Brain Dynamic States Analysis based 3D Convolutional Neural Network

Yu-Chia Hung, Yu-Kai Wang, Mukesh Prasad, Chin-Teng Lin
Centre for Artificial Intelligence, FEIT, University of Technology Sydney, Australia

Abstract—Drowsiness driving is one major factor of traffic accident. Monitoring the changes of brain signals provides an effective and direct way for drowsiness detection. One 3D convolutional neural network (3D CNN)-based forecasting system has been proposed to monitor electroencephalography (EEG) signals and predict fatigue level during driving. The limited weight sharing and channel-wise convolution were both applied to extract the significant phenomenon in various frequency bands of brain signals and the spatial information of EEG channel location, respectively. The proposed 3D CNN with limited weight sharing and channel-wise convolution has been demonstrated to predict reaction time (RT) of driving with low root mean square error (RMSE) through the brain dynamics. This proposed approach outperforms with the state-of-the-art algorithms, such as traditional CNN, Neural Network (NN), and support vector regression (SVR). Compared with traditional CNN and Artificial Neural Network, the RMSE of 3D CNN-based RT prediction has been improved 9.5% (RMSE from 0.6322 to 0.5720) and 8% (RMSE from 0.6217 to 0.5720), respectively. We envision that this study might open a new branch between deep learning application in neuro-cognitive analysis and real world application.

Keywords — driving safety, drowsiness, deep learning, convolutional neural network, Electroencephalography.

I. Introduction

Driving safety becomes one of major concern in our daily life. However, drowsiness driving is a major factor of traffic accident due to the increasing tiredness and stress level of the drivers. Some behavior-based technologies have been developed to predict fatigue level of drivers. In particular, monitoring brain signals provides an effective way to predict the changes of fatigue during driving [1] [2].Recently, many researchers leverage the deep learning algorithm to solve the image recognition and signal processing [5] [8]. Many researchers have been used deep learning methods to analyze electroencephalography (EEG) signals [1] [3] [4] and shown the improved results compared with the traditional approaches. Convolutional neural network (CNN) has very strong feature to extract spatial information in different areas, such as image recognition [5]. Although CNN is a powerful algorithm for spatial information capturing, spatial and temporal information of brain dynamics cannot be captured well [6] [7]. In video analysis, the module is designed to extract the feature between adjacent frames, since the frames of a video are a trend demonstrates a continuous movement. EEG signals have the similar characteristic because it is continuous, which brain state varies dynamically over time, and the brain states are related closely, especially in adjacent time frames.

Recently, deep learning is used into EEG signals processing [1] [3] [4]. CNN is one of the powerful deep learning method for analysing spatial information because of its convolutional data extraction in spatial domain. Each convolutional layer extracts 2D input feature maps, and only the spatial information is considered. Compared to traditional CNN, 3D CNN has ability to analyse temporal and spatial information based on 3D filters which extract both temporal and spatial information. In other words, 3D CNN extracts both temporal and spatial information by feeding 3D input data. Therefore, 3D CNN has been widely used in analysing the data which includes spatial and temporal data, such as video [6] [7].The results in the previous studies showed that 3D CNN can reach the higher performance in video recognition [6] [7]. The characteristics including continuous time signal in EEG signals (temporal information) are similar with that in video. In particular, the spatial information in every single time window (frame) is also one important characteristic for drowsiness detection [9]. Therefore, applying 3D CNN to multi-frame EEG signals is one potential way for drowsiness prediction. Furthermore, channel-wise convolution and limited weight sharing mechanisms were both applied to enhance the spatial relationships of EEG signals and capture the phenomenon from all EEG frequency bins, respectively [8]. The extracted features by these two mechanisms were then the inputs of 3D CNN.

In this study, we demonstrated a 3D CNN based monitoring system to predict reaction time (RT) during driving. The novel channel-wise convolution and limited weight sharing mechanisms also significantly improve the overall performance of prediction. Compared with traditional CNN and other machine learning algorithms including neural network (NN) and support vector regression (SVR), the proposed 3D CNN can reach the minimum root mean square error (RMSE).

