, no 6 (décembre 2021): 1442-54. https://doi.org/10.1016/j.jestch.2021.03.012.

[7] Koelstra, S., C. Muhl, M. Soleymani, Jong-Seok Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, et I. Patras. DEAP: A Database for Emotion Analysis;Using Physiological Signals . IEEE Transactions on Affective Computing 3, no 1 (janvier 2012): 18-31. https://doi.org/10.1109/T-AFFC.2011.15.

[8] Shen, Fangyao, Guojun Dai, Guang Lin, Jianhai Zhang, Wanzeng Kong, et Hong Zeng. EEG-Based Emotion Recognition Using 4D Convolutional Recurrent Neural Network . Cognitive Neurodynamics 14, no 6 (décembre 2020): 815-28. https://doi.org/10.1007/s11571-020-09634-1.

[9] Yang, Yilong, Qingfeng Wu, Ming Qiu, Yingdong Wang, et Xiaowei Chen. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network . In 2018 International Joint Conference on Neural Networks (IJCNN), 1-7. Rio de Janeiro: IEEE, 2018. https://doi.org/10.1109/IJCNN.2018.8489331.

[10] Sarma, Parthana, et Shovan Barma. Emotion Recognition by Distinguishing Appropriate EEG Segments Based on Random Matrix Theory . Biomedical Signal Processing and Control 70 (septembre 2021): 102991. https://doi.org/10.1016/j.bspc.2021.102991.

[11] Yang, Heekyung, Jongdae Han, et Kyungha Min. A Multi-Column CNN Model for Emotion Recognition from EEG Signals . Sensors 19, no 21 (31 octobre 2019): 4736. https://doi.org/10.3390/s19214736.

[12] Xing, Xiaofen, Zhenqi Li, Tianyuan Xu, Lin Shu, Bin Hu, et Xiangmin Xu. SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG . Frontiers in Neurorobotics 13 (12 juin 2019): 37. https://doi.org/10.3389/fnbot.2019.00037.

[13] Pan, Bo, et Wei Zheng. Emotion Recognition Based on EEG Using Generative Adversarial Nets and Convolutional Neural Network . Édité par Raul Alcaraz. Computational and Mathematical Methods in Medicine 2021 (11 octobre 2021): 1-11. https://doi.org/10.1155/2021/2520394.

[14] Jia, Ziyu, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, et Jing Wang. SST-EmotionNet: Spatial-Spectral-Temporal Based Attention 3D Dense Network for EEG Emotion Recognition . In Proceedings of the 28th ACM International Conference on Multimedia, 2909-17. Seattle WA USA: ACM, 2020. https://doi.org/10.1145/3394171.3413724.

[15] Caruana, Rich. Multitask Learning . Machine Learning 28, no 1 (1997): 41-75. https://doi.org/10.1023/A:1007379606734.