II. Experimental Setup

A. Data pre-proccessing

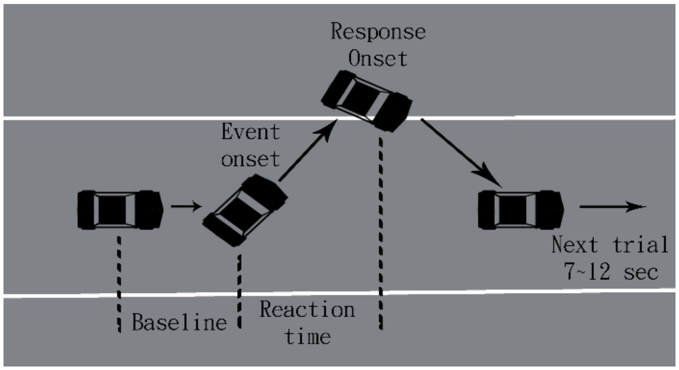

Figure 1: We use the signal before event onset 6 second as baseline signal, and the time between event onset and response onset as reaction time.

Reaction time (RT) is the duration between the onset of deviation and the time that participants started to drive the vehicle back to the original lane. To evaluate driver's brain states, RT represents driver's drowsiness and alertness during driving (Figure 1). Divers might be drowsy while the RT was high. In contrast, drivers might be alert while the RT was short.The brain dynamics in the baseline was extracted (as shown in Figure 1), and 6s data was extracted. Eighty subjects involved in this drowsiness driving. There were nine subjects abandoned because some EEG channels of these subjects were broken. Additionally, some trials with noise or high RT (> 10s) were removed manually. There are total 24,203 trials from 71 subjected. For normalization, all RTs was divided by the baseline (the shortest 10% RT). In particular, the maximum normalized RTs were set to three.

B. *Experiment*

The experiment dataset was collected by a 360-degree virtual reality environment with motion platform, which simulates a driving environment in reality. There were 80 healthy subjects without any history of psychological or sleep disorders. All the subjects were collected with a 32-channel EEG equipment, which includes 30 channels and 2 reference channels. The impedances of all EEG channels were all less than 5 kΩ and the sampling rate of EEG signals were 250 Hz. We measured human's brain states during driving by land-keeping task, which can measure driver's drowsiness and alertness. In lane-keeping task, the participants were asked to drive vehicles along a lane, however, the vehicle were designed to deviate forward left or right lanes from their original lane automatically. When the participants detected the vehicles deviated from the lane, they were asked to drive the car back to the original lane through turning the steering wheel. In this study, one trial was defined as from the onset of deviation to the offset of turning the steering wheel (back to the original lane). There was a random 7-12s break between two continuous trials (as shown in Figure 1). The whole experiment continued 90 minutes, and there were approximately 400 trials.

III. METHOD

A. *Algorithm*

In this research, 3D convolution neural network was majorly applied to analyse the recorded brain activities during driving. To extract the spatial features and frequency features from EEG signals efficiently, two algorithms including channel-wise convolution and limited weight sharing were both applied to reach higher performance.

Firstly, the structure of 3D convolutional layers and traditional (2D) convolutional layers are listed in Figure 2. Briefly, traditional CNN equips 2D filters, comprised of 2D weights, to convolute 2D input feature maps or input data at 2 axis (usually spatial axis), and produce 2D output feature maps too. The feature maps also are compressed by 2D pooling layers. Finally, the feature maps are flattened to 1D vectors and connected to fully connected layers. In 3D CNN, input data and input feature maps are 3D, and the filters are 3D cubes comprised of N X N X N weights, which can convolute input data or input feature maps at 3 axis (usually spatial and temporal axis). Following, each output feature maps of 3D convolutional layers are compressed by 3D pooling layers. Finally, the 3D feature maps are flattened to 1D vectors and connected to fully connected layers.

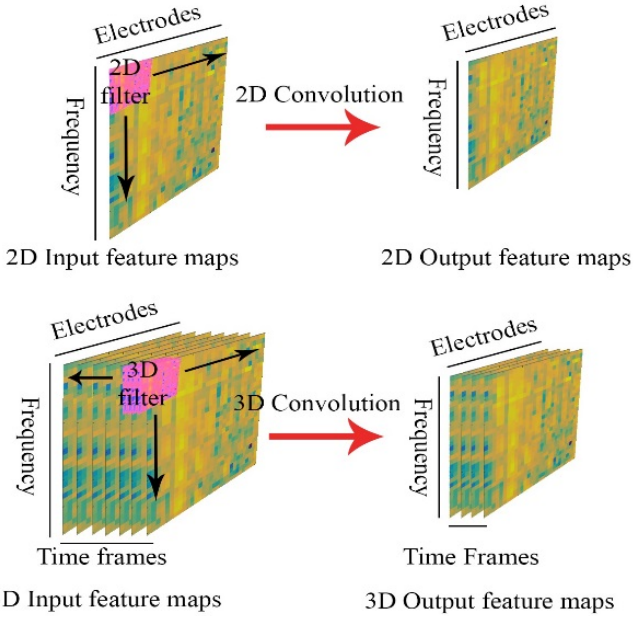

Figure 2: 3D convolutional layer and 2D convolutional layer, 3D convolutional layer extract 3D input feature maps with 3D filter, 2D convolutional layer extract 2D input feature maps with 2D filter

In terms of the function of 3D CNN,

$$v_{ij}^{xyz} = relu\left(b_{ij} + \sum_{m} \sum_{p=0}^{P_i - 1} \sum_{q=0}^{Q_i - 1} \sum_{r=0}^{R_i - 1} w_{i_{ijm}}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} \right)$$

which v_{ij}^{xyz} is the unit at x, y and z position at feature map between i^{th} and j^{th} layers, b_{ij} is the bias between the two layers, m is the m^{th} filter in convolutional layer, P, Q and R are the kernel size of filters, $w_{i_{ijm}}^{pqr}$ is the weight at pqr position of m^{th} filter, and *relu* is an activation function.

B. *Architecture*

The 3D conv layers extract 3D input feature maps at first and the max-pooling layers compress input feature maps to reduce data's dimensions. After three convolutional and max-pooling layers, the output feature maps are flattened and feed into a 512-node fully-connected layer, which followed by an output layer. In every convolutional layer, we adopted "padding", which pad zeros at input feature maps to make sure the input size and output size are equal. All configurations for 3D CNN are listed in Table 1.The dimension of input layer is 18 (frequency power,1-18 Hz) X 30 (EEG channel numbers) X 26 (6s frames) X one input feature map. We adopted a 3 (EEG channel) X 3 (frequency) X 3 (time frame) filter in all three convolutional layers, in which there were 32, 64 and 128 filters, respectively. We also adopted one 2 (EEG channel) X 2 (frequency) X 2 (time frame) max-pooling in all three pooling layers. Therefore, the structure of the proposed 3D CNN included Input-layer (18-30-26-1) - Conv1 (3-3-3-32) - Maxpool1 (2-2-2) - Conv2 (3-3-3-64) - Maxpool2 (2-2-2) -

Conv3 (3-3-3-64) - Maxpool1 (2-2-2) - Conv3 (3-3-3-128) - Maxpool1 (2-2-2) - Fully Connected Layer (512) – Output-layer (1).

Table 1 configuration of 3D CNN and 2D CNN

Layer	3D CNN	CNN
Input layer	18-30-26-1	18-30-1
Conv1	3-3-3-32	3-3-32
Maxpool1	2-2-2	2-2
Conv2	3-3-3-64	3-3-64
Maxpool2	2-2-2	2-2
Conv3	3-3-3-128	3-3-128
Maxpool3	2-2-2	2-2
Fully connected	512	512
Output layer	1	1

We then adopted one 3 (EEG channel) X 3 (frequency) filter in all three convolutional layers, in which there were 32, 64 and 128 filters, respectively. We also adopted 2 (EEG channel) X 2 (frequency) max-pooling in all three pooling layers. The input data of CNN are 2D and the conv layers and max-pooling layers are also two dimension. The structure of CNN is Input-layer (18-30-1) - Conv1 (3-3-32) - Maxpool1 (2-2) - Conv2 (3-3-64) - Maxpool2 (2-2) - Conv3 (3-3-128) – Maxpool3 (2-2) - Fully Connected Layer (512) – Output-layer (1). The configuration of neural network is listed below. Input-layer (18-30-1) - Fully Connected Layer (512) – Output-layer (1).

Furthermore, we proposed a novel approach, channel-wise convolution. Compared to other dataset, images and videos analysed with CNN modules, EEG data signals have some unique attributes should be considered into CNN modules, for example EEG signals demonstrate different phenomenon according to the position of EEG channels. Therefore, the position of EEG channels should be considered into the CNN module to extract the full spatial information.

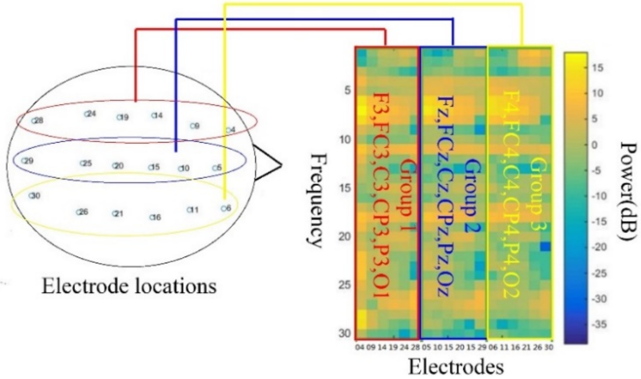

Figure 3: Channel-wise convolution. The EEG channels are divided into three groups and sorted from the frontal region to posterior region. After grouping the EEG channels by spatial location (channel location), the designed filters only convolute the information in each group.

There are high spatial relationships among the EEG signals from the adjacent EEG channels. Therefore, special filters are able to capture the information from the different EEG channels. A channel-wised convolution was applied to extract the spatial information according to the channel location of EEG equipment as shown in Figure 3. Firstly, the EEG channels are divided into three groups (as shown in Figure 3) according to their physical locations. Furthermore, the EEG channels in each group were sorted from the front of a scalp to the behind of a scalp sequentially. After doing that, a channel-wise convolutional layer scan frequency power with adjacent EEG channels together, group by group. In other words, those groups share the same filter, however, the filters do not convolute the frequency power across different groups. Therefore, the filters can extract the information between each EEG channel according to their physical locations. Moreover, the reason why we just used 18 EEG channels in channel-wise convolution here, instead of 30, is that these three groups of EEG channels have more regular phenomenon in power spectrum here, compared to the rest of the EEG channels.

In addition, we adopt a mechanism – limited weight sharing which is widely applied in deep learning methods from speech processing. Since the phenomenon of frequency power between different frequency bands are related to distinct cognitive functions, the weight of each band is supposed to be shared in the specific bands. The extracted features should also be trained individually. It means there is a specific filter for each frequency band to extract the features form the frequency information. The idea is from speech processing [1]. In speech processing, the bandwidth is quite wide and the attributes in different frequency bands are different. Therefore, each frequency band were trained individually to extract meaningful features. Since there are the same attributes between EEG signals and speech information, the limited weight sharing was also applied for EEG processing in the current study. There are frequencies bands (Delta band: 1-4 Hz, Theta band: 4-7 Hz, Alpha band: 8-15 Hz, Beta band: 16-30 Hz) in EEG analysis. Therefore, we adopt limited weight sharing in these four bands in the first convolutional layer as shown in Figure 4. In other words, the weights (filters) of one specific band are only shared inside this band. After applying the limited weight sharing, the band-based filters can majorly extract the features in each frequency band.

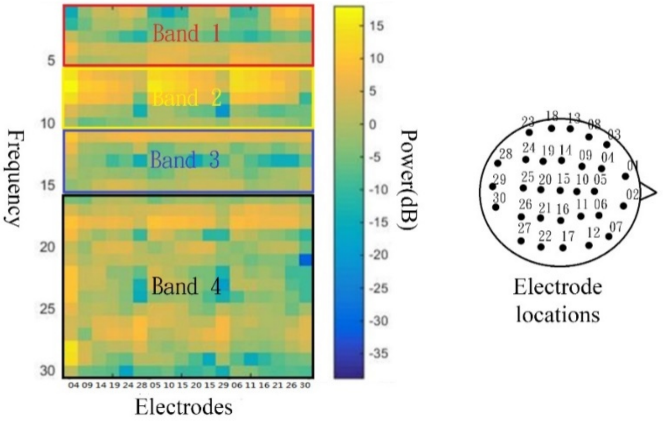

Figure 4: Limited Weight sharing in first convolutional layer.The color in power spectrum represents power (dB) in each frequency bin. The frequencies are divided into 4 bands. The weights (filters) are only shared inside a single band. In other words, the filters are trained individually inside their band without overlapping the other bands.

In this study, the channel-wise convolution and the limited weight sharing were applied in the conv1 layer and all three conv layers, respectively. In particular, these two mechanisms were both applied in 3D CNN and traditional CNN. In the limited weight sharing mechanism, the frequency in conv1 layer were grouped into four bands (band 1:1-4 Hz, band 2: 5-7 Hz, band 3: 8-15 Hz, band 4: 16-30 Hz), and three bands in conv2 layer (band 1: 1-5 Hz, band 2: 6-10 Hz, band 3: 11-15 Hz) and two bands (band 1: 1-5 Hz, band 2: 6-10 Hz) in conv3 layer. Similarly, the channel-wise convolution mechanism was applied at conv1 layer, which could avoid the over restricting convolution in single EEG channel group (as shown in Figure 4).

IV. Results and Discussion

Comparing with the result (as shown in Table 2) of 3D CNN and CNN, the average RMSE is improved 9.5% (from 0.6322 to 0.5720).The reason is 3D CNN is able to extract the temporal information by convoluting multi frames of EEG signals. Comparing with the results of 3D CNN, NN and SVR, 3D CNN still achieves the better performance. Compared with the results of SVR, the average and standard deviation RMSE of 3D CNN are improving 4.2% (from 0.5973 to 0.5720) and 10% (from 0.1636 to 0.1488), respectively.

Table 2: Results of 3D CNN, CNN, NN and SVR

RMSE	3D CNN	CNN	NN	SVR
Mean	0.5720	0.6322	0.6217	0.5973
STD	0.1488	0.1752	0.1428	0.1636

The performance of 3D CNN with channel-wise convolution and limited weight sharing is shown in Table 3. These two mechanisms both achieve higher performance. In this study, we adopt channel-wise convolution and limited weight sharing at conv1 layer and all three conv layers, respectively. These two mechanism can decrease the average RMSE by 1.9% (from 0.5834 to 0.5720) as comparing with the performance of 3D CNN. According to our results, these two mechanisms should be applied simultaneously to achieve the best performance. Otherwise, the performance is worse than that trained by original 3D CNN.

Table 3: Results of 3D CNN with & without channel-wise convolution and limited weight sharing

3D CNN	With limited weight sharing	Without limited weight sharing
With channel-wise convolution	**0.5720 ±0.1488**	0.5924±0.1464
Without channel-wise convolution	0.5909±0.1441	0.5834 ±0.1449

We also set two different structures of 3D CNN as shown in Table 4. The configurations of seven-layer 3D CNN are the same as those of three-layer 3D CNN. The structure of three-layer 3D CNN is Input-layer (18-30-26-1) - Conv1 (3-3-3-32) - Maxpool1 (2-2-2) -) - Maxpool1 (2-2-2) - Fully Connected Layer (512) – Output-layer (1). In particular, the channel-wise convolution and limited weight sharing at conv1 layer in three-layer 3D CNN. The results show that seven-layer 3D CNN reaches better performance than three-layer 3D CNN by decreasing average RMSE 7.7% (from 0.5729 to 0.6210).

Table 4: Results of different structures of 3D CNN

	seven-layer 3D CNN	three-layer 3D CNN
RMSE	0.5720±0.1488	0.6210±0.1382

V. Conclusions

The frequency information of EEG signal is continuous data. Therefore, traditional CNN did not get a good result in this study compared to the other algorithms since traditional CNN might not be able to capture brain dynamic state. In this study, 3D CNN has ability to extract multi-frame data, which content temporal information among different time frame, and the extracted features from various time frame are essential for analyzing the brain signals. The phenomenon in different EEG frequency bands are supposed to be convoluted with different filters by limited weight sharing. Furthermore, channel-wise convolution can extract more spatial information of EEG signals, compared to the normal convolution. In conclusion, 3D CNN actually captures brain dynamic state better than traditional CNN by decreasing RMSE from 0.6322 to 0.5834. Moreover, 3D CNN adopted channel-wise convolution and limited weight sharing together can also reach better performance in EEG signals, compared to original 3D CNN, by decreasing RMSE from 0.5834 to 0.5720. Based on the completive performance, we envision that deep learning might open a new branch between translation neuroscience and real world application.

References

[1] Mehdi Hajinoroozi, Zijing Mao, Yufei Huang, “Prediction of Driver’s Drowsy and Alert States From EEG Signals with Deep Learning” *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 13-16 Dec. 2015, Montpellier, France.

[2] Yu-Ting Liu, Yang-Yin Lin, Shang-Lin Wu, Tsung-Yu Hsieh, Chin-Teng Lin, “Assessment of Mental Fatigue: An EEG-based Forecasting System for Driving Safety”, *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 9-12 Oct. 2015, Kowloon, China.

[3] Yuanfang Ren and Yan Wu, “Convolutional Deep Belief Networks for Feature Extraction of EEG Signal” *2014 International Joint Conference on Neural Networks (IJCNN)*, July 6-11, 2014, Beijing, China.

[4] Suwicha Jirayucharoensak, Setha Pan-Ngum, and Pasin Israsena, “EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation”, *The Scientific World Journal*, 1 September 2014.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton., “ImageNet Classification with Deep Convolutional Neural Networks”, *2012 Advances in neural information processing systems(NIPS),* pp.1097-1105, 2012

[6] Shuiwang Ji, Wei Xu, Ming Yang, “3D Convolutional Neural Networks for Human Action Recognition”, 2013 *IEEE Trsaction on pattern analysis and maching intellengence*, 06 March 2012.

[7] Andrej Karpathy, George Toderici1, Sanketh Shetty, “Large-scale Video Classification with Convolutional Neural Networks”,*2014 Computer Vision and Pattern Recognition (CVPR)*, 23-28 June 2014, Columbus, OH, USA.

[8] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Li Deng, “Convolutional Neural Networks for Speech Recognition”, *IEEE/ACM Trsaction on audio, speech, and language processing*., vol. 22, no. 10, pp.1533-1545, October 2014.

[9] Kuan-Chih Huang, Teng-Yi Huang, Chun-Hsiang Chuang, Jung-Tai King, Yu-Kai Wang, Chin-Teng Lin, and Tzyy-Ping Jung, “An EEG-Based Fatigue Detection and Mitigation System,” *International journal of neural systems*, vol. 28, no. 4, 2016.